Stochastic models for the spread of infectious diseases on finite contact networks: exact results and representations

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy

by

Robert R. Wilkinson

June 2015

Acknowledgement

I would like to thank my supervisor, Dr. Kieran Sharkey, for his unwavering support and encouragement while conducting this work. From him, I have learned much about mathematical epidemiology, the academic life, and also to believe in my abilities.

Without the reassurance of family and friends, particularly my partner, Jocelyn, and my mother, Jennifer, I am sure this would not have been possible.

Abstract

Mathematical models for the spread of infectious diseases between living organisms are crucial to humanity's endeavour to understand and control its environment. The threat posed by communicable diseases is great. For example, the 1918 flu pandemic resulted in the deaths of over 50 million people and the HIV/AIDS pandemic is still under way with 2.3 million new cases in 2012. Mathematical models allow us to make predictions about the likelihood, impact and time scale of possible epidemics, and to devise effective intervention programmes, e.g. mass vaccination.

This thesis considers various stochastic models of disease propagation which utilise the concept of a finite contact (social) network. For such models, we investigate ways in which important information can be extracted without a full mathematical 'solution' (often unavailable) or numerous time consuming simulations (often inefficient and uninformative). For example, we consider the probability that a large scale outbreak will occur when a single infected individual is introduced to a susceptible population, and the expected number of infected individuals at time t.

Although we focus on the context of epidemiology, the models under investigation are elementary and will be applicable to other domains, such as the spread of computer viruses, the spread of ideas, chemical reactions, and interacting particle systems.

Contents

1 Introduction **1**

1.1 Overview . . . 1
1.2 Compartmental modelling . . . 2
1.3 Early deterministic models in mathematical epidemiology - construction and key results . . . 2
1.3.1 The deterministic SIR model . . . 2
1.3.2 The deterministic SIS model . . . 5
1.3.3 Remarks . . . 6
1.4 The standard SIR stochastic model (Markovian and non-Markovian versions) . . . 6
1.4.1 Defining invasion via coupling to branching processes . . . 8
1.4.2 Computing invasion probability . . . 9
1.5 The standard SIS stochastic model (Markovian) . . . 11
1.6 Quasi-stationary distributions (for conditional final outcomes) . . . 12
1.7 Networks (graphs representing populations with contact stucture) . . . 14
1.7.1 Erdös-Rényi random graphs . . . 16
1.7.2 Generalised random graphs . . . 17
1.8 Invasion and final outcome for epidemics on generalised random graphs (via percolation theory) . . . 19
1.9 Capturing the dynamics of epidemics on generalised random graphs . . . 21
1.10 Epidemics on finite non-random networks . . . 23
1.10.1 Capturing the dynamics by a moment closure approach . . . 24
1.10.2 Capturing the dynamics by a message passing approach . . . 29
1.11 Stochastic simulation . . . 30

2 Invasion and endemicity in the Markovian network-based SIS model **32**

2.1 Introduction . . . 32
2.2 Numerical investigations . . . 34
2.3 Graphical representation and duality . . . 37
2.4 Endemicity and quasi-prevalence . . . 41
2.5 Quasi-invasion . . . 43

2.6 The prevalence-invasion relationship . . . 44
2.7 Simulations on a small square lattice . . . 46
2.8 Computational efficiency in the measurement of invasion probability and endemic prevalence - a new perspective . . . 46
2.9 Weakly connected networks . . . 47
2.10 Discussion . . . 51

3 Moment-closure for Markovian epidemic dynamics on networks 53
3.1 Introduction . . . 53
3.2 Pair-based systems at the population level . . . 54
3.3 Exact systems for tree networks . . . 60
3.4 Exact systems for non-tree networks . . . 64
3.4.1 Exact closure theorem . . . 64
3.4.2 Examples . . . 69
3.4.3 System size . . . 71
3.5 Hierarchies of approximate systems . . . 75
3.5.1 Cycle-partitioning . . . 78
3.5.2 Size-partitioning . . . 82
3.5.3 Hybrid-partitioning . . . 84
3.5.4 Alternative closure . . . 85
3.6 Discussion . . . 87

4 The message passing approach to general epidemics on networks 89
4.1 Introduction . . . 89
4.2 The model . . . 90
4.3 Exact systems for tree networks . . . 91
4.3.1 Exact message passing equations for pair-states . . . 93
4.3.2 The case of Poisson contact processes; novel pair-based systems emerge . . . 93
4.4 Approximate systems and a rigorous bound for non-tree networks . . . 99
4.5 Generalising to SEIR dynamics . . . 102
4.6 A note on initial conditions . . . 104
4.7 Discussion . . . 107

5 Using the message passing formalism to prove new results for classic models 110
5.1 The deterministic SIR model provides a rigorous bound on the Markovian standard SIR model . . . 110
5.2 General epidemics on homogeneous graphs . . . 114
5.3 The general epidemic on an infinite complete graph (mean-field case) . . . 115

5.4 General epidemics on configuration networks with two levels of mixing (superimposed even mixing) 117

6 Final summary and discussion **121**

7 Appendix **125**

A.1 Proof of the underpinning results for Theorem 3.4.3 125

A.2 Proof of general results on cycle-partitioning 128

Bibliography **129**

Chapter 1

Introduction

1.1 Overview

This thesis is concerned with the construction and analysis of mathematical models for the spread of infectious diseases. The purpose of this introductory chapter is therefore to provide an idiosyncratic summary of some key methodologies and results within mathematical epidemiology, such that the ideas underlying the rest of the thesis are then familiar to the reader. These ideas will be presented, approximately, in an order of increasing complexity where the complexity arises as extra aspects of the real world processes are incorporated. For example, deterministic models will be followed by stochastic models, and homogeneous host populations will be considered before populations with more heterogeneous contact structures.

Throughout this chapter, various models will be defined, and their assumptions about the real world will be examined. In addition, existing methods of analysis and computation relating to the following important concepts/questions will be presented:

- Invasion: Given the introduction of an infectious agent into a population of susceptible hosts, what is the probability of a large scale outbreak (invasion)? How does one distinguish between 'large' and 'small' outbreaks?
- Final outcome: Given that invasion occurs, can we make predictions about the impact on the host population?
- Dynamics: Can we track the progress of the infection in time?

The thesis as a whole represents further work in this direction and is, to a large extent, based on the following papers (to which the authors made equal contributions):

- Chapter 2: Wilkinson, R.R and Sharkey, K.J. (2013) 'An exact relationship between invasion probability and endemic prevalence for Markovian SIS dynamics on networks', *PLoS ONE* **8**, e69028.

- Chapter 3: Sharkey, K.J. and Wilkinson, R.R. (2015) 'Complete hierarchies of SIR models on arbitrary networks with exact and approximate moment closure', arXiv:1501.06353 [q-bio.PE].

- Chapter 4: Wilkinson, R.R and Sharkey, K.J. (2014) 'Message passing and moment closure for susceptible-infected-recovered epidemics on finite networks', *Phys. Rev. E* **89**, 022808.

Indeed, some sections of the thesis are reproduced verbatim from these papers.

1.2 Compartmental modelling

The typical approach in mathematical epidemiology is to assume that at any point in time the host population can be divided into a small number of 'compartments', where a given compartment represents all individuals of a particular health status, e.g. the 'infected' compartment (Kermack and McKendrick, 1927). Deterministic models can then be defined as mathematical dynamical systems (ordinary differential equations), where the state of the system is a list of 'densities' for all the compartments. For example, the density of the infected compartment represents the total number of infected individuals. Note that under this set-up, any 'internal structure' of a given compartment is ignored: only the compartment's density is kept track of.

The models which we consider in this thesis are primarily stochastic (probabilistic not deterministic), and the individual hosts are treated explicitly. Thus, in general, the list of compartment densities at a given time does not allow us to 'evolve' the model according to its prescribed rules: the *particular* set of individuals belonging to each compartment must be known, and possibly even the length of time for which individuals have been in a given compartment (and other information which is ignored in the traditional dynamical systems approach). However, if we are still primarily concerned with keeping track of the compartment densities, and whether given individuals belong to given compartments, then we will still refer to the model as compartmental. In this wider sense, all of the models which we consider are compartmental.

The concept of a compartment is similar to that of an individual-state. Indeed, we will refer to an individual in compartment X as being in state X.

1.3 Early deterministic models in mathematical epidemiology - construction and key results

1.3.1 The deterministic SIR model

Some of the earliest and most influential work in mathematical epidemiology can be found in the seminal papers of Kermack and McKendrick (the first published in 1927). In 'Contributions to the mathematical theory of epidemics - I' the authors construct

a model by first defining three compartments/states: 'susceptible', 'infected' and 'removed'. In the susceptible state, an individual (host) is at risk of becoming infected if there are other individuals in the population that are infected. If an individual becomes infected it inevitably recovers (is removed from the process) at some later time. Thus, individuals are only permitted to transition between states in the following way:

$$S(\text{susceptible}) \rightarrow I(\text{infected}) \rightarrow R(\text{removed}).$$

The authors then discretise time and define how the change in the total number of individuals in a given state, from time t to time $t + \Delta$, is determined by the total number of susceptible individuals at time t and the total number of infected individuals at time t that have been infected for θ time steps (for all $\theta \in \{0, \Delta, 2\Delta, \ldots, t\}$). Implicit in this construction is the assumption that all individuals are the same in terms of their interaction with the infectious agent, i.e. all individuals are equally predisposed to catching, transmitting and recovering from the infection, and the assumption that all individuals interact equally with all other individuals, i.e. all susceptible individuals have the same risk of being infected at the next time step. Both of these assumptions are often relaxed or removed in more modern constructions which incorporate, for example, heterogeneous contact structure. The authors are able to arrive at a particularly appealing and simple model by also assuming the time for which individuals have been in their current state to be irrelevant, this corresponding to 'constant' rates of recovery and infectivity. This final assumption makes the mathematics so much more tractable that it is still ubiquitous in epidemiological modelling. Allowing the time step Δ to tend towards zero, the model is then expressible as a system of ordinary differential equations (ODEs):

$$\begin{aligned} \dot{S(t)} &= -(\beta N) S(t) I(t), \\ \dot{I(t)} &= (\beta N) S(t) I(t) - \gamma I(t), \\ \dot{R(t)} &= \gamma I(t), \end{aligned} \tag{1.1}$$

where $S(t)$, $I(t)$ and $R(t)$ are the fraction of a population of size N that are susceptible, infected and recovered respectively at time t, while γ represents the rate at which infected individuals recover and β represents the rate at which a given infected individual makes 'infectious contacts' to another given individual (upon receiving an infectious contact, a susceptible individual immediately becomes infected). However, in modelling a given disease in populations of different sizes it is common to keep βN constant, letting $\beta' = \beta N$, such that we get the same trajectories for a given disease in populations of different sizes. This is justified by the notion that a larger population is not necessarily more dense and so the rate at which an infected individual makes infectious contacts should, in general, be kept constant. In practice this can lead to very small values of β when N is large.

Jumping ahead somewhat, note that if the model were stochastic and at time t the fraction of the population in each compartment was $S(t)$, $I(t)$ and $R(t)$ (with probability 1), then plugging these numbers into the right-hand-sides of system (1.1) would give the exact rates of change of the expected fraction in each compartment (at time t) if infectious periods were exponentially distributed with parameter γ and if any given infected individual made infectious contacts to any given susceptible individual according to a Poisson process of rate β (see the Markovian stochastic model in section 1.4). This holds since, in this case, the product $S(t)I(t)N^2$ is exactly the number of susceptible-infected pairings at time t. Thus, γ may be interpreted as the reciprocal of the average infectious period, and β as the reciprocal of the average time between infectious contacts (from an infected individual to another given individual).

We will refer to the above system (1.1) as the deterministic SIR model. Note that this model is most appropriate for relatively fast moving infections where change in the size of the host population (caused by births, (unrelated) deaths, migration etc.) can be assumed to be negligible, and where infected individuals will either recover with long-term immunity from further infection or else die from the disease (in either case they are removed from the infection process). System (1.1) is deterministic in the sense that for any initial system state, given as the fractional size of each compartment at $t = 0$, the state of the system at time t is then fixed. However, it is not usually possible to obtain closed-form solutions for non-linear systems such as (1.1), and in these cases software packages such as MATLAB can be employed for numerical integration.

Although the non-linearity in system (1.1) makes a closed form solution 'impossible', Kermack and McKendrick identified several important features or characteristics of the model. Possibly the most famous observation is that there exists a threshold parameter $\beta N/\gamma$ such that, assuming the fraction of the population that are initially infected is small (and the rest are susceptible), then the size of the infected compartment is initially increasing if and only if $\beta N/\gamma > 1$, otherwise the infected compartment decreases monotonically. Therefore, when a small number of initially infected individuals are 'introduced' to the population, we say that an 'epidemic' or 'invasion' will occur if and only if $\beta N/\gamma > 1$.

Here, the number $\beta N/\gamma$ is called the 'basic reproductive ratio' and is denoted R_0. It can be interpreted as the 'expected number of secondary cases caused by a single primary case in an otherwise susceptible population' and, indeed, this is the more general definition of R_0. Note that under the Markovian stochastic interpretation of the deterministic model, mentioned above, this expected value is precisely $\beta(N-1)/(\beta+\gamma)$ (and this tends towards $\beta N/\gamma$ as $N \to \infty, \beta \to 0$). Clearly, if this expected number is less than 1, then the expected number of 'new' cases caused by a single infected individual, at any point during the process, is also less than 1. Thus, if $R_0 < 1$, then the expected number of n^{th} generation cases decreases monotonically with n, while if

$R_0 > 1$ then it is initially increasing with n (assuming a single initial infected).

For the deterministic SIR model, Kermack and McKendrick also derived a transcendental equation for the limit of $R(t)$ as $t \to \infty$ (denoted $R(\infty)$), for the case where $R(0) = 0$, this giving the overall fraction of the population affected by the disease (Kermack and McKendrick (1927)):

$$R(\infty) = 1 - S(0)e^{-\frac{\beta N}{\gamma}R(\infty)}. \tag{1.2}$$

Note that we now have a threshold parameter relating to the initial behaviour of the system and an expression for the final state ($S(\infty) = 1 - R(\infty), I(\infty) = 0$) which both involve the number $R_0 = \beta N/\gamma$. Indeed, if $S(0) \to 1, I(0) \to 0$, such that the initial presence of the infection is infinitesimal, then the above equation has a single solution for $R(\infty) \in [0, 1)$ if and only if $\beta N/\gamma > 1$ (otherwise the only solution is $R(\infty) = 0$).

