An Analytical Approach to Modelling Epidemics on Networks.

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Applied Mathematics

at Massey University, Albany,
New Zealand.

Karen McCulloch

December 21, 2016

Abstract

A significant amount of effort has been directed at understanding how the structure of a contact network can impact the spread of an infection through a population. This thesis is focused on obtaining tractable analytic results to aid our understanding of how infections spread through contact networks and to contribute to the existing body of research that is aimed at determining exact epidemic results on finite networks. We use SIR (Susceptible-Infected-Recovered) and SIS (Susceptible-Infected-Susceptible) models to investigate the impact network topology has on the spread of an infection through a population.

For an SIR model, the probability mass functions of the final epidemic size are derived for eight small networks of different topological structure. Results from the small networks are used to illustrate how it is possible to describe how an infection spreads through a larger network, namely a line of triangles network. The key here is to correctly decompose the larger network into an appropriate assemblage of small networks so that the results are exact.

We use Markov Chain theory to derive results for an SIS model on eight small networks such as the expected time to absorption, the expected number of times each individual is infected and the cumulative incidence of the epidemic. An algorithm to derive the transition matrix for any small network structure is presented, from which, in theory, all other results for the SIS model can be obtained using Markov Chain theory. In theory, this algorithm is applicable to networks of any size, however in practice it is too computationally intensive to be practical for larger networks than those presented in this thesis.

We give examples for both types of model and illustrate how to parameterise the small networks to investigate the spread of influenza, measles, rabies and chlamydia through a small community or population.

Acknowledgements

I would like to thank my supervisor, Professor Mick Roberts; for his guidance, support and patience throughout this project; without which this research would not have been possible.

I would also like to thank my co-supervisor, Associate Professor Carlo Laing, for his guidance, support and patience throughout this project.

Thirdly, I would like to thank all my colleagues whom I have met and had inspirational discussions with over the past few years. In particular, I would like to thank Dr Roslyn Hickson for her help with the Gillespie algorithm code.

Lastly, I would like to thank my partner, family and friends who have supported me throughout this endeavour; for that I am very grateful.

A paper based on the work presented in Chapter 2 and Appendix A of this thesis has been published in the following.
K. McCulloch, M.G. Roberts and C. R. Laing, Exact analytical expressions for the final epidemic size of an SIR model on small networks, *ANZIAM Journal*, 2016, 57 (4), p429-444.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Epidemics and global pandemics have been occurring throughout history and will continue to do so as new strains of viruses emerge, such as the ever changing influenza virus. Some infectious diseases present today have no known cure, yet others are preventable with vaccines (such as measles and rabies) but they still persist due to various reasons such as lack of education, cost and availability of vaccines and often a lack of willingness to be vaccinated. While there are infectious diseases causing morbidity and mortality around the world, there will be research focused on understanding the biology of each pathogen and how it spreads between hosts. Other research is aimed at understanding why the same infection can exhibit very different behaviour in separate populations. One common goal of this type of research is to gain further insight into the infection dynamics, with the ultimate goal for many being a successful control campaign that allows elimination of an infectious disease.

Arguably, the most common types of model used to describe the spread of an infection through a population are compartmental models. These models divide the individuals within a population into classes based on infection status. Such models include the SIR (Susceptible - Infectious - Recovered) model where Susceptible individuals are susceptible to infection; Infectious individuals are infected and can transmit the infection to others and Recovered (sometimes referred to as Removed) individuals are those who are either immune to the infection (for example, due to vaccination) or who have been removed from the population due to death and no longer take part in the infection life cycle. Influenza and measles are two examples of infections that are well known to follow an SIR model for the host infection lifetime [34, 41, 59]. Other types of compartmental models include the SIS (Susceptible - Infectious - Susceptible) and $SEIVR$ (Susceptible - Exposed - Infectious - Vaccinated - Recovered) models. In the SIS model, upon contact with an infectious individual, susceptible individuals move into the infectious class. When an infectious individual recovers they move back into the susceptible class. This makes multiple infections possible for each individual, which adds some complexity when analysing the infection dynamics. Typically, SIS models are used to investigate the spread of bacterial infections or sexually transmitted infections (such as chlamydia) that have no long-lasting immunity.

There are a number of different methods that have been used to model the spread of an infection through populations. Differential equations are the most commonly used type of model for describing the spread of infection within a population (see, for example, [16, 34]). A set of ordinary differential equations (ODEs) can be used to describe how a population of individuals moves between classes, such as in an SIR or SIS model. These types of models are useful for gaining an understanding of the infection dynamics. One of the most well known models for the spread of an infection was published in 1927 by Kermack and McKendrick [36]. The SIR model is often referred to as 'the Kermack-McKendrick model', however it is actually a special case of the Kermack-McKendrick model.

Traditional models for the spread of an infection assume a fully mixed (homogeneous) population where an infectious individual has a fixed rate, β, of transmitting infection to any other susceptible individual within the population. For the SIR model without demography, the infectious individuals also have a fixed recovery rate, γ, per unit time. These simplifying assumptions allow us to model the transmission of infection through a closed population by the well known set of ordinary differential equations [16, 34, 36, 67]:

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

In this model, S, I and R represent the *proportion* of the population in each class. There are many variations that can be made to this basic model such as incorporating births and deaths, including differences in susceptibility and infectivity, or variations in infectious period which would all create a more detailed model but also add to the complexity of it. In a closed population, such as in the model above, the infection in an SIR model will always eventually go extinct, whereas in the SIS model as the individuals do not recover but become susceptible again it is possible for the infection to become endemic within the population. If births and deaths are included in the SIR model an endemic infection is possible as the births provide a source to replenish the susceptible population.

Differential equation models allow us to derive various analytical results such as expressions for the basic reproduction number, R_0, and critical vaccination threshold in terms of key biological parameters such as the rate of infection and infectious period [62]. The basic reproduction number, R_0, is defined as the expected number of secondary cases that arise from a single primary case in a wholly susceptible population [16]. For $R_0 < 1$ the probability that a major epidemic will occur is low, but it is still possible, and for $R_0 > 1$ the probability that a major epidemic will occur is high but it is not guaranteed. For these reasons the value that R_0 takes is of significant interest in infectious disease modelling. However, differential equation models are usually based on some strong simplifying assumptions about the behaviour of individuals and the biology of the disease [62].

Another method for using differential equations to model the spread of an infection through a population is the use of Markov Chains. In Markov Chain methodology, the future state of the population depends only on the current state. It is a more rigorous modelling approach as discussed by Keeling and Ross [35], see also [62]. In this model the probability of the population being in a particular state at a given point in time is described by a system of linear differential equations that are derived from a transition matrix. The transition matrix contains the probability of transition between states.

There are many aspects that should be investigated when trying to understand the dynamics of an infection spreading through a population (animals or humans), one of which is the way individuals interact with each other. Contact networks[1] are often used to represent how individuals within a population are connected to each other. In this thesis we use a mathematical approach to investigate the dynamics of SIR and SIS epidemic models on small networks with different topologies, as a stepping stone to determining how the structure of a contact network impacts the transmission of infection through a population.

1.1 Network models for the spread of infections

The use of contact networks to model the spread of an infection through a population has become increasingly popular in recent years [8, 14, 40, 45, 54]. It is well known that models for the spread of an infection that assume the population to be homogeneously mixed are only appropriate for some situations. Such models often overestimate the proportion of the population that becomes infected during an epidemic; this overestimation occurs when there is heterogeneity in the number of contacts each individual has within the population.

Pair approximation models [28, 32] improve on the accuracy of the homogeneous

---

[1] A network is a collection of nodes, connected by edges if two nodes have contact with each other.

models by focusing on the interaction between pairs of individuals and assume that a change in the state of an individual depends on the state of its neighbours. However, contact networks provide more detailed model populations; connections between individuals can be permanent (static networks) or the number and duration of connections each individual has may change over time (dynamic networks). In a contact network each node represents an individual (or group of individuals) and an edge between two nodes represents a connection between individuals. A connection is defined depending on the context, e.g. an individual within a shared office would have connections between all other individuals within the same office for the transmission of influenza, or a connection could represent an individual's partners in a sexual contact network for HIV.

There has been considerable effort put into understanding how an infection spreads through different network structures such as scale-free, small-world, lattice and random networks [9, 14, 33, 40, 53, 65, 66, 70]. Ongoing research efforts are aimed at finding the right balance between capturing the population structure in enough detail to provide more accurate predictions, while keeping the mathematics tractable so that the derivation of analytical results is possible. There have been a number of different models developed with the aim of addressing this issue. There is often a trade off between model complexity and mathematical tractability, thus most of the models developed so far are tailored to specific assumptions about the population structure or infection dynamics. For example, consider the effective degree network models, where nodes are classified by their infection state and by the infection state of their neighbours. These models produce accurate predictions for SIR and SIS type models on large random static [5, 20, 39] and dynamic [42, 63] networks without clustering. Analytic expressions for the basic reproduction number, R_0 , were found for effective degree models on static networks and approximations to the final proportion of the population infected were found to be in agreement with stochastic simulations.

Other research has been focused on finding results for the spread of infection through networks by breaking up the networks into smaller specific motifs [37, 55, 62]. These models demonstrate the importance of understanding higher-order structure (such as larger network motifs) by comparing the spread of an infection between networks with the same degree distribution and clustering coefficient. Because these methods are mathematically difficult, results are obtained via numerical simulation.

Edge-based compartmental models (EBCM) [44, 46, 47, 65] begin to incorporate model complexity while keeping the mathematical concept and derivation tractable. EBCMs are based on the idea of determining the probability that a given node's neighbour is susceptible, infectious or recovered at time t. These models incorporate social heterogeneity and variation in partnership duration by investigating an SIR model on static and dynamic networks. EBCMs accurately predict the time evolution of the SIR model and provide a means for finding an analytic expression for the expected final epidemic size (in most cases) and the basic reproduction number, R_0.

Research that investigates the behaviour of probabilistic epidemic processes in the limit of large networks is another area which has gained much attention and many of the published works include detailed mathematical proofs, which while elegant can be conceptually challenging. Volz [65] derived deterministic equations by increasing the size of the network, they stated, but did not prove, the convergence of the continuous-time stochastic SIR model to its deterministic limit for large graphs. The mathematical proof of this is given in [15] by providing rigorous individual based description of the epidemic on a random graph. Janson et al [29] described a Markovian SIR model on a random network with constant infection and recovery rates and showed that conditional on a large outbreak occurring, the evolution of quantities such as the fraction of infective nodes converges to the deterministic functions of time. In [7] a stochastic SIR model on a random network of social contacts is derived, where the population is also partitioned into households. In this type of model infectious contacts are made at a higher frequency within households than those made between households. The behaviour of the model as the population size tends to infinity was investigated and results such as the probability a major outbreak would occur and the final epidemic size were derived. Monte Carlo simulations demonstrate that these asymptotic quantities accurately reflect the behaviour of finite populations, even for only moderately sized finite populations.

Branching processes have also been used to describe a stochastic SIR epidemic model with two levels of mixing in order to move away from the homogeneous mixing assumption and to incorporate some realism into the model [5, 6, 3, 4]. The two levels of mixing generally referred to individuals having local contacts (i.e. neighbours in a network) and global contacts which may be assigned randomly from the network.

Karrer and Newman [31] derived a generalized form of an SIR model on a contact network which allowed for a non-constant infection and recovery probability (i.e. infectious period is not exponentially distributed, which is a common assumption made). They illustrated that it is possible to derive useful results for SIR epidemics on contact networks when the infectious period is not exponentially distributed. As a result of the non-constant infection and recovery probabilities they could not use standard differential equation techniques and thus developed a message passing formulation of an epidemic. They found exact results for an SIR model on tree (un-clustered) networks and tree-like graphs and determined rigorous bounds on the probabilities for non-tree-like networks. In their message passing formulation they derived an equation for the 'message' that is being passed among network nodes which is the probability that a node j has not passed infection to a neighbouring node i by absolute time t. From the information derived they were able to calculate the probability that a node is susceptible, infected or recovered at time t. Exploration into non-Markovian methods are becoming increasingly more apparent in the network model literature in order to make the models more realistic.

Other research efforts have focused their attention on deriving exact results on finite networks. In Sharkey et al. [61] they proved that for SIR dynamics on fixed tree graphs (networks without clustering), assuming an exponentially distributed transmission and recovery process, a pair-based model provides an exact determination of the infection probability time course for each individual in the network. They also illustrated that the dynamics of some networks with cycles can also be represented exactly by the pair-based method under given initial conditions. Kiss et al [37] extended the modelling framework presented in [61] for certain classes of networks exhibiting loops. Their findings lead to a generalisation of closures that are based on partitioning the network around nodes that they refer to as cut vertices, these are essentially nodes which connect smaller subgraphs in the network. They illustrated that closures where loops are kept intact are exact (that is, the cut vertex must not break a loop) and gave a general method of how to apply the reduction by closures technique for arbitrary networks and provided an upper bound on the number of equations required for an exact representation.

For recent overviews of some future challenges in analysing epidemics on networks see [45] and [54]. Keeling and Eames [33] review networks and epidemic models with a more specific focus on different ways to collect data in order to construct real networks, including infection tracing, contact tracing and diary-based studies. Danon et al [14] focus on reviewing the interplay between network theory and epidemiology. For a comprehensive review of complex networks and epidemic dynamics including detailed sections covering the relevant introductory network theory, epidemic processes in heterogeneous networks and modelling realistic epidemics (non-Markovian properties) see [52].

It is well known that considering only the degree distribution of a network is not enough to accurately predict the outcome of an epidemic [41]; two networks with the same degree distribution can exhibit different epidemic outcomes due to topological differences. It has also been shown that networks with the same degree distribution and clustering coefficient[2] can exhibit very different epidemic behaviour [55]. There are many properties that describe the structure of a network such as the

---

[2] The clustering coefficient is the ratio of triangles to triples within a network [51]

above mentioned degree distribution and clustering coefficient [51]. The purpose of this thesis is to start addressing the question of which network properties matter in terms of the spread of infection and whether these properties differ between network structures and/or model types (such as SIR and SIS).

1.2 Mathematical Preliminaries

One of the most common assumptions made when defining an epidemic model is that of constant infection and recovery rates. More formally, for the SIR and SIS models discussed in this thesis we define the following:

- Susceptible nodes acquire infection at a rate β per $S-I$ edge, that is we assume the infection process is a Poisson process with rate β.
- Infectious nodes recover at a rate γ per node and therefore, we assume the recovery process is a Poisson process with rate γ.

This section focuses on introducing the fundamental concepts which underpin the above two assumptions, including illustrating how exponential distributions, Poisson processes and Markov chains are related. The majority of the theory presented here was drawn from the extensive descriptions in 'Introduction to Probability Models' by Sheldon M. Ross [60]. Throughout this section we elaborate on the above two definitions and derive some key results for the infection and recovery processes on networks which are used throughout this thesis.

1.2.1 The Exponential Distribution

Definition 1. Ross [60] gives the following explanation of the exponential distribution and some of it's properties.
Let X be a continuous random variable. Then, X follows an exponential distribution with rate λ ($\lambda > 0$), if its probability density function is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The cumulative density function (cdf) is given by:

$$F(x) \quad = \quad P(X \leq x) = \int_{-\infty}^{x} f(y)dy \quad = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The expected value or mean of the exponential distribution is denoted $E[X]$ and is given by:

$$E[X] \quad = \quad \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{\infty} x \lambda e^{-\lambda x} dx \qquad (1.1)$$

Integrating Equation 1.1 by parts yields $E[X] = \frac{1}{\lambda}$.

A random variable X is said to be memory-less if the following statement holds true:

$$P(X > s + t | X > t) = P(X > s), \quad \text{for all } s, t \geq 0 \tag{1.2}$$

In other words, the future state of X depends only on the present state and not on the previous states. Equation 1.2 is equivalent to:

$$P(X > s + t) = P(X > s)P(X > t) \tag{1.3}$$

If X is exponentially distributed, it is easy to show that Equation 1.3 is satisfied, and hence an exponentially distributed random variable is memory-less. That is, $e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t}$. This is an important property which will become apparent in the sections that follow.

1.2.2 Poisson Processes

A Poisson process is a counting process. Ross [60] gives multiple formal definitions for a Poisson process, below we present two that we think are the most intuitive and then discuss an example which illustrates its usefulness.

Definition 2. The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process having rate λ, $\lambda > 0$ if:

(i) $N(0) = 0$;

(ii) The process has stationary and independent increments.

(iii) For small h, $P(N(h) = 1) = \lambda h + o(h)$.

(iv) For small h, $P(N(h) \geq 2) = o(h)$.

The first condition in Definition 2 simply ensures that at time $t = 0$ the Poisson process is equal to 0. The second condition requires further explanation as follows. A counting process, $N(t)$, is said to have independent increments if the number of events that occur in disjoint time intervals are independent [60]. Furthermore, Ross [60] states that $N(t)$ is said to have stationary increments if the distribution of the number of events which occur in any interval of time depends only on the length of the time interval. For example, if the number of events that occur within the time interval $(t_1 + s, t_2 + s)$ has the same distribution as the number of events occurring in the time interval (t_1, t_2); that is $N(t_2 + s) - N(t_1 + s)$ has the same distribution as $N(t_2) - N(t_1)$, for all $t_1 < t_2$ and $s > 0$. It follows from the third condition in Definition 2 above that a Poisson process has stationary increments. The third condition states that the probability one event occurs in a small time step h is approximately the rate of the process multiplied by the length of the time interval, h. Where a function $f(\cdot)$ is said to be $o(h)$ if

$$\lim_{h \to 0} \frac{f(h)}{h} = 0$$

[60]. Finally, the fourth condition states that it is highly unlikely that more than one event will occur in a given small time interval, h. The expected value of the Poisson process, $N(t)$, is equal to λt which gives rise to why λ is called the rate of the process.

The second, perhaps less formal, way of thinking about a Poisson process is described by Ross [60] as follows.

Definition 3. Let $\{T_n, n \geq 1\}$ be a sequence of independent identically distributed exponential random variables each having mean $1/\lambda$. Then define a counting process by saying that the n^{th} event of this process occurs at time $S_n \equiv T_1 + T_2 + \cdots + T_n$. The resulting process will be a Poisson process with rate λ [60].

This definition of a Poisson process naturally leads us to define some key probabilities that are used throughout this thesis which we derive in the following.

Probability an $S - I$ edge transitions to an $I - I$ edge in a network

Let $N_I(t), t \geq 0$ and $N_R(t), t \geq 0$, be two independent Poisson processes with rates β and γ respectively which describe the infection and recovery processes in a network epidemic model. Furthermore, S_n^I denotes the time of the n^{th} event of the infection process, and S_m^R denotes the time of the m^{th} event of the recovery process. We want to determine the probability that an event occurs in the infection process before an event occurs in the recovery process, that is we want to determine $P(S_n^I < S_m^R)$, where $n = m = 1$. From Definition 3, as S_1^I is the time of the first event of the $N_I(t)$ process and S_1^R is the time of the first event of the $N_R(t)$ process then it follows that both $S_I(t)$ and $S_R(t)$ are exponentially distributed random variables with mean $1/\beta$ and $1/\gamma$ respectively. To determine $P(S_1^I < S_1^R)$ we must condition on S_1^R, therefore we end up with the following:

$$\begin{aligned}P(S_1^I < S_1^R) &= \int_0^\infty P(S_1^I < S_1^R | S_1^R = x)\gamma e^{-\gamma x} dx \\ &= \int_0^\infty P(S_1^I < x)\gamma e^{-\gamma x} dx \\ &= \int_0^\infty (1 - e^{-\beta x})\gamma e^{-\gamma x} dx \\ &= \int_0^\infty \gamma e^{-\gamma x} dx - \gamma \int_0^\infty e^{-(\beta+\gamma)x} dx \\ &= 1 - \frac{\gamma}{\beta + \gamma} \\ &= \frac{\beta}{\beta + \gamma} \qquad (1.4)\end{aligned}$$

In the SIR and SIS network models discussed in this thesis, it is assumed that only one event (infection or recovery) can occur at a time. Therefore, the probability an $S - I$ edge transits to an $I - I$ edge in a network is equivalent to asking what the probability is that an infection event will occur before a recovery event. That is,

$$\frac{\beta}{\beta + \gamma}$$

Similarly, the probability an $S - I$ edge transits to an $S - R$ is given by:

$$\frac{\gamma}{\beta + \gamma}$$

1.2.3 Discrete Time Markov Chains

This section describes the mathematical theory that forms the basis of the SIS network model. We present general concepts here and leave the illustrations to be discussed further in Chapter 5. A Markov Chain is described by a set of states, $X = \{X_1, X_2, X_3, \ldots, X_n\}$, where the system moves between successive states with a given probability. If the system is currently in state X_i then the probability that the system transitions to state X_j when the next event occurs is P_{ij}. In a Markov Chain, the future state of the system depends only upon the current state. That is, the probability P_{ij} does not depend upon which states the system was in before state X_i. The probabilities P_{ij} are called transition probabilities and are often stored in a matrix of transition probabilities or a transition matrix, $\mathbf{P}$. More formally we have

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_, X_0 = i_0) = P(X_{n+1} = j | X_n = i) = P_{ij}$$

for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and $n \geq 0$ [60].

As probabilities are non-negative and the process must make a transition into some state, we have:

$$P_{ij} \geq 0, \quad i, j \geq 0; \quad \sum_{j=0}^{\infty} P_{ij} = 1, \quad i = 0, 1, \ldots$$

The matrix, **P** denotes the one-step transition probabilities P_{ij} [60], that is the entries of **P** are of the form

$$P_{ij} \quad = \quad P(\text{system is in state j after one event} \mid \text{system is currently in state i})$$

This is often referred to as the embedded Discrete Time Markov Chain (DTMC). Furthermore, $\mathbf{P}^n$ denotes the n-step probability matrix where it's entries, P_{ij}^n define the probability that the process is in state j after n steps given that it started in state i. The Chapman-Kolmogorov Equations provide a method for computing these n step probabilities as follows.

Chapman-Kolmogorov Equations

To determine the n step probabilities we wish to find the probability that a process in state i will be in state j after n additional transitions, that is

$$P_{ij}^n = P(X_{n+m} = j | X_m = i), \quad n \geq 0, i, j \geq 0$$

The Chapman-Kolmogorov Equations are defined as

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m \quad \text{for all } n, m \geq 0, \text{ all } i, j$$

where $P_{ik}^n P_{kj}^m$ gives the probability that starting in state i, the process will transition to state j in $n+m$ transitions through a path which takes the process into state k at the n^{th} transition. Therefore, by summing over all possible intermediate states k we obtain the probability that the process will be in state j after $n+m$ transitions. See Ross [60] for a full proof.

For a specified initial condition we can also determine the unconditional distribution of the state at time n as follows [60] Let α_i denote the probability that the initial state of the process is state i, that is $\alpha_i = P(X_0 = i)$ for $i \geq 0$ and $\sum_{i=0}^{\infty} \alpha_i = 1$. Then we have

$$\begin{aligned} P(X_n = j) &= \sum_{i=0}^{\infty} P(X_n = j | X_0 = i) P(X_0 = i) \\ &= \sum_{i=0}^{\infty} P_{ij}^n \alpha_i \end{aligned}$$

The above theory underpins the results which we derive in Chapter 5 to describe the SIS epidemic process on networks.

1.2.4 Continuous Time Markov Chains

A Continuous Time Markov Chain (CTMC) is a stochastic process having the Markovian property that the conditional distribution of the future state, say $X(t+s)$, given the current state, $X(s)$ and the past state, $X(u)$, $0 \leq u < s$, depends only on the current state, $X(s)$. That is $X(t+s)$ is independent of the past state $X(u)$. More formally,

$$P(X(t+s) = j | X(s) = i, X(u) = x(u), 0 \leq u < s) = P(X(t+s) = j | X(s) = i)$$

Perhaps not surprisingly due to this memory-less property, a good example of a CTMC is the Poisson process. The following properties also hold for a CTMC each time the process enters state i [60]:

- The amount of time spent in that state before making a transition is exponentially distributed with mean $1/v_i$ (where v_i is the rate the process makes a transition) and,
- When the process leaves state i, it next enters state j with probability P_{ij} where P_{ij} must satisfy:
    - ◇ $P_{ii} = 0 \quad \forall i$
    - ◇ $\sum_j P_{ij} = 1 \quad \forall i$

Therefore, a CTMC is a stochastic process that moves from state to state in accordance with a DTMC, as discussed in Section 1.2.3, but is such that the amount of time spent in each state before transitioning to the next state is exponentially distributed. The amount of time spent in each state is also independent of the amount of time spent in any other state [60]. An important distinction between the DTMC and CTMC should be noted here. The matrix $\mathbf{P}$ describes the probabilistic behaviour of the embedded DTMC, however it does not fully capture the behaviour of the continuous time process because it does not specify the timescale at which the transitions occur.

Kolmogorov Differential Equations

In this section we summarise the theory upon which the methods used to describe the SIR model on networks in this thesis is based.

For any pair of states i and j, q_{ij} is defined as the rate at which the process makes a transition into state j given that it was in state i and takes the following form

$$q_{ij} = v_i P_{ij}$$

where v_i is the rate the process makes a transition when in state i and P_{ij} is the probability that this transition is into state j. The above quantities are linked by the following relationship:

$$v_i = \sum_j v_i P_{ij} = \sum_j q_{ij}$$

and

$$P_{ij} = \frac{q_{ij}}{v_i} = \frac{q_{ij}}{\sum_j q_{ij}}$$

By specifying the above rates we can determine the parameters of the process. Let $P_{ij}(t)$ be the probability that a process will be in state j after some time, t, given that it was in state i at time s.

$$P_{ij}(t) = P(X(t+s) = j | X(s) = i)$$

In what follows we derive a set of differential equations for these transition probabilities. However first we require the following lemmas as described in [60].

Lemma 1 (Ross [60])**.**

$$\begin{aligned} \lim_{h \to 0} \frac{1 - P_{ij}}{h} &= v_i && (1.5) \\ \lim_{h \to 0} \frac{P_{ij}}{h} &= q_{ij} \quad \textit{when } i \neq j && (1.6) \end{aligned}$$

Proof. First note that as the amount of time until a transition occurs is exponentially distributed it follows that the probability of two or more transitions occurring in a time interval h is $o(h)$. Therefore, the probability that a process in state i will not be in state i at time h,$1 - P_{ii}(h)$, is equal to the probability that a transition occurs within the time interval h plus something small compared to h. Thus,

$$1 - P_{ii}(h) = v_i h + o(h)$$

which proves Equation 1.5 above. Secondly, note that $P_{ij}(h)$ (the probability that the process goes from state i to state j in a time h, equals the probability that a transition occurs in this time multiplied by the probability that the transition is into state j, plus something small compared to h. Therefore,

$$P_{ij}(h) = h v_i P_{ij} + o(h)$$

and thus Equation 1.6 is proven [60]. □

Lemma 2 (Ross [60])**.** *For all $s \geq 0$, $t \geq 0$,*

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(s) \tag{1.7}$$

Proof. In order for the process to go from state i to state j in time $t+s$, it must be somewhere at time t and thus

$$\begin{aligned}
P_{ij}(t+s) &\equiv P(X(t+s)=j|X(0)=i) \\
&= \sum_{k=0}^{\infty} P(X(t+s)=j, X(t)=k|X(0)=i) \\
&= \sum_{k=0}^{\infty} P(X(t+s)=j|X(t)=k, X(0)=i) \times P(X(t)=k|X(0)=i) \\
&= \sum_{k=0}^{\infty} P(X(t+s)=j|X(t)=k) \times P(X(t)=k|X(0)=i) \\
&= \sum_{k=0}^{\infty} P_{kj}(s)P_{ik}(t)
\end{aligned}$$

which completes the proof [60]. □

From Lemmas 1 and 2 we can derive the following equation.

$$\begin{aligned}
P_{ij}(h+t) - P_{ij}(t) &= \sum_{k=0}^{\infty} P_{ik}(h)P_{kj}(t) - P_{ij}(t) \\
&= \sum_{k\neq i} P_{ik}(h)P_{kj}(t) - [1-P_{ii}(h)]P_{ij}(t)
\end{aligned}$$

and taking the limit of both sides as h tends towards zero gives

$$\lim_{h\to 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \lim_{h\to 0} \left(\sum_{k\neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) - \left[\frac{1-P_{ii}(h)}{h}\right] P_{ij}(t) \right)$$

Noting that in the above equation it can be justified to interchange the limit and the summation, we arrive at the following:

$$\dot{P}_{ij}(t) = \sum_{k\neq i} q_{ik}P_{kj}(t) - v_i P_{ij}(t) \tag{1.8}$$

where $\dot{P} = \frac{dP}{dt}$. The above Equation 1.9 holds for all states i, j, and times $t \geq 0$ and is more formally known as Kolmogorov's Backward Equations.

We can also derive the Kolmogorov Forward Equations following a similar method. From Lemmas 1 and 2 we can derive the following equation.

$$\begin{aligned}
P_{ij}(h+t) - P_{ij}(t) &= \sum_{k=0}^{\infty} P_{ik}(h)P_{kj}(t) - P_{ij}(t) \\
&= \sum_{k\neq j} P_{ik}(h)P_{kj}(t) - [1-P_{jj}(h)]P_{ij}(t)
\end{aligned}$$

and taking the limit of both sides as h tends towards zero gives

$$\lim_{h\to 0}\frac{P_{ij}(t+h)-P_{ij}(t)}{h}=\lim_{h\to 0}\left(\sum_{k\neq j}P_{ik}(h)\frac{P_{kj}(t)}{h}-\left[\frac{1-P_{jj}(h)}{h}\right]P_{ij}(t)\right)$$

Noting that in the above equation it can be justified to interchange the limit and the summation, we arrive at the following:

$$\dot{P}_{ij}(t)=\sum_{k\neq j}q_{kj}P_{ik}(t)-v_jP_{ij}(t) \tag{1.9}$$

In the above equation interchanging the limit and the summation is not always valid, however it does hold for most models including all finite state models and birth and death processes [60].

Equations for the SIR and SIS Models

We define the SIR epidemic on a network using a continuous time Markov chain where the infection and recovery processes are modelled as independent Poisson processes (as discussed above). This naturally leads us to define the Kolmogorov differential equations for an SIR epidemic on a network.

For the SIR and SIS network models we derive a system of Kolmogorov differential equations which describes how the infection progresses through a given network over time. We denote P_i as the probability that the network is in state i at time t; q_{ij} is the rate the system makes a transition from state i to state j and v_i is the rate the system transitions out of state i. Therefore, in general, we have:

$$\dot{\mathsf{P}}_j = \sum_{i\neq j}q_{ij}\mathsf{P}_i - v_i\mathsf{P}_j \tag{1.10}$$

For the SIR model only we also derive a set of equations describing the probability that the network is ever in state j denoted by $\mathcal{P}_j$. Note that these probabilities are independent of time and are later used to derive the final size probability mass functions.

$$\mathcal{P}_j = \sum_{i\neq j}P_{ij}\mathcal{P}_i \tag{1.11}$$

where P_{ij} is the probability that the transition is from state i to state j. We discuss these equations and the state space of the SIR and SIS network models in more detail in Chapters 2 and 5 of this thesis.

1.3 Overview

This thesis is structured as follows. Part I focuses on a SIR compartmental model for the spread of an infection through a population. In Chapter 2 we outline the methods used to find the analytic probability mass function (PMF) of the final epidemic size of small networks consisting of $N = 3$ and $N = 4$ nodes. Details are given for the simplest and most complex small networks to illustrate. This is followed by the presentation and discussion of results for the SIR model on all small networks considered. In Chapter 3 we illustrate how we can use known results found for the small networks in order to analytically describe the spread of an infection through a larger network. We detail methods and results for an SIR model on a line of triangles (LoT) network with $N = 6$ nodes; followed by generalising the analytic results, including the PMF of the final epidemic size, for a LoT with an arbitrary number of nodes, N. In Chapter 4 we discuss the ambiguity that surrounds the definition of the basic reproduction number, R_0, and give an overview of how R_0 has been defined for network models. This is followed by our interpretation of R_0 for small networks using the Next Generation Matrix approach [16, 18, 57]. We then illustrate how our SIR model can be used to investigate the spread of influenza and measles on small networks which could represent a small population or community.

Part II focuses on a SIS model for the spread of an infection through a population. In Chapter 5 we use Markov Chain methodology to derive results such as the expected time to absorption, the expected number of times each individual is infected and the cumulative incidence of the epidemic for the SIS model on the simplest and most complex small networks. We derive results for a SIS model on a LoT network consisting of $N = 6$ and $N = 9$ nodes in Chapter 6, this is followed by the derivation of a generalised transition matrix for a SIS model on any given network. This chapter ends with a discussion of results and differences between the SIR and SIS models on the LoT network. In Chapter 7 we illustrate how to parameterise the SIS small network models using rabies and chlamydia as examples. This is followed by the presentation and discussion of results.

Appendix A contains the detailed methods for the SIR model on the remaining six small networks. In Appendix B we give tables comparing the derived analytic results with stochastic realisations for the SIR model on a LoT network. We also present figures that show the difference between analytic and stochastic results converges to zero as the number of realisations increases. Appendix C contains the detailed methods for the SIS model on the remaining six small networks. In Appendix D we show that the difference between analytic and stochastic results converges to zero as the number of realisations increases for the SIS model on small networks and LoT network with $N = 6$ nodes.

Part I

SIR Model

Chapter 2

SIR epidemics on small networks

In this chapter we focus on an SIR compartmental model to describe the spread of an infection through a population. We consider eight small networks of varying topological structure (see Figure 2.1) and investigate the dynamics of the model on them. First, we define the notation used throughout this thesis. Susceptible nodes acquire infection at rate β per $S-I$ edge and infectious nodes recover at rate γ. As all of our results depend on the ratio β/γ, we define the variable $\mathcal{R} = \beta/\gamma$ to simplify our expressions. We denote the state that a network of three nodes is in by XYZ, where X, Y and Z denote the infection state (S, I or R) that nodes a, b and c are in respectively. Similarly, we denote the state that a network of four nodes is in by $WXYZ$.

For an SIR model on a network of N nodes there are 3^N possible states in which the network can be. Thus, for an SIR model on networks of size $N = 3$ and $N = 4$ nodes there are 27 and 81 possible states in which the networks can be respectively. To simplify our analysis we group states together based on symmetries of the network [62]; in the simplest case of a complete network we can group states together that have the same number of S, I and R nodes (A complete network is a network in which each node is connected to every other node within the network, see for example Figure 2.1 A and C). For networks that are not complete we must look at the topology and the number of nodes in each infection state before determining which network states can be grouped together. Thus, for each network we group together the states that are topologically equivalent and have the same number of nodes in each infection state. See Figure 2.2 for illustrations of topologically equivalent states. Once we have the reduced number of states we draw a transition diagram which shows how the infection moves through the network.

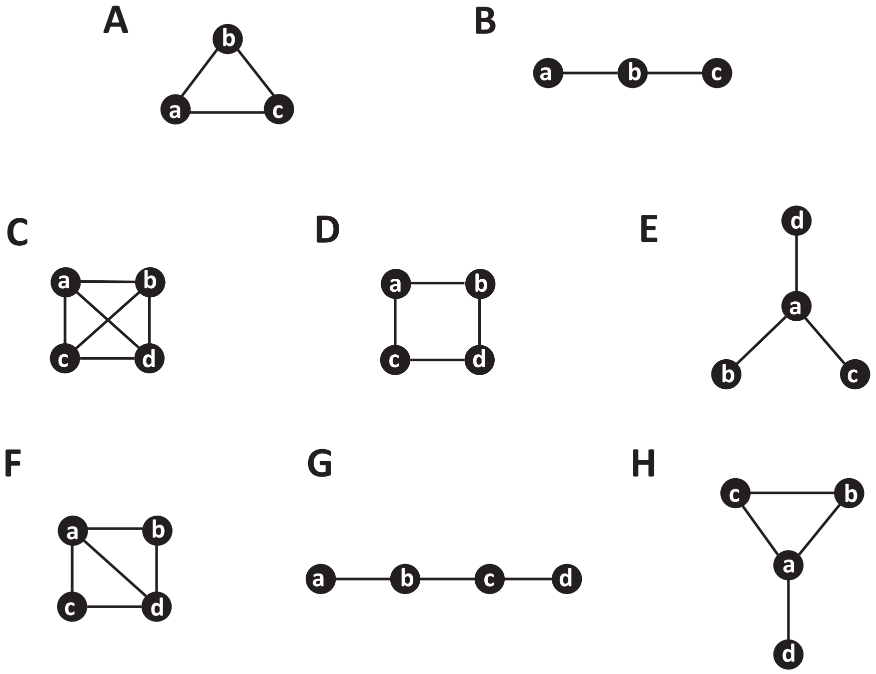


Figure 2.1: Network diagrams in order of increasing complexity. **A**. Triangle Network. **B**. Line Network, $N = 3$. **C**. Complete Network, $N = 4$. **D**. Square Network. **E**. Star Network. **F**. Toast Network. **G**. Line Network, $N = 4$. **H**. Lollipop Network.

Each box in the transition diagram represents a different state of the network and each arrow represents either recovery or infection of a node. Coloured outlines represent topologically different states with the same number of nodes in each infection state. Transition between states is only possible if there is a directed arrow from one to the other. Movement between states in the transition diagram shows how the infection spreads through the network. Once the network reaches an absorbing state the epidemic is over and the infection has died out. From the transition diagram we derive the individual transition probabilities between network states.

From the transition diagrams we can calculate the probability of obtaining different final sizes which depends only on the initial state and the infection parameters. The final size is the total number of nodes that acquire infection at some point during the epidemic. For an SIR model this is simply the number of nodes that end up in the recovered state. We obtain these final sizes by summing up all possible infection paths to the appropriate absorbing states, given the infection starts in one of the initial states. Once the network reaches an absorbing state the epidemic is over and the infection has died out. The analytic expression for the expected final size of the epidemic is also determined. The expected final size is defined by

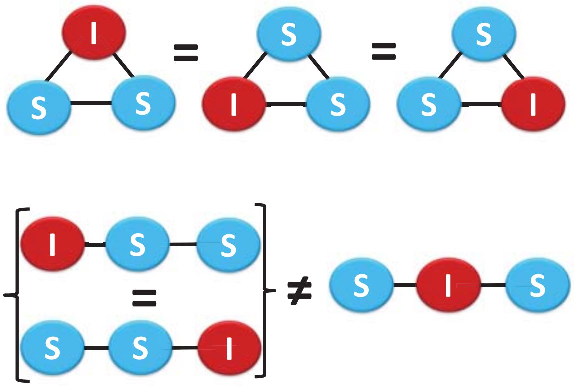


Figure 2.2: Example of grouping together topologically equivalent states for networks of size $N = 3$. The three triangle network states (top) are equivalent and thus can be grouped together into one state. However, for the line network (bottom) the two states on the left, where the end node is infectious, are equivalent and thus are grouped together to form one state. The state on the bottom right, where the central node of the line network is infectious, is not topologically equivalent and so it forms another state of its own.

$\mathbb{E}[\text{Final Size}] = \sum_{i=1}^{N} i\mathbb{P}(\text{Final Size} = i)$. We can also show how the infection progresses through the network by writing down a set of differential equations; each equation describing the evolution of the probability that the network is in a given state at time t. We then numerically solve the equations with the same initial conditions and infection parameters that we used to find the analytical expressions of the probability mass function for the final epidemic size. To independently verify our results we also ran stochastic realisations on each of the small networks for the same initial conditions and infection parameters. The Gillespie algorithm was used for the stochastic SIR model (See Section 2.3 and also [34] for details of the algorithm) which was implemented using MATLAB software. To calculate the final size probabilities we ran 2×10^5 stochastic realisations on each network with the initial conditions and parameters as specified, and counted the frequency of each final size occurring. We also calculated the clustering coefficient, ϕ, for each network to see if any patterns emerged between the final epidemic size and clustering coefficient of the networks. The clustering coefficient is the ratio of triangles to triples in a network; a triple is any three nodes that are connected which could be either a triangle or a line of three nodes.

In the following sections of this chapter we illustrate our methods with the simplest and most complex small networks, the triangle and lollipop networks respectively. We then outline the Gillespie algorithm for the stochastic SIR model used to verify our results. This is followed by the presentation and discussion of our results from all small networks shown in Figure 2.1. The detailed methods for each of the remaining small networks are given in Appendix A, page 196.

2.1 Triangle Network

The triangle network is the simplest of the eight small networks we investigated, being the smallest complete network consisting of three nodes of degree two. For an SIR model on a triangle network there are 27 possible states in which the network can be. As we are dealing with a complete network, we can group states together based on those with the same number of nodes in each infection state. By grouping the appropriate states together, the triangle network can be reduced to 10 sets of states shown by the transition diagram in Figure 2.3.

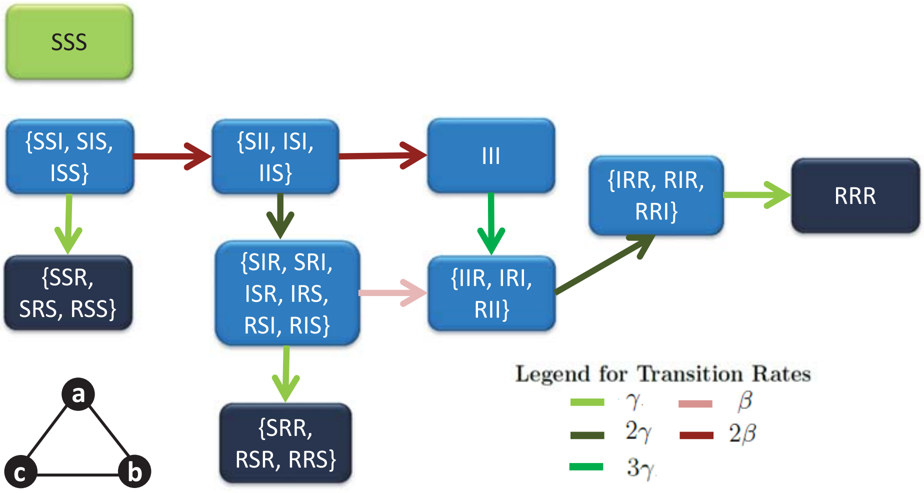


Figure 2.3: Transition diagram for the triangle (complete) network with $N = 3$ nodes.

In the triangle network, if the epidemic is started with one initial infectious node it does not matter if that is node a, b or c as each node is topologically equivalent. From the transition diagram of the SIR model on the triangle network (Figure 2.3) we derive the individual transition probabilities between network states. We obtained the final size probability mass functions for the SIR epidemic process on the triangle network, given that the epidemic was started with one initial infectious node shown in Table 2.1. Note that the probability of the final size equaling one is the same as the probability of the infection not taking off, that is the one infectious node recovers before it could infect one of its neighbours.

2.1.1 Catalogue of transition probabilities for the triangle network

From the transition diagram of the SIR model on the triangle network we derive the individual transition probabilities between network states. For example, given the system is in state SSI, the events that could happen in the next step are a) infection of one of the two susceptible nodes (as they are both neighbours of the infectious node), this will happen at a rate 2β as infection occurs at a rate β per $S-I$ edge as discussed in the introduction; b) recovery of the infectious node which happens at a rate γ. Therefore, the total rate that the system leaves state SSI is equal to the sum of the rates of these events, that is $2\beta+\gamma$. Then, we can write down the probabilities of the system moving from state SSI to SII (infection event occured) and SSR (recovery event occured):

$$\begin{aligned} P(SSI \to SII) &= \frac{2\beta}{2\beta+\gamma} \\ P(SSI \to SSR) &= \frac{\gamma}{2\beta+\gamma} \end{aligned}$$

It is easy to see that the sum of these probabilities equals one, which is expected. To obtain the transition probabilities which we give below, we set the variable $\mathcal{R}=\beta/\gamma$ in order to simplify our expressions. In the following $\mathcal{P}_{XYZ}$ denotes the probability that the network is ever in state XYZ, where X, Y and Z denote the infection state (S, I or R) that nodes a, b and c are in respectively. These probabilities are independent of time and depend only on the infection parameters. To specifiy which initial state the system starts in we define indicator variables, E_{XYZ}, which are set to equal one if the system begins in state XYZ and zero otherwise. From the absorbing state probabilities we find the final epidemic size probabilities, $\mathbb{P}$.

Possible initial state:

$$\mathrm{E}_{SSI} \quad = \quad 1$$

Probability of passing through transient states:

$$\begin{aligned} \mathcal{P}_{SII} &= \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{SSI} \\ \mathcal{P}_{SIR} &= \frac{2}{2\mathcal{R}+2}\mathcal{P}_{SII} \\ \mathcal{P}_{III} &= \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{SII} \\ \mathcal{P}_{IIR} &= \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SIR}+\mathcal{P}_{III} \\ \mathcal{P}_{IRR} &= \mathcal{P}_{IIR} \end{aligned}$$

Probability of terminating in absorbing states:

$$\begin{aligned}
\mathcal{P}_{SSR} &= \frac{1}{2\mathcal{R}+1}\mathrm{E}_{SSI} \\
\mathcal{P}_{SRR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SIR} \\
\mathcal{P}_{RRR} &= \mathcal{P}_{IRR}
\end{aligned}$$

To find the equations for the final size probabilities we evaluated the following:

$$\begin{aligned}
\mathbb{P}(\text{Final Size} = 1) &= \mathcal{P}_{SSR} \\
\mathbb{P}(\text{Final Size} = 2) &= \mathcal{P}_{SRR} \\
\mathbb{P}(\text{Final Size} = 3) &= \mathcal{P}_{RRR}
\end{aligned}$$

Simplifying the above we derive the analytic expressions for the final size probabilities for the triangle network as shown in Table 2.1. For example, to illustrate how we found $\mathbb{P}(\text{Final Size} = 3)$, we make the following substitutions:

$$\begin{aligned}
\mathbb{P}(\text{Final Size} = 3) &= \mathcal{P}_{RRR} \\
&= \mathcal{P}_{IRR} \\
&= \mathcal{P}_{IIR} \\
&= \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SIR} + \mathcal{P}_{III} \\
&= \frac{\mathcal{R}}{\mathcal{R}+1}\frac{2}{2\mathcal{R}+2}\mathcal{P}_{SII} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{SII} \\
&= \left(\frac{\mathcal{R}}{(\mathcal{R}+1)^2} + \frac{\mathcal{R}}{\mathcal{R}+1}\right)\mathcal{P}_{SII} \\
&= \frac{\mathcal{R}^2+2\mathcal{R}}{(\mathcal{R}+1)^2}\left(\frac{2\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{SSI}\right) \\
&= \frac{2\mathcal{R}^2\left(\mathcal{R}+2\right)}{\left(\mathcal{R}+1\right)^2\left(2\,\mathcal{R}+1\right)}\mathrm{E}_{SSI}
\end{aligned}$$

The probability mass function of the final size distribution for the triangle network is shown in Figure 2.9.

2.1.2 Progression of infection over time

In the following we use P_{XYZ} to denote the probability that the triangle network is in the state XYZ at time t, where X, Y and Z denote the infection state (S, I or R) that nodes a, b and c are in respectively. Thus, the equation for the time derivative $\dot{\mathsf{P}}_{XYZ}$ shows how the network can enter and leave the state XYZ. The rate at which the network enters and leaves each state can be found from the transition diagram. For example, to determine the equation for the time derivative $\dot{\mathsf{P}}_{SII}$ we take the

Table 2.1: Triangle network final size PMFs

Initial State	SSI
$\mathbb{P}$**(Final Size =1)**	$\frac{1}{2\mathcal{R}+1}$
$\mathbb{P}$**(Final Size =2)**	$\frac{2\mathcal{R}}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$
$\mathbb{P}$**(Final Size =3)**	$\frac{2\mathcal{R}^2(\mathcal{R}+2)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$
Expected FS	$\frac{6\mathcal{R}^3+13\mathcal{R}^2+6\mathcal{R}+1}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$

following steps. From the transition diagram we can see that the only way to enter the state SII is from state SSI at rate 2β; similarly there are two ways to leave state SII and that is to enter state III at rate 2β or to enter state SIR at rate 2γ. Putting this together into an equation yields:

$$\dot{\mathsf{P}}_{SII} = 2\beta\mathsf{P}_{SSI} - (2\beta + 2\gamma)\mathsf{P}_{SII}$$

We rescale time so $\gamma = 1$ in the equations before substituting in $\mathcal{R} = \beta/\gamma$ to obtain:

$$\begin{aligned}\dot{\mathsf{P}}_{SII} &= 2\beta\mathsf{P}_{SSI} - (2\beta + 2\gamma)\mathsf{P}_{SII} \\ &= \frac{2\beta}{\gamma}\mathsf{P}_{SSI} - \frac{(2\beta + 2\gamma)}{\gamma}\mathsf{P}_{SII} \\ &= 2\mathcal{R}\mathsf{P}_{SSI} - 2(\mathcal{R}+1)\mathsf{P}_{SII}\end{aligned}$$

These equations allow us to simulate the time course of the epidemic and to check our final size calculations. Equations describing the probability that the network is in a given state at time t for an SIR model on the triangle network are:

$$\begin{aligned}
&\text{Initial states:} && \\
&\dot{\mathsf{P}}_{SSS} &=& \; 0 \qquad (2.1)\\
&\dot{\mathsf{P}}_{SSI} &=& \; -(2\mathcal{R}+1)\mathsf{P}_{SSI}\\
&\text{Transient states:} && \\
&\dot{\mathsf{P}}_{SII} &=& \; 2\mathcal{R}\mathsf{P}_{SSI} - 2(\mathcal{R}+1)\mathsf{P}_{SII}\\
&\dot{\mathsf{P}}_{SIR} &=& \; 2\mathsf{P}_{SII} - (\mathcal{R}+1)\mathsf{P}_{SIR}\\
&\dot{\mathsf{P}}_{IIR} &=& \; \mathcal{R}\mathsf{P}_{SIR} + 3\mathsf{P}_{III} - 2\mathsf{P}_{IIR}\\
&\dot{\mathsf{P}}_{III} &=& \; 2\mathcal{R}\mathsf{P}_{SII} - 3\mathsf{P}_{III}\\
&\dot{\mathsf{P}}_{IRR} &=& \; 2\mathsf{P}_{IIR} - \mathsf{P}_{IRR}\\
&\text{Absorbing states:} && \\
&\dot{\mathsf{P}}_{SSR} &=& \; \mathsf{P}_{SSI}\\
&\dot{\mathsf{P}}_{SRR} &=& \; \mathsf{P}_{SIR}\\
&\dot{\mathsf{P}}_{RRR} &=& \; \mathsf{P}_{IRR}
\end{aligned}$$

We have included the equation for the initial state SSS for completeness, even though it is disjoint from the transition diagram as no infection is present. As $t \to \infty$, the infection dies out and the system ends up in one of the absorbing states. For a given transition rate, the initial state determines the probability of tending to each of the possible absorbing states. To find the final size probabilities we numerically solve the system of differential equations (see Figure 2.4) with specified initial conditions and sum up the appropriate absorbing state probabilities once a steady state has been reached. Numerical results confirm our analytical expressions for the final epidemic size. For the SIR model, $\mathcal{R} = \beta/\gamma$ can be thought of as being a naive estimate for R_0 (the basic reproduction number). It is well known that the basic reproduction number has a threshold at $R_0 = 1$; if $R_0 < 1$ the infection is likely to die out and if $R_0 > 1$ an epidemic is likely to occur. Therefore, we have chosen to compare results with $\mathcal{R} = 1$ and $\mathcal{R} = 2$ to investigate the dynamics of an SIR model. This allows us to compare results found with $\beta = \gamma$ to results found with $\beta = 2\gamma$.

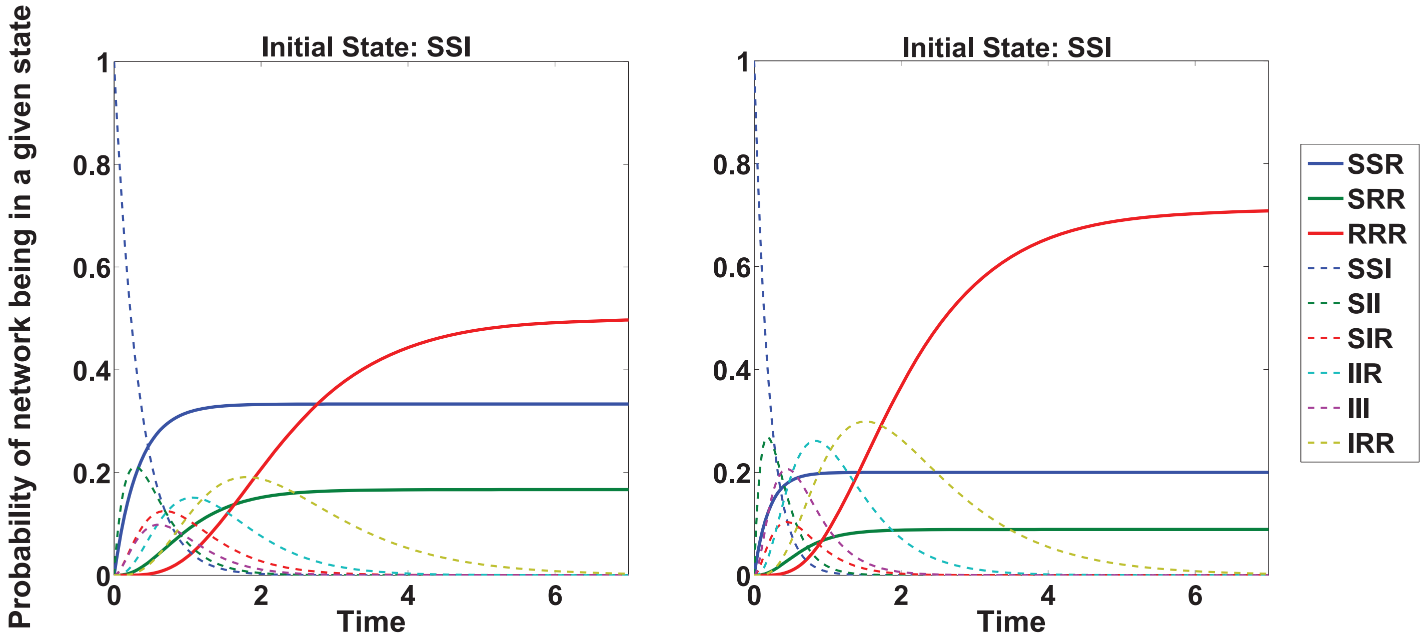


Figure 2.4: Numerical solution of the system of Equations 2.1 that describes the progression of infection over time for an SIR model on a triangle network with $N = 3$ nodes. Left and right columns contain graphical results for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively for starting the epidemic in the state SSI with one infectious node. The numerical results are in agreement with the analytical expressions for the same set of initial conditions.

2.2 Lollipop Network

We now consider the Lollipop network of four nodes which is a triangle network with one extra node connected to it (see Figure 2.1 H). For the lollipop network there are three different types of nodes; node a has degree 3, nodes b and c have degree 2 and node d has degree 1. Therefore, the final size probabilities vary depending on which type of node is the initial infectious node. We denote $ISSS$ as the initial state in which node a is infectious; $SSSI$ as the initial state in which node d is infectious; $SISS$ as the initial state in which either node b or node c is infectious and $SSII$ as the initial state in which nodes b and d are infectious. To simplify our analysis we group states together that are topologically equivalent and that have the same number of nodes in each infection state. By grouping the appropriate states together, the lollipop network can be reduced to a system of 54 states shown by the transition diagram in Figure 2.5. The probability mass functions of the final size for the lollipop network found with four initial conditions are shown in Table 2.2.

2.2.1 Catalogue of transition probabilities

Possible initial state indicator variables:

$$\mathrm{E}_{SISS} = \begin{cases} 1, & \text{if initial state is } SISS. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{ISSS} = \begin{cases} 1, & \text{if initial state is } ISSS. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{SSSI} = \begin{cases} 1, & \text{if initial state is } SSSI. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{SSII} = \begin{cases} 1, & \text{if initial state is } SSII. \\ 0, & \text{otherwise.} \end{cases}$$

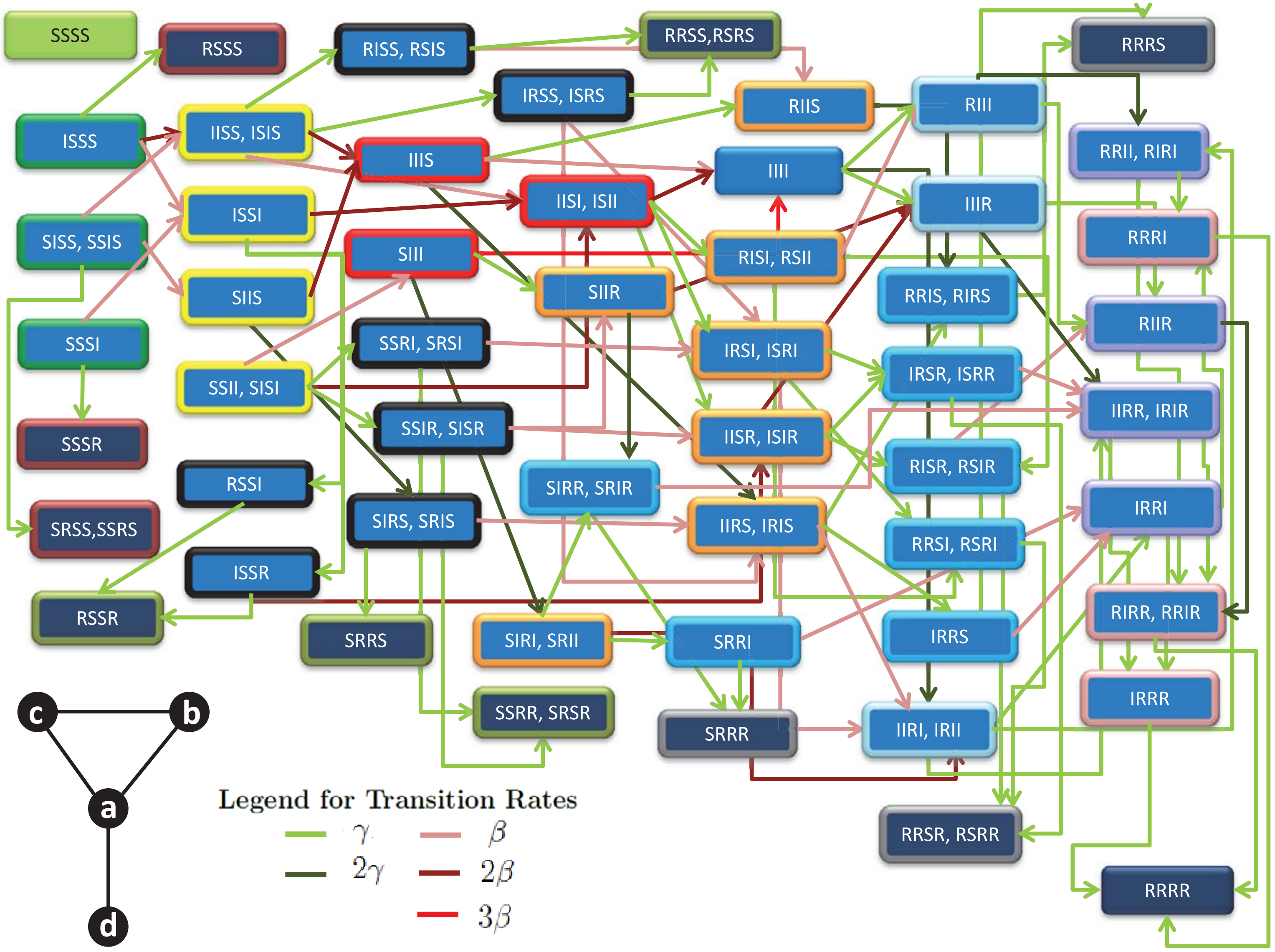


Figure 2.5: Transition diagram for the lollipop network with $N = 4$ nodes.

Probability of passing through transient states:

$$\begin{aligned}
\mathcal{P}_{IISS} &= \frac{2\mathcal{R}}{3\mathcal{R}+1}\mathrm{E}_{ISSS} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{SISS} \\
\mathcal{P}_{ISSI} &= \frac{\mathcal{R}}{3\mathcal{R}+1}\mathrm{E}_{ISSS} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathrm{E}_{SSSI} \\
\mathcal{P}_{SIIS} &= \frac{\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{SISS} \\
\mathcal{P}_{RSSI} &= \frac{1}{2\mathcal{R}+2}\mathcal{P}_{ISSI} \\
\mathcal{P}_{ISSR} &= \frac{1}{2\mathcal{R}+2}\mathcal{P}_{ISSI} \\
\mathcal{P}_{RISS} &= \frac{1}{3\mathcal{R}+2}\mathcal{P}_{IISS} \\
\mathcal{P}_{IRSS} &= \frac{1}{3\mathcal{R}+2}\mathcal{P}_{IISS} \\
\mathcal{P}_{IIIS} &= \frac{2\mathcal{R}}{3\mathcal{R}+2}\mathcal{P}_{IISS} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{SIIS} \\
\mathcal{P}_{IISI} &= \frac{\mathcal{R}}{3\mathcal{R}+2}\mathcal{P}_{IISS} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{ISSI} + \frac{2\mathcal{R}}{3\mathcal{R}+2}\mathrm{E}_{SSII} \\
\mathcal{P}_{SIII} &= \frac{\mathcal{R}}{3\mathcal{R}+2}\mathrm{E}_{SSII} \\
\mathcal{P}_{SSRI} &= \frac{1}{3\mathcal{R}+2}\mathrm{E}_{SSII} \\
\mathcal{P}_{SSIR} &= \frac{1}{3\mathcal{R}+2}\mathrm{E}_{SSII} \\
\mathcal{P}_{SIRS} &= \frac{2}{2\mathcal{R}+2}\mathcal{P}_{SIIS} \\
\mathcal{P}_{SIIR} &= \frac{1}{3\mathcal{R}+3}\mathcal{P}_{SIII} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{SSIR} \\
\mathcal{P}_{SIRI} &= \frac{2}{3\mathcal{R}+3}\mathcal{P}_{SIII} \\
\mathcal{P}_{SIRR} &= \frac{2}{2\mathcal{R}+2}\mathcal{P}_{SIIR} + \frac{1}{2\mathcal{R}+2}\mathcal{P}_{SIRI} \\
\mathcal{P}_{RIIS} &= \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{RISS} + \frac{1}{\mathcal{R}+3}\mathcal{P}_{IIIS} \\
\mathcal{P}_{IIII} &= \frac{\mathcal{R}}{\mathcal{R}+3}\mathcal{P}_{IIIS} + \frac{2\mathcal{R}}{2\mathcal{R}+3}\mathcal{P}_{IISI} + \frac{3\mathcal{R}}{3\mathcal{R}+3}\mathcal{P}_{SIII} \\
\mathcal{P}_{RISI} &= \frac{1}{2\mathcal{R}+3}\mathcal{P}_{IISI} \\
\mathcal{P}_{IRSI} &= \frac{1}{2\mathcal{R}+3}\mathcal{P}_{IISI} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{IRSS} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SSRI} \\
\mathcal{P}_{IISR} &= \frac{1}{2\mathcal{R}+3}\mathcal{P}_{IISI} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{SSIR} + \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{ISSR} \\
\mathcal{P}_{IIRS} &= \frac{2}{\mathcal{R}+3}\mathcal{P}_{IIIS} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SIRS} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{IRSS} \\
\mathcal{P}_{SRRI} &= \frac{1}{2\mathcal{R}+1}\mathcal{P}_{SIRI} \\
\mathcal{P}_{RRIS} &= \mathcal{P}_{RIIS} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{IIRS}
\end{aligned}$$

$$\begin{aligned}
\mathcal{P}_{IRSR} &= \frac{1}{2\mathcal{R}+2}\mathcal{P}_{IISR} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{IRSI} \\
\mathcal{P}_{RISR} &= \frac{1}{2\mathcal{R}+2}\mathcal{P}_{IISR} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{RISI} \\
\mathcal{P}_{RRSI} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{IRSI} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{RISI} \\
\mathcal{P}_{IRRS} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{IIRS} \\
\mathcal{P}_{IIRI} &= \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{IIRS} + \frac{1}{2}\mathcal{P}_{IIII} + \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{IRSI} + \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{SIRI} \\
\mathcal{P}_{RIII} &= \frac{1}{4}\mathcal{P}_{IIII} + \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{RISI} \\
\mathcal{P}_{IIIR} &= \frac{1}{4}\mathcal{P}_{IIII} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{SIIR} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{IISR} \\
\mathcal{P}_{RRII} &= \frac{2}{3}\mathcal{P}_{RIII} + \frac{1}{3}\mathcal{P}_{IIRI} \\
\mathcal{P}_{RIIR} &= \frac{1}{3}(\mathcal{P}_{IIIR} + \mathcal{P}_{RIII}) + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{RISR} \\
\mathcal{P}_{IIRR} &= \frac{\mathcal{R}}{\mathcal{R}+1}(\mathcal{P}_{IRSR} + \mathcal{P}_{SIRR}) + \frac{2}{3}\mathcal{P}_{IIIR} + \frac{1}{3}\mathcal{P}_{IIRI} \\
\mathcal{P}_{IRRI} &= \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SRRI} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{IRRS} + \frac{1}{3}\mathcal{P}_{IIRI} \\
\mathcal{P}_{RRRI} &= \frac{1}{2}(\mathcal{P}_{RRII} + \mathcal{P}_{IRRI}) \\
\mathcal{P}_{RIRR} &= \mathcal{P}_{RIIR} + \frac{1}{2}(\mathcal{P}_{IIRR} + \mathcal{P}_{RRII}) \\
\mathcal{P}_{IRRR} &= \frac{1}{2}(\mathcal{P}_{IIRR} + \mathcal{P}_{IRRI})
\end{aligned}$$

Probability of terminating in absorbing states:

$$\begin{aligned}
\mathcal{P}_{RSSS} &= \frac{1}{3\mathcal{R}+1}\mathrm{E}_{ISSS} \\
\mathcal{P}_{SSSR} &= \frac{1}{\mathcal{R}+1}\mathrm{E}_{SSSI} \\
\mathcal{P}_{SRSS} &= \frac{1}{2\mathcal{R}+1}\mathrm{E}_{SISS} \\
\mathcal{P}_{RSSR} &= \frac{1}{2\mathcal{R}+1}\mathcal{P}_{ISSR} + \mathcal{P}_{RSSI} \\
\mathcal{P}_{RRSS} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{RISS} + \frac{1}{2\mathcal{R}+1}\mathcal{P}_{IRSS} \\
\mathcal{P}_{SRRS} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SIRS} \\
\mathcal{P}_{SSRR} &= \frac{1}{2\mathcal{R}+1}\mathcal{P}_{SSIR} + \frac{1}{\mathcal{R}+1}\mathcal{P}_{SSRI} \\
\mathcal{P}_{SRRR} &= \frac{1}{\mathcal{R}+1}(\mathcal{P}_{SRRI} + \mathcal{P}_{SIRR}) \\
\mathcal{P}_{RRSR} &= \frac{1}{\mathcal{R}+1}(\mathcal{P}_{RISR} + \mathcal{P}_{IRSR}) + \mathcal{P}_{RRSI} \\
\mathcal{P}_{RRRS} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{IRRS} + \mathcal{P}_{RRIS} \\
\mathcal{P}_{RRRR} &= \mathcal{P}_{IRRR} + \mathcal{P}_{RIRR} + \mathcal{P}_{RRRI}
\end{aligned}$$

To find the equations for the final size probabilities we evaluated the following:

$$\begin{aligned}
\mathbb{P}(\text{Final Size} = 1) &= \mathcal{P}_{RSSS} + \mathcal{P}_{SSSR} + \mathcal{P}_{SRSS} \\
\mathbb{P}(\text{Final Size} = 2) &= \mathcal{P}_{RSSR} + \mathcal{P}_{RRSS} + \mathcal{P}_{SRRS} + \mathcal{P}_{SSRR} \\
\mathbb{P}(\text{Final Size} = 3) &= \mathcal{P}_{SRRR} + \mathcal{P}_{RRSR} + \mathcal{P}_{RRRS} \\
\mathbb{P}(\text{Final Size} = 4) &= \mathcal{P}_{RRRR}
\end{aligned}$$

Simplifying the above we obtained the final size equations for the lollipop network with four nodes as shown in Table 2.2 on page 45. The probability mass function of the final size distribution for the lollipop network is shown in Figure 2.14.

2.2.2 Progression of infection over time

Equations describing the probability that the network is in a given state at time t for an SIR model on the lollipop network are:

Initial states:

$$\dot{\mathsf{P}}_{SSSS} = 0 \tag{2.2}$$

$$\begin{aligned}
\dot{\mathsf{P}}_{ISSS} &= -(3\mathcal{R}+1)\mathsf{P}_{ISSS} \\
\dot{\mathsf{P}}_{SISS} &= -(2\mathcal{R}+1)\mathsf{P}_{SISS} \\
\dot{\mathsf{P}}_{SSSI} &= -(\mathcal{R}+1)\mathsf{P}_{SSSI} \\
\dot{\mathsf{P}}_{SSII} &= -(3\mathcal{R}+2)\mathsf{P}_{SSII}
\end{aligned}$$

Transient states:

$$\begin{aligned}
\dot{\mathsf{P}}_{IISS} &= 2\mathcal{R}\mathsf{P}_{ISSS} + \mathcal{R}\mathsf{P}_{SISS} - (3\mathcal{R}+2)\mathsf{P}_{IISS} \\
\dot{\mathsf{P}}_{ISSI} &= \mathcal{R}(\mathsf{P}_{ISSS} + \mathsf{P}_{SSSI}) - 2(\mathcal{R}+1)\mathsf{P}_{ISSI} \\
\dot{\mathsf{P}}_{SIIS} &= \mathcal{R}\mathsf{P}_{SISS} - 2(\mathcal{R}+1)\mathsf{P}_{SIIS} \\
\dot{\mathsf{P}}_{RSSI} &= \mathsf{P}_{ISSI} - \mathsf{P}_{RSSI} \\
\dot{\mathsf{P}}_{ISSR} &= \mathsf{P}_{ISSI} - \mathsf{P}_{ISSR} \\
\dot{\mathsf{P}}_{RISS} &= \mathsf{P}_{IISS} - (\mathcal{R}+1)\mathsf{P}_{RISS} \\
\dot{\mathsf{P}}_{IIIS} &= 2\mathcal{R}(\mathsf{P}_{IISS} + \mathsf{P}_{SIIS}) - (\mathcal{R}+3)\mathsf{P}_{IIIS} \\
\dot{\mathsf{P}}_{SIII} &= \mathcal{R}\mathsf{P}_{SSII} - 3(\mathcal{R}+1)\mathsf{P}_{SIII} \\
\dot{\mathsf{P}}_{SSRI} &= \mathsf{P}_{SSII} - (\mathcal{R}+1)\mathsf{P}_{SSRI} \\
\dot{\mathsf{P}}_{SSIR} &= \mathsf{P}_{SSII} - (2\mathcal{R}+1)\mathsf{P}_{SSIR} \\
\dot{\mathsf{P}}_{SIRS} &= 2\mathsf{P}_{SIIS} - (\mathcal{R}+1)\mathsf{P}_{SIRS} \\
\dot{\mathsf{P}}_{IRSS} &= \mathsf{P}_{IISS} - (2\mathcal{R}+1)\mathsf{P}_{IRSS} \\
\dot{\mathsf{P}}_{IISI} &= 2\mathcal{R}(\mathsf{P}_{ISSI} + \mathsf{P}_{SSII}) + \mathcal{R}\mathsf{P}_{IISS} - (2\mathcal{R}+3)\mathsf{P}_{IISI}
\end{aligned}$$

$$\begin{aligned}
\dot{\mathsf{P}}_{SIIR} &= \mathcal{R}\mathsf{P}_{SSIR} + \mathsf{P}_{SIII} - 2(\mathcal{R}+1)\mathsf{P}_{SIIR} \\
\dot{\mathsf{P}}_{SIRR} &= 2\mathsf{P}_{SIIR} + \mathsf{P}_{SIRI} - (\mathcal{R}+1)\mathsf{P}_{SIRR} \\
\dot{\mathsf{P}}_{SIRI} &= 2\mathsf{P}_{SIII} - 2(\mathcal{R}+1)\mathsf{P}_{SIRI} \\
\dot{\mathsf{P}}_{RIIS} &= \mathsf{P}_{IIIS} + \mathcal{R}\mathsf{P}_{RISS} - 2\mathsf{P}_{RISS} \\
\dot{\mathsf{P}}_{IIII} &= \mathcal{R}(\mathsf{P}_{IIIS} + 2\mathsf{P}_{IISI} + 3\mathsf{P}_{SIII}) - 4\mathsf{P}_{IIII} \\
\dot{\mathsf{P}}_{RISI} &= \mathsf{P}_{IISI} - (\mathcal{R}+2)\mathsf{P}_{RISI} \\
\dot{\mathsf{P}}_{IRSI} &= \mathcal{R}(\mathsf{P}_{IRSS} + \mathsf{P}_{SSRI}) + \mathsf{P}_{IISI} - (\mathcal{R}+2)\mathsf{P}_{IRSI} \\
\dot{\mathsf{P}}_{IISR} &= \mathcal{R}\mathsf{P}_{SSIR} + \mathsf{P}_{IISI} - 2(\mathcal{R}+1)\mathsf{P}_{IISR} \\
\dot{\mathsf{P}}_{IIRS} &= 2\mathsf{P}_{IIIS} + \mathcal{R}(\mathsf{P}_{SIRS} + \mathsf{P}_{IRSS}) - (\mathcal{R}+2)\mathsf{P}_{IIRS} \\
\dot{\mathsf{P}}_{SRRI} &= \mathsf{P}_{SIRI} - (\mathcal{R}+1)\mathsf{P}_{SRRI} \\
\dot{\mathsf{P}}_{RIII} &= \mathsf{P}_{IIII} + \mathcal{R}\mathsf{P}_{RISI} - 3\mathsf{P}_{RIII} \\
\dot{\mathsf{P}}_{IIIR} &= \mathsf{P}_{IIII} + 2\mathcal{R}(\mathsf{P}_{SIIR} + \mathsf{P}_{IISR}) - 3\mathsf{P}_{IIIR} \\
\dot{\mathsf{P}}_{RRIS} &= 2\mathsf{P}_{RIIS} + \mathsf{P}_{IIRS} - \mathsf{P}_{RRIS} \\
\dot{\mathsf{P}}_{IRSR} &= \mathsf{P}_{IRSI} + \mathsf{P}_{IISR} - (\mathcal{R}+1)\mathsf{P}_{IRSR} \\
\dot{\mathsf{P}}_{RISR} &= \mathsf{P}_{IISR} + \mathsf{P}_{RSIR} - (\mathcal{R}+1)\mathsf{P}_{RISR} \\
\dot{\mathsf{P}}_{RRSI} &= \mathsf{P}_{IRSI} + \mathsf{P}_{RISI} - \mathsf{P}_{RRSI} \\
\dot{\mathsf{P}}_{IRRS} &= \mathsf{P}_{IIRS} - (\mathcal{R}+1)\mathsf{P}_{IRRS} \\
\dot{\mathsf{P}}_{IIRI} &= 2\mathcal{R}\mathsf{P}_{SIRI} + \mathcal{R}(\mathsf{P}_{IIRS} + \mathsf{P}_{IRSI}) + 2\mathsf{P}_{IIII} - 3\mathsf{P}_{IIRI} \\
\dot{\mathsf{P}}_{RRII} &= 2\mathsf{P}_{RIII} + \mathsf{P}_{IRII} - 2\mathsf{P}_{RRII} \\
\dot{\mathsf{P}}_{RRRI} &= \mathsf{P}_{RRII} + \mathsf{P}_{IRRI} - \mathsf{P}_{RRRI} \\
\dot{\mathsf{P}}_{RIIR} &= \mathsf{P}_{RIII} + \mathsf{P}_{IIIR} + \mathcal{R}\mathsf{P}_{RISR} - 2\mathsf{P}_{RIIR} \\
\dot{\mathsf{P}}_{IIRR} &= 2\mathsf{P}_{IIIR} + \mathcal{R}(\mathsf{P}_{IRSR} + \mathsf{P}_{SIRR}) - 2\mathsf{P}_{IIRR} + \mathsf{P}_{IIRI} \\
\dot{\mathsf{P}}_{IRRI} &= \mathcal{R}(\mathsf{P}_{SRRI} + \mathsf{P}_{IRRS}) + \mathsf{P}_{IIRI} - 2\mathsf{P}_{IRRI} \\
\dot{\mathsf{P}}_{RIRR} &= 2\mathsf{P}_{RIIR} + \mathsf{P}_{RRII} + \mathsf{P}_{IIRR} - \mathsf{P}_{RIRR} \\
\dot{\mathsf{P}}_{IRRR} &= \mathsf{P}_{IRRI} + \mathsf{P}_{IIRR} - \mathsf{P}_{IRRR}
\end{aligned}$$

Absorbing states:

$$\begin{aligned}
\dot{\mathsf{P}}_{RSSS} &= \mathsf{P}_{ISSS} \\
\dot{\mathsf{P}}_{SSSR} &= \mathsf{P}_{SSSI} \\
\dot{\mathsf{P}}_{SRSS} &= \mathsf{P}_{SISS} \\
\dot{\mathsf{P}}_{RSSR} &= \mathsf{P}_{ISSR} + \mathsf{P}_{RSSI} \\
\dot{\mathsf{P}}_{RRSS} &= \mathsf{P}_{RISS} + \mathsf{P}_{IRSS} \\
\dot{\mathsf{P}}_{SRRS} &= \mathsf{P}_{SIRS}
\end{aligned}$$

$$
\begin{aligned}
\dot{\mathsf{P}}_{SSRR} &= \mathsf{P}_{SSRI} + \mathsf{P}_{SSIR} \\
\dot{\mathsf{P}}_{SRRR} &= \mathsf{P}_{SIRR} + \mathsf{P}_{SRRI} \\
\dot{\mathsf{P}}_{RRSR} &= \mathsf{P}_{RISR} + \mathsf{P}_{RRSI} + \mathsf{P}_{IRSR} \\
\dot{\mathsf{P}}_{RRRS} &= \mathsf{P}_{RRIS} + \mathsf{P}_{IRRS} \\
\dot{\mathsf{P}}_{RRRR} &= \mathsf{P}_{IRRR} + \mathsf{P}_{RRRI} + \mathsf{P}_{RIRR}
\end{aligned}
$$

We numerically solve the system of Equations 2.2 (see Figure 2.6) to confirm our analytical expressions for the final epidemic size of the lollipop network

Table 2.2: Lollipop network final size PMFs

Initial State	$SSSI$	$SISS$	$ISSS$	$SSII$
$\mathbb{P}$(Final Size =1)	$\frac{1}{\mathcal{R}+1}$	$\frac{1}{2\,\mathcal{R}+1}$	$\frac{1}{3\,\mathcal{R}+1}$	0
$\mathbb{P}$(Final Size =2)	$\frac{\mathcal{R}}{(\mathcal{R}+1)\,(2\,\mathcal{R}+1)}$	$\frac{\mathcal{R}\,(3\,\mathcal{R}+2)}{(\mathcal{R}+1)^2(2\,\mathcal{R}+1)^2}$	$\frac{\mathcal{R}\,(\mathcal{R}+3)}{(\mathcal{R}+1)\,(2\,\mathcal{R}+1)\,(3\,\mathcal{R}+1)}$	$\frac{1}{(\mathcal{R}+1)\,(2\,\mathcal{R}+1)}$
$\mathbb{P}$(Final Size =3)	$\frac{2\,\mathcal{R}^2}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}$	$\frac{\mathcal{R}^2\,\left(4\,\mathcal{R}^2+10\,\mathcal{R}+5\right)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}$	$\frac{4\,\mathcal{R}^2\,(\mathcal{R}+2)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)\,(3\,\mathcal{R}+1)}$	$\frac{\mathcal{R}\,(2\,\mathcal{R}+3)}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}$
$\mathbb{P}$(Final Size =4)	$\frac{2\,\mathcal{R}^3\,(\mathcal{R}+2)}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}$	$\frac{\mathcal{R}^3\,\left(4\,\mathcal{R}^2+10\,\mathcal{R}+5\right)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}$	$\frac{2\,\mathcal{R}^3\,(3\,\mathcal{R}+5)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)\,(3\,\mathcal{R}+1)}$	$\frac{\mathcal{R}^2\,(2\,\mathcal{R}+3)\,(\mathcal{R}+2)}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}$
Expected FS	$4-\frac{8\,\mathcal{R}^3+21\,\mathcal{R}^2+14\,\mathcal{R}+3}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}$	$4-\frac{5\,\mathcal{R}^3+16\,\mathcal{R}^2+13\,\mathcal{R}+3}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}$	$4-\frac{4\,\mathcal{R}^2+9\,\mathcal{R}+3}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}$	$4-\frac{4\,\mathcal{R}^2+7\,\mathcal{R}+2}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}$

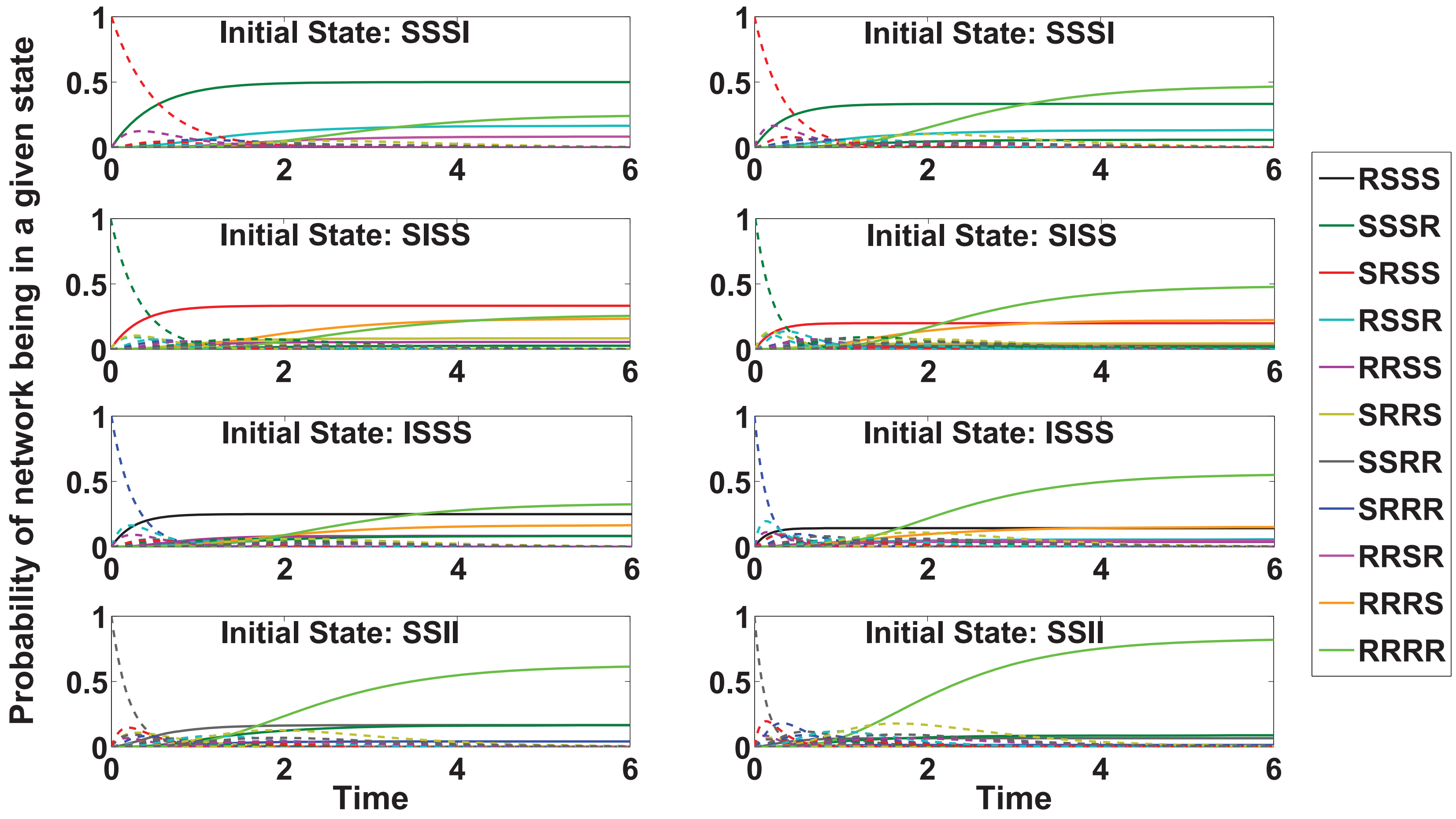


Figure 2.6: Numerical solution of the system of Equations (2.2) - (2.3) that describes the progression of infection over time for an SIR model on a line network with $N = 4$ nodes. Left and right columns contain graphical results for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively for the specified initial conditions. Dashed lines represent transient states and solid lines are the absorbing states. The numerical results are in agreement with the analytical expressions for the same set of initial conditions.

2.3 Epidemics on networks of three or four nodes

Here, we detail the Gillespie algorithm which was used to independently verify our analytic results. We then proceed to present and discuss results found for the SIR model on all small networks shown in Figure 2.1.

2.3.1 Stochastic Model

In previous sections the methods used to find and evaluate analytic results for an SIR model on small networks were detailed. In this section we will explain the methods used to check our results independently and confirm that they are correct. The Gillespie algorithm was used to simulate a stochastic SIR model on each of the small networks considered. We ran realisations of the stochastic model with the same initial conditions and infection parameters used to compute the analytic results. In the Gillespie algorithm, for each node, a random event time is generated by adding to the current time an exponentially distributed time step with rate either $i\beta$ (if node is in state S and where i is the number of infectious neighbours the node has) or γ (if node is in state I). Then, repeatedly, the node with the smallest time step is chosen, the current time step is updated to the event time of the chosen node, and the state of the chosen node is changed accordingly. The event times of the chosen node and all of its neighbours are then updated. This repetition continues until there are no more infectious nodes in the contact network. For the SIR model the epidemic ends when all nodes are in either state S or R.

The final size of the epidemic realisation is calculated by summing up the total number of nodes in the recovered class at the end of the epidemic. To determine the final size probabilities for given initial condition and infection parameters we ran 2×10^5 stochastic realisations on each network and calculated the frequency of each final size occurring.