1.3.2 The deterministic SIS model

Bearing in mind the construction process for the deterministic SIR model, it is now possible to write down deterministic models, expressible as systems of ODEs, for diseases where different or extra compartments (for the hosts) need to be defined, and/or where movement between compartments may reflect different mechanisms. For example, in some cases infected hosts may become re-susceptible to the disease after their infectious period has terminated (consider sexually transmitted diseases or computer viruses). This notion leads to the deterministic SIS model in which individuals can move back and forth between the susceptible compartment and the infected compartment:

$$\begin{aligned} \dot{S(t)} &= -(\beta N)S(t)I(t) + \gamma I(t), \\ \dot{I(t)} &= (\beta N)S(t)I(t) - \gamma I(t), \end{aligned} \tag{1.3}$$

where the variables and parameters are defined as in the deterministic SIR model, except that γ is now the rate at which infected individuals recover and immediately become susceptible. Note that for system (1.3), assuming $\beta, \gamma, N > 0$, all of the equilibria lie along the lines $I = 0$ and $S = \gamma/(\beta N)$. However, under the constraint $S + I = 1$, there are exactly two equilibria and these correspond to the points $(I = 0, S = 1)$ (disease-free equilibrium) and $(I = 1 - \gamma/(\beta N), S = \gamma/(\beta N))$ (endemic equilibrium). Note that the first of these is biologically plausible while the second is only plausible if $\beta N/\gamma > 1$. Moreover, it can be shown that if this condition for plausibility is met then the disease-free equilibrium is unstable and the endemic equilibrium is stable, and vice versa when the condition for plausibility is not met. Thus, as in the deterministic SIR model, there exists the same threshold parameter $R_0 = \beta N/\gamma$, such that if $R_0 > 1$ then a stable endemic equilibrium emerges whereby the infection can persist indefinitely. The size of the infected compartment at this stable endemic equilibrium is then related

to the threshold, i.e. $I_{\text{endemic}} = 1 - \gamma/(\beta N)$. If $\beta N/\gamma < 1$ then the infected compartment monotonically tends to zero (assuming $I(0), S(0) > 0$ and $S(0) + I(0) = 1$). Therefore, similarly to the deterministic SIR model, we can say that an 'epidemic' or 'invasion' will occur if and only if $R_0 > 1$ (and $I(0) > 0$).

1.3.3 Remarks

The threshold and final outcome results support the idea that reducing the average infectious period (by medicinal intervention) and/or the rate at which infectious contacts are made (by modification of human behaviour) can both reduce the likelihood of a significant outbreak and mitigate the impact of the disease in the event of such an outbreak. Indeed, the threshold results imply that if a fraction greater than $1-\gamma/(\beta N)$ are vaccinated then an 'epidemic' will no longer be possible (see, for example, Keeling and Rohani (2007)).

There have been many deterministic models constructed which emphasise different aspects of real-world disease propagation according to their perceived importance in different contexts. All of the following concepts (and more besides) have been incorporated into deterministic epidemiological models: age structured populations, metapopulations and households, vector-borne transmission, local contact structure, seasonal variation, heterogeneity in host susceptibility and infectivity. For examples, see respectively: McKendrick (1926); House and Keeling (2008); Ross (1911); Keeling (1999); Keeling, Rohani and Grenfell (2000); May and Anderson (1988).

The majority of this thesis will involve analysis of stochastic models but in almost all cases the considerations which led to the construction of the deterministic models will be highly relevant. Moreover, the endeavour to produce mathematically precise and practically relevant results, without finding exact closed form solutions to the dynamics, will be reflected in many of the narratives of this thesis.

1.4 The standard SIR stochastic model (Markovian and non-Markovian versions)

In the stochastic version of the deterministic SIR model (see, for example, Bailey's (1975) 'general stochastic epidemic'), which we will refer to as the Markovian standard SIR model, we have a population consisting of a set V of $N = |V|$ discrete individuals. An individual, while infected, makes infectious contacts to any other given individual according to a Poisson process of rate β. If a susceptible individual receives an infectious contact, it immediately becomes infected for an exponentially distributed period with parameter γ, after which it ceases making contacts and becomes permanently recovered (all individual level recovery and contact processes are independent). Such dynamics are Markovian, as a result of all the individual level processes being Poisson, and this

means that for a given present state, i.e. knowledge of which individuals are in which states, then the future and the past are independent. This independence of the future from the past, given some present state, is also inherent in the deterministic SIR model since it was there assumed that the history of any given individual in a given state was irrelevant. In fact, all of the assumptions are the same except that here we allow the process to evolve stochastically such that the sizes of the compartments are discrete and 'jump' up or down in single units (they are not continuous in time). If the infectious periods are i.i.d. (independent identically distributed) random variables distributed as $\mathcal{R}$, then we will refer to the resulting non-Markovian model as the standard SIR model.

The way in which the individuals transition between states, under the model outlined above, can be completely described by a continuous time Markov chain $\{\sigma(t)\}$ where $\sigma(t)$ denotes the (random) configuration of the system/population at time t, taking values in $\mathcal{S} = \mathcal{S}_{ind}^V = \{S, I, R\}^V$, such that $\sigma_i(t)$ denotes the corresponding (random) status of individual $i \in V$ at time t (S-susceptible, I-infected, R-removed/recovered). The transition rates for this Markov chain are given in table 1.1, in which σ is an arbitrary system configuration (not a random element) and σ_i is the state of individual i implied by σ. The configuration $\sigma^{i\to X}$ is the same as σ except with the state of individual i set to $X \in \{S, I, R\}$. Given this setup, a parameter set (V, β, γ) and a probability distribution for $\sigma(0)$, uniquely determines the probability distribution for $\sigma(t)$ (see, for example, Grimmett and Stirzaker (1982) for theory of Markov chains).

Letting $X(t)$ and $Y(t)$ denote the random number of susceptible individuals at time t and the random number of infected individuals at time t respectively, then the continuous time Markov chain $\{(X(t), Y(t))\}$, with transition rates as in table 1.2 with the constraint that $x + y \leq N$ and $x, y \geq 0$, is also consistent with the dynamics in the Markovian standard SIR model.

Table 1.1: Individual level transitions for the Markovian standard SIR model

from	to	at rate
$\sigma : \sigma_i = S$	$\sigma^{i\to I}$	$\beta \sum_{j\in V} \mathbb{1}(\sigma_j = I)$
$\sigma : \sigma_i = I$	$\sigma^{i\to R}$	γ

Table 1.2: Population level transitions for the Markovian standard SIR model

from	to	at rate
(x, y)	$(x-1, y+1)$	βxy
(x, y)	$(x, y-1)$	γy

Many important results for the Markovian standard SIR model have been proved and many also generalise to the case where infectious periods are arbitrary i.i.d. random variables, i.e. to the standard SIR model. One result which is important to this thesis,

and which we will state without proof, is that in the limit of large population size (as $N \to \infty$) the Markovian standard SIR model converges to the deterministic SIR model by a law of large numbers, i.e. $(X(t)/N, Y(t)/N) \to (S(t), I(t))$ uniformly on bounded intervals for agreeing parametrisation and initial conditions (the reader is directed to Ethier and Kurtz (1986) and the lecture notes of Andersson and Britton (2000)).

1.4.1 Defining invasion via coupling to branching processes

The early stages of the standard SIR model can be approximated as a 'branching process' (Bartlett, 1955; Ball and Donnelly, 1995) such that the approximation 'improves' as N increases. This result also allows a computation of the probability of an 'epidemic' or 'invasion', for the standard SIR model, when there are initially a small number of infected individuals, or just one. Interestingly, this does not require an 'epidemic' or 'invasion' to be mathematically defined such that it, or its absence, can be identified in a single stochastic realisation. The result relies on the method of coupling in which, for this context, a number of different processes are constructed on the same probability space in such a way as to be illuminating. As a simple example, it is possible to construct two standard SIR models, one with $\beta = a$ the other with $\beta = b > a$ (and no other differences), on the same probability space such that $X(t)_{\beta=a} \geq X(t)_{\beta=b}$ while $X(t)_{\beta=a}$ and $X(t)_{\beta=b}$ remain correctly distributed, thus proving that $P(X(t)_{\beta=a} \leq n) \leq P(X(t)_{\beta=b} \leq n) \forall n$ (Andersson and Britton, 2000). This is intuitive since one would expect there to be less susceptibles at a given time t if, everything else being kept the same, the rate at which infected individuals make infectious contacts was increased.

Following Andersson and Britton (2000), the relevant continuous time branching process is defined as follows: Let there be m live individuals (ancestors) at $t = 0$. The lifespans of all individuals are i.i.d. random variables distributed as $\mathcal{I}$. While alive, an individual gives birth to new individuals according to a Poisson process of rate λ, and all Poisson processes are independent. We will use $B(t)$ to denote the total number of births occurring before time t and $L(t)$ to denote the total number of living individuals at time t.

We will now describe a method of constructing the standard SIR model for a population of size N and m initial infecteds and the branching process defined above (with m ancestors), setting $\beta = \lambda/(N - m)$ and $\mathcal{R} = \mathcal{I}$, on the same probability space. Note that the rate at which an initially infected individual makes infectious contacts to initially susceptible individuals in the epidemic model has been set to be the same as the rate at which a live individual gives birth in the branching process, and the distribution of the infectious period has been set to match the distribution of an individual lifespan. Also note that, in the standard SIR model, the infectious contacts made by an infected individual are made to individuals chosen uniformly at random from the rest of the

population.

We now draw a sequence of integers $U_i, i \geq 1$, where each is drawn uniformly at random from the set $\{1, 2, \ldots, (N-m)\}$ (with replacement), and where the length of the sequence is equal to the total number of births that occur in the branching process (possibly infinite). We then assign the sequence of numbers U_i, $i \geq 1$, to the sequence of births in the branching process in the order in which they occur (in time). Now, for every birth which is assigned a number which is also assigned to an earlier birth we remove it, and all future births resulting from it, from the time line of the branching process (this constrains the branching process equivalently to how the epidemic process is constrained by the decreasing number of susceptibles). After doing this, the new number $L'(t)$ of individuals alive at time t in the (modified) branching process has the same distribution as the number $Y(t)$ of infected individuals at time t in the epidemic process, and the total number $B'(t)$ of births before time t now has the same distribution as the total number of successful infections in the epidemic process, $X(0) - X(t)$. Importantly, if we let T be the time of the first birth which has a number which is also assigned to an earlier birth (before the procedure of removing births from the branching process time line), then it is clear that up until this time the (unmodified) branching process and the epidemic process must have an exact correspondence, i.e. $P(L(t) = L'(t)|T > t) = 1$ and $P(B(t) = B'(t) \mid T > t) = 1$. It is straightforward to then check that for finite t we have $P(T > t) \to 1$ as $N \to \infty$ and so, in this limit, we have $L(t) \stackrel{D}{=} Y(t)$ and $B(t) \stackrel{D}{=} X(0) - X(t)$ (Ball and Donnelly, 1995; Andersson and Britton, 2000).

1.4.2 Computing invasion probability

We have shown that the early stage of the standard SIR model can be well approximated by a continuous time branching process when the population size is large. For the branching process, the probability of ultimate extinction (all individuals eventually die out) can be computed exactly. It is thus natural to define the probability of an 'epidemic' or 'invasion' for the standard SIR model as one minus the probability of ultimate extinction in the corresponding branching process (Ball and Donnelly, 1995). For the branching process described in the previous section, consider that the expected number of offspring produced by a single individual is $\lambda E[\mathcal{I}]$ and the expected number of individuals in generation n is given by $m(\lambda E[\mathcal{I}])^n$ (where m is the number of individuals at $t = 0$). There is thus a threshold value ($= 1$) for the expected number of offspring of an individual such that above the threshold the expected number in generation n grows geometrically with n, but below the threshold the expected size of generation n tends towards zero (for such generations, we know that the births of individuals into generation n occur after the birth of the first individual into generation $n-1$). However, this description is misleading since even above the threshold there is still the possibility

that the population will become extinct.

Let there be a single initial individual (ancestor) and let q be the probability of ultimate extinction, then q must satisfy the self consistency relationship:

$$q = \sum_{k=0}^{\infty} q^k P(D_0 = k), \tag{1.4}$$

where $P(D_0 = k)$ is the probability that an individual will give birth to k individuals in total. In fact, q is the smallest non-negative root of equation 1.4 (see, for example, Grimmett and Stirzaker (1982)). The self consistency relationship can be understood by noting that all individuals behave independently, and for extinction to occur we need all of the independent 'lines' from each of the offspring of the initial individual to also go extinct. If we have m initial individuals then the probability of ultimate extinction is q^m (again due to independence). Since the right-hand-side of equation (1.4) is a convex and non-decreasing function of q, for $q \in [0, 1]$, which passes through 1 at $q = 1$, then it is straightforward to show that there is at most one solution in $[0, 1)$ and it exists only if the gradient at $q = 1$ is greater than 1 (otherwise the only solution is $q = 1$). This is the case when the expected number of offspring for an individual is greater than 1 (and hence taking the $q = 1$ solution would be inconsistent with the exploding *expected* population size).

If the continuous time branching process is constructed to be Markovian, by setting $\mathcal{I}$ to be exponential with parameter μ, then the process will evolve according to the transition rates in table 1.3. Note that the expected number of offspring for an individual is now given by λ/μ such that the threshold is at $\lambda/\mu = 1$. For the Markovian case, given some present state, the future depends only on the current number of living individuals L (and not their histories).

Table 1.3: Transitions for the Markovian branching process

from	to	at rate
$L = x$	$L = x + 1$	λx
$L = x$	$L = x - 1$	μx

The probability of ultimate extinction q_M (from a single ancestor) in the Markovian branching process is simple to compute since there now exists a simpler self consistency relationship:

$$\begin{aligned} q_M &= P(D_0 = 0) + P(D_0 \neq 0) q_M^2 \\ &= \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} q_M^2. \end{aligned} \tag{1.5}$$

This can be understood by considering that as soon as the ancestor gives birth, the system then essentially consists of two 'new' individuals (since individual histories are now irrelevant). Indeed, we can now think of individuals splitting in two rather than

giving birth, and the probability of it going extinct from this new state, $L = 2$, is then q_M^2. The probability that an individual manages to produce offspring (split) before dying is here the probability of one exponential random variable (time of first birth/split) being less than another independent exponential random variable (time of death). This explains the second line in equation (1.5). The probability of ultimate extinction when there are m ancestors is thus q_M^m. Equation (1.5) can be solved as a quadratic function to give:

$$q_M \in \{\mu/\lambda, 1\}. \tag{1.6}$$

However, note that if the expected number of offspring of an individual is above 1, then the lower value, μ/λ, is the consistent solution. Otherwise, ultimate extinction is certain.

By the correspondence between the Markovian branching process and the Markovian standard SIR model (which gives $\lambda = \beta(N-m)$ and $\mu = \gamma$) we can now write

$$\begin{aligned} P^m(\text{epidemic/invasion}) &= 1 - q_M^m \\ &= \begin{cases} 1 - \left[\frac{\gamma}{\beta(N-m)}\right]^m & \text{if } \frac{\beta(N-m)}{\gamma} > 1 \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{1.7}$$

where P^m(epidemic/invasion) is the probability of a 'major' epidemic in the Markovian standard SIR model when there are initially m infecteds and the rest of the population are susceptible. This definition is more valid for larger populations since then the correspondence between the epidemic process and the branching process is more exact (for the early stage of the process). Note that since the ultimate extinction threshold for the Markovian branching process is at $\lambda/\mu = 1$ then the corresponding threshold for the Markovian standard SIR model is at $\beta(N-m)/\gamma = 1$. This agrees exactly with the same threshold in the deterministic SIR model, $R_0 = \beta N/\gamma = 1$, for the case where βN is held constant as $N \to \infty$.

1.5 The standard SIS stochastic model (Markovian)

In the stochastic version of the deterministic SIS model, which we call the Markovian standard SIS model (proposed by Weiss and Dishon (1971)), individuals behave in exactly the same way as in the Markovian standard SIR model except that after their infectious periods individuals return to the susceptible state (there is no recovered/removed state). The transitions for the Markovian standard SIS model are shown in tables 1.4 and 1.5, which can be understood with reference to tables 1.1 and 1.2 for the Markovian standard SIR model (and the explanation given there).

It can be proved that in the limit of large population size the Markovian standard SIS model converges to the deterministic SIS model, i.e. $Y(t)/N \to I(t)$ uniformly on

Table 1.4: Individual level transitions for the Markovian standard SIS model

from	to	at rate
$\sigma : \sigma_i = S$	$\sigma^{i \to I}$	$\beta \sum_{j \in V} \mathbb{1}(\sigma_j = I)$
$\sigma : \sigma_i = I$	$\sigma^{i \to S}$	γ

Table 1.5: Population level transitions for the Markovian standard SIS model

from	to	at rate
(x, y)	$(x - 1, y + 1)$	βxy
(x, y)	$(x + 1, y - 1)$	γy

bounded intervals for agreeing parametrisation and initial conditions (Ethier and Kurtz, 1986; Andersson and Britton, 2000). Also, we can couple the Markovian standard SIS model to the same branching process that we coupled to the Markovian standard SIR model. The logic follows through in exactly the same way except that the time T (see section 1.4.2) now gives a lower bound on the actual time at which the correspondence breaks down. This is because an individual can be infected more than once in the SIS model, and it is only when an individual that is *currently* infected receives an extra infectious contact that the correspondence breaks. Therefore, we get the same threshold and probability of an epidemic/invasion as for the Markovian standard SIR model (again, valid for large populations). Note that the probability of an epidemic/invasion from a single initial infected in the Markovian standard SIS model is thus equal to $I(\infty)$ in the endemic equilibrium of the deterministic SIS model (when above the threshold). It should be stressed that for the stochastic SIS model the eventual outcome is always extinction of the infection (given enough time) since, given any present state, the probability of the all-susceptible (absorbing) state arising in any future time interval is positive.

1.6 Quasi-stationary distributions (for conditional final outcomes)

Let us consider a continuous-time absorbing Markov chain, with finite state space $\mathcal{S}$ and generator Q, in which there is a single absorbing state labelled 0 and the set $\mathcal{S} \setminus \{0\}$ is a communicating class of transient states (as in the Markovian standard SIS model). In this case, there exists a unique quasi-stationary distribution (for the transient states) such that, if the system is initiated in a transient state and we condition on non-absorption, then the system (restricted to $\mathcal{S} \setminus \{0\}$) tends towards this distribution (Daroch and Seneta, 1965, 1967). For a set $A \subset \mathcal{S}$ of absorbing states, conditioning on non-absorption yields a unique quasi stationary distribution (QSD) if $\mathcal{S} \setminus A$ is a communicating class of transient states.

Following Nåsell (1996), let $p(t)$ be the row vector such that $p_1(t)$ is the probability that the system is in state 1 (after some arbitrary ordering) at time t. We will assume that there is a single absorbing state labelled 0 and that the system is initiated in a transient state. Let $p_Q(t)$ be the same as $p(t)$ but with the first component $p_0(t)$ removed. We can now define

$$q(t) = \frac{p_Q(t)}{1 - p_0(t)}, \tag{1.8}$$

such that $q_1(t)$ is the probability that the system is in state 1 at time t given that absorption does not occur.

From the theory of continuous time Markov chains (Grimmett and Stirzaker, 1982), we have

$$p(t) = p(0)e^{Qt}, \tag{1.9}$$

and so $p_Q(t) = q(0)e^{A_Q t}$, where A_Q is the same as Q but with the first row and column removed. We can now write:

$$\begin{aligned} q(0)e^{A_Q t} &= (1 - p_0(t))q(t) \\ &= (q(0)e^{A_Q t}\mathbb{1})q(t), \end{aligned} \tag{1.10}$$

where $\mathbb{1}$ is an $(|\mathcal{S}| - 1) \times 1$ column vector of ones. The quasi-stationary distribution q^* is defined to satisfy:

$$q^* e^{A_Q t} = (q^* e^{A_Q t}\mathbb{1})q^* \quad \forall t > 0, \tag{1.11}$$

and therefore it must be an eigenvector of $e^{A_Q t}$. Since A_Q and $e^{A_Q t}(t > 0)$ share the same eigenvectors and $e^{A_Q t}(t > 0)$ is a positive matrix, due to our transient states forming a communicating class, we can make use of the Perron-Frobenius theorem (see, for example, Grimmett and Stirzaker (1982)). Specifically, $e^{A_Q t}$ has just one left eigenvector v such that its components are all real and non-negative (equivalently non-positive), and is therefore the only candidate for a probability distribution. This eigenvector corresponds to a real eigenvalue $r(t)$ which is positive and greater in absolute value than any other eigenvalue. This being the case, we have $r = e^{\rho_1 t}$ where ρ_1 is the eigenvalue of A_Q which corresponds to v. In other words, if q^* exists, it must be proportional to v and $q^* e^{A_Q t}\mathbb{1} = e^{\rho_1 t}$. However, since it is clear that $v e^{A_Q t}\mathbb{1} = e^{\rho_1 t}$ (after we have normalised v such that $\sum_k v_k = 1$) then we do indeed have a unique quasi-stationary distribution $q^*(\propto v)$. Note that $q^* e^{A_Q t}\mathbb{1} = e^{\rho_1 t}$ implies that the time until absorption, when $p(0) = q^*$, is an exponential random variable with parameter ρ_1.

It is also straightforward to show that q^* is a limiting distribution. In fact, we can write

$$q(t) = q^* + O(e^{t(\rho' - \rho_1)}), \tag{1.12}$$

where $\rho' < \rho_1 < 0$ and so q^* is the limit as $t \to \infty$. We can see that q^* may have great practical relevance when $\rho' << \rho_1$ and $\rho_1 \approx 0$ since then q^* can potentially approximate $p_Q(t)$ for a significant time period. This would correspond to the case where the expected time to absorption is long and the rate of convergence of $q(t)$ to q^* is rapid.

1.7 Networks (graphs representing populations with contact stucture)

The models so far described have all made use of the assumption that populations are 'evenly mixed' such that, at any given time, all susceptible individuals are equally at risk of being infected. In reality, certain individuals are more isolated while others are better connected and it is intuitive that some individuals may play a more important role in spreading the infection than others. For example, consider the role played by promiscuous individuals in the case of sexually transmitted diseases (STDs). Additionally, the existence of communities or households, within which the infection may spread more rapidly, will clearly affect real world disease dynamics.

In models which assume 'even mixing', the large population limit is usually taken in order to simplify the mathematics. This limit is taken in such a way that two types of event, which may have important ramifications in the real world, are not captured by the models: repeated infectious contacts between a given pair of individuals and, on short time scales, infectious contacts to individuals who have already been infected by different individuals. By ignoring such events the models may overestimate the initial progress and final impact of the disease. These models also ignore the existence of individuals who are likely to remain susceptible for a long period, or indefinitely, primarily due to their 'distance' from initially infected individuals.

In order to avoid these limitations, and to allow more realistic heterogeneity, the idea of a 'social network' can be introduced. The social network is conceptualised as a set of individuals or 'social actors' where the presence of a relationship between any two specific individuals is recorded, and its 'strength' is quantified (the individuals in such a relationship are then 'neighbours'). Indeed, the real-world propagation of disease is an inherently stochastic phenomenon in which the 'infection event' is fundamental. Such an event (generally) involves just two individuals i.e. an infectious individual transmits the infection to a previously uninfected, yet susceptible, individual. Therefore it is natural to try and quantify the pair-wise relationships in a given population. Note also that by introducing a finite network which models each individual explicitly, we can avoid the assumption that all individual hosts interact with the infectious agent in the same way (in terms of infectivity and recovery profiles).

In theory, the process of disease propagation could be described (probabilistically) in terms of a time-dependent matrix $T(t)$ where the product $T_{ij}(t)\Delta$ is the probability

that individual j will make a contact (sufficient to transmit the disease) to individual i between time t and time $t+\Delta$ (for $\Delta \to 0$). It would then be necessary to superimpose the propagation of the infectious agent on to this existing behaviour, resulting in an extremely complex system. However, even at this level of generality we have assumed that the behaviour of the host population is largely independent of the effects of the propagating infectious agent. In light of these issues it may be preferable, for example, to let $T_{ij}(t)\Delta$ instead represent the probability that j will make a sufficient contact to i between time t and time $t+\Delta$, where time is measured from the moment that j becomes infected (assuming that it does). In this case T captures the behaviour of the population which is most relevant to the propagation of the disease, and allows more interaction between the behaviour of the host population and the presence of the infectious agent, while also offering a way around the problem of superimposing the dynamics of the disease on to some prescribed behaviour for the host population. Indeed, this kind of approach is adopted by Karrer and Newman (2010). Note that if the system is assumed to be such that, given some present state, the future and the past are independent, i.e. the Markovian case, then the difficulties discussed here are ameliorated.

Throughout this thesis, social networks will be defined as directed graphs $D = (V, A)$ where V is the set of all individuals (vertices/nodes) in the network (population) and A is a set of arcs (ordered pairs). The existence of an arc $(i, j) \in A$ corresponds to the ability of $i \in V$ to make direct contacts to $j \in V$. We also assume an arbitrary labelling such that the set V is given a one-to-one correspondence to the integers $\{1, 2, \ldots, N = |V|\}$. Similarly, we will sometimes define undirected networks as undirected graphs $G = (V, E)$ where E is a set of edges (unordered pairs) such that $(i, j) \in E$ indicates that i and j are in direct contact. By incorporating networks into the models it is possible to represent disease transmission as a collection of processes which take place between interacting pairs of individuals such that an individual's position in the network is highly relevant.

We note that contact structure is itself random and changing over time. However, it may be that the propagation process is sufficiently rapid such that the assumption of a static contact structure (during the relevant time period) is less strong. Some types of networks are naturally more static, such as networks of sexual partners in non-promiscuous societies, or networks of computer nodes in technological systems.

For a given network/directed graph $D = (V, A)$, we make the following definitions:

Definition 1.7.1. *The 'adjacency matrix' is a $|V| \times |V|$ matrix where the $(i, j)^{th}$ entry is 1 if $(j, i) \in A$ and zero otherwise.*

Definition 1.7.2. *Individual $i \in V$ has a set of 'in-neighbours' $\{j : (j, i) \in A\}$, a set of 'out-neighbours' $\{j : (i, j) \in A\}$ and a set of 'neighbours' $\{j : (i, j) \in A$ or $(j, i) \in A\}$.*

Definition 1.7.3. *A 'path' from $i \in V$ to $j \in V$ is a sequence of individuals $i = v_1, v_2, v_3, \ldots, j = v_n$ where $(v_k, v_{k+1}) \in A$ and where no individual appears more than once. We say that $i \in V$ can reach $j \in V$ iff there exists some path from i to j. Similarly, we say that $i \in V$ can be reached from $j \in V$ iff there exists some path from j to i. A 'cycle' is any path where there exists an arc from the last individual in the sequence to the first.*

Definition 1.7.4. *The 'in-component' of $A \subset V$, denoted $In(A)$, is the set of all individuals from which at least one member of A can be reached. The 'out-component' of A, denoted $Out(A)$, is the set of all individuals which can be reached from at least one member of A. We assume that $A \subset In(A)$ and $A \subset Out(A)$.*

Definition 1.7.5. *a 'subgraph' is a graph with a vertex set which is a subset of V, say $B \subset V$, and an arc set which is a subset of A (restricted to arcs where both ends are in B). The vertex induced subgraph $D[B]$, where $B \subset V$, is the graph consisting of vertex set B where there is an arc from $i \in B$ to $j \in B$ if and only if there is an arc from i to j in the original network D.*

Definition 1.7.6. *The 'underlying graph/network' is the graph with vertex set V where there is an arc from $i \in V$ to $j \in V$ iff there is an arc from i to j, or from j to i, in the original network D. We will refer to this underlying graph as undirected, since the existence of an arc from i to j implies the existence of an arc from j to i.*

Definition 1.7.7. *The network is 'strongly connected' iff every individual can be reached from every other individual. A strongly connected component is a subgraph which is strongly connected and maximal with respect to this property. The network is 'weakly connected' iff its underlying graph is strongly connected.*

1.7.1 Erdös-Rényi random graphs

Early work on random graphs was conducted by Erdös and Rényi (1959). They defined their random graph via the following construction 'recipe': 1) Let V be a collection of $N = |V|$ vertices 2) for every unordered pair of vertices, place an edge connecting them together with probability p (independently for each pair). It is immediately obvious that the expected number z of edges emanating from a randomly selected vertex, i.e. its expected degree, will then be equal to $p(N-1)$. By taking the limit as N tends to infinity, whilst keeping z constant, many important results concerning the structure of the random graph can be proved. The Erdös-Rényi random graph, of size N, is the graph-valued random element with possible values and distribution determined by the above recipe, i.e. the probability of it taking a particular structure is precisely the probability of that structure arising when following the recipe.

The probability p_k that a vertex will be of degree k is given by

$$\begin{aligned} p_k &= \binom{N-1}{k} p^k (1-p)^{N-1-k} \\ &\approx \frac{e^{-z} z^k}{k!}, \end{aligned} \tag{1.13}$$

where the approximation becomes exact as $N \to \infty$ (and $p \to 0$, since z is kept constant).

Modelling a population by this random graph is arguably more realistic than the approach which assumes that all individuals interact equally with all others. Indeed, it allows us to obtain results for the case where, in the limit of large population size, the number of neighbours of individuals is kept small, and the contact structure is heterogeneous. However, the Poisson degree distribution is not considered realistic for most social or technological networks. In addition, the local structure around a given node is almost surely tree like. The first of these shortcomings is overcome by way of generalised random networks.

1.7.2 Generalised random graphs

Generalised random graphs can be used to approximate 'real' networks when there is limited information. For example, the fraction of the population p_k consisting of individuals who commonly interact with precisely k others may be known (or might be confidently estimated), while other information is lacking. Given this distribution over k, one can consider properties of the random graph of (population) size N, where the probability of any given individual having k neighbours is p_k, but which is in all other respects uniformly random. We will refer to such a generalised random graph as a 'configuration network' (Bender and Canfield, 1978; Bollobás, 1985; Molloy and Reed, 1995; Newman et al., 2001). It is amenable to mathematical analysis (especially as $N \to \infty$) but has the limitation of not being able to capture the higher order structure of the real network, such as the likelihood that two neighbours of a given individual will be neighbouring to each other, or the correlation in degree between neighbouring individuals, e.g. the fraction of all individuals that are neighbouring to individuals of degree k that are also of degree k.

The 'recipe' for generating a realisation of a configuration network goes as follows: 1) Let V be a collection of $N = |V|$ vertices. 2) attach to each individual a number of 'stubs', where the number is an independent draw from the degree distribution p_k; if the total number of stubs is odd then start again. 3) select a pair of stubs uniformly at random from the set of all such pairs, and connect them via an edge, and keep doing this until there are no more available stubs. Thus, the probability of a randomly selected vertex having precisely k neighbours is p_k. However, an extremely important characteristic, resulting from the form of the above recipe, is that the probability of

a randomly selected neighbour of a randomly selected individual (given that it has at least one neighbour) having degree k is not given by p_k, but by kp_k/z (consider that the expected number of stubs attached to individuals with k stubs is kp_kN and the expected total number of stubs is zN). This means that 'your friend has more friends than you do' and this is reflected in real world networks. Indeed, for a finite simple graph (non-random), it can be shown that the expected degree of an individual at the (uniformly) randomly chosen end of a (uniformly) randomly chosen edge is greater than the average degree (Feld, 1991). Another important characteristic of configuration networks is that in the limit as $N \to \infty$ the structure is locally tree-like (Newman et al., 2001), in the sense that the probability of a randomly selected individual being contained in a cycle (loop) of finite size $n \in \mathbb{N}$ tends to zero as $N \to \infty$ (assuming that the expectation and variance of the degree of a randomly selected individual remain finite: Dorogovtsev, 2010; Newman et al., 2001).