2.3.2 Results

We analysed an SIR model on each of the small networks of three and four nodes (see Figure 2.1) following the method illustrated above for the triangle and lollipop networks. Here, we present the results from all networks considered and refer the reader to Appendix A for the full details of the analysis for each network. In Table 2.3 we give expressions for the expected final epidemic size for all networks of size $N = 3$ and $N = 4$ starting the epidemic with one infectious node. In Figures 2.7 and 2.8 we compare each of these expressions for the expected final size by plotting them over a range of $\mathcal{R}$ values. It can be seen clearly in Figure 2.7 that the expected final size for the triangle network is always higher than the expected final size of the line

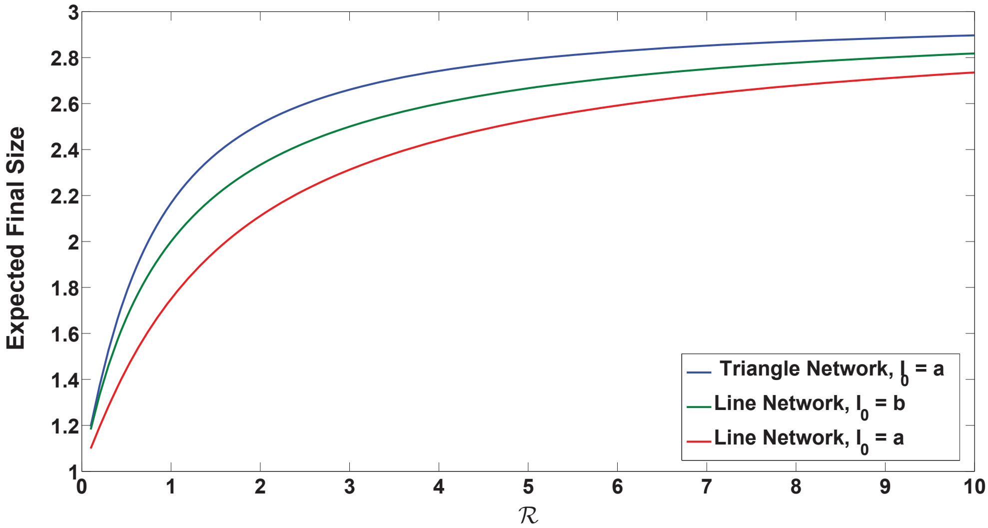


Figure 2.7: Expected final size functions of $\mathcal{R}$ for all networks of size $N = 3$; these were evaluated for the range of values $\mathcal{R} = 0.1$ to $\mathcal{R} = 10$. Each SIR epidemic started with one infectious node, $\mathbf{I}_0$, as specified.

network with $N = 3$ nodes starting with one infectious node. This result is expected as the triangle network is better connected than the line network; for example for node c to become infected in the triangle network node a or b first needs to be infected, however for node c to become infected in the line network node b must first be infected. Similarly in Figure 2.8 the expected final size for the complete network is always higher than that for every other network of size $N = 4$ for all values of $\mathcal{R}$. From Figure 2.8 we can also see that there are six intersections that occur when two expressions for the expected final size, of two different networks, are equal. The value of $\mathcal{R}$ where these intersections occur can be found analytically by equating the appropriate expressions found in Table 2.3 and solving for $\mathcal{R}$. Table 2.4 summarises these results. In Tables 2.3 and 2.4 we use I_0 to denote the initial infectious node. The expressions for the expected final size of the complete network and the line network of $N = 4$ nodes do not intersect with expressions for any other network for all values of $\mathcal{R}$.

Figures 2.9 to 2.14 show the probability mass functions for the final epidemic size of each network considered. For each network there is a probability mass function corresponding to each initial state. In Table 2.5 we give the expected final size corresponding to each of the probability mass functions in Figures 2.9 to 2.14. Table 2.5 also shows the degree of the initial infectious node(s) and clustering coefficient for each of the small networks of three and four nodes considered. Not surprisingly, the triangle network has a clustering coefficient of $\phi = 1$ and all complete networks

Table 2.3: Expressions for the Expected Final Size ($\mathbb{E}$[Final Size]) of an SIR epidemic starting with one infectious node.

Networks with $N = 3$			
	Initial state and initial infectious node, $\mathbf{I}_0$	**Degree of $\mathbf{I}_0$**	$\mathbb{E}$[**Final Size**]
A. Triangle	SSI ($\mathrm{I}_0 = a$)	2	$3 - \frac{2}{\mathcal{R}+1} + \frac{2\mathcal{R}^2}{(\mathcal{R}+1)(2\mathcal{R}+1)}$
B. Line	SSI ($\mathrm{I}_0 = a$ or c)	1	$3 - \frac{2}{\mathcal{R}+1} - \frac{\mathcal{R}}{(\mathcal{R}+1)^2}$
	SIS ($\mathrm{I}_0 = b$)	2	$3 - \frac{2}{\mathcal{R}+1}$
Networks with $N = 4$			
C. Complete	$SSSI$ ($\mathrm{I}_0 = a$, b, c or d)	3	$4 - \frac{3(4\mathcal{R}^5 + 18\mathcal{R}^4 + 37\mathcal{R}^3 + 29\mathcal{R}^2 + 9\mathcal{R} + 1)}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)^2\,(3\,\mathcal{R}+1)}$
D. Square	$SSSI$ ($\mathrm{I}_0 = a$, b, c or d)	2	$4 - \frac{3\mathcal{R}^3 + 17\mathcal{R}^2 + 13\mathcal{R} + 3}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$
E. Star	$SSSI$ ($\mathrm{I}_0 = b$, c or d)	1	$4 - \frac{3}{\mathcal{R}+1} - \frac{2\mathcal{R}}{(\mathcal{R}+1)^2}$
	$ISSS$ ($\mathrm{I}_0 = a$)	3	$4 - \frac{3}{\mathcal{R}+1}$
F. Toast	$SSIS$ ($\mathrm{I}_0 = b$ or c)	2	$4 - \frac{6\,\mathcal{R}^4 + 29\,\mathcal{R}^3 + 4\,\mathcal{R}^2 + 19\,\mathcal{R} + 3}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}$
	$SSSI$ ($\mathrm{I}_0 = a$ or d)	3	$4 - \frac{6\,\mathcal{R}^4 + 33\,\mathcal{R}^3 + 47\,\mathcal{R}^2 + 21\,\mathcal{R} + 3}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)\,(3\,\mathcal{R}+1)}$
G. Line	$SSSI$ ($\mathrm{I}_0 = a$ or d)	1	$4 - \frac{4\mathcal{R}+3}{(\mathcal{R}+1)^2} - \frac{\mathcal{R}(2\mathcal{R}+1)}{(\mathcal{R}+1)^3}$
	$SSIS$ ($\mathrm{I}_0 = b$ or c)	2	$4 - \frac{4\mathcal{R}+3}{(\mathcal{R}+1)^2}$
H. Lollipop	$SSSI$ ($\mathrm{I}_0 = d$)	1	$4 - \frac{8\,\mathcal{R}^3 + 21\,\mathcal{R}^2 + 14\,\mathcal{R} + 3}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}$
	$SISS$ ($\mathrm{I}_0 = b$ or c)	2	$4 - \frac{5\,\mathcal{R}^3 + 16\,\mathcal{R}^2 + 13\,\mathcal{R} + 3}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}$
	$ISSS$ ($\mathrm{I}_0 = a$)	3	$4 - \frac{4\,\mathcal{R}^2 + 9\,\mathcal{R} + 3}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}$

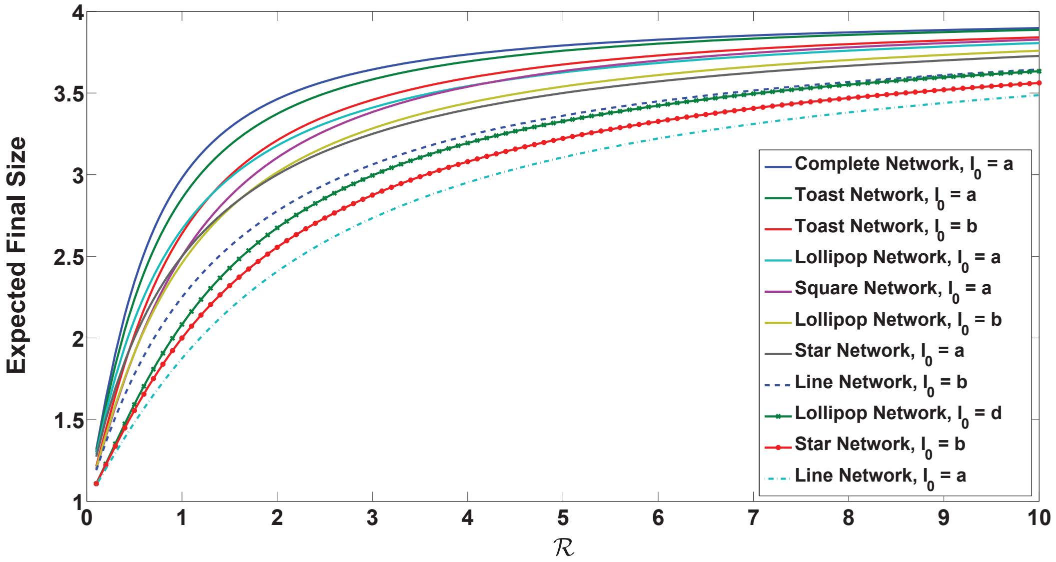


Figure 2.8: Expected final size functions of $\mathcal{R}$ for all networks of size $N = 4$; these were evaluated for the range of values $\mathcal{R} = 0.1$ to $\mathcal{R} = 10$. Each SIR epidemic started with one infectious node, $\mathbf{I}_0$, as specified.

Table 2.4: Intersections of Expected Final Size expressions.

$\mathbb{E}$[**Final Size**] **of this network** >	$\mathbb{E}$[**Final Size**] **of this network**	**for** $\mathcal{R} >$
Toast network ($\mathbf{I_0} = b$ or c)	Star network ($\mathbf{I_0} = a$)	0.2841
Toast network ($\mathbf{I_0} = b$ or c)	Lollipop network ($\mathbf{I_0} = a$)	1.3131
Square network ($\mathbf{I_0} = a, b, c$ or d)	Lollipop network ($\mathbf{I_0} = a$)	$2 + \sqrt{5} \approx 4.2361$
Square network ($\mathbf{I_0} = a, b, c$ or d)	Lollipop network ($\mathbf{I_0} = b$ or c)	$\frac{1}{2}$
Lollipop network ($\mathbf{I_0} = b$ or c)	Star network ($\mathbf{I_0} = a$)	$\frac{1+\sqrt{5}}{2} \approx 1.6180$
Square network ($\mathbf{I_0} = a, b, c$ or d)	Star network ($\mathbf{I_0} = a$)	1

have a clustering coefficient of $\phi = 1$. All line networks have a clustering coefficient of $\phi = 0$ as there are no closed loops of any size in a line network. Note that the clustering coefficient counts the number of closed loops of order 3 in the network and does not take into account closed loops of order 4 (or higher) which are present in the complete, square and toast networks with $N = 4$ nodes.

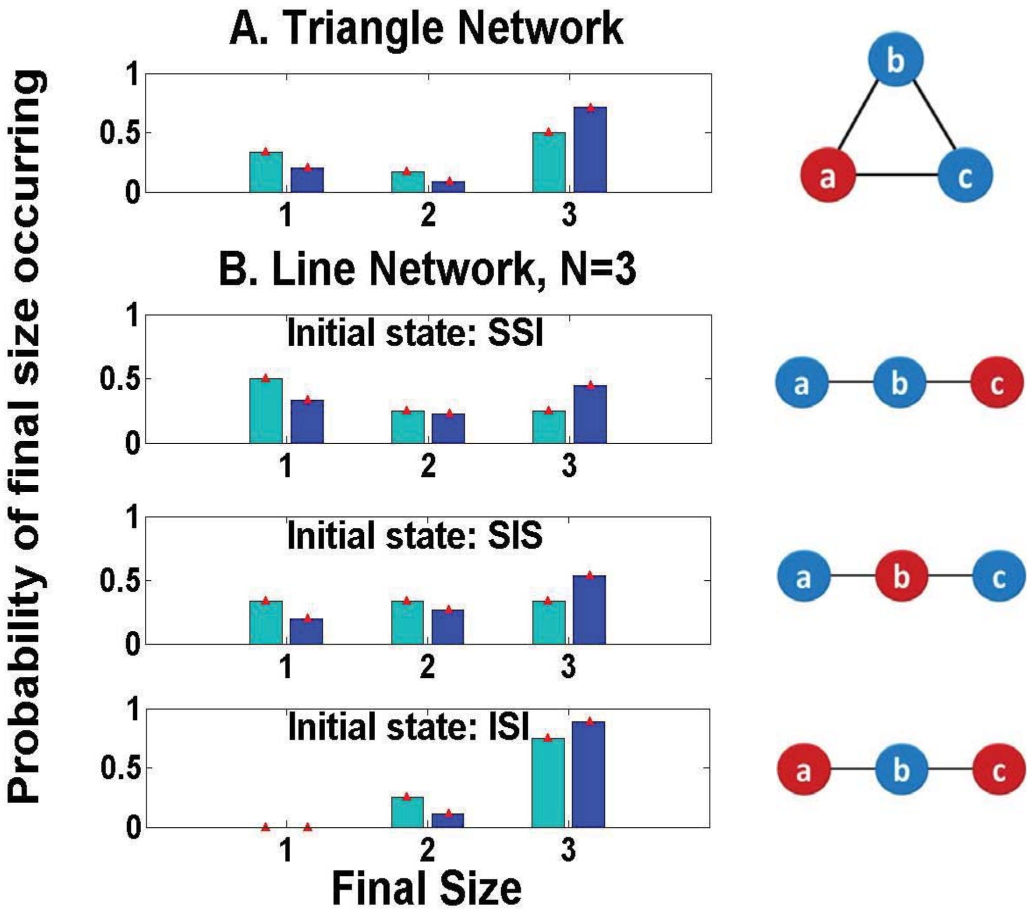


Figure 2.9: Probability mass functions for the triangle and line networks with $N = 3$. The PMF's in light and dark blue are for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively. Red triangles represent the corresponding final size probabilities found from stochastic realisations. Schematics of the initial state are shown on the right.

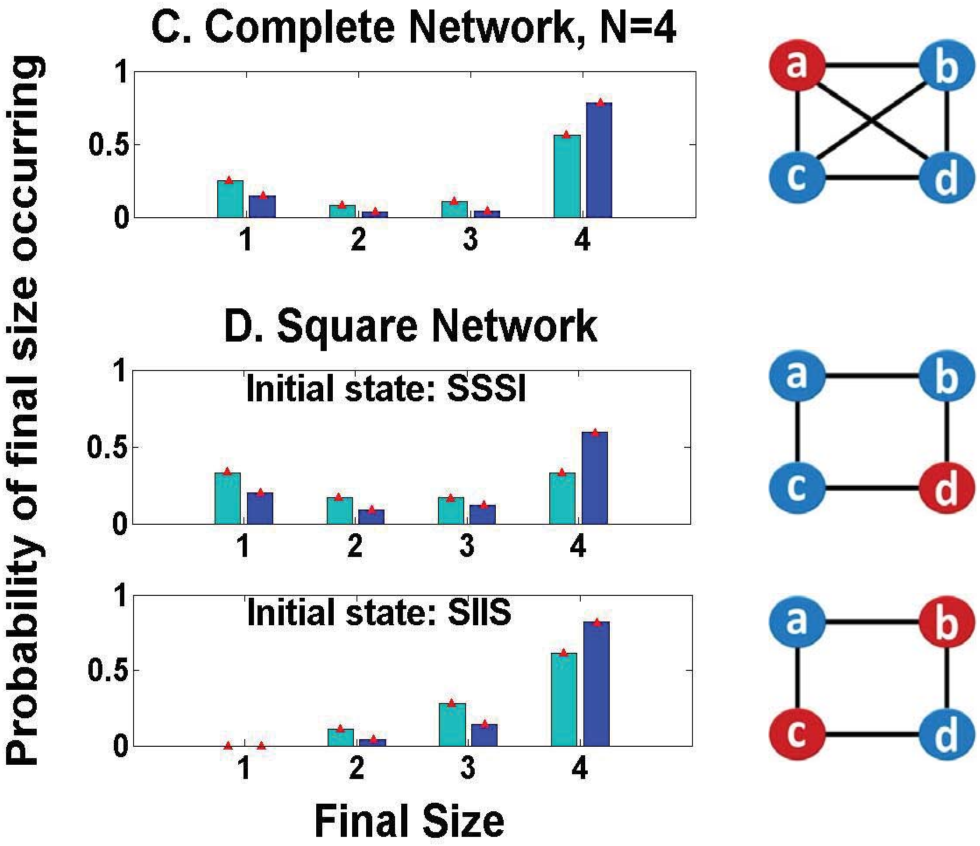


Figure 2.10: Probability mass functions for the complete and square networks with $N = 4$. The PMF's in light and dark blue are for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively. Red triangles represent the corresponding final size probabilities found from stochastic realisations. Schematics of the initial state are shown on the right.

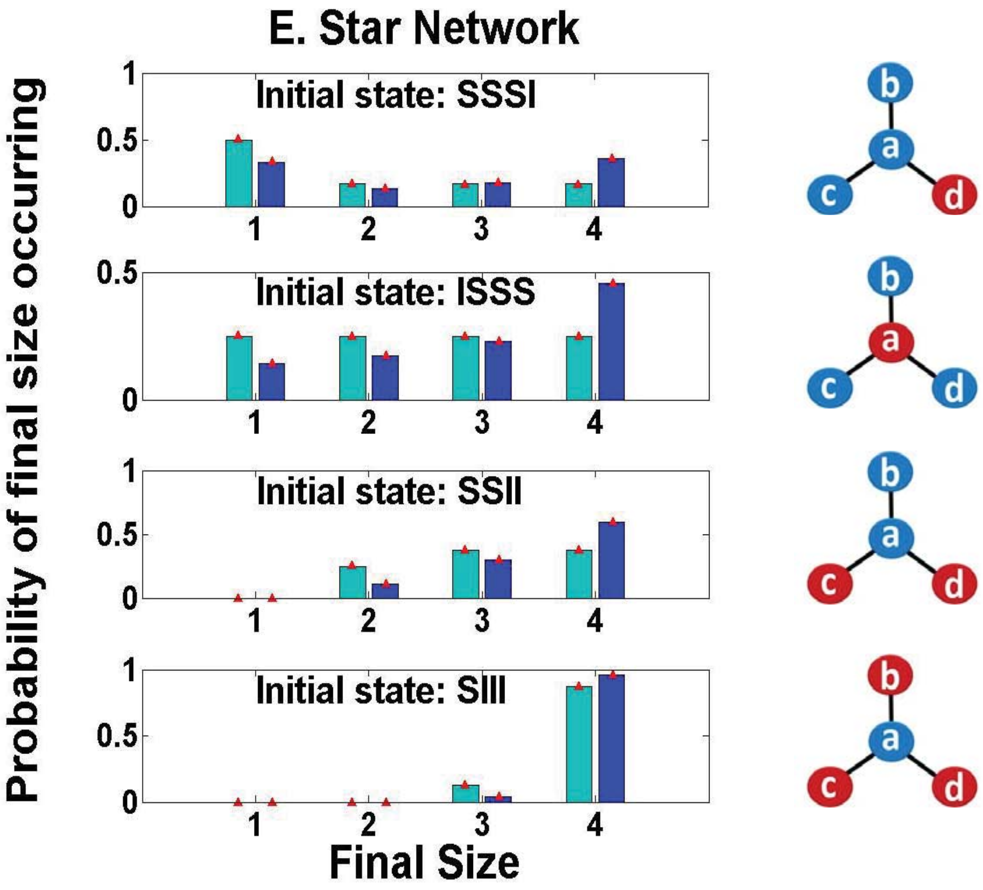


Figure 2.11: Probability mass functions for the star network with $N = 4$. The PMF's in light and dark blue are for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively. Red triangles represent the corresponding final size probabilities found from stochastic realisations. Schematics of the initial state are shown on the right.

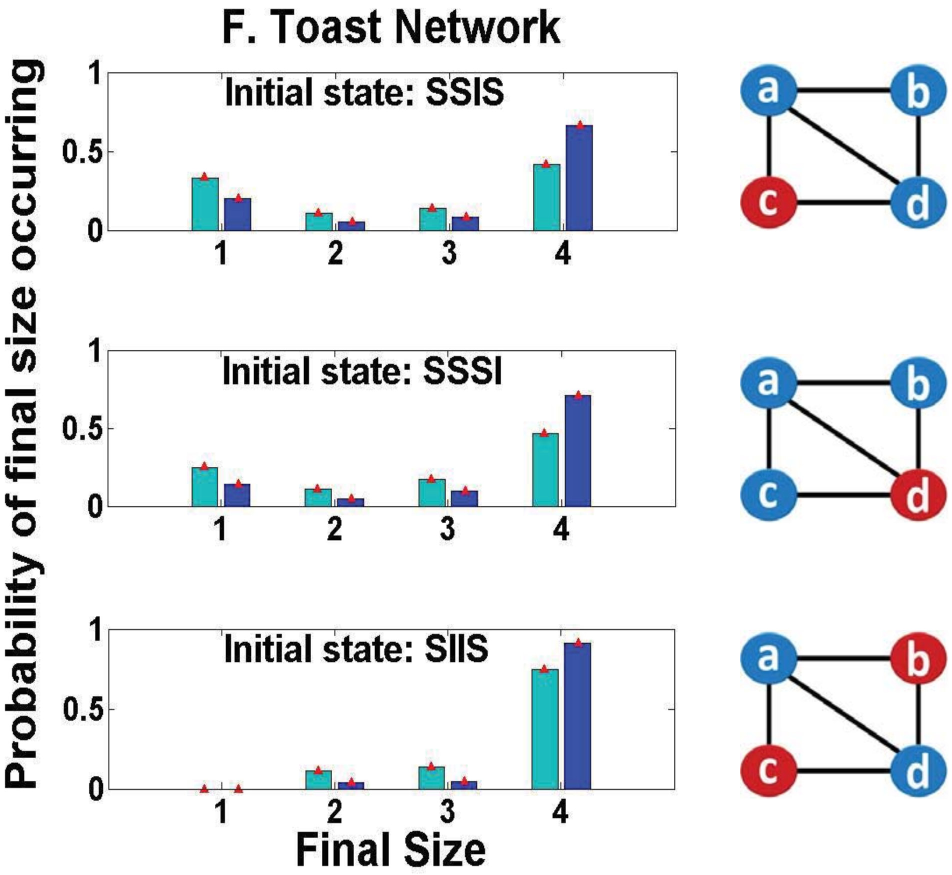


Figure 2.12: Probability mass functions for the toast network with $N = 4$. The PMF's in light and dark blue are for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively. Red triangles represent the corresponding final size probabilities found from stochastic realisations. Schematics of the initial state are shown on the right.

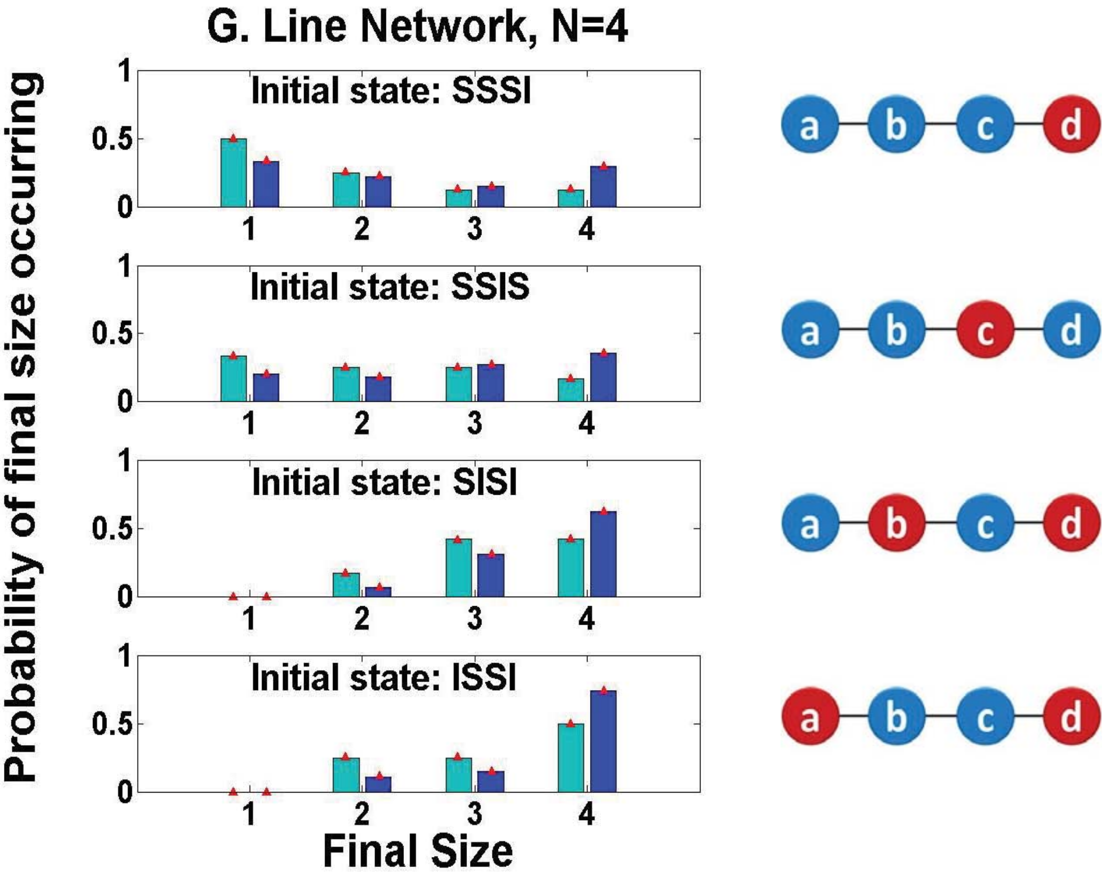


Figure 2.13: Probability mass functions for the line network with $N = 4$. The PMF's in light and dark blue are for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively. Red triangles represent the corresponding final size probabilities found from stochastic realisations. Schematics of the initial state are shown on the right.

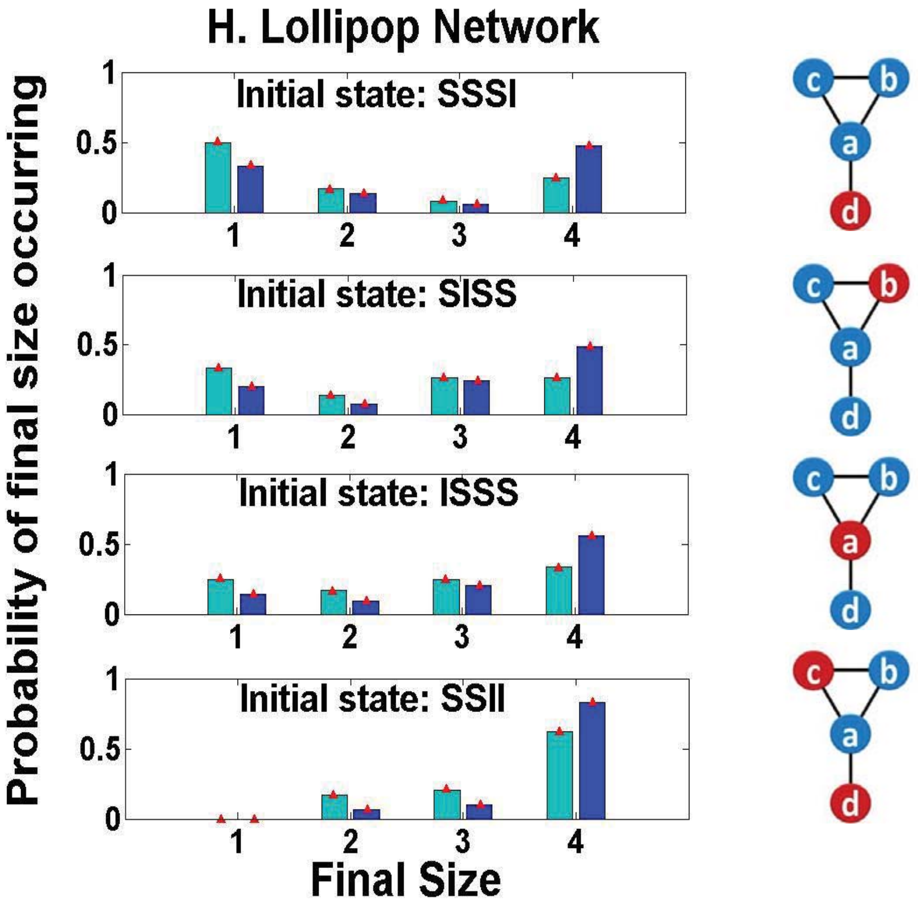


Figure 2.14: Probability mass functions for the lollipop network with $N = 4$. The PMF's in light and dark blue are for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively. Red triangles represent the corresponding final size probabilities found from stochastic realisations. Schematics of the initial state are shown on the right.

Table 2.5: Expected Final Size for an SIR epidemic and network clustering coefficients

Networks with $N = 3$				$\mathbb{E}$[**Final Size**]	
	Clustering Coefficient, ϕ	**Initial state**	**Degree of Initial Infectious Node(s)**	$\mathcal{R} = 1$	$\mathcal{R} = 2$
A. Triangle	1	SSI	2	2.17	2.51
B. Line	0	SSI	1	1.75	2.11
		SIS	2	2	2.33
		ISI	1 & 1	2.75	2.89
Networks with $N = 4$					
C. Complete	1	$SSSI$	3	2.98	3.46
D. Square	0	$SSSI$	2	2.5	3.1
		$SIIS$	2 & 2	3.5	3.78
E. Star	0	$SSSI$	1	2	2.56
		$ISSS$	3	2.5	3
		$SSII$	1 & 1	3.125	3.48
		$SIII$	1, 1 & 1	3.875	3.96
F. Toast	0.75	$SSIS$	2	2.64	3.21
		$SSSI$	3	2.85	3.37
		$SIIS$	2 & 2	3.64	3.87
G. Line	0	$SSSI$	1	1.88	2.41
		$SSIS$	2	2.25	2.78
		$SISI$	2 & 1	3.25	3.56
		$ISSI$	1 & 1	3.25	3.63
H. Lollipop	0.6	$SSSI$	1	2.08	2.67
		$SISS$	2	2.46	3.02
		$ISSS$	3	2.67	3.18
		$SSII$	2 & 1	3.46	3.76

2.4 Discussion

We have derived exact analytical expressions for the probability mass functions of the final epidemic size on eight small networks and investigated the effect that network structure and the degree of the initial infectious node has on the spread of an infection. Increasing the complexity of the network structure reduced the effect of grouping states together based on symmetries of the network. Consequently this increased the complexity of the analysis which is evident in the probability mass function expressions and is due to the increasing number of possible infection paths. The results presented here form the basis for finding tractable analytic results which describe the spread of an infection through large networks which are composed of the small networks discussed in this chapter. We calculated the clustering coefficient, ϕ, for each of the small networks in order to see if any correlation between ϕ and the final size of the epidemic emerged. The expected final size for the triangle network is higher than that for the line network of $N = 3$ nodes for all parameter values, see Appendix A for analytical proof. Small networks with a higher clustering coefficient had a higher expected final size when the epidemic was started with one infectious node. Although the square, line and star networks with $N = 4$ nodes all have $\phi = 0$; our results show that the square network has a higher expected final size than the star and line networks for starting the epidemic with an I_0 of the same degree for all values of $\mathcal{R}$. By looking at the topology of the networks this is intuitive as the square network is better connected, due to the closed loop of order 4. This reinforces the well known result that when comparing how an infection spreads on different networks we must consider more than just the clustering coefficient. The intersections of the final size curves in Figure 2.8 illustrate that at the beginning of the epidemic $\mathcal{R}$ needs to be higher in some networks for the infection to spread due to the low connectivity of the network. The probability that each node in a network will acquire infection at some point during the epidemic increased when the degree of the initial infectious node or the transmission parameter (β, and hence $\mathcal{R}$) was increased. Similarly, as expected, the probability that the infection would die out before infecting an initially susceptible node was higher when the degree of the initial infectious node was 1 when compared with initial infectious nodes of degree 2 and 3. Our results give us a good indication that both the network topology and degree of the initial infectious node are key factors in understanding how an infection might spread through small networks which could represent small populations or communities. Some of the results presented here are intuitive; however, we emphasize the importance of having found exact analytic expressions for the probability mass functions of the final epidemic size of the small networks. Having the probability mass function for the final epidemic size provides us with more detail than a single expression for the expected final size of an epidemic.

Chapter 3

An SIR model on a Line of Triangles Network.

In this chapter we investigate an SIR model on a larger network, in particular we focus on the line of triangles (LoT) network shown in Figure 3.1. We illustrate how we can use the results from small networks (see Chapter 2) to analytically describe how an infection spreads through a LoT network. The key here is to correctly decompose the LoT network into an appropriate assemblage of small networks (see Figure 2.1, page 30) so that the results are exact. The decomposition of the larger network into the correct assemblage of smaller networks to generalise the probability mass function of the final epidemic size is crucial. For example, if we were to join two triangle networks together to form a larger network there are three ways we could do this; the first way is by joining the two triangle networks at an apex node, this would create a bow tie network of 5 nodes. We could also join two triangle networks together by allowing them to share two nodes, this would create a toast network (see Figure 2.1 F). Thirdly, we could have two triangle networks connected by one edge which would create a line of triangles network of 6 nodes. Connecting small networks together with overlapping edges becomes problematic, as discussed in Kiss et al [37], when trying to generalise results based on those already found for the small networks. Therefore, we take the third approach and start by investigating how an infection can spread through a LoT composed of $N = 6$ nodes.

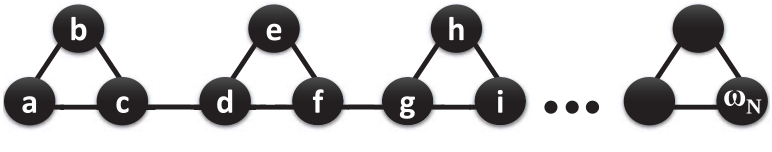


Figure 3.1: Schematic of a line of triangles network.

In Section 3.1 we derive the final size probability mass functions, the probability each node in the network ever gets infected, paths of infection from node a to ω_N

(where ω_N is the last node in a LoT network with N nodes, as illustrated in Figure 3.1) and the probability the infection ends at a given node. In Section 3.2 we generalise our results and show how to determine the probability the infection ends at a given node and the final size probability mass functions for a line of triangles network with N nodes. We conclude this chapter with a discussion of our results in Section 3.3. In our expressions throughout this chapter we continue to use the parameter $\mathcal{R} = \beta/\gamma$. In this chapter we derive results for two initial conditions, where the initial infectious node is denoted by $\mathbf{I_0}$. Figure 3.2 shows a schematic diagram of the two initial conditions, $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$, that are considered.

3.1 Line of Triangles with $N = 6$ nodes

In this section we present results found for the spread of an infection through the line of triangles network with $N = 6$ nodes, denoted LoT(6) for convenience. The derivation of results in the following subsections is aimed at understanding how infections can spread through networks, with the goal of analytically describing how an infection can spread through a line of triangles network of arbitrary size, N.

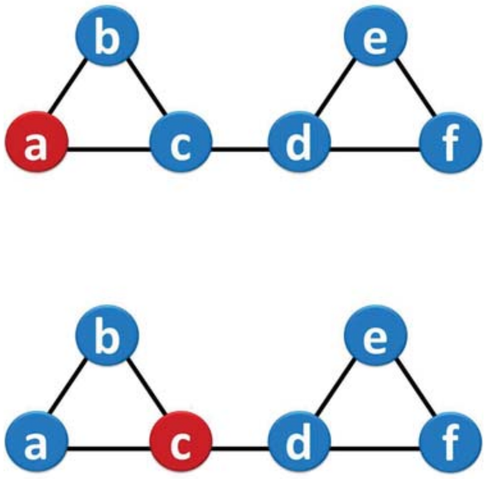


Figure 3.2: Initial conditions for the Line of Triangles Network with $N = 6$ nodes. Top: $\mathbf{I_0} = a$. Bottom: $\mathbf{I_0} = c$.

3.1.1 Probability mass function for the final epidemic size

For an SIR model on the LoT(6) network there are $3^N = 3^6 = 729$ possible states in which the network can be. Similarly to our analysis of the SIR model on small networks, we derive the set of equations that describe the transition probabilities for entering and leaving each state in the transition diagram. We do not present the transition diagram or equations here. From these we can find all the possible infection paths to each of the absorbing states of the network. We are able to find analytic expressions for the final epidemic size of the SIR model on the LoT(6) network by summing up the appropriate absorbing state probabilities. In Table 3.1 we give the expressions for the final size probability mass functions for the LoT(6) network with $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$. The expression for the expected final size is also given in Table 3.1. In Figure 3.3 we compare the expected final size of the LoT(6) network for $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$ by plotting them over a range of $\mathcal{R}$ values. We can see that the expected final size of the LoT(6) network with $\mathbf{I_0} = c$ is always higher than it is for the LoT(6) network with $\mathbf{I_0} = a$. This result was expected as node c has a higher degree than node a. Figure 3.4 shows the probability mass functions for the final epidemic size from Table 3.1 evaluated for $\mathcal{R} = 1$ and $\mathcal{R} = 2$.

We independently verify our analytic results by comparing them with results found using stochastic realisations. The Gillespie algorithm was used to generate the stochastic results with specified initial infectious node, transmission and recovery parameters. We ran 2×10^5 realisations of an SIR epidemic on the line of triangles network and for each realisation we recorded the total number of nodes that became infected at some point during the epidemic. To get the final size probabilities we calculated the frequency of obtaining a final size of one, two, three, four, five and six nodes infected over the 2×10^5 simulated epidemics.

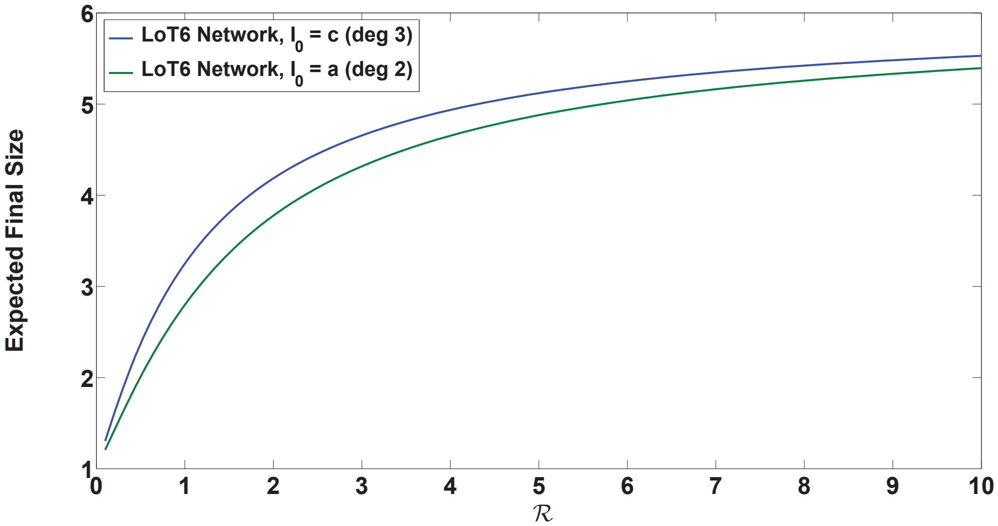


Figure 3.3: Expected final size functions of $\mathcal{R}$ for the LoT(6) network with $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$. See Table 3.1 for the expressions; these were evaluated for the range of values $\mathcal{R} = 0.1$ to $\mathcal{R} = 10$.

Table 3.1: Final size probability mass functions for LoT(6) with $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$.

Initial State	$\mathbf{I_0}$ = **node** a	$\mathbf{I_0}$ = **node** c
$\mathbb{P}$**(Final Size =1)**	$\frac{1}{2\mathcal{R}+1}$	$\frac{1}{3\mathcal{R}+1}$
$\mathbb{P}$**(Final Size =2)**	$\frac{\mathcal{R}(3\mathcal{R}+2)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)^2}$	$\frac{\mathcal{R}(5\mathcal{R}+3)}{(\mathcal{R}+1)(2\mathcal{R}+1)^2(3\mathcal{R}+1)}$
$\mathbb{P}$**(Final Size =3)**	$\frac{\mathcal{R}^2\left(8\mathcal{R}^3+22\mathcal{R}^2+18\mathcal{R}+5\right)}{(\mathcal{R}+1)^3(2\mathcal{R}+1)^3}$	$\frac{2\mathcal{R}^2\left(4\mathcal{R}^2+6\mathcal{R}+5\right)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)^2(3\mathcal{R}+1)}$
$\mathbb{P}$**(Final Size =4)**	$\frac{\mathcal{R}^3\left(4\mathcal{R}^3+14\mathcal{R}^2+15\mathcal{R}+7\right)}{(\mathcal{R}+1)^4(2\mathcal{R}+1)^3}$	$\frac{2\mathcal{R}^3\left(4\mathcal{R}^3+15\mathcal{R}^2+18\mathcal{R}+11\right)}{(\mathcal{R}+1)^4(2\mathcal{R}+1)^2(3\mathcal{R}+1)}$
$\mathbb{P}$**(Final Size =5)**	$\frac{2\mathcal{R}^4\left(5\mathcal{R}^2+13\mathcal{R}+7\right)}{(\mathcal{R}+1)^5(2\mathcal{R}+1)^3}$	$\frac{4\mathcal{R}^4(5\mathcal{R}+9)}{(\mathcal{R}+1)^4(2\mathcal{R}+1)^2(3\mathcal{R}+1)}$
$\mathbb{P}$**(Final Size =6)**	$\frac{2\mathcal{R}^5(\mathcal{R}+2)(4\mathcal{R}^2+10\mathcal{R}+5)}{(\mathcal{R}+1)^5(2\mathcal{R}+1)^3}$	$\frac{4\mathcal{R}^5(\mathcal{R}+2)(3\mathcal{R}+5)}{(\mathcal{R}+1)^4(2\mathcal{R}+1)^2(3\mathcal{R}+1)}$
Expected FS	$\frac{\left(6\mathcal{R}^3+13\mathcal{R}^2+6\mathcal{R}+1\right)\left(4\mathcal{R}^4+11\mathcal{R}^3+10\mathcal{R}^2+5\mathcal{R}+1\right)}{(\mathcal{R}+1)^5(2\mathcal{R}+1)^2}$	$\frac{\left(6\mathcal{R}^3+13\mathcal{R}^2+6\mathcal{R}+1\right)}{(\mathcal{R}+1)^3}$

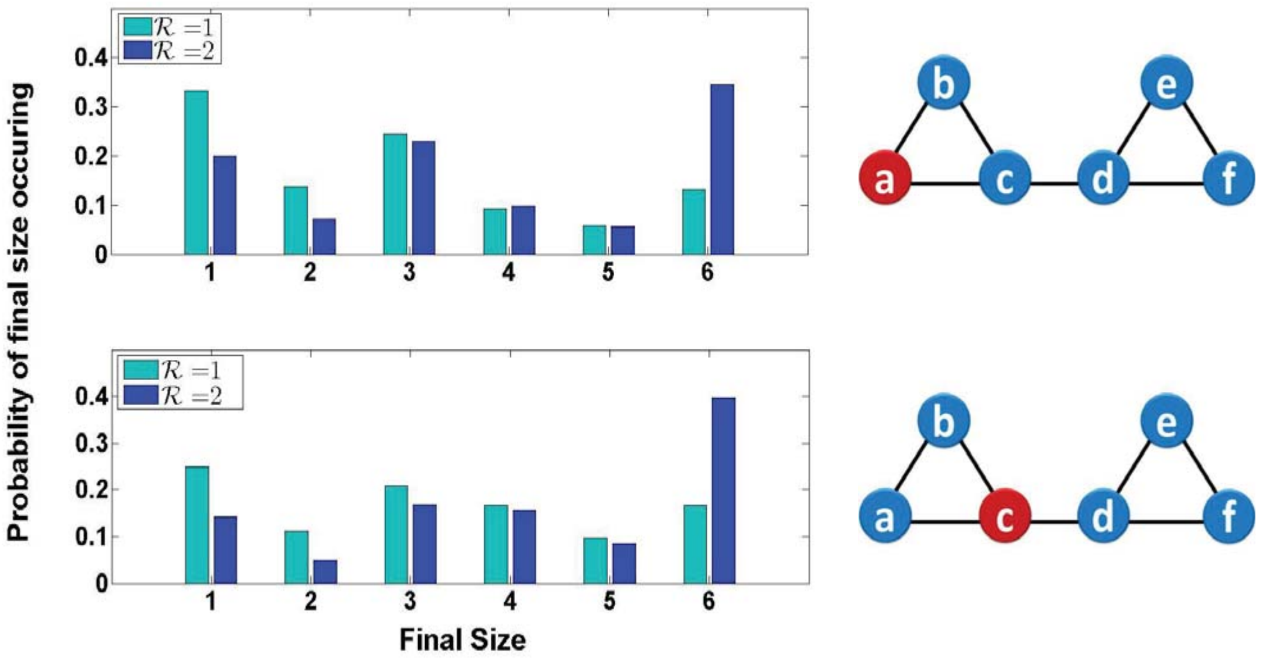


Figure 3.4: Final size probability mass function (PMF) for the LoT(6) network evaluated with $\mathcal{R} = 1$ and $\mathcal{R} = 2$. Initial conditions specified in diagrams on the right. The stochastic final size results were indistinguishable from the analytic results.

3.1.2 The probability that node i ever gets infected in a LoT(6) network

Here we derive the probability that node i is ever infected in the LoT(6) network, denoted by $P(i)$, where $i = a, b, c, d, e, f$. These probabilities were calculated by adding together the probability of reaching each absorbing state in which node i was in the recovered, R, class. The method is similar to how the probability of the final size in Sections 2.1.1 and 2.2.1 were derived.

For example, to find the probability that node b ever gets infected in the LoT(6) network we do the following. From the equations which describe the transition probabilities (not given as there are 729 equations) for the SIR model on a LoT(6) network we can obtain the probability that the network ends up in a given absorbing state at the end of the epidemic. We can then use this information to determine the probability that any node is ever infected. In the following equation $\mathcal{P}_{UVWXYZ}$ is the probability that the network is ever in state $UVWXYZ$ where U, V, W, X, Y and Z denote the infection state (S, I or R) that nodes a, b, c, d, e and f are in respectively. Note that all expressions for $\mathcal{P}_{UVWXYZ}$ are derived in the same way as those derived for the triangle and lollipop networks in Chapter 2. Therefore, the probability node b is ever infected is as follows.

$$\begin{aligned}
P(b) &= \mathcal{P}_{SRSSSS} + \mathcal{P}_{RRSSSS} + \mathcal{P}_{SRRSSS} + \mathcal{P}_{SRSRSS} + \mathcal{P}_{SRSSRS} \\
&\quad + \mathcal{P}_{SRSSSR} + \mathcal{P}_{RRRSSS} + \mathcal{P}_{SRRRSS} + \mathcal{P}_{SRRSSR} + \mathcal{P}_{SRSRRS} \\
&\quad + \mathcal{P}_{SRRSRS} + \mathcal{P}_{SRSRSR} + \mathcal{P}_{RRSSSR} + \mathcal{P}_{SRSSRR} + \mathcal{P}_{RRSSRS} \\
&\quad + \mathcal{P}_{RRSRSS} + \mathcal{P}_{RRRRSS} + \mathcal{P}_{RRSSRR} + \mathcal{P}_{RRRSSR} + \mathcal{P}_{RRRSRS} \\
&\quad + \mathcal{P}_{SRSRRR} + \mathcal{P}_{RRSRRS} + \mathcal{P}_{SRRSRR} + \mathcal{P}_{RRSRSR} + \mathcal{P}_{SRRRRS} \\
&\quad + \mathcal{P}_{SRRRSR} + \mathcal{P}_{RRRRRS} + \mathcal{P}_{SRRRRR} + \mathcal{P}_{RRRRSR} + \mathcal{P}_{RRSRRR} \\
&\quad + \mathcal{P}_{RRRSRR} + \mathcal{P}_{RRRRRR} \\
&= \frac{\mathcal{R}\left(2\mathcal{R}^2 + 4\mathcal{R} + 1\right)}{(\mathcal{R}+1)^2\,(2\mathcal{R}+1)} \qquad (3.1)
\end{aligned}$$

The expressions for the probability that each node in the LoT(6) network ever gets infected are given in Table 3.2. In Figure 3.5 we plot the results of these probabilities evaluated with $\mathcal{R} = 1$ and $\mathcal{R} = 2$.

Table 3.2: Probability each node ever gets infected in the LoT(6) network.

Initial State	$\mathbf{I_0}$ **= node** a	$\mathbf{I_0}$ **= node** c
$P(a)$	1	$\frac{\mathcal{R}\left(2\mathcal{R}^2+4\mathcal{R}+1\right)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$
$P(b)$	$\frac{\mathcal{R}\left(2\mathcal{R}^2+4\mathcal{R}+1\right)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$	$\frac{\mathcal{R}\left(2\mathcal{R}^2+4\mathcal{R}+1\right)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$
$P(c)$	$\frac{\mathcal{R}\left(2\mathcal{R}^2+4\mathcal{R}+1\right)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$	1
$P(d)$	$\frac{\mathcal{R}^2\left(2\mathcal{R}^2+4\mathcal{R}+1\right)}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$	$\frac{\mathcal{R}}{(\mathcal{R}+1)}$
$P(e)$	$\frac{\mathcal{R}^3\left(2\mathcal{R}^2+4\mathcal{R}+1\right)^2}{(\mathcal{R}+1)^5(2\mathcal{R}+1)^2}$	$\frac{\mathcal{R}^2\left(2\mathcal{R}^2+4\mathcal{R}+1\right)}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$
$P(f)$	$\frac{\mathcal{R}^3\left(2\mathcal{R}^2+4\mathcal{R}+1\right)^2}{(\mathcal{R}+1)^5(2\mathcal{R}+1)^2}$	$\frac{\mathcal{R}^2\left(2\mathcal{R}^2+4\mathcal{R}+1\right)}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$

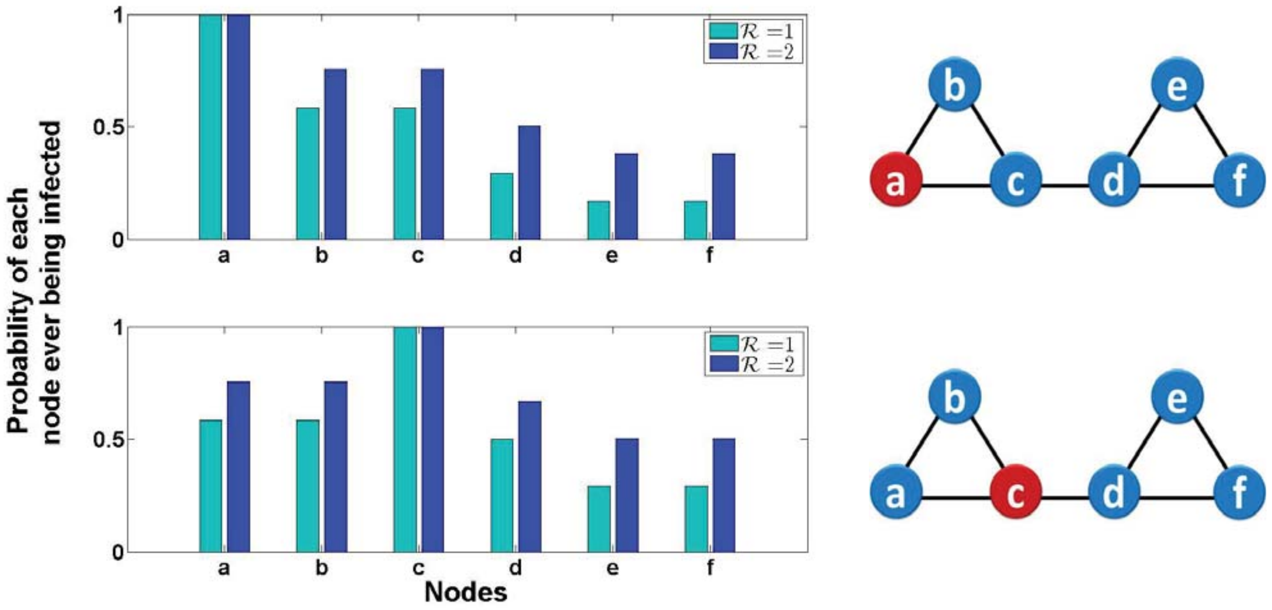


Figure 3.5: Probability each node is ever infected during an epidemic on the LoT(6) network. Initial conditions are shown on the right.

Given that $\mathbf{I_0} = a$, $P(b) = P(c)$ and $P(e) = P(f)$ due to the symmetry of these nodes in the network. Similarly, for $\mathbf{I_0} = c$ we have $P(a) = P(b)$ and $P(e) = P(f)$. Furthermore, we note that the probabilities $P(b)$ and $P(c)$ for $\mathbf{I_0} = a$ are equivalent to $P(a)$ and $P(b)$ for $\mathbf{I_0} = c$. To verify our analytical results (given in Table 3.2) we evaluated them for different values of $\mathcal{R}$ and then compared them with results obtained from stochastic realisations of an SIR epidemic on the same network with identical initial condition and parameter values. To obtain the stochastic results we did the following. After each stochastic epidemic realisation a list of the recovered nodes was recorded (i.e Rnodes =[1, 2, 3, 4] means that nodes a, b, c and d were infected and then recovered). The number of times each node was in the recovered state at the end of each realisation was counted and then divided by the total number of realisations to find the probability that each node was ever infected. For example,

$$\text{P(node } c \text{ was ever infected)} = \frac{\text{(number of times node } c \text{ was infected in } r \text{ runs)}}{r}$$

We can also determine which combination of nodes are ever infected during the epidemic and conditional probabilities to help us understand the infection dynamics. Here we make use of the conditional probability rule,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

and also the following definition

$$P(i|j) \quad = \quad \text{P(node } i \text{ is ever infected} \mid \text{node } j \text{ was infected)}$$

to derive probabilities of interest. To find $P(i \cap j)$ we sum up all absorbing states where nodes i and j are both in the recovered state. In some of the following calculations, we use $P(i')$ is used to denote the probability node i is never infected.

Conditional probabilities are useful if we want information about the probability that specific node(s) become infected given that we have information about the initial infectious node or current state that the network is in. The probabilities derived below will be useful for determining quantities such as the infection path probabilities in future sections.

For $\mathbf{I_0} = a$ we have

$$\begin{aligned} P(b|a) &= \frac{P(b \cap a)}{P(a)} \\ &= P(b \cap a) \text{ (as } P(a) = 1) \\ &= P(b) \\ &= \frac{\mathcal{R}\left(2\,\mathcal{R}^2 + 4\,\mathcal{R} + 1\right)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)} \end{aligned}$$

To find $P(b \cap a)$ we sum up the probability of the epidemic ending in the appropriate absorbing states where nodes a and b are in the recovered class. This is the same method used to determine $P(b)$ as illustrated in Equation 3.1. Due to the symmetry of the LoT(6) network, we also have the following results: $P(b|a) = P(c|a) = P(e|d) = P(f|d)$.

The following conditional probabilities are found in the same way as described above.

- P(Node a, c and d are ever infected while node b is not)

$$\begin{aligned} P(b' \cap c \cap d | \mathbf{I_0} = a) &= \frac{P(b' \cap c \cap d)}{P(a)} \\ &= \frac{\mathcal{R}^2}{(\mathcal{R}+1)(2\mathcal{R}+1)^2} \end{aligned}$$

- P(Node a, b, c and d are ever infected)

$$\begin{aligned} P(b \cap c \cap d | \mathbf{I_0} = a) &= \frac{P(b \cap c \cap d)}{P(a)} \\ &= \frac{\mathcal{R}^3(4\mathcal{R}^2 + 10\mathcal{R} + 5)}{(\mathcal{R}+1)^3(2\mathcal{R}+1)^2} \end{aligned}$$

For $\mathbf{I_0} = c$ we have

$$\begin{aligned} P(a|c) &= \frac{P(a \cap c)}{P(c)} \\ &= P(a \cap c) \text{ (as } P(c) = 1) \\ &= P(a) \\ &= \frac{\mathcal{R}\left(2\,\mathcal{R}^2 + 4\,\mathcal{R} + 1\right)}{\left(\mathcal{R}+1\right)^2\left(2\,\mathcal{R}+1\right)} \end{aligned}$$

Due to the symmetry of the network we also have the result:
$P(a|c) = P(e|d) = P(f|d)$.

- P(Node a, c and d are ever infected while node b is not)

$$P(a \cap b' \cap d | \mathbf{I_0} = c) = \frac{2\mathcal{R}^2}{(\mathcal{R}+1)^2(2\mathcal{R}+1)(3\mathcal{R}+1)}$$

- P(Node a, b, c and d are ever infected)

$$P(a \cap b \cap d | \mathbf{I_0} = c) = \frac{2\mathcal{R}^3(3\mathcal{R}+5)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)(3\mathcal{R}+1)}$$

The following probabilities are the same for the initial conditions, $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$.

- $P(d|c) = \dfrac{P(d)}{P(c)} = \dfrac{\mathcal{R}}{\mathcal{R}+1}$

- P(Node a, b and c are ever infected):

$$P(a \cap b \cap c) = \frac{2\,\mathcal{R}^2\ (\mathcal{R}+2)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}$$

- P(Nodes a and c are ever infected, while b is never infected):

$$\begin{aligned} P(a \cap b' \cap c) &= \frac{\mathcal{R}}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)} \\ &= P(a \cap b \cap c') \text{ due to network symmetry.} \end{aligned}$$

- P(Node d, e and f are ever infected, given that node d was infected):

$$P(e \cap f|d) = \frac{2\,\mathcal{R}^2\ (\mathcal{R}+2)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)} = P(a \cap b \cap c)$$

- P(Node d and f are ever infected, while node e is never infected, given that node d was infected):

$$P(e' \cap f)|d) = \frac{\mathcal{R}}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)} = P(a \cap b' \cap c)$$

We can check the above by ensuring

$$P(c) = P(a \cap b \cap c) + P(a \cap b' \cap c)$$

and

$$P(f|d) = P(e \cap f|d) + P(e' \cap f|d).$$

All analytic results were verified by comparing them to results found from stochastic realisations. In the next section we continue to derive and use conditional probabilities that help us describe the probability of the infection spanning the length (i.e. the probability the infection spreads to node a and to ω_N) of the LoT network.

3.1.3 Paths of infection from node a to f

We may wish to know the probability that the infection spreads to both ends of the LoT network. For $\mathbf{I_0} = a$ this involves simply finding the probability that the infection reaches node f (given by $P(f)$ in Table 3.2). For $\mathbf{I_0} = c$ this translates to the probability that the infection spreads to node a and to node f. However, we are also interested in which path the infection takes to reach node f (or nodes a and f for $\mathbf{I_0} = c$). Therefore we derive equations for the probability that the infection takes each of the four paths (shown in Figure 3.6) to spread across nodes a to f for $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$. We give the analytical expression for each of the paths (1) - (4) below as found directly from the equations describing the probability of transitions for the LoT network. We also show how we can derive the same expression from probabilities previously found for the lollipop and triangle networks. Thirdly we show how we can obtain the expression from some of the smaller probabilities obtained from the LoT network. In the following derivations we use $P(a \cap b \cap c)$ to denote the probability that nodes a,b and c are ever infected and $P(a \cap b' \cap c)$ to denote the probability that nodes a and c are ever infected, while node b is never infected. The derivation of probabilities in this section is aimed towards generalising infection path probabilities for a LoT network with N nodes.

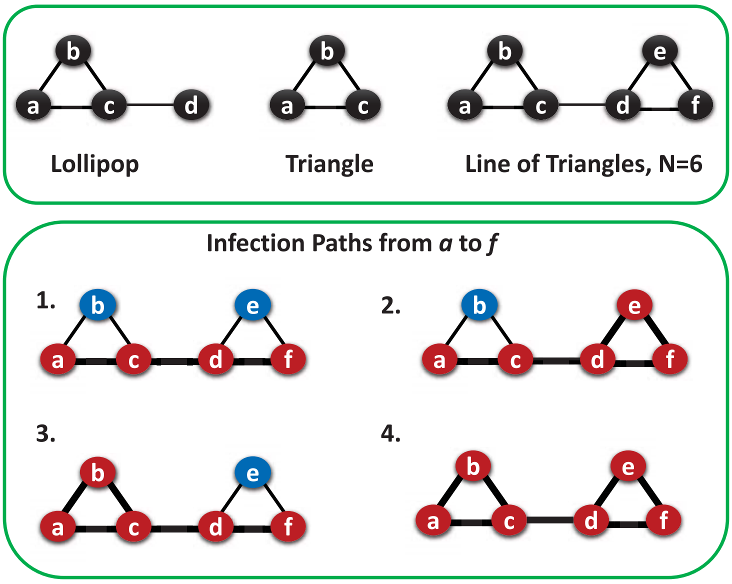


Figure 3.6: Top: Lollipop, Triangle & LoT network structures with node labels. Bottom: Four paths of infection from node a to node f in the LoT(6) network. Blue nodes = susceptible, Red nodes = recovered. Bold edges show infection paths.

We can split the LoT(6) network into two smaller networks which are the lollipop (for nodes a to d) and a triangle (nodes d to f), see Figure 3.6. We use this knowledge when formulating and checking the equations below.
The four paths the infection can take to spread from node a to f starting with the specified $\mathbf{I_0}$ are the following:

1. The infection progresses only along the line, that is only nodes a,c,d and f acquire the infection. This is the same as finding the probability of reaching the absorbing state where node a,c,d and f are all in the 'R' class and nodes b and e are in the 'S' class.

   - For $\mathbf{I_0} = a$, we have:

$$\begin{aligned} P(b' \cap c \cap d | \mathbf{I_0} = a) \times P(e' \cap f | d) &= \frac{\mathcal{R}^3}{(\mathcal{R}+1)^4 (2\,\mathcal{R}+1)^3} \\ &= \mathsf{P}_{RSRRSR} \end{aligned}$$

where $P(b' \cap c \cap d | \mathbf{I_0} = a)$ is derived from the lollipop network and $P(e' \cap f | d)$ is derived from the triangle network. We also note that $P(b' \cap c \cap d | \mathbf{I_0} = a)$ can be further decomposed as follows:

$$\begin{aligned} P(b' \cap c \cap d | \mathbf{I_0} = a) &= P(a \cap b' \cap c) \\ &\quad \times P(\text{node } c \text{ infects node } d \text{ but not node } b) \\ &= P(a \cap b' \cap c) \times \left(\frac{\mathcal{R}}{2\mathcal{R}+1} \right) \end{aligned}$$

Also note due to the network symmetry we have $P(e' \cap f | d) = P(a \cap b' \cap c)$. Therefore, the probability that the infection spreads only along the line is:

$$\begin{aligned} \mathsf{P}_{RSRRSR} &= P(b' \cap c \cap d | \mathbf{I_0} = a) \times P(e' \cap f | d) \\ &= P(b' \cap c \cap d | \mathbf{I_0} = a) \times P(a \cap b' \cap c) \qquad (3.2) \end{aligned}$$

- For $\mathbf{I_0} = c$:

$$\begin{aligned} P(a \cap b' \cap d | \mathbf{I_0} = c) \times P(e' \cap f | d) &= \frac{2\mathcal{R}^3}{(\mathcal{R}+1)^4 (2\,\mathcal{R}+1)^2\,(3\mathcal{R}+1)} \\ &= \mathsf{P}_{RSRRSR} \end{aligned}$$

where $P(a \cap b' \cap d | \mathbf{I_0} = c)$ can also be decomposed into the following:

$$P(a \cap b' \cap d | \mathbf{I_0} = c) = P(a \cap b' \cap c) \times \left(\frac{2\mathcal{R}}{3\mathcal{R}+1} \right)$$

Note, $P(a \cap b' \cap c)$ doesn't take into account the fact that the initial infectious node c has degree 3 (and not degree 2 like $\mathbf{I_0} = a$). Thus, we must multiply $P(a \cap b' \cap c)$ by the probability that node c infects nodes a and d but never node b, $\left(\frac{2\mathcal{R}}{3\mathcal{R}+1} \right)$, to get $P(a \cap b' \cap d | \mathbf{I_0} = c)$. Finally, the probability only nodes a, c, d and f are ever infected given that $\mathbf{I_0} = c$ is:

$$\mathsf{P}_{RSRRSR} = P(a \cap b' \cap d | \mathbf{I_0} = c) \times P(a \cap b' \cap c) \qquad (3.3)$$

2. Only node b escapes infection.

- For $\mathbf{I_0} = a$:

$$\begin{aligned} P(b' \cap c \cap d | \mathbf{I_0} = a) \times P(e \cap f | d) &= \frac{2\,\mathcal{R}^4\,(\mathcal{R}+2)}{(\mathcal{R}+1)^4 (2\,\mathcal{R}+1)^3} \\ &= \mathsf{P}_{RSRRRR} \end{aligned}$$

where $P(b'\cap c\cap d|\mathbf{I_0}=a)$ is derived from the lollipop network and $P(e\cap f|d)=P(a\cap b\cap c)$ is derived from the triangle network. Therefore, we have:

$$\mathsf{P}_{RSRRRR} = P(b'\cap c\cap d|\mathbf{I_0}=a)\times P(a\cap b\cap c) \tag{3.4}$$

- For $\mathbf{I_0}=c$:

$$\begin{aligned} P(a\cap b'\cap d|\mathbf{I_0}=c)\times P(e\cap f|d) &= \frac{4\mathcal{R}^4\,(\mathcal{R}+2)}{(\mathcal{R}+1)^4(2\mathcal{R}+1)^2\,(3\mathcal{R}+1)} \\ &= \mathsf{P}_{RSRRRR} \end{aligned}$$

Therefore,

$$\mathsf{P}_{RSRRRR} = P(a\cap b'\cap d|\mathbf{I_0}=c)\times P(a\cap b\cap c) \tag{3.5}$$

3. Only node e escapes infection.

- For $\mathbf{I_0}=a$:

$$\begin{aligned} P(b\cap c\cap d|\mathbf{I_0}=a)\times P(e'\cap f|d) &= \frac{\mathcal{R}^4\,\left(4\mathcal{R}^2+10\mathcal{R}+5\right)}{(\mathcal{R}+1)^5(2\mathcal{R}+1)^3} \\ &= \mathsf{P}_{RRRRSR} \end{aligned}$$

Note that to get $P(b\cap c\cap d|\mathbf{I_0}=a)$ we must take into account the two different pathways to get from node a to node d. The first pathway is when nodes a,b and c are infected first and then node d becomes infected, this is represented by the first part of Equation 3.6 below. The second pathway is when nodes a,c and d are infected first and then node b is infected last, this is represented by the second part of Equation 3.6.

$$\begin{aligned} P(b\cap c\cap d|\mathbf{I_0}=a) &= \left(\frac{\mathcal{R}}{\mathcal{R}+1}\right)P(a\cap b\cap c) \\ &\quad+\left(\frac{\mathcal{R}}{\mathcal{R}+1}\right)P(b'\cap c\cap d|\mathbf{I_0}=a) \end{aligned} \tag{3.6}$$

Finally, we have:

$$\mathsf{P}_{RRRRSR} = P(b\cap c\cap d|\mathbf{I_0}=a)\times P(a\cap b'\cap c) \tag{3.7}$$

- For $\mathbf{I_0}=c$:

$$\begin{aligned} P(a\cap b\cap d|\mathbf{I_0}=c)\times P(e'\cap f|d) &= \frac{2\mathcal{R}^4\,(3\mathcal{R}+5)}{(\mathcal{R}+1)^4(2\mathcal{R}+1)^2(3\mathcal{R}+1)} \\ &= \mathsf{P}_{RRRRSR} \end{aligned}$$

where $P(a \cap b \cap d | \mathbf{I_0} = c)$ can be decomposed in a similar way to Equation 3.6 resulting in the following:

$$\begin{aligned} P(a \cap b \cap d | \mathbf{I_0} = c) &= \frac{\mathcal{R}}{\mathcal{R}+1} P(a \cap b \cap c) \\ &+ \frac{\mathcal{R}(\mathcal{R}+3)}{\mathcal{R}+1} P(a \cap b' \cap d | \mathbf{I_0} = c) \end{aligned} \tag{3.8}$$

Therefore, we have:

$$\mathsf{P}_{RRRRSR} = P(a \cap b \cap d | \mathbf{I_0} = c) \times P(a \cap b' \cap c) \tag{3.9}$$

4. All nodes aquire the infection.

   - For $\mathbf{I_0} = a$:

$$\begin{aligned} P(b \cap c \cap d | \mathbf{I_0} = a) \times P(e \cap f | d) &= \frac{2\,\mathcal{R}^5\,\left(4\,\mathcal{R}^3 + 18\,\mathcal{R}^2 + 25\,\mathcal{R} + 10\right)}{(\mathcal{R}+1)^5 (2\,\mathcal{R}+1)^3} \\ &= \mathsf{P}_{RRRRRR} \end{aligned}$$

   Therefore, we have:

$$\mathsf{P}_{RRRRRR} = P(b \cap c \cap d | \mathbf{I_0} = a) \times P(a \cap b \cap c) \tag{3.10}$$

   - For $\mathbf{I_0} = c$:

$$\begin{aligned} P(a \cap b \cap d | \mathbf{I_0} = c) \times P(e \cap f | d) &= \frac{4\mathcal{R}^5(\mathcal{R}+2)(3\mathcal{R}+5)}{(\mathcal{R}+1)^4 (2\,\mathcal{R}+1)^2 (3\mathcal{R}+1)} \\ &= \mathsf{P}_{RRRRRR} \end{aligned}$$

   Finally, we have:

$$\mathsf{P}_{RRRRRR} = P(a \cap b \cap d | \mathbf{I_0} = c) \times P(a \cap b \cap c) \tag{3.11}$$

Figure 3.7 shows each of the above infection pathway probabilities evaluated for a range of $\mathcal{R}$ values. It is important to note that the formulae derived for different networks (triangle, lollipop and LoT) can be used to find $P(a \cap b' \cap c)$ and $P(a \cap b \cap c)$ and they all give the same result, as expected. Similarly, we can find $P(b' \cap c \cap d | \mathbf{I_0} = a)$, $P(a \cap b' \cap d | \mathbf{I_0} = c)$, $P(b \cap c \cap d | \mathbf{I_0} = a)$, $P(a \cap b \cap d | \mathbf{I_0} = c)$ using the lollipop or LoT network and the final equation for each probability are the same. This provides a good way to cross check that our expression for each of these probabilities are correct. We also note that $P(a \cap b' \cap c)$ and $P(a \cap b \cap c)$ remain the same for $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$, however $P(b' \cap c \cap d | \mathbf{I_0} = a)$, $P(a \cap b' \cap d | \mathbf{I_0} = c)$, $P(b \cap c \cap d | \mathbf{I_0} = a)$, $P(a \cap b \cap d | \mathbf{I_0} = c)$ clearly depend on the initial condition.
Equations 3.2 to 3.5 can be written down in terms of probabilities that can be found from the triangle network ($P(a \cap b' \cap c)$ and $P(a \cap b \cap c)$), this is because there

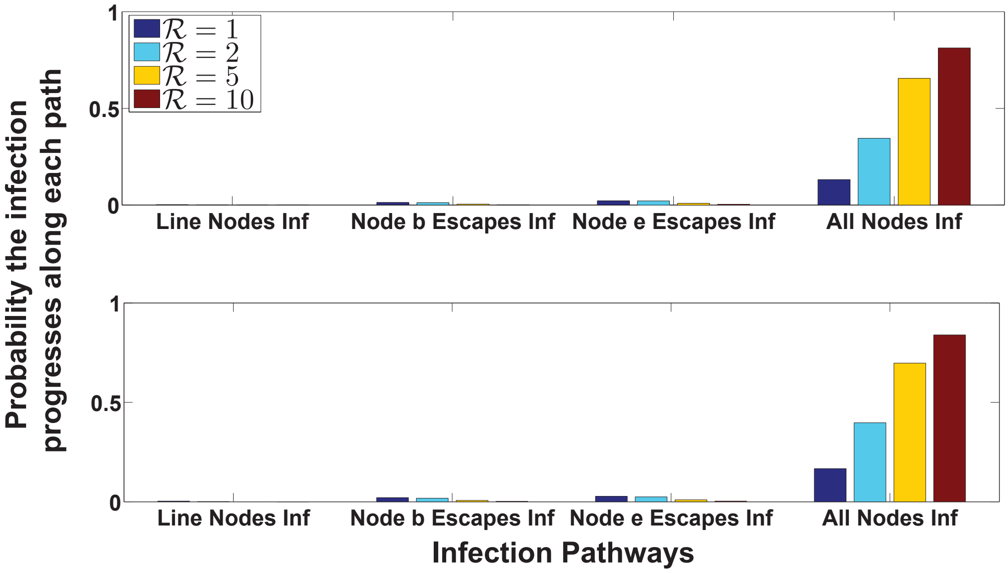


Figure 3.7: Probability of infection spreading along a given pathway from node a to f in the LoT(6) network. Top graph is for $\mathbf{I_0} = a$ and bottom graph is for $\mathbf{I_0} = c$.

is only one way the infection can get from node a to d (as node b does not get infected here). Once the infection reaches node d we can use the probabilities from the triangle network to find out if nodes e and f or just node f then become infected. However, in Equations 3.7 to 3.11 we must use probabilities that are derived from the lollipop network $[P(b' \cap c \cap d|\mathbf{I_0} = a)$, $P(b \cap c \cap d|\mathbf{I_0} = a)$ or $P(a \cap b' \cap d|\mathbf{I_0} = c)$, $P(a \cap b \cap d|\mathbf{I_0} = c)]$ as well as those for the triangle network. This is because there are multiple pathways the infection can spread across node a to d, this is demonstrated in Equation 3.6.

Therefore, to generalise a line of triangles network of N nodes we only require information from the triangle network and the lollipop network in order to find how far through the network the infection spreads before it dies out. We independently verified our analytic results for the infection path probabilities by comparing them to results found via stochastic realisations. We give figures and tables displaying results comparing the stochastic and analytic results for $\mathbf{I_0} = a$ in Appendix B on page 237.

3.1.4 Probabilities for how far along the network the infection spreads

We may want to find how far along the network the infection is likely to progress. Thus, we find analytic expressions that describe the probability that the infection ends at a given node. These are evaluated for specified parameters and compared to the stochastic realisation results in Table B.3 on page 239. Here we also derive results for the LoT(9) network after the LoT(6) network in order to see if any patterns emerge in the equations. Nodes a to i in Figure 3.1 represents the LoT(9) network. Throughout the following sections, we use $P(i|\mathbf{I_0} = j)$ to denote the probability node i is ever infected given that node j was the initial infectious node in a LoT network. For example, $P(c|\mathbf{I_0} = a)$ is the probability node c is ever infected given that $\mathbf{I_0} = a$. Note that $P(c|\mathbf{I_0} = a)$ gives the probability that the infection spreads from one side of the triangle network to the other side.

For the LoT(6) we define $P_{end}(i|\mathbf{I_0} = a)$ to be the probability that the infection reaches node i and no further given that $\mathbf{I_0} = a$, so all nodes to the right of node i must be susceptible. For example, $P_{end}(e|\mathbf{I_0} = a)$ is the probability that the infection starts at node a and ends at node e before node f could become infected. $P_{end}(i|\mathbf{I_0} = a)$ is derived by adding up all the probabilities of possible infection paths starting at node a and ending at node i. The probability that nodes a and b are the only two nodes ever infected is denoted by $P(a \cap b \text{ only})$. Note that $P_{end}(a|\mathbf{I_0} = a)$ is simply the probability that the infection does not spread. Therefore, we have:

$$\begin{aligned}
P_{end}(a|\mathbf{I_0} = a) &= \frac{1}{2\mathcal{R}+1} \\
P_{end}(b|\mathbf{I_0} = a) &= P(a \cap b \text{ only}) = \frac{\mathcal{R}}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)} \\
P_{end}(c|\mathbf{I_0} = a) &= \mathsf{P}_{RRRSSS} + \mathsf{P}_{RSRSSS} = P(c|\mathbf{I_0} = a) \times \left(\frac{1}{\mathcal{R}+1}\right) \\
&= \frac{\mathcal{R}\,\left(2\,\mathcal{R}^2 + 4\mathcal{R} + 1\right)}{(\mathcal{R}+1)^3\,(2\mathcal{R}+1)}
\end{aligned}$$

where $\dfrac{1}{\mathcal{R}+1}$ is the probability that node c recovers before it can transmit infection to node d.

$$\begin{aligned}
P_{end}(d|\mathbf{I_0} = a) &= \mathsf{P}_{RRRRSS} + \mathsf{P}_{RSRRSS} = P(d|\mathbf{I_0} = a) \times \left(\frac{1}{2\mathcal{R}+1}\right) \\
&= \frac{\mathcal{R}^2\,\left(2\,\mathcal{R}^2 + 4\,\mathcal{R} + 1\right)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}
\end{aligned}$$

where $\dfrac{1}{2\mathcal{R}+1}$ is the probability that node d recovers before it can transmit infection to node e or f.

$$\begin{aligned}
P_{end}(e|\mathbf{I_0} = a) &= P(d|\mathbf{I_0} = a) \times P(a \cap b \text{ only}) = \frac{\mathcal{R}^3\,\left(2\,\mathcal{R}^2 + 4\,\mathcal{R} + 1\right)}{(\mathcal{R}+1)^5(2\,\mathcal{R}+1)^2} \\
P_{end}(f|\mathbf{I_0} = a) &= P(f) = P(d|\mathbf{I_0} = a) \times P(c|\mathbf{I_0} = a) \\
&= \frac{\mathcal{R}^3\,\left(2\,\mathcal{R}^2 + 4\,\mathcal{R} + 1\right)^2}{(\mathcal{R}+1)^5(2\,\mathcal{R}+1)^2}
\end{aligned}$$

For the LoT(9) we follow the same method as shown above for the LoT(6) to obtain:

$$\begin{aligned}
P_{end}(a|\mathbf{I_0}=a) &= \frac{1}{2\mathcal{R}+1}\\
P_{end}(b|\mathbf{I_0}=a) &= P(a\cap b \text{ only})\\
P_{end}(c|\mathbf{I_0}=a) &= P(c|\mathbf{I_0}=a)\times\left(\frac{1}{\mathcal{R}+1}\right)\\
P_{end}(d|\mathbf{I_0}=a) &= P(d|\mathbf{I_0}=a)\times\left(\frac{1}{2\mathcal{R}+1}\right)\\
P_{end}(e|\mathbf{I_0}=a) &= P(d|\mathbf{I_0}=a)\times P(a\cap b \text{ only})
\end{aligned}$$

In the following, $P_{end}(f|\mathbf{I_0}=a)$ for the LoT(9) is the same as for the LoT(6) multiplied by the probability that node f recovers before it can transmit the infection further.

$$\begin{aligned}
P_{end}(f|\mathbf{I_0}=a) &= P(f)\left(\frac{1}{\mathcal{R}+1}\right)\\
&= P(d|\mathbf{I_0}=a)\times P(c|\mathbf{I_0}=a)\times\left(\frac{1}{\mathcal{R}+1}\right)\\
&= \frac{\mathcal{R}^3\left(2\,\mathcal{R}^2+4\,\mathcal{R}+1\right)^2}{(\mathcal{R}+1)^6(2\,\mathcal{R}+1)^2}
\end{aligned}$$

To find the probability the infection ends at node g we make use of what we know about the topology of the LoT(9) network. Thus, the probability of the infection reaching node g is simply the probability of reaching node d ($P(d|\mathbf{I_0}=a)$) multiplied by the probability of reaching node g given that node d is already infected (which is equal to $P(d|\mathbf{I_0}=a)$). Then we multiply this by the probability that node g recovers before it transmits the infection further to obtain:

$$P_{end}(g|\mathbf{I_0}=a) = P(d|\mathbf{I_0}=a)^2\left(\frac{1}{2\mathcal{R}+1}\right)$$

The following expressions are derived similarly.

$$\begin{aligned}
P_{end}(h|\mathbf{I_0}=a) &= P(d|\mathbf{I_0}=a)^2 P(a\cap b \text{ only})\\
P_{end}(i|\mathbf{I_0}=a) &= P(d|\mathbf{I_0}=a)^2 P(c|\mathbf{I_0}=a)
\end{aligned}$$

Figure 3.8 shows the results found for $\mathbf{I_0}=a$, where the probability that the infection ends on an apex node (nodes b, e, h etc) is much lower than the probability the infection ends on a node that is along the base of the LoT network (nodes a, c, d, f, g, i and so on). As expected, we can also see that as $\mathcal{R}$ increases the probability the infection ends before reaching ω_N decreases while the probability the infection ends at ω_N increases.

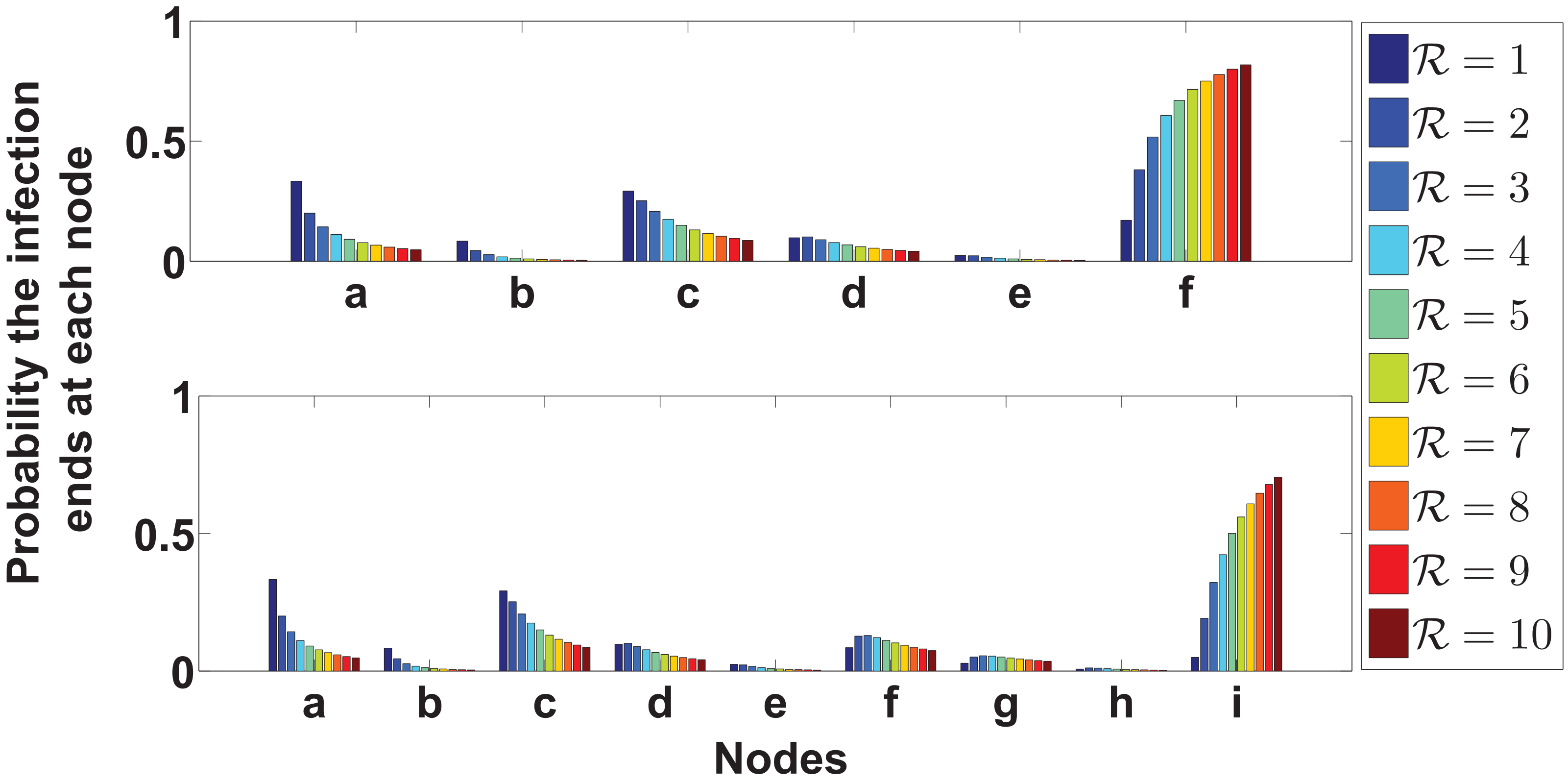


Figure 3.8: Probability of infection ending at each node in the LoT(6) and LoT(9) networks with $\mathbf{I_0} = a$. Top: LoT(6). Bottom: LoT(9).

For the LoT(6) with $\mathbf{I_0} = c$ there are now two directions in which the infection can spread; it can spread to the left and infect nodes a and b or to the right and infect nodes d, e and f. Therefore, below we give the probabilities that describe the infection spreading from node c in only one (left or right) or both directions. Thus, in the following we use $P_{left}(i)$ to denote the probability that infection only spreads to the left and ends at node i. Similarly, $P_{right}(i)$ denotes the probability that infection only spreads to the right and ends at node i. The aim of this section is to illustrate the different infection pathways that are possible for the LoT(6) and LoT(9) networks given that the epidemic starts with $\mathbf{I_0} = c$. The probability of the infection following a given path is determined using similar methods to those illustrated to obtain the final size probabilities in Chapter 2.

- P(Infection spreads to the left only and reaches node a).
  Note that this is the same as determining the probability that node c infects only node a (P_{RSRSSS}) or nodes a and b (P_{RRRSSS}) before the infection dies out. Therefore, we have:

$$\begin{aligned} \text{P(node } c \text{ only infects node } a) &= P(a \text{ only } |\mathbf{I_0} = c) \\ &= P(a \cap b' \cap c) - P(a \cap b' \cap d|\mathbf{I_0} = c) \\ &= P(a \cap b' \cap c)\left(1 - \frac{2\mathcal{R}}{3\mathcal{R}+1}\right) \\ &= \mathsf{P}_{RSRSSS} \end{aligned}$$

$$\begin{aligned} \text{P(node } c \text{ only infects node } a \ \& \ b) &= P(a \cap b \text{ only } |\mathbf{I_0} = c) \\ &= P(a \cap b \cap c) - P(a \cap b \cap d|\mathbf{I_0} = c) \\ &= P(a \cap b \cap c)\left(1 - \frac{\mathcal{R}(3\mathcal{R}+5)}{(\mathcal{R}+2)(3\mathcal{R}+1)}\right) \\ &= \mathsf{P}_{RRRSSS} \end{aligned}$$

  Therefore, P(Infection spreads to the left only and reaches node a):

$$\begin{aligned} P_{left}(a) &= P(a \text{ only } |\mathbf{I_0} = c) + P(a \cap b \text{ only } |\mathbf{I_0} = c) \\ &= (P(a \cap b' \cap c) + P(a \cap b \cap c)) \\ &\qquad -(P(a \cap b' \cap d|\mathbf{I_0} = c) + P(a \cap b \cap d|\mathbf{I_0} = c)) \\ &= \frac{\mathcal{R}(4\mathcal{R}+1)}{(\mathcal{R}+1)(2\mathcal{R}+1)(3\mathcal{R}+1)} \\ &= \mathsf{P}_{RRRSSS} + \mathsf{P}_{RSRSSS} \end{aligned}$$

- P(Infection only spreads to the left and only reaches node b).
  That is, we want to find the probability that only nodes b and c are infected

before the infection dies out given that $\mathbf{I_0} = c$.

$$\begin{aligned} P_{left}(b) &= P(b \text{ only } |\mathbf{I_0} = c) \\ &= \frac{\mathcal{R}}{(\mathcal{R}+1)(2\mathcal{R}+1)(3\mathcal{R}+1)} \\ &= \mathsf{P}_{SRRSSS} \end{aligned}$$

Note, due to the symmetry of the nodes, this is also equal to $P(a \text{ only } |\mathbf{I_0} = c) = \mathsf{P}_{RSRSSS}$.