As $N \to \infty$, the expected number of individuals that are at a distance of (finite) n edges from a randomly selected individual is

$$z\left[\frac{\sum_{k=0}^{\infty} k(k-1)p_k}{z}\right]^{n-1}.$$

Thus, there is a threshold where the quantity in square brackets is equal to 1, this quantity being the expected number of *extra* edges belonging to an individual arrived at via an edge emanating from a randomly selected individual. Below the threshold, the expected fraction of the network that can be reached from a randomly selected individual is vanishingly small. Above the threshold, the expected number of individuals reachable from a randomly selected individual tends to infinity as $N \to \infty$. However, even above the threshold, there is still the possibility that an individual will only be able to reach a vanishingly small fraction of the network. This echoes our earlier description of the threshold for branching processes.

Similarly to the probability of ultimate extinction in branching processes, the probability S that a randomly selected individual is connected only to a vanishingly small fraction of the configuration network is given by (Newman et al., 2001):

$$S = \sum_{k=0}^{\infty} p_k u^k,$$

where u is the probability of being able to reach only a vanishingly small fraction after traversing one edge (and not coming back along it). This follows from the fact that if the individual has k neighbours, which occurs with probability p_k, then all k of its edges must lead to a vanishingly small fraction of the network, and there can be nothing which favours one edge over another with respect to this property. Since we have already computed the distribution for the number of *extra* edges emanating from a neighbour of a randomly selected individual, we know that u must satisfy the self

consistency relationship (Newman et al., 2001):

$$u = \frac{1}{z}\sum_{k=1}^{\infty} k p_k u^{k-1}.$$

Note that besides the 'permanent' $u = 1$ solution, there can be at most one other solution in [0,1] (due to the convexity of the right-hand-side for $u \in (0,1)$), and it is straightforward to show that it appears when above the threshold (Newman et al., 2001).

It is said that, above the threshold, the network/graph 'percolates' such that a single connected component appears that is of the order of N in size, this being the 'giant component' (Newman et al., 2001). Unlike the branching process, we therefore have two interpretations of the probability $1 - S$. It is the probability that a randomly selected vertex be connected to a positive fraction of the network, and it is also the relative size of the giant component. Indeed, there is a branch of mathematics named 'percolation theory' which addresses the issue of the existence and size of giant components for different random graphs (and the size distribution for non-giant connected components). This kind of analysis was especially simple for configuration networks because of the locally tree like structure. In contrast, starting with an infinite square lattice and then removing every edge independently with probability p results in a giant component if and only if $p < 1/2$ - this took decades of work to prove (see, for example, Grimmett (2010)).

1.8 Invasion and final outcome for epidemics on generalised random graphs (via percolation theory)

The idea that an infection spreading on a configuration network is equivalent to a percolation process was discussed by Newman (2002), but the relationship between the spread of disease on an arbitrary network and percolation was recognised earlier by Frisch and Hammersley (1963) and Grassberger (1983). The key to understanding the relationship between these processes is by noting that the marginal probability that a given individual, when it becomes infected, will make at least one infectious contact to another given individual can sometimes be calculated. Indeed, for the Markovian standard SIR model this probability is $\beta/(\beta+\gamma)$. Therefore, if we have a graph where each edge represents the possibility of an infectious contact, and we remove each edge according to the marginal probability that the infectious contact via that edge does not occur, then the probability that there is a path from an initially infected individual to some other given individual is the probability that the given individual gets infected (at some point in time). However, as was pointed out by Kenah and Robins (2007), this mapping to a percolation process is only completely accurate if the infectious period is not random but fixed, otherwise a slight modification to the percolation process is

required, which we shall now discuss.

The problem arises from the fact that infectious contacts from a given individual are correlated since the probability of each one depends on the infectious period of the given individual. The way to account for this correlation is to first assign infectious periods to all individuals from the appropriate distribution(s) before computing the marginal probabilities of infectious contacts across each edge given this set of infectious periods (every edge must first be replaced by two oppositely directed arcs since the marginal probability of an infectious contact in one direction can be different to the other direction). The random network implied by this recipe is called the epidemic percolation network (EPN) and is a powerful theoretical and numerical tool (Kenah and Robins, 2007; Miller, 2009). Note that a single realisation of an EPN gives a statistically accurate 'simulation' of the final outcome of the epidemic process for all possible initial configurations (whether or not a particular individual gets infected just depends on how we independently assign the initial infecteds).

Let us consider the case where the infectious period t_I is fixed and the infection is spreading on a configuration network, and so an exact mapping to standard percolation is possible. The marginal probability P_I that a given individual makes an infectious contact to a given neighbour is then:

$$P_I = \int_0^{t_I} \beta e^{-\beta\tau} d\tau = 1 - e^{-\beta t_I}.$$

This means that the expected number of individuals (in the EPN) that are a distance of n edges from a randomly selected individual is:

$$P_I z \left[P_I \frac{\sum_{k=0}^{\infty} k(k-1)p_k}{z} \right]^{n-1},$$

and so there is a threshold where the quantity in square brackets, this being the expected number of secondary cases caused by an infected individual near the start of the process (not the index case), is equal to 1. Below the threshold, in the limit of large population size, a vanishingly small fraction of the network are infected, while above the threshold it is possible that a positive fraction of the network will be infected. The probability S that a randomly selected individual will be reachable from (and can reach) a vanishingly small fraction of the network is (Newman, 2002; Kenah and Robins, 2007):

$$S = \sum_{k=0}^{\infty} p_k u_I^k,$$

where u_I (to be interpreted similarly to u in the previous section) satisfies

$$u_I = (1 - P_I) + \frac{P_I}{z} \sum_{k=1}^{\infty} k p_k u_I^{k-1}.$$

Thus, we can interpret $1 - S$ as simultaneously the probability of a major epidemic (seeded by a single initially infected individual selected at random) and the relative

size of a major epidemic. In fact, these last equations enable us to find the correct relative size of a major epidemic even when the infectious period is random. This is because a given individual receiving an infectious contact *from* another given individual is (marginally) independent of it receiving an infectious contact from a different individual (both contacts depend on independent infectious periods) (Kenah and Robins, 2007). This means that these last equations give us the correct probability of a randomly selected individual being reachable *from* a vanishingly small fraction of the network in the EPN. Note that if the degree distribution for the configuration network is set to $p_{N-1} = 1$, and $N \to \infty$ such that we have an infinite fully connected network (which is not locally tree like, but the corresponding EPN is, since the expected number of infectious contacts from a given individual is finite), then the probability S of a randomly selected individual being reachable from a vanishingly small fraction of the network in the EPN satisfies

$$\begin{aligned} S &\approx (1 - P_I + P_I S)^{N-1} \\ &= (1 - P_I(1 - S))^{N-1} \\ &\approx e^{-P_I N(1-S)}, \end{aligned} \tag{1.14}$$

where the approximations are exact in the limit as $N \to \infty$. The last line follows since $P_I N$ is kept constant while $N \to \infty$ and $P_I \to 0$. If, as for the Markovian standard SIR model, we have $P_I = \beta/(\beta + \gamma)$ (where βN is held constant as $N \to \infty$) then it is straightforward to show that S is here the same as the value $S(\infty)$ $(= 1 - R(\infty))$ given by equation (1.2) for the final outcome of the deterministic SIR model (in the case $S(0) \to 1$). This was to be expected since, as stated earlier, the Markovian standard SIR model and the deterministic SIR model 'coincide' in the limit of large population size (Ethier and Kurtz, 1986; Andersson and Britton, 2000).

1.9 Capturing the dynamics of epidemics on generalised random graphs

Equations determining the exact time series for epidemics on configuration networks were first given by Volz (2008), but we will discuss the equivalent and simpler system of equations derived by Miller (2011) (see also Eames and Keeling (2002) for a different approach to dynamics on heterogeneous networks). Miller has described his method as 'edge-based compartmental modelling'. It relies on the concept of the 'cavity state' (see, for example, Karrer and Newman (2010)) which allows a simpler conceptual framework and is defined as follows: an individual is placed in the cavity state by removing its ability to transmit the infection while leaving the process by which it becomes infected unchanged. Thus, whether or not a given individual will be infected and the random time at which it becomes infected are unchanged if we first place the

individual in the cavity state. If the network is a tree network, an initially-susceptible individual in the cavity state is susceptible at time t if it has not received an infectious contact from any of its neighbours and, crucially, it receiving no such contacts from one neighbour is statistically independent of it receiving no such contacts from a different neighbour. In configuration networks, which are locally tree-like, this assertion of statistical independence is valid after assuming that the infection survives long enough for a positive fraction of the population to become infected (or if this is the case at $t = 0$). Another feature of this approach which makes the mathematics tractable is that, no matter how many neighbours a given cavity-state individual has, the probability of it receiving no infectious contacts from a given neighbour (the degree of which is unknown) by time t is always the same.

Following Miller (2011), and letting $\theta(t)$ denote the probability that a randomly selected individual, after having been placed in the cavity state, does not receive an infectious contact from a given (unknown) neighbour by time t:

$$\begin{aligned} \langle S \rangle &= \sum_{k=0}^{\infty} p_k [\theta(t)]^k, \\ \langle I \rangle &= 1 - \langle S \rangle - \langle R \rangle, \\ \langle \dot{R} \rangle &= \gamma \langle I \rangle, \end{aligned} \tag{1.15}$$

where $\langle X \rangle$ is the fraction of the population in state X at time t, γ is the (exponential) rate at which an infected individual recovers, and we use 'dot' notation for time derivatives. Implicit in the above construction is the assumption that a randomly selected individual is susceptible at $t = 0$ with probability 1 i.e. a vanishingly small fraction of the population is initially infected. By similarly assuming that a neighbour of a randomly selected individual will be susceptible at $t = 0$, Miller then derives:

$$\dot{\theta(t)} = -\beta \left[\theta(t) - \frac{\sum_{k=0}^{\infty} k p_k [\theta(t)]^{k-1}}{\sum_{k=0}^{\infty} k p_k} \right] + \gamma(1 - \theta(t)), \tag{1.16}$$

where β is the rate at which an infected individual makes infectious contacts to one of its neighbours. The quantity in large square brackets is the probability that a neighbour of a randomly selected individual has received an infectious contact by time t but has not made an infectious contact to our selected individual (note that this probability has been derived by placing the neighbour also into the cavity state, with the exception that it is allowed to make contacts to our selected individual). Note that $\theta(0) = 1$ and $\dot{\theta}(0) = 0$, and so we need a system state at time $t + \Delta$, where $\theta(t + \Delta) = 1 - e$ and $0 < e << 1$, from which we can then compute the time series (Miller, 2011).

The edge-based approach can be applied to a wide range of generalised random networks and even dynamic networks (Miller et al., 2011). The principle of selecting a random individual, placing it in the cavity state, and then computing $\theta(t)$ remains the same but the computation of $\theta(t)$ itself (or its time derivative) always needs careful consideration.

1.10 Epidemics on finite non-random networks

We define the Markovian network-based SIR model to be the same as the Markovian standard SIR model (section 1.4), except that we let individual j make infectious contacts to individual i at rate $T_{ij} \geq 0$ (when j is infected), and individual i recovers at rate γ_i (when it is infected). We let Γ denote the vector $[\gamma_1, \gamma_2, \ldots, \gamma_N]$. The transition rates for the corresponding Markov chain are shown in table 1.6, which can be understood with reference to table 1.1 for the Markovian standard SIR model (and the explanation given there). Similarly, we define the Markovian network-based SIS model according to the transition rates in table 1.7. The network $D = (V, A)$ on which the disease

Table 1.6: Transitions for the Markovian network-based SIR model

from	to	at rate
$\sigma : \sigma_i = S$	$\sigma^{i \to I}$	$\sum_{j \in V} T_{ij} \mathbb{1}(\sigma_j = I)$
$\sigma : \sigma_i = I$	$\sigma^{i \to R}$	γ_i

Table 1.7: Transitions for the Markovian network-based SIS model

from	to	at rate
$\sigma : \sigma_i = S$	$\sigma^{i \to I}$	$\sum_{j \in V} T_{ij} \mathbb{1}(\sigma_j = I)$
$\sigma : \sigma_i = I$	$\sigma^{i \to S}$	γ_i

spreads is thus implied by the matrix T, i.e. $T_{ij} > 0 \Leftrightarrow (j, i) \in A(D)$. Note that, since we allow T to be asymmetric, this model is able to effectively model heterogeneity in the innate susceptibility and infectivity of individuals. The case where $T_{ij} > T_{ji}$ may also be employed to capture the situation where i adopts an anti-spreading behaviour, like frequent hand-washing, whereas j does not, or, if i and j represent communities, the situation where there is more migration from j to i than from i to j (Sharkey et al. (2006)).

The introduction of the contact network has broken the symmetry which existed for the Markovian standard SIR and SIS models such that it now matters *which* individuals are in which states, i.e. there is no corresponding Markov chain where the state of the system can be defined simply in terms of the number of individuals in each compartment. This means that numerical simulation for such network-based models is more computationally intensive since, for the Markovian network-based SIR (SIS) model, there are $3^N (2^N)$ states to keep track of. Nonetheless, the models are still continuous time Markov chains and so relatively simple to simulate. Note also that the generator matrices will be extremely sparse since, for a given present state, we know that the next state can only differ by the status of a single individual. The rate of change of the status of a single individual is easily computed from the current state of the system and table 1.6 or 1.7.

It is possible to generalise these models such that a given infected individual makes infectious contacts to a given neighbour according to some arbitrary non-Markovian stochastic process, and/or a given individual's infectious period has an arbitrary distribution. Extra compartments, such as an 'exposed' class, can also be added. However, the network on which the disease spreads still defines (and is defined by) which individuals are in direct contact, and the network remains static in time.

For arbitrary finite networks, each individual is (potentially) unique in its ability to spread the infection and so there is no simple branching process to which the progress of the infection can be coupled. The expected number of nth generation infecteds depends on which individuals are initially infected and no longer grows (or reduces) geometrically with n. There is no invasion threshold and, indeed, what an invasion would 'look like' is now obscure.

Definition 1.10.1. *Let X_iY_j denote the indicator variable for the event that $\sigma_i(t) = X, \sigma_j(t) = Y$, where $\sigma(t)$ is the random state of a stochastic process at time t, taking values in $\mathcal{S} = \mathcal{S}_{ind}^V$ $(i, j \in V)$, and where X, Y are individual-states belonging to $\mathcal{S}_{ind}$. Thus we can write:*

$$\begin{aligned}\sum_{\sigma\in\mathcal{S}:\sigma_i=X,\sigma_j=Y} \mathbb{P}(\sigma(t)=\sigma) &= \mathbb{P}(\sigma_i(t)=X,\sigma_j(t)=Y)\\ &= \langle X_iY_j\rangle,\end{aligned}$$

where the angled brackets denote expectation. In other words, $\langle X_iY_j\rangle$ is the probability that, at time t, individual i is in state X and individual j is in state Y. This generalises to any number of individuals $X_iY_jZ_k\ldots$

Remark. When referring to such indicator variables, the probability space on which they are defined will be clear from the context, i.e. the model under consideration. For example, in the Markovian standard SIR model, $\{\sigma(t)\}$ is a continuous time Markov chain where $\mathcal{S}_{ind} = \{S, I, R\}$ and $\mathcal{S} = \{S, I, R\}^V$. Unless otherwise stated, all the random elements in any equation or system of equations are defined on a single (arbitrary) probability space which is consistent with a 'run' of the model. Hence, reference to initial conditions and parametrisation will usually be dropped.

1.10.1 Capturing the dynamics by a moment closure approach

After careful consideration of the transition rates in table 1.1 (section 1.4), it is straightforward to see that for the Markovian standard SIR model:

$$\begin{aligned}\langle\dot{S}_i\rangle &= -\beta\sum_{j\in V\setminus i}\langle S_iI_j\rangle,\\ \langle\dot{I}_i\rangle &= \beta\sum_{j\in V\setminus i}\langle S_iI_j\rangle - \gamma\langle I_i\rangle,\\ \langle\dot{R}_i\rangle &= \gamma\langle I_i\rangle,\end{aligned} \tag{1.17}$$

where $\langle S_i \rangle$, $\langle I_i \rangle$ and $\langle R_i \rangle$ are the probabilities that individual i is susceptible, infected and recovered respectively at time t and $\langle S_i I_j \rangle$ is the probability that i is susceptible and j is infected at time t (see definition 1.10.1). We can now sum over i to obtain the rates of change of the expected number in each compartment:

$$\begin{aligned} [\dot{S}] &= -\beta[SI], \\ [\dot{I}] &= \beta[SI] - \gamma[I], \\ [\dot{R}] &= \gamma[I], \end{aligned} \tag{1.18}$$

where $[S] = E[X(t)]$, $[I] = E[Y(t)]$ and $[R] = E[N - X(t) - Y(t)]$ are the expected number that are susceptible, infected and recovered respectively at time t, and $[SI]$ is the expected number of ordered pairs (of individuals) where the first is susceptible and the second is infected at time t. System (1.18), although exact (consistent with the Markovian standard SIR model), is not very useful without also having an equation which governs the change in the expected number of the pairs. However, note that the number of such ordered pairs at time t is given by $X(t)Y(t)$, and if we assume that $E[X(t)Y(t)] = E[X(t)]E[Y(t)]$, as would be the case if $X(t)$ and $Y(t)$ were statistically independent, then we get:

$$\begin{aligned} [\dot{S}]^* &= -\beta[S]^*[I]^*, \\ [\dot{I}]^* &= \beta[S]^*[I]^* - \gamma[I]^*, \\ [\dot{R}]^* &= \gamma[I]^*. \end{aligned} \tag{1.19}$$

After expressing in terms of fractions (of N), this is exactly the same system as the deterministic SIR model (the 'stars' indicate that the variables in the system are now approximations of the corresponding quantities for the underlying stochastic model). System (1.19) demonstrates the principle of moment closure (for this context) since it was generated from an exact (but not closed) system of equations by approximating the higher-order quantities (pairs of individuals) in terms of lower order quantities (individuals).

A slightly weaker assumption is that the states of individuals are statistically independent. This is weaker since it is (in some sense) implied by statistical independence of $X(t)$ and $Y(t)$, but does not imply statistical independence of $X(t)$ and $Y(t)$, e.g. statistical independence of individuals implies $E[X(t)Y(t)] = \sum_{i \in V} \sum_{j \in V \setminus i} \langle S_i \rangle \langle I_j \rangle = E[X(t)]E[Y(t)] - \sum_{i \in V} \langle S_i \rangle \langle I_i \rangle$. This assumption leads to:

$$\begin{aligned} \langle \dot{S_i} \rangle^* &= -\beta \sum_{j \in V \setminus i} \langle S_i \rangle^* \langle I_j \rangle^*, \\ \langle \dot{I_i} \rangle^* &= \beta \sum_{j \in V \setminus i} \langle S_i \rangle^* \langle I_j \rangle^* - \gamma \langle I_i \rangle^*, \\ \langle \dot{R_i} \rangle^* &= \gamma \langle I_i \rangle^*, \end{aligned} \tag{1.20}$$

which is a closed solvable system with a large number of variables (order N). To reduce the size of this system, one can make the individual level homogeneity assumption that the trajectories of all individuals are the same and so, for example, $\langle S_i \rangle = [S]/N \forall i$. Note that, due to symmetry, this assumption holds true if the probabilistic states of individuals are all the same at $t = 0$ (and the states of individuals are initially independent), e.g. $\langle S_i \rangle = \langle S_j \rangle \forall i, j$ at $t = 0$. Also, if we start the system in any pure state then all initially susceptible individuals follow the same trajectories, and similarly for initially infected individuals. After summing over all $i \in V$ in equation (1.20), this homogenity assumption allows us to write:

$$\begin{aligned} [\dot{S}]^* &= -\beta \frac{(N-1)}{N} [S]^* [I]^*, \\ [\dot{I}]^* &= \beta \frac{(N-1)}{N} [S]^* [I]^* - \gamma [I]^*, \\ [\dot{R}]^* &= \gamma [I]^*, \end{aligned} \tag{1.21}$$

which provides an alternative to system (1.19) which assumed statistical independence of $X(t)$ and $Y(t)$.

For the Markovian network-based SIR model, the same assumptions can be used in order to produce small solvable systems. Consideration of the transition rates for this model (section 1.10, table 1.6) leads to:

$$\begin{aligned} \langle \dot{S}_i \rangle &= -\sum_{j \in V \setminus i} T_{ij} \langle S_i I_j \rangle, \\ \langle \dot{I}_i \rangle &= \sum_{j \in V \setminus i} T_{ij} \langle S_i I_j \rangle - \gamma_i \langle I_i \rangle, \\ \langle \dot{R}_i \rangle &= \gamma_i \langle I_i \rangle, \end{aligned} \tag{1.22}$$

which are exact equations, in terms of the underlying stochastic model, but they do not form a closed system. To close the system, one can assume statistical independence for the states of individuals:

$$\begin{aligned} \langle \dot{S_i} \rangle^* &= -\sum_{j \in V \setminus i} T_{ij} \langle S_i \rangle^* \langle I_j \rangle^*, \\ \langle \dot{I_i} \rangle^* &= \sum_{j \in V \setminus i} T_{ij} \langle S_i \rangle^* \langle I_j \rangle^* - \gamma_i \langle I_i \rangle^*, \\ \langle \dot{R_i} \rangle^* &= \gamma_i \langle I_i \rangle^*, \end{aligned} \tag{1.23}$$

and this is a closed solvable system of order N. The assumption of statistical independence in the states of individuals for this model is generally stronger than for the Markovian standard SIR model since each individual (potentially) has a small number of neighbours. For an infected individual, at least one of this small number of neighbours is not susceptible - the individual who infected it.

To reduce the system size, the individual level homogeneity assumption can be made, giving:

$$\begin{aligned}
[\dot{S}]^* &= -\bar{T}[S]^*[I]^*, \\
[\dot{I}]^* &= \bar{T}[S]^*[I]^* - \bar{\gamma}[I]^*, \\
[\dot{R}]^* &= \bar{\gamma}[I]^*,
\end{aligned} \tag{1.24}$$

where

$$\bar{T} = \frac{1}{N^2}\sum_{i,j\in V} T_{ij} = \frac{n\bar{\beta}}{N}, \tag{1.25}$$

$$\bar{\gamma} = \frac{1}{N}\sum_{i\in V}\gamma_i, \tag{1.26}$$

and n is the average number of out-neighbours (= average number of in-neighbours) of an individual in the population, while $\bar{\beta}$ is the average rate at which an infected individual makes infectious contacts to one of its out-neighbours. The network structure, encoded in the matrix T, drastically reduces the symmetry of the system and makes the homogeneity assumption (generally) much stronger.

Going back to the exact equations (1.22), let us now write down exact expressions for the rates of change of $\langle S_i I_j\rangle$ and $\langle S_i S_j\rangle \forall i,j : T_{ij} > 0$:

$$\begin{aligned}
\dot{\langle S_i I_j\rangle} &= -(T_{ij}+\gamma_j)\langle S_i I_j\rangle \\
&\quad + \sum_{k\in N_j\setminus i} T_{jk}\langle S_i S_j I_k\rangle \\
&\quad - \sum_{k\in N_i\setminus j} T_{ik}\langle I_k S_i I_j\rangle, \\
\dot{\langle S_i S_j\rangle} &= -\sum_{k\in N_j\setminus i} T_{jk}\langle S_i S_j I_k\rangle \\
&\quad - \sum_{k\in N_i\setminus j} T_{ik}\langle I_k S_i S_j\rangle.
\end{aligned} \tag{1.27}$$

Here, we can achieve a closed system by attempting to represent the triples, e.g. $\langle S_i S_j I_k\rangle$, in terms of pairs and individuals. First, we assume that for any given individual in a given state, knowledge of the state of one of its neighbours cannot provide any extra information about the states of any of the other neighbours. Note that this does not preclude correlations in the states of neighbouring individuals. This is the heart of many so-called 'pair-wise' (or pair-based) approximation models, where correlations in the states of neighbouring individuals are attempted to be preserved while correlations between individuals further apart are ignored. This assumption leads to

the following closures (see Sharkey (2008) and references therein):

$$\begin{aligned}
\langle A_i B_j C_k \rangle &\approx \langle A_i B_j \rangle \frac{\langle B_j C_k \rangle}{\langle B_j \rangle} \qquad G[i,j,k] \text{ connected open triple, } j \text{ central ,} \\
\langle A_i B_j C_k \rangle &\approx \langle A_i B_j \rangle \frac{\langle B_j C_k \rangle}{\langle B_j \rangle} \frac{\langle A_i C_k \rangle}{\langle A_i \rangle \langle C_k \rangle} \qquad G[i,j,k] \text{ connected triangle,}
\end{aligned} \tag{1.28}$$

where $A, B, C \in \{S, I, R\}$ and G is the underlying undirected graph of the network D (recall, from definition 1.7.5, that $G[i,j,k]$ is the subgraph formed from individuals i, j and k and any edges which may exist between them in G). We use the term 'connected open triple' to describe a connected undirected subgraph of three individuals where only one (central) individual has two neighbours, and similarly for 'connected triangle' except here all three individuals have two neighbours. The first closure follows directly from the beyond-pair independence assumption discussed above. The second is similar but attempts to account for the extra correlation due to all three individuals being directly connected to one another, and is an example of the Kirkwood (1935) approximation; see also Rand (1999) for a detailed discussion.

Applying these approximations in equation (1.27), and after making pair level homogeneity assumptions, e.g. $\langle S_i I_j \rangle = [SI]/(Nn) \forall i, j : T_{ij} \neq 0$ (and for simplicity assuming $T = T^T, T_{ij} \in \{0, \beta\}, \gamma_i = \gamma$), yields the population-level closed system (see Sharkey (2008) and references therein):

$$\begin{aligned}
[\dot{S}] &= -\beta[SI], \\
[\dot{I}] &= \beta[SI] - \gamma[I], \\
[\dot{SI}] &= -(\beta+\gamma)[SI] \\
&\quad + \mathcal{C}\beta \frac{[SS][SI]}{[S]} \left((1-\phi) + \phi \frac{N[SI]}{n[S][I]} \right) \\
&\quad - \mathcal{C}\beta \frac{[SI][SI]}{[S]} \left((1-\phi) + \phi \frac{N[II]}{n[I][I]} \right), \\
[\dot{SS}] &= -2\mathcal{C}\beta \frac{[SS][SI]}{[S]} \left((1-\phi) + \phi \frac{N[SI]}{n[S][I]} \right), \\
[\dot{II}] &= -2\gamma[II] \\
&\quad + \mathcal{C}\beta \frac{[SI][SI]}{[S]} \left((1-\phi) + \phi \frac{N[II]}{n[I][I]} \right),
\end{aligned} \tag{1.29}$$

where $\mathcal{C}$ and ϕ are constants determined by the network's adjacency matrix. $\mathcal{C}$ is the number of triples divided by Nn^2, while ϕ is the fraction of triples which are also

triangles (fully connected triples):

$$\begin{aligned} \mathcal{C} &= \frac{||G^2|| - \text{trace}(G^2)}{Nn^2}, \\ \phi &= \frac{\text{trace}(G^3)}{||G^2|| - \text{trace}(G^2)}. \end{aligned} \tag{1.30}$$

Here, G is the network's adjacency matrix and $||A||$, where A is an arbitrary matrix, denotes the sum of all elements in A. Note that $\phi = 0$ when there are no triangles in the network, and this makes the equation for $[II]$ unnecessary for a closed system. The approach outlined above is also applicable to other stochastic models such as the Markovian network-based SIS model (see also the introduction of chapter 3, and Sharkey et al. (2006), for application to directed networks with asymmetric T).

1.10.2 Capturing the dynamics by a message passing approach

The message passing approach of Karrer and Newman (2010) is somewhat similar to Miller's (2011) edge-based approach, but more general, in that they first consider the case of an arbitrary undirected network $G = (V, E)$ and allow the individual-level transmission and recovery processes to be non-Poisson. They then also apply their approach to configuration networks with Markovian individual-level processes and obtain a system which is equivalent to Miller's (2011), with a small discrepancy relating to initial conditions. Analogous to the quantity $\theta(t)$, the message passing approach relies on the probability $H^{i \leftarrow j}(t)$ that i does not receive an infectious contact from j by time t, given that i has been placed in the cavity state. It can be approximated by the quantity $F^{i \leftarrow j}(t)$ (Karrer and Newman, 2010):

$$F^{i \leftarrow j}(t) = 1 - \int_0^t f(\tau)\Big[1 - z \prod_{k \in N_j \setminus i} F^{j \leftarrow k}(t - \tau)\Big] \mathrm{d}\tau, \tag{1.31}$$

where $\int_0^t f(\tau) d\tau$ is the probability that an infected individual makes an infectious contact to a given neighbour within time period t of first becoming infected (it is generally not equal to 1 in the limit as $t \to \infty$). N_j is the set of neighbours of j and z is the probability that an individual is susceptible at $t = 0$ (it is assumed that the states of individuals are statistically independent at $t = 0$ and there are initially no recovered individuals). This means that the expected number of infected individuals at $t = 0$ is $(1 - z)N$.

If the network is a tree network then it is straightforward to show that equation (1.31) is in fact exact, i.e. $F^{i \leftarrow j}(t) = H^{i \leftarrow j}(t)$. Otherwise, it can be shown that $F^{i \leftarrow j}(t) \leq H^{i \leftarrow j}(t)$. The dynamics can then be approximated by the system (Karrer

and Newman, 2010):

$$\langle S_i\rangle^* = z\prod_{j\in N_i} F^{i\leftarrow j}(t), \tag{1.32}$$

$$\langle R_i\rangle^* = \int_0^t r_i(\tau)\Big[1-\langle S_i\rangle^*_{t-\tau}\Big]\mathrm{d}\tau, \tag{1.33}$$

$$\langle I_i\rangle^* = 1-\langle S_i\rangle^*-\langle R_i\rangle^*, \tag{1.34}$$

where $\int_0^t r_i(\tau)d\tau$ is the probability that individual i recovers within time period t of having become infected (assuming it does become infected). The above system is exact, in terms of the underlying stochastic model, when the network is a tree network since then the probability of a cavity state individual receiving no infectious contacts from one neighbour is statistically independent of it receiving no infectious contacts from a different neighbour (similar to Miller's (2011) edge-based approach). Otherwise it can be shown that $\langle S_i\rangle \geq \langle S_i\rangle^*$ (Karrer and Newman, 2010).

1.11 Stochastic simulation

All stochastic simulations carried out in the service of this thesis were programmed using the Gillespie (1976) algorithm, i.e.

1. Choose an initial pure system state (the system is in this state with probability 1 at $t=0$) and set $i=1$.

2. Use a random number generator to determine the next system state Γ_i and the waiting time Δ_i until transition.

3. Increase the simulated time by Δ_i and update the system state to Γ_i.

4. If Γ_i is not an absorbing state and further simulation is required then return to step 2 and set $i=i+1$.

For continuous time Markov chains, it is well known that the waiting time for a given state is exponentially distributed where the parameter is equal to the sum of the rates at which the system transitions away from that state into others (Grimmett and Stirzaker, 1982). The probability that it transitions into a given state is equal to the ratio of the rate at which it transitions into that state to the sum of rates just described.