- P(Infection spreads to the right only and reaches node d).
  Note that here we are determining the probability that only node d becomes infected, that is nodes a,b, e and f do not become infected.

$$\begin{aligned} P_{right}(d) &= P(d \text{ only } |\mathbf{I_0} = c) \\ &= \frac{\mathcal{R}}{(2\mathcal{R}+1)^2(3\mathcal{R}+1)} \\ &= \mathsf{P}_{SSRRSS} \end{aligned}$$

- P(Infection only spreads to the right and only reaches node e).
  Here, we want the probability nodes d and e become infected while node f does not given that $\mathbf{I_0} = c$.

$$\begin{aligned} P_{right}(e) &= P(d \cap e \text{ only } |\mathbf{I_0} = c) \\ &= \frac{\mathcal{R}^2}{(\mathcal{R}+1)^2(2\mathcal{R}+1)^2(3\mathcal{R}+1)} \\ &= \mathsf{P}_{SSRRRS} \end{aligned}$$

- P(Infection spreads to the right only and reaches node f).
  For the infection to spread to the right only and reach node f there are two paths it could take, one where nodes c,d,e and f are infected and the other when nodes c,d and f are infected. Thus, $P_{right}(f)$ is the sum of these two infection path probabilities.

$$\begin{aligned} P_{right}(f) &= P(d \cap e \cap f \text{ only } |\mathbf{I_0} = c) + P(d \cap e' \cap f \text{ only } |\mathbf{I_0} = c) \\ &= \frac{\mathcal{R}^2(2\mathcal{R}^2+4\mathcal{R}+1)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)^2(3\mathcal{R}+1)} \\ &= \frac{\mathcal{R}}{(2\mathcal{R}+1)(3\mathcal{R}+1)} \times P(c|\mathbf{I_0} = a) \\ &= \mathsf{P}_{SSRRSR} + \mathsf{P}_{SSRRRR} \end{aligned}$$

where $\dfrac{\mathcal{R}}{(2\mathcal{R}+1)(3\mathcal{R}+1)}$ is the probability node c infects node d and recovers before it can infect nodes a or b. Note, we recall that $P(c|\mathbf{I_0} = a)$ is the probability that the infection spreads from one side of the triangle network to the other, that is $P(c|\mathbf{I_0} = a) = P(a \cap b \cap c) + P(a \cap b' \cap c)$.

- P(Infection reaches nodes a and d).

  In the following calculations we use $P_{both}(i \cap j)$ to denote the probability that the infection spreads to nodes i and j only. We also note an important observation here, if $\mathbf{I_0} = c$, the probability that the infection spreads to node a and to node d is **not** the product of $P_{left}(a) \times P_{right}(d)$. This is because both infection paths are intertwined due to the recovery rate of node c. Thus, the probability infection reaches nodes a and d only (i.e. nodes e and f do not become infected) is equal to the probability nodes a and d are ever infected given that $\mathbf{I_0} = c$ muliplied by the probability that node d recovers before transmitting the infection to nodes e or f, that is:

$$\begin{aligned} P_{both}(a \cap d) &= P(a \cap d | \mathbf{I_0} = c) \times \left(\frac{1}{2\mathcal{R}+1} \right) \\ &= \frac{2\mathcal{R}^2(3\mathcal{R}^2+5\mathcal{R}+1)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)^2(3\mathcal{R}+1)} \\ &= \mathsf{P}_{RSRRSS} + \mathsf{P}_{RRRRSS} \end{aligned}$$

  where $P(a \cap d | \mathbf{I_0} = c) = (P(a \cap b' \cap d | \mathbf{I_0} = c) + P(a \cap b \cap d | \mathbf{I_0} = c))$

- P(Infection reaches nodes a and e).

  For node e to become infected, node d must first be infected. Furthermore, the probability that node e is infected given node d is independent of if nodes a and b are infected given that node c is infected. Therefore, to find the probability that the infection reaches nodes a and e only (i.e. node f does not get infected) we do the following. We multiply the probability that the infection spreads to node a and d, $P(a \cap d | \mathbf{I_0} = c)$, by the probability that only node e is infected given node d, $P(e \cap f' | d)$. Note that in the following we use $P(a \cap b' \cap c)$ to keep our notation consistent as it is equivalent to $P(e \cap f' | d)$ due to the symmetry of the network.

$$\begin{aligned} P_{both}(a \cap e) &= P(a \cap d | \mathbf{I_0} = c) \times P(a \cap b' \cap c) \\ &= \frac{2\mathcal{R}^3 \left(3\mathcal{R}^2+5\mathcal{R}+1\right)}{(\mathcal{R}+1)^4 \left(2\mathcal{R}+1\right)^2 \left(3\mathcal{R}+1\right)} \\ &= \mathsf{P}_{RSRRRS} + \mathsf{P}_{RRRRRS} \end{aligned}$$

- P(Infection reaches nodes a and f).

  Here we use similar methods as discussed above and note that the probability node f becomes infected given node d is, $P(f|d)$, is equal to $P(c | \mathbf{I_0} = a)$.

$$\begin{aligned} P_{both}(a \cap f) &= P(a \cap d | \mathbf{I_0} = c) \times P(c | \mathbf{I_0} = a) \\ &= \frac{2\mathcal{R}^3(2\mathcal{R}^2+4\mathcal{R}+1)(3\mathcal{R}^2+5\mathcal{R}+1)}{(\mathcal{R}+1)^4(2\mathcal{R}+1)^2(3\mathcal{R}+1)} \\ &= \mathsf{P}_{RSRRSR} + \mathsf{P}_{RSRRRR} + \mathsf{P}_{RRRRSR} + \mathsf{P}_{RRRRRR} \end{aligned}$$

Note that this is simply the sum of the four infection paths discussed in Section 3.1.3 and shown in Figure 3.6, page 72.

For the LoT(9) with $\mathbf{I_0} = c$ we follow the same methods as used above for the LoT(6) network. We also note that $P_{left}(a)$, $P_{left}(b)$, $P_{right}(d)$ and $P_{right}(e)$ are identical for the LoT(6) and LoT(9) networks.

- P(Infection spreads to the left only and reaches node a).

$$\begin{aligned} P_{left}(a) &= P(a \text{ only } |\mathbf{I_0} = c) + P(a \cap b \text{ only } |\mathbf{I_0} = c) \\ &= (P(a \cap b' \cap c) + P(a \cap b \cap c)) \\ &\quad -(P(a \cap b' \cap d|\mathbf{I_0} = c) + P(a \cap b \cap d|\mathbf{I_0} = c)) \\ &= \mathsf{P}_{RRRSSSSSS} + \mathsf{P}_{RSRSSSSSS} \end{aligned}$$

- P(Infection only spreads to the left and only reaches node b).
  That is, we want to find the probability that only nodes b and c are infected while node a does not get infected given that $\mathbf{I_0} = c$.

$$\begin{aligned} P_{left}(b) &= P(b \text{ only } |\mathbf{I_0} = c) \\ &= \mathsf{P}_{SRRSSSSSS} \end{aligned}$$

  Note, due to the symmetry of the nodes, this is also equal to $P(a \text{ only } |\mathbf{I_0} = c) = \mathsf{P}_{RSRSSSSSS}$.

- P(Infection spreads to the right only and reaches node d).
  Note that here we are determining the probability that only node d becomes infected given $\mathbf{I_0} = c$, that is nodes a,b, e and f do not become infected.

$$\begin{aligned} P_{right}(d) &= P(d \text{ only } |\mathbf{I_0} = c) \\ &= \mathsf{P}_{SSRRSSSSS} \end{aligned}$$

- P(Infection only spreads to the right and only reaches node e).
  Here, we want the probability nodes d and e become infected while node f does not given that $\mathbf{I_0} = c$.

$$\begin{aligned} P_{right}(e) &= P(d \cap e \text{ only } |\mathbf{I_0} = c) \\ &= \mathsf{P}_{SSRRRSSSS} \end{aligned}$$

- P(Infection spreads to the right only and reaches node f).
  For the LoT(9) this is the same as $P_{right}(f)$ for LoT(6) except we now have to multiply it by the probability node f recovers before it can transmit the

infection further. This is because in the LoT(9) network node f is an internal node and is not equal to ω_N, as it is in the LoT(6) network. Thus, $P_{right}(f)$ for the LoT(9) is as follows:

$$\begin{aligned}P_{right}(f) &= \big(P(d \cap e \cap f \text{ only } |\mathbf{I_0} = c) + P(d \cap e' \cap f \text{ only } |\mathbf{I_0} = c)\big) \times \left(\frac{1}{\mathcal{R}+1}\right) \\ &= \frac{\mathcal{R}}{(2\mathcal{R}+1)(3\mathcal{R}+1)} \times P(c|\mathbf{I_0} = a) \times \left(\frac{1}{\mathcal{R}+1}\right) \\ &= \mathsf{P}_{SSRRSRSSS} + \mathsf{P}_{SSRRRRSSS}\end{aligned}$$

- P(Infection spreads to the right only and reaches node g).

  As discussed previously for the LoT(6), given $\mathbf{I_0} = c$, we want to determine the probability that nodes a and b do not get infected and the infection spreads to node g. Therefore, in the following we multiply the probability that node c transmits infection to node d and not nodes a and b, $\frac{\mathcal{R}}{(2\mathcal{R}+1)(3\mathcal{R}+1)}$, by the probability that node g becomes infected given that node d is, which is equal to $P(d|\mathbf{I_0} = a)$. Finally, we multiply this by the probability that node g recovers before transmitting the infection further, $\frac{1}{2\mathcal{R}+1}$.

  Therefore, we obtain:

$$P_{right}(g) = \frac{\mathcal{R}}{(2\mathcal{R}+1)(3\mathcal{R}+1)} \times P(d|\mathbf{I_0} = a) \times \left(\frac{1}{2\mathcal{R}+1}\right)$$

  The following expressions are derived similarly.

- P(Infection spreads to the right only and reaches node h)

$$P_{right}(h) = \frac{\mathcal{R}}{(2\mathcal{R}+1)(3\mathcal{R}+1)} \times P(d|\mathbf{I_0} = a) \times P(a \cap b' \cap c)$$

- P(Infection spreads to the right only and reaches node i)

$$P_{right}(i) = \frac{\mathcal{R}}{(2\mathcal{R}+1)(3\mathcal{R}+1)} \times P(d|\mathbf{I_0} = a) \times P(c|\mathbf{I_0} = a)$$

- P(Infection reaches nodes a and d)

$$P_{both}(a \cap d) = P(a \cap d|\mathbf{I_0} = c) \times \left(\frac{1}{2\mathcal{R}+1}\right)$$

  where we recall that $P(a \cap d|\mathbf{I_0} = c) = P(a \cap b' \cap d|\mathbf{I_0} = c) + P(a \cap b \cap d|\mathbf{I_0} = c)$ defines the two possible infection paths.

- P(Infection reaches nodes a and e)

$$P_{both}(a \cap e) = P(a \cap d|\mathbf{I_0} = c) \times P(a \cap b' \cap c)$$

- P(Infection reaches nodes a and f)

$$P_{both}(a \cap f) \quad = \quad P(a \cap d|\mathbf{I_0} = c) \times P(c|\mathbf{I_0} = a) \times \left(\frac{1}{\mathcal{R}+1}\right)$$

- P(Infection reaches nodes a and g)

$$P_{both}(a \cap g) \quad = \quad P(a \cap d|\mathbf{I_0} = c) \times P(d|\mathbf{I_0} = a) \times \left(\frac{1}{2\mathcal{R}+1}\right)$$

- P(Infection reaches nodes a and h)

$$P_{both}(a \cap h) \quad = \quad P(a \cap d|\mathbf{I_0} = c) \times P(d|\mathbf{I_0} = a) \times P(a \cap b' \cap c)$$

- P(Infection reaches nodes a and i)

$$P_{both}(a \cap i) \quad = \quad P(a \cap d|\mathbf{I_0} = c) \times P(d|\mathbf{I_0} = a) \times P(c|\mathbf{I_0} = a)$$

We can see from the equations above that the probability that the infection ends at node a,b,c,d and e for each initial condition are the same for the LoT(6) and LoT(9) networks. There are some patterns emerging in the equations for the different types of nodes too. For example, nodes c, f and i are all nodes of type 1; nodes d and g are all nodes of type 2 and b,e and h are all apex type nodes. The importance of determining the different types of nodes will become apparent in the next section.

Figures 3.9 and 3.10 show results for how far along the LoT(6) and LoT(9) networks the infection spreads, evaluated for a range of $\mathcal{R}$ values and with $\mathbf{I_0} = c$. Notice the difference in scale of the y-axis in the top graph compared to the bottom two graphs in Figure 3.9, this shows that the infection has a much higher probability of only spreading to the left of node c before it dies out. That is, the infection only spreads within nodes a, b and c. We can see from Figure 3.9 that as $\mathcal{R}$ increases, the probability that the infection only spreads in one direction (to the left or to the right of node c) decreases. Figure 3.10 shows that the probability the infection spreads in both directions and spreads across the length of the network (to nodes a and f for LoT(6) and to nodes a and i for LoT(9)) increases as $\mathcal{R}$ increases.

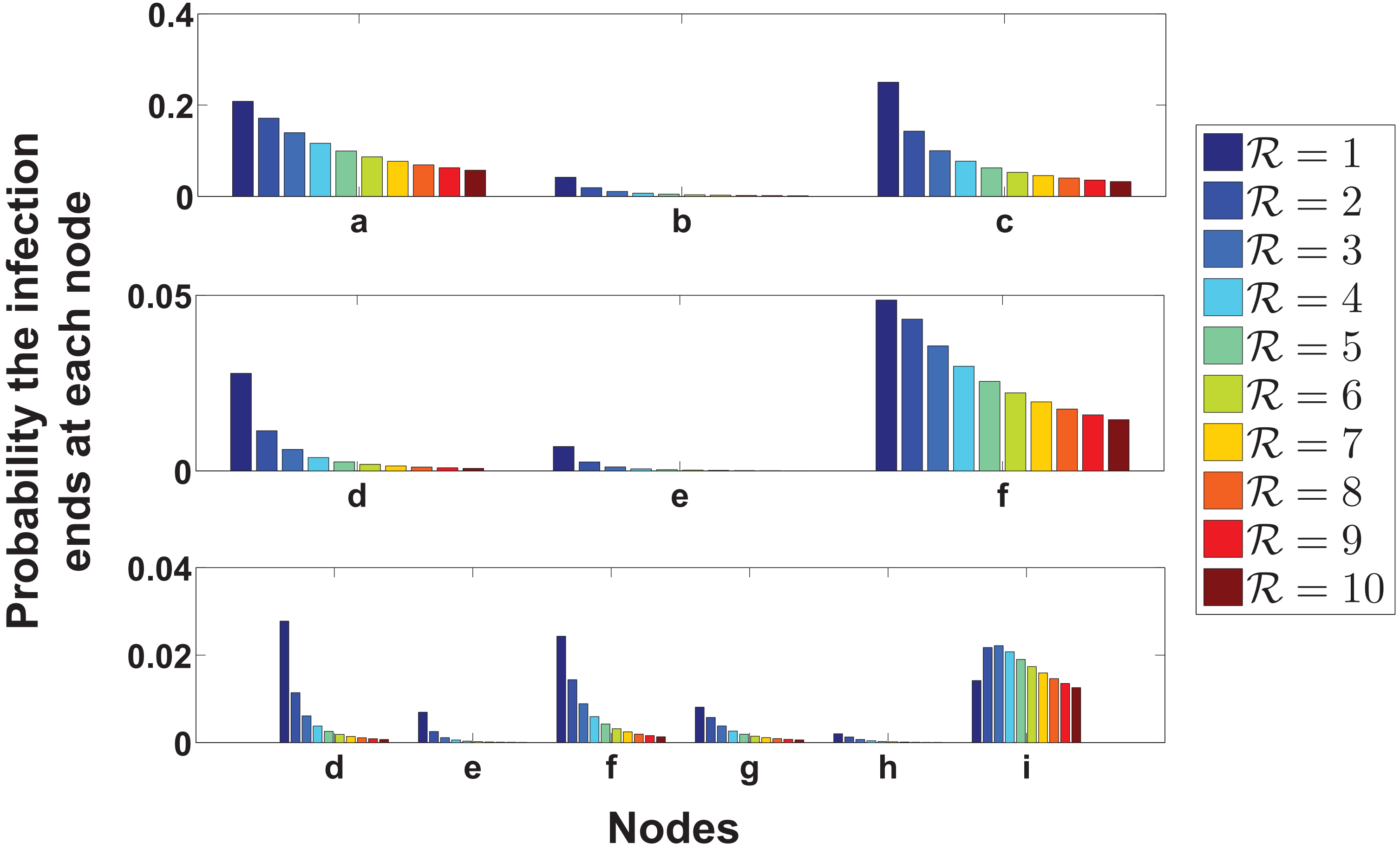


Figure 3.9: Top: Probability the infection only spreads to the left and ends at each node for LoT(6) and LoT(9). Middle: Probability infection only spreads to the right in the LoT(6). Bottom: Probability infection only spreads to the right in the LoT(9). All graphs have $\mathbf{I_0} = c$ and are evaluated for specified range of $\mathcal{R}$.

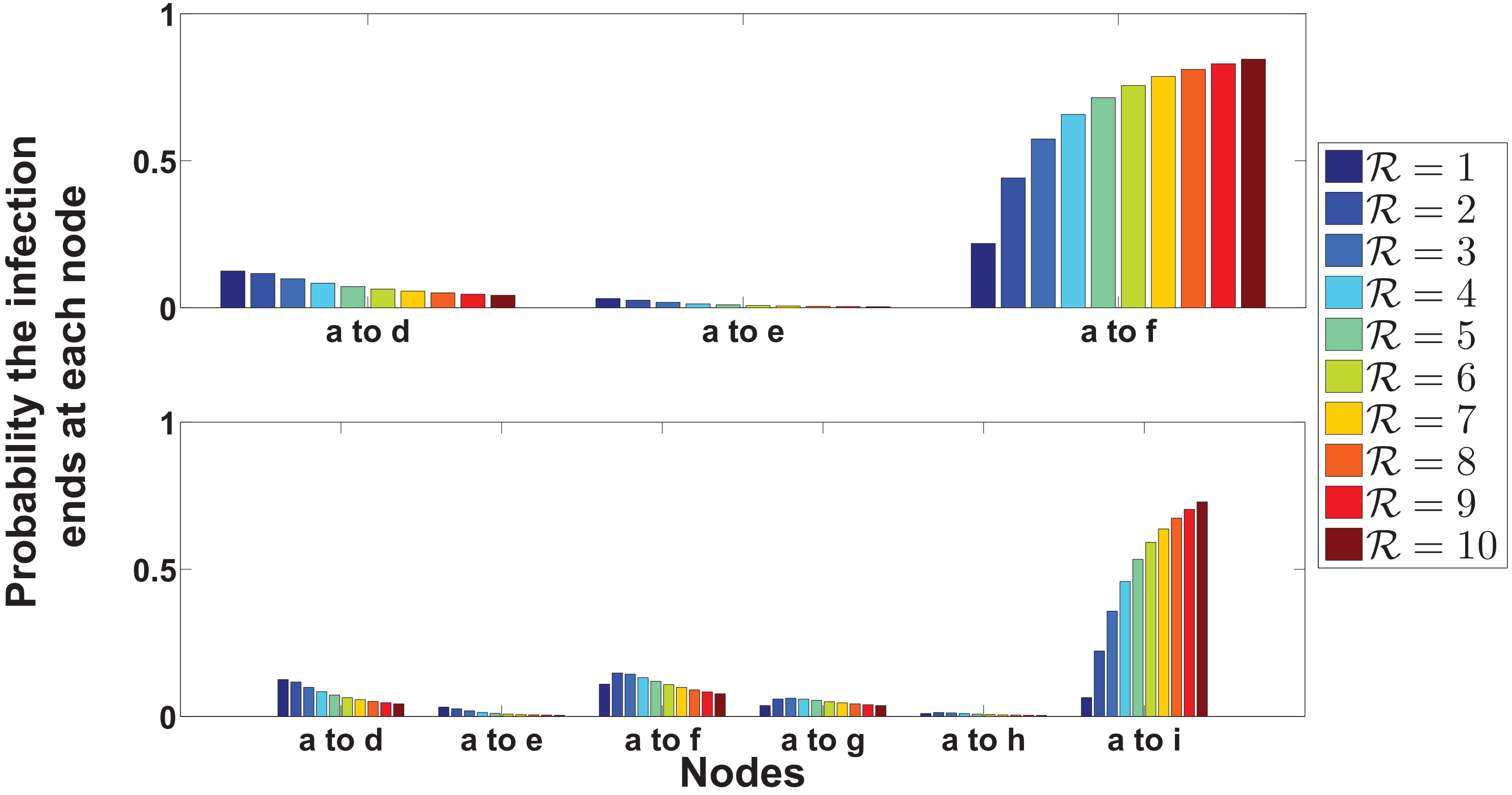


Figure 3.10: Probability the infection only spreads in both directions and ends at each node for LoT(6) (top) and LoT(9) (bottom). Both graphs have $\mathbf{I_0} = c$ and are evaluated for specified range of $\mathcal{R}$.

3.2 Generalising infection path probabilities for a LoT(N) network

As we saw in the previous section, all the infection path probabilities we found for the LoT(6) and LoT(9) depend on a combination of probabilities found for the triangle and lollipop networks. In the last section each of the infection paths from node a to node f in the LoT(6) network were made up of different combinations of these smaller probabilities. Here we show how we can use the same quantities in order to derive generalised analytic expressions for infection path probabilities of the LoT(N) network.

First we need to define some notation. We denote the number of triangles in a LoT(N) network by Δ, where

$$\Delta = \frac{N}{3}$$

Assuming $\mathbf{I_0} = a$, the minimum number of nodes infected in order for the infection to reach ω_N in LoT(N) is 2Δ, this corresponds to the infection progressing straight along the base without infecting any of the apex nodes. The maximum number of nodes that can be infected, assuming that $\mathbf{I_0} = a$, is $3\Delta = N$, in this case all nodes are infected. The number of possible infection paths from node a to ω_N is 2^{Δ}, as N increases the number of pathways for the infection to reach ω_N gets very large. Thus, we may not always want to know exactly which pathway the infection progressed along, although, in theory, we can find each analytic expression for any given infection path probability. We may however, be interested in finding quantities such as the probability the infection reaches ω_N (regardless of which pathway was taken), the probability that every node in the network acquires infection or the probability the infection ends at a given node. In the following sections we first derive generalised expressions for the probability the infection ends at a given node for $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$. We then use these results to generalise the final size probability mass functions for a LoT(N) for $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$.

3.2.1 Probability the infection ends at a given node

The probabilities in this section are found following similar methods to presented in Section 3.1.4 for the LoT(6) and LoT(9) networks, the only difference being that now we are finding the probabilities for a LoT network with N nodes. Throughout the sections that follow, our goal is to describe the different possible infection paths with the aim of using the results generated to derive the final size probability mass functions for a LoT network with N nodes.

Results for starting the epidemic with $\mathbf{I_0} = a$

We start with finding the probability that the infection reaches ω_N, given that it started at node a. For a LoT(N) network this is:

$$\begin{aligned} P_{end}(\omega_N) &= P(d|\mathbf{I_0} = a)^{\Delta-1} P(c|\mathbf{I_0} = a) \\ &= \left[\frac{\mathcal{R}^2\,\left(2\,\mathcal{R}^2 + 4\,\mathcal{R} + 1\right)}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}\right]^{\Delta-1} \left[\frac{\mathcal{R}\left(2\,\mathcal{R}^2 + 4\,\mathcal{R} + 1\right)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}\right] \end{aligned} \tag{3.12}$$

where $P_{end}(\omega_N)$ is the probability that the infection reaches the last node (ω_N) in a LoT network with N nodes. Note that this includes all possible infection pathways to ω_N.

We may also want to find the probability that every node in the network acquires infection at some point during the epidemic. This is the probability that the final size of the epidemic is equal to N nodes which is:

$$\begin{aligned} P(\text{Final Size} = N) &= P(b \cap c \cap d|\mathbf{I_0} = a)^{\Delta-1} P(a \cap b \cap c) \\ &= \left[\frac{\mathcal{R}^3\,\left(4\,\mathcal{R}^2 + 10\,\mathcal{R} + 5\right)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}\right]^{\Delta-1} \left[\frac{2\,\mathcal{R}^2\,(\mathcal{R}+2)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}\right] \end{aligned} \tag{3.13}$$

The above two quantities have been compared with stochastic realisation results set up with the same parameters on a LoT(N) network for $N = 6, 9, 12, \ldots 102$. See Figure 3.11 for graphical results of $P_{end}(\omega_N)$ and $P(\text{Final Size} = N)$ evaluated with different choices of infection parameters.

We can also find the probability that the infection ends at a given node in the LoT(N) network. There are four types of nodes which we must consider for each network; nodes of type 1 (last node in each triangle), type 2 (first node in each triangle), apex nodes in each triangle and ω_N, the last node in the network. See Figure 3.12 for a schematic diagram showing the different types of nodes in a LoT network. Therefore, the following probabilities are indexed by which triangle in the LoT(N) network the infection ends at. Note that the initial infectious node is a node of type 2 and thus the probability the infection ends at a node of type 2 in the first triangle is equal to the probability that the infection does not take off. Similarly, ω_N is a node of type 1 however we need a separate equation to find the probability that the infection reaches ω_N, this is discussed below.

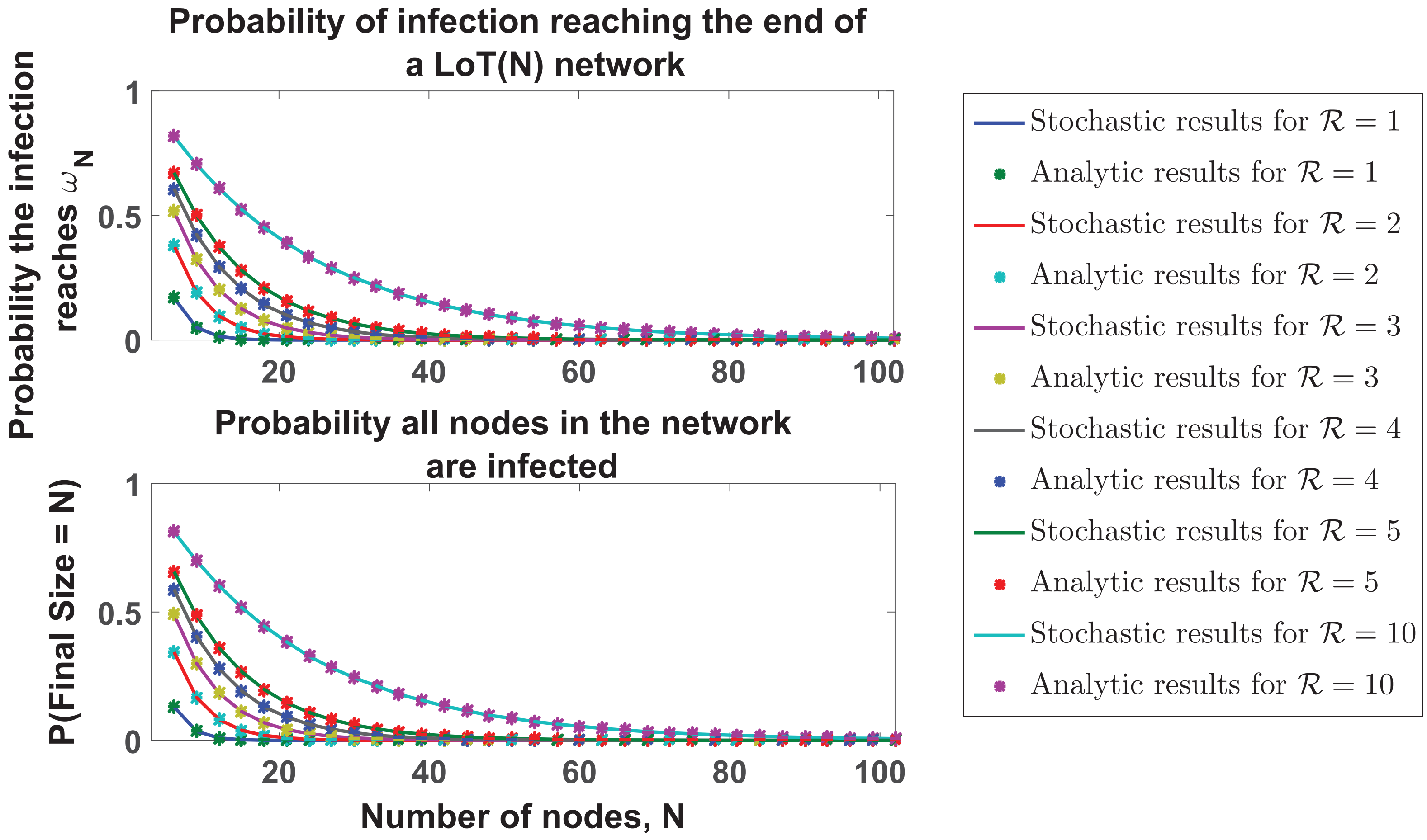


Figure 3.11: Probability the infection reaches ω_N and the probability of obtaining a final size equal to N for a LoT network with $\mathbf{I_0} = a$. Analytic results are compared to stochastic results derived using the Gillespie algorithm.

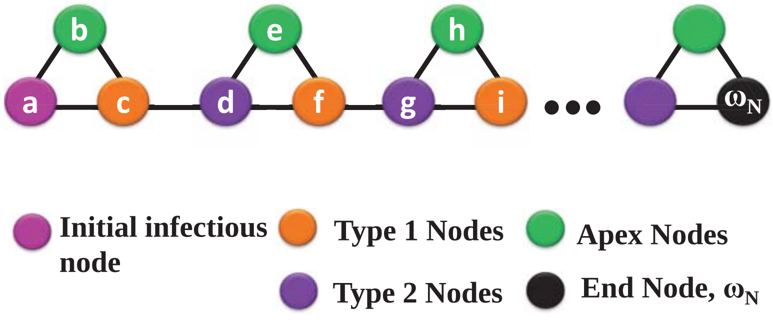


Figure 3.12: Types of nodes in a line of triangles network with $\mathbf{I_0} = a$.

1. The probability that the infection ends at the last node in triangle τ, (node of type 1) is:

$$P(d)^{\tau-1}P(c)\left(\frac{1}{\mathcal{R}+1}\right) \tag{3.14}$$

where $\tau = 1, 2, \ldots \Delta - 1$. Note, $\dfrac{1}{\mathcal{R}+1}$ is the probability that the node of type 1 in triangle τ recovers before it can infect another node in the network.

2. The probability that the infection ends at the first node in triangle τ, (node of type 2) is:

$$P(d)^{\tau-1}\left(\frac{1}{2\mathcal{R}+1}\right) \tag{3.15}$$

where $\tau = 2, \ldots \Delta$.

3. The probability that the infection ends at the apex node in triangle τ, (apex node) is:

$$P(a \cap b \text{ only})P(d)^{\tau-1} \tag{3.16}$$

where $P(a \cap b \text{ only})$ is the probability that only nodes a and b are ever infected (which is also equal to $P(a \cap b' \cap c)$) in a triangle network and $\tau = 1, 2, \ldots \Delta$.

4. The probability that the infection reaches ω_N, given that it started at node a, for a LoT(N) network is similar to Equation 3.14. ω_N is a node of type 1, however we keep it as a separate case because it is the last node in the network we do not need to include the probability that ω_N recovers before infecting a node further along the network. Thus, the probability the infection reaches ω_N is:

$$P(d)^{\Delta-1}P(c) \tag{3.17}$$

Note that this is the same as Equation 3.12 above.

Note that the probabilities desscribed in Equations 3.14 to 3.17 cover all possible infection pathways for the LoT network with N nodes, given that $\mathbf{I_0} = a$. We have verified that the sum of these probabilities is equal to 1, as expected. Figure 3.13 shows the probability the infection ends at each node in the LoT(N) network for N =12,15,18 and 21 for different transmission rates. We notice the same pattern in Figure 3.13 when compared to the probability the infection ends at each node in the LoT(6) and LoT(9) networks (see Figure 3.8); that is as $\mathcal{R}$ increases the probability the infection ends before ω_N decreases while the probability the infection ends at ω_N increases.

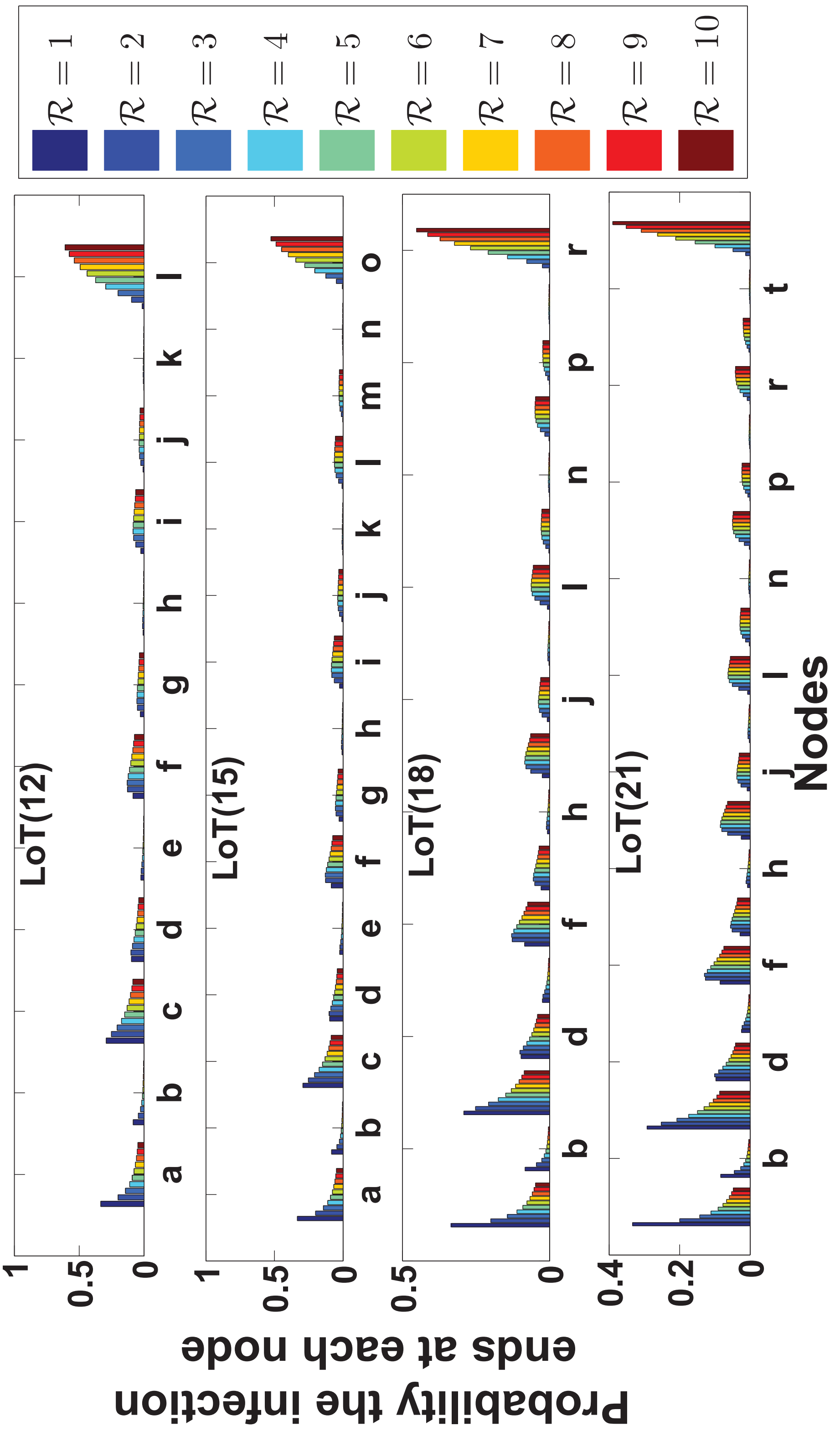


Figure 3.13: Probability the infection ends at a given node starting with $\mathbf{I_0} = a$ in a LoT network.

Results for starting the epidemic with $\mathbf{I_0} = c$

We start with finding the probability that the infection reaches ω_N, given that it started at node c. For a LoT(N) network this is:

$$\begin{aligned} P_{end}(\omega_N) &= P(a \cap d|\mathbf{I_0} = c)P(d|\mathbf{I_0} = a)^{\Delta-2}P(c|\mathbf{I_0} = a) \\ &= \left[\frac{2\mathcal{R}^2(3\mathcal{R}^2+5\mathcal{R}+1)}{(\mathcal{R}+1)^2(2\mathcal{R}+1)(3\mathcal{R}+1)}\right]\left[\frac{\mathcal{R}^2\left(2\,\mathcal{R}^2+4\,\mathcal{R}+1\right)}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)}\right]^{\Delta-2} \\ &\quad\times\left[\frac{\mathcal{R}\left(2\,\mathcal{R}^2+4\,\mathcal{R}+1\right)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}\right] \end{aligned} \tag{3.18}$$

Note that this includes all possible infection pathways to ω_N.

We may also want to find the probability that every node in the network acquires infection at some point during the epidemic. This is the probability that the final size of the epidemic is equal to N nodes which is:

$$\begin{aligned} &P(\text{Final Size} = N) \\ &\quad= P(a \cap b \cap d|\mathbf{I_0} = c)P(b \cap c \cap d|\mathbf{I_0} = a)^{\Delta-2}P(a \cap b \cap c) \\ &\quad= \left[\frac{2\mathcal{R}^3\,(3\mathcal{R}+5)}{(\mathcal{R}+1)^2\,(2\mathcal{R}+1)\,(3\mathcal{R}+1)}\right]\left[\frac{\mathcal{R}^3\,\left(4\,\mathcal{R}^2+10\,\mathcal{R}+5\right)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}\right]^{\Delta-2} \\ &\quad\quad\times\left[\frac{2\,\mathcal{R}^2\,(\mathcal{R}+2)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}\right] \end{aligned} \tag{3.19}$$

The above two quantities have been compared with stochastic realisation results set up with the same parameters on a LoT(N) network for $N = 6, 9, 12, \ldots 102$. See Figure 3.14 for graphical results of $P_{end}(\omega_N)$ and $P(\text{Final Size} = N)$ evaluated with different choices of infection parameters.

Next we find the probability that the infection ends at a given node in the LoT(N) network. See Figure 3.15 for a schematic diagram showing the different types of nodes in a LoT(N) network with $\mathbf{I_0} = c$. The following probabilities are indexed by the triangle in which the infection ends. Note that for $\mathbf{I_0} = c$, the initial infectious node is a node of type 1 in triangle $\tau = 1$. For the following probabilities we assume that the infection reaches node a in the first triangle ($\tau = 1$) and ends at the specified node in triangle $\tau = 2, 3, 4, \ldots \Delta$.

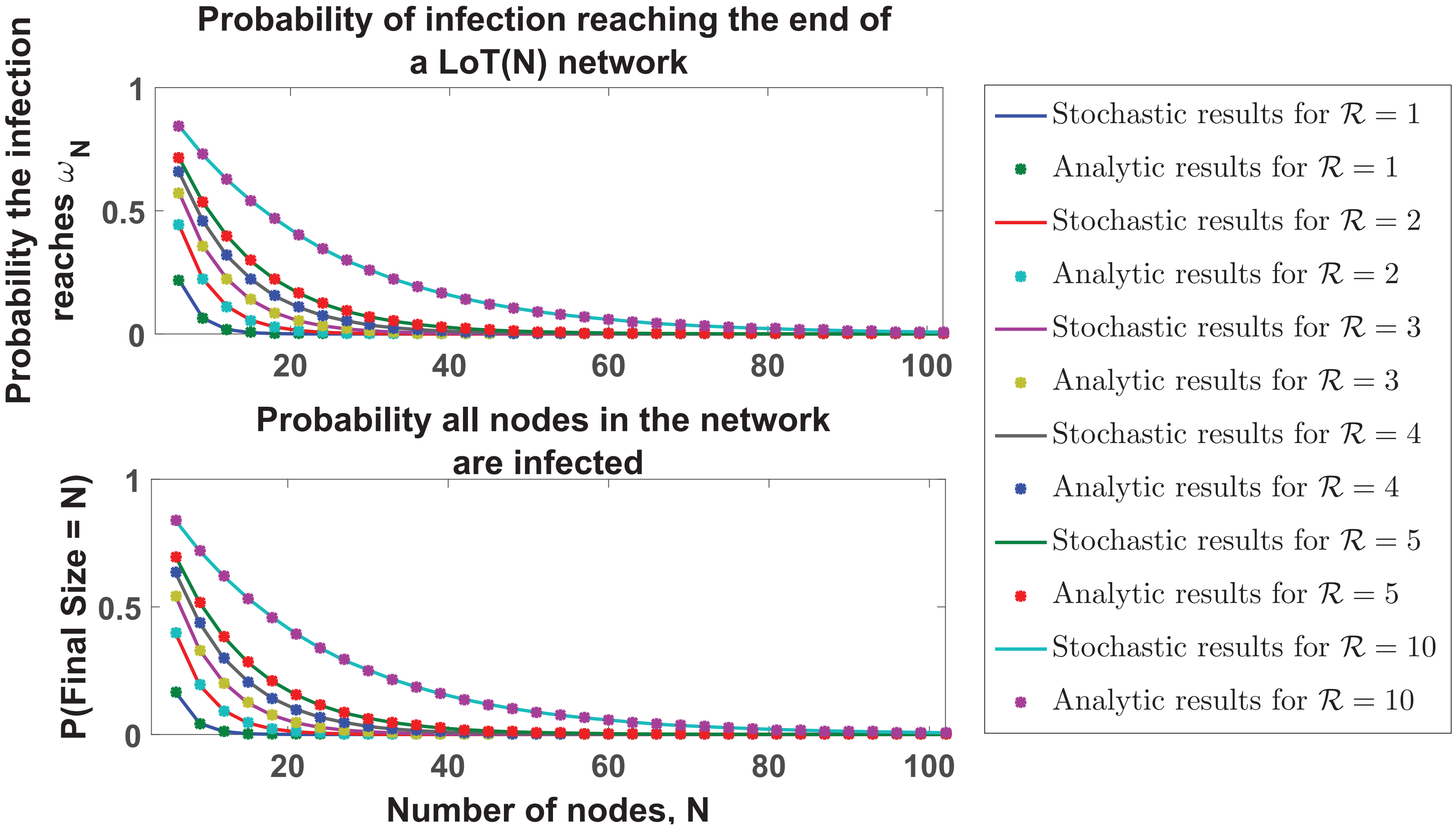


Figure 3.14: Probability the infection reaches ω_N and the probability of obtaining a final size equal to N for a LoT network with $\mathbf{I_0} = c$.

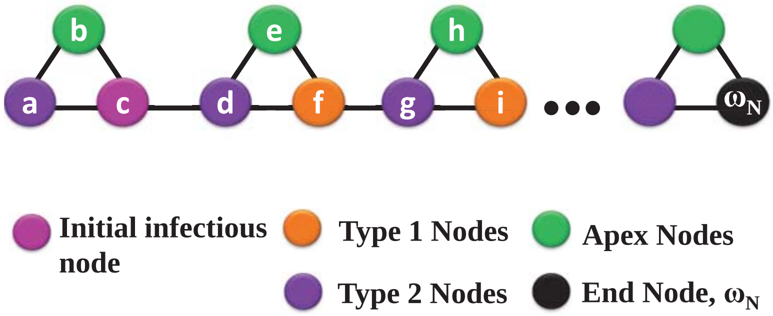


Figure 3.15: Types of nodes in a line of triangles network with $\mathbf{I_0} = c$.

1. The probability that the infection reaches node a and ends at the last node in triangle τ, (node of type 1) is as follows.
   If $\tau = 1$ this is the probability that the infection reaches node a on the left and does not spread into the second triangle, this is equal to $P_{left}(a)$ found previously. However, if $\tau > 1$, we have:
   $$P(a \cap d|\mathbf{I_0} = c)P(d|\mathbf{I_0} = a)^{\tau-2}P(c|\mathbf{I_0} = a)\left(\frac{1}{\mathcal{R}+1}\right) \tag{3.20}$$
   where $\tau = 2, 3, \dots \Delta - 1$ and recall that $P(a \cap d|\mathbf{I_0} = c) = P(a \cap b' \cap d|\mathbf{I_0} = c) + P(a \cap b \cap d|\mathbf{I_0} = c)$. Note, $\frac{1}{\mathcal{R}+1}$ is the probability that the node of type 1 in triangle τ recovers before it can infect another node in the network.

2. The probability that the infection spreads to node a and ends at the first node in triangle $\tau > 1$, (node of type 2) is:
   $$P(a \cap d|\mathbf{I_0} = c)P(d|\mathbf{I_0} = a)^{\tau-2}\left(\frac{1}{2\mathcal{R}+1}\right) \tag{3.21}$$
   where $\tau = 2, \dots \Delta$.

3. The probability that the infection spreads to node a and ends at the apex node in triangle $\tau > 1$, (apex node) is:
   $$P(a \cap d|\mathbf{I_0} = c)P(d|\mathbf{I_0} = a)^{\tau-2}P(a \cap b \text{ only}) \tag{3.22}$$
   where $P(a \cap b \text{ only})$ is the probability that only nodes a and b are ever infected in a triangle network and $\tau = 2, 3, \dots \Delta$.

4. The probabilitiy that the infection spreads to node a and reaches ω_N, given that it started at node c, for a LoT(N) network is similar to Equation 3.20. ω_N is a node of type 1, however we keep it as a separate case because it is the last

node in the network we do not need to include the probability that ω_N recovers before infecting a node further along the network. Thus, the probability the infection spreads to node a and reaches ω_N is:

$$P(a \cap d|\mathbf{I_0} = c)P(d|\mathbf{I_0} = a)^{\tau-2}P(c|\mathbf{I_0} = a) \tag{3.23}$$

Note that this is the same as Equation 3.18 above.

It is important to note that the above Equations 3.20 to 3.23 do not describe all possible infection pathways in the LoT(N) with $\mathbf{I_0} = c$ and therefore do not sum to equal 1 as they did for $\mathbf{I_0} = a$. This is because in each of the equations we have assumed that the infection always reaches node a (as mentioned previously) and thus the infection pathways in which node a does not get infected have not been accounted for. These probabilities were omitted here for simplicity. However, in the next section we discuss the derivation of the final size probability mass functions which ensures that we must take into account the other possible infection pathways in which node a does not get infected. Figure 3.16 shows the probability the infection reaches node a and ends at each node in the LoT(N) network for N =12,15,18 and 21 for different transmission rates.

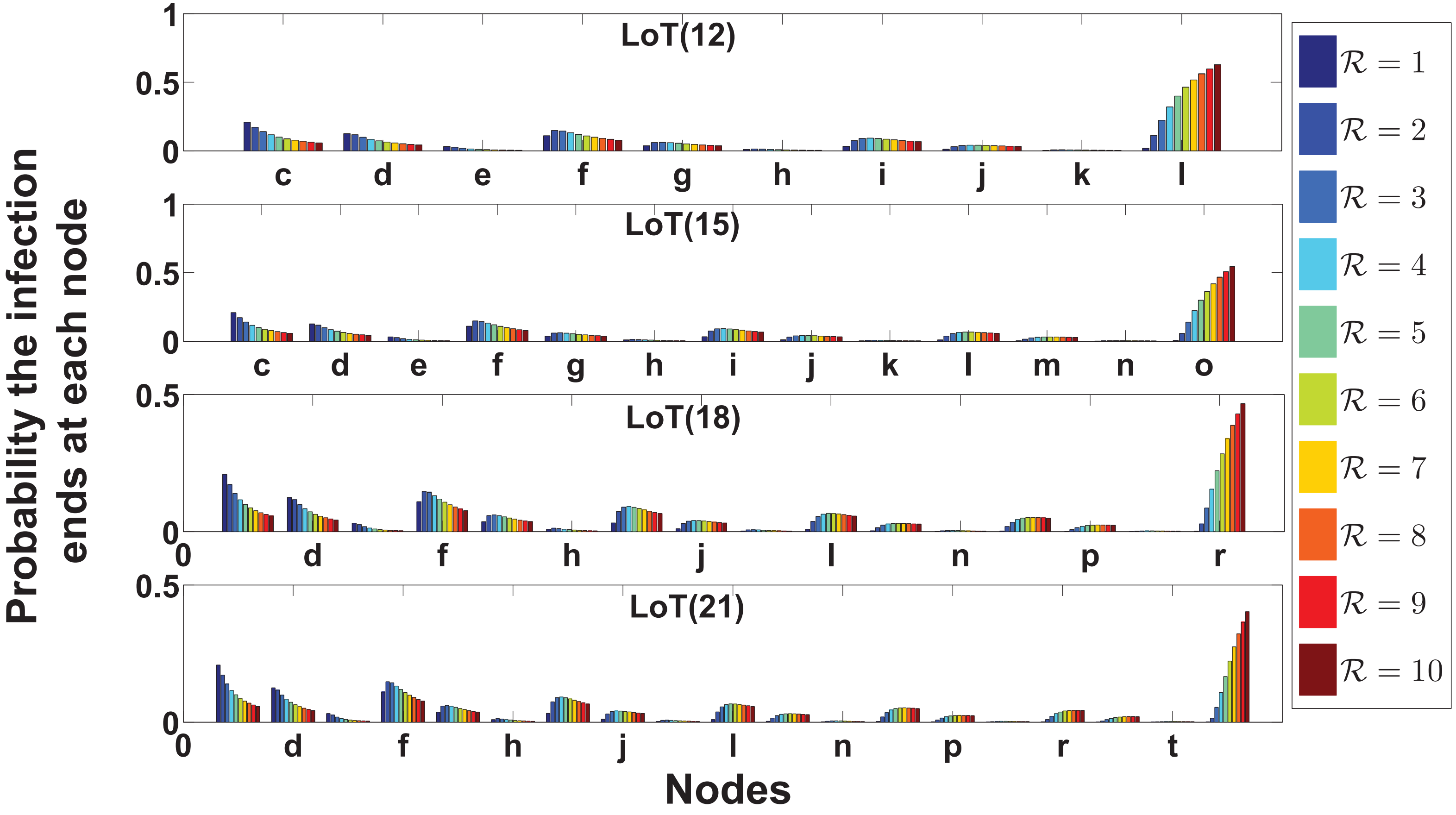


Figure 3.16: Probability the infection spreads to node a and then ends at a given node starting with $\mathbf{I_0} = c$ in a LoT network.

3.2.2 Final size PMFs for a Line of Triangles Network

In this section we derive the probability mass functions for the final epidemic size for a line of triangles network with N nodes. We split our derivations up into two parts; firstly we derive results for the initial condition $\mathbf{I_0} = a$, this is followed by the derivation of results for $\mathbf{I_0} = c$. We use similar methods for both derivations and illustrate our results by plotting the analytic and stochastic final size distribution for a LoT(21) evaluated for a range of $\mathcal{R}$ values.

Results for starting the epidemic with $\mathbf{I_0} = a$

Given that we know the infection ended at given node (node x), we want to find the expected number of apex nodes that were infected. To do this we first need to find:

$$\begin{aligned} P(y \text{ apex nodes infected} \mid \text{infection ended at node } x) &= P(y| \text{ ended at } x) \\ &= \frac{P(y \cap x)}{P(x)} \end{aligned}$$

Then can we evaluate the following

$$E[\# \text{ of apex nodes infected} \mid \text{infection ended at node } x] = \sum_y y P(y| \text{ ended at } x)$$

By finding all possible probabilities, $P(y|$ ended at $x)$, we can also compute the final size probability mass function (PMF) for the line of triangles network. As we have four equations (Equations 3.14 - 3.17) that describe the probability that the infection ends at a node of a specific type (type 1, 2, apex node and ω_N) in triangle τ, we will have four categories of expressions that describe $P(y|$ ends at $x)$.

1. Let the probability that y apex nodes were infected given that the infection ends at the last node in triangle τ (node of type 1), be denoted by:

   $$P(y|T1) = P(y \text{ apex nodes inf} \mid \text{inf ended at node of type 1 in the } \tau^{th} \text{ triangle})$$

   For the following we have $\tau = 1, 2, \ldots \Delta - 1$ and $y = 0, 1, 2, \ldots \tau$. Note that ω_N is a node of type 1 in triangle Δ, however we define a different set of probabilities for reaching ω_N. Thus, for the probabilities $P(y|T1)$ we only look at the probability of ending at node $T1$ in the first $\Delta - 1$ triangles.

   In the following we determine the probability of having either y of the first $\tau - 1$ triangle's apex nodes are infected and the final triangle's apex node is not, or the final triangle's apex node is infected and $y - 1$ of the first $\tau - 1$ apex nodes are infected. Therefore we make use of a binomial coefficient as in each case we get to choose which y (or $y-1$) triangles have an infected apex node.

$$\begin{aligned} P(y|T1) &= \binom{\tau-1}{y-1} P(b' \cap c \cap d|\mathbf{I_0}=a)^{(\tau-1)-(y-1)} P(b \cap c \cap d|\mathbf{I_0}=a)^{y-1} \\ &\quad \times P(b \cap c \text{ only } |\mathbf{I_0}=a) \\ &+ \binom{\tau-1}{y} P(b' \cap c \cap d|\mathbf{I_0}=a)^{(\tau-1)-y} P(b \cap c \cap d|\mathbf{I_0}=a)^{y} \\ &\quad \times P(c \text{ only } |\mathbf{I_0}=a) \end{aligned} \tag{3.24}$$

where $\binom{\tau-1}{y}$ is the binomial coefficient indexed by $\tau-1$ and y, that is:

$$\binom{\tau-1}{y} = \frac{(\tau-1)!}{y!(\tau-1-y)!}$$

We also use the convention that $\binom{n}{k}=0$ if $k<0$ or $k>n$ which allows for the cases when $y=0$ (i.e. zero apex nodes are infected) and $y=\tau$. In Equation 3.24 we also have $P(c \text{ only } |\mathbf{I_0}=a) = P(a \cap b' \cap c) - P(b' \cap c \cap d|\mathbf{I_0}=a)$ which is the probability nodes a and c are the only nodes in a network infected before the infection dies out. Similarly, $P(b \cap c \text{ only } |\mathbf{I_0}=a) = P(a \cap b \cap c) - P(b \cap c \cap d|\mathbf{I_0}=a)$ is the probability nodes a, b and c are the only nodes infected before the infection dies out.

2. The probability that y apex nodes were infected given that the infection ended at the first node in triangle τ (node of type 2), is denoted by:

   $P(y|T2) = P(y$ apex nodes inf | inf ended at node of type 2 in the τ^{th} triangle)

   Note that as our initial infectious node a is a type 2 node, for the following probabilities we do not include node a and thus we start with node d as the first type 2 node in any LoT network. Therefore, here we have $\tau = 2:\Delta$ and $y = 0:\Delta-1$.

   $$\begin{aligned} P(y|T2) &= \binom{\tau-1}{y} P(b' \cap c \cap d|\mathbf{I_0}=a)^{(\tau-1)-y} P(b \cap c \cap d|\mathbf{I_0}=a)^{y} \\ &\quad \times \left(\frac{1}{2\mathcal{R}+1}\right) \end{aligned} \tag{3.25}$$

   where $\left(\frac{1}{2\mathcal{R}+1}\right)$ is the probability that the node of type 2 in the τ^{th} triangle recovers before it can transmit the infection further.

3. The probability that y apex nodes were infected given that the infection ended at the apex node in triangle τ is denoted:

   $P(y|\text{apex } \tau) = P(y$ apex nodes inf | inf ended at an apex node in the τ^{th} triangle)

Note that here we have $y = 1, 2, \ldots \Delta$ as the infection ends on an apex node and $\tau = 1, 2, \ldots \Delta$.

$$\begin{aligned} P(y|\text{apex } \tau) \quad &= \quad \binom{\tau-1}{y-1} P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-1)-(y-1)} P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\qquad \times P(a \cap b \text{ only}) \end{aligned}$$

4. The probability that y apex nodes were infected given that the infection reaches ω_N is denoted by:

   $P(y|\omega_N) =$

   $P(y$ apex | inf ended at ω_N in the LoT(N) network with Δ triangles)

   In the following we have $y = 0, 1, 2, \ldots \Delta$.

$$\begin{aligned} P(y|\omega_N) \quad &= \quad \binom{\Delta-1}{y-1} P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-1)-(y-1)} P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\qquad \times P(a \cap b \cap c) \\ &+ \quad \binom{\Delta-1}{y} P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-1)-y} P(b \cap c \cap d|\mathbf{I_0} = a)^{y} \\ &\qquad \times P(a \cap b' \cap c) \end{aligned} \tag{3.26}$$

   We note that Equation 3.26 is the same as Equation 3.24 with $\tau = \Delta$ and where the last probability in each term is swapped out. That is, in Equation 3.26, $P(a \cap b \cap c)$ has replaced $P(b \cap c \text{ only } |\mathbf{I_0} = a)$ and $P(a \cap b' \cap c)$ has replaced $P(c \text{ only } |\mathbf{I_0} = a)$.

From the above equations we can find the final size of the network that corresponds to each probability. That is:

- $P(y|T1)$ gives a final size equal to $2\tau + y$.
- $P(y|T2)$ gives a final size equal to $2\tau - 1 + y$.
- $P(y|\text{apex } \tau)$ gives a final size equal to $2\tau - 1 + y$.
- $P(y|\omega_N)$ gives a final size equal to $2\Delta + y$.

To generate the final size PMF for a line of triangles network with N nodes we find all possible conditional probabilities as defined by Equations 3.24 to 3.26 and the final size to which each probability corresponds (see above). To find the final size PMF we sum up each of the equations which correspond to the same final size. This produces a set of equations that can be written in terms of $\mathcal{R} = \beta/\gamma$ and which describe the exact final size PMF for a LoT(N) network. The PMFs for a LoT network with $N = 21$ nodes are shown in Figure 3.17 for a range of $\mathcal{R}$ values.

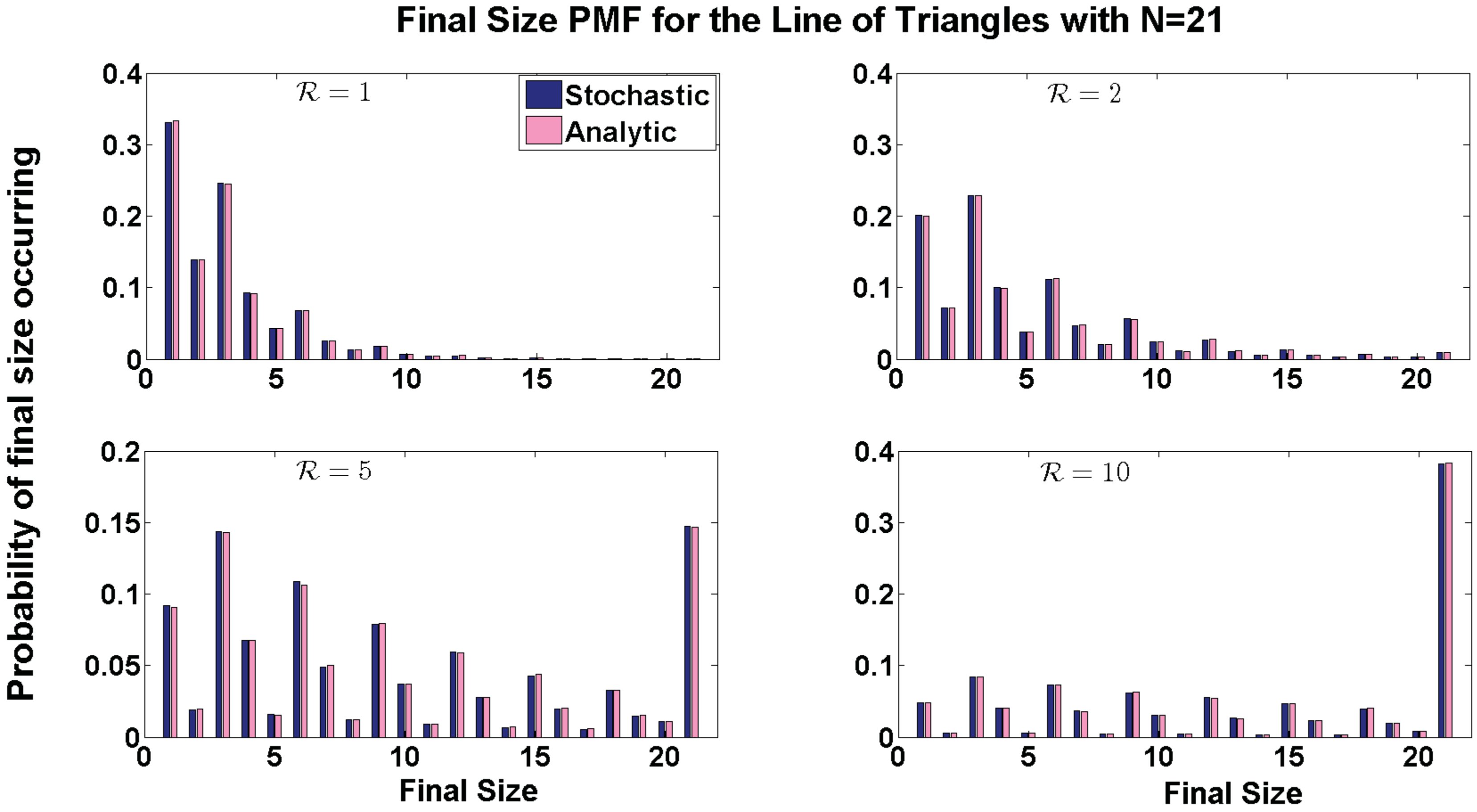


Figure 3.17: Probability mass functions for the final epidemic size with $\mathcal{R} = 1, 2, 5, 10$ for a line of triangles network with $N = 21$ nodes and $\mathbf{I_0} = a$. Dark blue and pink bars represent the stochastic and analytic results respectively.

Results for starting the epidemic with $\mathbf{I_0} = c$

Here we derive the probability that y apex nodes are infected, given that the infection ended at a given node (node x). Node x is one of four types of nodes as illustrated in Figure 3.15. Similarly to the previous section, here we derive the following probabilities for the LoT(N) network with $\mathbf{I_0} = c$.

$$P(y \text{ apex nodes infected} \mid \text{infection ended at node } x) \quad = \quad P(y| \text{ ended at } x)$$

As our new initial infectious node (node c) is not at the end of the line of triangles network, there is more to take into consideration than what was required to generalise the final size distributions starting with $\mathbf{I_0} = a$. In particular, there are now multiple ways we could define the probability the infection ends at node x. In this section we split the probability that the infection ends at node x into two equations as follows.

First, we have the probability that the infection ends at node x and node a is also infected. Secondly we have the probability that the infection ends at node x given that node a is never infected. Therefore, this probability includes all infection paths in which the infection spreads only to the right of node c and/or infection paths that reach node b but not a. As in the previous section, we have four categories that describe the probability the infection ends at a node of a specific type (type 1, type 2, apex node, ω_N), therefore we again have four categories here, however we will also have sub-categories. The two sub-categories will include probabilities where node a is infected and where node a is not infected. Keeping the probabilities separate that do and don't include node a is not required for finding the probability y apex nodes are infected given that the infection ends at node x, however, it is important for when we want to determine the probability of each final size occurring.

Throughout this section we will use the following notation to distinguish between probabilities where node a is and is not infected; $P(Y_a = y|x)$ denotes the probability that y apex nodes are infected given that the infection ends at node x in triangle τ and node a is infected. $P(Y = y|x)$ denotes the probability that y apex nodes are infected given that the infection ends at node x in triangle τ and node a is **not** infected. In the equations throughout this section we also continue to use the convention that $\binom{n}{k} = 0$ if $k < 0$ or $k > n$. As our aim of this section is to determine the final size probabilities we first show the contribution to the final size for each probability and then we derive the probabilities required. For $\mathbf{I_0} = c$, if node a was infected, then we obtain final sizes using the same equations as for $\mathbf{I_0} = a$, that is:

- $P(Y_a = y|T1)$ gives a final size equal to $2\tau + y$.
- $P(Y_a = y|T2)$ gives a final size equal to $2\tau - 1 + y$.

- $P(Y_a = y|\text{apex } \tau)$ gives a final size equal to $2\tau - 1 + y$.

- $P(Y_a = y|\omega_N)$ gives a final size equal to $2\Delta + y$.

For $\mathbf{I_0} = c$, if node a was not infected, we simply have to subtract one from each of the above equations to obtain:

- $P(Y = y|T1)$ gives a final size equal to $2\tau - 1 + y$.

- $P(Y = y|T2)$ gives a final size equal to $2\tau - 2 + y$.

- $P(Y = y|\text{apex } \tau)$ gives a final size equal to $2\tau - 2 + y$.

- $P(Y = y|\omega_N)$ gives a final size equal to $2\Delta - 1 + y$.

The probabilities required to determine the final size PMFs are as follows.

1. Given that the infection ended at a node of type 1 we define the following probabilities $P(Y_a = y|T1)$ and $P(Y = y|T1)$ below. In the following we have $\tau = 1, 2, \ldots \Delta - 1$ and $y = 0, 1, \ldots \tau$. Recall that ω_N is a node of type 1 in triangle Δ, however we define a different set of probabilities for reaching ω_N. Thus, for $P(Y_a = y|T1)$ and $P(Y = y|T1)$ we only look at the probability of ending at a node of type $T1$ in the first $\Delta - 1$ triangles.

$$P(Y_a = y|T1) = \sum_{i=1}^{4} P_i(Y_a = y|T1) \tag{3.27}$$

where $P_1(Y_a = y|T1)$ is the probability that the apex nodes in the first and $(\Delta - 1)^{th}$ triangles are not infected and y of the $\tau - 2$ apex nodes in the remaining triangles in the LoT network are infected.

$$\begin{aligned} P_1(Y_a = y|T1) &= \binom{\tau - 2}{y} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^y \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-y} P(c \text{ only } |\mathbf{I_0} = a) \end{aligned}$$

$P_2(Y_a = y|T1)$ is the probability that the apex node in the first triangle is not infected; the apex node in the $(\Delta - 1)^{th}$ triangle is infected and $y - 1$ of the $\tau - 2$ apex nodes in the remaining triangles in the LoT network are infected.

$$\begin{aligned} P_2(Y_a = y|T1) &= \binom{\tau - 2}{y - 1} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-1)} P(b \cap c \text{ only } |\mathbf{I_0} = a) \end{aligned}$$

$P_3(Y_a = y|T1)$ is the probability that the apex node in the first triangle is infected; the apex node in the $(\Delta - 1)^{th}$ triangle is not infected and $y - 1$

of the $\tau - 2$ apex nodes in the remaining triangles in the LoT network are infected.

$$\begin{aligned} P_3(Y_a = y|T1) &= \binom{\tau-2}{y-1} P(a \cap b \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-1)} P(c \text{ only } |\mathbf{I_0} = a) \end{aligned}$$

$P_4(Y_a = y|T1)$ is the probability that the apex nodes in the first and $(\Delta - 1)^{th}$ triangles are infected and $y - 2$ of the $\tau - 2$ apex nodes in the remaining triangles in the LoT network are infected.

$$\begin{aligned} P_4(Y_a = y|T1) &= \binom{\tau-2}{y-2} P(a \cap b \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-2} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-2)} P(b \cap c \text{ only } |\mathbf{I_0} = a) \end{aligned}$$

Similarly, we have

$$P(Y = y|T1) = \sum_{i=1}^{4} P_i(Y = y|T1) \tag{3.28}$$

where $P_1(Y = y|T1)$ is the probability that nodes a and b do not get infected; the apex node in the last triangle does not get infected and y of the $\tau - 2$ apex nodes in the remaining triangles in the LoT are infected.

$$\begin{aligned} P_1(Y = y|T1) &= \binom{\tau-2}{y} P(d \text{ only } |\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-y} P(c \text{ only } |\mathbf{I_0} = a) \end{aligned}$$

$P_2(Y = y|T1)$ is the probability that nodes a and b do not get infected; the apex node in the last triangle does get infected and $y - 1$ of the $\tau - 2$ apex nodes in the remaining triangles in the LoT are infected.

$$\begin{aligned} P_2(Y = y|T1) &= \binom{\tau-2}{y-1} P(d \text{ only } |\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-1)} P(b \cap c \text{ only } |\mathbf{I_0} = a) \end{aligned}$$

$P_3(Y = y|T1)$ is the probability that node a does not get infected and node b does get infected; the apex node in the last triangle does not get infected and $y - 1$ of the $\tau - 2$ apex nodes in the remaining triangles in the LoT are infected.

$$\begin{aligned} P_3(Y = y|T1) &= \binom{\tau-2}{y-1} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\times \quad P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-1)} P(c \text{ only } |\mathbf{I_0} = a) \end{aligned}$$

$P_4(Y = y|T1)$ is the probability that node a does not get infected and node b does get infected; the apex node in the last triangle does get infected and $y-2$ of the $\tau - 2$ apex nodes in the remaining triangles in the LoT are infected.

$$\begin{aligned} P_4(Y = y|T1) &= \binom{\tau-2}{y-2} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-2} \\ &\times \; P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-2)} P(b \cap c \text{ only } |\mathbf{I_0} = a) \end{aligned}$$

We note a special case in addition to the above equations is when $\tau = 1$, that is the infection is contained within the first triangle in the LoT network. In this case we have the following:

If $y = 0$ we have:

$$\begin{aligned} P(Y_a = 0|T1) &= P(a \text{ only } |\mathbf{I_0} = c) \\ P(Y = 0|T1) &= \frac{1}{3\mathcal{R}+1} \end{aligned} \tag{3.29}$$

where $P(a \text{ only } |\mathbf{I_0} = c)$ is the probability that only nodes a and c are ever infected (note: this is the same as the probability that only nodes b and c are ever infected, due to the symmetry of the nodes); $\frac{1}{3\mathcal{R}+1}$ is the probability that node c recovers before the infection spreads (which is equal to the probability of the final size being one).
If $y = 1$ we have:

$$\begin{aligned} P(Y_a = 1|T1) &= P(a \cap b \text{ only } |\mathbf{I_0} = c) \\ P(Y = 1|T1) &= P(a \text{ only } |\mathbf{I_0} = c) \end{aligned} \tag{3.30}$$

where $P(a \cap b \text{ only } |\mathbf{I_0} = c)$ is the probability that nodes a, b and c are the only nodes infected. Note, as mentioned above, in the second equation $P(a \text{ only } |\mathbf{I_0} = c)$ is equal to the probability that only nodes b and c are infected (i.e. node a is not infected).

2. Given that the infection ends at a node of type 2 we define the following probabilities $P(Y_a = y|T2)$ and $P(Y = y|T2)$ below. We continue to use the same notation as outlined above. For the following probabilities we do not include node a as a node of type 2 and therefore we start with node d as the first type 2 node in any LoT network. The probabilities that the infection is contained within the first triangle ($\tau = 1$) of the LoT network are included in the equations described above and as such we do not include them again here. Therefore, for the probability the infection ends at a node of type 2 in triangle τ, we have $\tau = 2, 3, \ldots \Delta$ and $y = 0, 1, \ldots \Delta - 1$.

$$\begin{aligned}P(Y_a = y|T2) &= \binom{\tau-2}{y} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-y} \\ &\quad \times P(b \cap c \cap d|\mathbf{I_0} = a)^{y} \left(\frac{1}{2\mathcal{R}+1}\right) \\ &+ \binom{\tau-2}{y-1} P(a \cap b \cap d|\mathbf{I_0} = c) P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-1)} \\ &\quad \times P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \left(\frac{1}{2\mathcal{R}+1}\right) \qquad (3.31)\end{aligned}$$

$$\begin{aligned}P(Y = y|T2) &= \binom{\tau-2}{y} P(d \text{ only } |\mathbf{I_0} = c) P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-y} \\ &\quad \times P(b \cap c \cap d|\mathbf{I_0} = a)^{y} \left(\frac{1}{2\mathcal{R}+1}\right) \\ &+ \binom{\tau-2}{y-1} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-1)} \\ &\quad \times P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \left(\frac{1}{2\mathcal{R}+1}\right) \qquad (3.32)\end{aligned}$$

where $\frac{1}{2\mathcal{R}+1}$ is the probability that last node of type 2 to be infected in the network recovers before the infection spreads further.

3. Given that the infection ends at an apex node in triangle τ, we define the following probabilities $P(Y_a = y|\text{apex } \tau)$ and $P(Y = y|\text{apex } \tau)$ below. As mentioned above for the probability the infection ends at a node of type 2, here we also only consider $\tau = 2 : \Delta$ as all probabilities that the infection is contained within the first triangle (nodes a, b and c) are included in our equations for $P(Y_a = y|T1)$ and $P(Y = y|T1)$ above. Therefore, the below equations are for $\tau = 2, 3, \ldots \Delta$ and $y = 1, 2, \ldots \tau$.

$$\begin{aligned}P(Y_a = y|\text{apex } \tau) &= \binom{\tau-2}{y-2} P(a \cap b \cap d|\mathbf{I_0} = c) P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-2)} \\ &\quad \times P(b \cap c \cap d|\mathbf{I_0} = a)^{y-2} P(a \cap b \text{ only}) \\ &+ \binom{\tau-2}{y-1} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-1)} \\ &\quad \times P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} P(a \cap b \text{ only}) \qquad (3.33)\end{aligned}$$

$$\begin{aligned}P(Y = y|\text{apex } \tau) &= \binom{\tau-2}{y-2} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-2)} \\ &\quad \times P(b \cap c \cap d|\mathbf{I_0} = a)^{y-2} P(a \cap b \text{ only}) \\ &+ \binom{\tau-2}{y-1} P(d \text{ only } |\mathbf{I_0} = c) P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\tau-2)-(y-1)} \\ &\quad \times P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} P(a \cap b \text{ only}) \qquad (3.34)\end{aligned}$$

4. Given that the infection ends at ω_N, we define the following probabilities $P(Y_a = y|\omega_N)$ and $P(Y = y|\omega_N)$ below.

$$P(Y_a = y|\omega_N) = \sum_{i=1}^{4} P_i(Y = y|\omega_N) \tag{3.35}$$

where

$$\begin{aligned} P_1(Y_a = y|\omega_N) &= \binom{\Delta - 2}{y} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^y \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-2)-y} P(a \cap b' \cap c) \end{aligned}$$

$$\begin{aligned} P_2(Y_a = y|\omega_N) &= \binom{\Delta - 2}{y-1} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-2)-(y-1)} P(a \cap b \cap c) \end{aligned}$$

$$\begin{aligned} P_3(Y_a = y|\omega_N) &= \binom{\Delta - 2}{y-1} P(a \cap b \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-2)-(y-1)} P(a \cap b' \cap c) \end{aligned}$$

$$\begin{aligned} P_4(Y_a = y|\omega_N) &= \binom{\Delta - 2}{y-2} P(a \cap b \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-2} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-2)-(y-2)} P(a \cap b \cap c) \end{aligned}$$

Similarly, we have

$$P(Y = y|\omega_N) = \sum_{i=1}^{4} P_i(Y = y|\omega_N) \tag{3.36}$$

where

$$\begin{aligned} P_1(Y = y|\omega_N) &= \binom{\Delta - 2}{y} P(d \text{ only } |\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^y \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-2)-y} P(a \cap b' \cap c) \end{aligned}$$

$$\begin{aligned} P_2(Y = y|\omega_N) &= \binom{\Delta - 2}{y-1} P(d \text{ only } |\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-2)-(y-1)} P(a \cap b \cap c) \end{aligned}$$

$$\begin{aligned} P_3(Y = y|\omega_N) &= \binom{\Delta - 2}{y-1} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-1} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-2)-(y-1)} P(a \cap b' \cap c) \end{aligned}$$

$$\begin{aligned} P_4(Y = y|\omega_N) &= \binom{\Delta - 2}{y-2} P(a \cap b' \cap d|\mathbf{I_0} = c) P(b \cap c \cap d|\mathbf{I_0} = a)^{y-2} \\ &\quad \times P(b' \cap c \cap d|\mathbf{I_0} = a)^{(\Delta-2)-(y-2)} P(a \cap b \cap c) \end{aligned}$$

We note that (as we saw for the equations with $\mathbf{I_0} = a$ in the previous section) the above two Equations 3.35 and 3.36 are the same as Equations 3.27 and 3.28 with the last probability swapped out.

To generate the final size PMF for a line of triangles network with N nodes we find all possible conditional probabilities as defined by Equations 3.29 to 3.36 and the final size that each probability corresponds to (see above). To find the final size PMF we sum up each of the equations that correspond to the same final size. This produces a set of equations that can be expressed in terms of $\mathcal{R} = \beta/\gamma$ and which describe the exact final size PMF for a LoT(N) network. The PMFs for a LoT network with $N = 21$ nodes are shown in Figure 3.18 for a range of $\mathcal{R}$ values.

3.3 Discussion

Figures 3.17 and 3.18 show the PMF for the final epidemic size of the line of triangles network with $N = 21$ nodes for different values of $\mathcal{R}$ with $\mathbf{I_0} = a$ and $\mathbf{I_0} = c$ respectively. The final size PMFs exhibit some interesting behaviour. We notice, as expected, that in the PMFs for $\mathcal{R} = 5$ and $\mathcal{R} = 10$ the probability that the final size is equal to 2,5,8,11,14,17 and 20 are much lower than the other probabilities. Similar behaviour can be seen in the PMFs for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ however as the transmission rate in these are lower the difference is less pronounced in the figures. Thus the difference between final size probabilities is more prominent in the PMFs with a higher transmission rate. The final size PMFs exhibit the same behaviour as distributions for the probability the infection ends at a given node. The probability that the infection ends at an apex node (i.e. nodes 2,5,8,11,14,17 and 20) in the LoT network with 21 nodes is much lower than the probability the infection ends at a node of type 1 or 2 (i.e. nodes along the baseline of the network). This behaviour is due to the baseline nodes, with the exception of node a and N, being more connected (having a higher degree) than the apex nodes and thus they have a higher probability of becoming infected at some point during an epidemic.

In this chapter we have shown that it is possible to describe the way an infection spreads through a line of triangles network based on our knowledge of the triangle and lollipop networks studied in an earlier chapter. We have used this knowledge to generalise the infection paths and PMF for the final epidemic size of a line of triangles network for an arbitrary number of nodes, N. We have illustrated a method to find the probability mass functions for the final epidemic size for a LoT(N) network starting with node a, b or c as the initial infectious node, $\mathbf{I_0}$. As a result we have found a way to obtain analytic results that describe the way an infection spreads through a LoT network. We compared our analytical results to results found from a stochastic model to ensure our methods are correct. We have shown that the difference between analytic and stochastic results tends toward zero as we increase the number of realisations the stochastic result is averaged over. For example, Figure 3.19 shows the difference between analytic and stochastic results for the probability the infection reaches a given node. Figure 3.20 illustrates the convergence of the difference between analytic and stochastic results for the probability k apex nodes were infected given that the infection ended at node j. See Appendix B on page 237 for more figures and tables.

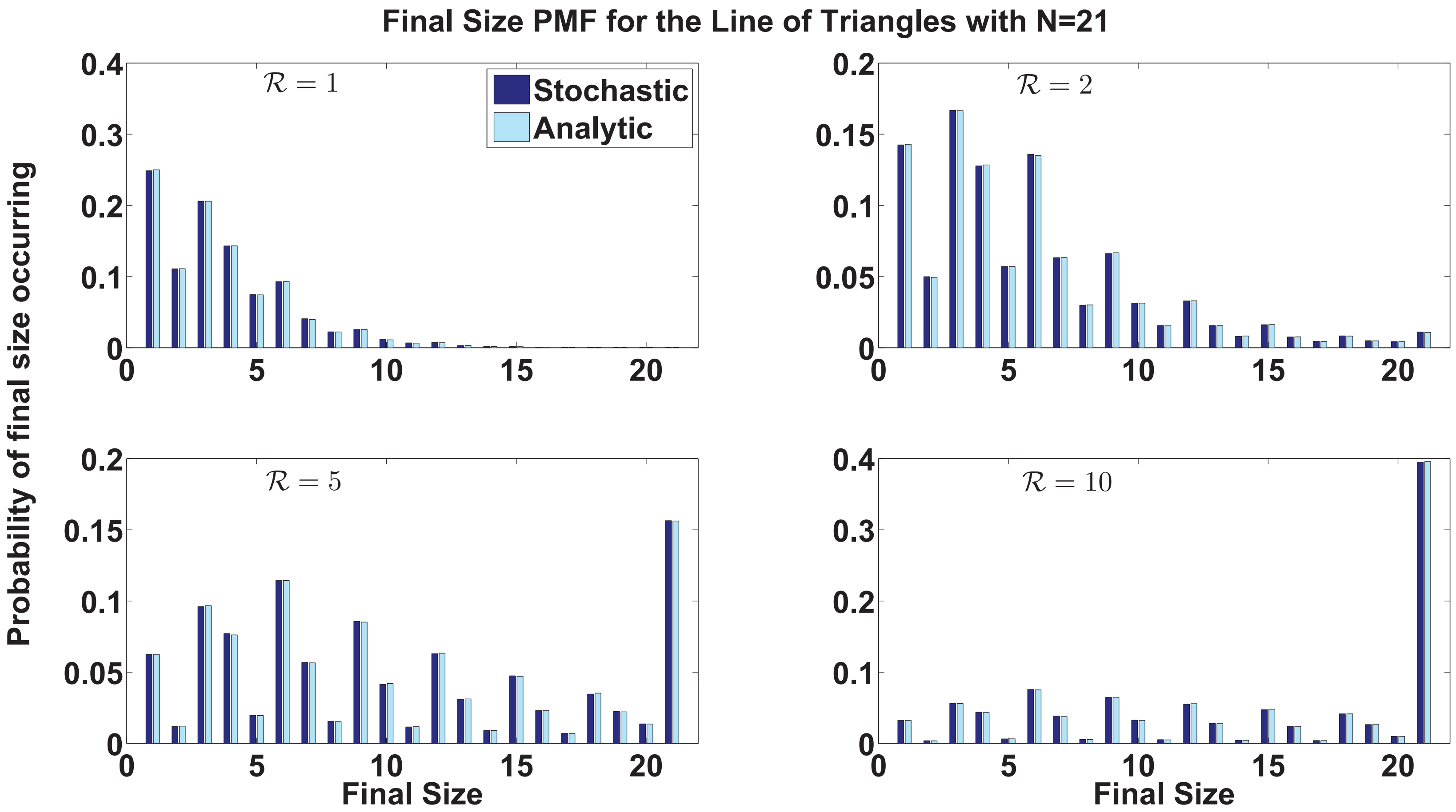


Figure 3.18: Probability mass functions for the final epidemic size with $\mathcal{R} = 1, 2, 5, 10$ for a line of triangles network with $N = 21$ nodes and $\mathbf{I_0} = c$. Dark blue and light blue bars represent the stochastic and analytic results respectively.

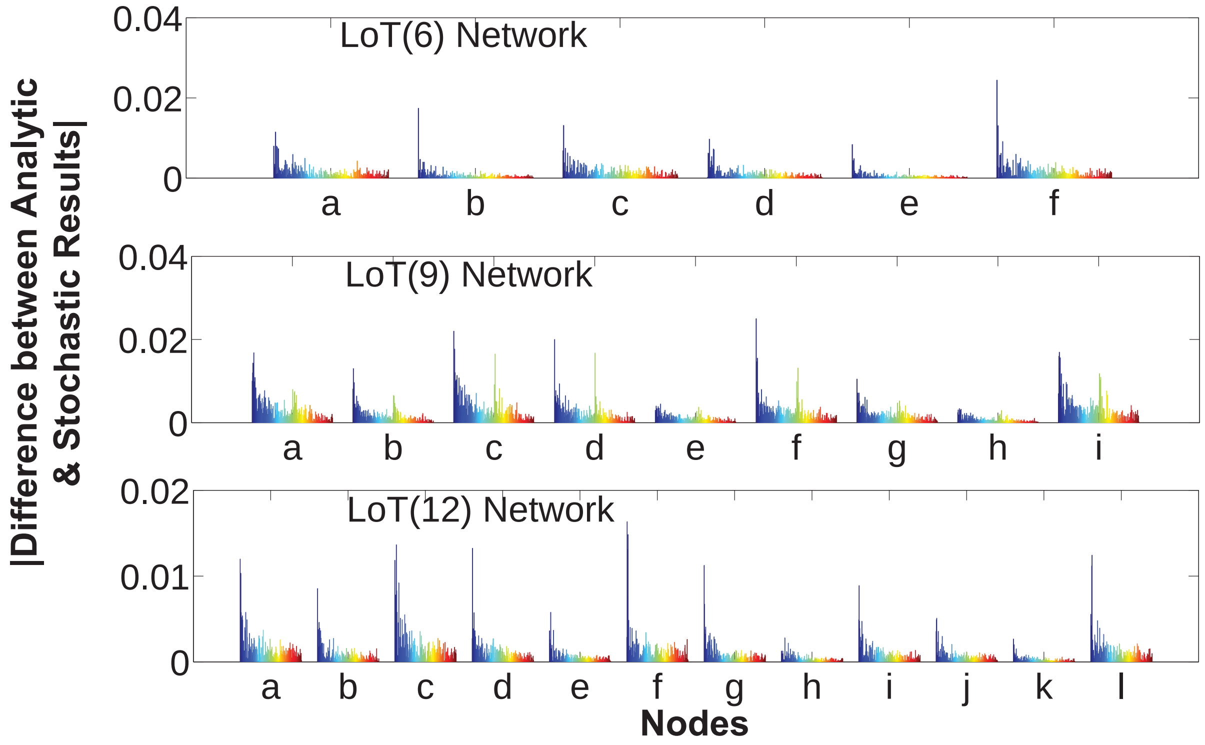


Figure 3.19: Absolute value of the difference between analytic and stochastic results for the probability infection ends at each node in a LoT network. For each node, from left to right (blue to red) we are increasing the number of stochastic realisations (from 1×10^3 to 2×10^5 in steps of 1×10^3) that our stochastic probabilities were averaged over. Thus, for each node we are looking for the difference to converge to zero if the accuracy of the stochastic realisation result gets closer to the exact analytic probability.