To simulate a random variable X with cumulative distribution function $F_X(x)$ we draw a number U uniformly from the interval $(0,1)$, using a random number generator, and then apply the generalised inverse $F^*_X(u)$ of $F_X(x)$ to the number U. The

generalised inverse is defined:

$$F_X^*(u) = \inf\{x : F_X(x) \geq u\} \qquad 0 < u < 1.$$

It is straightforward to show that $F_X^*(U)$ then has the same cumulative distribution function as X (U is a random variable distributed uniformly in (0,1)). If $X \sim \text{Exp}(\lambda)$ then

$$F_X^*(u) = -\frac{1}{\lambda}\ln(1-u),$$

and so, since U and $1-U$ have the same distribution, we can simulate X as simply $-\ln(U)/\lambda$. A discrete random variable X, for which $P(X = a_i) = p_i$, can also be simulated as $f(U)$ where $f(u) = a_i$ iff $u \in (\sum_{j=1}^{i-1} p_j, \sum_{j=1}^{i} p_j)$. This is useful for determining the next system state in a Markov chain simulation since, given some present state, the subsequent state can be represented as a state-valued discrete random variable.

Chapter 2

Invasion and endemicity in the Markovian network-based SIS model

2.1 Introduction

We will refer to any compartmental model in which the individuals of the population move back and forth between just two states: susceptible and infectious, as an SIS model. This framework has applications in the domains of sexually transmitted diseases (Hethcote and York, 1984; Garnett and Anderson, 1996; Eames and Keeling, 2002; Keeling and Eames, 2005) and computer viruses (Kephart and White, 1991, 1993; Pastor-Satorras and Vespignani, 2001; Balthrop et al., 2004), where infections are able to persist for long periods of time due to hosts being repeatedly infected. We say that an infection is 'endemic' when its presence in the population is self-sustained for a 'long period'. We say that 'invasion' occurs if the infection becomes endemic. See Gilligan and Bosch (2008) for an overview of invasion and persistence in epidemiological models (for the context of plant pathogens).

For the deterministic SIS model (described in section 1.3.2), endemicity occurs when $R_0 = \beta N/\gamma > 1$ and, for such values of R_0, the endemic state is represented by the equilibrium of the system in which the infection is present (complete absence of the infection is a trivial equilibrium). Therefore, invasion occurs if and only if $R_0 > 1$.

For the Markovian standard SIS model (first introduced by Weiss and Dishon (1971) and described in section 1.5), there is only one stationary distribution and this places probability 1 on the complete absence of the infection (the single absorbing state). However, there is also a unique quasi-stationary distribution (QSD) which is the natural description of endemicity for this stochastic model (see section 1.6). For methods of computation/approximation, see Clancy and Mendy (2011) and references therein. In the limit of large population size there is an exact correspondence between the stochastic and deterministic standard SIS models (Ethier and Kurtz, 1986; Andersson and Britton,

2000) and so, for $R_0 > 1$, the endemic equilibrium of the deterministic model must in some sense approximate the expected number of infected (or susceptible) individuals in the QSD.

Defining invasion for the Markovian standard SIS model poses difficulties since indefinite persistence is not a possibility. However, we may choose to define invasion *probability* as the probability that the corresponding branching process is infinite, as in section 1.4.2 (Ball and Donnelly, 1995; Ball, 1999), and this should roughly correspond to the probability that the time until the disease becomes extinct is long as opposed to short (recall that the validity of the coupling to the branching process increases with population size). Indeed, when running stochastic simulations it is usually straightforward to distinguish between long outbreaks (invasion) and short outbreaks (non-invasion) 'by eye'.

Jacquez and Simon (1993) have proposed that, for the Markovian standard SIS model, invasion can occur if and only if $\beta N/\gamma > N/(N-1)$ since below this threshold the expected number of infected individuals will monotonically decrease, while above the threshold it initially increases (considering a single initial infected). By a different line of thought, Nåsell (1995) proposes that the threshold should be determined by the point in parameter space which gives an expected time to extinction equal to $f(N)$, where f is some non-decreasing function of N (above the threshold, the expected time to extinction grows exponentially with N, while there is little dependence on N below the threshold).

The work of this chapter is motivated by a desire to understand why, for the Markovian standard SIS model, the probability of invasion from a single initial infected (computed from the corresponding branching process) is equal to the fraction of infected individuals in the endemic equilibrium of the deterministic SIS model (when parameters coincide), and to investigate the analogue of this relationship in a generalised version of the model which we call the Markovian network-based SIS model (defined in section 1.10). Along the way, we find a precise mathematical definition for invasion probability which generalises the (large population) definition via branching processes, similarly to the way in which the QSD generalises the (large population) definition of endemicity, i.e. equilibrium, in the deterministic model.

As discussed in section 1.10, the Markovian network-based SIS model can be described by a continuous time Markov chain $\{\sigma(t)\}$, where $\sigma(t)$ takes values in $\{S, I\}^V$, and, for convenience, we reproduce its transition rates in table 2.1 (S-susceptible, I-infected).

The contact network on which the disease spreads is encoded by the matrix T in the sense that $T_{ij} > 0 \Leftrightarrow (j,i) \in A(D)$ and $T_{ij} = 0 \Leftrightarrow (j,i) \notin A(D)$. Recall that T_{ij} is the rate at which j makes infectious contacts to i when j is infected, and γ_i as the rate at which i, once infected, recovers (returning to the susceptible state).

Table 2.1: Transitions for the Markovian network-based SIS model

from	to	at rate
$\sigma : \sigma_i = S$	$\sigma^{i\to I}$	$\sum_{j\in V} T_{ij}\mathbb{1}(\sigma_j = I)$
$\sigma : \sigma_i = I$	$\sigma^{i\to S}$	γ_i

In the context of an arbitrary finite network, the correspondence to a simple branching process is lost, even in the early stages, since each individual 'interacts' differently with its environment. Due to the population being finite, there are no non-trivial stationary distributions since the disease dies out almost surely as $t \to \infty$ (complete absence of the disease is the only absorbing state, reachable from all other states).

The network-based model is becoming more practically relevant as computing power increases, allowing efficient individual-based stochastic simulations to be performed. The increase in the amount of data on real-world network structures (cattle farms: Woolhouse et al., 2005; poultry farms: Sharkey et al., 2007; other examples: Danon et al., 2011) means that the dynamics of diseases, or computer viruses, on *particular* networks is now an area of real concern.

2.2 Numerical investigations

In order to carry out stochastic simulations we adopt the Gillespie (1976) algorithm (described in section 1.11). For a given matrix T and vector of recovery parameters $\Gamma = [\gamma_1, \gamma_2, \ldots \gamma_N]$, a simulation of the Markovian network-based SIS model is produced as follows:

1. Choose an initial system configuration (pure system state), $\sigma(0)$. Set $t = 0$.

2. Draw a random number U uniformly from (0,1). Set $\Delta = -\ln(U)/r$ where $r = \sum_{i=1}^{N}(\lambda_i + \mu_i)$ is the rate at which the system transitions away from its current state, and

$$\begin{aligned} \lambda_i &= \sum_{j=1}^{N} T_{ij}\mathbb{1}(\sigma_j = I)\mathbb{1}(\sigma_i = S), \\ \mu_i &= \gamma_i \mathbb{1}(\sigma_i = I), \end{aligned} \tag{2.1}$$

   are the rate at which individual i flips from susceptible to infected and from infected to susceptible respectively. Finally, update $t = t + \Delta$ (note that $\Delta \sim \text{Exp}(r)$).

3. Draw a random number U from $\{1, 2, \ldots, N\}$ where $i \in \{1, 2, \ldots, N\}$ is chosen with probability $(\lambda_i + \mu_i)/r$. Then, flip the state of individual U (to infected if it is currently susceptible, to susceptible otherwise). Update σ_U accordingly (note

that the probability of individual i's state being flipped is proportional to the rate at which i transitions away from its current state).

4. The system is 'now' in state σ at time t. If the system has not reached the absorbing state then return to step 2.

For the moment, we limit our investigations to contact networks which are strongly connected (see definition 1.7.7). This means that so long as the infection is present somewhere in the network then there is positive probability that all individuals are simultaneously in the infected state at some future time. From this 'all-infected' state any other system state can then be reached by a sequence of individual recoveries. Thus, all non-absorbing states communicate with one another, and the single absorbing 'all-susceptible' state can be reached from any non-absorbing state (all non-absorbing states are therefore 'transient'). The system which emerges when the network is strongly connected thus satisfies the criteria for a unique QSD (see section 1.6, and Daroch and Seneta (1967)).

Since the QSD is by definition the unique distribution towards which the system tends, given non-absorption, we can measure the QSD numerically by first allowing a simulation to run until it 'looks' as though the behaviour could be well described by some time-invariant distribution. Assuming that the remainder of the simulation is described by the QSD, we approximate the marginal probability that individual $i \in V$ is infected in the QSD as the fraction of time it then spends in this state. If absorption occurs (the disease dies out) then we can just restart the process, since what we are measuring is conditioned on non-absorption. Clearly, the longer the period of simulated time over which we take our measurement, the more accurate it will be. Similarly, the longer we 'wait' before starting our measurement, the closer the system will be to the QSD, and the more accurate our measurement will be. We define $P^i_{T,\Gamma}$(quasi-prevalence) to be the marginal probability that individual i is infected in the QSD and, by then summing over all $i \in V$, we define $P^{\text{global}}_{T,\Gamma}$(quasi-prevalence) to be the expected number of infected individuals in the QSD. This is how we define and quantify endemicity in this context, and this is what we attempt to approximate through numerical simulation.

Our measurement of invasion probability is (at this point) not a measurement of a precisely defined mathematical quantity, as is the case for the QSD. However, the intuitive notion of invasion is clear, and as long as we feel we can make a distinction between small and large outbreaks then the procedure is straightforward. The probability of invasion from individual $i \in V$ is computed by running a large number of simulations where the initial system state has i as the only infected individual. For efficiency, we stop the simulations at some maximum time or the time at which absorption occurs, whichever comes first. Then, we decide 'by eye' which simulations exhibited a large outbreak and which exhibited a short outbreak (by examining a histogram), and then compute invasion probability as the fraction of the simulations in which a large

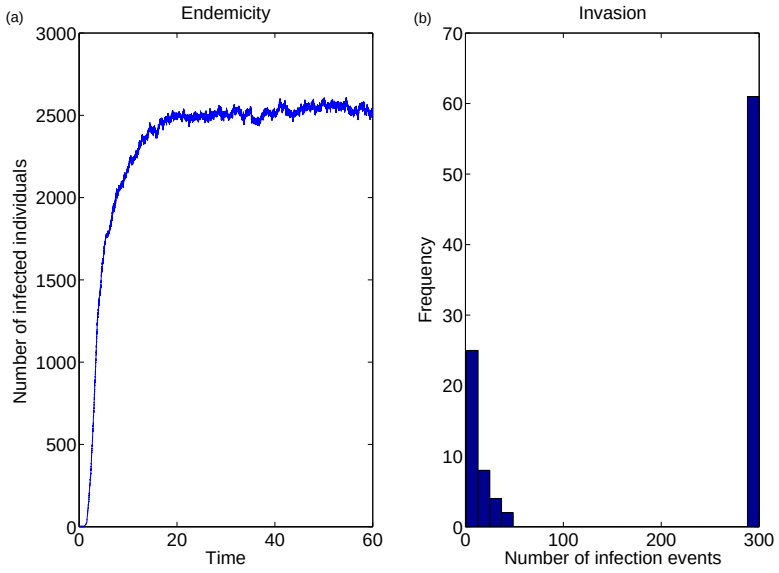

Figure 2.1: Numerical data from simulations of Markovian SIS dynamics on our example network T_{ex}. (a) is a plot of the total number of infected individuals against time in a simulation where the outbreak was initiated on a single infectious individual. (b) is a histogram of the number of infection events in 100 simulations of an outbreak, which were allowed to run up to a maximum of 300 infection events, initiated on the same individual each time. In both cases, the weighted network matrix was multiplied by 0.01 and the recovery rate was set to unity for all individuals.

outbreak occurred. Cases where a clear distinction between small and large outbreaks could be made were not difficult to find across a wide range of networks and parameters. However, we will later derive a mathematical definition of invasion probability which is independent of this numerical approach (and which can be applied to any parametrisation, even when the numerical approach does not allow a clear distinction between small and large outbreaks).

Figure 2.1 gives an example of the numerical measurement of endemicity and invasion probability for our example network T_{ex}. This network is the largest (5,119 node) strongly connected component of a network which was generated from simulations on a complex model of the spread of H5N1 avian influenza through the British poultry flock (Sharkey et al., 2008; Jonkers et al., 2010). It exhibits extensive heterogeneity including complex spatial structures, heterogeneous transmission strengths varying over many orders of magnitude, clustering and directed links. In other numerical investigations we measured and compared $P^i_{T,\Gamma}$(quasi-prevalence) with invasion probability from i in different networks and, for each network, the contact parameter matrix T was also multiplied by different scalar numbers.

In the case of undirected networks, i.e. symmetric T, the numerical results pointed towards an exact relationship between $P^i_{T,\Gamma}$(quasi-prevalence) and invasion probability

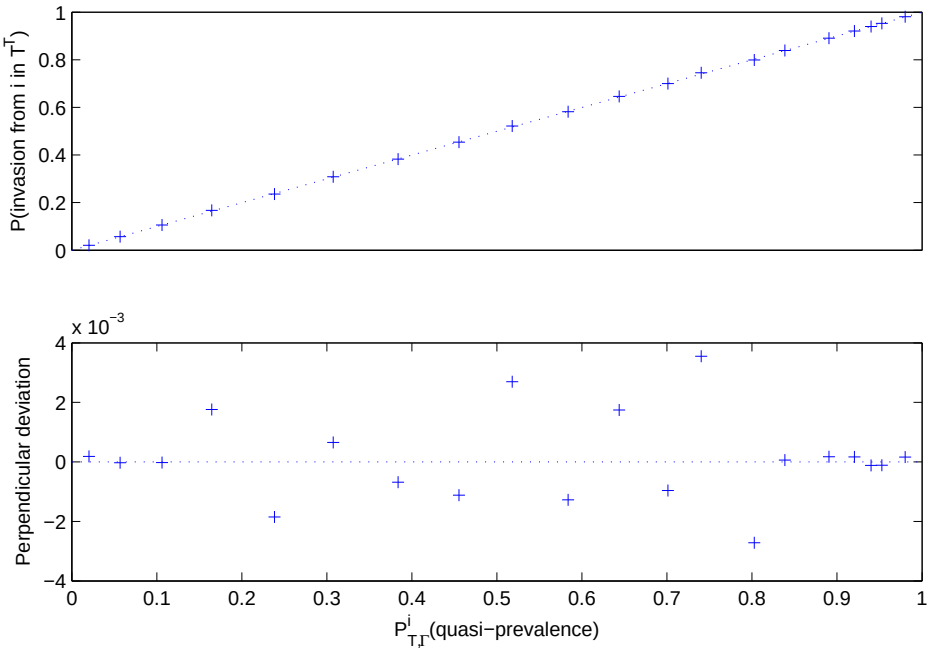

Figure 2.2: Measurements of $P^i_{T,\Gamma}$(quasi-prevalence) for a single individual (i = node 2332) in our example network T_{ex}, and of the probability of invasion from the same individual when network links are reversed, i.e. when T_{ex} is transposed. The recovery rate was set to unity for all individuals while a multiplier of the network matrix was varied. In (a), these two quantities are plotted against each other for each of 20 different multipliers of the network matrix. The faint dashed line indicates equality. On this scale it is not possible to determine any deviation from the equality of the two quantities. (b) is a 'zoomed-in' view of the perpendicular deviation of each of the data points from the straight line (equality), in the bottom right to top left direction.

from i (see figure 2.2). Note that the Markovian standard SIS model is actually a special case of the network-based version, where $T_{ij} = \beta \quad \forall i \neq j$ and $\gamma_i = \gamma \quad \forall i \in V$. Also note that any undirected network is necessarily strongly connected (we assume a single connected component since distinct connected components can be treated as separate systems).