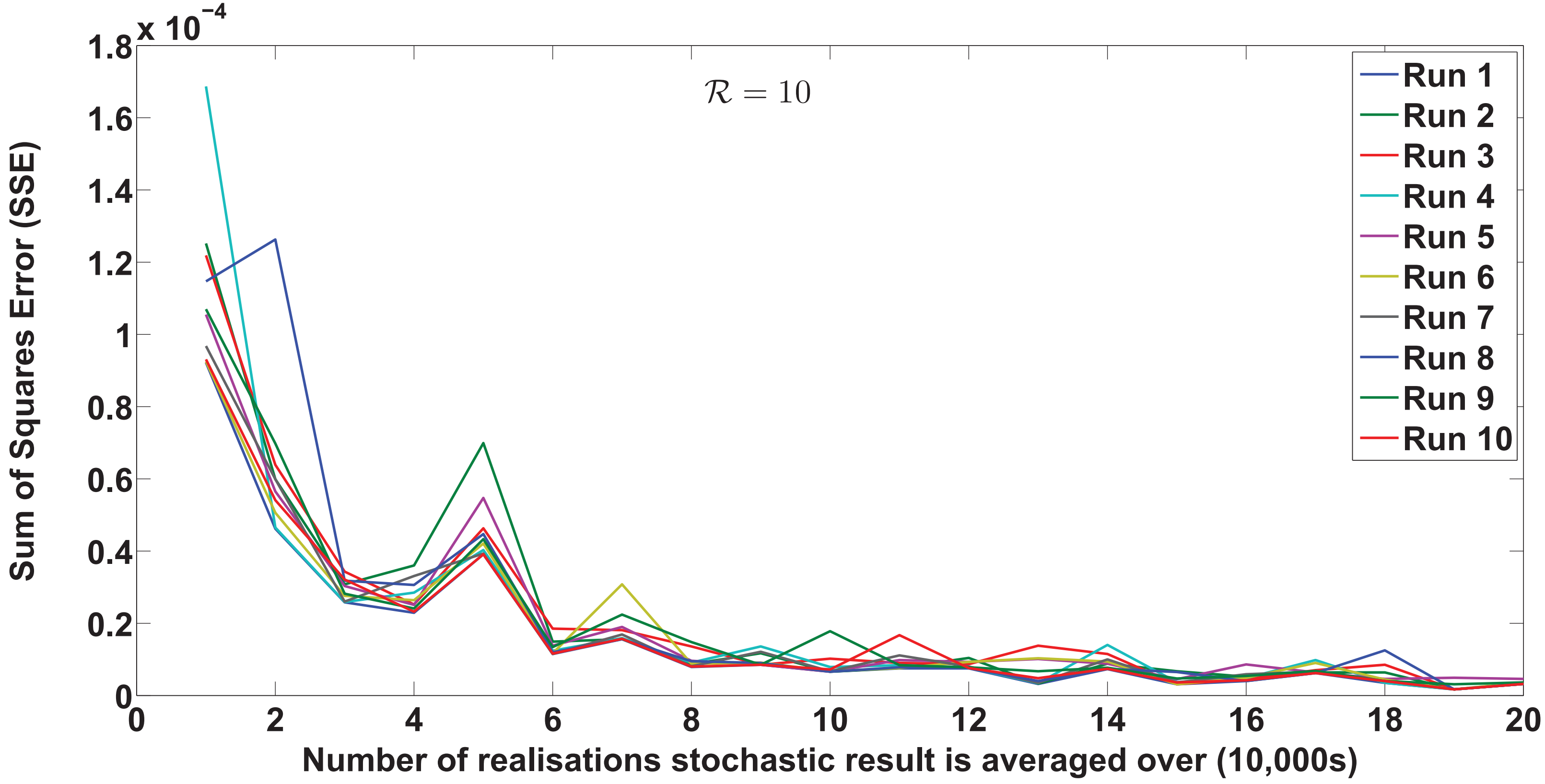


Figure 3.20: Convergence of the difference between analytic and stochastic results for the probability k apex nodes were infected given that the infection ended at node j in the LoT network with $N = 21$ nodes. For each run, the stochastic result was averaged over for an increasing number of realisations and the sum of squares error was computed.

Chapter 4

Illustrations of an SIR model on small networks

Here we illustrate how the network models presented in Chapter 2 can be used to investigate the dynamics of the spread of an infection through small populations that could be representative of a household or community. We will focus on two infections, influenza and measles, both of which are well known to follow an SIR model for the host infection lifetime [34, 41, 59]. In order to obtain results specific to influenza and measles from our network models we need to determine what the infection and recovery rates (β and γ respectively) are for both infections. In some cases we can obtain β and γ directly from previous studies; although in most studies the basic reproduction number, R_0, and the recovery rate are given. If this is the case, then it is necessary to be able to relate the basic reproduction number to the infection transmission rate (β) before we can investigate the infection dynamics using our network models. Therefore, in Section 4.1 we discuss the different ways that the basic reproduction number can be defined on a network, followed by the definition and derivation of R_0 for each of the small networks presented in previous chapters. In Sections 4.2 and 4.3 we parameterise the network models for influenza and measles respectively. In Section 4.4 we present and discuss the results.

4.1 A discussion about the basic reproduction number, R_0.

Arguably, one of the most important quantities in infectious disease modelling is the basic reproduction number, R_0. By definition R_0 is the expected number of secondary cases produced by a single 'typical' infectious individual in a wholly susceptible population [16]. It is often used to determine the severity of an emerging epidemic or predict what level of intervention (such as vaccination rates) would be required to contain and/or eliminate the infection. In general if $R_0 < 1$, it is expected that there will be a minor epidemic and if $R_0 > 1$ a major epidemic is possible. Having $R_0 < 1$ and $R_0 > 1$ equates to the 'typical' infectious individual transmitting infection to on average less than one or more than one other individual respectively.

However, there is much ambiguity in the definition and use of R_0, for example how does one define a 'typical' infectious individual? If we consider the population of New Zealand, would a typical infectious individual be someone who lives in the Auckland CBD, has a large number of social connections and travels frequently or would it be an individual who lives in a small rural community and doesn't travel past the surrounding suburbs? As one could imagine the expected number of secondary infections caused by the first individual would be considerably higher than that of the second individual. One question that arises when using R_0 to infer possible outcomes of an epidemic is how does one define a minor and a major epidemic? We could look at the fraction of susceptible individuals left in the population at the end of an epidemic to determine if it was a minor or major epidemic. For an SIR model in a homogeneously mixing population the infection peaks when the fraction of susceptible individuals is equal to $1/R_0$. If the fraction of susceptible individuals at the end of an epidemic is less than $1/R_0$ this would mean that the infection died out due to a lack of susceptible hosts available and thus this could be defined as a major epidemic. If the fraction of susceptible individuals at the end of an epidemic was greater than $1/R_0$ this would mean that the infection died out on its own, sometimes referred to as stochastic fade out, and it was not due to a lack of susceptible hosts. In this case we would refer to it as a minor epidemic.

Another important factor that will affect R_0 is differences in susceptibility and infectiousness of individuals within a population. For example, the average number of individuals a typical individual with measles could infect will be much higher in a population that has a low percentage of vaccinated individuals than in a population with a high percentage of vaccinated individuals. The definition of R_0 states that there is a single infectious individual in a wholly susceptible population. However, assuming that the entire population is susceptible is not always accurate. Therefore, having a similar quantity to R_0 defined when the population is not entirely susceptible can be useful. The effective reproduction number (R) is the expected number of secondary infections that arise from a single infectious individual in the population which is not necessarily entirely susceptible (e.g. when interventions such as vaccination campaigns are put in place some individuals will be immune) [12].

The structure of a given population will also have an effect on R_0, for large homogeneously mixing populations the derivation of R_0 is well known. However, it is also well known that assuming a population mixes homogeneously is inaccurate in many cases. So, the question arises, how can we define the basic reproduction number for heterogeneous populations? A number of definitions for R_0 have been derived for models that have a network representation of the given population.

4.1.1 Definitions of R_0 for network models

There are three definitions of the basic reproduction number, R_0, for network models which we discuss in this section. Expressions for these definitions are given in Table 4.1. We begin by explaining how the first equation for R_0 in Table 4.1 is derived. From the definition given in [16], it would seem logical to define R_0 for the initial infectious node in a network. As all k neighbours of the initial infectious node will be susceptible, we can define the average number of secondary cases produced by a 'typical' node as:

$$R_0 = \psi\langle K\rangle$$

where $\psi = \dfrac{\beta}{\beta+\gamma}$ is the per edge probability of transmission, the average degree of a network is denoted by $\langle K\rangle = \sum_k kP(k)$ and $P(k)$ is the probability that a given node has degree k [14]. Note here, we use the average degree of the network as the 'typical' initial infectious node would be randomly chosen.

However, the value of R_0 for the initial infectious node will differ from that of a 'typical' infectious node from the early stages of an epidemic. For this definition of R_0 we will need to take into account the fact that the chosen node would have already been infected by one of its neighbours and hence can only transmit the infection to at most $k-1$ nodes. The probability a newly infected node has degree k is $\dfrac{kP(k)}{\langle K\rangle}$ where $kP(k)$ is the expected number of nodes in the network with degree k. We can define the expected number of new infections caused by a newly infected node as:

$$\begin{aligned} R_0 &= \psi\sum_k \frac{kP(k)}{\langle K\rangle}(k-1) \\ &= \psi\sum_k \frac{k^2P(k)-kP(k)}{\langle K\rangle} \\ \therefore R_0 &= \psi\frac{\langle K(K-1)\rangle}{\langle K\rangle} \qquad (4.1) \end{aligned}$$

Note that this derivation assumes that clustering is unimportant and thus the only infectious neighbour the newly infected node has is the one that infected it. If clustering is present then the newly infected node could have more than one infectious

Table 4.1: **Comparison of R_0 for SIR models on heterogeneous populations.**

Equation for R_0	**Model assumptions**	**References**
$R_0 = \psi \frac{\langle K(K-1) \rangle}{\langle K \rangle}$	Large random networks **without** clustering generated using the configuration model. Small initial fraction of population infected. Miller et al [46] derived this expression for R_0 using the EBCM approach with the configuration and mixed Poisson network models.	[39], [40], [43], [46], [48], [68], [50]
$R_0 = (1 - \psi\phi)\psi \frac{\langle K(K-1) \rangle}{\langle K \rangle}$	Large random networks **with** clustering generated using the configuration model. Small initial fraction of population infected.	[48]
$R_0 = \frac{\langle K(K-1) \rangle}{\langle K \rangle}\psi - \frac{2\langle n_t \rangle}{\langle K \rangle}\psi^2 - \frac{\langle n_s \rangle}{\langle K \rangle}\psi^3 + \mathcal{O}\left(\frac{\psi^4}{\langle K \rangle}\right)$	Large random networks **with** clustering generated using the configuration model. Small initial fraction of population infected, where n_t and n_s are the number of triangles and squares respectively in the network.	[43]

neighbour which would reduce the effective reproduction number. Refer to Table 4.1 for other assumptions made for this derivation.

The second equation for R_0 in Table 4.1 was derived by Molina and Stone [48] and incorporates the effects that clustering may have on the basic reproduction number. The equation for R_0 that they derive is the same as in Equation 4.1 with an added correction factor of $(1 - \psi\phi)$. If the clustering coefficient, ϕ, is zero then this is in agreement with R_0 in Equation 4.1 for unclustered networks. Molina and Stone [48] show that clustering results in a reduced R_0 and increased epidemic threshold when compared with an infection spreading on a network where $\phi = 0$. The clustering coefficient counts the number of triangles in the network but does not account for loops of higher order. Miller [43] derived an equation for R_0 that incorporates the effects of higher order motifs (such as loops of order 4), as shown in Table 4.1.

The basic reproduction number, R_0, is usually calculated as an average over all individuals within a population where there is only one species of host. So, what happens if we are investigating the spread of an infection with multiple hosts such as the transmission of malaria? For malaria, there is transmission from mosquitos to humans and vice versa. By definition the basic reproduction number will find the average number of secondary infections caused by humans and mosquitos. One has to ask, why would you want to average over humans and mosquitos? This does not make biological sense as the transmission rate from mosquitos to humans would be different to the transmission rate from humans to mosquitos. The type reproduction number was introduced by Roberts and Heesterbeek [56] to account for individuals that can be distinguished by some epidemiologically significant characteristic such as species (See also [24]).

4.1.2 Defining R_0 for small networks

In light of the discussion in the previous sections, strictly speaking it does not make sense to define R_0 for small networks. However, as R_0 is one of the most commonly estimated quantities for infectious diseases we need some sort of approximation in order to incorporate the infection parameters with the small network models. Here, we will discuss how we define an approximation to the average number of secondary infections produced by a single 'typical' infectious node in each small network. We use a NGM (Next Generation Matrix) approach [16, 18, 57] and divide the nodes up into types based on their degree (or number of neighbours). For example, in a star network of four nodes there is one node of type 1 (central node of degree 3) and three nodes of type 2 (the outer nodes of degree 1). It should be clear that the expected number of neighbours that the central node infects will be much higher than the expected number of neighbours an outer node infects. The entries $\mathbf{K}_{ij}$ of the NGM define the expected number of individuals of type i infected by a 'typical' individual of type j. For our small networks we have the following.

$$\begin{aligned}
\mathbf{K}_{ij} &= \text{E}[\ \#\text{ nodes of type } i \text{ infected by a node of type } j\ ] \\
&= \text{P}(\text{transmission occurs across an } S-I \text{ edge}) \\
&\qquad \times \quad (\#\text{ neighbours of type } i \text{ a node of type } j \text{ has}) \\
\therefore \mathbf{K}_{ij} &= \psi \times (\#\text{ neighbours of type } i \text{ a node of type } j \text{ has})
\end{aligned}$$

We can define the basic reproduction number, R_0, as the maximum eigenvalue of the NGM, $\mathbf{K}$ (see Diekman et al [17]). This gives us the average number of secondary infections resulting from a randomly chosen initial infectious node in each small network. For example, the NGM for a line network with $N = 3$ nodes is constructed in the following way. There are two types of nodes in the line network (type 1 has degree 1 and type 2 has degree 2), so we first define the NGM $\mathbf{K}$:

$$\mathbf{K} = \psi \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$$

We note that the sum of each column gives the degree of each type of node in the network. Then, R_0 is the dominant eigenvalue of $\mathbf{K}$ and can be found by solving $\text{Det}\,(\mathbf{K} - \lambda\mathbf{I}) = 0$ for λ, where $\mathbf{I}$ is the identity matrix of the same order as $\mathbf{K}$. Therefore, we have

$$R_0 = \sqrt{2}\psi$$

We follow the above method to determine the NGM and R_0 for each small network of three and four nodes, as shown in Table 4.2.

4.2 Influenza

Influenza is a viral infection that usually lasts for one to two weeks; symptoms can include fever, muscle aches, sore throat and coughing. The virus spreads easily between people when infected individuals cough and sneeze, often resulting in seasonal epidemics. Substantial morbidity and mortality is caused annually by influenza epidemics all over the world [11]. Influenza viruses have caused global pandemics throughout history and will probably continue to do so as new strains of the viruses emerge. See Neumann et al [49] for a summary of the difference in origins and severity between the 'Spanish' influenza (H1N1) of 1918-1919, 'Asian' influenza (H2N2) of 1957, 'Hong Kong' influenza (H3N2) of 1968 and the 'Russian' influenza (H1N1) of 1977. In more recent history there was a pandemic of 'swine flu' (H1N1) in 2009 which led to a global effort to determine epidemiological parameters that could be used to predict the severity of the pandemic and control efforts required.

Roberts and Nishiura [58] analysed data from pandemic influenza H1N1 2009 in New Zealand. They estimated the effective reproduction number to be 1.25 with a 95% confidence interval of (1.07,1.47) and showed that earlier studies overestimated the effective reproduction number. The overestimation was a consequence of not accounting for factors such as the infection-age distribution of imported cases and the delay in transmission dynamics due to international travel. Tuite et al [64] estimated R_0 to be 1.31 with a 95% confidence interval of (1.25, 1.38) and the infectious period to be approximately 3.38 days with a 95% confidence interval of (2.06, 4.69) for the influenza A H1N1 pandemic in Ontario, Canada in 2009.

Table 4.2: Next Generation Matrices and R_0 for small networks.

Network	NGM (if applicable)	R_0
Triangle	Each node has degree $k = 2$.	$R_0 = 2\psi$
Line ($N = 3$)	$\mathbf{K} = \psi \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$	$R_0 = \sqrt{2}\psi$
Complete ($N = 4$)	Each node has degree $k = 3$.	$R_0 = 3\psi$
Square	Each node has degree $k = 2$.	$R_0 = 2\psi$
Star ($N = 4$)	$\mathbf{K} = \psi \begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix}$	$R_0 = \sqrt{3}\psi$
Toast	$\mathbf{K} = \psi \begin{pmatrix} 0 & 2 \\ 2 & 1 \end{pmatrix}$	$R_0 = \frac{1 + \sqrt{17}}{2}\psi$
Line ($N = 4$)	$\mathbf{K} = \psi \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$	$R_0 = \frac{1 + \sqrt{5}}{2}\psi$
Lollipop	$\mathbf{K} = \psi \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 1 & 0 \end{pmatrix}$	$R_0 = 2.1701\psi$

In order to parameterise our model to investigate how influenza spreads through small networks we need to determine what the transmission and recovery rates (β and γ respectively) are. We approximate R_0 by the effective reproduction number found in Roberts & Nishiura [58] and approximate the infectious period ($1/\gamma$) to be 3.88 days as found in Tuite et al [64]. As we were unable to find an estimate of the infectious period for the New Zealand influenza pandemic, we argue that influenza A H1N1 in New Zealand and Ontario have approximately the same infectious period.

As R_0 has been estimated for a large homogeneous population, we set $R_0 = 1.25$ for the complete network of $N = 4$. This allows us to rearrange the equation to determine an approximation for β, which we can then use to determine the probability of transmission given contact (ψ) for all small networks. Therefore, from Table 4.2, we rearrange the R_0 equation corresponding to the complete network of $N = 4$ nodes to find β. We then substitute in $R_0 = 1.25$ and $\gamma = 0.2577$ to obtain the following estimate for β.

$$\beta \quad = \quad \frac{\gamma R_0}{3 - R_0} = 0.1841$$

Using these parameters we estimate the probability of transmission given contact for influenza to be $\psi = 0.4167$. We present and discuss the results in Section 4.4.

4.2.1 Results

The expected number of secondary infections resulting from a single infectious individual, R_0, will differ between network structures as it depends on the probability of transmission given contact and number of contacts each individual has. Table 4.3 shows our estimates of the basic reproduction number, R_0, for each small network. To obtain results we evaluate (using the parameter values specified above) the parameter $\mathcal{R} = \beta/\gamma = 0.7144$. We then input these values of $\mathcal{R}$ into the final size equations for each network to find the final size distributions for each specified initial state. We also found the expected final size as shown in Table 4.4. Figures 4.1 to 4.6 show the final size probability distributions for influenza, along with the corresponding initial state.

Table 4.3: R_0 estimates for influenza.

Network	**Value of R_0** ($\psi = 0.4167$)
Triangle	0.83
Line ($N = 3$)	0.59
Complete ($N = 4$)	1.25
Square	0.83
Star ($N = 4$)	0.72
Toast	1.07
Line ($N = 4$)	0.67
Lollipop	0.9

Table 4.4: Expected Final Size (EFS) for influenza on small networks

Networks with $N = 3$				**EFS**
	Clustering Coefficient, ϕ	**Initial state**	**Degree of Initial Infectious Node(s)**	**Influenza**
A. Triangle	1	SSI	2	1.97
B. Line	0	SSI	1	1.59
		SIS	2	1.83
		ISI	1 & 1	2.66
Networks with $N = 4$				
C. Complete	1	$SSSI$	3	2.68
D. Square	0	$SSSI$	2	2.2
		$SIIS$	2 & 2	3.32
E. Star	0	$SSSI$	1	1.76
		$ISSS$	3	2.25
		$SSII$	1 & 1	2.93
		$SIII$	1, 1 & 1	3.8
F. Toast	0.75	$SSIS$	2	2.33
		$SSSI$	3	2.55
		$SIIS$	2 & 2	3.46
G. Line	0	$SSSI$	1	1.66
		$SSIS$	2	2.01
		$SISI$	2 & 1	3.08
		$ISSI$	1 & 1	3.04
H. Lollipop	0.6	$SSSI$	1	1.82
		$SISS$	2	2.18
		$ISSS$	3	2.39
		$SSII$	2 & 1	3.26

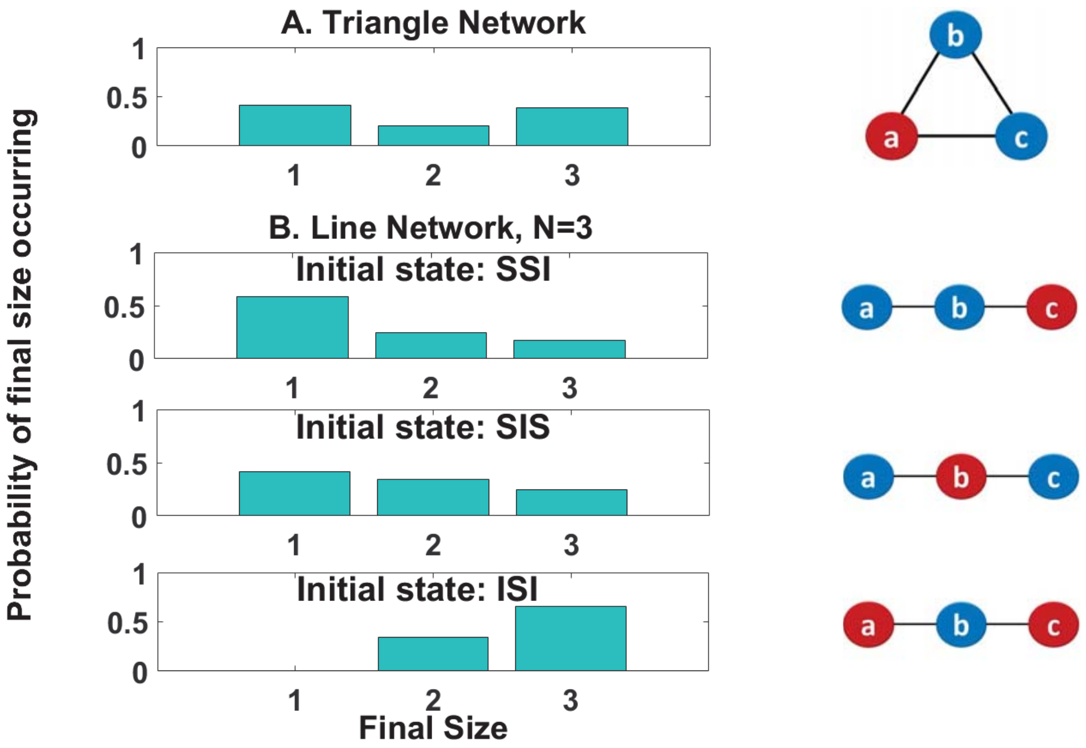


Figure 4.1: Probability mass functions for the spread of influenza through the triangle and line networks with $N = 3$. Schematics of the initial state are shown on the right.

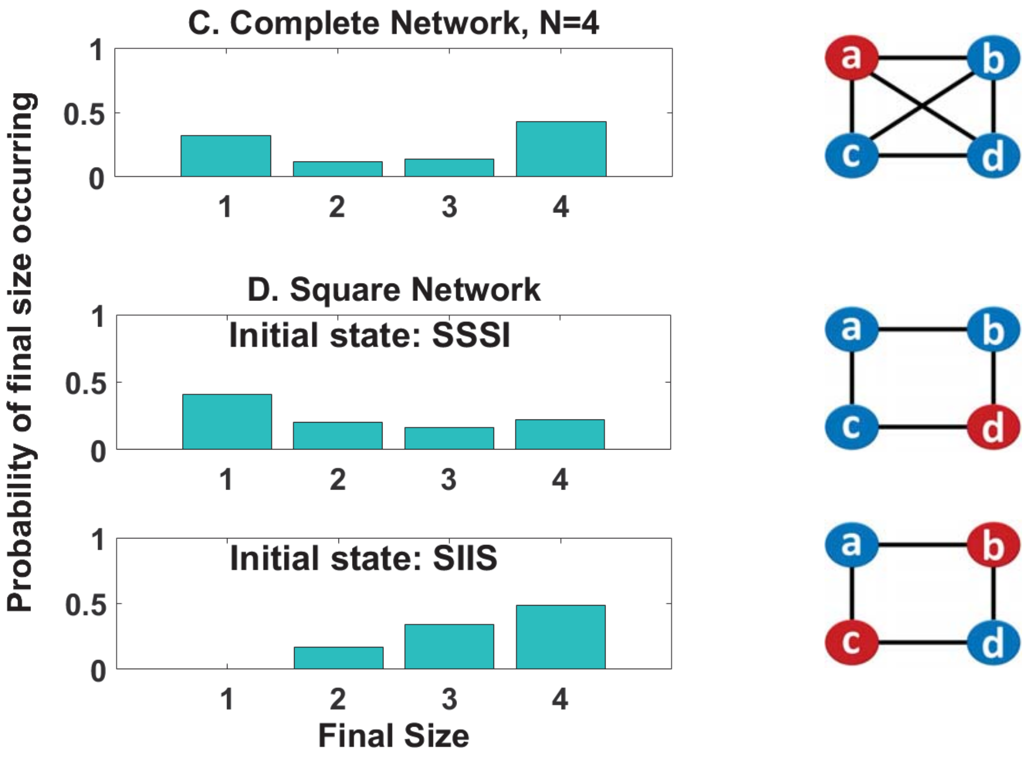


Figure 4.2: Probability mass functions for the spread of influenza through the complete and square networks with $N = 4$. Schematics of the initial state are shown on the right.

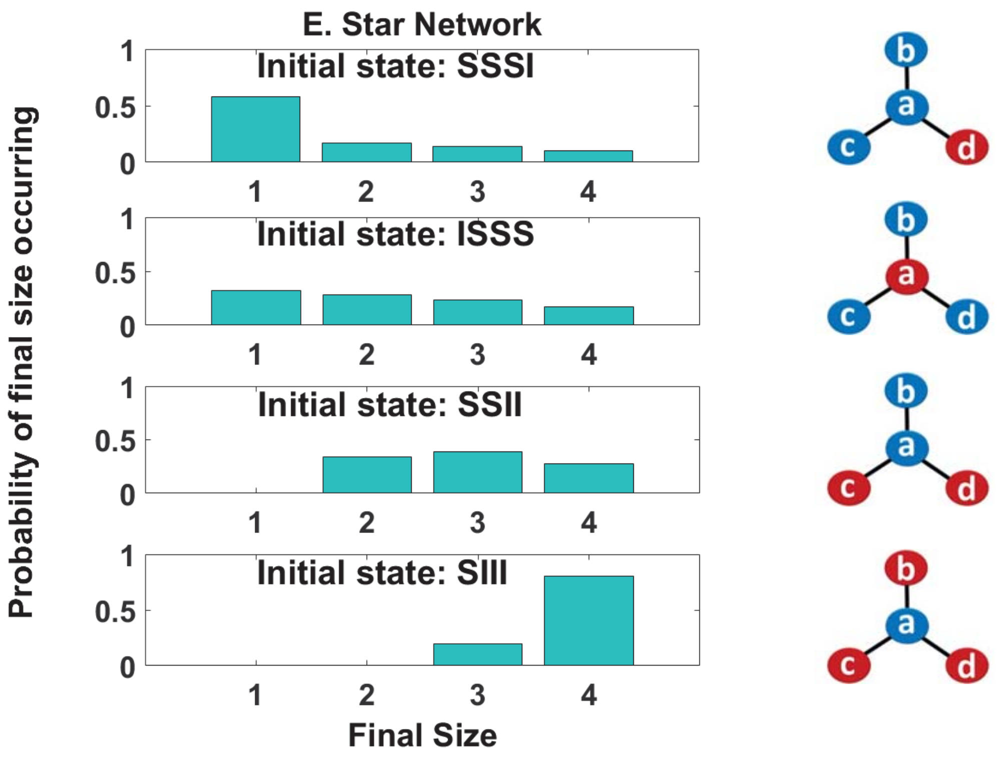


Figure 4.3: Probability mass functions for the spread of influenza through the star network with $N = 4$. Schematics of the initial state are shown on the right.

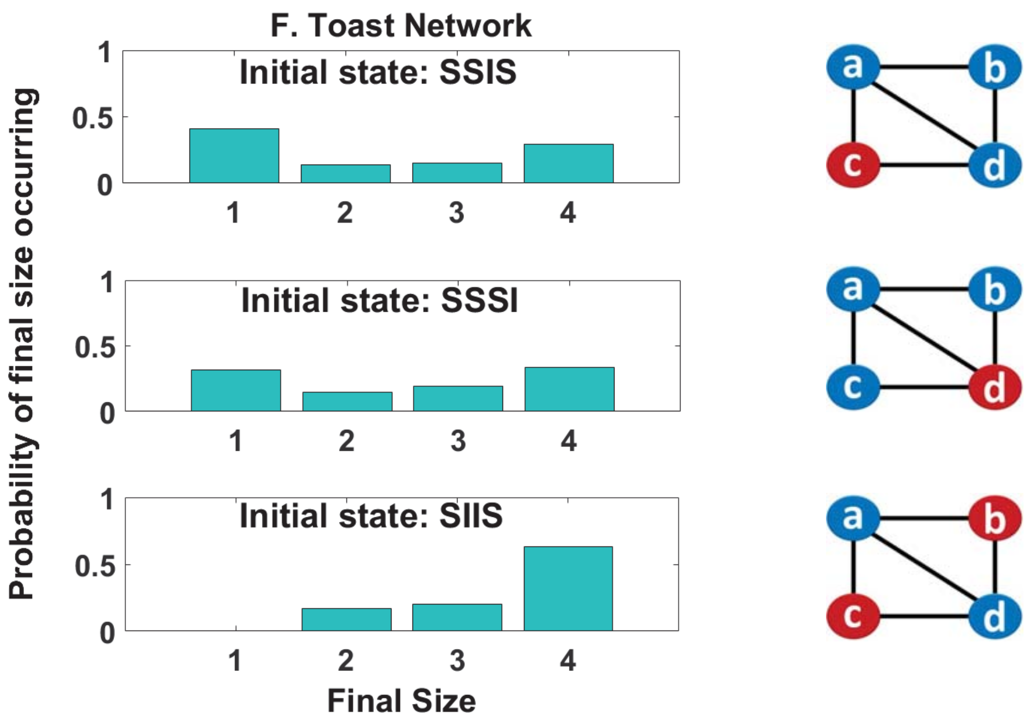


Figure 4.4: Probability mass functions for the spread of influenza through the toast network with $N = 4$. Schematics of the initial state are shown on the right.

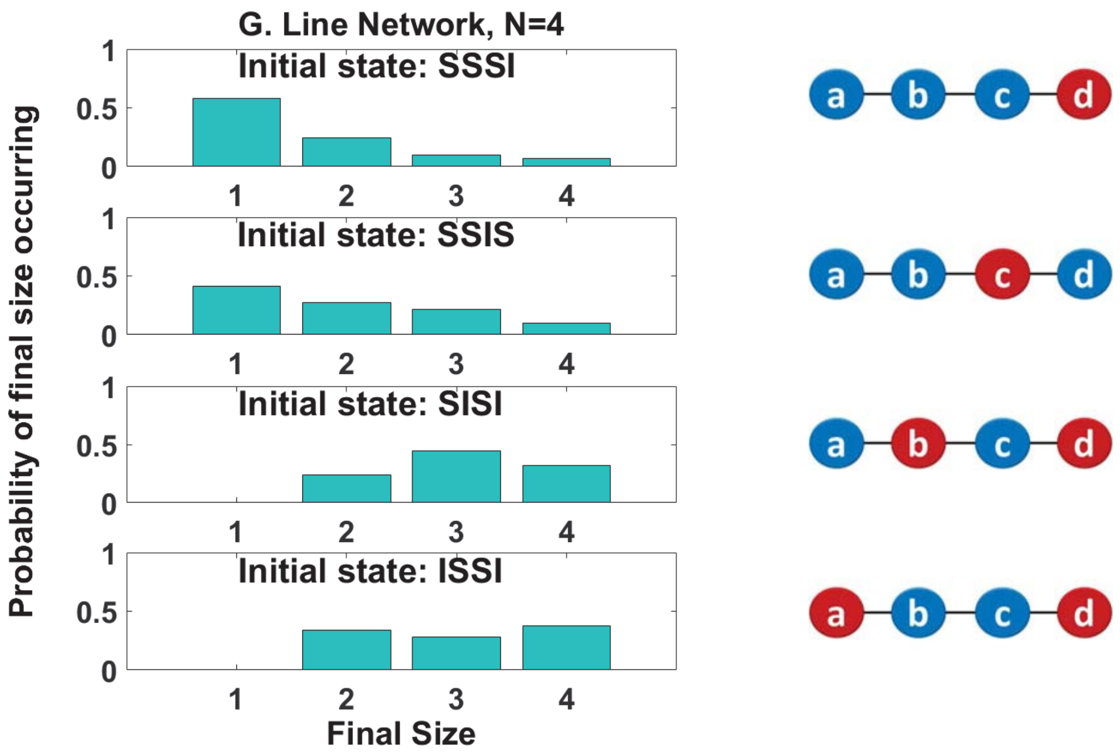


Figure 4.5: Probability mass functions for the spread of influenza through the line network with $N = 4$. Schematics of the initial state are shown on the right.

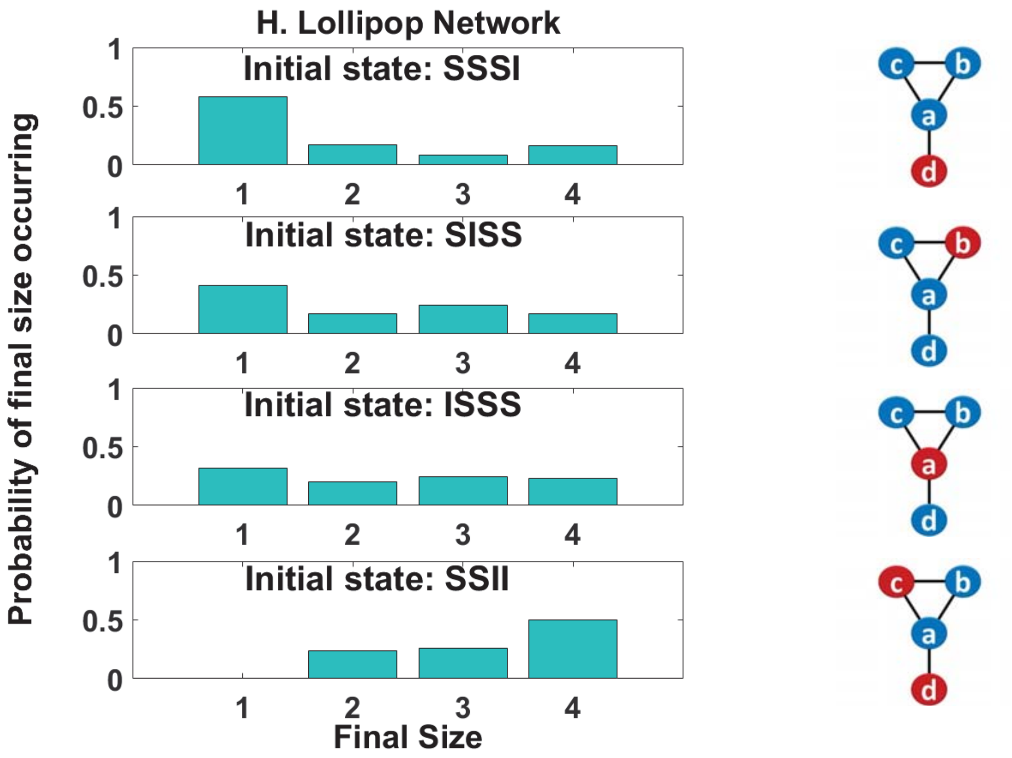


Figure 4.6: Probability mass functions for the spread of influenza through the lollipop network with $N = 4$. Schematics of the initial state are shown on the right.

4.3 Measles

In this section we present a case where an infectious disease has an R_0 value which exceeds the average degree of the small networks considered. This means that the use of the small network models, as illustrated above for influenza, is not possible. We end this section by briefly describing a scenario which would allow the use of small network models by making use of the effective reproduction number, R.

Measles is a highly contagious, yet preventable disease that can infect children and adults. Typical initial symptoms include a fever, cough, runny nose, sore and watery eyes, followed by the onset of a blotchy rash. The best prevention method is two doses of the measles, mumps and rubella (MMR) vaccine. Roberts and Tobias [59] provided evidence which suggests the optimal timing for the MMR vaccination is when children are aged 15 months and between 3 and 6 years. However, a high coverage of vaccination would be required to eliminate measles from a population completely. They estimated the basic reproduction number, R_0, to be approximately 12.8 for the 1997 measles epidemic in New Zealand. They also estimated γ (1/average infectious period) to be 52/year, that is the infectious period is approximately 0.019 of a year ($\approx$ 7.02 days). Therefore, in days, we have $\gamma = 1/7.02 = 0.1424$. Roberts and Tobias [59] used a population with similar size and age structure to that of New Zealand in 1996, they excluded individuals who were aged over 25 years. The population was composed of different age classes and the total population size used was approximately 1.5 million which was the sum of the number of individuals in each age class.

As discussed in the previous section, R_0 is a quantity that is derived for large populations and it does not make sense to use it for small populations. For example, in the 1997 measles epidemic described above, $R_0 = 12.8$ means that some individuals within the population had connections with more than 12 other individuals. If we were to represent this population as a network there would be nodes with more than 12 neighbours. Therefore, as our small network models consist of three and four individuals, it does not make sense to have an R_0 that is greater than the maximum degree of the network. For example, in a network of four nodes, any individual has at most three neighbours and thus can transmit infection to at most three nodes. Therefore, for the measles data we found, it is not possible to parameterize the small network models and illustrate how the infection could spread. However, we should note that, it is possible that our small networks of 4 nodes represent the unvaccinated individuals of a larger population. For example, if we had a population of $N = 40$ and 90% of the population were vaccinated with the MMR vaccine, then there would be $N = 4$ individuals left that could get infected and transmit measles to one another. In this case it would be possible that $R_0 = 12.8$, however the effective reproduction number, R, would be much smaller ($R = (1 - 0.9)R_0 \approx 1.2$).

4.4 Discussion

In this section we discuss the results found for the spread of influenza on small networks and comment on the difference between the basic reproduction number for influenza and measles. From Table 4.3 we can see that the networks with the two highest estimates for the basic reproduction number, R_0, for influenza are the complete and toast networks of size $N = 4$ respectively. For networks of size $N = 3$ and for a given probability of transmission given contact, the triangle network has an R_0 value greater than that for the line network. As R_0 depends on the average degree of the network we notice that the square and triangle networks have the same value for $R_0 = 0.83$. Overall, the line networks of size $N = 3$ and $N = 4$ had the lowest estimate of R_0. Table 4.4 shows the expected final size of an influenza epidemic for each small network and given initial condition. The triangle network has a higher expected final size than the line network of $N = 3$ nodes starting with one infectious node. For networks of size $N = 4$ the complete and toast networks had the highest expected final size of 2.68 and 2.55 for starting the epidemic with one infectious node. These results are in agreement with the results found for the probability mass functions of the final epidemic size in Chapter 2.

If we compare the infection parameters of influenza ($R_0 = 1.25$,$\gamma = 0.2577$) and measles ($R_0 = 12.8$, $\gamma = 0.1424$) we can see that the recovery rates of each infection are of the same magnitude. However, the basic reproduction number, R_0, for measles is ten times higher than the R_0 for influenza. Although we cannot parameterize our small network models for measles and compare the results to those found for influenza, we can see from the large difference in parameter values that if measles were introduced in a small network model it is highly likely that all individuals would become infected. Of course, there is still a non-zero probability that the infection does not take off and a minor epidemic occurs. However, if a major epidemic occurred, we would expect the final size of measles to be much higher than that of influenza due to the difference in orders of magnitude of R_0.

Part II

SIS Model

Chapter 5

SIS epidemics on small networks

In this chapter we focus on an SIS (Susceptible - Infectious - Susceptible) compartmental model to describe the spread of an infection through a population. When an infectious individual recovers they move back into the susceptible class. This makes multiple infections possible for each individual which adds some complexity to analysing the infection dynamics. Here, we consider eight small networks of varying topological structure and investigate the dynamics of the model on them. These small networks are the same eight as in Part I, repeated here for convenience.

First, we define the notation used throughout this chapter. Susceptible nodes acquire infection at rate β per $S-I$ edge and infectious nodes recover at rate γ. We denote the state that a network of three nodes is in by XYZ, where X, Y and Z denote the infection state (S or I) that nodes a, b and c are in respectively. Similarly, we denote the state that a network of four nodes is in by $WXYZ$. For an SIS model on a network of N nodes there are 2^N possible states in which the network can be. Thus, for an SIS model on networks of size $N = 3$ and $N = 4$ nodes there are 8 and 16 possible states which the networks can be in respectively. To simplify our analysis we group states together based on symmetries of the network [62]; in the most simple case of a complete network we can group states together that have the same number of S and I nodes (A complete network is a network in which each node is connected to every other node within the network; see, for example, Figures 5.1 A and C.). For networks that are not complete we must look at the topology and the number of nodes in each infection state before determining which network states can be grouped together. Thus, for each network we group together the states that are topologically equivalent and have the same number of nodes in each infection state. Once we have the reduced number of states we draw a transition diagram which

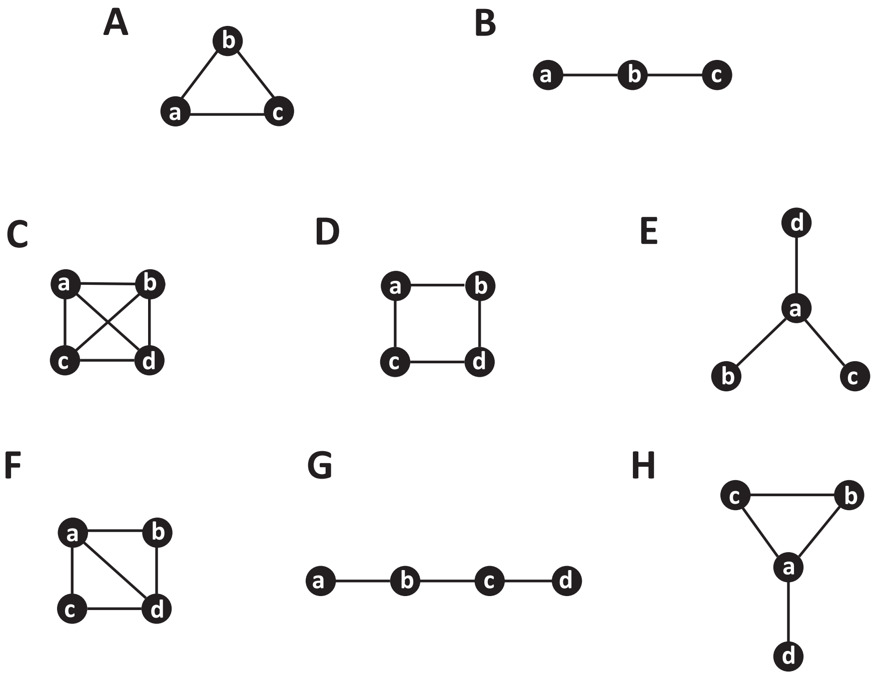


Figure 5.1: Network diagrams in order of increasing complexity. **A**. Triangle Network. **B**. Line Network, $N = 3$. **C**. Complete Network, $N = 4$. **D**. Square Network. **E**. Star Network. **F**. Toast Network. **G**. Line Network, $N = 4$. **H**. Lollipop Network.

shows how the infection moves through the network. From the transition diagrams we can show how the infection progresses through the network by writing down a set of differential equations; each equation describing the evolution of the probability that the network is in a given state at time t. We determine the Jacobian matrix of the system and the Routh-Hurwitz criterion is used for the triangle network to prove that the only stable steady state for the system is the infection free steady state.

Markov Chain theory is then used to describe the behaviour of SIS epidemics on small networks as shown in Figure 5.1. The results derived are largely based on a matrix which contains the probability of transition between states, which can be derived from the transition diagrams. In particular there are five quantities that we derive for the SIS model on small networks:

- Expected time to absorption (the expected time it takes for the system to reach the infection free steady state).
- Cumulative incidence (the number of new infections that occur during the epidemic).

- Expected number of times each individual is infected during the epidemic.

- Expected number of times the system is in each state (as shown in the transition diagrams).

- Expected time spent in each state for the duration of the epidemic.

All of these properties are governed by the initial conditions (initial state of the system) and by the transmission and recovery parameters, β and γ respectively.

To independently verify our results we also ran stochastic realisations on each of the small networks for the same initial conditions and infection parameters. The Gillepsie algorithm was used for the stochastic SIS model (See [34] for details of the algorithm) which was implemented using MATLAB software. In the following sections of this chapter we will illustrate our methods with the simplest and most complex small networks, the triangle and lollipop networks respectively. This is followed by the presentation and discussion of our results from all small networks shown in Figure 5.1. The detailed methods for each of the remaining small networks are given in Appendix C, page 242.

5.1 Triangle Network

The triangle network is the simplest network we consider; it is a complete network of size $N = 3$. We can reduce the full system of $2^N = 8$ equations which describe the SIS epidemic process on a triangle network by considering the network topology and adopting the method of lumping by graph automorphism approach, as detailed in Kiss et al [62]. As the triangle is a complete network we can lump states together which have the same number of S and I nodes; this reduces the system to one of four states. The transition state diagram is shown in Figure 5.2.

Each box in the transition diagram represents a different state of the network and each arrow represents either recovery or infection of a node. Transition between states is only possible if there is a directed arrow from one to the other. Movement between states in the transition diagram shows how the infection spreads through the network. Once the network reaches the absorbing state the epidemic is over and the infection has died out.

The lumped system of equations governing the SIS dynamics on a triangle network are given below. Each equation describes the probability that the triangle

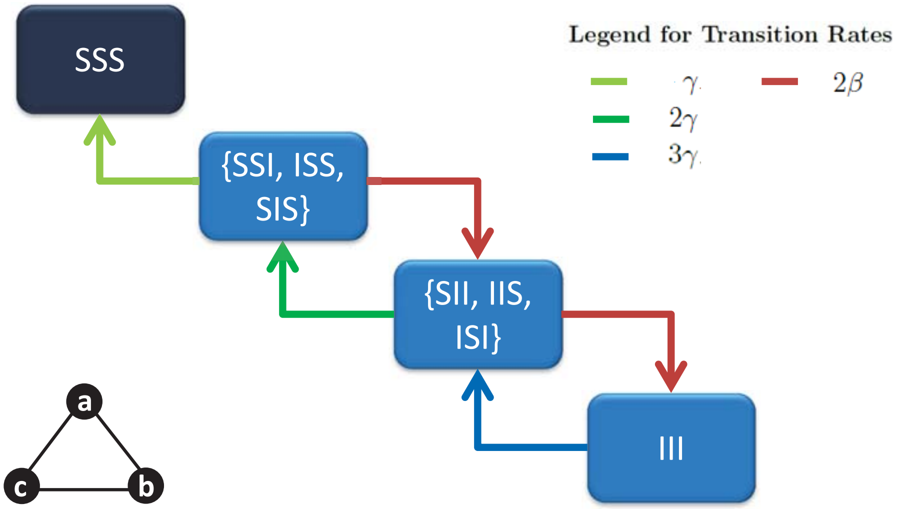


Figure 5.2: Transition diagram of an SIS model on a triangle network after lumping appropriate states together.

network is in a given state at time t.

$$\begin{aligned}
\dot{P}_{SSI} &= 2\gamma P_{SII} - (2\beta+\gamma)P_{SSI} \qquad (5.1)\\
\dot{P}_{SII} &= 2\beta P_{SSI} + 3\gamma P_{III} - 2(\beta+\gamma)P_{SII}\\
\dot{P}_{III} &= 2\beta P_{SII} - 3\gamma P_{III}\\
\dot{P}_{SSS} &= \gamma P_{SSI}
\end{aligned}$$

Here, for example, P_{SSI} denotes the probability that the system is in the state where one node (a, b or c) is infectious and the other two nodes are susceptible at time t. Therefore the equation for $\dot{P}_{SSI}$ describes how the system can enter and leave the state SSI.

5.1.1 The Routh-Hurwitz Criterion

In this section we investigate the stability of the system, described above, in order to prove that the infection free state (SSS) is the only stable steady state for the system. A well known method used to investigate the stability of linear systems is the Routh-Hurwitz criterion which utilizes the characteristic polynomial of the system to determine the stability of fixed points. Therefore in the following we first determine the Jacobian matrix of the system in order to obtain a characteristic polynomial before outlining the method for the Routh-Hurwitz criterion [19]. To investigate the stability of a fixed point of an n-dimensional system we want to find all eigenvalues, λ, which satisfy

$$\det(\mathbf{J} - \lambda \mathbf{I}) = 0$$

where $\mathbf{J}$ is the Jacobian matrix of the system evaluated at the fixed point which is of order n and $\mathbf{I}$ is the identity matrix of order n. From this we obtain the characteristic polynomial

$$\lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_n = 0$$

where the coefficients a_i, $i = 1...n$, are real constants. The fixed point is stable if all the roots of the characteristic polynomial have negative real part. The Routh-Hurwitz criterion provides a method of determining if the roots of the above polynomial have negative real part based on the coefficients of the polynomial [19]. For a three dimensional system, the characteristic equation is a cubic polynomial of the form

$$\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3$$

All roots of the polynomial have negative real part, if $a_1 > 0$, $a_3 > 0$ and $a_1 a_2 > a_3$.

The Jacobian matrix for the SIS model on a triangle network is defined by:

$$\mathbf{J} = \begin{bmatrix} -(2\beta+\gamma) & 2\gamma & 0 \\ 2\beta & -2(\beta+\gamma) & 3\gamma \\ 0 & 2\beta & -3\gamma \end{bmatrix}$$

where the columns and rows of $\mathbf{J}$ are indexed in the same order as the system of Equations 5.1, $\{SSI, SII, III\}$. Note that here we exclude the absorbing state SSS as the equation for $\dot{P}_{SSS}$ is redundant. Also note that as we have a linear system, the only steady state is the absorbing state with $P_{SSS} = 1$. The characteristic equation for the lumped system describing an SIS epidemic on a triangle network can be found by evaluating $\det(\mathbf{J} - \lambda\mathbf{I}) = 0$, where

$$\mathbf{J} - \lambda\mathbf{I} = \begin{bmatrix} -(2\beta+\gamma) - \lambda & 2\gamma & 0 \\ 2\beta & -2(\beta+\gamma) - \lambda & 3\gamma \\ 0 & 2\beta & -3\gamma - \lambda \end{bmatrix}$$

Therefore, finding the determinant for the above matrix gives the characteristic equation:

$$\lambda^3 + (4\beta + 6\gamma)\lambda^2 + (4\beta^2 + 8\beta\gamma + 11\gamma^2)\lambda + 6\gamma^3$$

In this case we have the following coefficients,

$$\begin{aligned} a_1 &= 4\beta + 6\gamma \\ a_2 &= 4\beta^2 + 8\beta\gamma + 11\gamma^2 \\ &= 4(\beta + \gamma)^2 + 7\gamma^2 \\ a_3 &= 6\gamma^3 \end{aligned}$$

Now, as β is the transmission rate and γ is the recovery rate they must both always be non-negative. Therefore, for the first two conditions we have $a_1 = 4\beta + 6\gamma > 0$ and $a_3 = 6\gamma^3 > 0$ which is always true. For the third condition we have:

$$\begin{aligned} a_1 &> 6\gamma > 0 \\ a_2 &> 6\gamma^2 > 0 \\ a_1 a_2 &> 6\gamma^3 \\ \therefore a_1 a_2 &> a_3 \end{aligned}$$

Hence this system satisfies the Routh-Hurwitz criterion which indicates that the steady state (the infection free state, SSS) of the system described by Equations 5.1 is stable.

5.1.2 Results from Markov Chain theory

As discussed in Chapter 1, we use a discrete time embedded markov chain to describe the SIS epidemic model on a network. For an SIS model we can show how the infection progresses through a network using the transition diagram (see Figure 5.2). The transition diagram shows the different states in which the network can be and the transition rates between states. These states can be represented as a particular type of Markov Chain called an absorbing Markov Chain . It is an absorbing Markov Chain because from any starting state the system will end up in the infection free state with all nodes susceptible. From this state the system cannot move anywhere else, thus the infection free state is called the absorbing state of this Markov Chain. All other states in this system are called transient states [21].

In discrete time Markov Chains at each step it is assumed that the system makes at most one transition (or one event occurs) [10]. This makes the analysis less complex, however it does not fully capture the continuous time dynamics as the n step transition matrix, $\mathbf{P}$, does not specify the rates at which transitions occur. Therefore, to approximate continuous time dynamics we assume that a time step, δt, is small such that $\delta t \to 0$ and only one event (infection or recovery of a node) can occur during each time step. That is, we assume $\delta t \to 0$ so that we are effectively averaging over an infinite number of networks with the same configuration to determine the average probability of transition between states. This also means that multiple events cannot happen simultaneously. These events could be thought of as infection generations as we are keeping track of how an infection progresses through each small network.

Here, we define some key theorems and definitions, using the SIS epidemic on a triangle network to illustrate. Following this we will describe the SIS epidemic dynamics on a lollipop network and present the results found for all small networks shown in Figure 5.1. For the triangle network we have the following transition probability matrix from which we can calculate the n^{th} event probabilities as follows. The set of states for the SIS process on a triangle network is $\mathbf{S} = \{SSI, SII, III, SSS\}$. Note that we have lumped states together which are topologically equivalent due to the symmetry of the network. The state SSI represents the state in which there is only one infected node (a, b or c) in the triangle network. For example, the probability the network is in this state is equal to the probability that the network is in any one of the states $\{SIS$,ISS and $SSI\}$. Similarly, the state SII is represents the state in which any two nodes are infected in the triangle network. The following matrices are indexed in the same order as $\mathbf{S}$.

$$\mathbf{P} \quad = \quad \begin{pmatrix} 0 & \dfrac{2\beta}{2\beta+\gamma} & 0 & \dfrac{\gamma}{2\beta+\gamma} \\ \dfrac{\gamma}{\beta+\gamma} & 0 & \dfrac{\beta}{\beta+\gamma} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Here, for example, P_{12} is the probability of transition from state SSI to SII, P_{14} is the probability of transition from state SSI to SSS, P_{23} is the probability of transition from state SII to III and so on. We can see that the sum across all rows of $\mathbf{P}$ equals one, as it should for a probability matrix.

Therefore, we can find the probability that the system is in a given state after n

events by evaluating the following matrix, $\mathbf{P^n}$.

$$\mathbf{P^n} = \begin{pmatrix} 0 & \frac{2\beta}{2\beta+\gamma} & 0 & \frac{\gamma}{2\beta+\gamma} \\ \frac{\gamma}{\beta+\gamma} & 0 & \frac{\beta}{\beta+\gamma} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^n$$

We can also find the probability that the system is in a given state after n events, given that the system started in state $\mathbf{v_0}$ by computing $\mathbf{v_0 P^n}$. Where $\mathbf{v_0}$ is the initial state vector of size $1 \times \mathbf{S}$ with $\mathbf{v_0}(i) = 1$ if the epidemic starts in state i and all other entries zero.

Expected number of visits to each state

Theorem 1. *Let* $\boldsymbol{I}$ *be the identity matrix of the same order as* $\boldsymbol{Q}$ *and let* $\boldsymbol{Q}$ *be the transition matrix for transient states only. Then, for an absorbing Markov Chain the matrix* $\boldsymbol{I}$ - $\boldsymbol{Q}$ *has an inverse* $\boldsymbol{N} = \boldsymbol{I} + \boldsymbol{Q} + \boldsymbol{Q}^2 + \ldots + \boldsymbol{Q}^k + \ldots$. *The* ij*-entry* $\mathbf{N}_{ij}$ *of the matrix* $\boldsymbol{N}$ *is the expected number of times the chain is in state* s_j*, given that it starts in state* s_i*. The initial state is counted if* $i = j$ *[21].*

Therefore, the matrix $\mathbf{N}$ is a geometric series of all the possible pathways to get from one state to another in a given network. It can be shown that the series converges to $(\mathbf{I}\text{-}\mathbf{Q})^{-1}$ which leads to the following formal definition for $\mathbf{N}$.

Definition 4. For an absorbing Markov Chain $\mathbf{P}$, the matrix $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$ is called the fundamental matrix for $\mathbf{P}$. The entry $\mathbf{N}_{ij}$ gives the expected number of times that the process is in the transient state s_j if it started in the transient state s_i [21].

The transition matrix for the transient states only of the SIS model on a triangle network is defined by $\mathbf{Q}$.

$$\mathbf{Q} = \begin{pmatrix} 0 & \frac{2\beta}{2\beta+\gamma} & 0 \\ \frac{\gamma}{\beta+\gamma} & 0 & \frac{\beta}{\beta+\gamma} \\ 0 & 1 & 0 \end{pmatrix}$$

From $\mathbf{Q}$ and using the above information, we can find the expected number of times the process visits each transient state from a given initial state. Note that $\mathbf{Q}$ is simply the matrix $\mathbf{P}$ with the row and column corresponding to the absorbing state deleted. For the remaining small networks we will present the matrix $\mathbf{Q}$ and omit the symbolic representation of $\mathbf{P}$.

$$
\begin{aligned}
\mathbf{N} &= (\mathbf{I}-\mathbf{Q})^{-1} \\
&= \begin{pmatrix} 1 & \dfrac{-2\beta}{2\beta+\gamma} & 0 \\ \dfrac{-\gamma}{\beta+\gamma} & 1 & \dfrac{-\beta}{\beta+\gamma} \\ 0 & -1 & 1 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \dfrac{2\beta+\gamma}{\gamma} & \dfrac{2\beta(\beta+\gamma)}{\gamma^2} & \dfrac{2\beta^2}{\gamma^2} \\ \dfrac{2\beta+\gamma}{\gamma} & \dfrac{(2\beta+\gamma)(\beta+\gamma)}{\gamma^2} & \dfrac{\beta(2\beta+\gamma)}{\gamma^2} \\ \dfrac{2\beta+\gamma}{\gamma} & \dfrac{(2\beta+\gamma)(\beta+\gamma)}{\gamma^2} & \dfrac{(2\beta^2+\beta\gamma+\gamma^2)}{\gamma^2} \end{pmatrix}
\end{aligned}
$$

Expected time spent in each state during the epidemic

One property of interest is the expected time spent in each of the transient states over the course of an epidemic. This can be determined as follows:

$$E[\text{time in given state}] = E[\#\text{visits to given state}] \times [\text{Duration of visits}]$$

where # refers to the *'number of'*. From the matrix $\mathbf{N}$ found above we have the expected number of visits to each state for a given initial state. The duration of visits is simply the average time spent in a state once the system has entered that particular state and this can be found from the transition diagram in Figure 5.2. The average time spent in each state of the triangle network is stored in a diagonal matrix, $\mathbf{T_S}$, where the entries are in the same order as $\mathbf{S} = \{SSI, SII, III, SSS\}$, excluding the state SSS. We rescaled time with $\gamma = 1$ as time has units γ^{-1} for the SIS model on small networks; this makes the resulting matrix entries dimensionless. We adopt this notation throughout this chapter for other results which also have dimension γ^{-1}. The entry $\mathbf{T_S}(1,1)$ gives the time spent in state SSI once the system has entered that state, this is simply

$$\frac{1}{(\text{Rate for leaving state } SSI)}$$

Similarly, $\mathbf{T_S}(2,2)$ and $\mathbf{T_S}(3,3)$ give the time spent in state SII and III once the system has entered each state respectively.

$$\mathbf{T_S} = \frac{1}{\gamma}\begin{pmatrix} \dfrac{\gamma}{2\beta+\gamma} & 0 & 0 \\ 0 & \dfrac{\gamma}{2\beta+2\gamma} & 0 \\ 0 & 0 & \dfrac{1}{3} \end{pmatrix}$$

Therefore, the expected time spent in each state is given by $\mathbf{E_T}$, where

$$\begin{aligned} \mathbf{E_T} &= \mathbf{NT_S} \\ &= \frac{1}{\gamma}\begin{pmatrix} 1 & \dfrac{\beta}{\gamma} & \dfrac{2\beta^2}{3\gamma^2} \\ \\ 1 & \dfrac{2\beta+\gamma}{2\gamma} & \dfrac{\beta(2\beta+\gamma)}{3\gamma^2} \\ \\ 1 & \dfrac{2\beta+\gamma}{2\gamma} & \dfrac{2\beta^2+\beta\gamma+\gamma^2}{3\gamma^2} \end{pmatrix} \end{aligned}$$

Here the element $\mathbf{E_T}(i,j)$ gives the expected time spent in state j given that state i was the initial starting state. For example, $\mathbf{E_T}(1,2) = \beta/\gamma^2$ is the expected time spent in state SII given that the epidemic started in state SSI. Recalling that our indexing for the matrices is in the order of $\mathbf{S} = \{SSI, SII, III\}$.

Time to absorption

The expected time to absorption is the expected time it takes for the system to reach the infection free steady state. For the triangle network this is the expected time it takes to reach the state SSS. Summing across the rows of $\mathbf{E_T}$ gives a vector containing the total expected time spent in the transient states combined for a given initial state. This is also the expected time to absorption. Therefore, in the following vector $\mathbf{T_A}$ the entries $\mathbf{T_A}(i)$ gives the expected time to absorption given that the epidemic started in state i.

$$\begin{aligned} \mathbf{T_A}(i) &= \sum_{\mathbf{j}} \mathbf{E_T}(i,j) \\ &= \frac{1}{\gamma}\begin{pmatrix} \dfrac{2\beta^2+3\beta\gamma+3\gamma^2}{3\gamma^2} \\ \\ \dfrac{4\beta^2+8\beta\gamma+9\gamma^2}{6\gamma^2} \\ \\ \dfrac{4\beta^2+8\beta\gamma+11\gamma^2}{6\gamma^2} \end{pmatrix} \end{aligned}$$

Table 5.1 (page 154) contains the time to absorption for each small network and initial condition considered, evaluated with $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

Number of times each node is infected

To determine the number of times node k is infected during the epidemic we first determine the expected time for which node k is infected during the epidemic. We do this by summing up the expected times spent in all states j where node k is infected. For this we need to use the matrix which contains the expected time spent in each state of the triangle network for the full system, not the lumped system which we have been using up until now. For example, in the lumped system, the expected time spent in state SSI for a given initial condition is giving the expected time for which there is only one infectious node during the epidemic. This does not give information about which node was infected. Similarly the expected time spent in SII gives the expected time for which there are two infectious nodes. By deriving the expected time spent in each state for the full system we are able to find out more detail such as the expected time for which each node (a, b and c) is infected for during the epidemic. The state space for the full system is $\mathbf{S_{Full}} = \{SSI, SIS, SII, ISS, ISI, IIS, III, SSS\}$. Therefore we have

$$\begin{aligned} \mathcal{T}(k|i) &= E[\text{time node } k \text{ is infectious} \mid \text{epidemic started in state } i] \qquad (5.2) \\ &= \sum_{j|k=\mathcal{I}} \mathbf{E_{T\_Full}}(i,j) \end{aligned}$$

where $i, j \in \mathbf{S_{Full}}$, $k \in \{a, b, c\}$ and $k = \mathcal{I}$ means node k is infected. We give the algebraic expression of the matrix $\mathbf{E_{T\_Full}}$ on page 144. We exclude the infection free state SSS in $\mathbf{E_{T\_Full}}$. In $\mathcal{T}(k|i)$ (below), row 1 corresponds to the initial state SSI where node c is infected; row 2 corresponds to the state SII where nodes b and c are infected and row 3 corresponds to the state III where all three nodes are infected. The columns correspond to nodes a, b and c respectively. Therefore, the entry

$$\mathcal{T}(1|1) = \mathbf{E_{T\_Full}}(1,4) + \mathbf{E_{T\_Full}}(1,5) + \mathbf{E_{T\_Full}}(1,6) + \mathbf{E_{T\_Full}}(1,7)$$

gives the expected time node a is infected given that the epidemic started in state SSI. Similarly

$$\mathcal{T}(2|1) = \mathbf{E_{T\_Full}}(1,2) + \mathbf{E_{T\_Full}}(1,3) + \mathbf{E_{T\_Full}}(1,6) + \mathbf{E_{T\_Full}}(1,7)$$

gives the expected time node b is infected given that the epidemic started in state SSI. The entry

$$\mathcal{T}(3|1) = \mathbf{E_{T\_Full}}(1,1) + \mathbf{E_{T\_Full}}(1,3) + \mathbf{E_{T\_Full}}(1,5) + \mathbf{E_{T\_Full}}(1,7)$$

gives the expected time node c is infected given that the epidemic started in state SSI.
$\mathcal{T}(k|i) =$

$$\begin{pmatrix} \frac{2\beta\,(4\beta^3+9\beta^2\gamma+9\beta\gamma^2+3\gamma^3)}{3\gamma^3\,(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{2\beta\,(4\beta^3+9\beta^2\gamma+9\beta\gamma^2+3\gamma^3)}{3\gamma^3\,(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{8\beta^4+18\beta^3\gamma+18\beta^2\gamma^2+15\beta\gamma^3+6\gamma^4}{3\gamma^3\,(4\beta^2+5\beta\gamma+2\gamma^2)} \\ \frac{\beta\,(8\beta^3+22\beta^2\gamma+27\beta\gamma^2+12\gamma^3)}{3\gamma^3\,(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{8\beta^4+22\beta^3\gamma+27\beta^2\gamma^2+18\beta\gamma^3+6\gamma^4}{3\gamma^3\,(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{8\beta^4+22\beta^3\gamma+27\beta^2\gamma^2+18\beta\gamma^3+6\gamma^4}{3\gamma^3\,(4\beta^2+5\beta\gamma+2\gamma^2)} \\ \frac{2\beta^2+3\beta\gamma+3\gamma^2}{3\gamma^3} & \frac{2\beta^2+3\beta\gamma+3\gamma^2}{3\gamma^3} & \frac{2\beta^2+3\beta\gamma+3\gamma^2}{3\gamma^3} \end{pmatrix}$$

For the SIS model we assume that the infectious period for each node in the network is constant and equal to $\frac{1}{\gamma}$. To determine the length of time each node is infected for the duration of the epidemic we simply mulitply the number of times each node is infected by the average infectious period. Therefore, we have:

$$\begin{aligned} \mathcal{T}(k|i) &= (\text{Number of times node } k \text{ is infected}) \times (\text{infectious period}) \qquad (5.3) \\ &= (\text{Number of times node } k \text{ is infected}) \times \frac{1}{\gamma} \end{aligned}$$

Finally, we obtain:

$$\text{Number of times node } k \text{ is infected} = \gamma\mathcal{T}(k|i)$$

In the triangle network, for example, to find the expected number of times node a is infectious given that the epidemic started in state SSI (where node c is infected) we add up the expected time spent in all states where node a is infected (ISS, ISI, IIS and III). To keep our results consistent we find the expected time each node is infected given that the epidemic started in states SSI, SII and III.

$$\mathbf{E_{T\_Full}} = \frac{1}{\gamma}\begin{pmatrix}
\frac{(4\beta^2+9\beta\gamma+6\gamma^2)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta+3\gamma)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta^2+5\beta\gamma+3\gamma^2)}{3\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta+3\gamma)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta^2+5\beta\gamma+3\gamma^2)}{3\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta^2(4\beta+5\gamma)}{3\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{2\beta^2}{3\gamma^2} \\
\frac{\beta(4\beta+3\gamma)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{(4\beta^2+9\beta\gamma+6\gamma^2)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta^2+5\beta\gamma+3\gamma^2)}{3\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta+3\gamma)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta^2(4\beta+5\gamma)}{3\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta^2+5\beta\gamma+3\gamma^2)}{3\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{2\beta^2}{3\gamma^2} \\
\frac{(4\beta^2+5\beta\gamma+3\gamma^2)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{(4\beta^2+5\beta\gamma+3\gamma^2)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{(2\beta+\gamma)(4\beta^2+5\beta\gamma+6\gamma^2)}{6\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta+5\gamma)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(2\beta+\gamma)(4\beta+5\gamma)}{6\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(2\beta+\gamma)(4\beta+5\gamma)}{6\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(2\beta+\gamma)}{3\gamma^2} \\
\frac{\beta(4\beta+3\gamma)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta+3\gamma)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta^2(4\beta+5\gamma)}{3\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{(4\beta^2+9\beta\gamma+6\gamma^2)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta^2+5\beta\gamma+3\gamma^2)}{3\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta^2+5\beta\gamma+3\gamma^2)}{3\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{2\beta^2}{3\gamma^2} \\
\frac{(4\beta^2+5\beta\gamma+3\gamma^2)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(4\beta+5\gamma)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(2\beta+\gamma)(4\beta+5\gamma)}{6\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{(4\beta^2+5\beta\gamma+3\gamma^2)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{(2\beta+\gamma)(4\beta^2+5\beta\gamma+6\gamma^2)}{6\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(2\beta+\gamma)(4\beta+5\gamma)}{6\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(2\beta+\gamma)}{3\gamma^2} \\
\frac{\beta(4\beta+5\gamma)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{(4\beta^2+5\beta\gamma+3\gamma^2)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(2\beta+\gamma)(4\beta+5\gamma)}{6\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{(4\beta^2+5\beta\gamma+3\gamma^2)}{3(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(2\beta+\gamma)(4\beta+5\gamma)}{6\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{(2\beta+\gamma)(4\beta^2+5\beta\gamma+6\gamma^2)}{6\gamma(4\beta^2+5\beta\gamma+2\gamma^2)} & \frac{\beta(2\beta+\gamma)}{3\gamma^2} \\
\frac{1}{3} & \frac{1}{3} & \frac{(2\beta+\gamma)}{6\gamma} & \frac{1}{3} & \frac{(2\beta+\gamma)}{6\gamma} & \frac{(2\beta+\gamma)}{6\gamma} & \frac{(2\beta^2+\beta\gamma+\gamma^2)}{3\gamma^2}
\end{pmatrix}$$

Cumulative Incidence

The cumulative incidence is the total number of new infections that occur during an epidemic. For the SIS model on a triangle network we find the cumulative incidence by summing up the vector containing the number of times each node is infected for a given initial state. Therefore, the cumulative incidence for an epidemic that started in state i can be found by evaluating the following:

$$\text{Cumulative Incidence } = \gamma \sum_{k=1}^{N} \mathcal{T}(k|i) \tag{5.4}$$

The expression for the cumulative incidence of an SIS epidemic on a triangle network starting in state SSI with one infectious node is given by

$$\gamma \sum_{k=1}^{N=3} \mathcal{T}(1,k) = \frac{\left(2\,\beta^2 + 2\,\beta\,\gamma + \gamma^2\right)}{\gamma^2}$$

Table 5.1 (page 154) gives the Cumulative Incidence for all of the small networks and different initial states considered for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

5.2 Lollipop Network

For a lollipop network we were able to reduce the system from 16 different states down to 12 via lumping due to the network symmetry. The reduced system consists of the state space, $\mathbf{S} = \{SSSI, SISS, ISSS, IISS, ISSI, SIIS, SSII, SIII, ISII, IIIS, IIII, SSSS\}$. Throughout this section the matrices are indexed in the same order as the state space $\mathbf{S}$. The transition state diagram is shown in Figure 5.3.

Each box in the transition diagram represents a different state of the network and each arrow represents either recovery or infection of a node. Coloured outlines represent topologically different states with the same number of nodes in each infection state. Transition between states is only possible if there is a directed arrow from one to the other. Movement between states in the transition diagram shows how the infection spreads through the network. Once the network reaches the absorbing state the epidemic is over and the infection has died out. From the transition diagram we derive the individual transition probabilities between network states. The

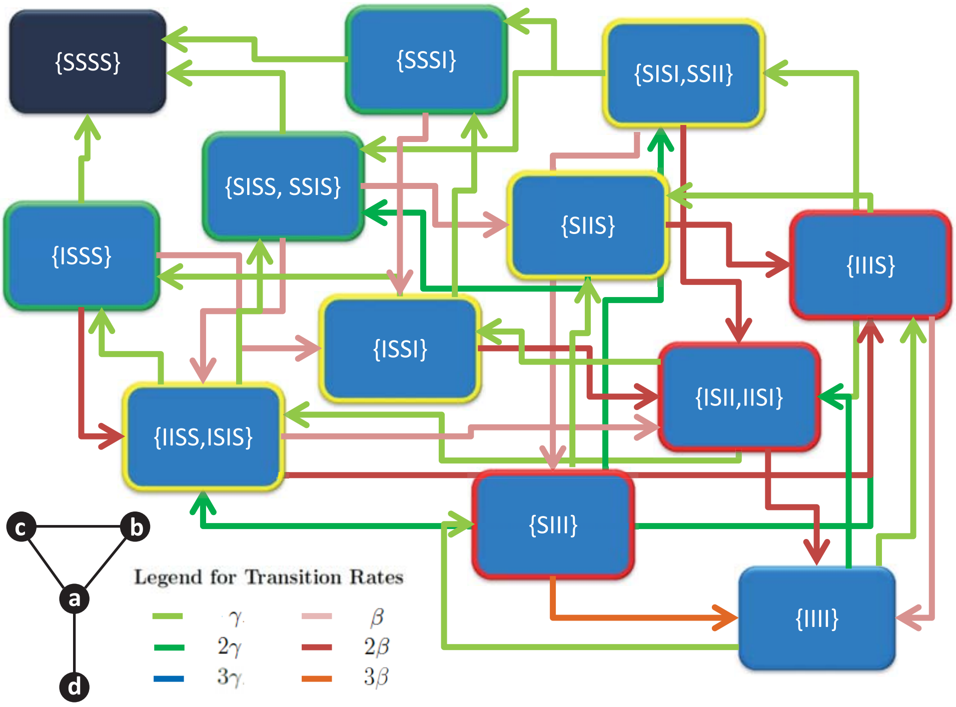


Figure 5.3: Transition diagram shows how the infection can spread through the lollipop network.

equations describing an SIS epidemic on a lollipop network with 4 nodes are:

$$\begin{aligned}
\dot{P}_{SSSI} &= \gamma(P_{ISSI} + P_{SSII}) - (\beta+\gamma)P_{SSSI} \qquad (5.5)\\
\dot{P}_{SISS} &= 2\gamma P_{SIIS} + \gamma(P_{IISS} + P_{SSII}) - (2\beta+\gamma)P_{SISS}\\
\dot{P}_{ISSS} &= \gamma(P_{ISSI} + P_{IISS}) - (3\beta+\gamma)P_{ISSS}\\
\dot{P}_{IISS} &= 2\beta P_{ISSS} + \beta P_{SISS} + 2\gamma P_{IIIS} + \gamma P_{ISII} - (3\beta+2\gamma)P_{IISS}\\
\dot{P}_{ISSI} &= \beta(P_{ISSS} + P_{SSSI}) + \gamma P_{ISII} - 2(\beta+\gamma)P_{ISSI}\\
\dot{P}_{SIIS} &= \gamma P_{IIIS} + \beta P_{SISS} + \gamma P_{SIII} - 2(\beta+\gamma)P_{SIIS}\\
\dot{P}_{SSII} &= 2\gamma P_{SIII} + \gamma P_{ISII} - (3\beta+2\gamma)P_{SSII}\\
\dot{P}_{SIII} &= \beta P_{SSII} + \gamma P_{IIII} - (3\beta+3\gamma)P_{SIII}\\
\dot{P}_{ISII} &= 2\gamma P_{IIII} + 2\beta P_{ISSI} + 2\beta P_{SSII} + \beta P_{IISS} - (2\beta+3\gamma)P_{ISII}\\
\dot{P}_{IIIS} &= 2\beta P_{IISS} + 2\beta P_{SIIS} + \gamma P_{IIII} - (\beta+3\gamma)P_{IIIS}\\
\dot{P}_{IIII} &= 3\beta P_{SIII} + 2\beta P_{ISII} + \beta P_{IIIS} - 4\gamma P_{IIII}\\
\dot{P}_{SSSS} &= \gamma(P_{SSSI} + P_{SSIS} + P_{ISSS})
\end{aligned}$$

The following matrix **Q**, which is used in the calculation of most of the results presented here, contains the transition probabilities for the *transient states only* of the lollipop network.

$\mathbf{Q} =$

$$\begin{bmatrix}
0 & 0 & 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{\beta}{2\beta+\gamma} & 0 & \frac{\beta}{2\beta+\gamma} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{2\beta}{3\beta+\gamma} & \frac{\beta}{3\beta+\gamma} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{\gamma}{3\beta+2\gamma} & \frac{\gamma}{3\beta+2\gamma} & 0 & 0 & 0 & 0 & 0 & \frac{\beta}{3\beta+2\gamma} & \frac{2\beta}{3\beta+2\gamma} & 0 \\
\frac{\gamma}{2\beta+2\gamma} & 0 & \frac{\gamma}{2\beta+2\gamma} & 0 & 0 & 0 & 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 & 0 \\
0 & \frac{\gamma}{\beta+\gamma} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 \\
\frac{\gamma}{3\beta+2\gamma} & \frac{\gamma}{3\beta+2\gamma} & 0 & 0 & 0 & 0 & 0 & \frac{\beta}{3\beta+2\gamma} & \frac{2\beta}{3\beta+2\gamma} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{\gamma}{3\beta+3\gamma} & \frac{2\gamma}{3\beta+3\gamma} & 0 & 0 & 0 & \frac{\beta}{\beta+\gamma} \\
0 & 0 & 0 & \frac{\gamma}{2\beta+3\gamma} & \frac{\gamma}{2\beta+3\gamma} & 0 & \frac{\gamma}{2\beta+3\gamma} & 0 & 0 & 0 & \frac{2\beta}{2\beta+3\gamma} \\
0 & 0 & 0 & \frac{2\gamma}{\beta+3\gamma} & 0 & \frac{\gamma}{\beta+3\gamma} & 0 & 0 & 0 & 0 & \frac{\beta}{\beta+3\gamma} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0
\end{bmatrix}$$

Expected number of visits to each state

Here, we omit the algebraic matrix expressions for $\mathbf{N}$ due to the size of each of the entries but note that it has been found by evaluating $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$. The determinant of $(\mathbf{I} - \mathbf{Q})$ is

$$\frac{\gamma^3\left(162\,\beta^7 + 806\,\beta^6\,\gamma + 1857\,\beta^5\,\gamma^2 + 2618\,\beta^4\,\gamma^3 + 2412\,\beta^3\,\gamma^4 + 1416\,\beta^2\,\gamma^5 + 480\,\beta\,\gamma^6 + 72\,\gamma^7\right)}{2\,(2\,\beta+\gamma)\,(\beta+3\,\gamma)\,(3\,\beta+\gamma)\,(2\,\beta+3\,\gamma)\,(3\,\beta+2\,\gamma)^2(\beta+\gamma)^4}$$

which gives an indication to the complexity of the matrix. In this case complexity refers to the order of the rational functions which are found in the non zero entries of the matrix, $\mathbf{N}$. These non-zero entries of the matrix are ratios of 7^{th} and 8^{th} order polynomials in β and γ. These polynomials have no common factor so the expression cannot be easily simplified. However, it is not necessary to have the explicit form of $\mathbf{N}$ as we can generate it from the transition matrix $\mathbf{Q}$ using the above formula. Therefore, provided we know how to find $\mathbf{Q}$ we can find $\mathbf{N}$.

Expected time spent in each state during the epidemic

To find the expected time spent in each state for the duration of the epidemic in the lollipop network we first find the time spent in each state once the system has entered that particular state. This can be found from the transition diagram in Figure 5.3 on page 146. Therefore, the time spent in each state of the lollipop network is stored in a diagonal matrix, $\mathbf{T_S}$, where the entries are in the same order as $\mathbf{S} = \{SSSI, SISS, ISSS, IISS, ISSI, SIIS, SSII, SIII, ISII, IIIS, IIII, SSSS\}$, excluding the state $SSSS$. In the following we use the notation $\mathsf{diag}(x, y, z)$ to denote a diagonal matrix with entries x,y and z.

$$\begin{aligned}\mathbf{T_S} &= \frac{1}{\gamma} \times \mathsf{diag}\left(\frac{\gamma}{\beta+\gamma}, \frac{\gamma}{2\beta+\gamma}, \frac{\gamma}{3\beta+\gamma}, \frac{\gamma}{3\beta+2\gamma}, \frac{\gamma}{2\beta+2\gamma}, \frac{\gamma}{2\beta+2\gamma}, \frac{\gamma}{3\beta+2\gamma},\right.\\ &\qquad\left.\frac{\gamma}{3\beta+3\gamma}, \frac{\gamma}{2\beta+3\gamma}, \frac{\gamma}{\beta+3\gamma}, \frac{1}{4}\right)\end{aligned}$$

To find the expected time spent in each state throughout the epidemic we compute the following.

$$\mathbf{E_T} = \mathbf{NT_S}$$

Again, we omit the algebraic expressions due to the size of the entries of the matrix $\mathbf{E_T}$ but note that it is similar in size to $\mathbf{N}$ as it is simply each entry in $\mathbf{N}$ multiplied by the corresponding value in $\mathbf{T_S}$.

Time to absorption

We determine the time to absorption for the lollipop network using the same method as outlined in Section 5.1.2 and present the results evaluated for different parameters in Table 5.1.

Number of times each node is infected

We determine the number of times each node is infected for the lollipop network using the same method as illustrated in Section 5.1.2.

We omit the algebraic expression for $\mathcal{T}(k|i)$ due to the size of the entries but note that it has been found. Each entry is a fraction where the numerator is a 10^{th} order polynomial in β and γ and the denominator is an 11^{th} order polynomial in β and γ. These polynomials have large coefficients (some in the tens of thousands) and cannot be factorized. Figure 5.11 on page 160 shows bar charts of the expected number of times each node is infected for the lollipop network.

Cumulative Incidence

The cumulative incidence for an epidemic which started in state i can be found by evaluating

$$\text{Cumulative Incidence } = \gamma \sum_{k=1}^{N} \mathcal{T}(k|i) \tag{5.6}$$

where $i \in \{SSSI, SISS, ISSS, SSII\}$ and $\sum_{k=1}^{N} \mathcal{T}(k|i)$ is shown on page 149. Table 5.1 gives the cumulative incidence for all of the small networks and different initial states considered for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

$$\sum_{k=1}^{N} \mathcal{T}(k|i) =$$

$$\frac{1}{\gamma} \begin{pmatrix} \frac{\left(648\,\beta^{10} + 4086\,\beta^{9}\,\gamma + 12299\,\beta^{8}\,\gamma^{2} + 23781\,\beta^{7}\,\gamma^{3} + 32597\,\beta^{6}\,\gamma^{4} + 32943\,\beta^{5}\,\gamma^{5} + 25170\,\beta^{4}\,\gamma^{6} + 14580\,\beta^{3}\,\gamma^{7} + 6120\,\beta^{2}\,\gamma^{8} + 1656\,\beta\,\gamma^{9} + 216\,\gamma^{10}\right)}{3\gamma^{3}\,(162\,\beta^{7} + 806\,\beta^{6}\,\gamma + 1857\,\beta^{5}\,\gamma^{2} + 2618\,\beta^{4}\,\gamma^{3} + 2412\,\beta^{3}\,\gamma^{4} + 1416\,\beta^{2}\,\gamma^{5} + 480\,\beta\,\gamma^{6} + 72\,\gamma^{7})} \\ \\ \frac{\beta\left(3564\,\beta^{9} + 20124\,\beta^{8}\,\gamma + 54005\,\beta^{7}\,\gamma^{2} + 92316\,\beta^{6}\,\gamma^{3} + 110785\,\beta^{5}\,\gamma^{4} + 97404\,\beta^{4}\,\gamma^{5} + 64769\,\beta^{3}\,\gamma^{6} + 32325\,\beta^{2}\,\gamma^{7} + 10782\,\beta\,\gamma^{8} + 1728\,\gamma^{9}\right)}{6\gamma^{3}\,(162\,\beta^{7} + 806\,\beta^{6}\,\gamma + 1857\,\beta^{5}\,\gamma^{2} + 2618\,\beta^{4}\,\gamma^{3} + 2412\,\beta^{3}\,\gamma^{4} + 1416\,\beta^{2}\,\gamma^{5} + 480\,\beta\,\gamma^{6} + 72\,\gamma^{7})} \\ \\ \frac{\beta\left(3564\,\beta^{9} + 20340\,\beta^{8}\,\gamma + 55001\,\beta^{7}\,\gamma^{2} + 94632\,\beta^{6}\,\gamma^{3} + 114349\,\beta^{5}\,\gamma^{4} + 100908\,\beta^{4}\,\gamma^{5} + 66677\,\beta^{3}\,\gamma^{6} + 32757\,\beta^{2}\,\gamma^{7} + 10782\,\beta\,\gamma^{8} + 1728\,\gamma^{9}\right)}{6\gamma^{3}\,(162\,\beta^{7} + 806\,\beta^{6}\,\gamma + 1857\,\beta^{5}\,\gamma^{2} + 2618\,\beta^{4}\,\gamma^{3} + 2412\,\beta^{3}\,\gamma^{4} + 1416\,\beta^{2}\,\gamma^{5} + 480\,\beta\,\gamma^{6} + 72\,\gamma^{7})} \\ \\ \frac{\beta\left(3240\,\beta^{9} + 19080\,\beta^{8}\,\gamma + 53002\,\beta^{7}\,\gamma^{2} + 93548\,\beta^{6}\,\gamma^{3} + 115523\,\beta^{5}\,\gamma^{4} + 102701\,\beta^{4}\,\gamma^{5} + 66220\,\beta^{3}\,\gamma^{6} + 30405\,\beta^{2}\,\gamma^{7} + 9054\,\beta\,\gamma^{8} + 1296\,\gamma^{9}\right)}{6\gamma^{3}\,(162\,\beta^{7} + 806\,\beta^{6}\,\gamma + 1857\,\beta^{5}\,\gamma^{2} + 2618\,\beta^{4}\,\gamma^{3} + 2412\,\beta^{3}\,\gamma^{4} + 1416\,\beta^{2}\,\gamma^{5} + 480\,\beta\,\gamma^{6} + 72\,\gamma^{7})} \end{pmatrix}$$

5.3 SIS epidemics on networks of three or four nodes

In this section we discuss how we computed the key results outlined in the previous sections which were derived using Markov Chain theory for all small networks shown in Figure 5.1. Following this we detail the stochastic SIS model which was used to independently verify our analytic results. We then proceed to present and discuss results found for the SIS model on all small networks shown in Figure 5.1.

5.3.1 MATLAB Implementation

We wrote a program in MATLAB to generate the five results for the SIS model as outlined in the previous sections. All results can be found analytically and displayed in symbolic form, however, as the network size increases and the network structure becomes less symmetric the expressions for the results become very large. Thus, the symbolic results become increasingly difficult to display. For this reason we wrote a MATLAB program to evaluate the results found using Markov Chain theory for different parameter values.

One of the inputs into the programs is the adjacency matrix, $\mathbf{A}$, of each network, where $\mathbf{A}_{ij} = 1$ if nodes i and j have an edge between them and $\mathbf{A}_{ij} = 0$ otherwise. The adjacency matrix is used to determine the neighbours of each node in the network. The degree of node i can also be found by computing $\sum_{j=1}^{N} \mathbf{A}_{ij}$. In an SIS model we have two infection states (S and I) in which each node can be, therefore we can construct a matrix, $\mathbf{V}$, which contains the binary representation of each network state. The columns of $\mathbf{V}$ represent different states in which the network can be. The rows corresponds to the infection state of each node in the network. Thus, $\mathbf{V}_{ij}$ gives the infection status of node i in state j.

To determine the transition probability from state i to state j of the network we look at the columns of $\mathbf{V}$. If the sum of the difference between two columns (states) is equal to -1 that means that an infection event can occur. If the sum of the difference between two columns (states) is equal to 1 that means that a recovery event can occur. For example, if $\sum(\mathbf{V}(:,i) - \mathbf{V}(:,j)) = -1$ then a susceptible node in state i is infectious in state j. Similarly, if $\sum(\mathbf{V}(:,i) - \mathbf{V}(:,j)) = 1$ then an infectious node in state i is recovered in state j.

Using this information and knowledge of each node's neighbours we are able to construct the transition matrix, $\mathbf{P}$, where the entries $\mathbf{P}_{ij}$ contains the probability of entering state j given that the system is in state i. From $\mathbf{P}$ we are able to find the transition matrix for the transient states only, $\mathbf{Q}$. Also, by using the information about the neighbours of each node we can construct the diagonal matrix, $\mathbf{T_S}$, for each network which stores the average duration of each visit to all states. We are then able to construct the matrices and vectors which define the five key results as discussed in the previous sections.

5.3.2 Stochastic Model

In previous sections the methods used to find and evaluate analytic results for an SIS model on small networks were detailed. In this section we will explain the methods used to check our results independently and show that they are correct. The Gillespie algorithm was used to simulate a stochastic SIS model on each of the small networks considered. We ran realisations of the stochastic model with the same initial conditions and infection parameters used to compute the analytic results. In the Gillespie algorithm, for each node, a random event time is generated by adding to the current time an exponentially distributed time step with rate either $i\beta$ (if node is in state S and where i is the number of infectious neighbours the node has) or γ (if node is in state I). Then, repeatedly, the node with the smallest time step is chosen, the current time step is updated to the event time of the chosen node, and the state of the chosen node is changed accordingly. The event time of the chosen node and all of its neighbours are then updated. This repetition continues until there are no more infectious nodes in the contact network.

The expected time to absorption is calculated by summing up each event time until the end of the epidemic. We keep a vector containing which nodes are currently infected throughout the epidemic, using this information after each event occurs we update a vector that contains the number of times each node is infected for the duration of the epidemic. The cumulative incidence for each realisation is the total number of new infections, which is obtained by summing up the entries in the vector containing the number of times each node is infected. Before starting the stochastic realisations, a matrix $\mathbf{V}$ is generated that contains the binary representation of all the possible states in which the network can be for an SIS model. Here, susceptible nodes correspond to entries that are zero and infectious nodes correspond to entries that are one. A vector containing the current infectious nodes is updated after each event occurs and thus after each event occurs we can determine the current state of the network. Using this information we record a count of the number of times the network is in each state, and also the total time spent in each state throughout the epidemic. To find the total time spent in each state we sum up the time to the next event once the network has entered a given state.

Comparison of Analytic and Stochastic Results

To compare the analytic and stochastic results we computed the following. For each initial condition and infection parameter combination we computed the average of each stochastic result over k realisations, we then incremented the number of realisations by 100, calculating the new average result each time until we had found the 100 averaged results for $k = \{100, 200, 300, \ldots, 10000\}$ realisations. We repeated this 10 times, so for each value of k we have 10 points. For each k we computed the ratio between each of the 10 stochastic values and the corresponding analytic value (S/A), we then calculated the variance of the 10 ratios. This gave us a vector of 100 variances, one corresponding to each k value. We then plot the variances on a log-log plot in order to show that as the number of realisations k, increases, $\log(S/A) \to 0$. In other words we show that as $k \to \infty$, $(S/A) \to 1$ and thus $\log(S/A) \to 0$. This shows that our analytic results are in agreement with the stochastic results which were found independently. We do this for each of the five key results discussed in this chapter. Figure 5.4 shows the convergence of the variance of S/A for the time to absorption and number of times each node was infected for the SIS model on a triangle network given that node a was the initial infectious node. Figure 5.5 shows the results for the expected number of visits to each state and the expected time spent in each state. More figures showing the variance of (S/A) as the number of realisations increases are given in Appendix D, page 273.

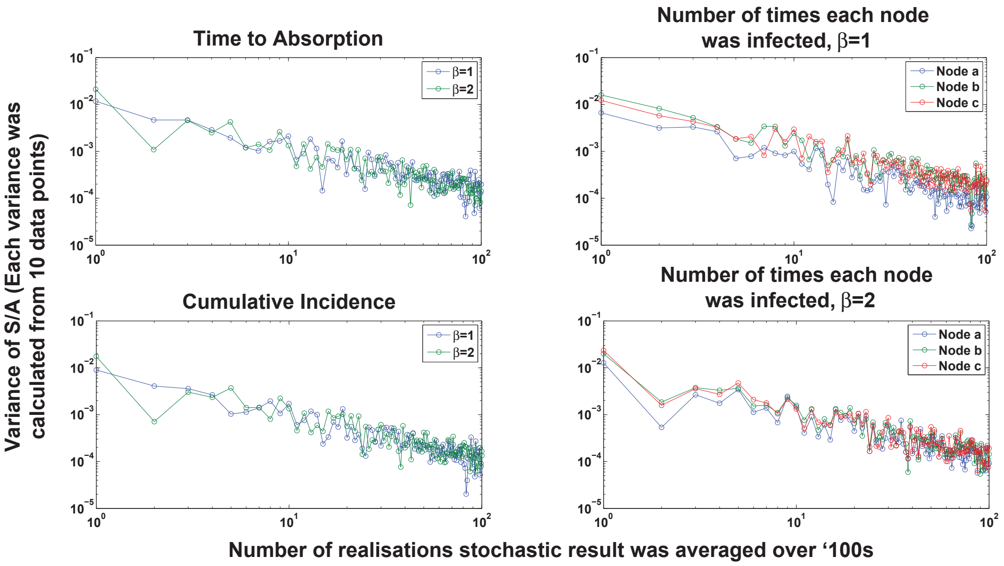


Figure 5.4: Convergence of the variance of $\log(S/A)$ for the Triangle Network with initial state ISS (where node a is infectious and nodes b and c are susceptible).

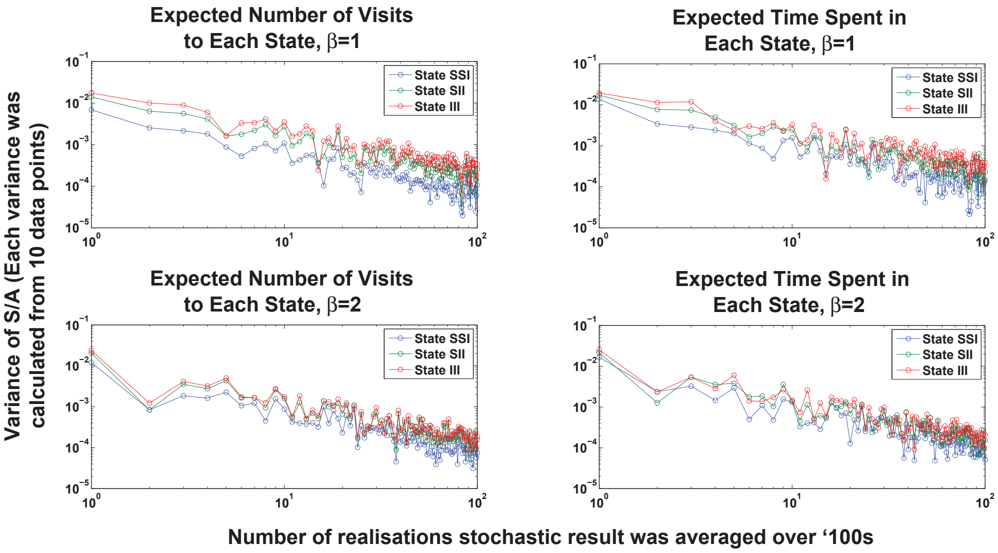


Figure 5.5: Convergence of the variance of $\log(S/A)$ for the Triangle Network with initial state ISS.

5.3.3 Results

We analyse an SIS model on each of the small networks of three and four nodes following the methods illustrated above for the triangle and lollipop networks. Here, we present the results from all networks considered and refer the reader to the appendices for the full details of the analysis for each network. In Figures 5.6 to 5.11 we show the distributions for the expected number of times each node is infected for the duration of the epidemic for each initial condition considered. In Table 5.1 we give the cumulative incidence and expected time to absorption results for parameter values $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ which correspond to each of the distributions shown in Figures 5.6 to 5.11. Table 5.1 also shows the clustering coefficient and degree of the initial infectious node(s) for each of the small networks of three and four nodes considered.

Table 5.1: Cumulative Incidence (C.I) and Expected Time to Absorption, E[Time to Abs], for an SIS epidemic on small networks.

Networks with $N = 3$				**C.I**		**E[Time to Abs]**[1]	
	ϕ	**Initial state**	**Degree of Initial Infectious Node(s)**	$\{\beta = 1, \gamma = 1\}$	$\{\beta = 2, \gamma = 1\}$	$\{\beta = 1, \gamma = 1\}$	$\{\beta = 2, \gamma = 1\}$
A. Triangle	1	SSI	2	5	13	2.67	5.67
B. Line	0	SSI	1	2.82	6.15	1.8	3.12
		SIS	2	3.43	7.18	2.07	3.55
		ISI	1 & 1	4.75	8.91	2.63	4.23
Networks with $N = 4$							
C. Complete	1	$SSSI$	3	16	79	6	24
D. Square	0	$SSSI$	2	6.75	26.54	3.13	9.15
		$SIIS$	2 & 2	10.25	34	4.38	11.5
E. Star	0	$SSSI$	1	3.97	12.45	2.16	4.92
		$ISSS$	3	5.46	15.73	2.74	6.05
		$SSII$	1 & 1	6.57	17.64	3.18	6.7
		$SIII$	1, 1 & 1	8.4	20.1	3.73	7.42
F. Toast	0.75	$SSIS$	2	9.99	45.44	4.14	14.58
		$SSSI$	3	11.24	48.86	4.57	15.62
		$SIIS$	2 & 2	14.65	56.87	5.74	18.05
G. Line	0	$SSSI$	1	3.46	10.7	2	4.4
		$SSIS$	2	4.52	13.16	2.41	5.24
		$SISI$	2 & 1	6.67	16.92	3.23	6.5
		$ISSI$	1 & 1	6.17	16.17	3.04	6.25
H. Lollipop	0.6	$SSSI$	1	5.23	21.1	2.57	7.49
		$SISS$	2	6.94	25.74	3.21	9
		$ISSS$	3	7.69	27.53	3.48	9.57
		$SSII$	2 & 1	9.73	31.7	4.21	10.89

---

[1] The expected time to absorption is in units of γ^{-1}.

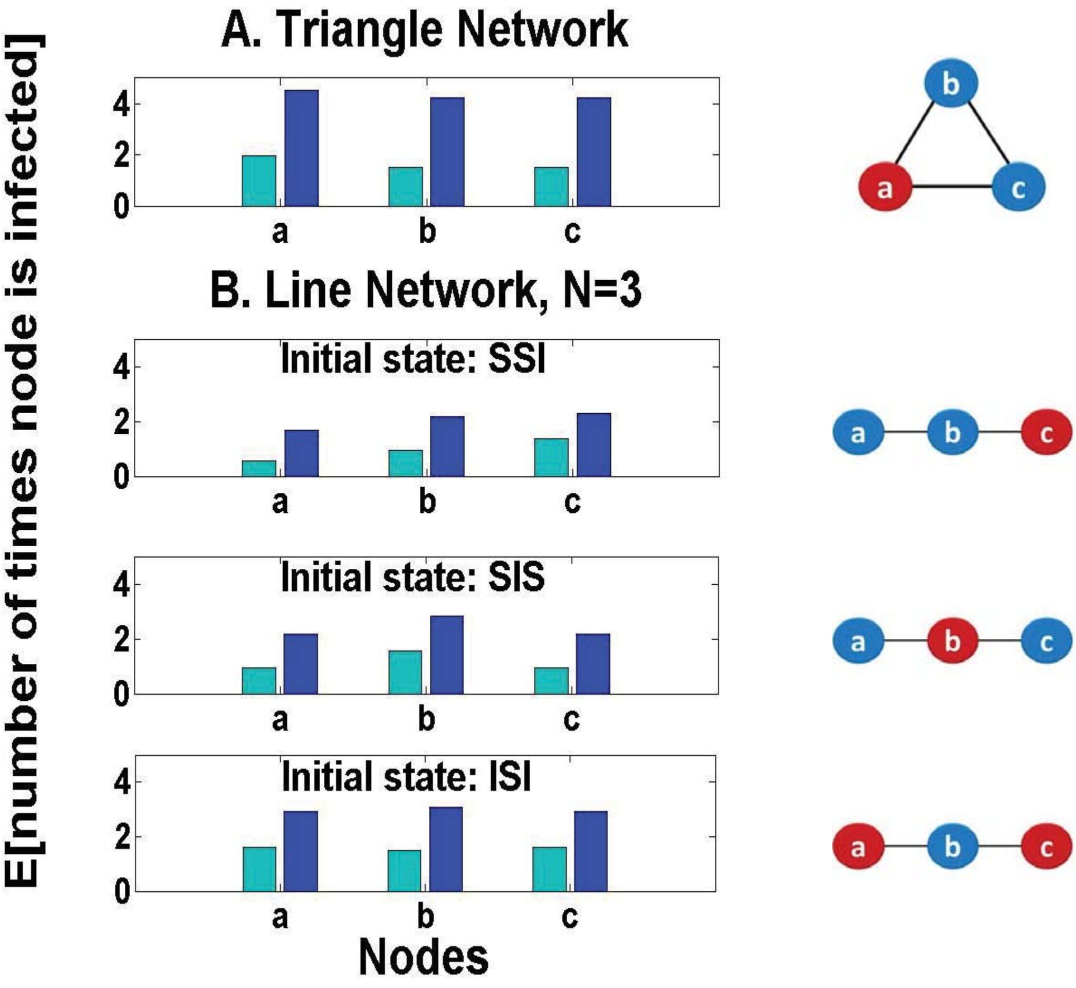


Figure 5.6: Expected number of times each node is infected during the epidemic for the triangle and line networks with $N = 3$ nodes. The bars in light and dark blue are for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ respectively. Schematic diagrams of the initial state are on the right.

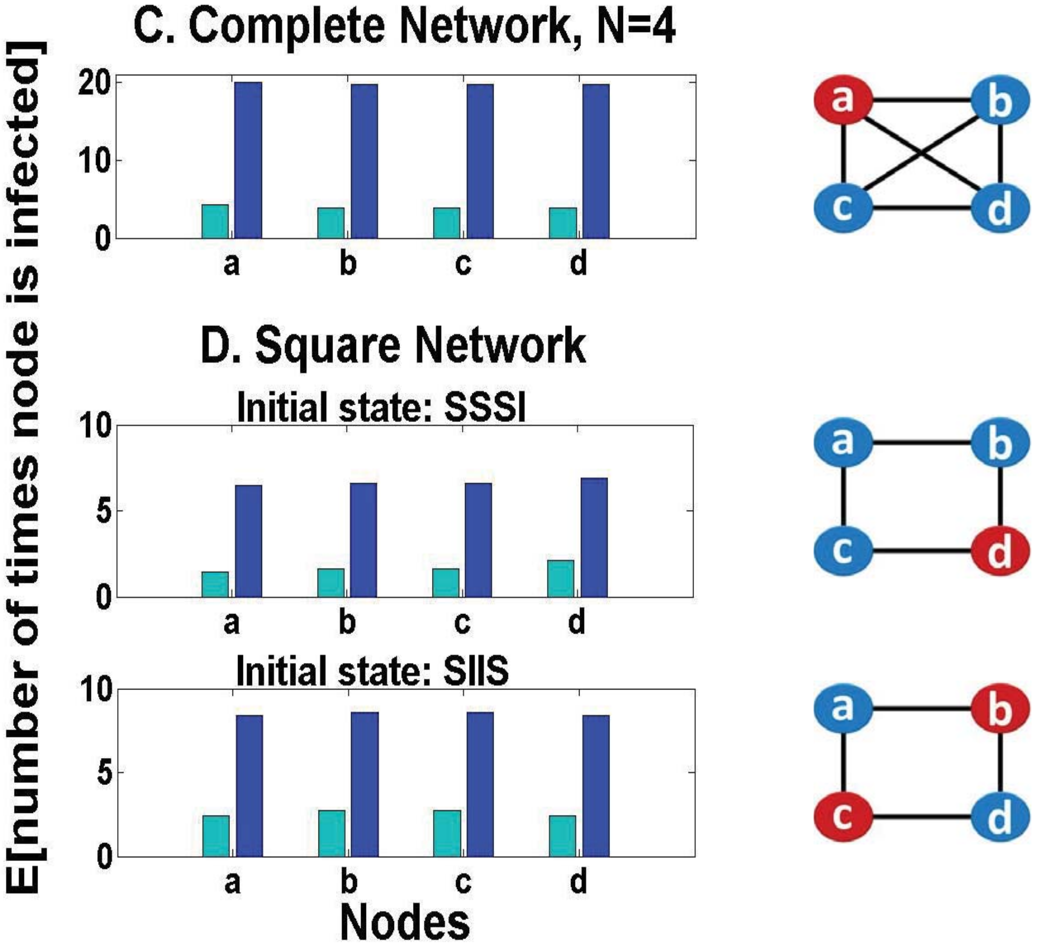


Figure 5.7: Expected number of times each node is infected during the epidemic for the complete and square networks with $N = 4$ nodes. The bars in light and dark blue are for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ respectively. Schematic diagrams of the initial state are on the right.

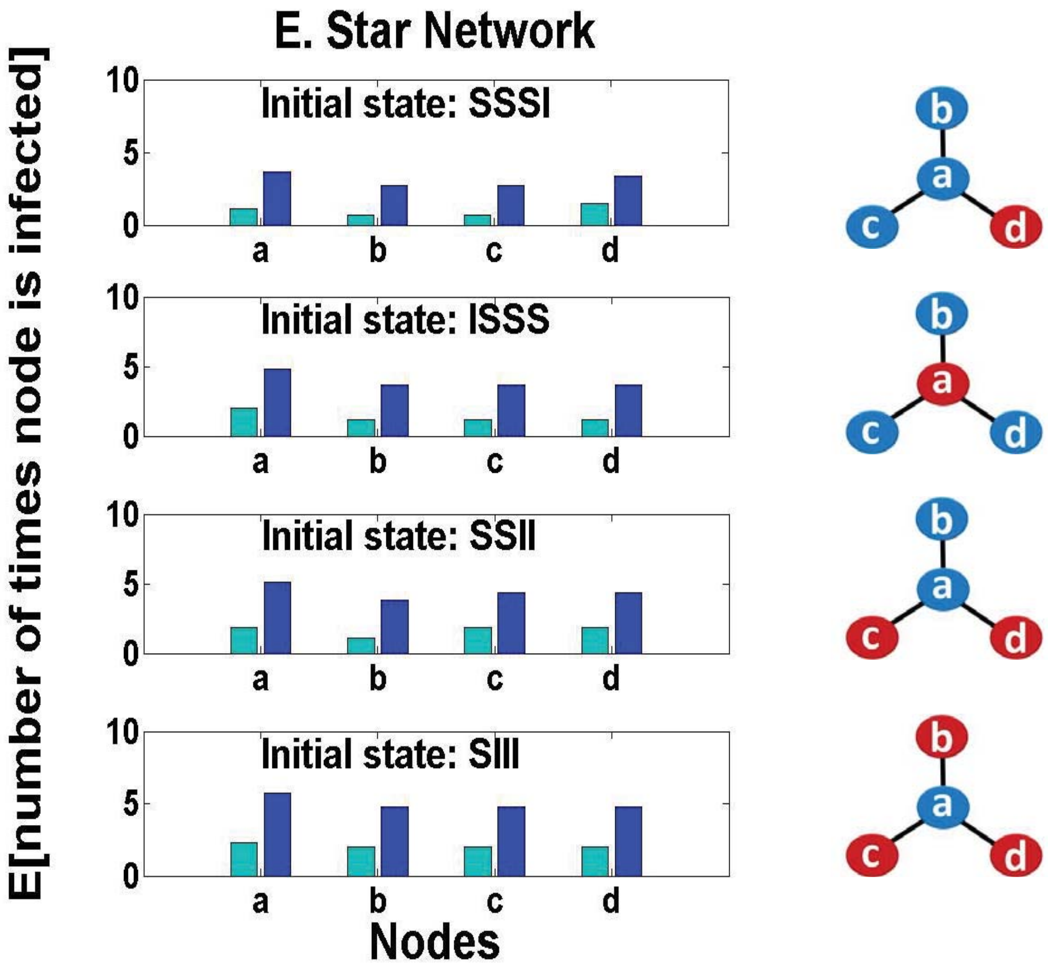


Figure 5.8: Expected number of times each node is infected during the epidemic for the star network with $N = 4$ nodes. The bars in light and dark blue are for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ respectively. Schematic diagrams of the initial state are on the right.

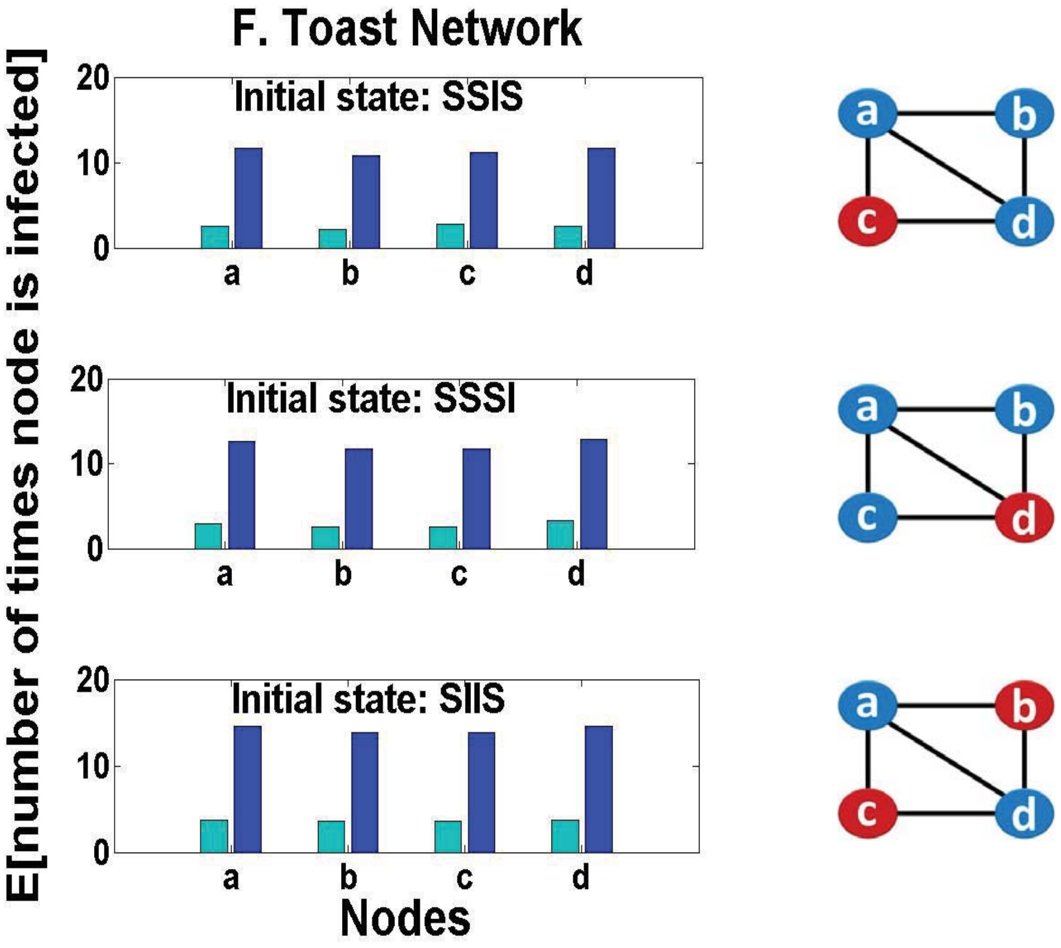


Figure 5.9: Expected number of times each node is infected during the epidemic for the toast network with $N = 4$ nodes. The bars in light and dark blue are for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ respectively. Schematic diagrams of the initial state are on the right.

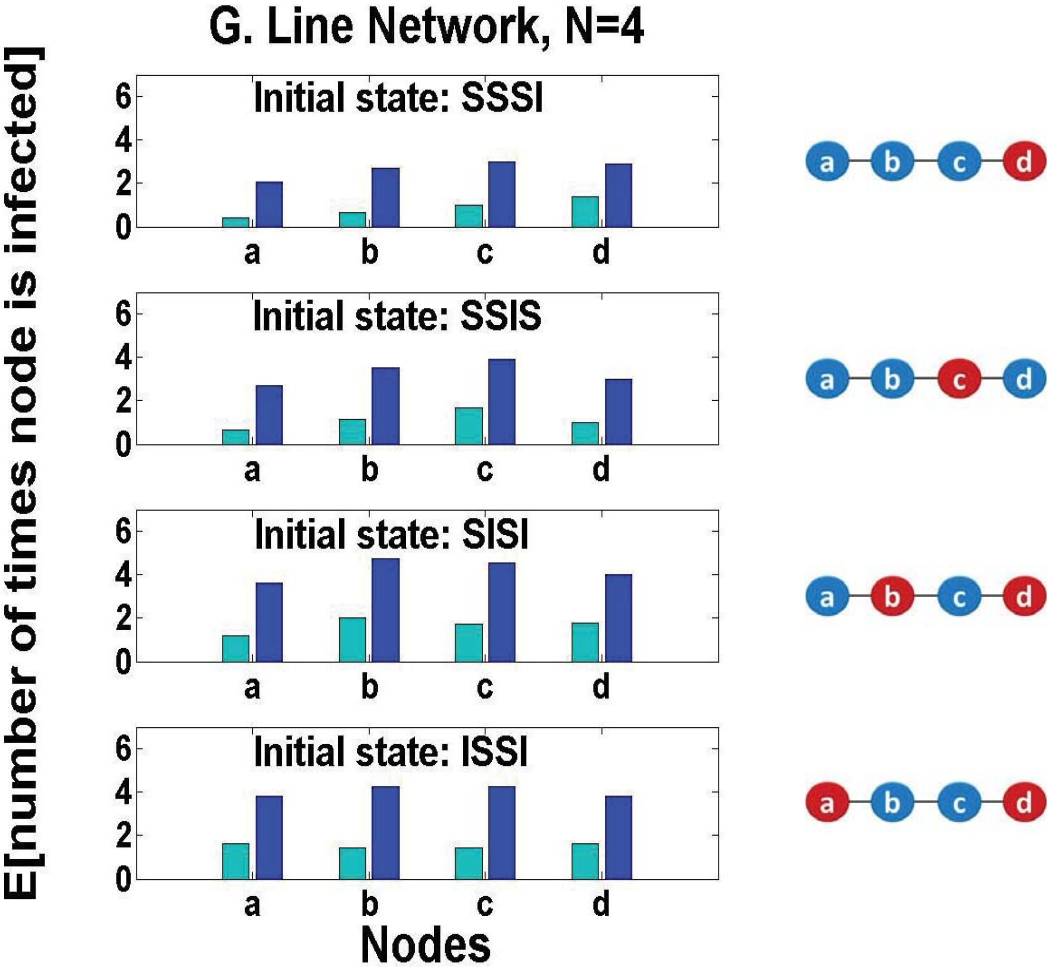


Figure 5.10: Expected number of times each node is infected during the epidemic for the line network with $N = 4$ nodes. The bars in light and dark blue are for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ respectively. Schematic diagrams of the initial state are on the right.

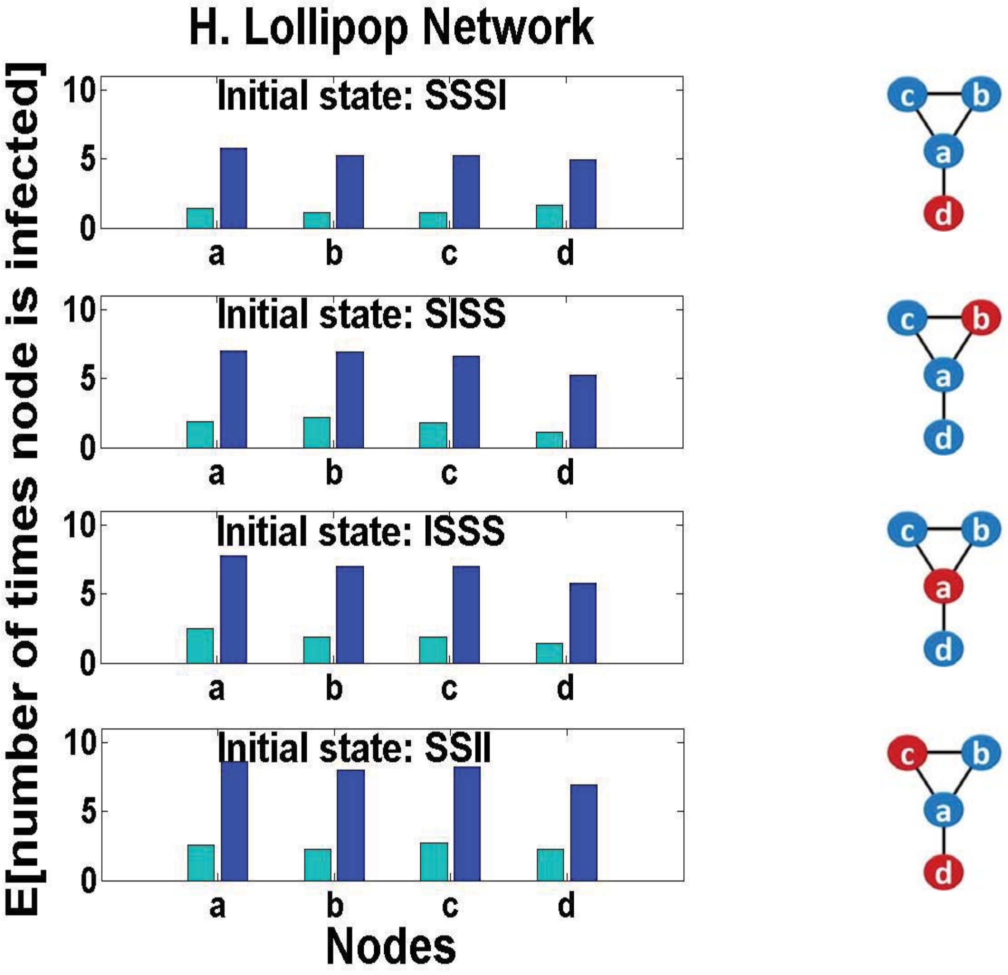


Figure 5.11: Expected number of times each node is infected during the epidemic for the lollipop network with $N = 4$ nodes. The bars in light and dark blue are for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ respectively. Schematic diagrams of the initial state are on the right.

5.4 Discussion

We have derived, using Markov Chain theory, analytic results that describe the behaviour of an SIS model on small networks. We have shown that while all results can be found analytically, some results can be displayed more easily in symbolic form than others, and all results can be computed exactly for given parameter values. Increasing the complexity of the network structure reduced the effect of grouping states together based on symmetries of the network. Consequently this increased the complexity of the analysis which is evident in the size of the matrices for various results. The results presented here form the basis for finding tractable analytic results that describe the spread of an infection through large networks composed of the small networks discussed in this chapter. We calculated the clustering coefficient, ϕ (see Table 5.1), for each of the small networks in order to see if there was any correlation between the results (such as the expected time to absorption) and how clustered the network is.

For the networks of size $N = 3$ we have the triangle network and the line network. The clustering coefficient for these networks is $\phi = 1$ and $\phi = 0$ respectively. From Table 5.1 we can see that the triangle network had a longer time to absorption and higher cumulative incidence than the line network with $N = 3$ nodes for all initial conditions considered. For the networks of size $N = 4$ we investigated the complete, square, star, toast, line and lollipop networks. The clustering coefficients for each of these networks are given in Table 5.1. The complete network, starting with one infectious node, had the longest expected time to absorption and highest cumulative incidence for all parameter values considered when compared to all other networks of size $N = 4$ across all initial states considered. This was followed by the toast network, however there was a large difference between the expected time to absorption and cumulative incidence when compared to the complete network. Interestingly, if we look at the results for the square network starting with initial state $SSSI$ (one infectious node of degree 2) and compare with the lollipop network starting with initial state $SISS$ (one infectious node of degree 2) we see that the results are similar even though the clustering coefficients for the two networks are 0 and 0.6 respectively.

When the initial number of infectious nodes is increased, the cumulative incidence and time to absorption increases, this was seen across all small networks. When starting with one infectious node, the degree of this node also affected the cumulative incidence and time to absorption. With an infectious node of higher degree, the cumulative incidence and time to absorption was increased. Our results give us a good indication that both the network topology and degree of the initial infectious node are key factors in understanding how an infection might spread through small networks that could represent small populations or communities. Some of the results presented here are intuitive; however, we emphasize the importance of being able to present results in symbolic form which are exact. Being able to do this allows us to gain more information about what the main driving factor of each result is (i.e transmission or recovery rate).

For the SIS epidemic process on small networks so far only a small range of parameter values have been considered, this could be extended and a sensitivity analysis on the parameter space might be worthwhile to see how different ranges of transmission and recovery rates affect the SIS epidemic process.

Chapter 6

An SIS model on a Line of Triangles Network.

In this chapter we explore the possibility of generalising results that describe how an SIS type infection spreads through a line of triangles (LoT) network (see Figure 6.1). We chose the LoT network as it can be assembled from a combination of lollipop and triangle networks. The SIS model on a LoT network needs to be approached in a different manner to the SIR model on a LoT network discussed in Chapter 3 (page 61). This is because re-infection of individuals is possible in an SIS model which is not possible in an SIR model and adds some complexity to the analysis on infection dynamics. In Section 5.3.1 (page 150) we gave an outline of a MATLAB program which was used to investigate the dynamics of an SIS model on small networks. Here, we make use of the same program incorporating a function for the adjacency matrix of a LoT network with a specified number of nodes, N. We discuss the methods used to investigate the dynamics of an SIS type infection on a LoT network and present results for the LoT network with $N = 6$ nodes. We then give details of an algorithm which can be used to generate the transition probability matrix, from which all results can be found, for any network of size N.

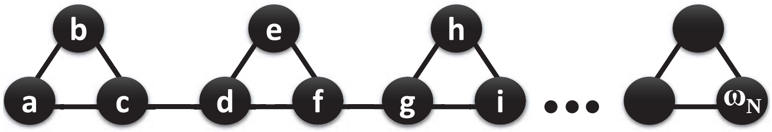


Figure 6.1: Schematic of a line of triangles network.

6.1 Methods

We approach analysing the line of triangles network in a similar way to the small networks in the previous chapter, except now we are more conscious of the size of the system. For an SIS model on a network with $N = 6$ nodes we have $2^N = 64$

possible states in which the network can be. We group states together based on the number of nodes in each infection state and network topology; this reduces our system to 21 states. For the LoT network with $N = 6$ nodes we are able to find the symbolic representation of the transition matrices, $\mathbf{P}$ and $\mathbf{Q}$, for the full and lumped systems respectively. The transition matrices are sparse and the non-zero entries are probabilities of transition from one state to another due to the infection or recovery of a node. The matrix $\mathbf{P}$ has entries which describe the probability of transition from one state to another for every state in the LoT network. The matrix $\mathbf{Q}$ has entries which describe the probability of transition from one state to another for every transient state in the LoT network; that is $\mathbf{Q}$ is the matrix $\mathbf{P}$ with the last column and row deleted which correspond to the absorbing state of the system. The non-zero entries of $\mathbf{P}$ and $\mathbf{Q}$ are ratios of two linear functions in β and γ. In the following sections we illustrate our methods using the LoT network with $N = 6$ nodes, however it is important to note that these methods can be used for a LoT network with an arbitrary number of nodes, N.

Expected number of visits to each state

The transition matrix for the transient states only of the SIS model on a LoT network is defined by $\mathbf{Q}$. From $\mathbf{Q}$ and using the above information, we can find the expected number of times the process visits each transient state from a given initial state.

$$\mathbf{N} \;=\; (\mathbf{I} - \mathbf{Q})^{-1}$$

We omit the algebraic expression of $\mathbf{N}$ due to the complexity of its entries. We are able to find the symbolic form of the matrix, $\mathbf{N}$, containing the expected number of visits to each state for a given initial state; however we can only display the whole matrix corresponding to the lumped system in MATLAB as the matrix for the full system is too large and as a result the results are truncated. However, the expression for the determinant of the matrix $(\mathbf{I} - \mathbf{Q})$ for the full system can be found. The numerator and denominator are polynomials of order 61. The numerator of the determinant can be factored into four polynomials of order 3, 15, 19 and 24. The largest coefficient in one of these polynomials is greater than $10^{1}0$. The denominator of the determinant can be factored into the following expression

$$\begin{aligned} & 12(\beta+\gamma)^7(\beta+2\gamma)^2(2\beta+\gamma)^9(\beta+3\gamma)^2(3\beta+\gamma)^2(3\beta+2\gamma)^4(2\beta+5\gamma)^4 \\ \times \quad & (3\beta+4\gamma)^4(4\beta+3\gamma)^8(5\beta+2\gamma)^4(3\beta+5\gamma)^2(5\beta+3\gamma)^9(5\beta+4\gamma)^4 \end{aligned}$$

Even though the numerator and denominator can be factorized separately, the numerator and denominator do not share further factors.

Expected time spent in each state during the epidemic

One property of interest is the expected time spent in each of the transient states over the course of an epidemic, defined in Chapter 5 and denoted by $\mathbf{E_T}$. Where the element $\mathbf{E_T}(i, j)$ gives the expected time spent in state j given that state i was the initial starting state. However, following on from the discussion in the previous section for the expected number of visits to each state, we do not give the symbolic representation of $\mathbf{E_T}$ due to the complexity of its entries.

Time to absorption

The expected time to absorption (see Chapter 5 for details) is the expected time it takes for the system to reach the infection free steady state. For the LoT network with $N = 6$ nodes, the expected time to absorption given that the SIS epidemic started in the state where node a is infectious is given by an expression that is the ratio of two 19^{th} order polynomials in β and γ. Table 6.1 contains the time to absorption for the LoT network with $N = 6$ and $N = 9$ for each initial condition considered, evaluated with $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

Number of times each node is infected

We determine the expected number of times each node is infected in the LoT network with $N = 6$ for two initial states; where node a and node c are the initial infectious nodes and the remaining $N-1$ nodes are susceptible (see Figure 6.2). For a specified initial state, the number of times each node is infected for the LoT network with $N = 6$ is given by a 1×6 vector. The entries are ratios of large polynomials in β and γ (of up to order 46) which cannot be factorised. Figures 6.2 and 6.3 show a distribution of the number of times each node is infected for the LoT network with $N = 6$ and $N = 9$ for each initial state and specified parameter values.

Cumulative Incidence

The cumulative incidence is the total number of new infections that occur during an epidemic. For an SIS model on the LoT network with $N = 6$ the expression for the cumulative incidence is the ratio of two polynomials in β and γ of order 19 and 20. Table 6.1 shows the cumulative incidence evaluated with $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ for the LoT network with $N = 6$ and $N = 9$.

Table 6.1: Cumulative Incidence (C.I) and Expected Time to Absorption for an SIS epidemic on a LoT network

Line of Triangles Network				**C.I**		**E[Time to Abs]**[1]	
Number of Nodes	ϕ	**Initial state**	**Degree of Initial Infectious Node**	$\{\beta = 1, \gamma = 1\}$	$\{\beta = 2, \gamma = 1\}$	$\{\beta = 1, \gamma = 1\}$	$\{\beta = 2, \gamma = 1\}$
6	0.6	$ISSSSS$	2	12.31	119.59	4.45	27.84
6	0.6	$SSISSS$	3	15.01	133.18	5.20	30.85
9	0.5294	$ISSSSSSSS$	2	20.49	766.44	5.88	118.4
9	0.5294	$SSISSSSSS$	3	26.08	860.55	7.14	132.7
9	0.5294	$SSSISSSSS$	3	29.28	893.95	7.79	137.66
9	0.5294	$SSSSISSSS$	2	25.23	822.88	6.84	126.81

---

[1] The expected time to absorption is in units of γ^{-1}.

6.2 Results for a line of triangles network with $N = 6$ and $N = 9$ nodes

Here we present some results for the LoT networks with $N = 6$ and $N = 9$ nodes and show that our methods produce accurate results. We compare our analytical results with results found from a stochastic SIS model on the two networks and show that they are correct. Table 6.1 gives the cumulative incidence and expected time to absorption results for the LoT with $N = 6$ and $N = 9$ nodes for all initial conditions considered. Figure 6.2 shows the distribution for the number of times each node is infected during the epidemic for the LoT with $N = 6$ nodes with specified initial conditions and infection parameters. Figure 6.3 shows the distribution for the number of times each node is infected during the epidemic for the LoT with $N = 9$ nodes with specified initial conditions and infection parameters.

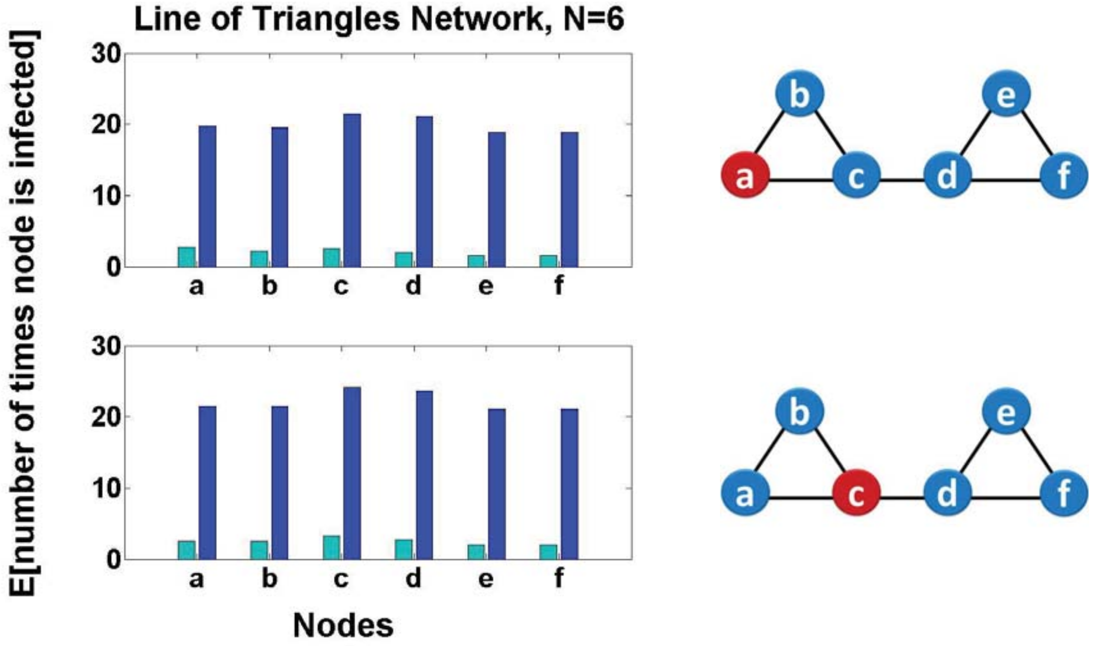


Figure 6.2: Expected number of times each node is infected during the epidemic for the line of triangles with $N = 6$ nodes with initial condition as shown on the right. Note that due to the symmetry of the network, nodes a and b are identical and as such we do not show the results for starting with node b as the initial infectious node. The bars in light and dark blue are for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ respectively.

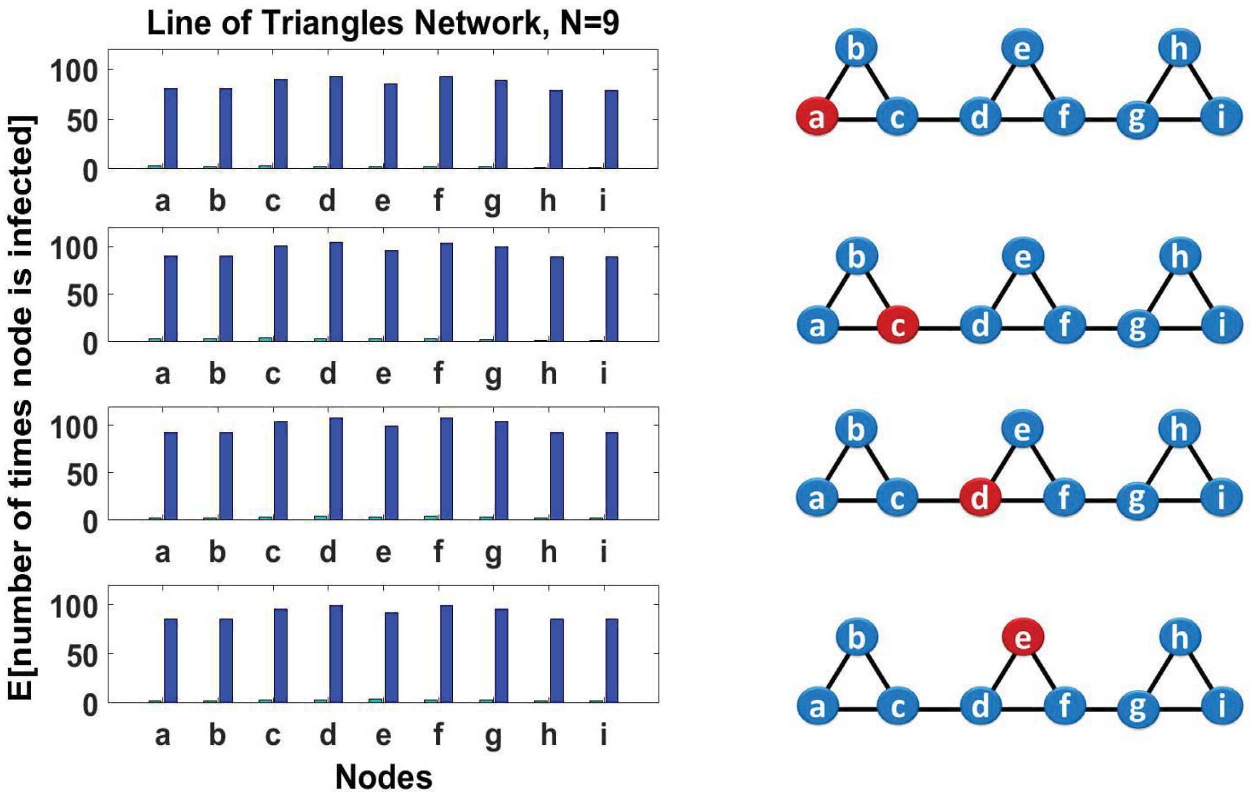


Figure 6.3: Expected number of times each node is infected during the epidemic for the line of triangles with $N = 9$ nodes with initial condition as shown on the right. Note that due to the symmetry of the network, nodes a and b are identical and as such we do not show the results for starting with node b as the initial infectious node. The bars in light and dark blue are for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$ respectively.

6.3 Transition matrix for an SIS model on any network

To construct the transition matrix for an SIS model on a network of N nodes we follow the steps below.

1. Find the adjacency matrix, $\mathbf{A}$, which contains entries $\mathbf{A}_{ij} = 1$ if nodes i and j in the network are connected, and zero otherwise.

2. Determine the degree distribution by summing up each column of $\mathbf{A}$. Thus the entries $\mathbf{d}_j = \sum_i \mathbf{A}_{ij}$ gives the degree of node j in the network.

3. Find the neighbours, j, of each node i by determining where $\mathbf{A}_{ij} = 1$.

4. Generate a binary matrix $\mathbf{V}$ of order $N \times 2^N$ which contains all possible states in which the network can be, where each entry $\mathbf{V}_{ij}$ corresponds to the infection status of node i in state j. If $\mathbf{V}_{ij} = 0$, node i is Susceptible in state j; if $\mathbf{V}_{ij} = 1$, node i is Infectious in state j.

5. Determine if an infection or recovery event is possible between states. To do this we find the absolute value of the sum of the difference of two states, if this is equal to 1 it means a transition event is possible. If the sum of the difference between two columns (states) is equal to -1 that means that an infection event can occur. If the sum of the difference between two columns (states) is equal to 1 that means that a recovery event can occur. That is, if $\sum(\mathbf{V}(:, s_1) - \mathbf{V}(:, s_2)) = -1$ then a susceptible node in state s_1 is infectious in state s_2. Similarly, if $\sum(\mathbf{V}(:, s_1) - \mathbf{V}(:, s_2)) = 1$ then an infectious node in state s_1 is recovered in state s_2.

   For example if we had the following two states for a network of $N = 6$ nodes,

   $$s_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad s_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

   to determine if the transition from state s_1 to state s_2 is possible we calculate

   $$\left|\sum(s_1 - s_2)\right| = |-1 + 0 + 0 + 0 + 0 + 0| = 1.$$

   Therefore, transition from state s_1 to state s_2 is possible. Moreover, as

   $$\sum(s_1 - s_2) = -1 + 0 + 0 + 0 + 0 + 0 = -1$$

we know that this event is an infection event where the first node moved from the susceptible to the infectious class.

6. We then find the total rate for leaving each state and store it in a vector, $\mathbf{Tr}$. The total rate for leaving state i is given by $\mathbf{Tr}_i = \mathcal{SI}\beta + \mathcal{I}\gamma$, where $\mathcal{SI}$ is the number of $S-I$ edges that are in state i and $\mathcal{I}$ is the number of infectious nodes in state i.

7. Next we construct the transition matrix, $\mathbf{P}$, which has entries $\mathbf{P}_{ij}$ corresponding to the probability of transition from state i to state j. We start by defining $\mathbf{P}$ as a zero matrix of order $i \times j$. Then, if the transition from state i to state j involves the infection of node k then we have the following probability of transition from state i to state j:

   $$\mathbf{P}_{ij} = \frac{\mathcal{I}\beta}{\mathbf{Tr}_i}$$

   where $\mathcal{I}$ is the number of infectious neighbours node k has.
   If the transition from state i to state j involves the recovery of node k then we have the following probability of transition from state i to state j:

   $$\mathbf{P}_{ij} = \frac{\gamma}{\mathbf{Tr}_i}$$

   All entries of $\mathbf{P}_{ij}$ that are zero indicate where transition from state i to state j is not possible.

From the transition matrix, $\mathbf{P}$, we can find the matrix $\mathbf{Q}$ by removing the row and column corresponding to the absorbing state from $\mathbf{P}$. From $\mathbf{Q}$ all of the results discussed in this chapter for the line of triangles network can be obtained by following the appropriate definitions given. By simply defining the appropriate adjacency matrix for a particular network and specifying the number of nodes, N, all of the results presented in this chapter can be adapted to investigate an SIS model any network structure.

6.4 Discussion

For the LoT network, each time we want to make the network larger by adding more triangles we increase the number of nodes in the network by a multiple of 3. Thus, going to a LoT network with $N = 9$ nodes (with $2^N = 512$ states) from a LoT network with $N = 6$ nodes (with $2^N = 64$ states) by adding just one triangle increases the number of possible states in which the network can be by a factor of 2^3. In other words, each time the LoT network has another triangle added to it, the number of possible states is multiplied by 8. So, this system increases in size rapidly as N increases. Due to this, the methods that have been used so far are not practical or computationally efficient enough to be useful in describing the dynamics of an SIS type infection on large networks. However, in theory, the algorithm given in Section 6.3 can be used for any network structure of any network size to generate the transition probability matrix. Once we have the transition probability matrix, all other results that were presented throughout this chapter can be derived for any network of any size.

This is in contrast to the SIR type infection results found in a previous chapter. This is largely due to the key difference between the two types of infections, in an SIR model once individuals are infected they can transmit infection for a time but they will end up in the recovered or removed class where they no longer take part in the epidemic. Whereas, in the SIS model once an individual is infected, they can transmit infection to other individuals before returning to the susceptible class where they are able to be infected again. This cyclic behaviour continues until there are no infected individuals in the network. This major difference in infection dynamics is the reason behind the different approaches used and subsequently the different results produced for the small and large networks discussed in this thesis.

Chapter 7

Illustrations of an SIS model on small networks

Here we illustrate how the network models presented in Chapter 5 can be used to investigate the dynamics of the spread of an infection through small populations that could be representative of a household or community. For the SIS model on small networks we focus on two infections, rabies and chlamydia trachomatis. In order to obtain results specific to rabies and chlamydia from our network models we need to determine what the infection and recovery rates (β and γ respectively) are for both infections. In Sections 7.1 and 7.2 we parameterise the network models for rabies and chlamydia respectively. For each infection we discuss the assumptions made and how we used data from previous studies to determine the parameters required for the small network models. In Section 7.3 we present and discuss the results.

7.1 Rabies

Rabies is an infectious zoonotic disease caused by a virus; it is endemic on every continent except Antarctica. Rabies is endemic in many wildlife populations (including bats, foxes and jackals), however, domestic dog populations are the main reservoir for the transmission of rabies to humans with over 90% of human cases caused by bites from domestic dogs. It is a horrific disease where the neurological phase can include increased aggression, the tendency to bite (and hence transmit) followed by a rapid deterioration of health and progression to death [23]. Rabies is a preventable disease and there are currently effective pre- and post - exposure vaccines available, however these vaccines are expensive and hard to obtain in the areas where they are needed most (rural areas in Africa and Asia). Recent research efforts currently focus on understanding the transmission dynamics of rabies among domestic dogs with the aim of eliminating human and canine cases of rabies [13, 22, 23, 27, 38, 69, 71].

Hampson et al [23] used contact-tracing methods to investigate the spread of rabies among domestic dogs in two districts of rural Tanzania, Serengeti and Ngorongoro. They estimated the basic reproduction number for Serengeti and Ngorongoro to be $R_0 = 1.1$ and $R_0 = 1.3$ respectively. They found no significant difference between these two estimates of R_0 even though there was a large difference in dog population density between the two districts (9.38 dogs/km^2 in Serengeti and 1.36 dogs/km^2 in Ngorongoro). The relationship between R_0 and population density for the transmission of rabies is not fully understood. While most studies to date assume a density dependent transmission rate (see, for example, [13, 22, 27, 38, 69, 71]), there is no empirical evidence to support such a claim. On the other hand, if the relationship between R_0 and population density is subtle it could be difficult to detect. Hence, this is an area of ongoing research aimed at understanding the transmission dynamics of rabies among domestic dogs more thoroughly. Hampson et al [23] give a table comparing historical values for the basic reproduction number that have been estimated for rabies epidemics in rural and urban areas around the world. They found $1 < R_0 < 2$ for all outbreaks of rabies in domestic dog populations.

Rabies is typically modelled as an SIR process as the infection almost certainly leads to the death of the host. Here we will use an SIS model for the spread of rabies in small networks. We argue that when one infectious dog dies from the infection it is replaced by another susceptible dog. We are interested in modelling the transmission of rabies among domestic dogs and therefore this replacement could happen, for example, if the owner gets a new dog or if its territory gets taken over. Hampson et al [23] found maximum likelihood estimates of infection parameters for the spread of rabies among domestic dogs in rural Tanzania. They estimated the infectious period ($1/\gamma$) to be 3.1 days and found the probability that an unvaccinated dog would develop rabies after being bitten to be 0.49. That is, the probability of transmission given contact (which in this case is a bite) is equal to 0.49. Therefore, we have the following parameters for the transmission of rabies using our small network models,

$$\psi = 0.49$$

recalling that $\psi = \dfrac{\beta}{\beta + \gamma}$, re-arranging the above equation and substituting $\gamma = 0.3226$ day^{-1} we find the transmission rate to be $\beta = 0.30993$ day^{-1} . We present and discuss the results for the transmission of rabies in small networks in Section 7.3.

7.2 Chlamydia

Chlamydia is a common sexually transmitted disease (STD) caused by infection with *Chlamydia trachomatis*. Many infected with chlamydia do not know they are infected as it is common for individuals to be asymptomatic. However, if left untreated it can lead to more serious health issues such as infertility, chronic pelvic pain or lymphogranuloma venereum which is another type of STD [1]. Transmission of chlamydia most commonly occurs via sexual contact with an infected partner but can also be passed from an untreated infected woman to her baby during childbirth. Individuals who have been infected with chlamydia become susceptible again after treatment.

Heijne et al [26] use a pair approximation model for the transmission of chlamydia in a sample population of individuals aged 16 to 24 in the United Kingdom (see [30] for data collection details). They argue that the basic reproduction number does not distinguish between the number of secondary infections and the number of secondary cases produced by a single infectious individual, where the number of secondary cases accounts for the possibility that a single individual can be infected more than once [26]. They derive a case reproduction number (R_c) which is the average number of secondary cases that a single infectious individual will infect during their infectious period. They also define the partnership reproduction number (R_p) which is *'the average number of secondary partnerships consisting of two infected individuals one typical infected individual will produce during their infectious lifetime'*. The main result the authors present is the distinction between the average number of secondary infections, R_0, and R_c. The impact of re-infections on the transmisison dynamics is able to be quantified by using the case reproduction number, R_c. Heijne et al [26] also show that having $R_c < 1$ does not necessarily mean that the epidemic is decreasing and R_0 can still be above 1. Previous studies that have confused the definitions of R_0 and R_c have falsely suggested that the epidemic was declining as $R_c < 1$. Hence, it is important to understand the effects of re-infection on transmission dynamics for chlamydia.

To investigate the spread of chlamydia using the small network models we use parameters as outlined in [2, 25, 26]. Thus, the probability of transmission per sex act is 0.1 and the duration of the infectious period is one year ($1/\gamma = 1$ year). For chlamydia, the probability of transmission given contact is equivalent to the probability of transmission per sex act, therefore

$$\psi \quad = \quad 0.1$$

re-arranging the above equation and substituting $\gamma = 1$ year^{-1} we find the transmission rate to be $\beta = 0.1111$ year^{-1}. We present and discuss the results for the transmission of chlamydia in small networks in Section 7.3.

7.3 Results & Discussion

In this section we present and discuss our results for the transmission of rabies and chlamydia through small networks. As we were able to find estimates of the infectious period and probability of transmission given contact for rabies and chlamydia, we did not need to use our definitions of the basic reproduction number, R_0, for each small network (as found in Section 4.1.2) to determine our infection parameters β and γ. However, in Table 7.1 we present the expressions for R_0 and evaluate R_0 for each small network for rabies and chlamydia to see how they compare. Note that in Table 7.1 the parameters used to find R_0 for chlamydia are $\beta = 0.1111$ year^{-1}, $\gamma = 1$ year^{-1} and the parameters used to find R_0 for rabies are $\beta = 0.30993$ day^{-1}, $\gamma = 0.3226$ day^{-1}

Table 7.2 shows the cumulative incidence and expected time to absorption (expected time it takes for the infection to die out) for rabies and chlamydia epidemics on small networks with specified initial conditions. Note that the expected time to absorption are in units of days for rabies and years for chlamydia. From Table 7.2 we can see that it takes 1-2 years for the chlamydia infection to die out in the network and it takes approximately 6 - 18 days for rabies to die out. So while the rabies epidemic is much shorter than an epidemic for chlamydia, the cumulative incidence is much higher for rabies due to the higher probability of transmission given contact.

The large difference between the probability of transmission given contact for rabies (0.49) and chlamydia (0.1) is reflected in Figures 7.1 to 7.6. These figures show the expected number of times each node becomes infected during an epidemic. Rabies has a relatively short infectious period of a few days but during this time on average 49% of bites from a rabid (infectious) dog result in a new infectious individual. In contrast, chlamydia has a longer infectious period of one year yet contact with an infectious individual only has a 10% chance of resulting in a new infectious individual. The plots in Figures 7.1 to 7.6 are dominated by the initial infection for chlamydia as the infection dies out quickly whereas this is not the case for rabies due to the higher probability of resulting secondary infections.

Table 7.1: R_0 for rabies and chlamydia trachomatis on small networks.

Network	R_0	R_0 **(rabies)**	R_0 **(chlamydia)**
Triangle	$R_0 = 2\psi$	0.98	0.2
Line $(N = 3)$	$R_0 = \sqrt{2}\psi$	0.69	0.14
Complete $(N = 4)$	$R_0 = 3\psi$	1.47	0.3
Square	$R_0 = 2\psi$	0.98	0.2
Star $(N = 4)$	$R_0 = \sqrt{3}\psi$	0.85	0.17
Toast	$R_0 = \frac{1+\sqrt{17}}{2}\psi$	2.51	0.51
Line $(N = 4)$	$R_0 = \frac{1+\sqrt{5}}{2}\psi$	1.59	0.32
Lollipop	$R_0 = 2.1701\psi$	1.06	0.22

Table 7.2: Cumulative Incidence (C.I) and Expected Time to Absorption, E[Time to Abs], for an SIS epidemic on small networks.

Networks with $N = 3$				C.I		E[Time to Abs]	
	ϕ	Initial state	Degree of Initial Infectious Node(s)	Rabies	Chlamydia	Rabies [1]	Chlamydia [2]
A. Triangle	1	SSI	2	4.77	1.25	7.99	1.12
B. Line	0	SSI	1	2.72	1.12	5.46	1.06
		SIS	2	3.31	1.22	6.27	1.11
		ISI	1 & 1	4.62	2.23	7.97	1.6
Networks with $N = 4$							
C. Complete	1	$SSSI$	3	14.74	1.42	17.41	1.19
D. Square	0	$SSSI$	2	6.32	1.25	9.25	1.12
		$SIIS$	2 & 2	9.7	2.47	13.01	1.69
E. Star	0	$SSSI$	1	3.77	1.13	6.48	1.07
		$ISSS$	3	5.19	1.34	8.22	1.16
		$SSII$	1 & 1	6.29	2.26	9.55	1.61
		$SIII$	1, 1 & 1	8.09	3.37	11.25	1.97
F. Toast	0.75	$SSIS$	2	9.26	1.28	12.12	1.13
		$SSSI$	3	10.45	1.39	13.4	1.18
		$SIIS$	2 & 2	13.72	2.52	16.9	1.71
G. Line	0	$SSSI$	1	3.29	1.12	6	1.06
		$SSIS$	2	4.3	1.24	7.24	1.11
		$SISI$	2 & 1	6.41	2.35	9.72	1.64
		$ISSI$	1 & 1	5.91	2.25	9.15	1.6
H. Lollipop	0.6	$SSSI$	1	4.89	1.14	7.61	1.07
		$SISS$	2	6.51	1.26	9.5	1.12
		$ISSS$	3	7.24	1.36	10.32	1.17
		$SSII$	2 & 1	9.2	2.38	12.53	1.66

---

[1] Rabies estimates are in units of days^{-1}

[2] Chlamydia estimates are in units of years^{-1}

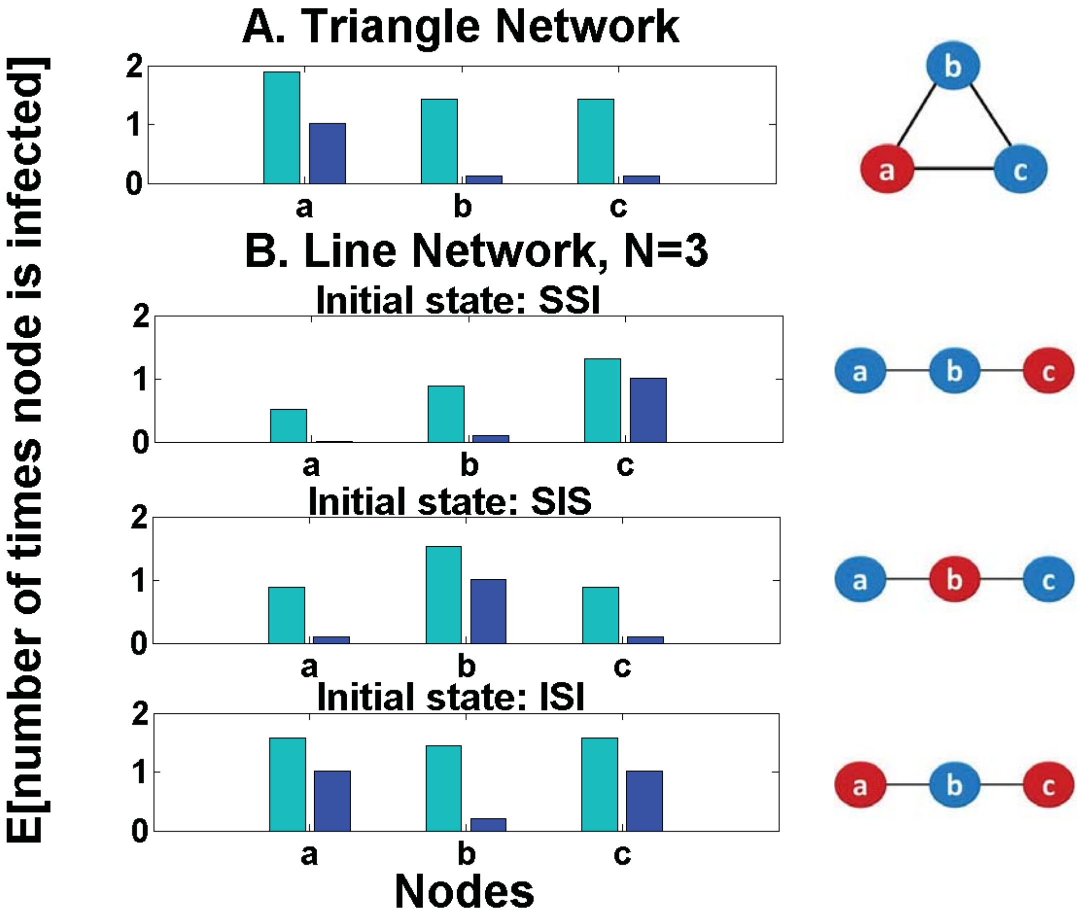


Figure 7.1: Expected number of times each node is infected during the epidemic for the triangle and line networks with $N = 3$ nodes. The bars in light and dark blue are for rabies and chlamydia respectively. Schematic diagrams of the initial state are on the right.

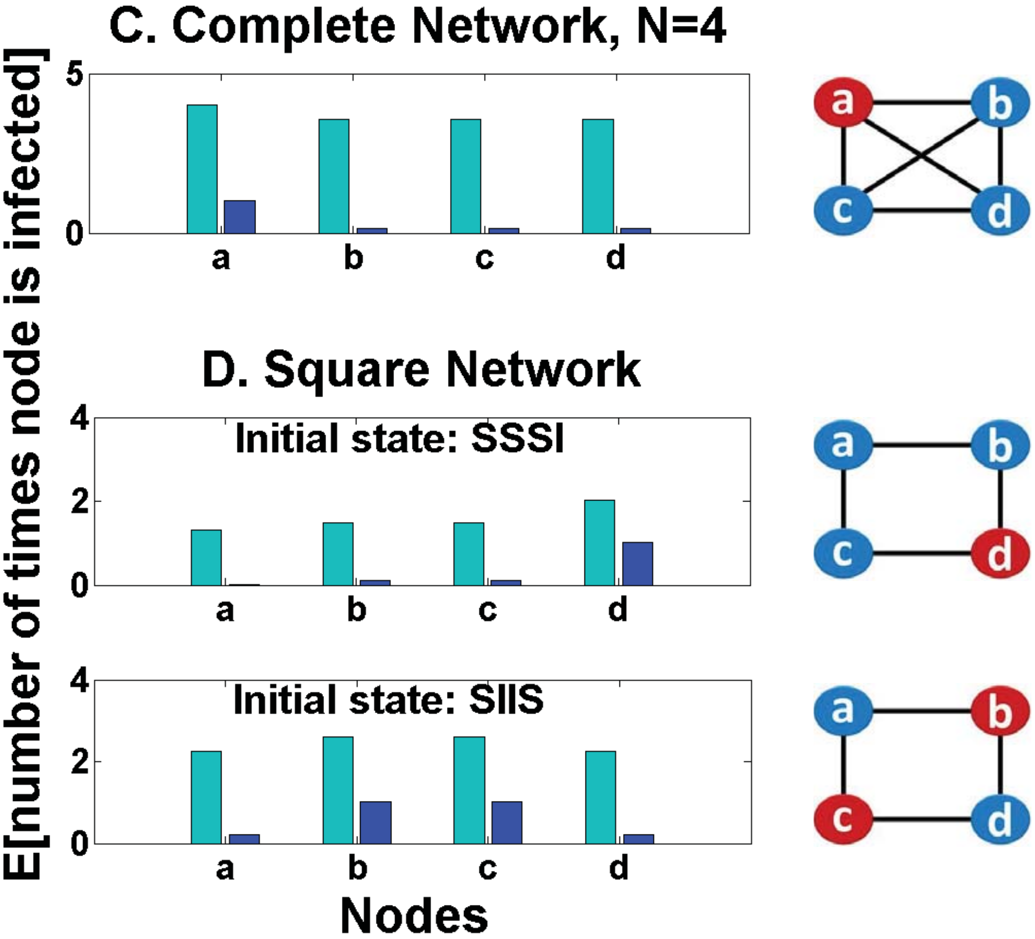


Figure 7.2: Expected number of times each node is infected during the epidemic for the complete and square networks with $N = 4$ nodes. The bars in light and dark blue are for rabies and chlamydia respectively. Schematic diagrams of the initial state are on the right.

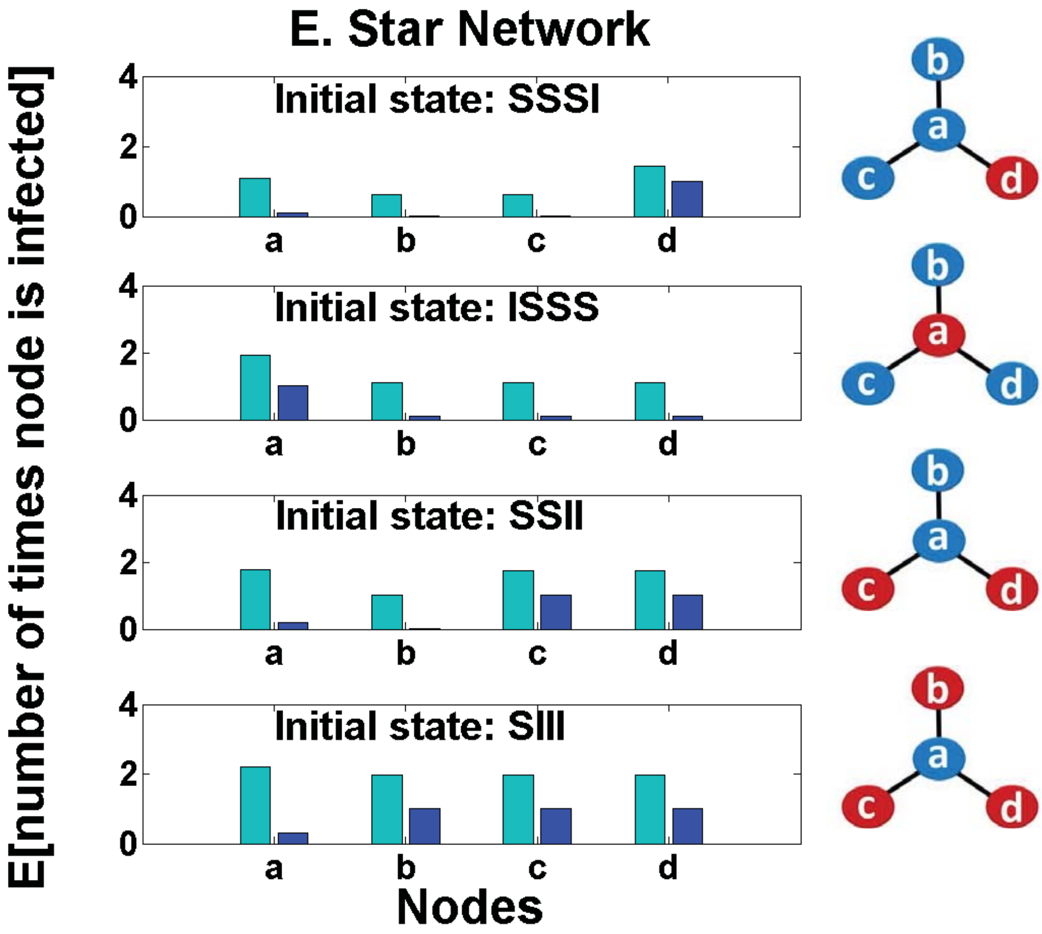


Figure 7.3: Expected number of times each node is infected during the epidemic for the star network with $N = 4$ nodes. The bars in light and dark blue are for rabies and chlamydia respectively. Schematic diagrams of the initial state are on the right.

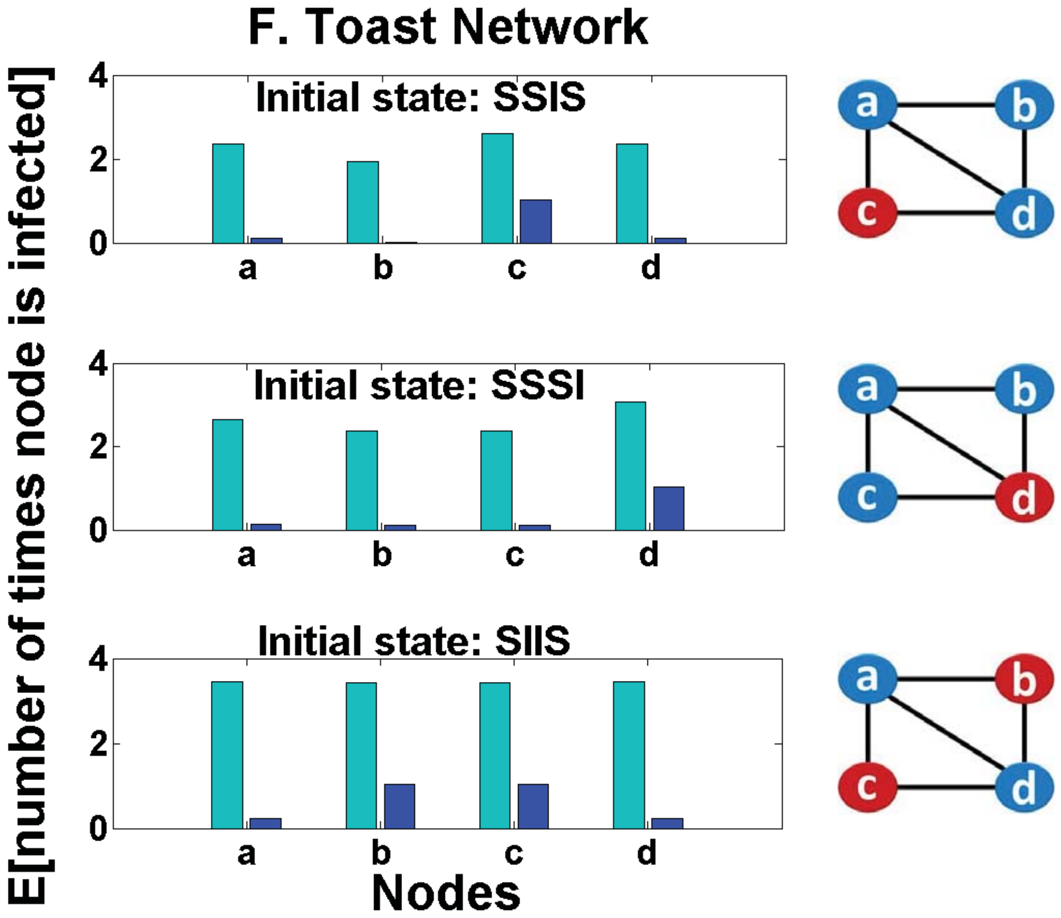


Figure 7.4: Expected number of times each node is infected during the epidemic for the toast network with $N = 4$ nodes. The bars in light and dark blue are for rabies and chlamydia respectively. Schematic diagrams of the initial state are on the right.

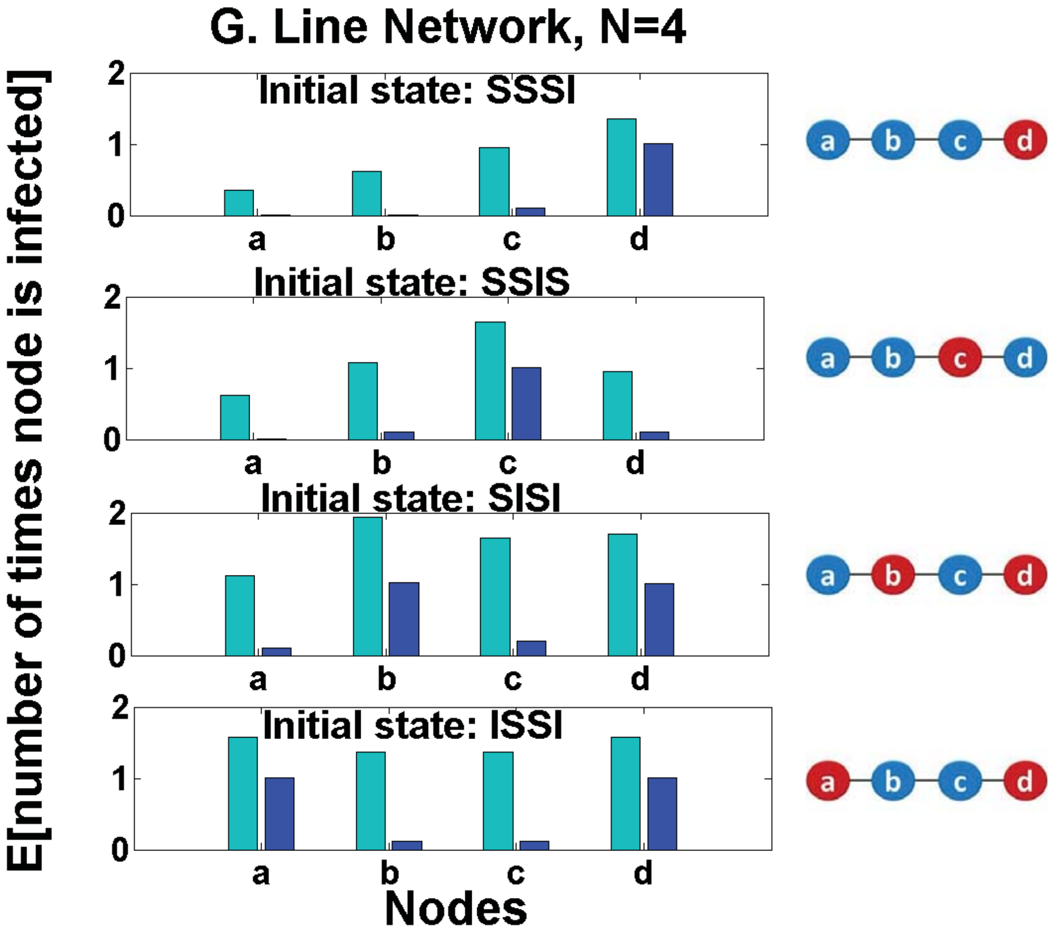


Figure 7.5: Expected number of times each node is infected during the epidemic for the line network with $N = 4$ nodes. The bars in light and dark blue are for rabies and chlamydia respectively. Schematic diagrams of the initial state are on the right.

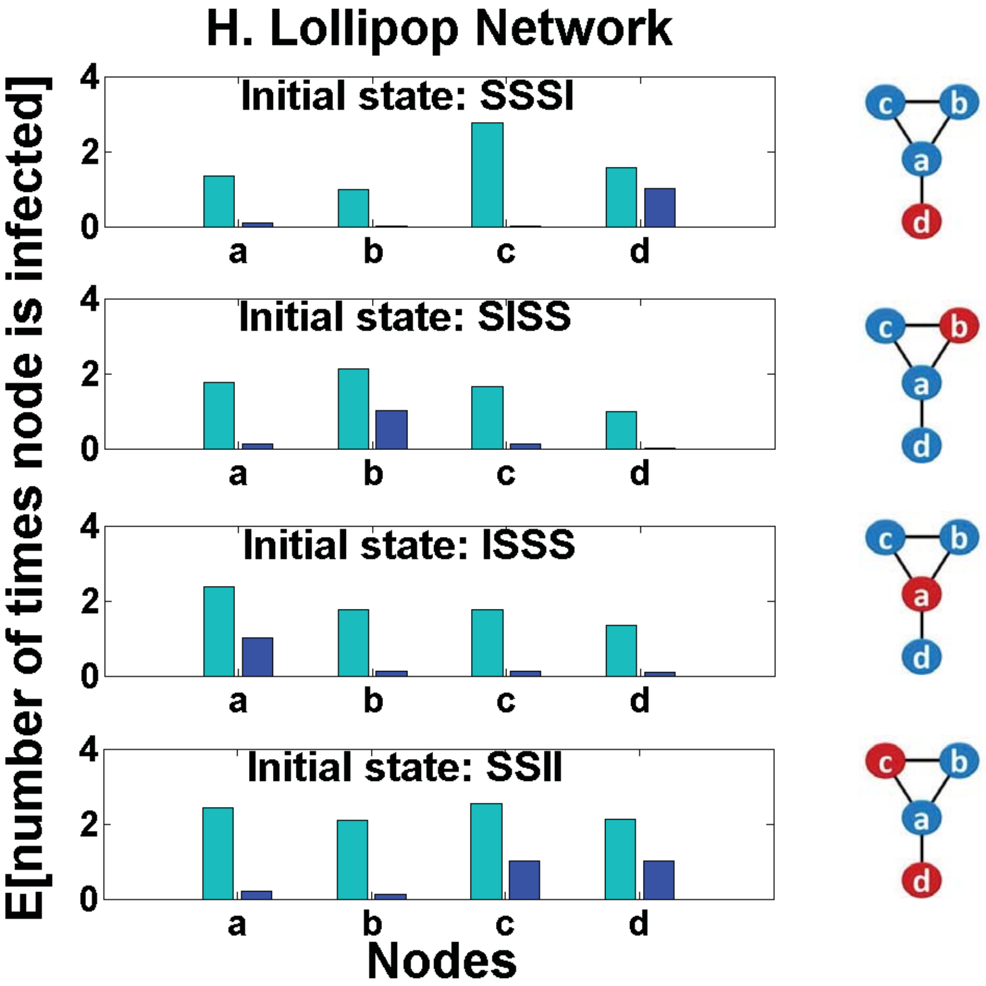


Figure 7.6: Expected number of times each node is infected during the epidemic for the lollipop network with $N = 4$ nodes. The bars in light and dark blue are for rabies and chlamydia respectively. Schematic diagrams of the initial state are on the right.

Chapter 8

Summary

We have derived exact analytical expressions for the probability mass functions (PMFs) of the final epidemic size on eight small networks and investigated the effect that network structure and the degree of the initial infectious node has on the spread of an infection. Increasing the complexity of the network structure reduced the effect of grouping states together based on symmetries of the network. Consequently this increased the complexity of the analysis which is evident in the probability mass function expressions and is due to the increasing number of possible infection paths. The results presented in Chapter 2 form the basis for finding tractable analytic results which describe the spread of an infection through large networks that are composed of the small networks given in Figures 2.1 and 5.1.

The clustering coefficient, ϕ, was determined for each of the small networks in order to see if there was any correlation between final size of the epidemic and how clustered the network is. The expected final size for the triangle network is higher than that for the line network of $N = 3$ nodes for all values of $\mathcal{R}$, which can be seen by comparing the expressions in Table 2.3. The expected final size for the complete network with $N = 4$ nodes is higher than the expected final size of every other small network of size $N = 4$ nodes. Small networks with a higher clustering coefficient have a higher expected final size when the epidemic is started with one infectious node. Here, we note that we simply observe how clustered each of the networks is to see if any patterns emerge in relation to the final epidemic size results. It should be understood that in order to make a more rigorous comparison between networks, ideally only one property (such as the degree distribution or clustering coefficient) would be varied at a time wherever possible. We also determined the value of $\mathcal{R}$ where the expected final size functions for different networks of size $N = 4$ intersect; this gave further insight into the effect of infection parameters on the final epidemic size. The probability that each node in a network will acquire infection at some point during the epidemic increases when the degree of the initial infectious node or the transmission parameter (β, and hence $\mathcal{R}$) increases for these small networks. Similarly, the probability that the infection would die out before infecting an initially susceptible node is higher when the degree of the initial infectious node is one, when compared with initial infectious nodes of degree two and three.

The results presented in Chapter 2 give us a good indication that both the network topology and degree of the initial infectious node are key factors in understanding how an infection might spread through small networks which could represent small populations or communities. Some of the results presented for the small networks are intuitive; however, we emphasize the importance of having found exact analytic expressions for the PMFs of the final epidemic size of the small networks. Having the PMF for the final epidemic size provides us with more detail than a single expression for the expected final size of an epidemic.

Furthermore, in Chapter 3 we have shown that it is possible to generalise the PMF of the final epidemic size of larger networks composed of the small networks studied in Chapter 2. The decomposition of the larger network into the correct assemblage of smaller networks to generalise the probability mass function of the final epidemic size is crucial. Connecting small networks together with overlapping edges becomes problematic, as discussed in Kiss et al [37], when trying to generalise results based on those already found for the small networks. However, we have shown that connecting small networks together with overlapping nodes is a feasible approach for the generalisation of some time independent results such as the final epidemic size which depends only upon the infection parameters and initial conditions. Based on this idea it will not be necessary to continue to construct the transition diagrams for each network.

More specifically, we have shown that it is possible to describe the way an infection spreads through a line of triangles (LoT) network based on our knowledge of the triangle and lollipop networks studied in Chapter 2. We have used this knowledge to generalise the infection paths and PMF for the final epidemic size of a LoT network for an arbitrary number of nodes, N. We have illustrated a method to find the probability mass functions for the final epidemic size for a LoT(N) network starting with node a, b or c as the initial infectious node, I_0. As a result we have found a way to obtain analytic results that describe the way an infection spreads through a LoT network.

In Chapter 4 we discussed the definition of the basic reproduction number, R_0, and provide an overview of methods used to derive R_0 for network models. This was followed by an illustration of how to parameterise the small network models from Chapter 2, which could be representative of a small population or community, for influenza and measles. We found that the networks with the two highest estimates for the basic reproduction number, R_0, for influenza are the complete and toast networks of size $N = 4$ respectively. For networks of size $N = 3$ and for a given probability of transmission given contact, the triangle network has an R_0 value greater than that for the line network. As R_0 depends on the average degree of the network we notice that the square and triangle networks have the same value for $R_0 = 0.83$. Overall, the line networks of size $N = 3$ and $N = 4$ had the lowest estimate of R_0.