2.3 Graphical representation and duality

For the Markovian network-based SIS model, the graphical representation of Harris (1978) allows a coupling together of all of the processes corresponding to all possible initial configurations on the same probability space. It does this in such a way that there is a large amount of statistical dependence between these different processes and thus several important features emerge. We follow the description of the graphical representation given by Grimmett (2010).

Descriptions of the graphical representation in the literature are commonly re-

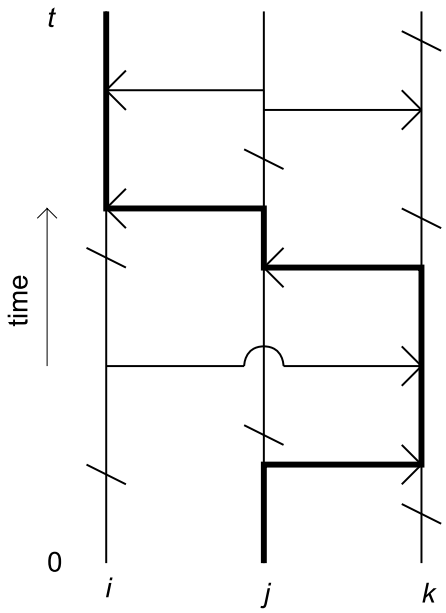

Figure 2.3: A realisation of the graphical representation of Markovian SIS dynamics on a fully connected network of three individuals i, j and k (up to time t). The vertical lines are the time lines corresponding to each individual. The short diagonal lines indicate the points of cure and the horizontal arrows are the arrows of infection. A path from 0 on j's time line to t on i's time line is shown in bold.

stricted to the following special case of the Markovian network-based SIS model: the network is undirected, all contact parameters are the same and all recovery parameters are the same, i.e. $T = T^T$, $T_{ij} \in \{\lambda, 0\}$ and $\gamma_i = \delta$ (Grimmett, 2010). We now consider this special case which is also known as the 'contact process' (it is also frequently restricted to the case of an infinite d-dimensional square lattice, often with $d = 1$ (Griffeath, 1981)). See Liggett (1999) for an overview, and Durrett and Levin (1994) for an ecological perspective.

Firstly, and following Grimmett (2010), we assign a non-negative real number line to each member $i \in V$ of the population and call these time lines. For each $i \in V$ we then place an independent Poisson point process of intensity δ on the time line corresponding to i, and call these points of cure (the 'time' or 'spacing' between two consecutive points of cure, on a single time line, is thus $\sim \text{Exp}(\delta)$). Finally, for each arc $(i, j) \in A(D)$ we place 'arrows of infection' from i's time line to j's time line according to an independent Poisson point process of intensity λ.

Adopting the notation of Harris (1974), we now define the set $\xi_t^A \subset V$, where $A \subset V$, such that $i \in \xi_t^A$ if and only if there is at least one path from 0 on a time line corresponding to an individual in A to t on i's time line (paths follow the direction of time and the directions of the arrows of infection and do not traverse points of cure). Due to its construction, ξ_t^A is precisely the same set-valued random variable (in terms of possible values and distribution) as the set of individuals infected at time t, in the

corresponding epidemiological model, when the set of initial infecteds is A. As such, we know that $\{\xi_t^A\}$ is a Markov chain with a unique QSD which is the same for all choices of $A \subset V (A \neq \emptyset)$.

The following property (monotonicity) is immediate (see, for example, Grimmett (2010)):

$$\xi_t^A \subset \xi_t^B \quad \text{if } A \subset B, \tag{2.2}$$

which implies for the contact process, amongst other things, that the probability of any given individual being infected at time t when A is the set of initial infecteds is less than or equal to the probability of that individual being infected at time t when $B(\supset A)$ is the set of initial infecteds. We say that ξ_t^B is stochastically greater than (or equal to) ξ_t^A, where $A \subset B$, since $P(\xi_t^B \cap C \neq \emptyset) \geq P(\xi_t^A \cap C \neq \emptyset) \quad \forall C \subset V$.

The above graphical representation also lends itself to the proving of the following important property known as 'duality' (Holley and Liggett, 1975; Harris, 1976):

$$P_{\lambda,\delta}(\xi_t^A \cap B \neq \emptyset) = P_{\lambda,\delta}(\xi_t^B \cap A \neq \emptyset), \tag{2.3}$$

which says that the probability of at least one member of B being infected at time t when the set A are initially infected is equal to the probability of at least one member of A being infected at time t when the set B are initially infected (for given transmission and recovery parameters, λ and δ).

Lemma 2.3.1. *(see, for example, Grimmett (2010))*

$$\lim_{t\to\infty} P_{\lambda,\delta}(\xi_t^{\{i\}} \cap V \neq \emptyset) = \lim_{t\to\infty} P_{\lambda,\delta}(\xi_t^V \cap \{i\} \neq \emptyset) \qquad \forall i \in V.$$

This lemma follows simply from the property of duality. The lemma says that the probability of indefinite persistence when only i is initially infected is equal to the probability that i is infected at time t, in the limit as $t \to \infty$, when all individuals are initially infected. For the case where the network is infinite, e.g. an infinite square lattice, it has been shown that these probabilities can be greater than zero (see, for example, Grimmett (2010)) (if the network is finite then they are zero). If the network is also homogeneous, as is the case for the infinite square lattice, then these probabilities are independent of the particular choice of $i \in V$.

The above lemma can be used to intuit the exact correspondence between invasion probability in the Markovian standard SIS model (when computed via the corresponding infinite branching process) and the endemic equilibrium of the deterministic SIS model. The Markovian standard SIS model is equivalent to the contact process on a homogeneous fully connected network (every individual interacts equally with every other individual). As the population size tends to infinity the deterministic model exactly corresponds with the Markovian model/contact process and the early stages of the Markovian model/contact process exactly correspond with a branching process.

Since, in this case, the probabilities in the lemma are the same for all choices of $i \in V$, the right-hand-side is also the 'final' expected fraction of the population in the infected state when all individuals are initially infected. In the limit as $N \to \infty$ this must equal the 'final' fractional size of the infected compartment in the deterministic SIS model, i.e. the endemic equilibrium (or the disease-free equilibrium if below the threshold). The left-hand-side of the lemma is the probability that the infection persists indefinitely when just one individual is initially infected in the Markovian model/contact process which, in the limit as $N \to \infty$, is the probability that the corresponding branching process is infinite.

Much of the analysis of the contact process in the literature focuses on the computation of a threshold value for the parameter λ/δ such that below the threshold eventual extinction of the disease is certain while above the threshold indefinite persistence is possible (for the case where the network is infinite, e.g. an infinite square lattice). For example, in the case of a network which is a $k-$regular Bethe lattice ($k > 2$), this threshold value is known to lie between $1/(2\sqrt{k-1})$ and $1/(k-2)$ (Liggett, 1996). For our analysis of the Markovian network-based SIS model, we will sidestep this question by assuming that endemicity (long but not indefinite persistence) is always possible, and then focus on the probability that endemicity occurs (invasion) and its relationship to the endemic state (QSD).

In order to proceed, we modify the above graphical representation such that it can be applied to the Markovian network-based SIS model in full generality (Wilkinson and Sharkey, 2013): For each $i \in V$ we place an independent Poisson point process of intensity γ_i on the time line corresponding to i. For each arc $(i, j) \in A(D)$ we place 'arrows of infection' from i's time line to j's time line according to an independent Poisson point process of intensity T_{ji}. We then define ξ_t^A as before and note that if the network is strongly connected then $\{\xi_t^A\}$ has a unique QSD that is the same for all choices of $A \subset V (A \neq \emptyset)$. The property of monotonicity still holds, while the property of duality needs some modification and is expressed as:

Theorem 2.3.1 (Duality for the Markovian network-based SIS model)**.**

$$P_{T,\Gamma}(\xi_t^A \cap B \neq \emptyset) = P_{T^T,\Gamma}(\xi_t^B \cap A \neq \emptyset),$$

where matrix T and vector Γ fully parametrise the model, and T^T is the transpose of T (directions of transmission processes are reversed). Here, T may be asymmetric and its elements may be any non-negative real numbers. Γ also consists of non-negative real numbers. The theorem says that the probability of at least one member of B being infected at time t when only the set A are initially infected is equal to, in the transposed network, the probability of at least one member of A being infected at time t when only the set B are initially infected (note that for undirected networks $T = T^T$). See Wilkinson and Sharkey (2013).

Proof. Consider first the case where the network is undirected (symmetric T) and where the subsets of the population are single individuals: $A = j, B = i$. The probability of a path (under the graphical representation) from 0 on j's time line to t on i's time line is then expressed by the left-hand-side. This is equal to the probability that i is infected at time t when only j is initially infected. With reference to figure 2.3, reversing the direction of time (turning the diagram upside down) does not alter the validity or interpretation of the graphical representation, in any way, since the probability of a point of cure or arrow of infection in any time interval just depends on the absolute size of the time interval (by the memoryless property of Poisson processes). Therefore, since by reversing the direction of time a path from 0 on j's time line to t on i's time line maps to a path from 0 on i's time line to t on j's time line, the theorem holds for the case where T is symmetric and A and B are single individuals. When T is asymmetric, reversing the direction of time produces a valid graphical representation for the transposed network and this explains the appearance of T^T in the right-hand-side. The case of arbitrary sets A and B follows through by exactly the same logic (see, for example, Grimmett (2010) for the case of symmetric T and homogeneous transmission and recovery rates). □

2.4 Endemicity and quasi-prevalence

We define $P^A_{T,\Gamma}$(quasi-prevalence) to be the marginal probability that at least one member of $A \subset V$ is infected in the QSD and so, by the definition of the QSD, we can write:

$$P^A_{T,\Gamma}(\text{quasi-prevalence}) = \lim_{t\to\infty} P_{T,\Gamma}(\xi^V_t \cap A \neq \emptyset \mid \xi^V_t \cap V \neq \emptyset) \tag{2.4}$$

$$= \lim_{t\to\infty} \frac{P_{T,\Gamma}(\xi^V_t \cap A \neq \emptyset)}{P_{T,\Gamma}(\xi^V_t \cap V \neq \emptyset)}, \tag{2.5}$$

where the second equality follows from the fact that $\xi^V_t \cap A \neq \emptyset$ implies $\xi^V_t \cap V \neq \emptyset$ (recall that we are assured a unique QSD by assuming the network to be strongly connected). Note that the choice of initial conditions in equation (2.4) is arbitrary since the QSD is independent of initial conditions (except that the initial state cannot be the absorbing state).

It can be argued that the QSD has practical relevance if the rate of convergence to this distribution, when conditioning on non-absorption, is rapid compared to the rate at which the probability of absorption by time t approaches 1 when the system is 'initiated' in the QSD (Daroch and Seneta, 1965). This corresponds to the case where, given that the infection survives the initial stages, the expected time to extinction is long and for most of its lifetime the system behaves as if it is well described by the QSD. This can often be the case for stochastic SIS dynamics where, according to Nåsell (1996) (on the Markovian standard SIS model), 'it is easy to find examples where the

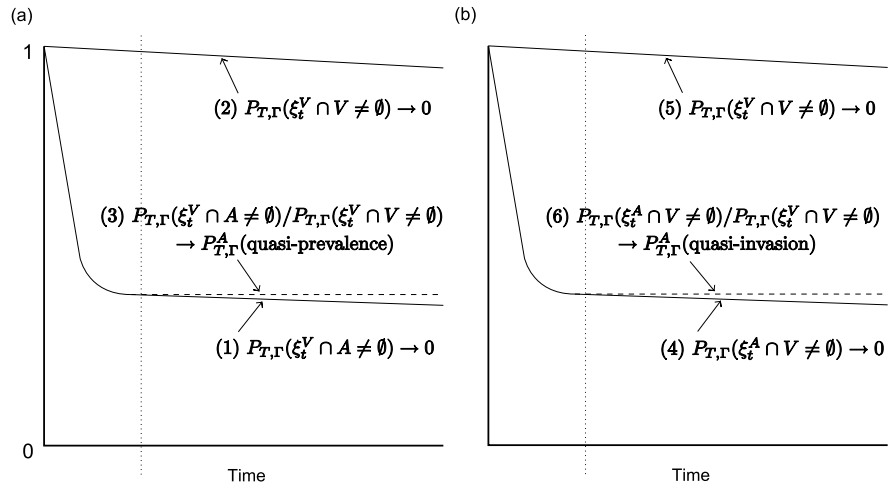

Figure 2.4: Here we illustrate how it is possible for the quantifiers $P^A_{T,\Gamma}$(quasi-prevalence) and $P^A_{T,\Gamma}$(quasi-invasion) to capture critical information about the model. If the network is undirected then these quantifiers are numerically the same and have equal practical relevance (as is seen by assuming that T is the same undirected network in (a) and (b), above).

expected time to extinction even for a rather small population exceeds the age of the universe'.

Let us consider the following quantities for the Markovian network-based SIS model, and a given matrix T and vector Γ:

1. $P_{T,\Gamma}(\xi_t^V \cap A \neq \emptyset)$ = The probability that at least one member of $A \subset V$ is infected at time t given that all individuals are infected at $t = 0$.

2. $P_{T,\Gamma}(\xi_t^V \cap V \neq \emptyset)$ = The probability that the infection survives to time t given that all individuals are infected at $t = 0$.

3. $P_{T,\Gamma}(\xi_t^V \cap A \neq \emptyset)/P_{T,\Gamma}(\xi_t^V \cap V \neq \emptyset)$ = The probability that at least one member of A is infected at time t given that all individuals are infected at $t = 0$ and given that the infection survives to time t.

Note that in the limit as $t \to \infty$ quantity 3 is equal to $P^A_{T,\Gamma}$(quasi-prevalence). In figure 2.4a, the way in which these three quantities vary with respect to time is illustrated for the scenario in which the QSD has practical relevance (discussed above). In this scenario, the quantifier $P^A_{T,\Gamma}$(quasi-prevalence) is able to capture the value at which $P_{T,\Gamma}(\xi_t^V \cap A \neq \emptyset)$ initially 'plateaus' before its slow decay to zero. Since the all-infected initial state gives the maximum expected time to absorption (via monotonicity) then it is natural to associate this plateau with endemicity.

2.5 Quasi-invasion

In the context of SIS dynamics on arbitrary finite networks, invasion and invasion probability have not been given rigorous theoretical definitions, even though attempts to establish an invasion-threshold in the Markovian standard SIS model have been made (Nåsell, 1995; Jacquez and Simon, 1993). As we have already discussed, invasion probability is often computed as the probability that the corresponding branching process is infinite, but even for fully connected networks this is only valid in the limit of large population size. For finite heterogeneous networks, finding an appropriate branching process which can reasonably approximate the early stages, while still being tractable, is not so easy (and the choice of branching process depends on which individual in the network initiates the epidemic). The finite situation demands the ability to distinguish between short outbreaks and long, but not indefinite, outbreaks.