Markov Chain theory was used to derive analytic results that describe the behaviour of an SIS model on small networks. Results such as the expected time to absorption, the expected number of times each individual is infected and the cumulative incidence were derived for all small networks given in Figures 2.1 and 5.1. We found that the triangle network had a longer time to absorption and higher cumulative incidence than the line network with $N = 3$ nodes for all initial conditions considered. The complete network, starting with one infectious node, had the longest expected time to absorption and highest cumulative incidence for all parameter values considered when compared to all other networks of size $N = 4$ across all initial states considered. This was followed by the toast network, however there was a large difference between the expected time to absorption and cumulative incidence when compared to the complete network.

When the initial number of infectious nodes is increased, the cumulative incidence and expected time to absorption increases, this was seen across all small networks. When starting with one infectious node, the degree of this node also affected the cumulative incidence and expected time to absorption. With an infectious node of higher degree, the cumulative incidence and expected time to absorption was increased. Our results give us a good indication that both the network topology and degree of the initial infectious node are key factors in understanding how an infection might spread through small networks that could represent small populations or communities. Some of the results presented here are intuitive; however, we emphasize the importance of being able to present results in symbolic form that are exact. Being able to do this allows us to gain more information about what the main driving factor of each result is (i.e transmission or recovery rate). We also note that the results found for the SIS model on small networks follow a similar pattern to what was found for the SIR model. That is, it was the same networks that had a higher cumulative incidence and time to absorption (for the SIS model) that had a higher expected final size (for the SIR model) when compared to the other small networks.

The dynamics of an SIS model on a LoT network of $N = 6$ and $N = 9$ nodes were investigated, however we were unable to describe the behaviour of an SIS

model on a LoT network in using results found from the small networks in Chapter 5. We derived an algorithm in Section 6.3 that, in theory, can be used for any network structure of any size to generate the transition probability matrix. Once we have the transition probability matrix, all other results that were presented for the SIS model on small networks can be derived for a network of arbitrary size, N, by following the methods outlined in Chapter 5.

We illustrated how to parameterise the small network models for two types of infections, rabies and chlamydia, that follow an SIS model for the host infection life cycle. The differences in infection parameters between rabies and chlamydia were emphasized in the results. The probability of transmission given contact for rabies and chlamydia was estimated to be 0.49 and 0.1 respectively. However, the infectious period for rabies is relatively short (approximately 4 days) compared to that of chlamydia (approximately 1 year). We found that it takes 1-2 years for the chlamydia infection to die out in the network and it takes approximately 6 - 18 days for rabies to die out. So while the rabies epidemic is much shorter than an epidemic for chlamydia, the cumulative incidence is much higher for rabies due to the higher probability of transmission given contact.

Despite the size of the SIS model system (2^N states) being smaller than the size for an SIR model (3^N states), analysing the infection dynamics is more complex. This added complexity is caused by the cyclic behaviour that an SIS model exhibits. Upon contact with an infectious individual, susceptible individuals move into the infectious class. When an infectious individual recovers they move back into the susceptible class. This makes multiple reinfections possible for each individual, furthermore the only stable steady state of the system is the one in which all individuals are susceptible. In contrast, for an SIR model each individual is infected at most once. This major difference in infection dynamics is the reason different approaches were required for analysing the two models.

One limitation of this study is the use of static networks when in many situations it would be more realistic to consider a dynamic network in which each individual's connections can change over time. Therefore, future work could include investigating the spread of an infection through small dynamic networks of three and four nodes with the intention of composing a larger dynamic network made up of the smaller networks. Another way this research could be extended would be to consider investigating how an infection spreads through small directed networks with the same topological structure as in Figures 2.1 and 5.1. There are many possibilities for extending the work presented in this thesis and we look forward to seeing how this area of research unfolds.

We have shown that it is possible to generalise some time-independent results for an SIR model such as the PMF of the final epidemic size of a larger network using smaller network results. While the LoT network is a specific network structure, future work could investigate how to generalise the PMF of the final epidemic size of a circle of triangles. The goal is to develop similar methods for generalising results for a wider class of networks, however we hope what we presented here is used as the foundation for improving research in this area.

The emphasis of this research is on finding tractable analytic expressions that describe how an infection might spread through a contact network. Numerical simulation methods are becoming increasingly appealing as the capacity of computers is improving and thus making large simulations more efficient to run. However, obtaining analytical results which can be compared to simulations could provide extra confidence that the model and parameters being used are meaningful and correct. Analytical results can provide further insight into the inputs and outputs of a given model. For example, sometimes we obtain results from simulations but it is impossible to determine how much a given parameter (or combination of parameters) is driving that result. These sorts of insights can be determined more easily from explicit analytical results. If we can understand which network properties impact the transmission of infection through a population then we can use this information to aid the planning and implementation of control strategies such as vaccination campaigns.

Bibliography

[1] Centre for Disease Control: Chlamydia Detailed Fact Sheet. `http://www.cdc.gov/std/chlamydia/stdfact-chlamydia-detailed.htm`. Accessed: 16-12-2015.

[2] C. Athaus, J. Heijne, S. Herzog, A. Roellin, and N. Low. Individual and population level effects of partner notification for chlamydia trachomatis. *PLoS ONE*, 7(12):e51438, 2012.

[3] F. Ball and D. Sirl et al. An SIR epidemic model on a population with random network and household structure, and several types of individuals. *Advances in Applied Probability*, 44(1):63–86, 2012.

[4] F. Ball and D. Sirl et al. Acquaintance vaccination in an epidemic on a random graph with specified degree distribution. *Journal of Applied Probability*, 50(4):1147–1168, 2013.

[5] F. Ball and P. Neal. Network epidemic models with two levels of mixing. *Mathematical Bioscience*, 212:69–87, 2008.

[6] F. Ball, D. Sirl, and P. Trapman. Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Advances in Applied Probability*, 41(3):765–796, 2009.

[7] F. Ball, D. Sirl, and P. Trapman. Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Mathematical Biosciences*, 224(2):53–73, 2010.

[8] S. Bansal, B. Grenfell, and L. Meyers. When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society Interface*, 4:879–891, 2007.

[9] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[10] D.J. Barnes and D. Chu. *Introduction to Modelling for Biosciences*. Springer, 2010.

[11] M. Biggerstaff, S. Cauchemez, C. Reed, M. Gambhir, and L. Finelli. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *PLoS ONE*, 6:e17835, 2011.

[12] P. Boelle, S. Ansart, A. Cori, and A. Valleron. Transmission parameters of the A/H1N1 (2009) influenza virus pandemic: a review. *Influenza and Other Respiratory Viruses*, 5:306–316, 2011.

[13] P.G. Coleman and C. Dye. Immunization coverage required to prevent outbreaks of dog rabies. *Vaccine*, 14:185–186, 1996.

[14] L. Danon, A. Ford, T. House, C. Jewell, M.J. Kelling, G. Roberts, J.V. Ross, and M. Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, 2011:1–28, 2011.

[15] L. Decreusefond, J.S. Dhersin, P. Moyal, and V. Tran. Large graph limit for an SIR process in a random network with heterogenous connectivity. *The Anals of Applied Probability*, 22(2):541–575, 2010.

[16] O. Diekmann, J.A.P Heesterbeek, and T. Britton. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, 2013.

[17] O. Diekmann, J.A.P Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:365–382, 1990.

[18] O. Diekmann, J.A.P Heesterbeek, and M.G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7:873–885, 2010.

[19] L. Edelstein-Keshet. *Mathematical Models in Biology*. Society for Industrial and Applied Mathematics, 2005.

[20] J. Gleeson. High-accuracy approximation of binary-state dynamics on networks. *Physical Review Letters*, 107:068701, 2011.

[21] C. Grinstead and J. Snell. *Introduction to Probability: Markov Chains*. Dartmouth College, 1997.

[22] K. Hampson, J. Dushoff, J. Bingham, G. Brúckner, Y.H. Ali, and A. Dobson. Synchronous cycles of domestic dog rabies in sub-Saharan Africa and the impact of control efforts. *PNAS Journal*, Vol 104, Issue 18:7717–7722, May 2007.

[23] K. Hampson, J. Dushoff, S. Cleaveland, D. T. Haydon, M. Kaare, C. Packer, and A. Dobson. Transmission dynamics and prospects for the elimination of canine rabies. *PLoS Biology*, 7:e1000053, 2009.

[24] J.A.P Heesterbeek and M.G. Roberts. The type-reproduction number T in models for infectious disease control. *Mathematical Biosciences*, 206:3–10, 2007.

[25] J. Heijne, C. Athaus, S. Herzog, M. Kretzschmar, and N. Low. The role of reinfection and partner notification in the efficacy of chlamydia screening programs. *Journal of Infectious Diseases*, 203:372–377, 2010.

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

[27] Q. Hou, Z. Jin, and S. Ruan. Dynamics of rabies epidemics and the impact of control efforts in Guangdong Province, China. *Journal of Theoretical Biology*, 300:39–47, 2012.

[28] T. House and M.J. Keeling. Epidemic prediction and control in clustered populations. *Journal of Theoretical Biology*, 272:1–7, 2010.

[29] S. Janson, M. Luczak, and P. Windridge. Law of large numbers for the SIR epidemic on a random graph with given degrees. *Random Structures and Algorithms*, 2014.

[30] A. Johnson, C. Mercer, B. Erens, A. Copas, S. McManus, K. Wellings, C. Korovessis K. Fenton, W. Macdowall, K. Nanchahal, S. Purdon, and J. Field. Sexual behaviour in britain: partnerships, practices, and HIV risk behaviours. *Lancet*, 358:1835–1842, 2001.

[31] B. Karrer and M. Newman. Message passing approach for general epidemic models. *Physical Review E*, 82(1):016101, 2010.

[32] M.J. Keeling. The effects of local spatial structures on epidemiological invasions. *Proceedings of The Royal Society London Series B*, 266:859–867, 1999.

[33] M.J. Keeling and K. Eames. Networks and epidemic models. *Journal of The Royal Society Interface*, 2:295–307, 2005.

[34] M.J. Keeling and P. Rohani. *Modelling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[35] M.J. Keeling and J. V. Ross. On methods for studying stochastic disease dynamics. *Journal of The Royal Society Interface*, 5:171–181, 2008.

[36] W.O. Kermack and A.G. McKendrick. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society London Series A*, 115:700–721, 1927.

[37] I.Z. Kiss, C. Morris, F. Selley, P. Simon, and R. Wilkinson. Exact deterministic representation of Markovian SIR epidemics on networks with and without loops. *Journal of Mathematical Biology*, 70:437–464, 2015.

[38] P.M. Kitala, J.J McDermott, P.G Coleman, and C. Dye. Comparison of vaccination strategies for the control of dog rabies in Machakos District, Kenya. *Epidemiology and Infection*, Vol 129, Issue 1:215–222, 2002.

[39] J. Lindquist, J. Ma, P. van den Driessche, and F. Willeboordse. Effective degree network disease models. *Journal of Mathematical Biology*, 62:143–164, 2011.

[40] A. Lloyd and S Valeika. *Network models in epidemiology: an overview. In: Complex population dynamics: nonlinear modeling in ecology, epidemiology and genetics*, pages 189–214. World Scientific, 2007.

[41] J Ma, P van den Driessche, and F. Willeboordse. The importance of contact network topology for the success of vaccination strategies. *Journal of Theoretical Biology*, 325:12–21, 2013.

[42] V. Marceau, P. Noel, L. Hebert-Dufresne, A. Allard, and L. Dube. Adaptive networks: Coevolution of disease and topology. *Physical Review E*, 82:1–9, 2010.

[43] J.C. Miller. Spread of infectious disease through clustered populations. *Journal of The Royal Society Interface*, 6:1121–1134, 2009.

[44] J.C. Miller. A note on a paper by Erik Volz: SIR dynamics in random networks. *Journal of Mathematical Biology*, 62:349–358, 2010.

[45] J.C. Miller and I.Z. Kiss. Epidemic spread in networks: Existing methods and current challenges. *Mathematical Modelling of Natural Phenomena*, 9:4–42, 2014.

[46] J.C. Miller, A. Slim, and E. Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of the Royal Society Interface*, 9:890–906, 2012.

[47] J.C. Miller and E. Volz. Incorporating disease and population structure into models of SIR disease in contact networks. *PLoS ONE*, 8:e69162, 2013.

[48] C. Molina and L. Stone. Modelling the spread of diseases in clustered networks. *Journal of Theoretical Biology*, 315:110–118, 2012.

[49] G. Neumann, T. Noda, and Y. Kawaoka. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature*, 459:931–939, 2009.

[50] M. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66:016128, 2002.

[51] M. Newman. *Networks: An Introduction.* Oxford University Press, 2010.

[52] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87:925–979, 2015.

[53] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63:066117:1–8, 2001.

[54] L. Pellis, F. Ball, S. Bansal, K. Eames, T. House, V. Isham, and P. Trapman. Eight challenges for network epidemic models. *Epidemics*, 10:58–62, 2015.

[55] M. Ritchie, L. Berthouze, T. House, and I.Z Kiss. Higher-order structure and epidemic dynamics in clustered networks. *Journal of Theoretical Biology*, 348:21–32, 2014.

[56] M.G. Roberts and J.A.P Heesterbeek. A new method for estimating the effort required to control an infectious disease. *Proceedings of the Royal Society London Series B*, 270:1359–1364, 2003.

[57] M.G. Roberts and J.A.P Heesterbeek. Characterizing the next-generation matrix and basic reproduction number in ecological epidemiology. *Mathematical Biology*, 66:1045–1064, 2013.

[58] M.G. Roberts and H. Nishiura. Early estimation of the reproduction number in the prescence of imported cases: Pandemic influenza H1N1-2009 in New Zealand. *PLoS ONE*, 6:e17835, 2011.

[59] M.G. Roberts and M. Tobias. Predicting and preventing measles epidemics in New Zealand: application of a mathematical model. *Epidemiology and Infection*, 124:279–287, 2000.

[60] Sheldon M. Ross. *Introduction to Probability Models*. Elsevier Science, 2007.

[61] K.J. Sharkey, I.Z. Kiss, R.R. Wilkinson, and P.L. Simon. Exact equations for SIR epidemics on tree graphs. *Bulletin of Mathematical Biology*, pages 1–32, 2013.

[62] P. Simon, M. Taylor, and I.Z. Kiss. Exact epidemic models on graphs using graph-automorphism driven lumping. *Journal of Mathematical Biology*, 62:479–508, 2010.

[63] M. Taylor, T. Taylor, and I.Z. Kiss. Epidemic threshold and control in a dynamic network. *Physical Review E*, 85:0161031, 2012.

[64] A. Tuite, A. Greer, M. Whelan, A. Winter, B. Lee, P. Yan, J. Wu, S. Moghadas, D. Buckeridge, B. Pourbohloul, and D. Fisman. Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. *Canadian Medical Association Journal*, 182:131–136, 2010.

[65] E. Volz. SIR dynamics in random networks with heterogenous connectivity. *Journal of Mathematical Biology*, 56:293–310, 2008.

[66] D. Watts and S. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442, 1998.

[67] G. Witten and G. Poulter. Simulations of infectious diseases on networks. *Computers in Biology and Medicine*, 37:195–205, 2007.

[68] M. Youssef and C. Scoglio. An individual-based approach to SIR epidemics in contact networks. *Journal of Theoretical Biology*, 283:136–144, 2011.

[69] J. Zhang, Z. Jin, G. Sun, T. Zhou, and S. Ruan. Analysis of rabies in China: Transmission dynamics and control. *PLoS One*, 6(7):e20891, June 2011.

[70] Y. Zhao and M.G. Roberts. Simulating epidemics on networks. *Research Letters in Information and Mathematical Sciences*, 6:101–103, 2007.

[71] J. Zinsstag, S. Durr, M.A. Penny, R. Mindekem, F. Roth, S. Menendez Gonzalez, S. Naissengar, and J. Hattendorf. Transmission dynamics and economics of rabies control in dogs and humans in an african city. *PNAS*, 106(35):14996–15001, September 2009.

Appendix A

A.1 Line Network, $N = 3$

For a line network with three nodes we have two nodes of degree 1 (a and c) and one node of degree 2 (b) (See also Figure 2.2). For an SIR model on a line network of three nodes there are 27 possible states in which the network can be. By grouping the appropriate states together, the line network can be reduced to 18 sets of states shown by the transition diagram in Figure A.1. In the line network we have three different initial states; two initial states have one infectious node (SSI and SIS) and one initial state has two infectious nodes (ISI). Initial states cannot be reached from any other state in the transition diagram. In states SSI and SIS, the infectious node has degree 1 and 2 respectively.

Each box in the transition diagram represents a different state of the network and each arrow represents either recovery or infection of a node. Transition between states is only possible if there is a directed arrow from one to the other. Movement between states in the transition diagram shows how the infection spreads through the network. Once the network reaches an absorbing state the epidemic is over and the infection has died out.

A.1.1 Catalogue of transition probabilities

From the transition diagram of the SIR model on the line network we derive the individual transition probabilities between network states. In the following $\mathcal{P}_{XYZ}$ denotes the probability that the network is ever in state XYZ. These probabilities

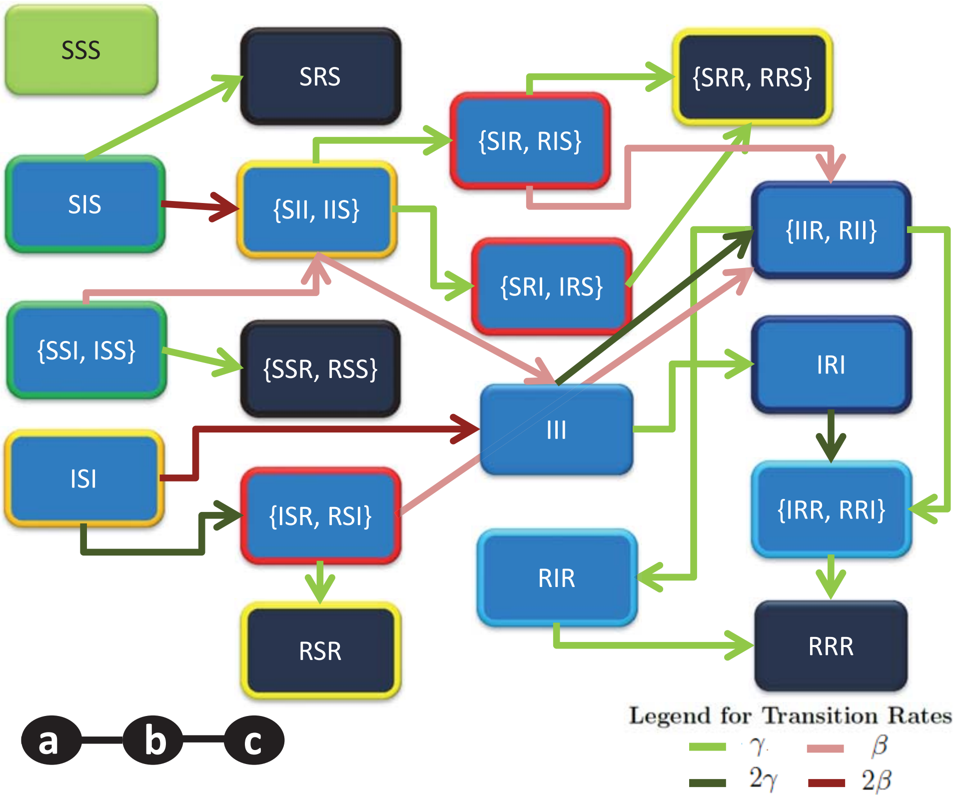


Figure A.1: Transition diagram showing how an infection can spread through the line network. Possible initial states SIS, SSI and ISI are shown on the left hand side. The absorbing states for this process are coloured in black. Coloured outlines identify states with the same number of nodes in each infection state but different network configuration.

are independent of time and depend only on the infection parameters. From the absorbing state probabilities we find the final epidemic size probabilities, $\mathbb{P}$.

Possible initial state indicator variables:

$$\mathrm{E}_{SIS} = \begin{cases} 1, & \text{if initial state is } SIS. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{SSI} = \begin{cases} 1, & \text{if initial state is } SSI. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{ISI} = \begin{cases} 1, & \text{if initial state is } ISI. \\ 0, & \text{otherwise.} \end{cases}$$

Probability of passing through transient states:

$$\begin{aligned}
\mathcal{P}_{SII} &= \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{SIS} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathrm{E}_{SSI} \\
\mathcal{P}_{ISR} &= \frac{2}{2\mathcal{R}+2}\mathrm{E}_{ISI} \\
\mathcal{P}_{SIR} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{SII} \\
\mathcal{P}_{SRI} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{SII} \\
\mathcal{P}_{III} &= \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{SII} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathrm{E}_{ISI} \\
\mathcal{P}_{IIR} &= \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SIR} + \frac{2}{3}\mathcal{P}_{III} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{ISR} \\
\mathcal{P}_{IRI} &= \frac{1}{3}\mathcal{P}_{III} \\
\mathcal{P}_{RIR} &= \frac{1}{2}\mathcal{P}_{IIR} \\
\mathcal{P}_{IRR} &= \mathcal{P}_{IRI} + \frac{1}{2}\mathcal{P}_{IIR}
\end{aligned}$$

Probability of terminating in absorbing states:

$$\begin{aligned}
\mathcal{P}_{SRS} &= \frac{1}{2\mathcal{R}+1}\mathrm{E}_{SIS} \\
\mathcal{P}_{SSR} &= \frac{1}{\mathcal{R}+1}\mathrm{E}_{SSI} \\
\mathcal{P}_{RSR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{ISR} \\
\mathcal{P}_{SRR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SIR} + \mathcal{P}_{SRI} \\
\mathcal{P}_{RRR} &= \mathcal{P}_{RIR} + \mathcal{P}_{IRR}
\end{aligned}$$

To find the equations for the final size probabilities we evaluated the following:

$$\begin{aligned}
\mathbb{P}(\text{Final Size} = 1) &= \mathcal{P}_{SRS} + \mathcal{P}_{SSR} \\
\mathbb{P}(\text{Final Size} = 2) &= \mathcal{P}_{RSR} + \mathcal{P}_{SRR} \\
\mathbb{P}(\text{Final Size} = 3) &= \mathcal{P}_{RRR}
\end{aligned}$$

Simplifying the above we derive three sets of analytic expressions for the final size probabilities (see Table A.1) which reflect the different initial conditions possible for the line network. For example, working backwards to illustrate how we found $\mathbb{P}(\text{Final Size} = 2)$, we make the following substitutions:

$$\begin{aligned}
\mathbb{P}(\text{Final Size} = 2) &= \mathcal{P}_{RSR} + \mathcal{P}_{SRR} \\
&= \frac{1}{\mathcal{R}+1}\mathcal{P}_{ISR} + \frac{1}{\mathcal{R}+1}\mathcal{P}_{SIR} + \mathcal{P}_{SRI} \\
&= \frac{2}{(\mathcal{R}+1)(2\mathcal{R}+2)}\mathrm{E}_{ISI} + \frac{1}{(\mathcal{R}+1)(\mathcal{R}+2)}\mathcal{P}_{SII} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{SII} \\
&= \frac{1}{(\mathcal{R}+1)^2}\mathrm{E}_{ISI} + \frac{2\mathcal{R}(\mathcal{R}+2)}{(\mathcal{R}+1)(\mathcal{R}+2)(2\mathcal{R}+1)}\mathrm{E}_{SIS} \\
&\quad + \frac{\mathcal{R}(\mathcal{R}+2)}{(\mathcal{R}+1)^2(\mathcal{R}+2)}\mathrm{E}_{SSI} \\
\therefore \mathbb{P}(\text{Final Size} = 2) &= \frac{1}{(\mathcal{R}+1)^2}\mathrm{E}_{ISI} + \frac{2\mathcal{R}}{(\mathcal{R}+1)(2\mathcal{R}+1)}\mathrm{E}_{SIS} + \frac{\mathcal{R}}{(\mathcal{R}+1)^2}\mathrm{E}_{SSI}
\end{aligned}$$

The probability mass function of the final size distribution for the line network is shown in Figure 2.9. We independently verified our final size calculations using a stochastic SIR model.

A.1.2 Progression of infection over time

In the following we use P_{XYZ} to denote the probability that the line network is in the state XYZ at time t. Thus, the equation for the time derivative $\dot{\mathsf{P}}_{XYZ}$ shows how the network can enter and leave the state XYZ. The rate the network enters and leaves each state can be found from the transition diagram. These equations allow us to simulate the time course of the epidemic and to check our final size calculations. Equations describing the probability that the network is in a given state at time t for an SIR model on the line network are as follows.

Table A.1: Final Size PMFs

Initial State	SSI	SIS	ISI
$\mathbb{P}$**(Final Size =1)**	$\frac{1}{\mathcal{R}+1}$	$\frac{1}{2\mathcal{R}+1}$	0
$\mathbb{P}$**(Final Size =2)**	$\frac{\mathcal{R}}{(\mathcal{R}+1)^2}$	$\frac{2\mathcal{R}}{(\mathcal{R}+1)(2\mathcal{R}+1)}$	$\frac{1}{(\mathcal{R}+1)^2}$
$\mathbb{P}$**(Final Size =3)**	$\frac{\mathcal{R}^2}{(\mathcal{R}+1)^2}$	$\frac{2\mathcal{R}^2}{(\mathcal{R}+1)(2\mathcal{R}+1)}$	$\frac{\mathcal{R}(\mathcal{R}+2)}{(\mathcal{R}+1)^2}$
Expected FS	$\frac{3\mathcal{R}^2+3\mathcal{R}+1}{(\mathcal{R}+1)^2}$	$\frac{3\mathcal{R}+1}{\mathcal{R}+1}$	$\frac{\left(3\mathcal{R}^2+6\mathcal{R}+2\right)}{(\mathcal{R}+1)^2}$

$$
\begin{aligned}
\text{Initial states:} & & \\
\dot{\mathsf{P}}_{SSS} &=& 0 \\
\dot{\mathsf{P}}_{SIS} &=& -(2\mathcal{R}+1)\mathsf{P}_{SIS} \\
\dot{\mathsf{P}}_{SSI} &=& -(\mathcal{R}+1)\mathsf{P}_{SSI} \\
\dot{\mathsf{P}}_{ISI} &=& -2(\mathcal{R}+1)\mathsf{P}_{ISI} \\
\text{Transient states:} & & \\
\dot{\mathsf{P}}_{SII} &=& 2\mathcal{R}\mathsf{P}_{SIS}+\mathcal{R}\mathsf{P}_{SSI}-(\mathcal{R}+2)\mathsf{P}_{SII} \\
\dot{\mathsf{P}}_{ISR} &=& 2\mathsf{P}_{ISI}-(\mathcal{R}+1)\mathsf{P}_{ISR} \\
\dot{\mathsf{P}}_{III} &=& \mathcal{R}\mathsf{P}_{SII}+2\mathcal{R}\mathsf{P}_{ISI}-3\mathsf{P}_{III} \\
\dot{\mathsf{P}}_{SIR} &=& \mathsf{P}_{SII}-(\mathcal{R}+1)\mathsf{P}_{SIR} \\
\dot{\mathsf{P}}_{SRI} &=& \mathsf{P}_{SII}-\mathsf{P}_{SRI} \\
\dot{\mathsf{P}}_{IIR} &=& \mathcal{R}\mathsf{P}_{SIR}+\mathcal{R}\mathsf{P}_{ISR}+2\mathsf{P}_{III}-2\mathsf{P}_{IIR} \\
\dot{\mathsf{P}}_{IRI} &=& \mathsf{P}_{III}-2\mathsf{P}_{IRI} \\
\dot{\mathsf{P}}_{RIR} &=& \mathsf{P}_{IIR}-\mathsf{P}_{RIR} \\
\dot{\mathsf{P}}_{IRR} &=& \mathsf{P}_{IIR}+2\mathsf{P}_{IRI}-\mathsf{P}_{IRR} \\
\text{Absorbing states:} & & \\
\dot{\mathsf{P}}_{SSR} &=& \mathsf{P}_{SSI} \\
\dot{\mathsf{P}}_{SRS} &=& \mathsf{P}_{SIS} \\
\dot{\mathsf{P}}_{RSR} &=& \mathsf{P}_{ISR} \\
\dot{\mathsf{P}}_{SRR} &=& \mathsf{P}_{SIR}+\mathsf{P}_{SRI} \\
\dot{\mathsf{P}}_{RRR} &=& \mathsf{P}_{IRR}+\mathsf{P}_{RIR}
\end{aligned}
$$

We have included the equation for the initial state SSS for completeness, even though it is disjoint from the transition diagram as no infection is present. As $t \to \infty$, the infection will die out and the system will end up in one of the absorbing states. For a given transition rate, the initial state determines the probability of tending to each of the possible absorbing states. To find the final size probabilities we numerically solve the system of differential equations with specified initial conditions and sum up the appropriate absorbing state probabilities once a steady state has been reached. See Figure A.2 for numerical results that confirm our analytical expressions for the final epidemic size.

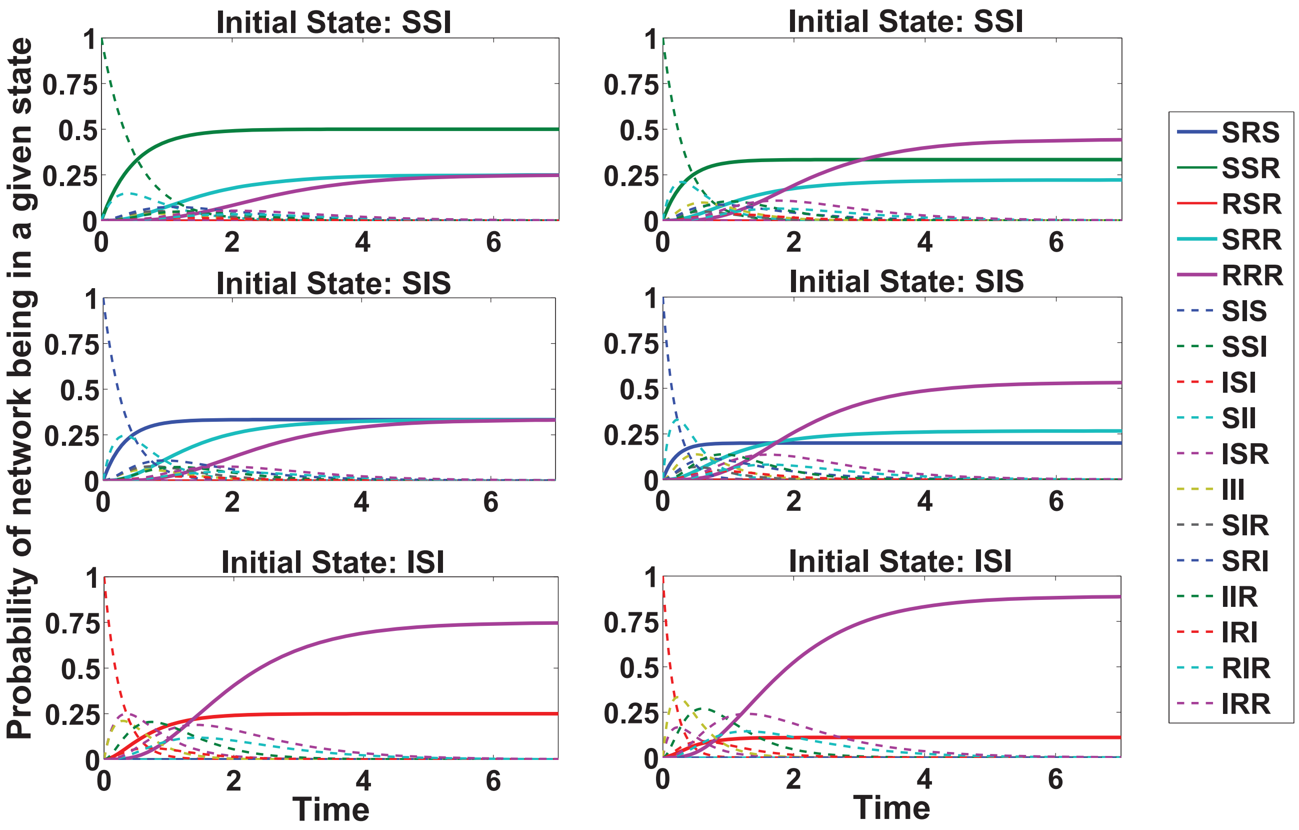


Figure A.2: Numerical solution of the system of differential equations described above for the progression of infection over time for an SIR model on a line network with $N = 3$ nodes. Left and right columns contain graphical results for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively for the specified initial conditions. The numerical results are in agreement with the analytical expressions for the same set of initial conditions.

A.1.3 Comparison of expected final size for the triangle and line networks of size $N = 3$.

Here we show that the expected final size for the triangle network is always higher than the expected final size of the line network for all parameter values, starting with one initial infectious node. The expected final size is defined by $\mathbb{E}[\text{Final Size}] = \sum_{i=1}^{N} i\mathbb{P}(\text{Final Size} = i)$. For the triangle and line networks we have:

$$\mathbb{E}[\text{Final Size of Triangle}] = \frac{6\,\mathcal{R}^3 + 13\,\mathcal{R}^2 + 6\,\mathcal{R} + 1}{(2\,\mathcal{R}+1)\,(\mathcal{R}+1)^2} \tag{A.1}$$

$$\mathbb{E}[\text{Final Size of Line, IC} = SSI] = \frac{3\,\mathcal{R}^2 + 3\,\mathcal{R} + 1}{(\mathcal{R}+1)^2} \tag{A.2}$$

$$\mathbb{E}[\text{Final Size of Line, IC} = SIS] = \frac{3\,\mathcal{R} + 1}{\mathcal{R}+1} \tag{A.3}$$

It can be shown that $\mathbb{E}[\text{Final Size of Triangle}] > \mathbb{E}[\text{Final Size of Line, IC} = SSI]$ iff

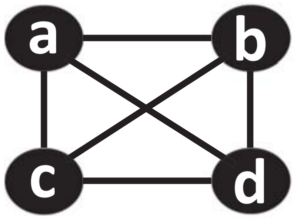
$$\mathcal{R}(4\mathcal{R}+1) > 0$$

which is true as $\mathcal{R} = \beta/\gamma$ is always positive.
Similarly, it can be shown that $\mathbb{E}[\text{Final Size of Triangle}] > \mathbb{E}[\text{Final Size of Line, IC} = SIS]$ iff

$$2\mathcal{R}^2 > 0$$

which is always true. Thus, when the epidemic is started with one infectious node, the expected final size of the triangle network is always larger than the expected final size of the line network.

A.2 Complete Network



We now examine networks of four nodes, again starting with the simplest case which is the complete network. For an SIR model on a network of $N = 4$ nodes there are 81 possible states in which the network can be. As all nodes in this network are topologically equivalent, we reduce the size of the system by grouping states together based on the number of nodes in each infection state. This results in 15 different states shown in the transition diagram in Figure A.3. We consider one initial condition for the SIR epidemic on a complete network where one node (either a, b, c or d) is the initial infectious node. The final size probabilities for the complete network of 4 nodes are shown in Table A.2.

Table A.2: Complete network final size PMFs

Initial State	$SSSI$
$\mathbb{P}$(**Final Size =1**)	$\frac{1}{3\,\mathcal{R}+1}$
$\mathbb{P}$(**Final Size =2**)	$\frac{3\,\mathcal{R}}{(2\,\mathcal{R}+1)^2\,(3\,\mathcal{R}+1)}$
$\mathbb{P}$(**Final Size =3**)	$\frac{6\mathcal{R}^2\,(3\mathcal{R}+2)}{(\mathcal{R}+1)^3\,(2\mathcal{R}+1)^2\,(3\mathcal{R}+1)}$
$\mathbb{P}$(**Final Size =4**)	$\frac{6\,\mathcal{R}^3\,\left(2\,\mathcal{R}^3+8\,\mathcal{R}^2+12\,\mathcal{R}+5\right)}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)^2\,(3\,\mathcal{R}+1)}$
Expected FS	$\frac{48\,\mathcal{R}^6+196\,\mathcal{R}^5+310\,\mathcal{R}^4+217\,\mathcal{R}^3+73\,\mathcal{R}^2+13\,\mathcal{R}+1}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)^2\,(3\,\mathcal{R}+1)}$

A.2.1 Catalogue of transition probabilities

In the following $\mathcal{P}_{WXYZ}$ denotes the probability that the network is ever in state $WXYZ$, where W, X, Y and Z denote the infection state (S, I or R) that nodes a, b, c and d are in respectively.

Possible initial state:

$$\mathrm{E}_{SSSI} = 1$$

Probability of passing through transient states:

$$\begin{aligned}
\mathcal{P}_{SSII} &= \frac{3\mathcal{R}}{3\mathcal{R}+1}\mathrm{E}_{SSSI} \\
\mathcal{P}_{SSIR} &= \frac{2}{4\mathcal{R}+2}\mathcal{P}_{SSII} \\
\mathcal{P}_{SIII} &= \frac{4\mathcal{R}}{4\mathcal{R}+2}\mathcal{P}_{SSII} \\
\mathcal{P}_{SIIR} &= \frac{3}{3\mathcal{R}+3}\mathcal{P}_{SIII} + \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{SSIR} \\
\mathcal{P}_{SIRR} &= \frac{2}{2\mathcal{R}+2}\mathcal{P}_{SIIR} \\
\mathcal{P}_{IIII} &= \frac{3\mathcal{R}}{3\mathcal{R}+3}\mathcal{P}_{SIII}
\end{aligned}$$

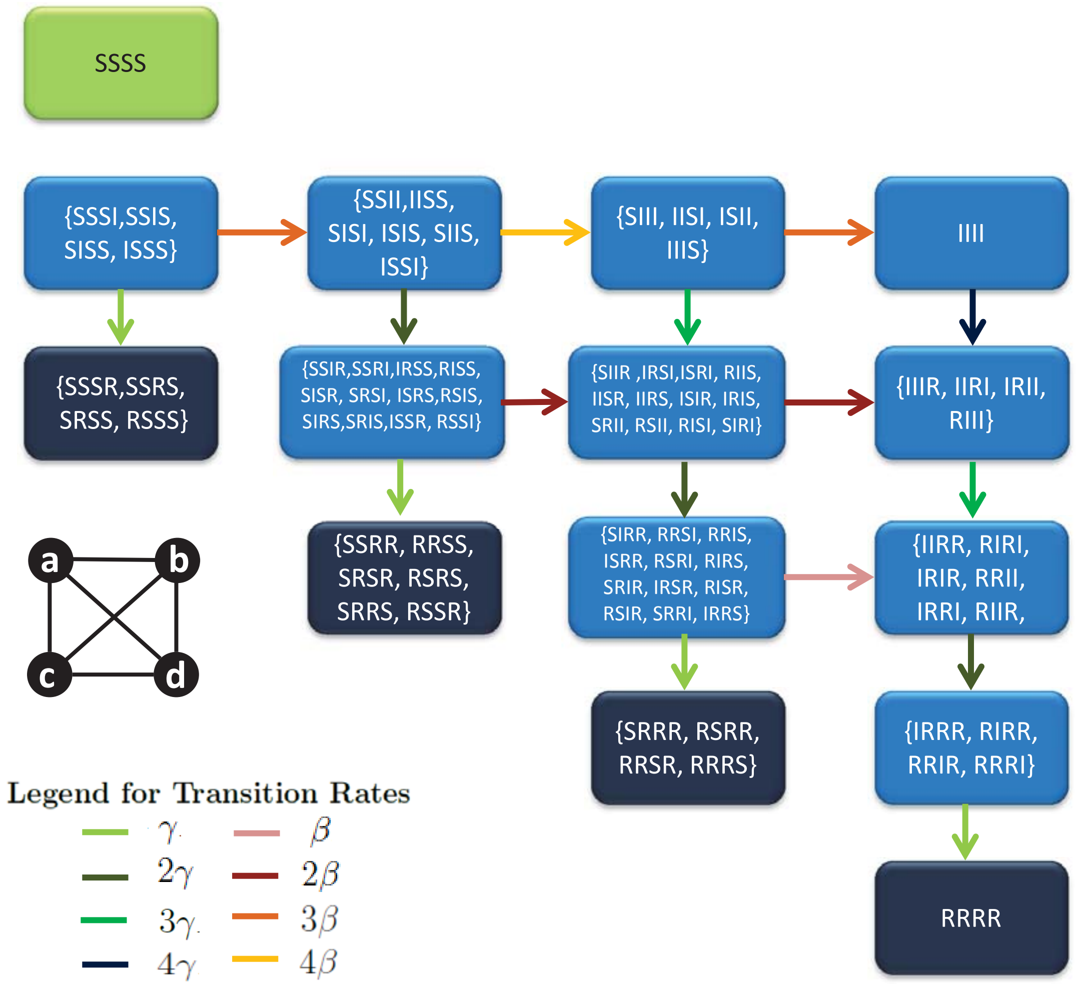


Figure A.3: Transition diagram for the complete network with $N = 4$ nodes.

$$\begin{aligned}\mathcal{P}_{IIIR} &= \mathcal{P}_{IIII} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{SIIR}\\ \mathcal{P}_{IIRR} &= \mathcal{P}_{IIIR} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SIRR}\\ \mathcal{P}_{IRRR} &= \mathcal{P}_{IIRR}\end{aligned}$$

Probability of terminating in absorbing states:

$$\begin{aligned}\mathcal{P}_{SSSR} &= \frac{1}{3\mathcal{R}+1}\mathrm{E}_{SSSI}\\ \mathcal{P}_{SSRR} &= \frac{1}{2\mathcal{R}+1}\mathcal{P}_{SSIR}\\ \mathcal{P}_{SRRR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SIRR}\\ \mathcal{P}_{RRRR} &= \mathcal{P}_{IRRR}\end{aligned}$$

To find the equations for the final size probabilities we evaluated the following:

$$\begin{aligned}\mathbb{P}(\text{Final Size} = 1) &= \mathcal{P}_{SSSR}\\ \mathbb{P}(\text{Final Size} = 2) &= \mathcal{P}_{SSRR}\\ \mathbb{P}(\text{Final Size} = 3) &= \mathcal{P}_{SRRR}\\ \mathbb{P}(\text{Final Size} = 4) &= \mathcal{P}_{RRRR}\end{aligned}$$

Simplifying the above we obtained the final size equations for the complete network with four nodes as shown in Table A.2. The probability mass function of the final size distribution for the complete network is shown in Figure 2.10.

A.2.2 Progression of infection over time

In the following we use P_{WXYZ} to denote the probability that the complete network is in the state $WXYZ$ at time t, where W, X, Y and Z denote the infection state (S, I or R) that nodes a, b, c and d are in respectively. Thus, the equation for the time derivative $\dot{\mathsf{P}}_{WXYZ}$ shows how the network can enter and leave the state $WXYZ$.

Equations describing the probability that the network is in a given state at time t for an SIR model on the complete network of $N = 4$ are:

Initial states:

$$\begin{aligned}\dot{\mathsf{P}}_{SSSS} &= 0\\ \dot{\mathsf{P}}_{SSSI} &= -(3\mathcal{R}+1)\mathsf{P}_{SSSI}\end{aligned}$$

Transient states:

$$\begin{aligned}
\dot{\mathsf{P}}_{SSII} &= 3\mathcal{R}\mathsf{P}_{SSSI} - (4\mathcal{R}+2)\mathsf{P}_{SSII}\\
\dot{\mathsf{P}}_{SSIR} &= 2\mathsf{P}_{SSII} - (2\mathcal{R}+1)\mathsf{P}_{SSIR}\\
\dot{\mathsf{P}}_{SIII} &= 4\mathcal{R}\mathsf{P}_{SSII} - 3(\mathcal{R}+1)\mathsf{P}_{SIII}\\
\dot{\mathsf{P}}_{SIIR} &= 3\mathsf{P}_{SIII} + 2\mathcal{R}\mathsf{P}_{SSIR} - 2(\mathcal{R}+1)\mathsf{P}_{SIIR}\\
\dot{\mathsf{P}}_{SIRR} &= 2\mathsf{P}_{SIIR} - (\mathcal{R}+1)\mathsf{P}_{SIRR}\\
\dot{\mathsf{P}}_{IIII} &= 3\mathcal{R}\mathsf{P}_{SIII} - 4\mathsf{P}_{IIII}\\
\dot{\mathsf{P}}_{IIIR} &= 2\mathcal{R}\mathsf{P}_{SIIR} + 4\mathsf{P}_{IIII} - 3\mathsf{P}_{IIIR}\\
\dot{\mathsf{P}}_{IIRR} &= \mathcal{R}\mathsf{P}_{SIRR} + 3\mathsf{P}_{IIIR} - 2\mathsf{P}_{IIRR}\\
\dot{\mathsf{P}}_{IRRR} &= 2\mathsf{P}_{IIRR} - \mathsf{P}_{IRRR}
\end{aligned}$$

Absorbing states:

$$\begin{aligned}
\dot{\mathsf{P}}_{SSSR} &= \mathsf{P}_{SSSI}\\
\dot{\mathsf{P}}_{SSRR} &= \mathsf{P}_{SSIR}\\
\dot{\mathsf{P}}_{SRRR} &= \mathsf{P}_{SIRR}\\
\dot{\mathsf{P}}_{RRRR} &= \mathsf{P}_{IRRR}
\end{aligned}$$

To find the final size probabilities we numerically solve the system of differential equations with specified initial conditions and sum up the appropriate absorbing state probabilities once a steady state has been reached. See Figure A.4 for numerical results that confirm our analytical expressions for the final epidemic size.

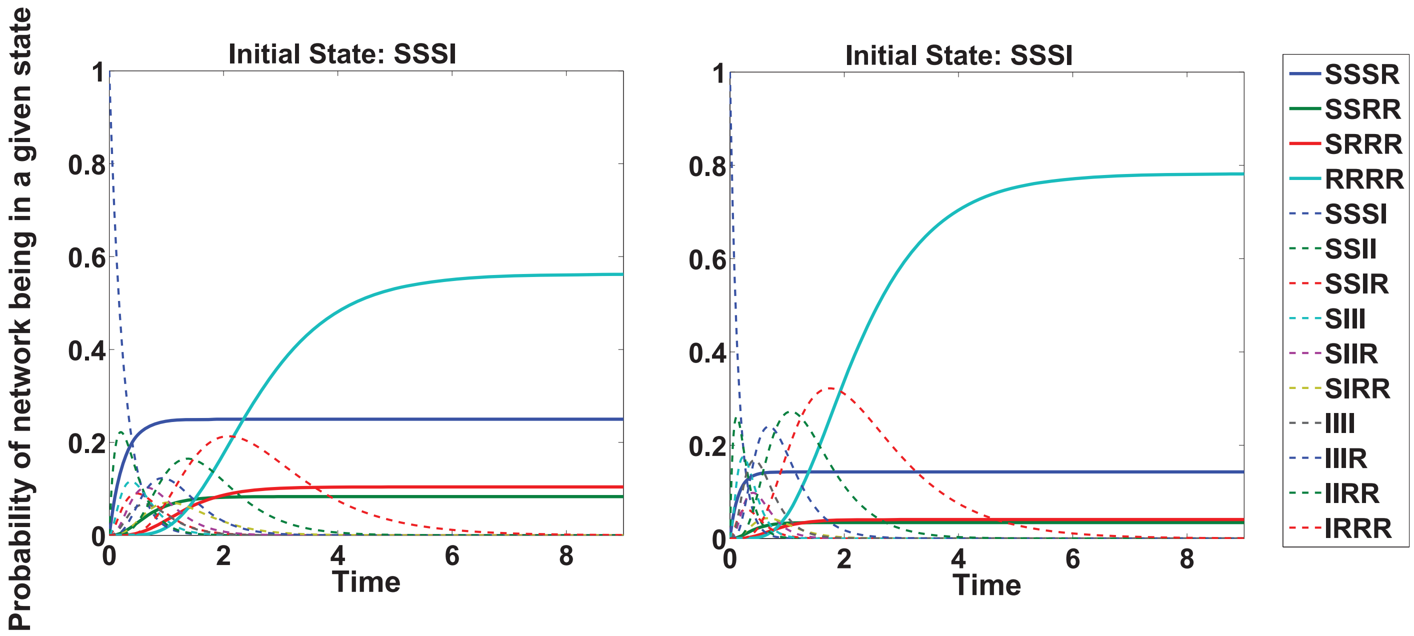
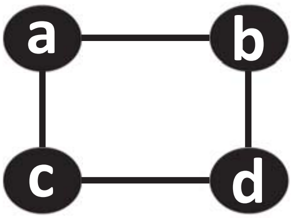


Figure A.4: Numerical solution of the system of differential equations described above for the progression of infection over time for an SIR model on a complete network with $N = 4$ nodes. Left and right columns contain graphical results for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively for the specified initial conditions. The numerical results are in agreement with the analytical expressions for the same set of initial conditions.

A.3 Square Network



The square network is a regular network in which each node has degree 2. As all nodes are topologically equivalent if we start the epidemic with one initial infectious node it can be either node a, b, c or d. The first initial state, $SSSI$, represents starting the epidemic with one infectious node. The second initial state, $SIIS$, represents starting the epidemic with two infectious nodes that are not neighbours; that is nodes a and d or nodes b and c are the initial infectious nodes (see Figure A.5). We decided to look into the case of starting with two initial infectious nodes (second initial state) within a network as it is not unlikely that two individuals within a population acquire an infection independently of each other and from which an epidemic may occur. We illustrate the probability mass functions of the final size for the square network with two initial conditions in Table A.3.

Table A.3: Square Network Final Size PMFs

Initial State	$SSSI$	$SIIS$
$\mathbb{P}$(Final Size =1)	$\frac{1}{2\mathcal{R}+1}$	0
$\mathbb{P}$(Final Size =2)	$\frac{2\mathcal{R}}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$	$\frac{1}{(2\mathcal{R}+1)^2}$
$\mathbb{P}$(Final Size =3)	$\frac{4\mathcal{R}^2}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$	$\frac{2\mathcal{R}(3\mathcal{R}^2+5\mathcal{R}+2)}{(\mathcal{R}+1)^3(2\mathcal{R}+1)^2}$
$\mathbb{P}$(Final Size =4)	$1-\frac{\mathcal{R}^3+9\mathcal{R}^2+5\mathcal{R}+1}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$	$1-\frac{7\mathcal{R}^2+6\mathcal{R}+1}{(\mathcal{R}+1)^2(2\mathcal{R}+1)^2}$
Expected FS	$4-\frac{3\mathcal{R}^3+17\mathcal{R}^2+13\mathcal{R}+3}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$	$4-\frac{2}{(\mathcal{R}+1)^2}$

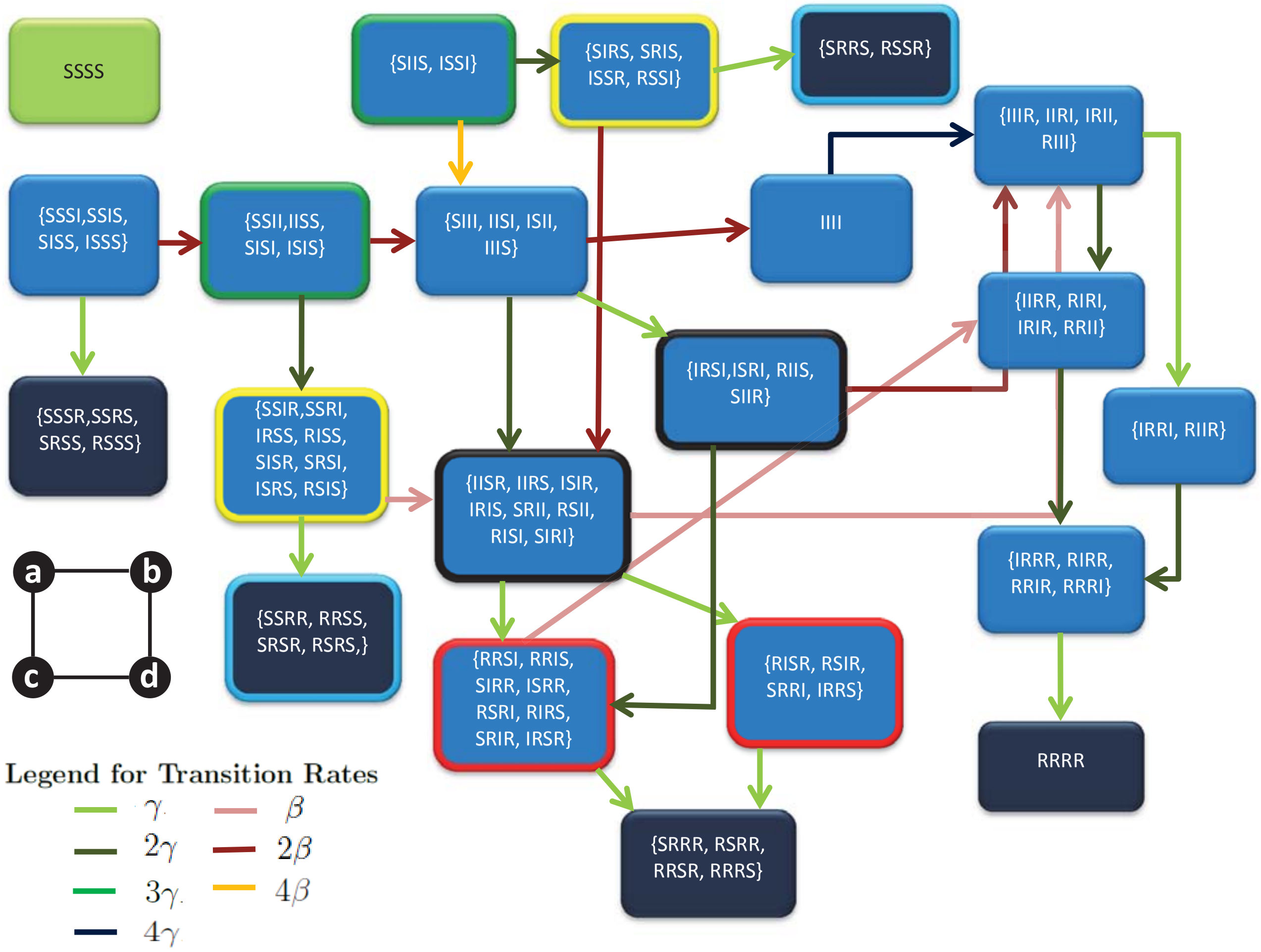


Figure A.5: Transition diagram for the square network with $N = 4$ nodes.

A.3.1 Catalogue of transition probabilities

Possible initial state indicator variables:

$$\mathrm{E}_{SSSI} = \begin{cases} 1, & \text{if initial state is } SSSI. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{SIIS} = \begin{cases} 1, & \text{if initial state is } SIIS. \\ 0, & \text{otherwise.} \end{cases}$$

Probability of passing through transient states:

$$\begin{aligned}
\mathcal{P}_{SSII} &= \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{SSSI} \\
\mathcal{P}_{SSIR} &= \frac{2}{2\mathcal{R}+2}\mathcal{P}_{SSII} \\
\mathcal{P}_{SIRS} &= \frac{2}{4\mathcal{R}+2}\mathrm{E}_{SIIS} \\
\mathcal{P}_{SIII} &= \frac{4\mathcal{R}}{4\mathcal{R}+2}\mathrm{E}_{SIIS} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{SSII} \\
\mathcal{P}_{IISR} &= \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{SIRS} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SSIR} + \frac{2}{2\mathcal{R}+2}\mathcal{P}_{SIII} \\
\mathcal{P}_{RRSI} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{IISR} + \frac{2}{2\mathcal{R}+2}\mathcal{P}_{IRSI} \\
\mathcal{P}_{IIII} &= \frac{2\mathcal{R}}{2\mathcal{R}+3}\mathcal{P}_{SIII} \\
\mathcal{P}_{IRSI} &= \frac{1}{2\mathcal{R}+3}\mathcal{P}_{SIII} \\
\mathcal{P}_{RISR} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{IISR} \\
\mathcal{P}_{IIIR} &= \mathcal{P}_{IIII} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{IRSI} + \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{IISR} \\
\mathcal{P}_{IIRR} &= \frac{2}{3}\mathcal{P}_{IIIR} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{RRSI} \\
\mathcal{P}_{IRRI} &= \frac{1}{3}\mathcal{P}_{IIIR} \\
\mathcal{P}_{IRRR} &= \mathcal{P}_{IRRI} + \mathcal{P}_{IIRR}
\end{aligned}$$

Probability of terminating in absorbing states:

$$\begin{aligned}
\mathcal{P}_{SSSR} &= \frac{1}{2\mathcal{R}+1}\mathrm{E}_{SSSI} \\
\mathcal{P}_{SSRR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SSIR} \\
\mathcal{P}_{SRRS} &= \frac{1}{2\mathcal{R}+1}\mathcal{P}_{SIRS} \\
\mathcal{P}_{SRRR} &= \mathcal{P}_{RRSI} + \mathcal{P}_{RISR} \\
\mathcal{P}_{RRRR} &= \mathcal{P}_{IRRR}
\end{aligned}$$

To find the equations for the final size probabilities we evaluated the following:

$$\begin{aligned}
\mathbb{P}(\text{Final Size} = 1) &= \mathcal{P}_{SSSR} \\
\mathbb{P}(\text{Final Size} = 2) &= \mathcal{P}_{SSRR} + \mathcal{P}_{SRRS} \\
\mathbb{P}(\text{Final Size} = 3) &= \mathcal{P}_{SRRR} \\
\mathbb{P}(\text{Final Size} = 4) &= \mathcal{P}_{RRRR}
\end{aligned}$$

Simplifying the above we obtained the final size equations for the square network with four nodes as shown in Table A.3. The probability mass function of the final size distribution for the square network is shown in Figure 2.10.

A.3.2 Progression of infection over time

Equations describing the probability that the network is in a given state at time t for an SIR model on the square network are:

$$\begin{aligned}
&\text{Initial states:} && \\
&\dot{\mathsf{P}}_{SSSS} &=&\ 0 \\
&\dot{\mathsf{P}}_{SSSI} &=&\ -(2\mathcal{R}+1)\mathsf{P}_{SSSI} \\
&\dot{\mathsf{P}}_{SIIS} &=&\ -(4\mathcal{R}+2)\mathsf{P}_{SIIS} \\
&\text{Transient states:} && \\
&\dot{\mathsf{P}}_{SSII} &=&\ 2\mathcal{R}\mathsf{P}_{SSSI} - 2(\mathcal{R}+1)\mathsf{P}_{SSII} \\
&\dot{\mathsf{P}}_{SSIR} &=&\ 2\mathsf{P}_{SSII} - (\mathcal{R}+1)\mathsf{P}_{SSIR} \\
&\dot{\mathsf{P}}_{SIII} &=&\ 2\mathcal{R}\mathsf{P}_{SSII} + 4\mathcal{R}\mathsf{P}_{SIIS} - (2\mathcal{R}+3)\mathsf{P}_{SIII} \\
&\dot{\mathsf{P}}_{SIRI} &=&\ 2\mathsf{P}_{SIII} + \mathcal{R}\mathsf{P}_{SSIR} + 2\mathcal{R}\mathsf{P}_{SIRS} \\
& && -(\mathcal{R}+2)\mathsf{P}_{SIRI} \\
&\dot{\mathsf{P}}_{SIRR} &=&\ \mathsf{P}_{SIRI} + 2\mathsf{P}_{SIIR} - (\mathcal{R}+1)\mathsf{P}_{SIRR} \\
&\dot{\mathsf{P}}_{SRRI} &=&\ \mathsf{P}_{SIRI} - \mathsf{P}_{SRRI} \\
&\dot{\mathsf{P}}_{SIIR} &=&\ \mathsf{P}_{SIII} - 2(\mathcal{R}+1)\mathsf{P}_{SIIR} \\
&\dot{\mathsf{P}}_{IIII} &=&\ 2\mathcal{R}\mathsf{P}_{SIII} - 4\mathsf{P}_{IIII} \\
&\dot{\mathsf{P}}_{IIIR} &=&\ 4\mathsf{P}_{IIII} + 2\mathcal{R}\mathsf{P}_{SIIR} + \mathcal{R}\mathsf{P}_{SIRI} - 3\mathsf{P}_{IIIR} \\
&\dot{\mathsf{P}}_{IIRR} &=&\ 2\mathsf{P}_{IIIR} + \mathcal{R}\mathsf{P}_{SIRR} - 2\mathsf{P}_{IIRR} \\
&\dot{\mathsf{P}}_{IRRI} &=&\ \mathsf{P}_{IIIR} - 2\mathsf{P}_{IRRI} \\
&\dot{\mathsf{P}}_{IRRR} &=&\ 2(\mathsf{P}_{IIRR} + \mathsf{P}_{IRRI}) - \mathsf{P}_{IRRR} \\
&\dot{\mathsf{P}}_{SIRS} &=&\ 2\mathsf{P}_{SIIS} - (2\mathcal{R}+1)\mathsf{P}_{SIRS}
\end{aligned}$$

Absorbing states:

$$\begin{aligned}
\dot{\mathsf{P}}_{SSSR} &= \mathsf{P}_{SSSI} \\
\dot{\mathsf{P}}_{SSRR} &= \mathsf{P}_{SSIR} \\
\dot{\mathsf{P}}_{SRRS} &= \mathsf{P}_{SIRS} \\
\dot{\mathsf{P}}_{SRRR} &= \mathsf{P}_{SIRR} + \mathsf{P}_{SRRI} \\
\dot{\mathsf{P}}_{RRRR} &= \mathsf{P}_{IRRR}
\end{aligned}$$

To find the final size probabilities we numerically solve the system of differential equations with specified initial conditions and sum up the appropriate absorbing state probabilities once a steady state has been reached. See Figure A.6 for numerical results that confirm our analytical expressions for the final epidemic size.

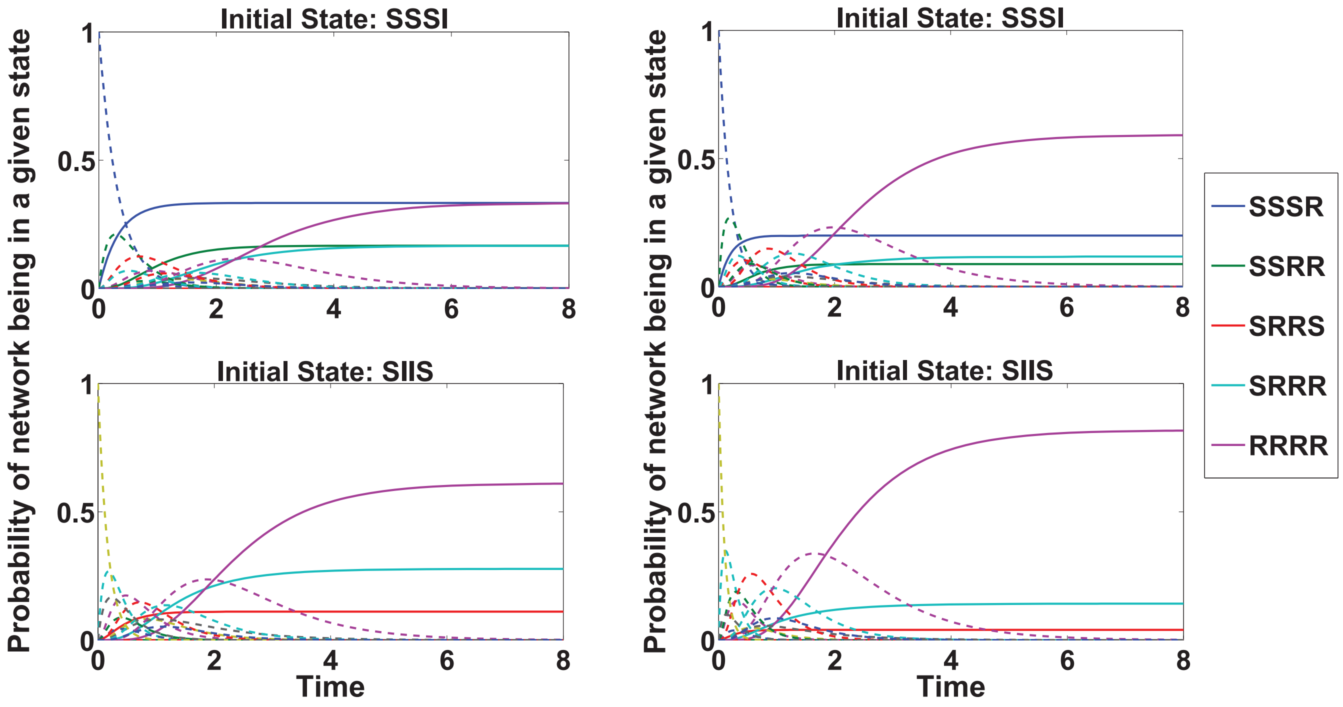
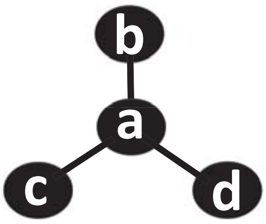


Figure A.6: Numerical solution of the system of differential equations described above for the progression of infection over time for an SIR model on a square network with $N = 4$ nodes. Left and right columns contain graphical results for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively for the specified initial conditions. Dashed lines represent transient states and solid lines are the absorbing states. The numerical results are in agreement with the analytical expressions for the same set of initial conditions.

A.4 Star Network



For a star network of N nodes there are two different types of nodes, the centre node with degree $N-1$ and the $N-1$ outer nodes each of which have degree one. Here we have a centre node with degree 3 and three outer nodes with degree 1. Therefore, the final size probabilities will again depend on which type of node is initially infected. We denote $ISSS$ as the initial state in which the centre node is infectious; $SSSI$, $SSII$ and $SIII$ denote the initial state in which 1, 2 and 3 of the outer nodes are infectious respectively. Figure A.7 shows the transition diagram of the star network. The probability mass functions of the final size for the star network found with four initial conditions are shown in Table A.4.

A.4.1 Catalogue of transition probabilities

Possible initial state indicator variables:

$$\mathrm{E}_{SSSI} = \begin{cases} 1, & \text{if initial state is } SSSI. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{ISSS} = \begin{cases} 1, & \text{if initial state is } ISSS. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{SSII} = \begin{cases} 1, & \text{if initial state is } SSII. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{SIII} = \begin{cases} 1, & \text{if initial state is } SIII. \\ 0, & \text{otherwise.} \end{cases}$$

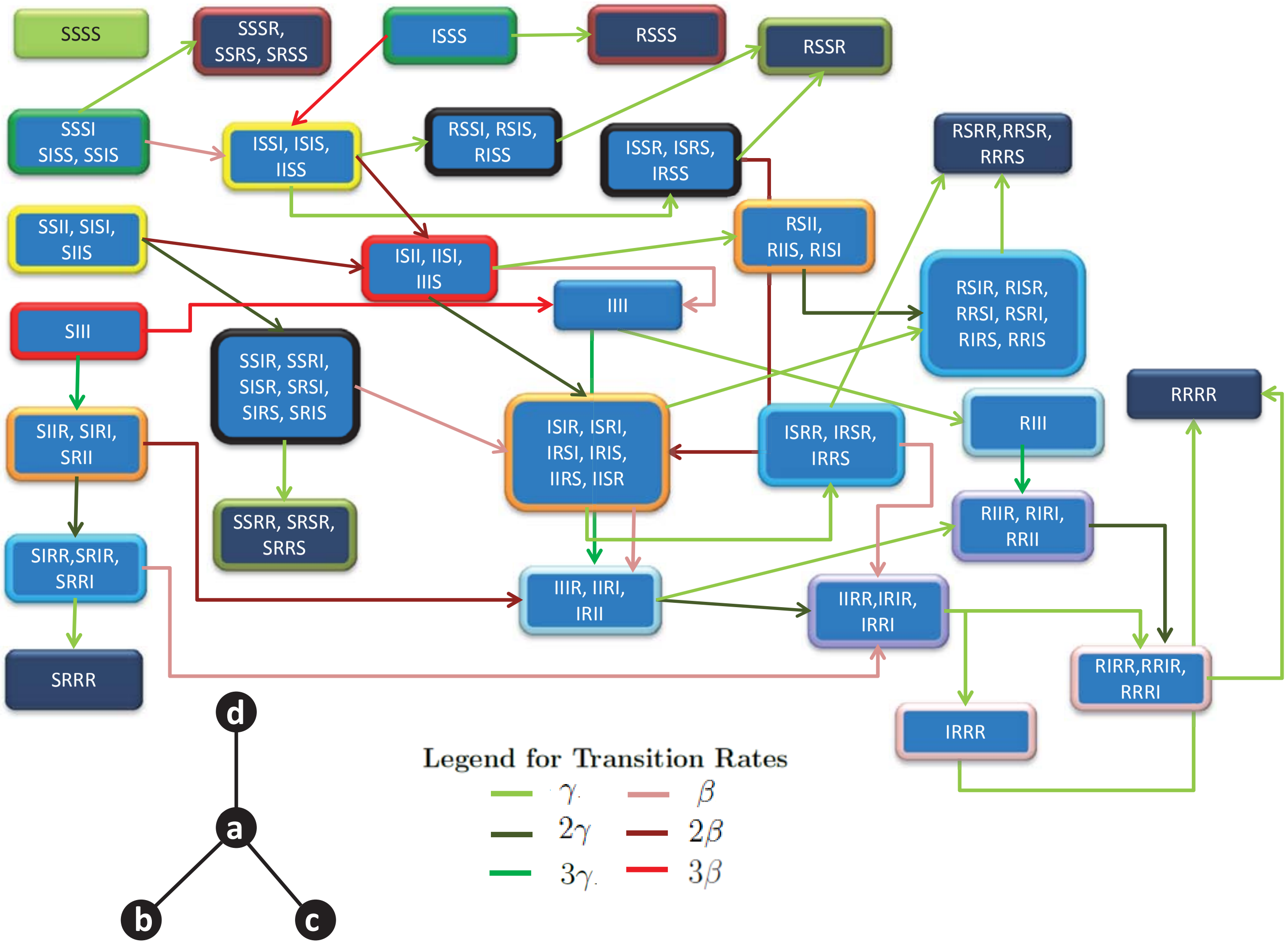


Figure A.7: Transition diagram for the star network with $N = 4$ nodes.

Table A.4: Star Network final size PMFs

Initial State	$SSSI$	$ISSS$	$SSII$	$SIII$
$\mathbb{P}$**(Final Size =1)**	$\frac{1}{\mathcal{R}+1}$	$\frac{1}{3\mathcal{R}+1}$	0	0
$\mathbb{P}$**(Final Size =2)**	$\frac{\mathcal{R}}{(\mathcal{R}+1)(2\mathcal{R}+1)}$	$\frac{3\mathcal{R}}{(2\mathcal{R}+1)(3\mathcal{R}+1)}$	$\frac{1}{(\mathcal{R}+1)^2}$	0
$\mathbb{P}$**(Final Size =3)**	$\frac{2\mathcal{R}^2}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$	$\frac{6\mathcal{R}^2}{(\mathcal{R}+1)(2\mathcal{R}+1)(3\mathcal{R}+1)}$	$\frac{\mathcal{R}}{(\mathcal{R}+1)^2}$	$\frac{1}{(\mathcal{R}+1)^3}$
$\mathbb{P}$**(Final Size =4)**	$\frac{6\mathcal{R}^4-5\mathcal{R}^2-4\mathcal{R}-1}{(\mathcal{R}+1)^2(2\mathcal{R}+1)(3\mathcal{R}+1)}$	$\frac{6\mathcal{R}^3}{(\mathcal{R}+1)(2\mathcal{R}+1)(3\mathcal{R}+1)}$	$\frac{\mathcal{R}^2(\mathcal{R}+2)}{(\mathcal{R}+1)^3}$	$1-\frac{1}{(\mathcal{R}+1)^3}$
Expected FS	$4-\frac{5\mathcal{R}+3}{(\mathcal{R}+1)^2}$	$4-\frac{3}{\mathcal{R}+1}$	$4-\frac{\mathcal{R}^2+4\mathcal{R}+2}{(\mathcal{R}+1)^3}$	$4-\frac{1}{(\mathcal{R}+1)^3}$

Probability of passing through transient states:

$$\begin{aligned}
\mathcal{P}_{SIIR} &= \frac{3}{3\mathcal{R}+3}\mathrm{E}_{SIII} \\
\mathcal{P}_{SIRR} &= \frac{2}{2\mathcal{R}+2}\mathcal{P}_{SIIR} \\
\mathcal{P}_{ISSI} &= \frac{\mathcal{R}}{\mathcal{R}+1}\mathrm{E}_{SSSI} + \frac{3\mathcal{R}}{3\mathcal{R}+1}\mathrm{E}_{ISSS} \\
\mathcal{P}_{SSIR} &= \frac{2}{2\mathcal{R}+2}\mathrm{E}_{SSII} \\
\mathcal{P}_{RSSI} &= \frac{1}{2\mathcal{R}+2}\mathcal{P}_{ISSI} \\
\mathcal{P}_{ISII} &= \frac{2\mathcal{R}}{2\mathcal{R}+2}(\mathcal{P}_{ISSI} + \mathrm{E}_{SSII}) \\
\mathcal{P}_{ISSR} &= \frac{1}{2\mathcal{R}+2}\mathcal{P}_{ISSI} \\
\mathcal{P}_{IIII} &= \frac{\mathcal{R}}{\mathcal{R}+3}\mathcal{P}_{ISII} + \frac{3\mathcal{R}}{3\mathcal{R}+3}\mathrm{E}_{SIII} \\
\mathcal{P}_{ISIR} &= \frac{2}{\mathcal{R}+3}\mathcal{P}_{ISII} + \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{ISSR} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SSIR} \\
\mathcal{P}_{RSII} &= \frac{1}{\mathcal{R}+3}\mathcal{P}_{ISII} \\
\mathcal{P}_{ISRR} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{ISIR} \\
\mathcal{P}_{RSIR} &= \mathcal{P}_{RSII} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{ISIR} \\
\mathcal{P}_{IIIR} &= \frac{3}{4}\mathcal{P}_{IIII} + \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{ISIR} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{SIIR} \\
\mathcal{P}_{RIII} &= \frac{1}{4}\mathcal{P}_{IIII} \\
\mathcal{P}_{RIIR} &= \mathcal{P}_{RIII} + \frac{1}{3}\mathcal{P}_{IIIR} \\
\mathcal{P}_{IIRR} &= \frac{\mathcal{R}}{\mathcal{R}+1}(\mathcal{P}_{ISRR} + \mathcal{P}_{SIRR}) + \frac{2}{3}\mathcal{P}_{IIIR} \\
\mathcal{P}_{IRRR} &= \frac{1}{2}\mathcal{P}_{IIRR} \\
\mathcal{P}_{RIRR} &= \frac{1}{2}\mathcal{P}_{IIRR} + \mathcal{P}_{RIIR}
\end{aligned}$$

Probability of terminating in absorbing states:

$$\begin{aligned}
\mathcal{P}_{SSSR} &= \frac{1}{\mathcal{R}+1}\mathrm{E}_{SSSI} \\
\mathcal{P}_{RSSS} &= \frac{1}{3\mathcal{R}+1}\mathrm{E}_{ISSS} \\
\mathcal{P}_{SSRR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SSIR} \\
\mathcal{P}_{RSSR} &= \mathcal{P}_{RSSI} + \frac{1}{2\mathcal{R}+1}\mathcal{P}_{ISSR} \\
\mathcal{P}_{SRRR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SIRR}
\end{aligned}$$

$$\begin{aligned}\mathcal{P}_{RSRR} &= \mathcal{P}_{RSIR} + \frac{1}{\mathcal{R}+1}\mathcal{P}_{ISRR}\\ \mathcal{P}_{RRRR} &= \mathcal{P}_{IRRR} + \mathcal{P}_{RIRR}\end{aligned}$$

To find the equations for the final size probabilities we evaluated the following:

$$\begin{aligned}\mathbb{P}(\text{Final Size} = 1) &= \mathcal{P}_{SSSR} + \mathcal{P}_{RSSS}\\ \mathbb{P}(\text{Final Size} = 2) &= \mathcal{P}_{SSRR} + \mathcal{P}_{RSSR}\\ \mathbb{P}(\text{Final Size} = 3) &= \mathcal{P}_{SRRR} + \mathcal{P}_{RSRRR}\\ \mathbb{P}(\text{Final Size} = 4) &= \mathcal{P}_{RRRR}\end{aligned}$$

Simplifying the above we obtained the final size equations for the star network with four nodes as shown in Table A.4. The probability mass function of the final size distribution for the star network is shown in Figure 2.11.

A.4.2 Progression of infection over time

Equations describing the probability that the network is in a given state at time t for an SIR model on the star network are:

Initial states:

$$\begin{aligned}\dot{\mathsf{P}}_{SSSS} &= 0\\ \dot{\mathsf{P}}_{SSSI} &= -(\mathcal{R}+1)\mathsf{P}_{SSSI}\\ \dot{\mathsf{P}}_{ISSS} &= -(3\mathcal{R}+1)\mathsf{P}_{ISSS}\\ \dot{\mathsf{P}}_{SSII} &= -2(\mathcal{R}+1)\mathsf{P}_{SSII}\\ \dot{\mathsf{P}}_{SIII} &= -3(\mathcal{R}+1)\mathsf{P}_{SIII}\end{aligned}$$

Transient states:

$$\begin{aligned}\dot{\mathsf{P}}_{ISSI} &= 3\mathcal{R}\mathsf{P}_{ISSS} + \mathcal{R}\mathsf{P}_{SSSI} - 2(\mathcal{R}+1)\mathsf{P}_{ISSI}\\ \dot{\mathsf{P}}_{ISII} &= 2\mathcal{R}(\mathsf{P}_{ISSI} + \mathsf{P}_{SSII}) - (\mathcal{R}+3)\mathsf{P}_{ISII}\\ \dot{\mathsf{P}}_{SSIR} &= 2\mathsf{P}_{SSII} - (\mathcal{R}+1)\mathsf{P}_{SSIR}\\ \dot{\mathsf{P}}_{RSSI} &= \mathsf{P}_{ISSI} - \mathsf{P}_{RSSI}\\ \dot{\mathsf{P}}_{ISSR} &= \mathsf{P}_{ISSI} - (2\mathcal{R}+1)\mathsf{P}_{ISSR}\\ \dot{\mathsf{P}}_{SIIR} &= 3\mathsf{P}_{SIII} - 2(\mathcal{R}+1)\mathsf{P}_{SIIR}\\ \dot{\mathsf{P}}_{RSII} &= \mathsf{P}_{ISII} - 2\mathsf{P}_{RSII}\\ \dot{\mathsf{P}}_{ISIR} &= 2\mathsf{P}_{ISII} + \mathcal{R}(\mathsf{P}_{SSIR} + 2\mathsf{P}_{ISSR}) - (\mathcal{R}+2)\mathsf{P}_{ISIR}\end{aligned}$$

$$
\begin{aligned}
\dot{\mathsf{P}}_{IIII} &= \mathcal{R}(3\mathsf{P}_{SIII} + \mathsf{P}_{ISII}) - 4\mathsf{P}_{IIII} \\
\dot{\mathsf{P}}_{SIRR} &= 2\mathsf{P}_{SIIR} - (\mathcal{R}+1)\mathsf{P}_{SIRR} \\
\dot{\mathsf{P}}_{RSIR} &= (2\mathsf{P}_{RSII} + \mathsf{P}_{ISIR}) - 1\mathsf{P}_{RSIR} \\
\dot{\mathsf{P}}_{ISRR} &= \mathsf{P}_{ISIR} - (\mathcal{R}+1)\mathsf{P}_{ISRR} \\
\dot{\mathsf{P}}_{IIIR} &= \mathcal{R}(2\mathsf{P}_{SIIR} + \mathsf{P}_{ISIR}) + 3\mathsf{P}_{IIII} - 3\mathsf{P}_{IIIR} \\
\dot{\mathsf{P}}_{RIII} &= \mathsf{P}_{IIII} - 3\mathsf{P}_{RIII} \\
\dot{\mathsf{P}}_{RIIR} &= (3\mathsf{P}_{RIII} + \mathsf{P}_{IIIR}) - 2\mathsf{P}_{RIIR} \\
\dot{\mathsf{P}}_{IIRR} &= \mathcal{R}(\mathsf{P}_{ISRR} + \mathsf{P}_{SIRR}) + 2\mathsf{P}_{IIIR} - 2\mathsf{P}_{IIRR} \\
\dot{\mathsf{P}}_{IRRR} &= \mathsf{P}_{IIRR} - \mathsf{P}_{IRRR} \\
\dot{\mathsf{P}}_{RIRR} &= (\mathsf{P}_{IIRR} + 2\mathsf{P}_{RIIR}) - \mathsf{P}_{RIRR}
\end{aligned}
$$

Absorbing states:

$$
\begin{aligned}
\dot{\mathsf{P}}_{SSSR} &= \mathsf{P}_{SSSI} \\
\dot{\mathsf{P}}_{RSSS} &= \mathsf{P}_{ISSS} \\
\dot{\mathsf{P}}_{SSRR} &= \mathsf{P}_{SSIR} \\
\dot{\mathsf{P}}_{RSSR} &= (\mathsf{P}_{RSSI} + \mathsf{P}_{ISSR}) \\
\dot{\mathsf{P}}_{SRRR} &= \mathsf{P}_{SIRR} \\
\dot{\mathsf{P}}_{RSRR} &= (\mathsf{P}_{ISRR} + \mathsf{P}_{RSIR}) \\
\dot{\mathsf{P}}_{RRRR} &= (\mathsf{P}_{IRRR} + \mathsf{P}_{RIRR})
\end{aligned}
$$

To find the final size probabilities we numerically solve the system of differential equations with specified initial conditions and sum up the appropriate absorbing state probabilities once a steady state has been reached. See Figure A.8 for numerical results that confirm our analytical expressions for the final epidemic size.

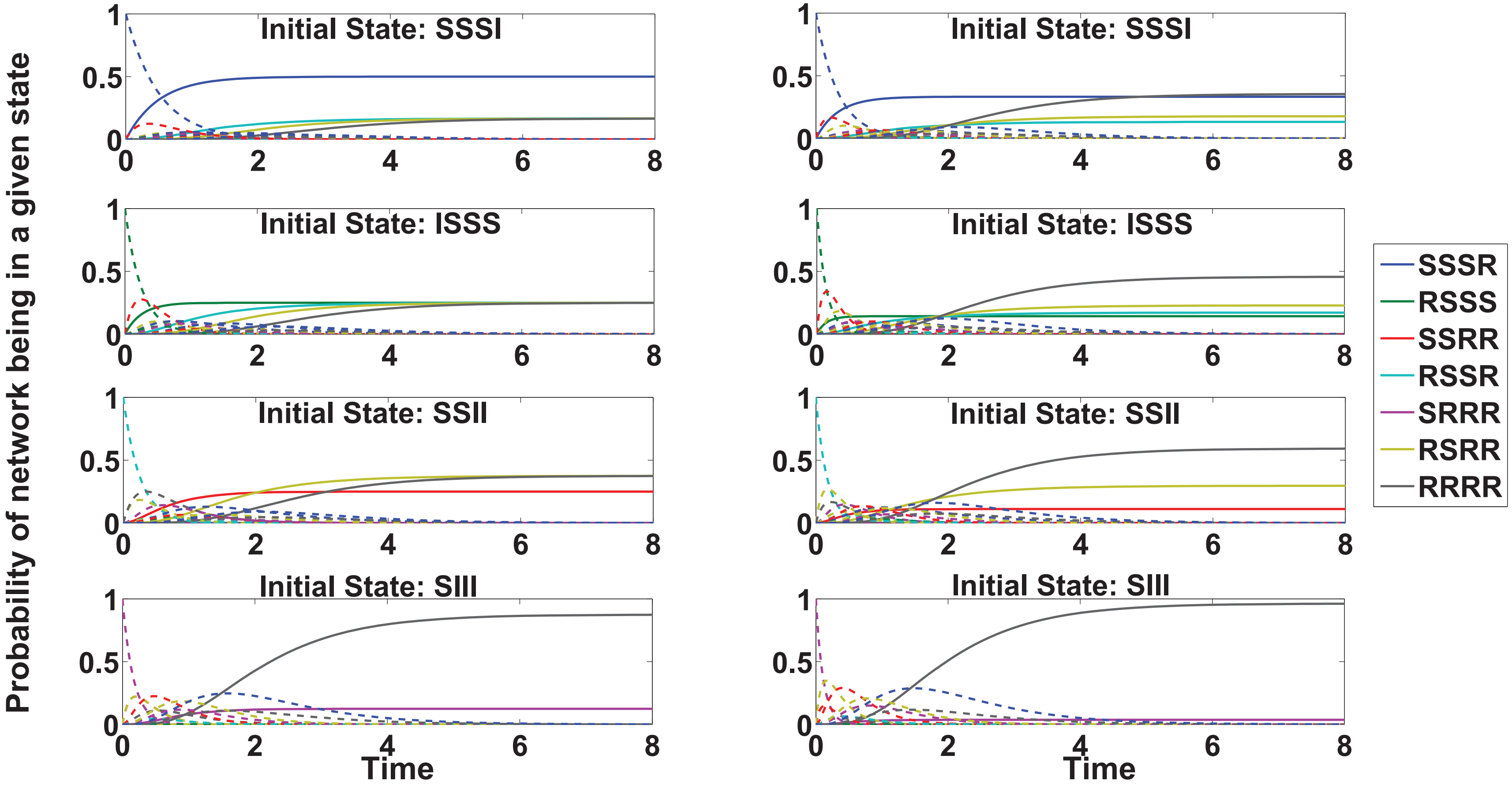
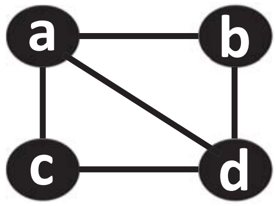


Figure A.8: Numerical solution of the system of differential equations described above for the progression of infection over time for an SIR model on a star network with $N = 4$ nodes. Left and right columns contain graphical results for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively for the specified initial conditions. Dashed lines represent transient states and solid lines are the absorbing states. The numerical results are in agreement with the analytical expressions for the same set of initial conditions.

A.5 Toast Network



The toast network is simply a square network with one diagonal edge through it. There are two different types of nodes in the toast network, two nodes of degree 2 (nodes b and c) and two nodes of degree 3 (nodes a and d). We denote $SSIS$ as the initial state in which node b or c is infectious; $SSSI$ as the initial state in which node a or d is infectious and $SIIS$ as the initial state in which nodes b and c are infectious. Figure A.9 shows the transition diagram of the toast network. The probability mass functions of the final size for the toast network found with three initial conditions are shown in Table A.5.

A.5.1 Catalogue of transition probabilities

Possible initial state indicator variables:

$$\mathrm{E}_{SISS} = \begin{cases} 1, & \text{if initial state is } SISS. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{ISSS} = \begin{cases} 1, & \text{if initial state is } ISSS. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{SIIS} = \begin{cases} 1, & \text{if initial state is } SIIS. \\ 0, & \text{otherwise.} \end{cases}$$

Probability of passing through transient states:

$$\begin{aligned}
\mathcal{P}_{IISS} &= \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{SISS} + \frac{2\mathcal{R}}{3\mathcal{R}+1}\mathcal{P}_{SSSI} \\
\mathcal{P}_{ISSI} &= \frac{\mathcal{R}}{3\mathcal{R}+1}\mathcal{P}_{SSSI} \\
\mathcal{P}_{RSSI} &= \frac{2}{4\mathcal{R}+2}\mathcal{P}_{ISSI} \\
\mathcal{P}_{SRIS} &= \frac{2}{4\mathcal{R}+2}\mathrm{E}_{SIIS} \\
\mathcal{P}_{IIIS} &= \frac{4\mathcal{R}}{4\mathcal{R}+2}\mathrm{E}_{SIIS} + \frac{\mathcal{R}}{3\mathcal{R}+2}\mathcal{P}_{IISS}
\end{aligned}$$

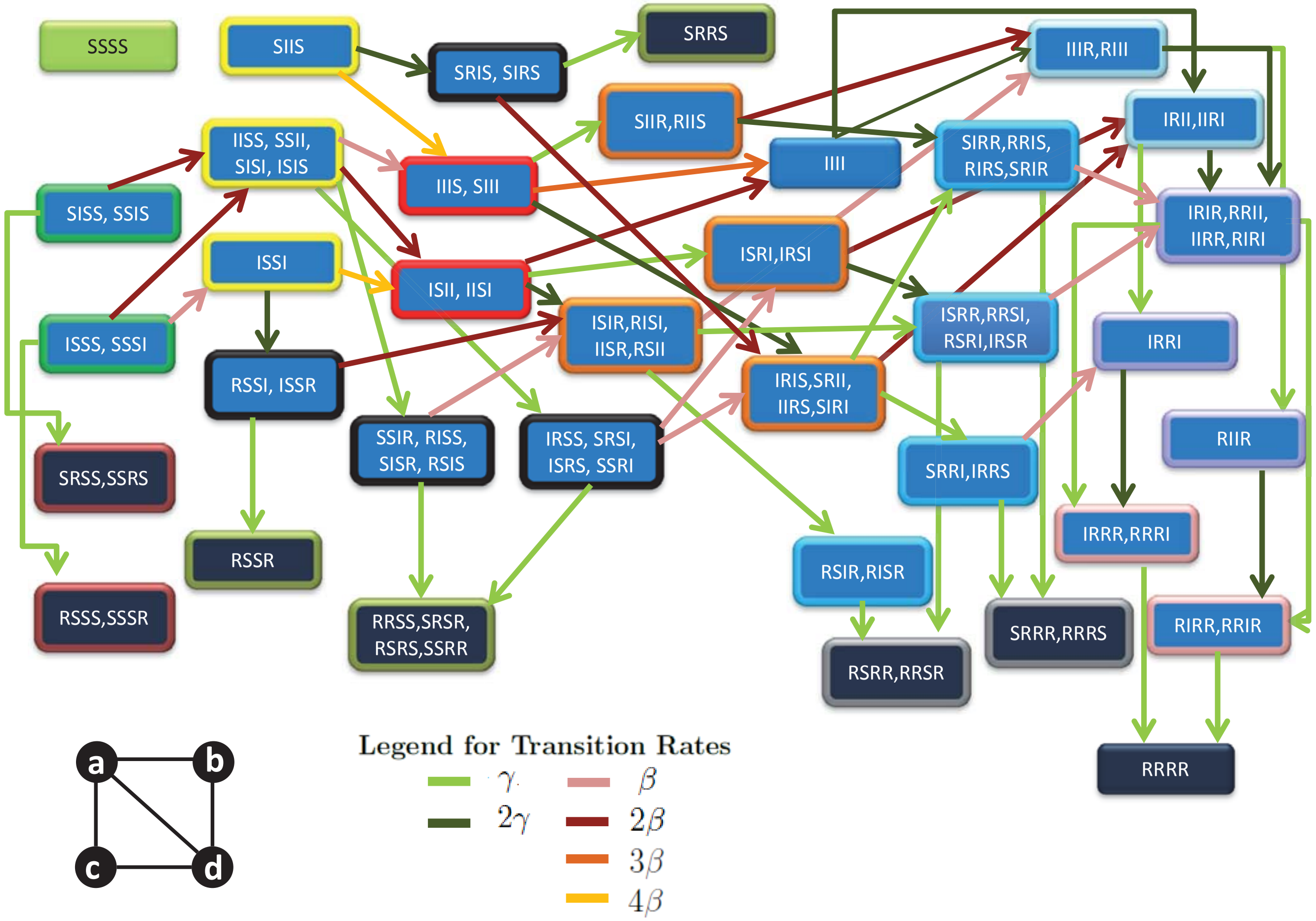


Figure A.9: Transition diagram for a toast network with $N = 4$ nodes.

Table A.5: Toast network final size PMFs

Initial State	$SSIS$	$SSSI$	$SIIS$
$\mathbb{P}$**(Final Size =1)**	$\frac{1}{2\mathcal{R}+1}$	$\frac{1}{3\mathcal{R}+1}$	0
$\mathbb{P}$**(Final Size =2)**	$\frac{2\,\mathcal{R}}{(\mathcal{R}+1)\,(2\,\mathcal{R}+1)^2}$	$\frac{\mathcal{R}\,(5\,\mathcal{R}+3)}{(\mathcal{R}+1)\,(2\,\mathcal{R}+1)^2\,(3\,\mathcal{R}+1)}$	$\frac{1}{(2\mathcal{R}+1)^2}$
$\mathbb{P}$**(Final Size =3)**	$\frac{2\,\mathcal{R}^2\,(2\,\mathcal{R}+3)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}$	$\frac{2\,\mathcal{R}^2\,\left(7\,\mathcal{R}^2+13\,\mathcal{R}+5\right)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2\,(3\,\mathcal{R}+1)}$	$\frac{2\,\mathcal{R}\,(3\,\mathcal{R}+2)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}$
$\mathbb{P}$**(Final Size =4)**	$\frac{2\,\mathcal{R}^3\,(2\,\mathcal{R}+3)\,(\mathcal{R}+2)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}$	$\frac{2\,\mathcal{R}^3\,\left(6\,\mathcal{R}^3+24\,\mathcal{R}^2+28\,\mathcal{R}+9\right)}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2\,(3\,\mathcal{R}+1)}$	$1-\frac{\mathcal{R}^3+9\,\mathcal{R}^2+7\,\mathcal{R}+1}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}$
Expected FS	$4-\frac{6\,\mathcal{R}^4+29\,\mathcal{R}^3+4\,\mathcal{R}^2+19\,\mathcal{R}+3}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}$	$4-\frac{6\,\mathcal{R}^4+33\,\mathcal{R}^3+47\,\mathcal{R}^2+21\,\mathcal{R}+3}{(\mathcal{R}+1)^3\,(2\,\mathcal{R}+1)\,(3\,\mathcal{R}+1)}$	$4-\frac{2\,\mathcal{R}^3+12\,\mathcal{R}^2+10\,\mathcal{R}+2}{(\mathcal{R}+1)^3(2\,\mathcal{R}+1)^2}$

$$\begin{aligned}
\mathcal{P}_{ISII} &= \frac{2\mathcal{R}}{3\mathcal{R}+2}\mathcal{P}_{IISS} + \frac{4\mathcal{R}}{4\mathcal{R}+2}\mathcal{P}_{ISSI} \\
\mathcal{P}_{SSIR} &= \frac{1}{3\mathcal{R}+2}\mathcal{P}_{IISS} \\
\mathcal{P}_{IRSS} &= \frac{1}{3\mathcal{R}+2}\mathcal{P}_{IISS} \\
\mathcal{P}_{SIIR} &= \frac{1}{3\mathcal{R}+3}\mathcal{P}_{IIIS} \\
\mathcal{P}_{ISRI} &= \frac{1}{2\mathcal{R}+3}\mathcal{P}_{ISII} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{IRSS} \\
\mathcal{P}_{ISIR} &= \frac{2}{2\mathcal{R}+3}\mathcal{P}_{ISII} + \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{RSSI} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SSIR} \\
\mathcal{P}_{IRIS} &= \frac{2}{3\mathcal{R}+3}\mathcal{P}_{IIIS} + \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{SRIS} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{IRSS} \\
\mathcal{P}_{IIII} &= \frac{3\mathcal{R}}{3\mathcal{R}+3}\mathcal{P}_{IIIS} + \frac{2\mathcal{R}}{2\mathcal{R}+3}\mathcal{P}_{ISII} \\
\mathcal{P}_{SIRR} &= \frac{2}{2\mathcal{R}+2}\mathcal{P}_{SIIR} + \frac{1}{2\mathcal{R}+2}\mathcal{P}_{IRIS} \\
\mathcal{P}_{ISRR} &= \frac{2}{2\mathcal{R}+2}\mathcal{P}_{ISRI} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{ISIR} \\
\mathcal{P}_{SRRI} &= \frac{1}{2\mathcal{R}+2}\mathcal{P}_{IRIS} \\
\mathcal{P}_{RSIR} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{ISIR} \\
\mathcal{P}_{IIIR} &= \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{SIIR} + \frac{1}{2}\mathcal{P}_{IIII} + \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{ISIR} \\
\mathcal{P}_{IRII} &= \frac{1}{2}\mathcal{P}_{IIII} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{ISRI} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{IRIS} \\
\mathcal{P}_{IRIR} &= \frac{\mathcal{R}}{\mathcal{R}+1}(\mathcal{P}_{SIRR} + \mathcal{P}_{ISRR}) + \frac{2}{3}(\mathcal{P}_{IRII} + \mathcal{P}_{IIIR}) \\
\mathcal{P}_{IRRI} &= \frac{1}{3}\mathcal{P}_{IRII} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SRRI} \\
\mathcal{P}_{RIIR} &= \frac{1}{3}\mathcal{P}_{IIIR} \\
\mathcal{P}_{IRRR} &= \frac{1}{2}\mathcal{P}_{IRIR} + \mathcal{P}_{IRRI} \\
\mathcal{P}_{RIRR} &= \frac{1}{2}\mathcal{P}_{IRIR} + \mathcal{P}_{RIIR}
\end{aligned}$$

Probability of terminating in absorbing states:

$$\begin{aligned}
\mathcal{P}_{SRSS} &= \frac{1}{2\mathcal{R}+1}\mathrm{E}_{SISS} \\
\mathcal{P}_{SSSR} &= \frac{1}{3\mathcal{R}+1}\mathcal{P}_{SSSI} \\
\mathcal{P}_{RSSR} &= \frac{1}{2\mathcal{R}+1}\mathcal{P}_{RSSI} \\
\mathcal{P}_{SRRS} &= \frac{1}{2\mathcal{R}+1}\mathcal{P}_{SRIS}
\end{aligned}$$

$$\begin{aligned}
\mathcal{P}_{RRSS} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SSIR} + \frac{1}{2\mathcal{R}+1}\mathcal{P}_{IRSS} \\
\mathcal{P}_{RSRR} &= \mathcal{P}_{RSIR} + \frac{1}{\mathcal{R}+1}\mathcal{P}_{ISRR} \\
\mathcal{P}_{SRRR} &= \frac{1}{\mathcal{R}+1}(\mathcal{P}_{SRRI} + \mathcal{P}_{SIRR}) \\
\mathcal{P}_{RRRR} &= \mathcal{P}_{IRRR} + \mathcal{P}_{RIRR}
\end{aligned}$$

To find the equations for the final size probabilities we evaluated the following:

$$\begin{aligned}
\mathbb{P}(\text{Final Size} = 1) &= \mathcal{P}_{SRSS} + \mathcal{P}_{SSSR} \\
\mathbb{P}(\text{Final Size} = 2) &= \mathcal{P}_{RSSR} + \mathcal{P}_{SRRS} + \mathcal{P}_{RRSS} \\
\mathbb{P}(\text{Final Size} = 3) &= \mathcal{P}_{RSRR} + \mathcal{P}_{SRRR} \\
\mathbb{P}(\text{Final Size} = 4) &= \mathcal{P}_{RRRR}
\end{aligned}$$

Simplifying the above we obtained the final size equations for the toast network with four nodes as shown in Table A.5. The probability mass function of the final size distribution for the toast network is shown in Figure 2.12.

A.5.2 Progression of infection over time

Equations describing the probability that the network is in a given state at time t for an SIR model on the toast network are:

$$\begin{aligned}
&\text{Initial states:} \\
\dot{\mathsf{P}}_{SSSS} &= 0 \\
\dot{\mathsf{P}}_{SSIS} &= -(2\mathcal{R}+1)\mathsf{P}_{SSIS} \\
\dot{\mathsf{P}}_{SSSI} &= -(3\mathcal{R}+1)\mathsf{P}_{SSSI} \\
\dot{\mathsf{P}}_{SIIS} &= -(2+4\mathcal{R})\mathsf{P}_{SIIS} \\
&\text{Transient states:} \\
\dot{\mathsf{P}}_{SSII} &= -(3\mathcal{R}+2)\mathsf{P}_{SSII} + 2\mathcal{R}\mathsf{P}_{SSIS} + 2\mathcal{R}\mathsf{P}_{SSSI} \\
\dot{\mathsf{P}}_{ISSI} &= \mathcal{R}\mathsf{P}_{SSSI} - (4\mathcal{R}+2)\mathsf{P}_{ISSI} \\
\dot{\mathsf{P}}_{RSSI} &= 2\mathsf{P}_{ISSI} - (2\mathcal{R}+1)\mathsf{P}_{RSSI} \\
\dot{\mathsf{P}}_{SRIS} &= 2\mathsf{P}_{SIIS} - \mathsf{P}_{SRIS} \\
\dot{\mathsf{P}}_{SIII} &= 4\mathcal{R}\mathsf{P}_{SIIS} + \mathcal{R}\mathsf{P}_{SSII} - (3\mathcal{R}+3)\mathsf{P}_{SIII}
\end{aligned}$$

$$
\begin{aligned}
\dot{\mathsf{P}}_{ISII} &= 2\mathcal{R}\mathsf{P}_{SSII} + 4\mathcal{R}\mathsf{P}_{ISSI} - (2\mathcal{R}+3)\mathsf{P}_{ISII} \\
\dot{\mathsf{P}}_{SSIR} &= \mathsf{P}_{SSII} - (\mathcal{R}+1)\mathsf{P}_{SSIR} \\
\dot{\mathsf{P}}_{SRSI} &= \mathsf{P}_{SSII} - (2\mathcal{R}+1)\mathsf{P}_{SRSI} \\
\dot{\mathsf{P}}_{RISI} &= 2\mathsf{P}_{ISII} + 2\mathcal{R}\mathsf{P}_{RSSI} + \mathcal{R}\mathsf{P}_{SSIR} - (\mathcal{R}+2)\mathsf{P}_{RISI} \\
\dot{\mathsf{P}}_{SIIR} &= \mathsf{P}_{SIII} - 2(\mathcal{R}+1)\mathsf{P}_{SIIR} \\
\dot{\mathsf{P}}_{IRSI} &= \mathsf{P}_{ISII} + \mathcal{R}\mathsf{P}_{SRSI} - 2(\mathcal{R}+1)\mathsf{P}_{IRSI} \\
\dot{\mathsf{P}}_{IRIS} &= 2\mathsf{P}_{SIII} + \mathcal{R}\mathsf{P}_{SRSI} - 2(\mathcal{R}+1)\mathsf{P}_{IRIS} \\
\dot{\mathsf{P}}_{IIII} &= 3\mathcal{R}\mathsf{P}_{SIII} + 2\mathcal{R}\mathsf{P}_{ISII} - 4\mathsf{P}_{IIII} \\
\dot{\mathsf{P}}_{SIRR} &= 2\mathsf{P}_{SIIR} + 2\mathsf{P}_{IRIS} - (\mathcal{R}+1)\mathsf{P}_{SIRR} \\
\dot{\mathsf{P}}_{ISRR} &= \mathsf{P}_{RISI} + 2\mathsf{P}_{IRSI} - (\mathcal{R}+1)\mathsf{P}_{ISRR} \\
\dot{\mathsf{P}}_{SRRI} &= \mathsf{P}_{IRIS} - (\mathcal{R}+1)\mathsf{P}_{SRRI} \\
\dot{\mathsf{P}}_{RSIR} &= \mathsf{P}_{RISI} - \mathsf{P}_{RSIR} \\
\dot{\mathsf{P}}_{IIIR} &= 2\mathcal{R}\mathsf{P}_{SIIR} + \mathcal{R}\mathsf{P}_{RISI} + 2\mathsf{P}_{IIII} - 3\mathsf{P}_{IIIR} \\
\dot{\mathsf{P}}_{IRII} &= 2\mathcal{R}(\mathsf{P}_{IRSI} + \mathsf{P}_{IRIS}) + 2\mathsf{P}_{IIII} - 3\mathsf{P}_{IRII} \\
\dot{\mathsf{P}}_{RIRI} &= 2(\mathsf{P}_{IIIR} + \mathsf{P}_{IRII}) + \mathcal{R}(\mathsf{P}_{SIRR} + \mathsf{P}_{ISRR}) - 2\mathsf{P}_{RIRI} \\
\dot{\mathsf{P}}_{IRRI} &= \mathsf{P}_{IRII} + \mathcal{R}\mathsf{P}_{SRRI} - 2\mathsf{P}_{IRRI} \\
\dot{\mathsf{P}}_{RIIR} &= \mathsf{P}_{IIIR} - 2\mathsf{P}_{RIIR} \\
\dot{\mathsf{P}}_{IRRR} &= 2\mathsf{P}_{IRRI} + \mathsf{P}_{RIRI} - \mathsf{P}_{IRRR} \\
\dot{\mathsf{P}}_{RIRR} &= 2\mathsf{P}_{RIIR} + \mathsf{P}_{RIRI} - \mathsf{P}_{RIRR}
\end{aligned}
$$

Absorbing states:

$$
\begin{aligned}
\dot{\mathsf{P}}_{SRSS} &= \mathsf{P}_{SSIS} \\
\dot{\mathsf{P}}_{SSSR} &= \mathsf{P}_{SSSI} \\
\dot{\mathsf{P}}_{RSSR} &= \mathsf{P}_{RSSI} \\
\dot{\mathsf{P}}_{SRRS} &= \mathsf{P}_{SRIS} \\
\dot{\mathsf{P}}_{SSRR} &= \mathsf{P}_{SSIR} + \mathsf{P}_{SRSI} \\
\dot{\mathsf{P}}_{RSRR} &= \mathsf{P}_{RSIR} + \mathsf{P}_{ISRR} \\
\dot{\mathsf{P}}_{SRRR} &= \mathsf{P}_{SRRI} + \mathsf{P}_{SIRR} \\
\dot{\mathsf{P}}_{RRRR} &= \mathsf{P}_{IRRR} + \mathsf{P}_{RIRR}
\end{aligned}
$$

To find the final size probabilities we numerically solve the system of differential equations with specified initial conditions and sum up the appropriate absorbing state probabilities once a steady state has been reached. See Figure A.10 for numerical results that confirm our analytical expressions for the final epidemic size.

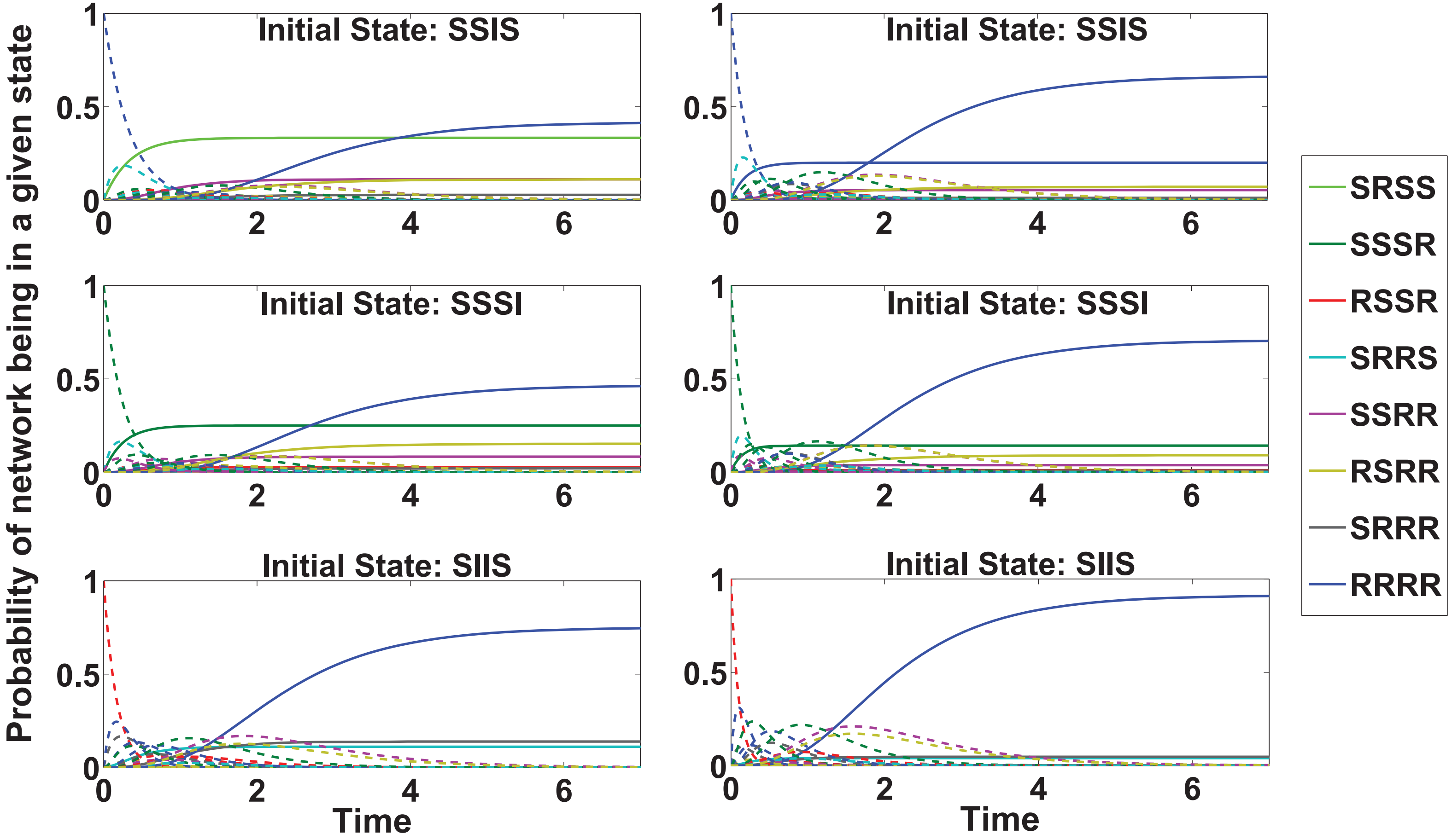
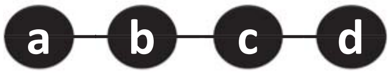


Figure A.10: Numerical solution of the system of differential equations described above for the progression of infection over time for an SIR model on a toast network with $N = 4$ nodes. Left and right columns contain graphical results for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively for the specified initial conditions. Dashed lines represent transient states and solid lines are the absorbing states. The numerical results are in agreement with the analytical expressions for the same set of initial conditions.

A.6 Line Network, $N = 4$



Similarly to a line network of $N = 3$ nodes, a line network of $N = 4$ nodes has two different types of nodes, the end nodes of degree 1 (nodes a and d) and the centre nodes of degree 2 (nodes b and c). We denote $SSIS$ as the initial state in which node b or c is infectious; $SSSI$ as the initial state in which node a or d is infectious; $SISI$ as the initial state in which nodes b and d are infectious and $ISSI$ as the initial state in which nodes a and d are infectious. The transition diagram of the line network is shown in Figure A.9. The probability mass functions of the final epidemic size for the line network of four nodes found with four initial conditions are shown in Table A.6.

A.6.1 Catalogue of transition probabilities

Possible initial state indicator variables:

$$\mathrm{E}_{SISS} = \begin{cases} 1, & \text{if initial state is } SISS. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{SSSI} = \begin{cases} 1, & \text{if initial state is } SSSI. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{SISI} = \begin{cases} 1, & \text{if initial state is } SISI. \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathrm{E}_{ISSI} = \begin{cases} 1, & \text{if initial state is } ISSI. \\ 0, & \text{otherwise.} \end{cases}$$

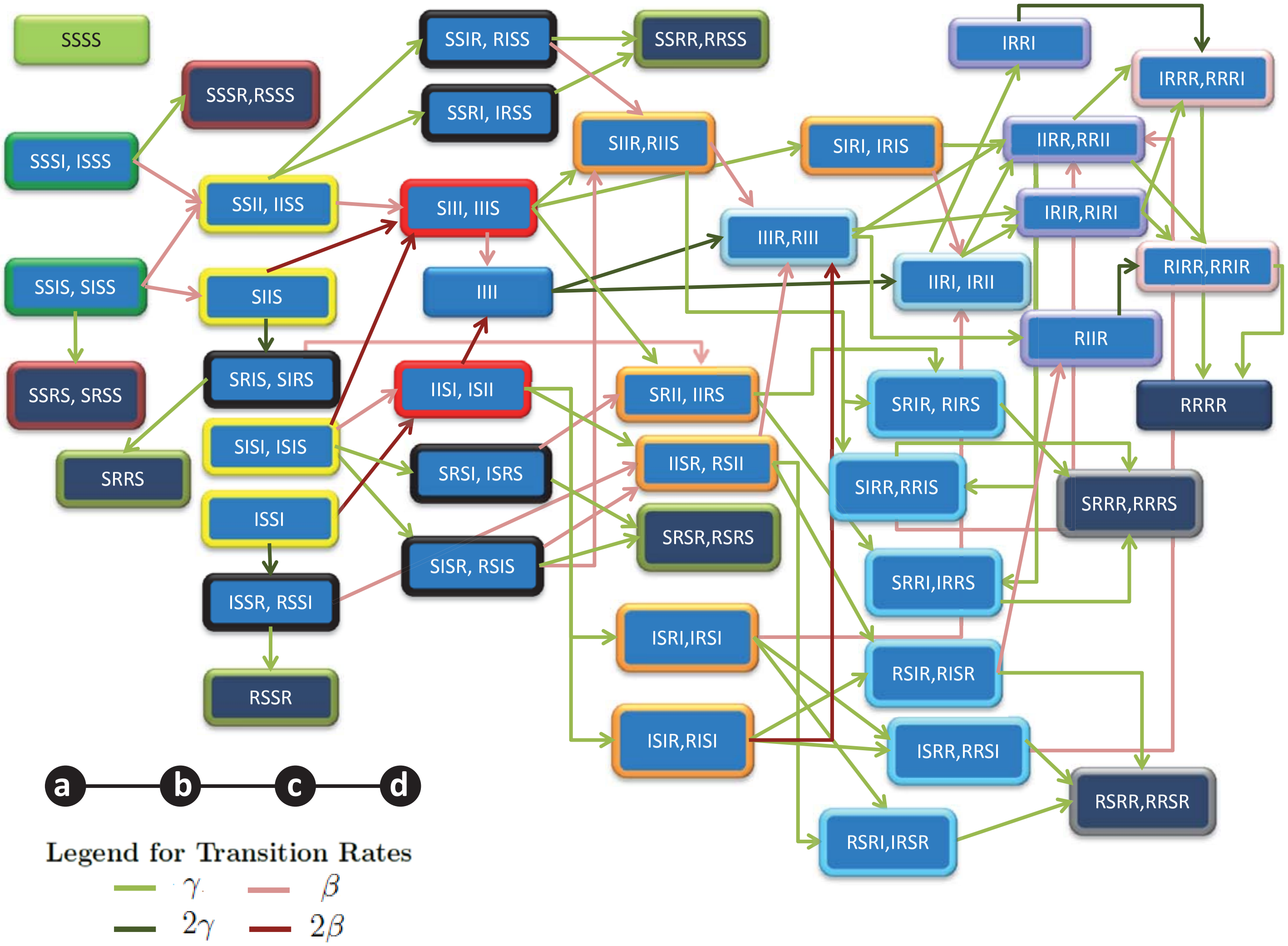


Figure A.11: Transition diagram for a line network with $N = 4$ nodes.

Table A.6: Line network final size PMFs

Initial State	$SSSI$	$SSIS$	$ISSI$	$SISI$
$\mathbb{P}$**(Final Size =1)**	$\frac{1}{\mathcal{R}+1}$	$\frac{1}{2\,\mathcal{R}+1}$	0	0
$\mathbb{P}$**(Final Size =2)**	$\frac{\mathcal{R}}{(\mathcal{R}+1)^2}$	$\frac{\mathcal{R}\,(\mathcal{R}+2)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}$	$\frac{1}{(\mathcal{R}+1)^2}$	$\frac{1}{(\mathcal{R}+1)\,(2\,\mathcal{R}+1)}$
$\mathbb{P}$**(Final Size =3)**	$\frac{\mathcal{R}^2}{(\mathcal{R}+1)^3}$	$\frac{3\mathcal{R}^2}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}$	$\frac{2\,\mathcal{R}}{(\mathcal{R}+1)^3}$	$\frac{\mathcal{R}\,(2\,\mathcal{R}+3)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}$
$\mathbb{P}$**(Final Size =4)**	$\frac{\mathcal{R}^3}{(\mathcal{R}+1)^3}$	$\frac{2\,\mathcal{R}^3}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}$	$1-\frac{3\,\mathcal{R}+1}{(\mathcal{R}+1)^3}$	$\frac{\mathcal{R}^2\,(2\,\mathcal{R}+3)}{(\mathcal{R}+1)^2\,(2\,\mathcal{R}+1)}$
Expected FS	$4-\frac{6\,\mathcal{R}^2+8\,\mathcal{R}+3}{(\mathcal{R}+1)^3}$	$\frac{(2\,\mathcal{R}+1)^2}{(\mathcal{R}+1)^2}$	$4-\frac{2\,(2\,\mathcal{R}+1)}{(\mathcal{R}+1)^3}$	$4-\frac{\mathcal{R}+2}{(\mathcal{R}+1)^2}$

Probability of passing through transient states:

$$
\begin{aligned}
\mathcal{P}_{SSII} &= \frac{\mathcal{R}}{\mathcal{R}+1}\mathrm{E}_{SSSI} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{SISS} \\
\mathcal{P}_{SIIS} &= \frac{\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{SISS} \\
\mathcal{P}_{SRIS} &= \frac{2}{2\mathcal{R}+2}\mathcal{P}_{SIIS} \\
\mathcal{P}_{ISSR} &= \frac{2}{2\mathcal{R}+2}\mathrm{E}_{ISSI} \\
\mathcal{P}_{SSIR} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{SSII} \\
\mathcal{P}_{SSRI} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{SSII} \\
\mathcal{P}_{SIII} &= \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{SSII} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{SIIS} + \frac{2\mathcal{R}}{3\mathcal{R}+2}\mathrm{E}_{SISI} \\
\mathcal{P}_{ISII} &= \frac{\mathcal{R}}{3\mathcal{R}+2}\mathrm{E}_{SISI} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathrm{E}_{ISSI} \\
\mathcal{P}_{IIII} &= \frac{\mathcal{R}}{\mathcal{R}+3}\mathcal{P}_{SIII} + \frac{2\mathcal{R}}{2\mathcal{R}+3}\mathcal{P}_{ISII} \\
\mathcal{P}_{SRSI} &= \frac{1}{3\mathcal{R}+2}\mathrm{E}_{SISI} \\
\mathcal{P}_{SISR} &= \frac{1}{3\mathcal{R}+2}\mathrm{E}_{SISI} \\
\mathcal{P}_{SIIR} &= \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SSIR} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{SISR} + \frac{1}{\mathcal{R}+3}\mathcal{P}_{SIII} \\
\mathcal{P}_{SIRI} &= \frac{1}{\mathcal{R}+3}\mathcal{P}_{SIII} \\
\mathcal{P}_{SRII} &= \frac{1}{\mathcal{R}+3}\mathcal{P}_{SIII} + \frac{\mathcal{R}}{\mathcal{R}+1}(\mathcal{P}_{SRIS} + \mathcal{P}_{SRSI}) \\
\mathcal{P}_{IISR} &= \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{ISSR} + \frac{\mathcal{R}}{2\mathcal{R}+1}\mathcal{P}_{SISR} + \frac{1}{2\mathcal{R}+3}\mathcal{P}_{IISI} \\
\mathcal{P}_{ISRI} &= \frac{1}{2\mathcal{R}+3}\mathcal{P}_{IISI} \\
\mathcal{P}_{ISIR} &= \frac{1}{2\mathcal{R}+3}\mathcal{P}_{IISI} \\
\mathcal{P}_{SRIR} &= \frac{1}{2}\mathcal{P}_{SRII} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{SIIR} \\
\mathcal{P}_{SIRR} &= \frac{1}{\mathcal{R}+2}(\mathcal{P}_{SIIR} + \mathcal{P}_{SIRI}) \\
\mathcal{P}_{SRRI} &= \frac{1}{2}\mathcal{P}_{SRII} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{SIRI} \\
\mathcal{P}_{RSIR} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{IISR} + \frac{1}{2\mathcal{R}+2}\mathcal{P}_{ISIR} \\
\mathcal{P}_{ISRR} &= \frac{1}{\mathcal{R}+2}\mathcal{P}_{ISRI} + \frac{1}{2\mathcal{R}+2}\mathcal{P}_{ISIR} \\
\mathcal{P}_{RSRI} &= \frac{1}{\mathcal{R}+2}(\mathcal{P}_{ISRI} + \mathcal{P}_{IISR}) \\
\mathcal{P}_{IIIR} &= \frac{\mathcal{R}}{\mathcal{R}+2}(\mathcal{P}_{SIIR} + \mathcal{P}_{IISR}) + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathcal{P}_{ISIR} + \frac{1}{2}\mathcal{P}_{IIII}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{P}_{IIRI} &= \frac{1}{2}\mathcal{P}_{IIII} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{SIRI} + \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{ISRI} \\
\mathcal{P}_{IRRI} &= \frac{1}{3}\mathcal{P}_{IIRI} \\
\mathcal{P}_{IIRR} &= \frac{1}{3}(\mathcal{P}_{IIRI} + \mathcal{P}_{IIIR}) + \frac{\mathcal{R}}{\mathcal{R}+1}(\mathcal{P}_{SIRR} + \mathcal{P}_{ISRR}) \\
\mathcal{P}_{IRIR} &= \frac{1}{3}(\mathcal{P}_{IIRI} + \mathcal{P}_{IIIR}) \\
\mathcal{P}_{RIIR} &= \frac{1}{3}\mathcal{P}_{IIRI} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{RSIR} \\
\mathcal{P}_{IRRR} &= \frac{1}{2}(\mathcal{P}_{IIRR} + \mathcal{P}_{IRIR}) + \mathcal{P}_{IRRI} \\
\mathcal{P}_{RIRR} &= \frac{1}{2}(\mathcal{P}_{IIRR} + \mathcal{P}_{IRIR}) + \mathcal{P}_{RIIR}
\end{aligned}
$$

Probability of terminating in absorbing states:

$$
\begin{aligned}
\mathcal{P}_{SSSR} &= \frac{1}{\mathcal{R}+1}\mathrm{E}_{SSSI} \\
\mathcal{P}_{SRSS} &= \frac{1}{2\mathcal{R}+1}\mathrm{E}_{SISS} \\
\mathcal{P}_{SRRS} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SRIS} \\
\mathcal{P}_{RSSR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{ISSR} \\
\mathcal{P}_{SSRR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SSIR} + \mathcal{P}_{SSRI} \\
\mathcal{P}_{SRSR} &= \frac{1}{\mathcal{R}+1}\mathcal{P}_{SRSI} + \frac{1}{2\mathcal{R}+1}\mathcal{P}_{SISR} \\
\mathcal{P}_{SRRR} &= \mathcal{P}_{SRIR} + \mathcal{P}_{SRRI} + \frac{1}{\mathcal{R}+1}\mathcal{P}_{SIRR} \\
\mathcal{P}_{RSRR} &= \mathcal{P}_{RSRI} + \frac{1}{\mathcal{R}+1}(\mathcal{P}_{ISRR} + \mathcal{P}_{RSIR}) \\
\mathcal{P}_{RRRR} &= \mathcal{P}_{IRRR} + \mathcal{P}_{RIRR}
\end{aligned}
$$

To find the equations for the final size probabilities we evaluated the following:

$$
\begin{aligned}
\mathbb{P}(\text{Final Size} = 1) &= \mathcal{P}_{SSSR} + \mathcal{P}_{SRSS} \\
\mathbb{P}(\text{Final Size} = 2) &= \mathcal{P}_{SRRS} + \mathcal{P}_{RSSR} + \mathcal{P}_{SSRR} + \mathcal{P}_{SRSR} \\
\mathbb{P}(\text{Final Size} = 3) &= \mathcal{P}_{SRRR} + \mathcal{P}_{RSRR} \\
\mathbb{P}(\text{Final Size} = 4) &= \mathcal{P}_{RRRR}
\end{aligned}
$$

Simplifying the above we obtained the final size equations for the line network with four nodes as shown in Table A.6. The probability mass function of the final size distribution for the line network is shown in Figure 2.13.

A.6.2 Progression of infection over time

Equations describing the probability that the network is in a given state at time t for an SIR model on the line network of $N = 4$ are:

Initial states:

$$\begin{aligned}
\dot{\mathsf{P}}_{SSSS} &= 0 \\
\dot{\mathsf{P}}_{SSSI} &= -(\mathcal{R}+1)\mathsf{P}_{SSSI} \\
\dot{\mathsf{P}}_{SSIS} &= -(2\mathcal{R}+1)\mathsf{P}_{SSIS} \\
\dot{\mathsf{P}}_{SISI} &= -(3\mathcal{R}+2)\mathsf{P}_{SISI} \\
\dot{\mathsf{P}}_{ISSI} &= -2(\mathcal{R}+1)\mathsf{P}_{ISSI}
\end{aligned} \tag{A.4}$$

Transient states:

$$\begin{aligned}
\dot{\mathsf{P}}_{SSII} &= \mathcal{R}(\mathsf{P}_{SSSI}+\mathsf{P}_{SSIS}) - (\mathcal{R}+2)\mathsf{P}_{SSII} \\
\dot{\mathsf{P}}_{SIIS} &= \mathcal{R}\mathsf{P}_{SSIS} - 2(\mathcal{R}+1)\mathsf{P}_{SIIS} \\
\dot{\mathsf{P}}_{SRIS} &= 2\mathsf{P}_{SIIS} - (\mathcal{R}+1)\mathsf{P}_{SRIS} \\
\dot{\mathsf{P}}_{ISSR} &= 2\mathsf{P}_{ISSI} - (\mathcal{R}+1)\mathsf{P}_{ISSR} \\
\dot{\mathsf{P}}_{SSIR} &= \mathsf{P}_{SSII} - (\mathcal{R}+1)\mathsf{P}_{SSIR} \\
\dot{\mathsf{P}}_{SSRI} &= \mathsf{P}_{SSII} - \mathsf{P}_{SSRI} \\
\dot{\mathsf{P}}_{SRSI} &= \mathsf{P}_{SISI} - (\mathcal{R}+1)\mathsf{P}_{SRSI} \\
\dot{\mathsf{P}}_{SISR} &= \mathsf{P}_{SISI} - (2\mathcal{R}+1)\mathsf{P}_{SISR} \\
\dot{\mathsf{P}}_{SIII} &= \mathcal{R}\mathsf{P}_{SSII} + 2\mathcal{R}(\mathsf{P}_{SIIS}+\mathsf{P}_{SISI}) - (\mathcal{R}+3)\mathsf{P}_{SIII} \\
\dot{\mathsf{P}}_{IISI} &= \mathcal{R}\mathsf{P}_{SISI} + 2\mathcal{R}\mathsf{P}_{ISSI} - (2\mathcal{R}+3)\mathsf{P}_{IISI} \\
\dot{\mathsf{P}}_{IIII} &= \mathcal{R}\mathsf{P}_{SIII} + 2\mathcal{R}\mathsf{P}_{IISI} - 4\mathsf{P}_{IIII} \\
\dot{\mathsf{P}}_{SIIR} &= \mathcal{R}(\mathsf{P}_{SSIR}+\mathsf{P}_{SISR}) + \mathsf{P}_{SIII} - (\mathcal{R}+2)\mathsf{P}_{SIIR} \\
\dot{\mathsf{P}}_{SIRI} &= \mathsf{P}_{SIII} - (\mathcal{R}+2)\mathsf{P}_{SIRI} \\
\dot{\mathsf{P}}_{SRII} &= \mathsf{P}_{SIII} + \mathcal{R}(\mathsf{P}_{SRIS}+\mathsf{P}_{SRSI}) - 2\mathsf{P}_{SRII} \\
\dot{\mathsf{P}}_{IISR} &= \mathsf{P}_{IISI} + \mathcal{R}(\mathsf{P}_{SISR}+\mathsf{P}_{ISSR}) - (\mathcal{R}+2)\mathsf{P}_{IISR} \\
\dot{\mathsf{P}}_{ISRI} &= \mathsf{P}_{IISI} - (\mathcal{R}+2)\mathsf{P}_{ISRI} \\
\dot{\mathsf{P}}_{ISIR} &= \mathsf{P}_{IISI} - (2\mathcal{R}+2)\mathsf{P}_{ISIR} \\
\dot{\mathsf{P}}_{SRIR} &= \mathsf{P}_{SIIR} + \mathsf{P}_{SRII} - \mathsf{P}_{SRIR} \\
\dot{\mathsf{P}}_{SIRR} &= \mathsf{P}_{SIIR} + \mathsf{P}_{SIRI} - (\mathcal{R}+1)\mathsf{P}_{SIRR} \\
\dot{\mathsf{P}}_{SRRI} &= \mathsf{P}_{SRII} + \mathsf{P}_{SIRI} - \mathsf{P}_{SRRI} \\
\dot{\mathsf{P}}_{RSIR} &= \mathsf{P}_{IISR} + \mathsf{P}_{ISIR} - (\mathcal{R}+1)\mathsf{P}_{RSIR} \\
\dot{\mathsf{P}}_{ISRR} &= \mathsf{P}_{ISIR} + \mathsf{P}_{ISRI} - (\mathcal{R}+1)\mathsf{P}_{ISRR} \\
\dot{\mathsf{P}}_{RSRI} &= \mathsf{P}_{ISRI} + \mathsf{P}_{IISR} - \mathsf{P}_{RSRI}
\end{aligned}$$

$$
\begin{aligned}
\dot{\mathsf{P}}_{IIIR} &= 2\mathsf{P}_{IIII} + \mathcal{R}(\mathsf{P}_{SIIR} + \mathsf{P}_{IISR} + 2\mathsf{P}_{ISIR}) - 3\mathsf{P}_{IIIR} \\
\dot{\mathsf{P}}_{IIRI} &= 2\mathsf{P}_{IIII} + \mathcal{R}(\mathsf{P}_{SIRI} + \mathsf{P}_{ISRI}) - 3\mathsf{P}_{IIRI} \\
\dot{\mathsf{P}}_{IRRI} &= \mathsf{P}_{IIRI} - 2\mathsf{P}_{IRRI} \\
\dot{\mathsf{P}}_{IIRR} &= \mathsf{P}_{IIIR} + \mathsf{P}_{IIRI} + \mathcal{R}(\mathsf{P}_{SIRR} + \mathsf{P}_{ISRR}) - 2\mathsf{P}_{IIRR} \\
\dot{\mathsf{P}}_{IRIR} &= \mathsf{P}_{IIIR} + \mathsf{P}_{IIRI} - 2\mathsf{P}_{IRIR} \\
\dot{\mathsf{P}}_{RIIR} &= \mathsf{P}_{IIIR} + \mathcal{R}\mathsf{P}_{RSIR} - 2\mathsf{P}_{RIIR} \\
\dot{\mathsf{P}}_{IRRR} &= 2\mathsf{P}_{IRRI} + \mathsf{P}_{IIRR} + \mathsf{P}_{IRIR} - \mathsf{P}_{IRRR} \\
\dot{\mathsf{P}}_{RIRR} &= 2\mathsf{P}_{RIIR} + \mathsf{P}_{IIRR} + \mathsf{P}_{IRIR} - \mathsf{P}_{RIRR}
\end{aligned}
$$

Absorbing states:

$$
\begin{aligned}
\dot{\mathsf{P}}_{SSSR} &= \mathsf{P}_{SSSI} \\
\dot{\mathsf{P}}_{SSRS} &= \mathsf{P}_{SSIS} \\
\dot{\mathsf{P}}_{SRRS} &= \mathsf{P}_{SRIS} \\
\dot{\mathsf{P}}_{RSSR} &= \mathsf{P}_{ISSR} \\
\dot{\mathsf{P}}_{SSRR} &= \mathsf{P}_{SSIR} + \mathsf{P}_{SSRI} \\
\dot{\mathsf{P}}_{SRSR} &= \mathsf{P}_{SRSI} + \mathsf{P}_{SISR} \\
\dot{\mathsf{P}}_{SRRR} &= \mathsf{P}_{SRIR} + \mathsf{P}_{SIRR} + \mathsf{P}_{SRRI} \\
\dot{\mathsf{P}}_{RSRR} &= \mathsf{P}_{RSIR} + \mathsf{P}_{ISRR} + \mathsf{P}_{RSRI} \\
\dot{\mathsf{P}}_{RRRR} &= \mathsf{P}_{IRRR} + \mathsf{P}_{RIRR}
\end{aligned}
$$

To find the final size probabilities we numerically solve the system of differential equations with specified initial conditions and sum up the appropriate absorbing state probabilities once a steady state has been reached. See Figure A.12 for numerical results that confirm our analytical expressions for the final epidemic size.

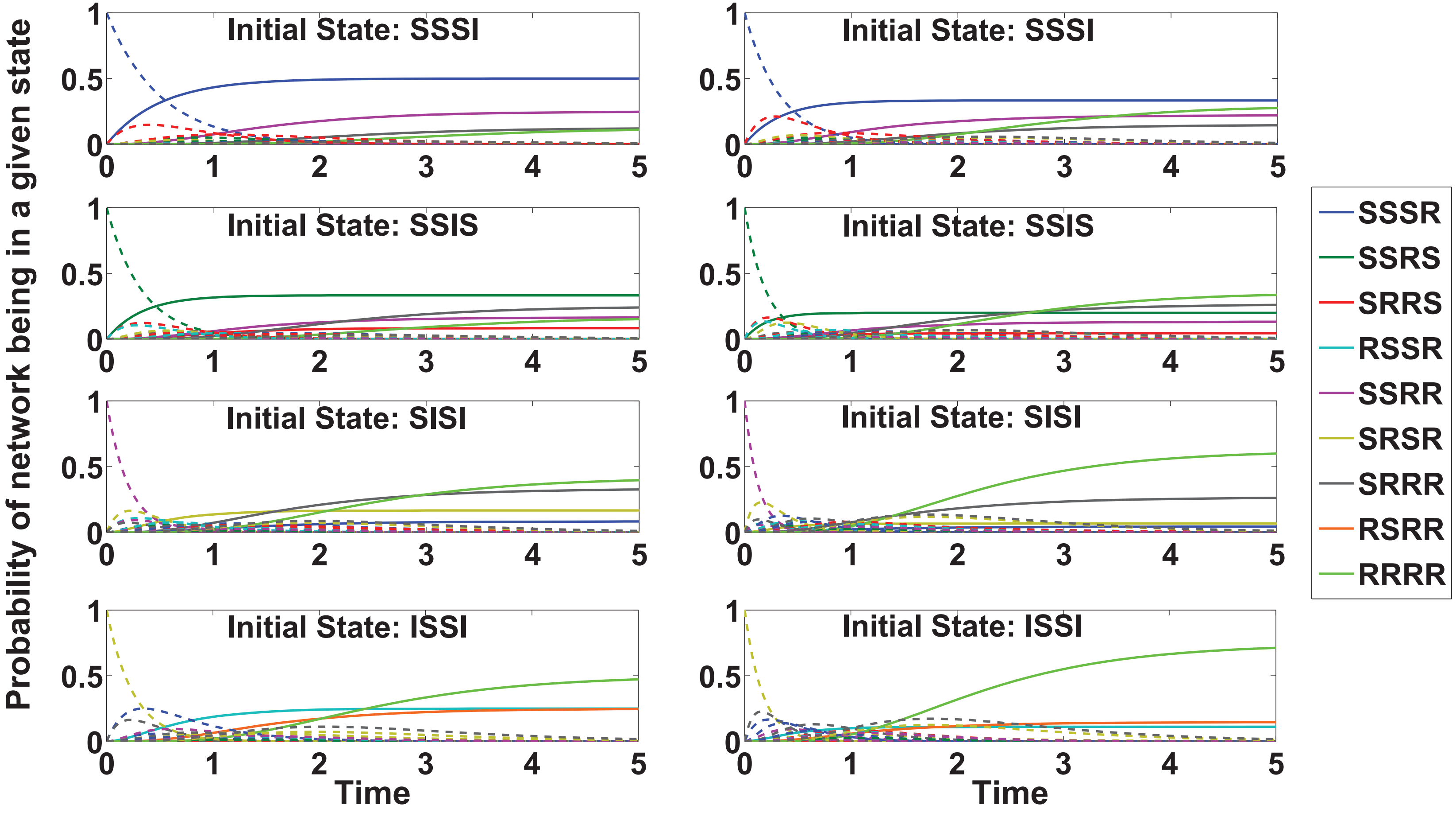


Figure A.12: Numerical solution of the system of differential equations (A.4) which describes the progression of infection over time for an SIR model on a line network with $N = 4$ nodes. Left and right columns contain graphical results for $\mathcal{R} = 1$ and $\mathcal{R} = 2$ respectively for the specified initial conditions. Dashed lines represent transient states and solid lines are the absorbing states. The numerical results are in agreement with the analytical expressions for the same set of initial conditions.

Appendix B

B.1 Comparison of analytic and stochastic results:

Here we show tables and graphs comparing the analytic and stochastic results for a LoT network. All results given here are for starting the epidemic with $I_0 = a$. The stochastic result given in each table was found by averaging over the results of 2×10^5 realisations using the Gillespie algorithm for a stochastic SIR model. Table B.1 compares the analytic and stochastic results for the final size probabilities for a LoT(6) network. Results for the probability each node is ever infected in the LoT(6) network are given in Table B.2.

Table B.1: Stochastic Vs Analytic Final Size Results for LoT Network with $N = 6$ nodes

Final Size	**Parameters**	**Stochastic**	**Analytic**
1	$\mathcal{R} = 1$	0.3335	0.3333
	$\mathcal{R} = 2$	0.1994	0.2000
2	$\mathcal{R} = 1$	0.1383	0.1389
	$\mathcal{R} = 2$	0.0707	0.0711
3	$\mathcal{R} = 1$	0.2460	0.2454
	$\mathcal{R} = 2$	0.2280	0.2287
4	$\mathcal{R} = 1$	0.0932	0.0926
	$\mathcal{R} = 2$	0.0996	0.0988
5	$\mathcal{R} = 1$	0.0571	0.0579
	$\mathcal{R} = 2$	0.0560	0.0558
6	$\mathcal{R} = 1$	0.1318	0.1319
	$\mathcal{R} = 2$	0.3463	0.3455

Table B.3 compares analytic and stochastic results for the probability the infection ends at each node for the LoT(6) and LoT(9) networks, given that the initial infectious node, $I_0 = a$. The absolute value of the difference between analytic and stochastic results for the probability the final size equals N for LoT networks

Table B.2: Probability that each node ever becomes infected in the LoT network with $N = 6$.

Probability	**Parameters**	**Stochastic**	**Analytic**
P_a	$\mathcal{R} = 1$	1	1
	$\mathcal{R} = 2$	1	1
P_b	$\mathcal{R} = 1$	0.5837	0.5833
	$\mathcal{R} = 2$	0.7558	0.7556
P_c	$\mathcal{R} = 1$	0.5836	0.5833
	$\mathcal{R} = 2$	0.7556	0.7556
P_d	$\mathcal{R} = 1$	0.2931	0.2917
	$\mathcal{R} = 2$	0.5035	0.5037
P_e	$\mathcal{R} = 1$	0.1711	0.1701
	$\mathcal{R} = 2$	0.3797	0.3806
P_f	$\mathcal{R} = 1$	0.1710	0.1701
	$\mathcal{R} = 2$	0.3795	0.3806

with $N = 6, 9, 18, 36, 81, 102$ are shown in Figure B.1. Figures 3.19 and B.2 show the convergence of the absolute value of the difference between the analytic and stochastic probabilities for the infection ending at node x for a LoT network with $N = 6, 9, 12, 15, 18, 21$. The sum of squares error (SSE) for the probability k apex nodes are infected given that the infection ends at node j in the LoT(21) network is shown in Figure 3.20.

Table B.3: The probability infection reaches each node and then stops, given $I_0 = a$.

$P_{end}(x)$ is the probability that the infection ends at node x.

LoT(6)				
	$\mathcal{R} = 1$		$\mathcal{R} = 2$	
	Analytic	**Stochastic**	**Analytic**	**Stochastic**
$P_{end}(a)$	0.3333	0.3335	0.2	0.2002
$P_{end}(b)$	0.0833	0.0835	0.0444	0.0446
$P_{end}(c)$	0.2917	0.2923	0.2519	0.2531
$P_{end}(d)$	0.0972	0.0965	0.1007	0.1001
$P_{end}(e)$	0.0243	0.0243	0.0224	0.0221
$P_{end}(f)$	0.1701	0.1702	0.3806	0.3807

LoT(9)				
	$\mathcal{R} = 1$		$\mathcal{R} = 2$	
	Analytic	**Stochastic**	**Analytic**	**Stochastic**
$P_{end}(a)$	0.3333	0.3335	0.2	0.2002
$P_{end}(b)$	0.0833	0.0835	0.0444	0.0446
$P_{end}(c)$	0.2917	0.2923	0.2519	0.2531
$P_{end}(d)$	0.0972	0.0965	0.1007	0.1001
$P_{end}(e)$	0.0243	0.0243	0.0224	0.0221
$P_{end}(f)$	0.0851	0.0844	0.1269	0.1253
$P_{end}(g)$	0.0284	0.0277	0.0507	0.0506
$P_{end}(h)$	0.0071	0.0070	0.0113	0.0116
$P_{end}(i)$	0.0496	0.0508	0.1917	0.1925

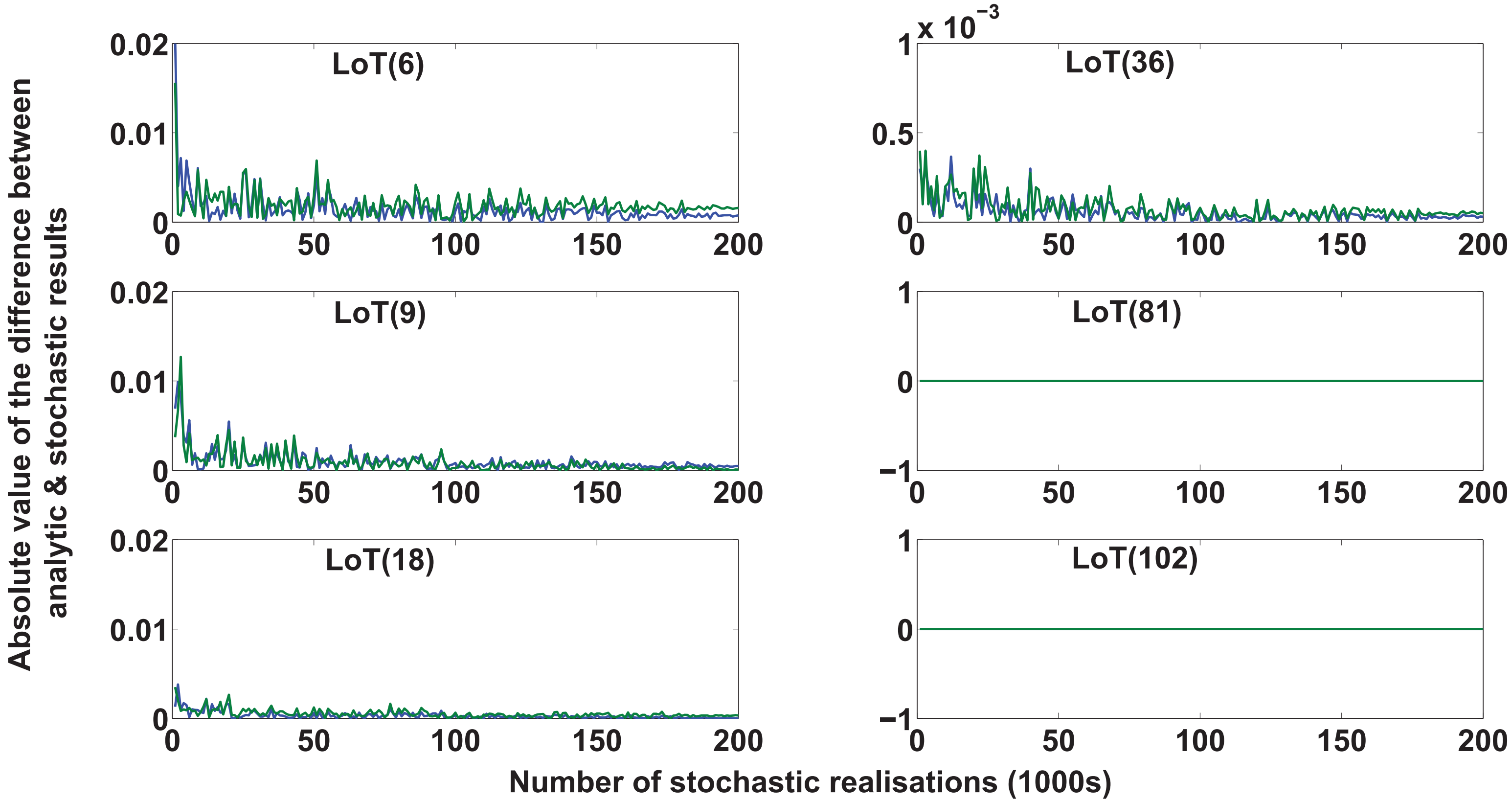


Figure B.1: Blue: shows the absolute value of the difference between the analytic and stochastic results for the probability that the final size is equal to N for increasing number of stochastic realisations. Green: shows the absolute value of the difference between the analytic and stochastic results for the probability that the infection reaches the N^{th} node for increasing number of stochastic realisations.

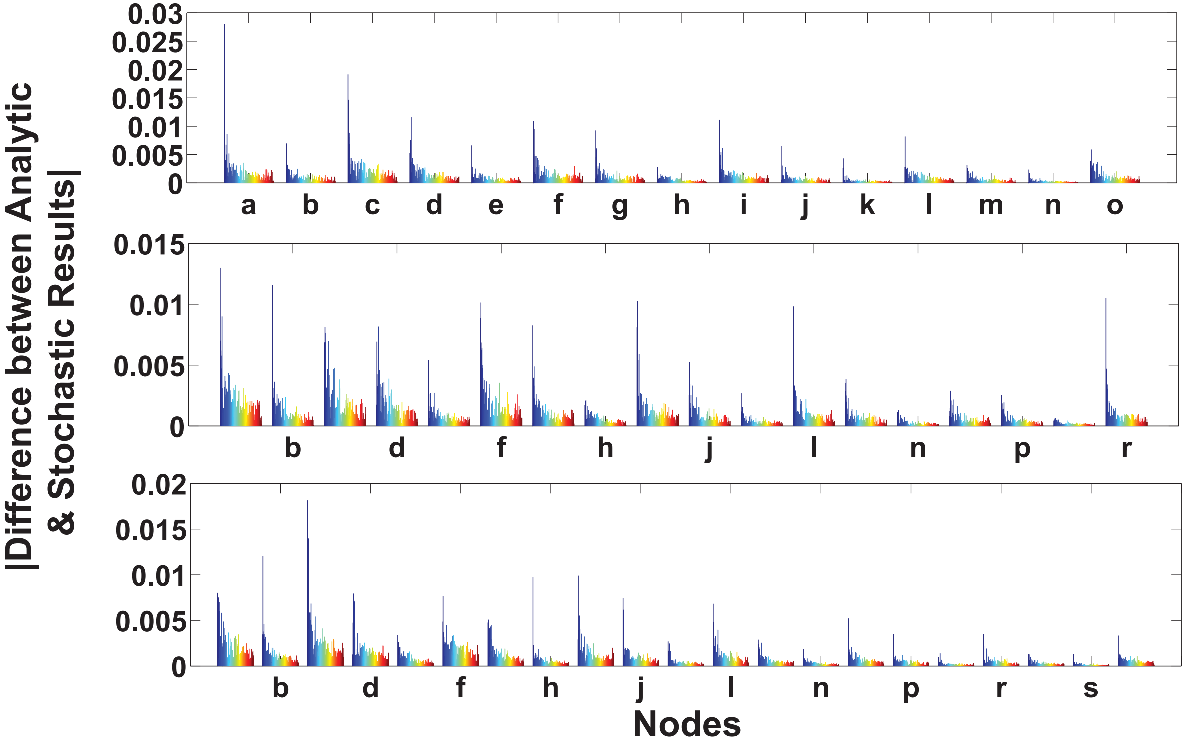


Figure B.2: Absolute value of the difference between analytic and stochastic results for the probability infection ends at each node in a LoT network. For each node, from left to right (blue to red) we are increasing the number of stochastic realisations (from 1×10^3 to 2×10^5 in steps of 1×10^3) that our stochastic probabilities were averaged over. Thus, for each node we are looking for the difference to converge to zero if the accuracy of the stochastic realisation result gets closer to the exact analytic probability. Top: LoT(15) network. Middle: LoT(18) network. Bottom: LoT(21) network.

Appendix C

Here we present the results for an SIS model on the remaining small networks, however we omit the derivation of results and refer the reader to Chapter 5 in the main text for full details of the methods used.

C.1 Line Network, $N = 3$

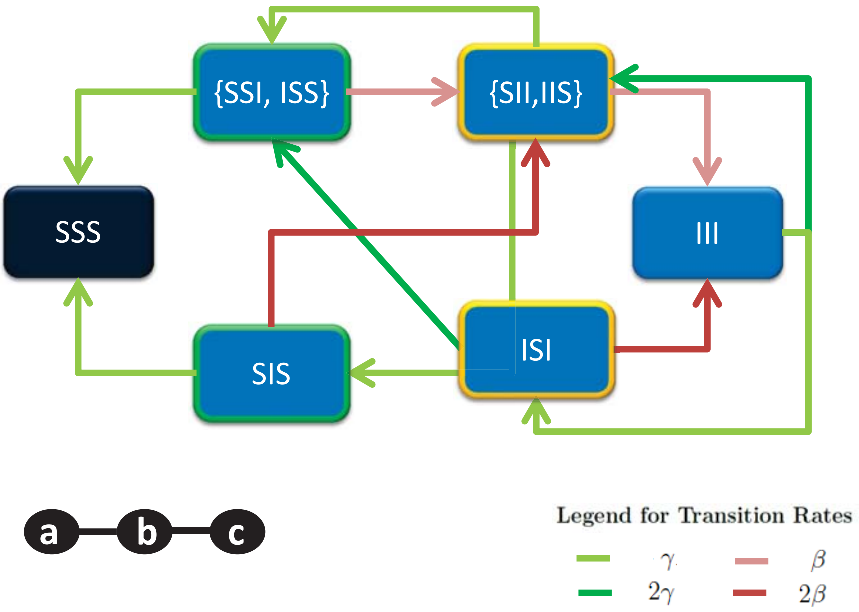


Figure C.1: Transition state diagram for the SIS model on a line network of $N = 3$ nodes.

Here we have a line network of 3 nodes, the full state space for an SIS model on this network is $2^N = 8$. As we now have two different types of nodes (two end nodes of degree one and one central node of degree two) we must consider the topology of the network as well as the number of nodes in each infection state when reducing the system via the lumping technique. From Figure C.1 it can be seen that the reduced system has the following state space $\mathbf{S} = \{SSI, SIS, SII, ISI, III, SSS\}$. The following matrices are indexed in the same order as $\mathbf{S}$.

The equations describing an SIS epidemic on a line network with 3 nodes are:

$$\begin{aligned}
\dot{P}_{SSI} &= \gamma P_{SII} + 2\gamma P_{ISI} - (\beta+\gamma)P_{SSI} \\
\dot{P}_{SIS} &= \gamma P_{SII} - (2\beta+\gamma)P_{SIS} \\
\dot{P}_{SII} &= 2\beta P_{SIS} + 2\gamma P_{III} + \beta P_{SSI} - (2\gamma+\beta)P_{SII} \\
\dot{P}_{ISI} &= \gamma P_{III} - 2(\beta+\gamma)P_{ISI} \\
\dot{P}_{III} &= 2\beta P_{ISI} + \beta P_{SII} - 3\gamma P_{III} \\
\dot{P}_{SSS} &= \gamma(P_{SSI} + P_{SIS})
\end{aligned}$$

The following matrix, $\mathbf{Q}$, contains the transition probabilities for the *transient states only* of the line network.

$$\mathbf{Q} = \begin{pmatrix}
0 & 0 & \dfrac{\beta}{\beta+\gamma} & 0 & 0 \\
0 & 0 & \dfrac{2\beta}{2\beta+\gamma} & 0 & 0 \\
\dfrac{\gamma}{\beta+2\gamma} & \dfrac{\gamma}{\beta+2\gamma} & 0 & 0 & \dfrac{\beta}{\beta+2\gamma} \\
\dfrac{\gamma}{\beta+\gamma} & 0 & 0 & 0 & \dfrac{\beta}{\beta+\gamma} \\
0 & 0 & \dfrac{2}{3} & \dfrac{1}{3} & 0
\end{pmatrix}$$

Expected number of visits to each state

The expected number of times the system visits each state from a given initial state is given by the matrix $\mathbf{N}$:

$$
\mathbf{N} = \begin{pmatrix}
1 & 0 & -\frac{\beta}{\beta+1} & 0 & 0 \\
0 & 1 & -\frac{2\beta}{2\beta+1} & 0 & 0 \\
-\frac{1}{\beta+2} & -\frac{1}{\beta+2} & 1 & 0 & -\frac{\beta}{\beta+2} \\
-\frac{1}{\beta+1} & 0 & 0 & 1 & -\frac{\beta}{\beta+1} \\
0 & 0 & -\frac{2}{3} & -\frac{1}{3} & 1
\end{pmatrix}^{-1}
$$

$$
= \begin{pmatrix}
\frac{6\beta^2+11\beta+6}{2(4\beta+3)} & \frac{\beta(2\beta+1)(2\beta+3)}{2(\beta+1)(4\beta+3)} & \frac{\beta(2\beta+1)(2\beta+3)(\beta+2)}{2(\beta+1)(4\beta+3)} & \frac{\beta^2(2\beta+1)}{2(4\beta+3)} & \frac{3\beta^2(2\beta+1)}{2(4\beta+3)} \\
\frac{3\beta(\beta+1)}{4\beta+3} & \frac{(2\beta+1)(\beta+3)}{4\beta+3} & \frac{\beta(2\beta+3)(\beta+2)}{4\beta+3} & \frac{\beta^2(\beta+1)}{4\beta+3} & \frac{3\beta^2(\beta+1)}{4\beta+3} \\
\frac{3(\beta+1)(2\beta+1)}{2(4\beta+3)} & \frac{(2\beta+1)(2\beta+3)}{2(4\beta+3)} & \frac{(2\beta+1)(2\beta+3)(\beta+2)}{2(4\beta+3)} & \frac{\beta(\beta+1)(2\beta+1)}{2(4\beta+3)} & \frac{3\beta(\beta+1)(2\beta+1)}{2(4\beta+3)} \\
\frac{3(2\beta^2+3\beta+2)}{2(4\beta+3)} & \frac{\beta(2\beta+1)(2\beta+5)}{2(\beta+1)(4\beta+3)} & \frac{\beta(2\beta+1)(2\beta+5)(\beta+2)}{2(\beta+1)(4\beta+3)} & \frac{2\beta^3+3\beta^2+10\beta+6}{2(4\beta+3)} & \frac{3\beta(2\beta^2+3\beta+2)}{2(4\beta+3)} \\
\frac{6\beta^2+9\beta+4}{2(4\beta+3)} & \frac{(2\beta+1)^2(\beta+2)}{2(\beta+1)(4\beta+3)} & \frac{(2\beta+1)^2(\beta+2)^2}{2(\beta+1)(4\beta+3)} & \frac{2\beta^3+3\beta^2+4\beta+2}{2(4\beta+3)} & \frac{3(2\beta^3+3\beta^2+4\beta+2)}{2(4\beta+3)}
\end{pmatrix}
$$

Expected time spent in each state during the epidemic

We can determine the total expected time spent in each state for the duration of the epidemic as follows,

$$E[\text{time in given state}] = E[\#\text{visits to given state}] \times [\text{Duration of visits}]$$

The diagonal matrix, $\mathbf{T_S}$, contains entries which give the duration of each visit to the states in the line network. The entries of $\mathbf{T_S}$ are in the same order as $\mathbf{S} = \{SSI, SIS, SII, ISI, III, SSS\}$, excluding the state SSS.

$$\mathbf{T_S} \;=\; \frac{1}{\gamma} \times \mathsf{diag}\left(\frac{\gamma}{\beta+\gamma}, \frac{\gamma}{2\beta+\gamma}, \frac{\gamma}{\beta+2\gamma}, \frac{\gamma}{2\beta+2\gamma}, \frac{1}{3}]\right)$$

Therefore to find the expected time spent in each state throughout the epidemic we compute:

$$\mathbf{E_T} \;=\; \mathbf{N T_S}$$

Here the element $\mathbf{E_T}(i, j)$ gives the expected time spent in state j given that state i was the initial starting state. We give the algebraic expression for $\mathbf{E_T}$ on page 246.

If the epidemic starts in the state SSI (SIS) when the transmission rate β is increased and keeping the recovery rate fixed, the expected time spent in SSI (SIS) decreases and the expected time spent in states SIS (SSI), ISI, SII and III all increase. If the system starts in the state SSI or SIS when γ is decreased and β fixed, the expected time spent in states SSI, SIS, SII and ISI increases while the expected time spent in state III decreases.

$$\mathbf{E_T} = \frac{1}{\gamma}\begin{pmatrix} \frac{\left(6\,\beta^2+11\,\beta\,\gamma+6\,\gamma^2\right)}{2\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{\beta\,(2\,\beta+3\,\gamma)}{2\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{\beta\,(2\,\beta+\gamma)\,(2\,\beta+3\,\gamma)}{2\gamma\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{\beta^2\,(2\,\beta+\gamma)}{4\gamma\,(\beta+\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta^2\,(2\,\beta+\gamma)}{2\gamma^2\,(4\,\beta+3\,\gamma)} \\ \frac{3\beta}{(4\,\beta+3\,\gamma)} & \frac{(\beta+3\,\gamma)}{(4\,\beta+3\,\gamma)} & \frac{\beta\,(2\,\beta+3\,\gamma)}{\gamma\,(4\,\beta+3\,\gamma)} & \frac{\beta^2}{2\gamma\,(4\,\beta+3\,\gamma)} & \frac{\beta^2\,(\beta+\gamma)}{\gamma^2\,(4\,\beta+3\,\gamma)} \\ \frac{3\,(2\,\beta+\gamma)}{2\,(4\,\beta+3\,\gamma)} & \frac{(2\,\beta+3\,\gamma)}{2\,(4\,\beta+3\,\gamma)} & \frac{(2\,\beta+\gamma)\,(2\,\beta+3\,\gamma)}{2\gamma\,(4\,\beta+3\,\gamma)} & \frac{\beta\,(2\,\beta+\gamma)}{4\gamma\,(4\,\beta+3\,\gamma)} & \frac{\beta\,(2\,\beta+\gamma)\,(\beta+\gamma)}{2\gamma^2\,(4\,\beta+3\,\gamma)} \\ \frac{3\,\left(2\,\beta^2+3\,\beta\,\gamma+2\,\gamma^2\right)}{2\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{\beta\,(2\,\beta+5\,\gamma)}{2\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{\beta\,(2\,\beta+\gamma)\,(2\,\beta+5\,\gamma)}{2\gamma\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{\left(2\,\beta^3+3\,\beta^2\,\gamma+10\,\beta\,\gamma^2+6\,\gamma^3\right)}{4\gamma\,(\beta+\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta\,\left(2\,\beta^2+3\,\beta\,\gamma+2\,\gamma^2\right)}{2\gamma^2\,(4\,\beta+3\,\gamma)} \\ \frac{\left(6\,\beta^2+9\,\beta\,\gamma+4\,\gamma^2\right)}{2\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{(2\,\beta+\gamma)\,(\beta+2\,\gamma)}{2\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{(2\,\beta+\gamma)^2\,(\beta+2\,\gamma)}{2\gamma\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{\left(2\,\beta^3+3\,\beta^2\,\gamma+4\,\beta\,\gamma^2+2\,\gamma^3\right)}{4\gamma\,(\beta+\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\left(2\,\beta^3+3\,\beta^2\,\gamma+4\,\beta\,\gamma^2+2\,\gamma^3\right)}{2\gamma^2\,(4\,\beta+3\,\gamma)} \end{pmatrix}$$

Time to absorption

In the following vector $\mathbf{T_A}$ the entries $\mathbf{T_A}(i)$ gives the expected time to absorption given that the epidemic started in state i.

$$\mathbf{T_A} = \frac{1}{\gamma}\left(\begin{array}{c} \dfrac{\left(4\,\beta^4+16\,\beta^3\,\gamma+35\,\beta^2\,\gamma^2+34\,\beta\,\gamma^3+12\,\gamma^4\right)}{4\gamma^2\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} \\ \\ \dfrac{\left(2\,\beta^3+7\,\beta^2\,\gamma+14\,\beta\,\gamma^2+6\,\gamma^3\right)}{\gamma^2\,(4\,\beta+3\,\gamma)} \\ \\ \dfrac{\left(4\,\beta^3+16\,\beta^2\,\gamma+35\,\beta\,\gamma^2+18\,\gamma^3\right)}{2\gamma^2\,(4\,\beta+3\,\gamma)} \\ \\ \dfrac{\left(4\,\beta^4+20\,\beta^3\,\gamma+53\,\beta^2\,\gamma^2+52\,\beta\,\gamma^3+18\,\gamma^4\right)}{4\gamma^2\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} \\ \\ \dfrac{\left(4\,\beta^4+20\,\beta^3\,\gamma+57\,\beta^2\,\gamma^2+62\,\beta\,\gamma^3+22\,\gamma^4\right)}{4\gamma^2\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} \end{array}\right)$$

Number of times each node is infected

Recalling that we have:

$$\text{Number of times node } k \text{ is infected} = \gamma\mathcal{T}(k|i)$$

We give the algebraic expression of the matrix $\mathcal{T}(k|i)$ on page 248. In the line network, for example, to find the expected number of times node a is infectious given that the epidemic started in state SSI (where node c is infected) we add up the expected time spent in all states where node a is infected (ISS, ISI, IIS and III). Each row of $\mathcal{T}(k|i)$ corresponds to the expected time each node is infected given that the epidemic started in states SSI, SIS and ISI respectively. Figure 5.6 on page 155 shows bar charts of the expected number of times each node is infected for the line network with $N = 3$ nodes.

$$\mathcal{T}(k|i) = \begin{pmatrix} \frac{20\,\beta^6+86\,\beta^5\,\gamma+192\,\beta^4\,\gamma^2+271\,\beta^3\,\gamma^3+238\,\beta^2\,\gamma^4+116\,\beta\,\gamma^5+24\,\gamma^6}{4\,\gamma^2\,\left(4\,\beta+3\,\gamma\right)\,\left(\beta+\gamma\right)\,\left(\beta^2+2\,\beta\,\gamma+2\,\gamma^2\right)} & \frac{\beta\,\left(2\,\beta+\gamma\right)\,\left(5\,\beta^2+12\,\beta\,\gamma+9\,\gamma^2\right)}{2\,\gamma^2\,\left(4\,\beta+3\,\gamma\right)\,\left(\beta+\gamma\right)} & \frac{\beta^2\,\left(20\,\beta^4+86\,\beta^3\,\gamma+160\,\beta^2\,\gamma^2+135\,\beta\,\gamma^3+42\,\gamma^4\right)}{4\,\gamma^2\,\left(4\,\beta+3\,\gamma\right)\,\left(\beta+\gamma\right)\,\left(\beta^2+2\,\beta\,\gamma+2\,\gamma^2\right)} \\ \frac{\beta\,\left(10\,\beta^2+18\,\beta\,\gamma+9\,\gamma^2\right)}{2\,\gamma^2\,\left(4\,\beta+3\,\gamma\right)} & \frac{5\,\beta^3+12\,\beta^2\,\gamma+13\,\beta\,\gamma^2+3\,\gamma^3}{\gamma^2\,\left(4\,\beta+3\,\gamma\right)} & \frac{\beta\,\left(10\,\beta^2+18\,\beta\,\gamma+9\,\gamma^2\right)}{2\,\gamma^2\,\left(4\,\beta+3\,\gamma\right)} \\ \frac{20\,\beta^4+66\,\beta^3\,\gamma+100\,\beta^2\,\gamma^2+69\,\beta\,\gamma^3+18\,\gamma^4}{4\,\gamma^2\,\left(4\,\beta+3\,\gamma\right)\,\left(\beta+\gamma\right)} & \frac{\beta\,\left(10\,\beta^3+39\,\beta^2\,\gamma+56\,\beta\,\gamma^2+21\,\gamma^3\right)}{2\,\gamma^2\,\left(4\,\beta+3\,\gamma\right)\,\left(\beta+\gamma\right)} & \frac{20\,\beta^4+66\,\beta^3\,\gamma+100\,\beta^2\,\gamma^2+69\,\beta\,\gamma^3+18\,\gamma^4}{4\,\gamma^2\,\left(4\,\beta+3\,\gamma\right)\,\left(\beta+\gamma\right)} \end{pmatrix}$$

Cumulative Incidence

The vector containing expressions for the cumulative incidence of an SIS epidemic on a line network starting in a given initial state is

$$\gamma \sum_{k=1}^{N=3} \mathcal{T}(k|i) = \begin{pmatrix} \dfrac{30\,\beta^4 + 75\,\beta^3\,\gamma + 74\,\beta^2\,\gamma^2 + 32\,\beta\,\gamma^3 + 6\,\gamma^4}{2\,\gamma\,(4\,\beta + 3\,\gamma)\,(\beta + \gamma)} \\ \\ \dfrac{15\,\beta^3 + 30\,\beta^2\,\gamma + 22\,\beta\,\gamma^2 + 3\,\gamma^3}{\gamma\,(4\,\beta + 3\,\gamma)} \\ \\ \dfrac{30\,\beta^4 + 105\,\beta^3\,\gamma + 156\,\beta^2\,\gamma^2 + 90\,\beta\,\gamma^3 + 18\,\gamma^4}{2\,\gamma\,(4\,\beta + 3\,\gamma)\,(\beta + \gamma)} \end{pmatrix}$$

where $i \in \{SSI, SIS, ISI\}$. Table 5.1 (page 154) gives the cumulative incidence for all of the small networks and different initial states considered for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

In the next sections we consider networks of size $N = 4$, where there are $2^N = 16$ possible states in which the system can be. We again use the technique of lumping with the symmetry of the networks to reduce the number of differential equations needed to describe the SIS epidemic dynamics.

C.2 Complete Network

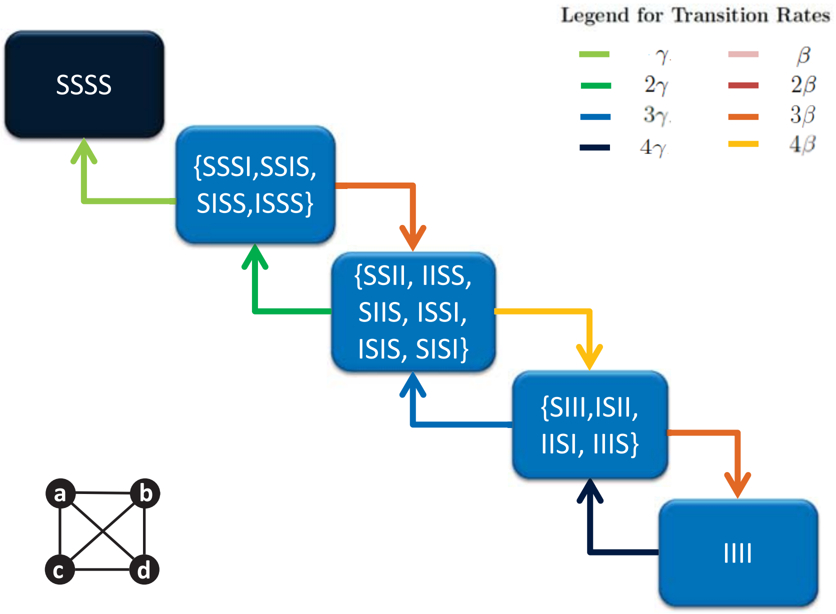


Figure C.2: Transition state diagram for the SIS model on a complete network of $N = 4$ nodes.

The first network of size $N = 4$ which we consider is the complete network. As we have a complete network, in the above figure we have lumped together all the states with the same number of S and I nodes and hence reduced the system from 16 different states to 8 (see Figure C.2). Therefore the reduced state space for the SIS epidemic on a complete network of 4 nodes is $\mathbf{S} = \{SSSI, SSII, SIII, IIII, SSSS\}$, the following matrices are in the same order as $\mathbf{S}$. The equations describing an SIS epidemic on a complete network with 4 nodes are:

$$\begin{aligned}
\dot{P}_{SSSI} &= 2\gamma P_{SSII} - (3\beta + \gamma)P_{SSSI} \\
\dot{P}_{SSII} &= 3\beta P_{SSSI} + 3\gamma P_{SIII} - 2(2\beta + \gamma)P_{SSII} \\
\dot{P}_{SIII} &= 4\beta P_{SSII} + 4\gamma P_{IIII} - 3(\beta + \gamma)P_{SIII} \\
\dot{P}_{IIII} &= 3\beta P_{SIII} - 4\gamma P_{IIII} \\
\dot{P}_{SSSS} &= \gamma P_{SSSI}
\end{aligned}$$

The following matrix, $\mathbf{Q}$, contains the transition probabilities for the *transient states only* of the line network.

$$\mathbf{Q} = \begin{pmatrix} 0 & \dfrac{3\beta}{3\beta+\gamma} & 0 & 0 \\ \dfrac{\gamma}{2\beta+\gamma} & 0 & \dfrac{2\beta}{2\beta+\gamma} & 0 \\ 0 & \dfrac{\gamma}{\beta+\gamma} & 0 & \dfrac{\beta}{\beta+\gamma} \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Expected number of visits to each state

To find the expected number of times the system visits each state from a given initial state we compute the matrix $\mathbf{N}$ as follows, where $\mathbf{I}$ is the identity matrix:

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$$

$$= \begin{pmatrix} \dfrac{3\beta+\gamma}{\gamma} & \dfrac{3\beta(2\beta+\gamma)}{\gamma^2} & \dfrac{6\beta^2(\beta+\gamma)}{\gamma^3} & \dfrac{6\beta^3}{\gamma^3} \\ \dfrac{3\beta+\gamma}{\gamma} & \dfrac{(2\beta+\gamma)(3\beta+\gamma)}{\gamma^2} & \dfrac{2\beta(\beta+\gamma)(3\beta+\gamma)}{\gamma^3} & \dfrac{2\beta^2(3\beta+\gamma)}{\gamma^3} \\ \dfrac{3\beta+\gamma}{\gamma} & \dfrac{(2\beta+\gamma)(3\beta+\gamma)}{\gamma^2} & \dfrac{(\beta+\gamma)(6\beta^2+2\beta\gamma+\gamma^2)}{\gamma^3} & \dfrac{\beta(6\beta^2+2\beta\gamma+\gamma^2)}{\gamma^3} \\ \dfrac{3\beta+\gamma}{\gamma} & \dfrac{(2\beta+\gamma)(3\beta+\gamma)}{\gamma^2} & \dfrac{(\beta+\gamma)(6\beta^2+2\beta\gamma+\gamma^2)}{\gamma^3} & \dfrac{6\beta^3+2\beta^2\gamma+\beta\gamma^2+\gamma^3}{\gamma^3} \end{pmatrix}$$

Expected time spent in each state during the epidemic

The time spent in each state per visit for the complete network of four nodes is stored in the following matrix $\mathbf{T_S}$, where $\mathbf{T_S}(1,1)$ corresponds to the time spent in state $SSSI$, $\mathbf{T_S}(2,2)$ corresponds to the time spent in state $SSII$ and so on following the order of the state space $\mathbf{S} = \{SSSI, SSII, SIII, IIII\}$.

$$\mathbf{T_S} = \frac{1}{\gamma} \times \mathsf{diag}\left(\frac{\gamma}{3\beta+\gamma}, \frac{\gamma}{4\beta+2\gamma}, \frac{\gamma}{3\beta+3\gamma}, \frac{1}{4}\right)$$

$$\mathbf{E_T} = \mathbf{NT_S}$$

$$= \frac{1}{\gamma} \begin{pmatrix} 1 & \dfrac{3\beta}{2\gamma} & \dfrac{2\beta^2}{\gamma^2} & \dfrac{3\beta^3}{2\gamma^3} \\ 1 & \dfrac{3\beta+\gamma}{2\gamma} & \dfrac{2\beta(3\beta+\gamma)}{3\gamma^2} & \dfrac{\beta^2(3\beta+\gamma)}{2\gamma^3} \\ 1 & \dfrac{3\beta+\gamma}{2\gamma} & \dfrac{6\beta^2+2\beta\gamma+\gamma^2}{3\gamma^2} & \dfrac{\beta(6\beta^2+2\beta\gamma+\gamma^2)}{4\gamma^3} \\ 1 & \dfrac{3\beta+\gamma}{2\gamma} & \dfrac{6\beta^2+2\beta\gamma+\gamma^2}{3\gamma^2} & \dfrac{6\beta^3+2\beta^2\gamma+\beta\gamma^2+\gamma^3}{4\gamma^3} \end{pmatrix}$$

Time to absorption

The expected time to absorption is the expected time it takes for the system to reach the infection free steady state. For the complete network with $N = 4$ nodes this is the expected time it takes to reach the state $SSSS$. The following vector $\mathbf{T_A}$ the entries $\mathbf{T_A}(i)$ gives the expected time to absorption given that the epidemic started in state i.

$$\mathbf{T_A} \;=\; \frac{1}{\gamma}\begin{pmatrix} \dfrac{(\beta+\gamma)\,\left(3\,\beta^2+\beta\,\gamma+2\,\gamma^2\right)}{2\,\gamma^3} \\ \dfrac{9\,\beta^3+15\,\beta^2\,\gamma+13\,\beta\,\gamma^2+9\,\gamma^3}{6\,\gamma^3} \\ \dfrac{18\,\beta^3+30\,\beta^2\,\gamma+29\,\beta\,\gamma^2+22\,\gamma^3}{12\,\gamma^3} \\ \dfrac{18\,\beta^3+30\,\beta^2\,\gamma+29\,\beta\,\gamma^2+25\,\gamma^3}{12\,\gamma^3} \end{pmatrix}$$

Number of times each node is infected

We use methods as discussed in previous sections to determine the number of times each node is infected during the epidemic for the complete network with $N = 4$. Therefore we have the following,

$$\text{Number of times node } k \text{ is infected} = \gamma\mathcal{T}(k|i)$$

We have found the algebraic expression of the matrix $\mathcal{T}(k|i)$ for starting in any state i, however here we present the expression for $\mathcal{T}(k|i)$ where $i = SSSI$ which corresponds to $\mathcal{T}(k|1)$. Below we have taken the transpose of $\mathcal{T}(k|1)$ to make it easier to display.

$$\mathcal{T}(k,1) \quad = \quad \begin{pmatrix} \dfrac{54\,\beta^6+126\,\beta^5\,\gamma+138\,\beta^4\,\gamma^2+93\,\beta^3\,\gamma^3+57\,\beta^2\,\gamma^4+26\,\beta\,\gamma^5+6\,\gamma^6}{2\,\gamma^4\,\left(18\,\beta^3+24\,\beta^2\,\gamma+13\,\beta\,\gamma^2+3\,\gamma^3\right)} \\ \\ \dfrac{\beta\,\left(54\,\beta^5+126\,\beta^4\,\gamma+138\,\beta^3\,\gamma^2+93\,\beta^2\,\gamma^3+35\,\beta\,\gamma^4+6\,\gamma^5\right)}{2\,\gamma^4\,\left(18\,\beta^3+24\,\beta^2\,\gamma+13\,\beta\,\gamma^2+3\,\gamma^3\right)} \\ \\ \dfrac{\beta\,\left(54\,\beta^5+126\,\beta^4\,\gamma+138\,\beta^3\,\gamma^2+93\,\beta^2\,\gamma^3+35\,\beta\,\gamma^4+6\,\gamma^5\right)}{2\,\gamma^4\,\left(18\,\beta^3+24\,\beta^2\,\gamma+13\,\beta\,\gamma^2+3\,\gamma^3\right)} \\ \\ \dfrac{\beta\,\left(54\,\beta^5+126\,\beta^4\,\gamma+138\,\beta^3\,\gamma^2+93\,\beta^2\,\gamma^3+35\,\beta\,\gamma^4+6\,\gamma^5\right)}{2\,\gamma^4\,\left(18\,\beta^3+24\,\beta^2\,\gamma+13\,\beta\,\gamma^2+3\,\gamma^3\right)} \end{pmatrix}$$

Therefore, to clarify in the above vector each entry corresponds to the number of times node a, b, c and d were infected respectively during the epidemic given that the system started with node d as the initial infectious node (state $SSSI$). Figure 5.7 shows bar charts of the expected number of times each node is infected for the complete network with $N = 4$ nodes.

Cumulative Incidence

The cumulative incidence is the total number of new infections that occur during an epidemic. The cumulative incidence for an epidemic that started in state i can be found by evaluating the following:

$$\text{Cumulative Incidence } = \gamma \sum_{k=1}^{N} \mathcal{T}(k|i) \tag{C.1}$$

The expression for the cumulative incidence of an SIS epidemic on a complete network starting in state $SSSI$ with one infectious node is given by

$$\gamma \sum_{k=1}^{N=3} \mathcal{T}(k|1) = \frac{6\,\beta^3+6\,\beta^2\,\gamma+3\,\beta\,\gamma^2+\gamma^3}{\gamma^3}$$

Table 5.1 gives the cumulative incidence for all of the small networks and different initial states considered for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

C.3 Square Network

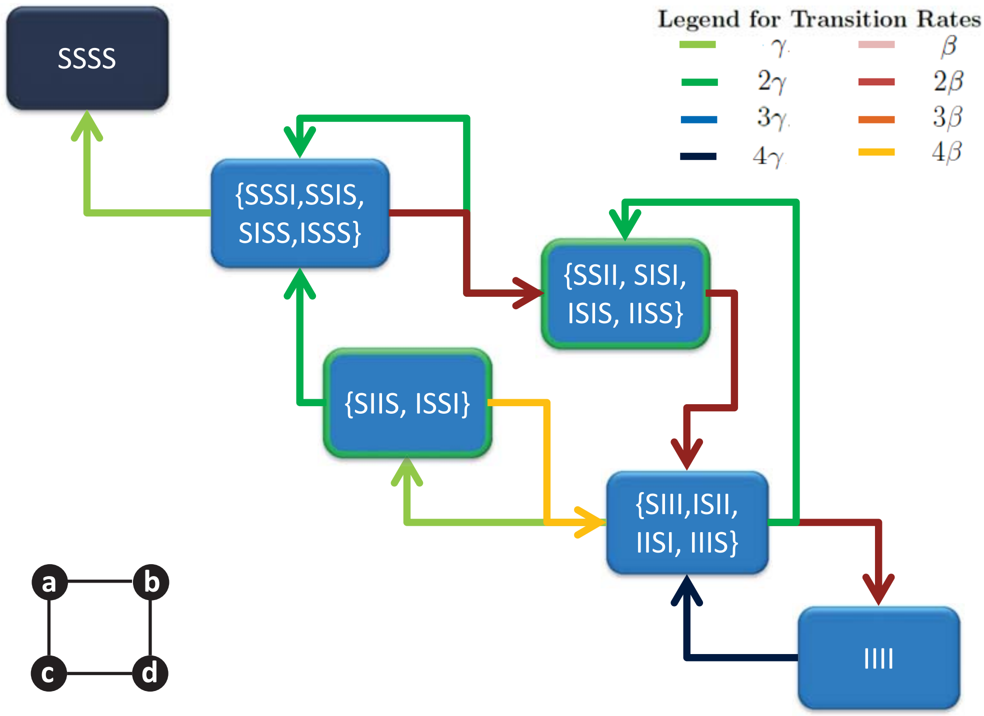


Figure C.3: Transition state diagram for the *SIS* model on a square network of $N = 4$ nodes.

For a square network we were able to reduce the system from 16 states down to 6 via lumping due to the network symmetry. Figure C.3 shows the transition diagram for the square network. The reduced system consists of the state space, $\mathbf{S} = \{SSSI, SIIS, SSII,$ $SIII, IIII, SSSS\}$. Again, throughout this section the order of the matrices will correspond to the order of the state space $\mathbf{S}$ (i.e. the first row in each matrix will correspond to $SSSI$ and so on). The equations describing an SIS epidemic on a square network with 4 nodes are:

$$\begin{aligned}
\dot{P}_{SSSI} &= 2\gamma(P_{SIIS} + P_{SSII}) - (2\beta + \gamma)P_{SSSI} \\
\dot{P}_{SIIS} &= \gamma P_{SIII} - 2(2\beta + \gamma)P_{SIIS} \\
\dot{P}_{SSII} &= 2\gamma P_{SIII} + 2\beta P_{SSSI} - 2(\beta + \gamma)P_{SSII} \\
\dot{P}_{SIII} &= 4\beta P_{SIIS} + 2\beta P_{SSII} + 4\gamma P_{IIII} - (2\beta + 3\gamma)P_{SIII} \\
\dot{P}_{IIII} &= 2\beta P_{SIII} - 4\gamma P_{IIII} \\
\dot{P}_{SSSS} &= \gamma P_{SSSI}
\end{aligned}$$

The following matrix, **Q**, contains the transition probabilities for the *transient states only* of the line network.

$$\mathbf{Q} = \begin{pmatrix} 0 & 0 & \frac{2\beta}{2\beta+\gamma} & 0 & 0 \\ \frac{\gamma}{2\beta+\gamma} & 0 & 0 & \frac{2\beta}{2\beta+\gamma} & 0 \\ \frac{\gamma}{\beta+\gamma} & 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 \\ 0 & \frac{\gamma}{2\beta+3\gamma} & \frac{2\gamma}{2\beta+3\gamma} & 0 & \frac{2\beta}{2\beta+3\gamma} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Expected number of visits to each state

To find the expected number of times the system visits each state from a given initial state we compute the matrix **N** as follows,

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$$

The symbolic form of the matrix **N** for the square network is given on the following page.

$$\mathbf{N} = \begin{pmatrix} \frac{2\beta+\gamma}{\gamma} & \frac{2\beta^2\,(2\beta+\gamma)}{\gamma^2\,(5\beta+3\gamma)} & \frac{2\beta\,(4\beta+3\gamma)\,(\beta+\gamma)}{\gamma^2\,(5\beta+3\gamma)} & \frac{\beta^2\,(2\beta+\gamma)\,(4\beta+6\gamma)}{\gamma^3\,(5\beta+3\gamma)} & \frac{4\beta^3\,(2\beta+\gamma)}{\gamma^3\,(5\beta+3\gamma)} \\ \frac{2\beta+\gamma}{\gamma} & \frac{(2\beta+\gamma)\,\left(2\beta^2+\beta\gamma+3\gamma^2\right)}{\gamma^2\,(5\beta+3\gamma)} & \frac{2\beta\,(4\beta+5\gamma)\,(\beta+\gamma)}{\gamma^2\,(5\beta+3\gamma)} & \frac{\beta\,(4\beta+6\gamma)\,\left(2\beta^2+2\beta\gamma+\gamma^2\right)}{\gamma^3\,(5\beta+3\gamma)} & \frac{4\beta^2\,\left(2\beta^2+2\beta\gamma+\gamma^2\right)}{\gamma^3\,(5\beta+3\gamma)} \\ \frac{2\beta+\gamma}{\gamma} & \frac{\beta\,(2\beta+\gamma)^2}{\gamma^2\,(5\beta+3\gamma)} & \frac{(2\beta+\gamma)\,(4\beta+3\gamma)\,(\beta+\gamma)}{\gamma^2\,(5\beta+3\gamma)} & \frac{\beta\,(2\beta+\gamma)^2\,(2\beta+3\gamma)}{\gamma^3\,(5\beta+3\gamma)} & \frac{2\beta^2\,(2\beta+\gamma)^2}{\gamma^3\,(5\beta+3\gamma)} \\ \frac{2\beta+\gamma}{\gamma} & \frac{(2\beta+\gamma)\,\left(2\beta^2+\beta\gamma+\gamma^2\right)}{\gamma^2\,(5\beta+3\gamma)} & \frac{2\,(4\beta+\gamma)\,(\beta+\gamma)^2}{\gamma^2\,(5\beta+3\gamma)} & \frac{(2\beta+\gamma)\,(2\beta+3\gamma)\,\left(2\beta^2+\beta\gamma+\gamma^2\right)}{\gamma^3\,(5\beta+3\gamma)} & \frac{2\beta\,(2\beta+\gamma)\,\left(2\beta^2+\beta\gamma+\gamma^2\right)}{\gamma^3\,(5\beta+3\gamma)} \\ \frac{2\beta+\gamma}{\gamma} & \frac{(2\beta+\gamma)\,\left(2\beta^2+\beta\gamma+\gamma^2\right)}{\gamma^2\,(5\beta+3\gamma)} & \frac{2\,(4\beta+\gamma)\,(\beta+\gamma)^2}{\gamma^2\,(5\beta+3\gamma)} & \frac{(2\beta+\gamma)\,(2\beta+3\gamma)\,\left(2\beta^2+\beta\gamma+\gamma^2\right)}{\gamma^3\,(5\beta+3\gamma)} & \frac{8\beta^4+8\beta^3\gamma+6\beta^2\gamma^2+7\beta\gamma^3+3\gamma^4}{\gamma^3\,(5\beta+3\gamma)} \end{pmatrix}$$

Expected time spent in each state during the epidemic

The time spent in each state per visit for the square network of four nodes is stored in the following matrix $\mathbf{T_S}$, where $\mathbf{T_S}(1,1)$ corresponds to the time spent in state $SSSI$, $\mathbf{T_S}(2,2)$ corresponds to the time spent in state $SIIS$ and so on following the order of the state space $\mathbf{S} = \{SSSI, SIIS, SSII, SIII, IIII\}$.

$$\mathbf{T_S} \quad = \quad \frac{1}{\gamma} \times \mathsf{diag}\left(\frac{\gamma}{2\beta+\gamma}, \frac{\gamma}{4\beta+2\gamma}, \frac{\gamma}{2\beta+2\gamma}, \frac{\gamma}{2\beta+3\gamma}, \frac{1}{4}\right)$$

The matrix for the expected time spent in each state, $\mathbf{E_T} = \mathbf{NT_S}$ is given on the following page.

Time to absorption

The following vector $\mathbf{T_A}$ the entries $\mathbf{T_A}(i)$ gives the expected time to absorption given that the epidemic started in state i. We used the same methodology here as in previous sections to find $\mathbf{T_A}$.

$$\mathbf{T_A} \quad = \quad \frac{1}{\gamma}\begin{pmatrix} \dfrac{(2\,\beta+3\,\gamma)\left(\beta^3+\beta^2\,\gamma+2\,\beta\,\gamma^2+\gamma^3\right)}{\gamma^3\,(5\,\beta+3\,\gamma)} \\[2em] \dfrac{\left(4\,\beta^4+12\,\beta^3\,\gamma+20\,\beta^2\,\gamma^2+25\,\beta\,\gamma^3+9\,\gamma^4\right)}{2\gamma^3\,(5\,\beta+3\,\gamma)} \\[2em] \dfrac{\left(4\,\beta^4+12\,\beta^3\,\gamma+19\,\beta^2\,\gamma^2+23\,\beta\,\gamma^3+9\,\gamma^4\right)}{2\gamma^3\,(5\,\beta+3\,\gamma)} \\[2em] \dfrac{\left(4\,\beta^4+12\,\beta^3\,\gamma+21\,\beta^2\,\gamma^2+28\,\beta\,\gamma^3+11\,\gamma^4\right)}{2\gamma^3\,(5\,\beta+3\,\gamma)} \\[2em] \dfrac{\left(8\,\beta^4+24\,\beta^3\,\gamma+42\,\beta^2\,\gamma^2+61\,\beta\,\gamma^3+25\,\gamma^4\right)}{4\gamma^3\,(5\,\beta+3\,\gamma)} \end{pmatrix}$$

$$\mathbf{E_T} = \frac{1}{\gamma}\begin{pmatrix} 1 & \frac{\beta^2}{\gamma\,(5\,\beta+3\,\gamma)} & \frac{\beta\,(4\,\beta+3\,\gamma)}{\gamma\,(5\,\beta+3\,\gamma)} & \frac{2\,\beta^2\,(2\,\beta+\gamma)}{\gamma^2\,(5\,\beta+3\,\gamma)} & \frac{\beta^3\,(2\,\beta+\gamma)}{\gamma^3\,(5\,\beta+3\,\gamma)} \\ 1 & \frac{2\,\beta^2+\beta\,\gamma+3\,\gamma^2}{2\,\gamma\,(5\,\beta+3\,\gamma)} & \frac{\beta\,(4\,\beta+5\,\gamma)}{\gamma\,(5\,\beta+3\,\gamma)} & \frac{2\,\beta\,\left(2\,\beta^2+2\,\beta\,\gamma+\gamma^2\right)}{\gamma^2\,(5\,\beta+3\,\gamma)} & \frac{\beta^2\,\left(2\,\beta^2+2\,\beta\,\gamma+\gamma^2\right)}{\gamma^3\,(5\,\beta+3\,\gamma)} \\ 1 & \frac{\beta\,(2\,\beta+\gamma)}{2\,\gamma\,(5\,\beta+3\,\gamma)} & \frac{(2\,\beta+\gamma)\,(4\,\beta+3\,\gamma)}{2\,\gamma\,(5\,\beta+3\,\gamma)} & \frac{\beta\,(2\,\beta+\gamma)^2}{\gamma^2\,(5\,\beta+3\,\gamma)} & \frac{\beta^2\,(2\,\beta+\gamma)^2}{2\,\gamma^3\,(5\,\beta+3\,\gamma)} \\ 1 & \frac{2\,\beta^2+\beta\,\gamma+\gamma^2}{2\,\gamma\,(5\,\beta+3\,\gamma)} & \frac{(4\,\beta+\gamma)\,(\beta+\gamma)}{\gamma\,(5\,\beta+3\,\gamma)} & \frac{(2\,\beta+\gamma)\,\left(2\,\beta^2+\beta\,\gamma+\gamma^2\right)}{\gamma^2\,(5\,\beta+3\,\gamma)} & \frac{\beta\,(2\,\beta+\gamma)\,\left(2\,\beta^2+\beta\,\gamma+\gamma^2\right)}{2\,\gamma^3\,(5\,\beta+3\,\gamma)} \\ 1 & \frac{2\,\beta^2+\beta\,\gamma+\gamma^2}{2\,\gamma\,(5\,\beta+3\,\gamma)} & \frac{(4\,\beta+\gamma)\,(\beta+\gamma)}{\gamma\,(5\,\beta+3\,\gamma)} & \frac{(2\,\beta+\gamma)\,\left(2\,\beta^2+\beta\,\gamma+\gamma^2\right)}{\gamma^2\,(5\,\beta+3\,\gamma)} & \frac{8\,\beta^4+8\,\beta^3\,\gamma+6\,\beta^2\,\gamma^2+7\,\beta\,\gamma^3+3\,\gamma^4}{4\,\gamma^3\,(5\,\beta+3\,\gamma)} \end{pmatrix}$$

Number of times each node is infected

We omit the algebraic expression for $\mathcal{T}(k|i)$ but note it can be easily computed following the above definition and methods outlined previously. Figure 5.7 shows bar charts of the expected number of times each node is infected for the square network with $N = 4$ nodes.

Cumulative Incidence

The vector containing expressions for the cumulative incidence of an SIS epidemic on a square network starting in a given initial state is

$$\gamma \sum_{k=1}^{N=4} \mathcal{T}(k|i) = \left(\begin{array}{c} \dfrac{(2\,\beta+\gamma)\,(\beta+\gamma)\,\left(4\,\beta^2+2\,\beta\,\gamma+3\,\gamma^2\right)}{\gamma^3\,(5\,\beta+3\,\gamma)} \\ \\ \dfrac{8\,\beta^4+20\,\beta^3\,\gamma+26\,\beta^2\,\gamma^2+22\,\beta\,\gamma^3+6\,\gamma^4}{\gamma^3\,(5\,\beta+3\,\gamma)} \end{array} \right)$$

where $i \in \{SSSI, SIIS\}$. Table 5.1 gives the cumulative incidence for all of the small networks and different initial states considered for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

C.4 Star Network

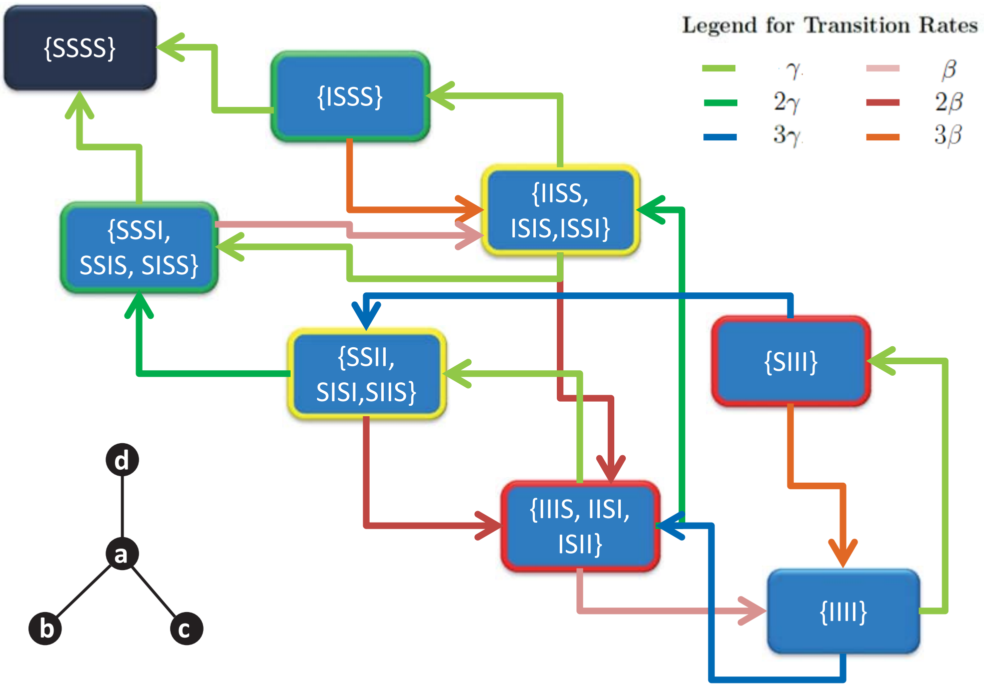


Figure C.4: Transition state diagram for the SIS model on a star network of $N = 4$ nodes.

For a star network we were able to reduce the system from 16 different states down to 8 via lumping due to the network symmetry. The reduced system consists of the state space, $\mathbf{S} = \{SSSI, ISSS, IISS, SSII, SIII, IIIS, IIII, SSSS\}$. Throughout this section the matrices are indexed in the order that the state space $\mathbf{S}$ is in. The transition diagram for the star network is shown in Figure C.4. The equations describing an SIS epidemic on a star network with 4 nodes are:

$$\begin{aligned}
\dot{P}_{SSSI} &= 2\gamma P_{SSII} + \gamma P_{IISS} - (\beta+\gamma)P_{SSSI} \\
\dot{P}_{ISSS} &= \gamma P_{IISS} - (3\beta+\gamma)P_{ISSS} \\
\dot{P}_{IISS} &= \beta P_{SSSI} + 3\beta P_{ISSS} + 2\gamma P_{IIIS} - 2(\beta+\gamma)P_{IISS} \\
\dot{P}_{SSII} &= \gamma P_{IIIS} + 3\gamma P_{SIII} - 2(\beta+\gamma)P_{SSII} \\
\dot{P}_{SIII} &= \gamma P_{IIII} - 3(\beta+\gamma)P_{SIII} \\
\dot{P}_{IIIS} &= 2\beta(P_{IISS} + P_{SSII}) + 3\gamma P_{IIII} - (\beta+3\gamma)P_{IIIS} \\
\dot{P}_{IIII} &= 3\beta P_{SIII} + \beta P_{IIIS} - 4\gamma P_{IIII} \\
\dot{P}_{SSSS} &= \gamma(P_{ISSS} + P_{SSSI})
\end{aligned}$$

The following matrix, **Q**, contains the transition probabilities for the *transient states only* of the line network.

$$\mathbf{Q} = \begin{pmatrix} 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{3\beta}{3\beta+\gamma} & 0 & 0 & 0 & 0 \\ \frac{\gamma}{2\beta+2\gamma} & \frac{\gamma}{2\beta+2\gamma} & 0 & 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 \\ \frac{\gamma}{\beta+\gamma} & 0 & 0 & 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 \\ 0 & 0 & 0 & \frac{\gamma}{\beta+\gamma} & 0 & 0 & \frac{\beta}{\beta+\gamma} \\ 0 & 0 & \frac{2\gamma}{\beta+3\gamma} & \frac{\gamma}{\beta+3\gamma} & 0 & 0 & \frac{\beta}{\beta+3\gamma} \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{pmatrix}$$

Expected number of visits to each state

To find the expected number of times the system visits each state from a given initial state we compute the matrix $\mathbf{N}$ using the same methods as discussed in previous sections. Due to the size of the matrix $\mathbf{N}$ for the star network, it is given over the next two pages.

$$\mathbf{N} = \left(\begin{array}{cccc}
\frac{(\beta+\gamma)\,\left(21\,\beta^2+28\,\beta\,\gamma+12\,\gamma^2\right)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta\,(\beta+2\,\gamma)\,(9\,\beta+3\,\gamma)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{3\,\beta\,(\beta+2\,\gamma)\,(3\,\beta+\gamma)\,(\beta+\gamma)}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{2\,\beta^2\,(3\,\beta+\gamma)\,(\beta+\gamma)}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\
\frac{3\,\beta\,(7\,\beta+6\,\gamma)\,(\beta+\gamma)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{(3\,\beta+\gamma)\,\left(3\,\beta^2+16\,\beta\,\gamma+12\,\gamma^2\right)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{9\,\beta\,(\beta+2\,\gamma)\,(\beta+\gamma)^2}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{6\,\beta^2\,(\beta+\gamma)^2}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\
\frac{(3\,\beta+\gamma)\,(7\,\beta+6\,\gamma)\,(\beta+\gamma)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{(\beta+2\,\gamma)\,(9\,\beta+3\,\gamma)\,(\beta+\gamma)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{(\beta+2\,\gamma)\,(9\,\beta+3\,\gamma)\,(\beta+\gamma)^2}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{2\,\beta\,(3\,\beta+\gamma)\,(\beta+\gamma)^2}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\
\frac{21\,\beta^3+46\,\beta^2\,\gamma+36\,\beta\,\gamma^2+12\,\gamma^3}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta\,(3\,\beta+\gamma)\,\left(3\,\beta^2+12\,\beta\,\gamma+10\,\gamma^2\right)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{\beta\,(3\,\beta+\gamma)\,\left(3\,\beta^2+12\,\beta\,\gamma+10\,\gamma^2\right)}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{6\,\beta^4+14\,\beta^3\,\gamma+20\,\beta^2\,\gamma^2+19\,\beta\,\gamma^3+6\,\gamma^4}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \quad \cdots \\
\frac{(3\,\beta+4\,\gamma)\,\left(7\,\beta^2+6\,\beta\,\gamma+3\,\gamma^2\right)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta\,(3\,\beta+\gamma)\,\left(3\,\beta^2+12\,\beta\,\gamma+13\,\gamma^2\right)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{\beta\,(3\,\beta+\gamma)\,\left(3\,\beta^2+12\,\beta\,\gamma+13\,\gamma^2\right)}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{6\,\beta^4+14\,\beta^3\,\gamma+14\,\beta^2\,\gamma^2+16\,\beta\,\gamma^3+6\,\gamma^4}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\
\frac{(\beta+\gamma)\,\left(21\,\beta^2+25\,\beta\,\gamma+8\,\gamma^2\right)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{(3\,\beta+\gamma)\,\left(3\,\beta^2+9\,\beta\,\gamma+4\,\gamma^2\right)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{(3\,\beta+\gamma)\,(\beta+\gamma)\,\left(3\,\beta^2+9\,\beta\,\gamma+4\,\gamma^2\right)}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{(2\,\beta+2\,\gamma)\,\left(3\,\beta^3+4\,\beta^2\,\gamma+3\,\beta\,\gamma^2+\gamma^3\right)}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\
\frac{21\,\beta^3+46\,\beta^2\,\gamma+33\,\beta\,\gamma^2+9\,\gamma^3}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{(3\,\beta+\gamma)\,\left(3\,\beta^3+12\,\beta^2\,\gamma+13\,\beta\,\gamma^2+3\,\gamma^3\right)}{2\,\gamma\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{(3\,\beta+\gamma)\,\left(3\,\beta^3+12\,\beta^2\,\gamma+13\,\beta\,\gamma^2+3\,\gamma^3\right)}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{6\,\beta^4+14\,\beta^3\,\gamma+14\,\beta^2\,\gamma^2+10\,\beta\,\gamma^3+3\,\gamma^4}{\gamma^2\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)}
\end{array} \right.$$

$$\left. \cdots \quad \begin{matrix} \frac{\beta^3\,(3\,\beta+\gamma)\,(\beta+\gamma)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta^2\,(\beta+3\,\gamma)\,(3\,\beta+\gamma)\,(3\,\beta+4\,\gamma)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{2\,\beta^3\,(3\,\beta+\gamma)\,(\beta+\gamma)}{\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\ \frac{3\,\beta^3\,(\beta+\gamma)^2}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta^2\,(3\,\beta+4\,\gamma)\,(3\,\beta+9\,\gamma)\,(\beta+\gamma)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{6\,\beta^3\,(\beta+\gamma)^2}{\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\ \frac{\beta^2\,(3\,\beta+\gamma)\,(\beta+\gamma)^2}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta\,(\beta+3\,\gamma)\,(3\,\beta+\gamma)\,(3\,\beta+4\,\gamma)\,(\beta+\gamma)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{2\,\beta^2\,(3\,\beta+\gamma)\,(\beta+\gamma)^2}{\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\ \frac{\beta^2\,\left(3\,\beta^3+7\,\beta^2\,\gamma+4\,\beta\,\gamma^2+\gamma^3\right)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta\,(\beta+3\,\gamma)\,(3\,\beta+4\,\gamma)\,\left(3\,\beta^3+7\,\beta^2\,\gamma+4\,\beta\,\gamma^2+\gamma^3\right)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{2\,\beta^2\,\left(3\,\beta^3+7\,\beta^2\,\gamma+4\,\beta\,\gamma^2+\gamma^3\right)}{\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\ \frac{3\,\beta^5+7\,\beta^4\,\gamma+7\,\beta^3\,\gamma^2+32\,\beta^2\,\gamma^3+37\,\beta\,\gamma^4+12\,\gamma^5}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{\beta\,(\beta+3\,\gamma)\,\left(9\,\beta^4+33\,\beta^3\,\gamma+49\,\beta^2\,\gamma^2+28\,\beta\,\gamma^3+7\,\gamma^4\right)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{2\,\beta\,\left(3\,\beta^4+7\,\beta^3\,\gamma+7\,\beta^2\,\gamma^2+8\,\beta\,\gamma^3+3\,\gamma^4\right)}{\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\ \frac{\beta\,(\beta+\gamma)\,\left(3\,\beta^3+4\,\beta^2\,\gamma+3\,\beta\,\gamma^2+\gamma^3\right)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{(\beta+3\,\gamma)\,(3\,\beta+4\,\gamma)\,\left(3\,\beta^3+4\,\beta^2\,\gamma+3\,\beta\,\gamma^2+\gamma^3\right)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{2\,\beta\,(\beta+\gamma)\,\left(3\,\beta^3+4\,\beta^2\,\gamma+3\,\beta\,\gamma^2+\gamma^3\right)}{\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \\ \frac{3\,\beta^5+7\,\beta^4\,\gamma+7\,\beta^3\,\gamma^2+11\,\beta^2\,\gamma^3+10\,\beta\,\gamma^4+3\,\gamma^5}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} & \frac{(\beta+3\,\gamma)\,\left(9\,\beta^5+33\,\beta^4\,\gamma+49\,\beta^3\,\gamma^2+37\,\beta^2\,\gamma^3+16\,\beta\,\gamma^4+3\,\gamma^5\right)}{2\,\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)\,(\beta+\gamma)} & \frac{6\,\beta^5+14\,\beta^4\,\gamma+14\,\beta^3\,\gamma^2+22\,\beta^2\,\gamma^3+20\,\beta\,\gamma^4+6\,\gamma^5}{\gamma^3\,(3\,\beta+2\,\gamma)\,(4\,\beta+3\,\gamma)} \end{matrix} \right)$$

Expected time spent in each state during the epidemic

The time spent in each state per visit for the star network of four nodes is stored in the following matrix $\mathbf{T_S}$, where $\mathbf{T_S}(1,1)$ corresponds to the time spent in state $SSSI$, $\mathbf{T_S}(2,2)$ corresponds to the time spent in state $ISSS$ and so on following the order of the state space $\mathbf{S} = \{SSSI, ISSS, IISS, SSII, SIII, IIIS, IIII\}$.

$$\mathbf{T_S} \quad = \quad \frac{1}{\gamma} \times \mathsf{diag}\left(\frac{\gamma}{\beta+\gamma}, \frac{\gamma}{3\beta+\gamma}, \frac{\gamma}{2\beta+3\gamma}, \frac{\gamma}{2\beta+2\gamma}, \frac{\gamma}{3\beta+3\gamma}, \frac{\gamma}{\beta+3\gamma}, \frac{1}{4}\right)$$

From $\mathbf{T_S}$ we can find the expected time spent in each state throughout the epidemic using methods discussed in previous sections. However,we omit the symbolic representation of the matrix $\mathbf{E_T}$ due to the large size of the entries it contains.

Time to absorption

The following vector $\mathbf{T_A}$ the entries $\mathbf{T_A}(i)$ gives the expected time to absorption for the star network given that the epidemic started in state i.

$$\mathbf{T_A} \quad = \quad \frac{1}{\gamma} \begin{pmatrix} \dfrac{\left(3\,\beta^3 + 10\,\beta^2\,\gamma + 15\,\beta\,\gamma^2 + 18\,\gamma^3\right)\,\left(6\,\beta^3 + 17\,\beta^2\,\gamma + 18\,\beta\,\gamma^2 + 6\,\gamma^3\right)}{6\,\gamma^3\,\left(2\,\beta + 3\,\gamma\right)\,\left(3\,\beta + 2\,\gamma\right)\,\left(4\,\beta + 3\,\gamma\right)} \\[2ex] \dfrac{6\,\beta^6 + 41\,\beta^5\,\gamma + 128\,\beta^4\,\gamma^2 + 249\,\beta^3\,\gamma^3 + 284\,\beta^2\,\gamma^4 + 162\,\beta\,\gamma^5 + 36\,\gamma^6}{2\,\gamma^3\,\left(2\,\beta + 3\,\gamma\right)\,\left(3\,\beta + 2\,\gamma\right)\,\left(4\,\beta + 3\,\gamma\right)} \\[2ex] \dfrac{18\,\beta^6 + 129\,\beta^5\,\gamma + 425\,\beta^4\,\gamma^2 + 875\,\beta^3\,\gamma^3 + 1053\,\beta^2\,\gamma^4 + 630\,\beta\,\gamma^5 + 144\,\gamma^6}{6\,\gamma^3\,\left(2\,\beta + 3\,\gamma\right)\,\left(3\,\beta + 2\,\gamma\right)\,\left(4\,\beta + 3\,\gamma\right)} \\[2ex] \dfrac{18\,\beta^7 + 147\,\beta^6\,\gamma + 548\,\beta^5\,\gamma^2 + 1265\,\beta^4\,\gamma^3 + 1835\,\beta^3\,\gamma^4 + 1611\,\beta^2\,\gamma^5 + 789\,\beta\,\gamma^6 + 162\,\gamma^7}{6\,\gamma^3\,\left(\beta + \gamma\right)\,\left(2\,\beta + 3\,\gamma\right)\,\left(3\,\beta + 2\,\gamma\right)\,\left(4\,\beta + 3\,\gamma\right)} \\[2ex] \dfrac{18\,\beta^7 + 147\,\beta^6\,\gamma + 566\,\beta^5\,\gamma^2 + 1412\,\beta^4\,\gamma^3 + 2209\,\beta^3\,\gamma^4 + 1985\,\beta^2\,\gamma^5 + 969\,\beta\,\gamma^6 + 198\,\gamma^7}{6\,\gamma^3\,\left(\beta + \gamma\right)\,\left(2\,\beta + 3\,\gamma\right)\,\left(3\,\beta + 2\,\gamma\right)\,\left(4\,\beta + 3\,\gamma\right)} \\[2ex] \dfrac{18\,\beta^6 + 129\,\beta^5\,\gamma + 437\,\beta^4\,\gamma^2 + 951\,\beta^3\,\gamma^3 + 1202\,\beta^2\,\gamma^4 + 765\,\beta\,\gamma^5 + 186\,\gamma^6}{6\,\gamma^4\,\left(2\,\beta + 3\,\gamma\right)\,\left(3\,\beta + 2\,\gamma\right)\,\left(4\,\beta + 3\,\gamma\right)} \\[2ex] \dfrac{18\,\beta^7 + 147\,\beta^6\,\gamma + 566\,\beta^5\,\gamma^2 + 1430\,\beta^4\,\gamma^3 + 2308\,\beta^3\,\gamma^4 + 2171\,\beta^2\,\gamma^5 + 1077\,\beta\,\gamma^6 + 216\,\gamma^7}{6\,\gamma^3\,\left(\beta + \gamma\right)\,\left(2\,\beta + 3\,\gamma\right)\,\left(3\,\beta + 2\,\gamma\right)\,\left(4\,\beta + 3\,\gamma\right)} \end{pmatrix}$$

Number of times each node is infected

The number of times node k is infected during the epidemic given that the epidemic started in state i, denoted $\gamma\mathcal{T}(k|i)$, can be found by following methods as outlined in previous sections.

We have found the algebraic expression of the matrix $\mathcal{T}(k|i)$ for starting in any state i, however we omit presenting the matrix here as some of the entries are ratios between polynomials of up to 9^{th} order in β and γ that can not be simplified. Figure

5.8 shows bar charts of the expected number of times each node is infected for the star network with $N = 4$ nodes.

Cumulative Incidence

The vector containing expressions for the cumulative incidence of an SIS epidemic on a star network starting in a given initial state is

$$\gamma \sum_{k=1}^{N=4} \mathcal{T}(k|i) = \begin{pmatrix} \dfrac{6\,\beta^5 + 23\,\beta^4\,\gamma + 40\,\beta^3\,\gamma^2 + 41\,\beta^2\,\gamma^3 + 23\,\beta\,\gamma^4 + 6\,\gamma^5}{\gamma^3\,(3\,\beta + 2\,\gamma)\,(4\,\beta + 3\,\gamma)} \\[2ex] \dfrac{6\,\beta^5 + 27\,\beta^4\,\gamma + 54\,\beta^3\,\gamma^2 + 63\,\beta^2\,\gamma^3 + 35\,\beta\,\gamma^4 + 6\,\gamma^5}{\gamma^3\,(3\,\beta + 2\,\gamma)\,(4\,\beta + 3\,\gamma)} \\[2ex] \dfrac{6\,\beta^5 + 29\,\beta^4\,\gamma + 61\,\beta^3\,\gamma^2 + 76\,\beta^2\,\gamma^3 + 46\,\beta\,\gamma^4 + 12\,\gamma^5}{\gamma^3\,(3\,\beta + 2\,\gamma)\,(4\,\beta + 3\,\gamma)} \\[2ex] \dfrac{6\,\beta^5 + 29\,\beta^4\,\gamma + 67\,\beta^3\,\gamma^2 + 105\,\beta^2\,\gamma^3 + 69\,\beta\,\gamma^4 + 18\,\gamma^5}{\gamma^3\,(3\,\beta + 2\,\gamma)\,(4\,\beta + 3\,\gamma)} \end{pmatrix}$$

where $i \in \{SSSI, ISSS, SSII, SIII\}$. Table 5.1 (page 154) gives the cumulative incidence for all of the small networks and different initial states considered for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

C.5 Toast Network

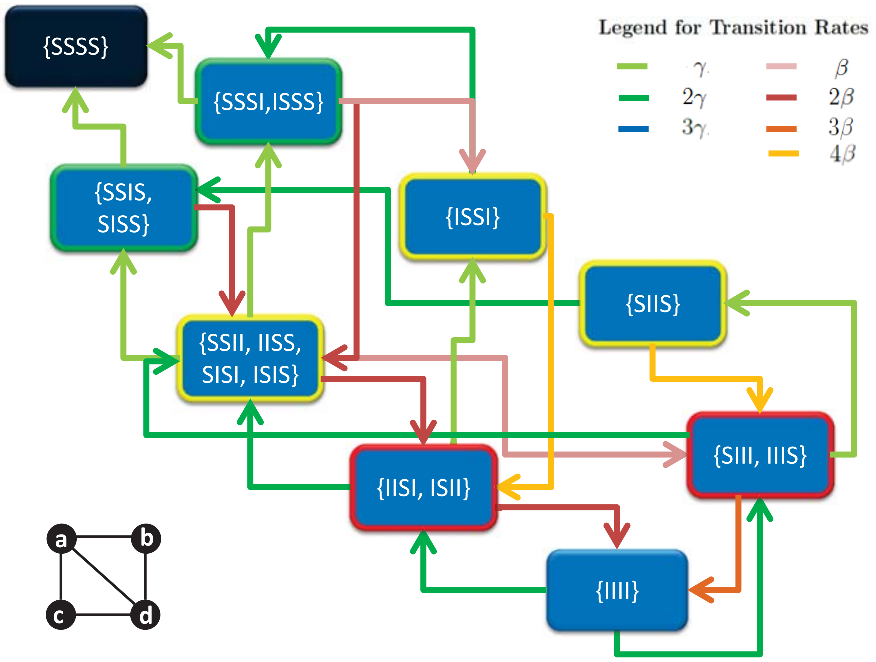


Figure C.5: Transition state diagram for the SIS model on a toast network of $N = 4$ nodes.

For a toast network we were able to reduce the system from 16 different states down to 9 via lumping due to the network symmetry (see Figure C.5). The reduced system consists of the state space, $\mathbf{S} = \{SSSI, SSIS, SSII, ISSI, IISI, SIIS, SIII, IIII, SSSS\}$. Throughout this section the matrices are indexed in the order that the state space $\mathbf{S}$ is in. The equations describing an SIS epidemic on a toast network with 4 nodes are:

$$\begin{aligned}
\dot{P}_{SSSI} &= \gamma P_{SSII} + 2\gamma P_{ISSI} - (3\beta + \gamma) P_{SSSI} \\
\dot{P}_{SSIS} &= \gamma P_{SSII} + 2\gamma P_{SIIS} - (2\beta + \gamma) P_{SSIS} \\
\dot{P}_{SSII} &= 2\beta (P_{SSIS} + P_{SSSI}) + 2\gamma (P_{IISI} + P_{SIII}) - (3\beta + 2\gamma) P_{SSII} \\
\dot{P}_{ISSI} &= \beta P_{SSSI} + \gamma P_{IISI} - (4\beta + 2\gamma) P_{ISSI} \\
\dot{P}_{IISI} &= 2\beta P_{SSII} + 4\beta P_{ISSI} + 2\gamma P_{IIII} - (2\beta + 3\gamma) P_{IISI} \\
\dot{P}_{SIIS} &= \gamma P_{SIII} - (4\beta + 2\gamma) P_{SIIS} \\
\dot{P}_{SIII} &= \beta P_{SSII} + 4\beta P_{SIIS} + 2\gamma P_{IIII} - 3(\beta + \gamma) P_{SIII} \\
\dot{P}_{IIII} &= 2\beta P_{IISI} + 3\beta P_{SIII} - 4\gamma P_{IIII} \\
\dot{P}_{SSSS} &= \gamma (P_{SSSI} + P_{SSIS})
\end{aligned}$$

The following matrix, $\mathbf{Q}$, contains the transition probabilities for the *transient states only* of the line network.
$\mathbf{Q} =$

$$\begin{pmatrix}
0 & 0 & \frac{2\beta}{3\beta+\gamma} & \frac{\beta}{3\beta+\gamma} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{2\beta}{2\beta+\gamma} & 0 & 0 & 0 & 0 & 0 \\
\frac{\gamma}{3\beta+2\gamma} & \frac{\gamma}{3\beta+2\gamma} & 0 & 0 & \frac{2\beta}{3\beta+2\gamma} & 0 & \frac{\beta}{3\beta+2\gamma} & 0 \\
\frac{\gamma}{2\beta+\gamma} & 0 & 0 & 0 & \frac{2\beta}{2\beta+\gamma} & 0 & 0 & 0 \\
0 & 0 & \frac{2\gamma}{2\beta+3\gamma} & \frac{\gamma}{2\beta+3\gamma} & 0 & 0 & 0 & \frac{2\beta}{2\beta+3\gamma} \\
0 & \frac{\gamma}{2\beta+\gamma} & 0 & 0 & 0 & 0 & \frac{2\beta}{2\beta+\gamma} & 0 \\
0 & 0 & \frac{2\gamma}{3\beta+3\gamma} & 0 & 0 & \frac{\gamma}{3\beta+3\gamma} & 0 & \frac{\beta}{\beta+\gamma} \\
0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0
\end{pmatrix}$$

Expected number of visits to each state

The expected number of times the system visits each state from a given initial state is given by the matrix $\mathbf{N}$. However, we omit the algebraic expression for the matrix $\mathbf{N}$ due to size but note that they have been found.

Expected time spent in each state during the epidemic

The time spent in each state per visit for the toast network of four nodes is stored in the following matrix $\mathbf{T_S}$, where $\mathbf{T_S}(1,1)$ corresponds to the time spent in state $SSSI$, $\mathbf{T_S}(2,2)$ corresponds to the time spent in state $SSIS$ and so on following the order of the state space $\mathbf{S}$.

$$\mathbf{T_S} \quad = \quad \frac{1}{\gamma} \times \mathsf{diag}\left(\frac{\gamma}{3\beta+\gamma}, \frac{\gamma}{2\beta+\gamma}, \frac{\gamma}{3\beta+2\gamma}, \frac{\gamma}{4\beta+2\gamma}, \frac{\gamma}{2\beta+3\gamma}, \frac{\gamma}{4\beta+2\gamma}, \frac{\gamma}{3\beta+3\gamma}, \frac{1}{4}\right)$$

The expected time spent in each state throughout the epidemic, $\mathbf{E_T}$, can be found following the methods of previous sections. However, again, we omit the algebraic expression for the matrix $\mathbf{E_T}$ due to the size of the entries it contains

Time to absorption

The following vector $\mathbf{T_A}$ the entries $\mathbf{T_A}(i)$ gives the expected time to absorption given that the epidemic started in state i.

$$\mathbf{T_A} =$$

$$\frac{1}{\gamma}\begin{pmatrix}
\dfrac{432\,\beta^7 + 1680\,\beta^6\,\gamma + 3152\,\beta^5\,\gamma^2 + 3916\,\beta^4\,\gamma^3 + 3285\,\beta^3\,\gamma^4 + 1762\,\beta^2\,\gamma^5 + 540\,\beta\,\gamma^6 + 72\,\gamma^7}{2\,\gamma^3\,\left(272\,\beta^4 + 584\,\beta^3\,\gamma + 515\,\beta^2\,\gamma^2 + 216\,\beta\,\gamma^3 + 36\,\gamma^4\right)} \\[2em]
\dfrac{216\,\beta^7 + 804\,\beta^6\,\gamma + 1446\,\beta^5\,\gamma^2 + 1737\,\beta^4\,\gamma^3 + 1441\,\beta^3\,\gamma^4 + 785\,\beta^2\,\gamma^5 + 252\,\beta\,\gamma^6 + 36\,\gamma^7}{\gamma^3\,\left(272\,\beta^4 + 584\,\beta^3\,\gamma + 515\,\beta^2\,\gamma^2 + 216\,\beta\,\gamma^3 + 36\,\gamma^4\right)} \\[2em]
\dfrac{432\,\beta^7 + 1824\,\beta^6\,\gamma + 3696\,\beta^5\,\gamma^2 + 4920\,\beta^4\,\gamma^3 + 4347\,\beta^3\,\gamma^4 + 2427\,\beta^2\,\gamma^5 + 774\,\beta\,\gamma^6 + 108\,\gamma^7}{2\,\gamma^3\,\left(272\,\beta^4 + 584\,\beta^3\,\gamma + 515\,\beta^2\,\gamma^2 + 216\,\beta\,\gamma^3 + 36\,\gamma^4\right)} \\[2em]
\dfrac{432\,\beta^7 + 1824\,\beta^6\,\gamma + 3744\,\beta^5\,\gamma^2 + 5060\,\beta^4\,\gamma^3 + 4533\,\beta^3\,\gamma^4 + 2549\,\beta^2\,\gamma^5 + 804\,\beta\,\gamma^6 + 108\,\gamma^7}{2\,\gamma^3\,\left(272\,\beta^4 + 584\,\beta^3\,\gamma + 515\,\beta^2\,\gamma^2 + 216\,\beta\,\gamma^3 + 36\,\gamma^4\right)} \\[2em]
\dfrac{432\,\beta^7 + 1824\,\beta^6\,\gamma + 3816\,\beta^5\,\gamma^2 + 5356\,\beta^4\,\gamma^3 + 4969\,\beta^3\,\gamma^4 + 2881\,\beta^2\,\gamma^5 + 940\,\beta\,\gamma^6 + 132\,\gamma^7}{2\,\gamma^3\,\left(272\,\beta^4 + 584\,\beta^3\,\gamma + 515\,\beta^2\,\gamma^2 + 216\,\beta\,\gamma^3 + 36\,\gamma^4\right)} \\[2em]
\dfrac{432\,\beta^7 + 1824\,\beta^6\,\gamma + 3708\,\beta^5\,\gamma^2 + 4988\,\beta^4\,\gamma^3 + 4388\,\beta^3\,\gamma^4 + 2417\,\beta^2\,\gamma^5 + 768\,\beta\,\gamma^6 + 108\,\gamma^7}{2\,\gamma^3\,\left(272\,\beta^4 + 584\,\beta^3\,\gamma + 515\,\beta^2\,\gamma^2 + 216\,\beta\,\gamma^3 + 36\,\gamma^4\right)} \\[2em]
\dfrac{432\,\beta^7 + 1824\,\beta^6\,\gamma + 3816\,\beta^5\,\gamma^2 + 5396\,\beta^4\,\gamma^3 + 5009\,\beta^3\,\gamma^4 + 2878\,\beta^2\,\gamma^5 + 934\,\beta\,\gamma^6 + 132\,\gamma^7}{2\,\gamma^3\,\left(272\,\beta^4 + 584\,\beta^3\,\gamma + 515\,\beta^2\,\gamma^2 + 216\,\beta\,\gamma^3 + 36\,\gamma^4\right)} \\[2em]
\dfrac{432\,\beta^7 + 1824\,\beta^6\,\gamma + 3816\,\beta^5\,\gamma^2 + 5512\,\beta^4\,\gamma^3 + 5281\,\beta^3\,\gamma^4 + 3137\,\beta^2\,\gamma^5 + 1045\,\beta\,\gamma^6 + 150\,\gamma^7}{2\,\gamma^3\,\left(272\,\beta^4 + 584\,\beta^3\,\gamma + 515\,\beta^2\,\gamma^2 + 216\,\beta\,\gamma^3 + 36\,\gamma^4\right)}
\end{pmatrix}$$

Number of times each node is infected

The number of times node k is infected during the epidemic given that the epidemic started in state i, denoted $\gamma\mathcal{T}(k|i)$, can be found by following methods as outlined in previous sections. We have found the algebraic expression of the matrix $\mathcal{T}(k|i)$ for starting in any state i but omit presenting it here due to the size of the entries. Figure 5.9 shows bar charts of the expected number of times each node is infected for the toast network with $N = 4$ nodes.

Cumulative Incidence

The cumulative incidence is the total number of new infections that occur during an epidemic. For the SIS model on a toast network with $N = 4$ nodes we find the cumulative incidence by summing up the vector containing the number of times each node is infected. Therefore, the cumulative incidence for an epidemic that started in state i can be found by evaluating

$$\text{Cumulative Incidence } = \gamma \sum_{k=1}^{N} \mathcal{T}(k|i) \tag{C.2}$$

The vector containing expressions for the cumulative incidence of an SIS epidemic on a toast network starting in a given initial state is

$$\gamma \sum_{k=1}^{N=4} \mathcal{T}(k|i) =$$

$$\begin{pmatrix} \dfrac{(2\,\beta+\gamma)\,(3\,\beta+2\,\gamma)\,\left(144\,\beta^5+308\,\beta^4\,\gamma+322\,\beta^3\,\gamma^2+208\,\beta^2\,\gamma^3+81\,\beta\,\gamma^4+18\,\gamma^5\right)}{\gamma^3\,\left(272\,\beta^4+584\,\beta^3\,\gamma+515\,\beta^2\,\gamma^2+216\,\beta\,\gamma^3+36\,\gamma^4\right)} \\ \\ \dfrac{864\,\beta^7+3000\,\beta^6\,\gamma+4836\,\beta^5\,\gamma^2+4774\,\beta^4\,\gamma^3+3101\,\beta^3\,\gamma^4+1307\,\beta^2\,\gamma^5+324\,\beta\,\gamma^6+36\,\gamma^7}{\gamma^3\,\left(272\,\beta^4+584\,\beta^3\,\gamma+515\,\beta^2\,\gamma^2+216\,\beta\,\gamma^3+36\,\gamma^4\right)} \\ \\ \dfrac{(4\,\beta+2\,\gamma)\,\left(216\,\beta^6+714\,\beta^5\,\gamma+1100\,\beta^4\,\gamma^2+1058\,\beta^3\,\gamma^3+623\,\beta^2\,\gamma^4+216\,\beta\,\gamma^5+36\,\gamma^6\right)}{\gamma^3\,\left(272\,\beta^4+584\,\beta^3\,\gamma+515\,\beta^2\,\gamma^2+216\,\beta\,\gamma^3+36\,\gamma^4\right)} \end{pmatrix}$$

where $i \in \{SSIS, SSSI, SIIS\}$. Table 5.1 gives the cumulative incidence for all of the small networks and different initial states considered for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

C.6 Line Network, $N = 4$

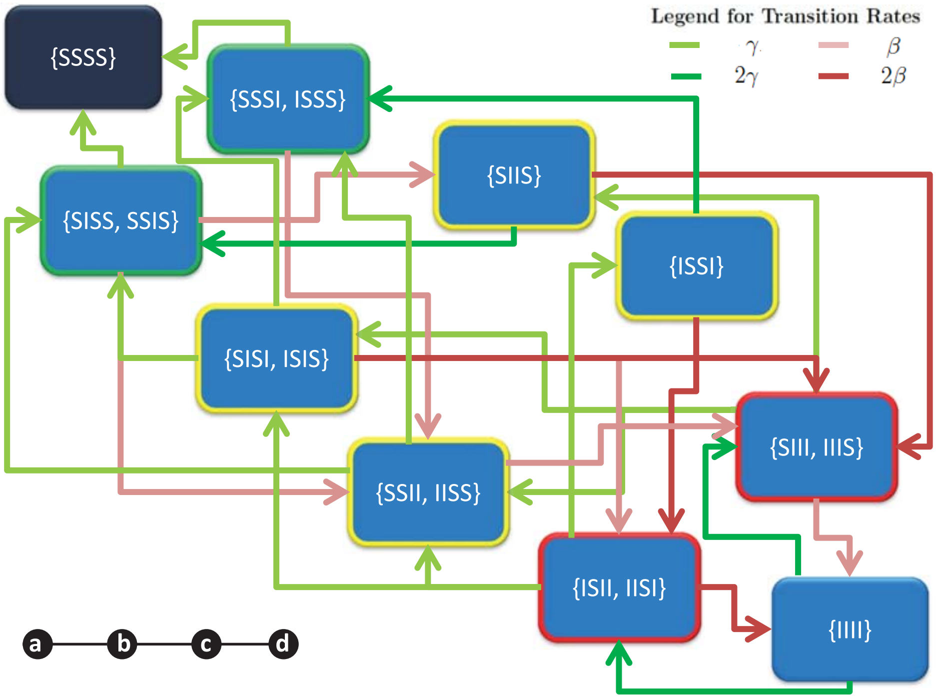


Figure C.6: Transition state diagram for the *SIS* model on a line network of $N = 4$ nodes.

For a line network we were able to reduce the system from 16 different states down to 10 via lumping due to the network symmetry. Figure C.6 shows the transition diagram for line network of $N = 4$. The reduced system consists of the state space, $\mathbf{S} = \{SSSI, SSIS, SIIS, SISI, ISSI, SSII, SIII, ISII, IIII, SSSS\}$. Throughout this section the matrices are indexed in the order that the state space $\mathbf{S}$ is in. The equations describing an *SIS* epidemic on a line network with 4 nodes are:

$$\begin{aligned}
\dot{P}_{SSSI} &= \gamma(P_{SISI} + P_{SSII}) + 2\gamma P_{ISSI} - (\beta + \gamma)P_{SSSI} \\
\dot{P}_{SSIS} &= \gamma(P_{SISI} + P_{SSII}) + 2\gamma P_{SIIS} - (2\beta + \gamma)P_{SSIS} \\
\dot{P}_{SIIS} &= \beta P_{SSIS} + \gamma P_{SIII} - 2(\beta + \gamma)P_{SIIS} \\
\dot{P}_{SISI} &= \gamma(P_{ISII} + P_{SIII}) - (3\beta + 2\gamma)P_{SISI} \\
\dot{P}_{ISSI} &= \gamma P_{ISII} - 2(\beta + \gamma)P_{ISSI} \\
\dot{P}_{SSII} &= \beta(P_{SSSI} + P_{SSIS}) + \gamma(P_{ISII} + P_{SIII}) - (\beta + 2\gamma)P_{SSII} \\
\dot{P}_{SIII} &= \beta P_{SSII} + 2\beta(P_{SISI} + P_{SIIS}) + 2\gamma P_{IIII} - (\beta + 3\gamma)P_{SIII} \\
\dot{P}_{ISII} &= \beta P_{SISI} + 2\beta P_{ISSI} + 2\gamma P_{IIII} - (2\beta + 3\gamma)P_{ISII} \\
\dot{P}_{IIII} &= 2\beta P_{ISII} + \beta P_{SIII} - 4\gamma P_{IIII} \\
\dot{P}_{SSSS} &= \gamma(P_{SSIS} + P_{SSSI})
\end{aligned}$$

The following matrix, $\mathbf{Q}$, contains the transition probabilities for the *transient states only* of the line network.

$$\mathbf{Q} =$$

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 & 0 & 0 \\
0 & 0 & \frac{\beta}{2\beta+\gamma} & 0 & 0 & \frac{\beta}{2\beta+\gamma} & 0 & 0 & 0 \\
0 & \frac{\gamma}{\beta+\gamma} & 0 & 0 & 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 & 0 \\
\frac{\gamma}{3\beta+2\gamma} & \frac{\gamma}{3\beta+2\gamma} & 0 & 0 & 0 & 0 & \frac{2\beta}{3\beta+2\gamma} & \frac{\beta}{3\beta+2\gamma} & 0 \\
\frac{\gamma}{\beta+\gamma} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\beta}{\beta+\gamma} & 0 \\
\frac{\gamma}{\beta+2\gamma} & \frac{\gamma}{\beta+2\gamma} & 0 & 0 & 0 & 0 & \frac{\beta}{\beta+2\gamma} & 0 & 0 \\
0 & 0 & \frac{\gamma}{\beta+3\gamma} & \frac{\gamma}{\beta+3\gamma} & 0 & \frac{\gamma}{\beta+3\gamma} & 0 & 0 & \frac{\beta}{\beta+3\gamma} \\
0 & 0 & 0 & \frac{\gamma}{2\beta+3\gamma} & \frac{\gamma}{2\beta+3\gamma} & \frac{\gamma}{2\beta+3\gamma} & 0 & 0 & \frac{2\beta}{2\beta+3\gamma} \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0
\end{pmatrix}$$

Expected number of visits to each state

The expected number of times the system visits each state from a given initial state is given by the matrix $\mathbf{N}$ as discussed in previous sections. However, we omit the algebraic expression for the matrix $\mathbf{N}$ due to size but note that they have been found.

Expected time spent in each state during the epidemic

The time spent in each state per visit for the line network of four nodes is stored in the following matrix $\mathbf{T_S}$, where $\mathbf{T_S}(1,1)$ corresponds to the time spent in state $SSSI$, $\mathbf{T_S}(2,2)$ corresponds to the time spent in state $SSIS$ and so on following the order of the state space $\mathbf{S}$.

$$\mathbf{T_S} \;=\; \frac{1}{\gamma} \times \mathsf{diag}\left(\frac{\gamma}{\beta+\gamma}, \frac{\gamma}{2\beta+\gamma}, \frac{\gamma}{2\beta+2\gamma}, \frac{\gamma}{3\beta+2\gamma}, \frac{\gamma}{2\beta+2\gamma}, \frac{\gamma}{\beta+2\gamma}, \frac{\gamma}{\beta+3\gamma}, \frac{\gamma}{2\beta+3\gamma}, \frac{1}{4}\right)$$

The expected time spent in each state throughout the epidemic, $\mathbf{E_T}$, can be found following the methods of previous sections. However, again, we omit the algebraic expression for the matrix $\mathbf{E_T}$ due to the size of the entries it contains.

Time to absorption

The following vector $\mathbf{T_A}$ the entries $\mathbf{T_A}(i)$ gives the expected time to absorption given that the epidemic started in state i.

$$\mathbf{T_A} =$$

$$\frac{1}{\gamma}\begin{pmatrix} \dfrac{12\,\beta^8 + 88\,\beta^7\,\gamma + 321\,\beta^6\,\gamma^2 + 800\,\beta^5\,\gamma^3 + 1420\,\beta^4\,\gamma^4 + 1742\,\beta^3\,\gamma^5 + 1404\,\beta^2\,\gamma^6 + 672\,\beta\,\gamma^7 + 144\,\gamma^8}{2\,\gamma^3\,\left(4\,\beta^2 + 9\,\beta\,\gamma + 6\,\gamma^2\right)\left(12\,\beta^3 + 31\,\beta^2\,\gamma + 32\,\beta\,\gamma^2 + 12\,\gamma^3\right)} \\[2ex] \dfrac{12\,\beta^8 + 94\,\beta^7\,\gamma + 368\,\beta^6\,\gamma^2 + 975\,\beta^5\,\gamma^3 + 1783\,\beta^4\,\gamma^4 + 2176\,\beta^3\,\gamma^5 + 1680\,\beta^2\,\gamma^6 + 744\,\beta\,\gamma^7 + 144\,\gamma^8}{2\,\gamma^3\,\left(4\,\beta^2 + 9\,\beta\,\gamma + 6\,\gamma^2\right)\left(12\,\beta^3 + 31\,\beta^2\,\gamma + 32\,\beta\,\gamma^2 + 12\,\gamma^3\right)} \\[2ex] \dfrac{12\,\beta^8 + 100\,\beta^7\,\gamma + 421\,\beta^6\,\gamma^2 + 1203\,\beta^5\,\gamma^3 + 2328\,\beta^4\,\gamma^4 + 2955\,\beta^3\,\gamma^5 + 2358\,\beta^2\,\gamma^6 + 1080\,\beta\,\gamma^7 + 216\,\gamma^8}{2\,\gamma^3\,\left(4\,\beta^2 + 9\,\beta\,\gamma + 6\,\gamma^2\right)\left(12\,\beta^3 + 31\,\beta^2\,\gamma + 32\,\beta\,\gamma^2 + 12\,\gamma^3\right)} \\[2ex] \dfrac{12\,\beta^8 + 100\,\beta^7\,\gamma + 415\,\beta^6\,\gamma^2 + 1168\,\beta^5\,\gamma^3 + 2209\,\beta^4\,\gamma^4 + 2739\,\beta^3\,\gamma^5 + 2174\,\beta^2\,\gamma^6 + 1020\,\beta\,\gamma^7 + 216\,\gamma^8}{2\,\gamma^3\,\left(4\,\beta^2 + 9\,\beta\,\gamma + 6\,\gamma^2\right)\left(12\,\beta^3 + 31\,\beta^2\,\gamma + 32\,\beta\,\gamma^2 + 12\,\gamma^3\right)} \end{pmatrix}$$

Number of times each node is infected

The number of times node k is infected during the epidemic given that the epidemic started in state i, denoted $\gamma\mathcal{T}(k|i)$, can be found by following methods as outlined in previous sections. We have found the algebraic expression of the matrix $\mathcal{T}(k|i)$ for starting in any state i however we omit presenting it here. Figure 5.10 on page 159 shows bar charts of the expected number of times each node is infected for the line network with $N = 4$ nodes.

Cumulative Incidence

The vector containing expressions for the cumulative incidence of an SIS epidemic on a line network starting in a given initial state is

$$\gamma\sum_{k=1}^{N=4}\mathcal{T}(k|i) =$$

$$\begin{pmatrix} \dfrac{(\beta+\gamma)^2\,\left(24\,\beta^6 + 110\,\beta^5\,\gamma + 251\,\beta^4\,\gamma^2 + 377\,\beta^3\,\gamma^3 + 366\,\beta^2\,\gamma^4 + 228\,\beta\,\gamma^5 + 72\,\gamma^6\right)}{\gamma^3\,\left(4\,\beta^2 + 9\,\beta\,\gamma + 6\,\gamma^2\right)\left(12\,\beta^3 + 31\,\beta^2\,\gamma + 32\,\beta\,\gamma^2 + 12\,\gamma^3\right)} \\[2ex] \dfrac{24\,\beta^8 + 170\,\beta^7\,\gamma + 580\,\beta^6\,\gamma^2 + 1261\,\beta^5\,\gamma^3 + 1862\,\beta^4\,\gamma^4 + 1859\,\beta^3\,\gamma^5 + 1194\,\beta^2\,\gamma^6 + 444\,\beta\,\gamma^7 + 72\,\gamma^8}{\gamma^3\,\left(4\,\beta^2 + 9\,\beta\,\gamma + 6\,\gamma^2\right)\left(12\,\beta^3 + 31\,\beta^2\,\gamma + 32\,\beta\,\gamma^2 + 12\,\gamma^3\right)} \\[2ex] \dfrac{24\,\beta^8 + 182\,\beta^7\,\gamma + 677\,\beta^6\,\gamma^2 + 1630\,\beta^5\,\gamma^3 + 2654\,\beta^4\,\gamma^4 + 2890\,\beta^3\,\gamma^5 + 2016\,\beta^2\,\gamma^6 + 816\,\beta\,\gamma^7 + 144\,\gamma^8}{\gamma^3\,\left(4\,\beta^2 + 9\,\beta\,\gamma + 6\,\gamma^2\right)\left(12\,\beta^3 + 31\,\beta^2\,\gamma + 32\,\beta\,\gamma^2 + 12\,\gamma^3\right)} \\[2ex] \dfrac{(\beta+\gamma)\,\left(24\,\beta^7 + 158\,\beta^6\,\gamma + 507\,\beta^5\,\gamma^2 + 1062\,\beta^4\,\gamma^3 + 1414\,\beta^3\,\gamma^4 + 1188\,\beta^2\,\gamma^5 + 600\,\beta\,\gamma^6 + 144\,\gamma^7\right)}{\gamma^3\,\left(4\,\beta^2 + 9\,\beta\,\gamma + 6\,\gamma^2\right)\left(12\,\beta^3 + 31\,\beta^2\,\gamma + 32\,\beta\,\gamma^2 + 12\,\gamma^3\right)} \end{pmatrix}$$

where $i \in \{SSSI, SSIS, SISI, ISSI\}$. Table 5.1 gives the cumulative incidence for all of the small networks and different initial states considered for $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

Appendix D

Here we present figures showing a selected number of the networks presented throughout this thesis to illustrate that the variance of S/A (ratio between the stochastic, S, and analytic, A, results) tends toward zero as the number of realisations increases. Refer to Section 5.3.2 on page 152 for the detailed discussion. All results presented here are evaluated for the infection parameter values of $\{\beta = 1, \gamma = 1\}$ and $\{\beta = 2, \gamma = 1\}$.

D.1 Line Network, $N = 3$

Figure D.2 shows the convergence of the variance of S/A for the time to absorption and number of times each node was infected for the SIS model on a line network given that the initial state is as shown in Figure D.1. Figure D.3 shows the results for the expected number of visits to each state and the expected time spent in each state.

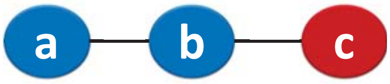


Figure D.1: Schematic of the initial state SSI for the Line Network with $N = 3$ nodes.

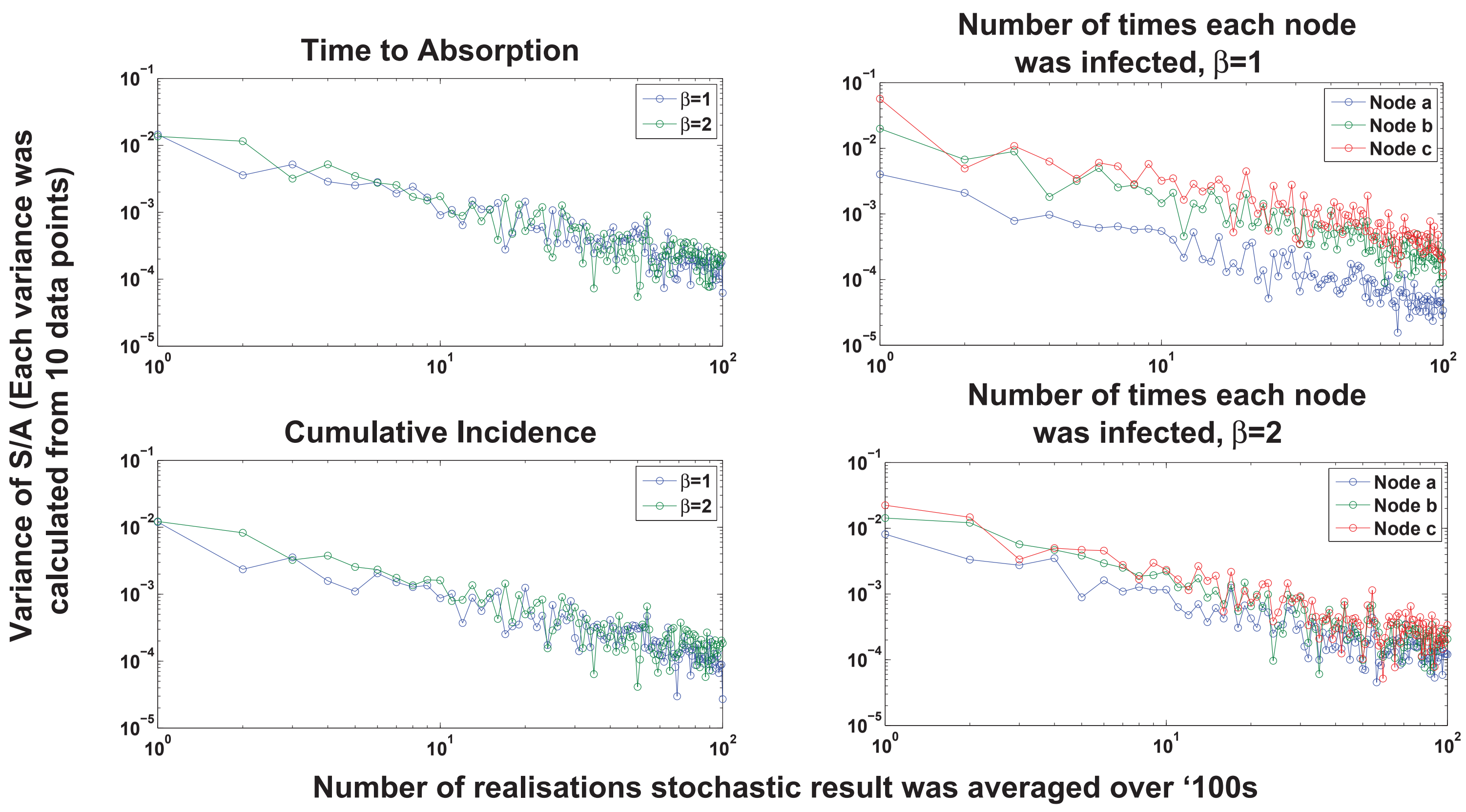


Figure D.2: Convergence of the variance of $\log(S/A)$ for the Line Network with initial state SSI

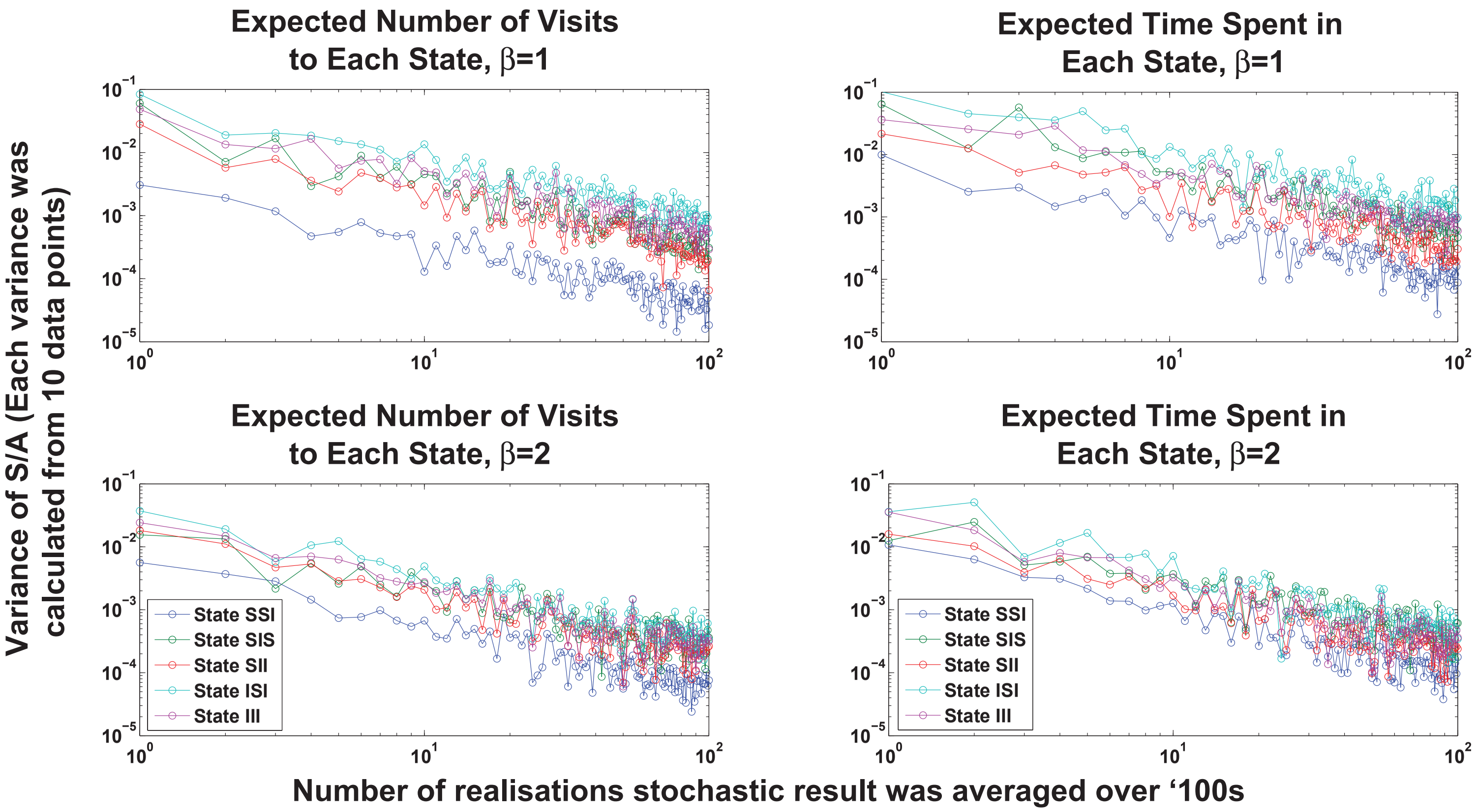


Figure D.3: Convergence of the variance of $\log(S/A)$ for the Line Network with initial state SSI

D.2 Complete Network, $N = 4$

Figure D.5 shows the convergence of the variance of S/A for the time to absorption and number of times each node was infected for the SIS model on a complete network given that the initial state is as shown in Figure D.4. Figure D.6 shows the results for the expected number of visits to each state and the expected time spent in each state.

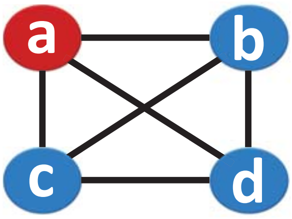


Figure D.4: Schematic of the initial state $ISSS$ for the Complete Network with $N = 4$ nodes.

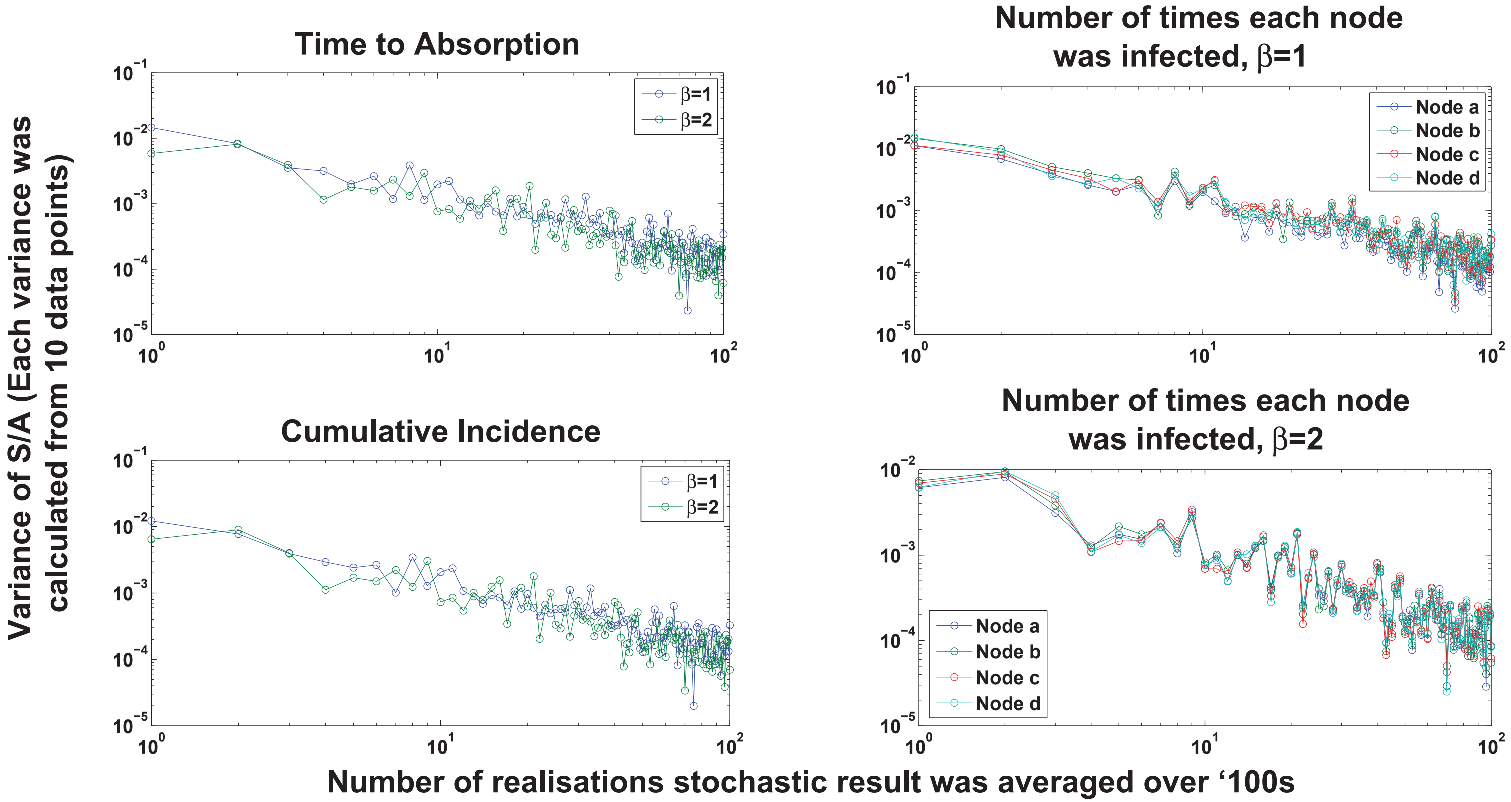


Figure D.5: Convergence of the variance of $\log(S/A)$ for the Complete Network with initial state $ISSS$

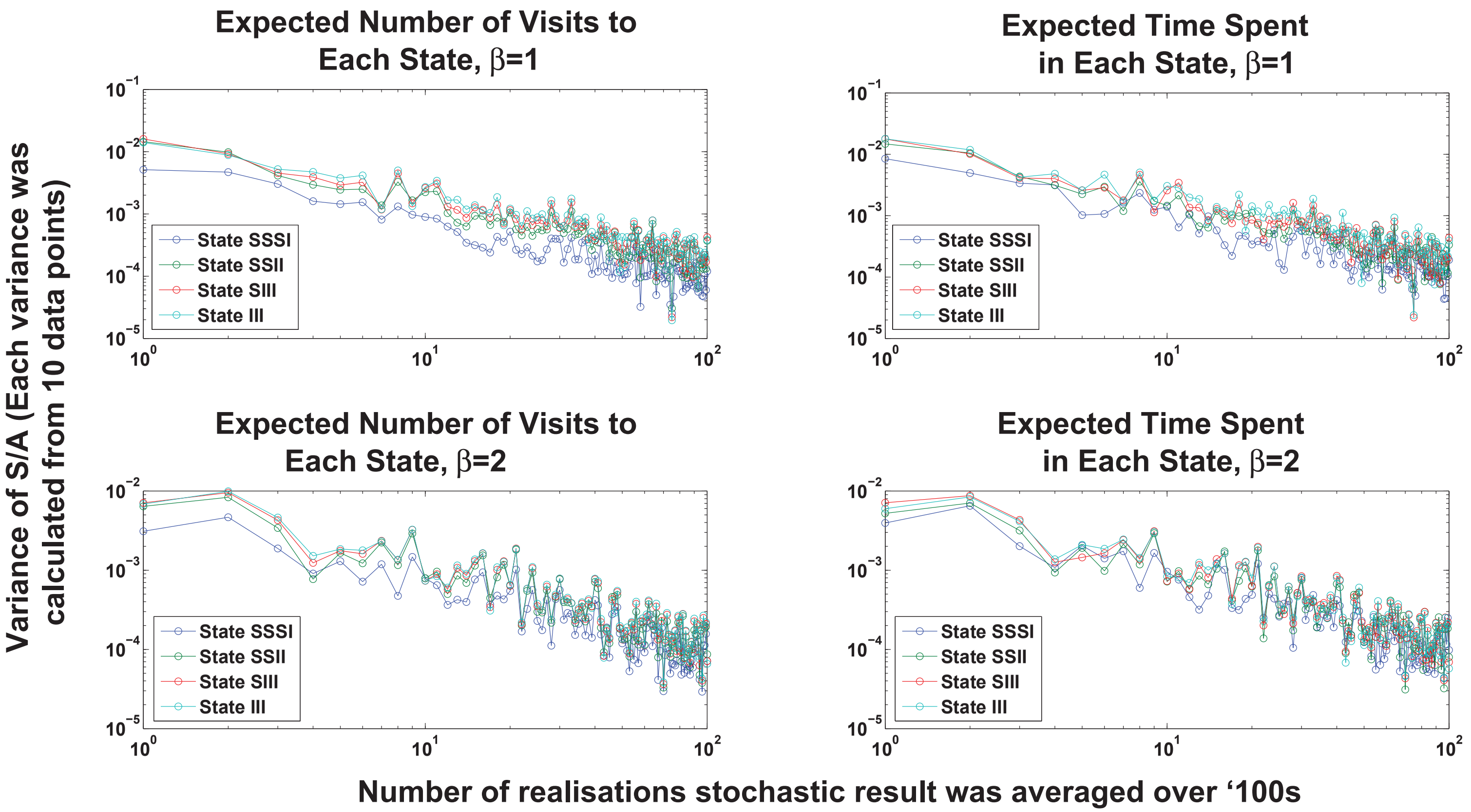


Figure D.6: Convergence of the variance of $\log(S/A)$ for the Complete Network with initial state $ISSS$

D.3 Line of Triangles Network, $N = 6$

Figure D.8 shows the convergence of the variance of S/A for the time to absorption and number of times each node was infected for the SIS model on a LoT(6) network given that the initial infectious node is a (see Figure D.7).

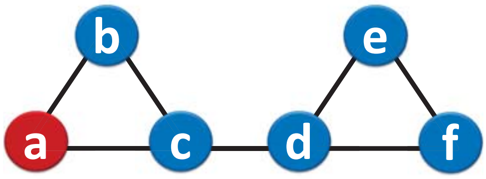


Figure D.7: Schematic of the initial state where node a is infectious for the Line of Triangles Network with $N = 6$ nodes.

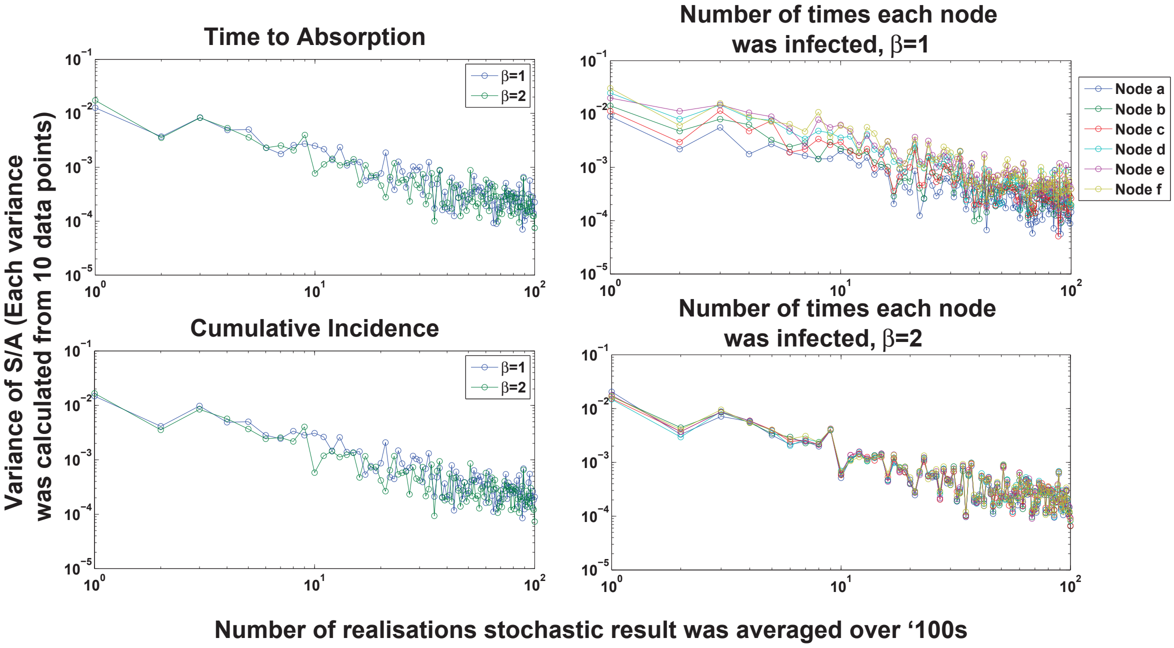


Figure D.8: Convergence of the variance of log(S/A) for the Line of Triangles Network with initial state $ISSSSS$

Appendix E

DRC 16

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Karen McCulloch

Name/Title of Principal Supervisor: Professor Mick Roberts

Name of Published Research Output and full reference:

K. McCulloch, M. G. Roberts and C. R. Laing, Exact analytical expressions for the final epidemic size of an SIR model on small networks, ANZIAM Journal, 2016, 57(4), p429-444.

In which Chapter is the Published Work: Chapter 2 and Appendix A

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate: 90

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  Candidate wrote paper based on research that was carried out as part of their PhD. Co-authors of the paper are the candidates research supervisors.

Karen McCulloch Digitally signed by Karen McCulloch Date: 2016.12.16 13:52:23 +13'00'

Candidate's Signature

16-12-2016

Date

Mick Roberts Digitally signed by Mick Roberts Date: 2016.12.16 14:16:36 +13'00'

Principal Supervisor's signature

16-12-2016

Date