In this section we show that the quantifier of invasion probability, which has equal practical relevance to the QSD and quasi-prevalence, for outbreaks initiated by the members of $A \subset V$ in a strongly connected network can be defined:

$$P^A_{T,\Gamma}(\text{quasi-invasion}) = \lim_{t\to\infty} \frac{P_{T,\Gamma}(\xi^A_t \cap V \neq \emptyset)}{P_{T,\Gamma}(\xi^V_t \cap V \neq \emptyset)}, \tag{2.6}$$

where the numerator is the probability of survival to time t when subset A are the initial infecteds, while the denominator is the probability of survival to time t when all individuals are initially infected. Immediately, there are some observations about this definition which correspond to our intuitive notions of invasion. The definition implies that invasion is certain when all individuals are initially infected, since the numerator and the denominator are then equal for all t. Also, we have that

$$P^A_{T,\Gamma}(\text{quasi-invasion}) \geq P^B_{T,\Gamma}(\text{quasi-invasion}) \qquad B \subset A,$$

by the property of monotonicity. In other words, when 'adding' more initial infecteds the probability of invasion is non-decreasing.

Let us now consider the quantities:

4. $P_{T,\Gamma}(\xi^A_t \cap V \neq \emptyset)$ = The probability that the infection survives to time t given that only the members of A are infected at $t = 0$.

5. $P_{T,\Gamma}(\xi^V_t \cap V \neq \emptyset)$ = The probability that the infection survives to time t given that all individuals are infected at $t = 0$.

6. $P_{T,\Gamma}(\xi^A_t \cap V \neq \emptyset)/P_{T,\Gamma}(\xi^V_t \cap V \neq \emptyset)$ = The quotient of quantities 4 and 5.

It follows from duality that the three quantities, 4, 5 and 6, are all equal respectively to the three quantities, 1, 2 and 3 (from section 2.4), provided that we transpose T. Note also that, in the limit as $t \to \infty$, quantity 6 is equal to $P^A_{T,\Gamma}(\text{quasi-invasion})$.

Quantity 4 denotes the survival probability up to time. We see in figure 2.4b that this plateaus in exactly the same way as quantity 1 for the transposed network (figure 2.4a). Thus, in exactly the same way in which $P^A_{T,\Gamma}$(quasi-prevalence) captures the value at which the probability of infection in A plateaus before its slow decay to zero, when all individuals are initially infected, $P^A_{T,\Gamma}$(quasi-invasion) must capture the value at which the probability of survival anywhere, when subset A are the initial infecteds, plateaus before its slow decay to zero. This is, after all, what we are interested in since the slow decay to zero represents the inevitability of absorption, regardless of initial conditions, while the initial plateau represents the effect of the much faster decay due to specific initial conditions. In some sense, we can say that quasi-invasion corresponds to the 'achievement' of the QSD. That the limit in the above definition exists, and is always positive (from the Perron-Frobenius theorem - see section 1.6), follows from the fact that due to the property of duality $P^A_{T,\Gamma}$(quasi-invasion) is equal to the probability that at least one member of A is infected in the QSD when the transmission processes are reversed, i.e. $T \to T^T$ (transposing T does not affect the network's being strongly connected and so a unique QSD is still ensured). We discuss this further in the next section.

Our quantifier of invasion probability can be generalised as:

$$P^X(\text{quasi-invasion}) = \lim_{t\to\infty} \frac{P^X_S(t)}{P^{\max}_S(t)}, \tag{2.7}$$

where X is the initial system state from which invasion may or may not occur. $P^X_S(t)$ is the probability that the infection survives to time t given that the system is initiated in state X, and $P^{\max}_S$ is the probability of survival to time t given that the initial state is such that the expected time to extinction is maximised (that this maximising state is the all-infected state, for Markovian network-based SIS dynamics, follows from monotonicity). In this form, the definition becomes applicable to other Markovian infection dynamics (on strongly connected networks) which permit endemic behaviour via a unique QSD, e.g. susceptible-infected-removed-susceptible (SIRS) dynamics (see figure 2.5). It is the existence of a unique QSD, to which the system tends regardless of initial conditions (given non-absorption), which enables our definition to capture invasion probability in the same way as for SIS dynamics. Note that the definition of quasi-prevalence can also be generalised to any infection dynamics with a unique QSD.

2.6 The prevalence-invasion relationship

Our main result, in this area, states the prevalence-invasion relationship and is presented as a theorem:

Theorem 2.6.1.

$$P^A_{T,\Gamma}(\textit{quasi-invasion}) = P^A_{T^T,\Gamma}(\textit{quasi-prevalence}),$$

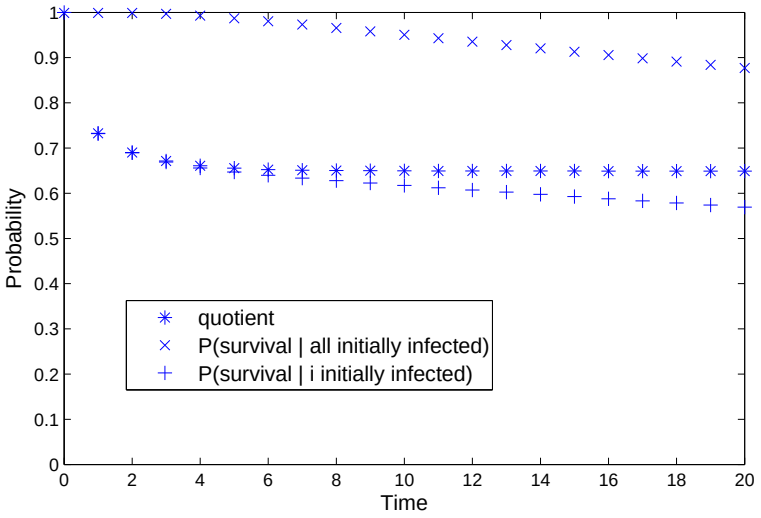

Figure 2.5: Here, we approximate $P(\text{quasi-invasion})$ for a single initial infected (i), in the case of Markovian SIRS dynamics on a small network of 25 individuals. The definition (see text) is seen to capture the 'plateau' of the survival probability.

for Markovian network-based SIS dynamics on strongly connected networks, and where A is any subset of the population.

The above theorem can be rewritten as:

$$\lim_{t\to\infty} \frac{P_{T,\Gamma}(\xi_t^A \cap V \neq \emptyset)}{P_{T,\Gamma}(\xi_t^V \cap V \neq \emptyset)} = \lim_{t\to\infty} \frac{P_{T^T,\Gamma}(\xi_t^V \cap A \neq \emptyset)}{P_{T,\Gamma}(\xi_t^V \cap V \neq \emptyset)}, \tag{2.8}$$

which holds because of the property of duality.

Note that for a single individual $i \in V$ we have:

$$P^i_{T,\Gamma}(\text{quasi-invasion}) = P^i_{T^T,\Gamma}(\text{quasi-prevalence}), \tag{2.9}$$

that is, the probability of invasion from individual i is equal to the probability that i is infected in the QSD (for the transposed network). By summing over all $i \in V$ and dividing by N we get

$$P^{\text{global}}_{T,\Gamma}(\text{quasi-invasion}) = P^{\text{global}}_{T^T,\Gamma}(\text{quasi-prevalence}), \tag{2.10}$$

where $P^{\text{global}}_{T,\Gamma}(\text{quasi-invasion})$ is the probability of invasion when there is a single initial infected chosen uniformly at random from the population. An implication of the global-level relationship is that, for strongly connected directed networks, reversing the transmission processes will result in an interchange between these two quantifiers without affecting the 'stability' of the QSD, i.e. the expected time to extinction when the system is initiated in its QSD is the same. This can be understood by observing that $P_{T,\Gamma}(\xi_t^V \cap V \neq \emptyset) = P_{T^T,\Gamma}(\xi_t^V \cap V \neq \emptyset)\ \forall t$. Another interesting observation is that

$P_{T,\Gamma}(\xi_t^i \cap i \neq \emptyset) = P_{T^T,\Gamma}(\xi_t^i \cap i \neq \emptyset)$, i.e. given the infection is initiated by individual i, the probability that i is infected at time t is the same for T and its transpose.

For the case where the network is undirected (symmetric T), theorem 2.6.1 implies that:

$$P^A_{T,\Gamma}(\text{quasi-invasion}) = P^A_{T,\Gamma}(\text{quasi-prevalence}), \tag{2.11}$$

and for a single individual $i \in V$:

$$P^i_{T,\Gamma}(\text{quasi-invasion}) = P^i_{T,\Gamma}(\text{quasi-prevalence}), \tag{2.12}$$

and globally:

$$P^{\text{global}}_{T,\Gamma}(\text{quasi-invasion}) = P^{\text{global}}_{T,\Gamma}(\text{quasi-prevalence}). \tag{2.13}$$

2.7 Simulations on a small square lattice

An undirected and homogeneously weighted ($T_{ij} \in \{0, \beta\}$) square lattice of 25 individuals was investigated (see figure 2.6). Due to the small population size, the probability of extinction on a relatively short timescale was significant, even when starting from all-infected. This network enables us to illustrate the numerical measurement of our quantifiers in a scenario where the QSD has less practical relevance, i.e. where endemic quasi-stationary behaviour and dichotomised persistence are not recognisable phenomena. In this case, we can compute $P^i_{T,\Gamma}(\text{quasi-invasion})(= P^i_{T,\Gamma}(\text{quasi-prevalence}))$ by directly measuring $P_{T,\Gamma}(\xi_t^i \cap V \neq \emptyset)/P_{T,\Gamma}(\xi_t^V \cap V \neq \emptyset)$ at increasing time points and then estimating its convergent value. Thus, for two different global transmission parameters (0.8, 0.5), and two different initial states (all-infected, one infected), 1 million simulations were allowed to run up to some specific point in simulated time (the global recovery parameter was always set to 1). For each simulation, the time at which extinction occurred was recorded so that the probability of survival up to increasing points in time could be measured.

In figure 2.6a, our quantifier is able to capture an important feature of the model, i.e. the value at which $P_{T,\Gamma}(\xi_t^i \cap V \neq \emptyset)(= P_{T,\Gamma}(\xi_t^V \cap i \neq \emptyset))$ plateaus before its inevitable decay to zero. Figure 2.6b gives an example of a scenario where, although our quasi-invasion and quasi-prevalence quantifiers are clearly defined, their practical relevance is less obvious. This is because the transmission parameter was sufficiently low such that early extinction was the dominant behaviour.

2.8 Computational efficiency in the measurement of invasion probability and endemic prevalence - a new perspective

Through duality, we can approximate $P^A_{T,\Gamma}(\text{quasi-prevalence})$ by measuring $P_{T^T,\Gamma}(\xi_t^A \cap V \neq \emptyset)$ at increasing points in time (as in figure 2.6) in order to estimate the value

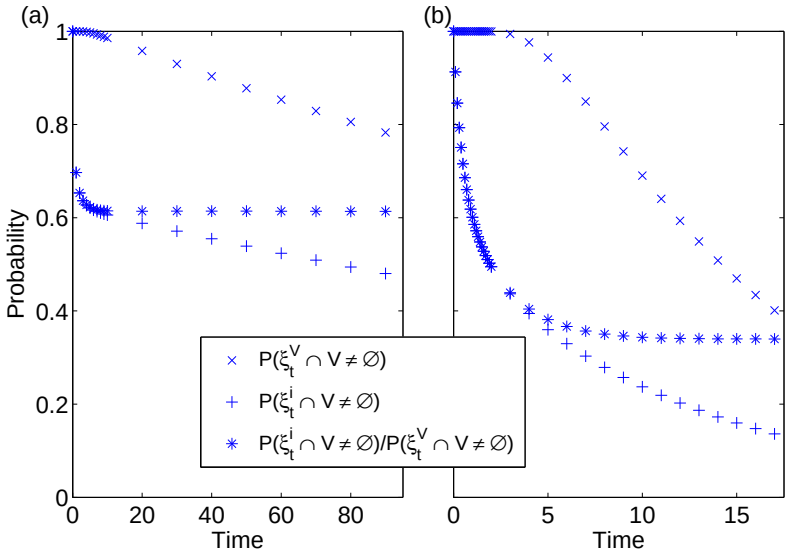

Figure 2.6: Here we illustrate a method of measurement, through stochastic simulation, for $P^i_{T,\Gamma}(\text{quasi-invasion})(= P^i_{T,\Gamma}(\text{quasi-prevalence}))$, where T is an undirected and homogeneously weighted square lattice of 25 individuals (we look for the value towards which $P_{T,\Gamma}(\xi^i_t \cap V \neq \emptyset)/P_{T,\Gamma}(\xi^V_t \cap V \neq \emptyset)$ converges). For (a), the global transmission parameter (β) was set to 0.8. For (b), the global transmission parameter was 0.5. The global recovery parameter (γ) was set to 1 in both cases. The figure illustrates how these quantifiers are well defined but not always practically relevant.

at which it may initially plateau. This could, in certain circumstances, be much more efficient than trying to establish global quasi-stationary behaviour and then computing the proportion of time for which the infection is present in A. Conversely, if we wish to approximate $P^i_{T,\Gamma}(\text{quasi-invasion})$, for all $i \in V$, it may be more computationally efficient to first establish global quasi-stationary behaviour in the transposed network and then measure the proportion of time each individual spends infected.

2.9 Weakly connected networks

In this section we consider the Markovian network-based SIS model in the case where the contact network is weakly connected (we still assume that the underlying undirected network is connected). This means that the sets of individuals reachable from, and that can be reached from, some individual $i \in V$ may be differing subsets of the population. Therefore, there can be no unique QSD since some areas of the network may remain completely disease-free depending on where the infection originates. Indeed, the criteria for a unique QSD that is independent of initial conditions is here absent. However, for a given initial distribution over the state space there is a unique distribution towards which the system, conditioned on non-absorption, converges (Daroch and Seneta, 1965).

As a first step towards understanding, observe that if the set of individuals reachable

from individual $i \in V$ is disjoint from the set of individuals from which i is reachable, then i's ability to spread the infection has no impact on the probability of it being infected at any future time. However, if there is some set of individuals (or just one) which i can reach, and from which i is reachable, then i's ability to spread the infection impacts on the probability of it being infected at future times (it indirectly causes individuals to become infected which then indirectly cause it to be infected). Therefore, we expect such individuals to be important in sustaining the disease and promoting endemicity. These individuals are in fact the members of the network's strongly connected *components* (SCCs). We can think of the SCCs as regions in which the presence of the infection can be self-sustained, or as the equivalence classes of the equivalence relation 'i can be reached from j and j can be reached from i'. As such, they are crucial to endemicity.

The importance of SCCs to endemicity leads us to define distinct QSDs for each. For example, let $D = (V, A)$ be the network and $D[B]$, where $B \subset V$, be an SCC of D. Now let S_B be the set of all system configurations where at least one member of B is infected and no members of $V \setminus \mathrm{Out}(B)$ are infected. If the system is in a configuration belonging to S_B then the probability that all individuals in $\mathrm{Out}(B)$ are simultaneously infected at some future time is positive. From this configuration, any configuration of S_B can then be achieved by a sequence of recoveries, and so all configurations in S_B communicate with one another. No configuration in which at least one member of $V \setminus \mathrm{Out}(B)$ is infected can be reached from a configuration in S_B, and no configuration of S_B can be reached from a configuration where there is no infection in $B \cup (V \setminus \mathrm{Out}(B))$. This means that S_B is a communicating class of states. Therefore, by treating all system configurations not in S_B as absorbing states, and then conditioning on non-absorption, we generate a unique QSD for S_B. In other words, $\{\xi_t^A \mid \xi_t^A \cap B \neq \emptyset\}$ converges to a unique distribution that is the same for all choices of $A \subset \mathrm{Out}(B) : A \cap B \neq \emptyset$. Similarly, we can define unique QSDs for sets of SCCs, e.g. by defining S_{B_1,B_2}, where $D[B_1]$ and $D[B_2]$ are SCCs, as the set of system configurations where at least one member of B_1 and at least one member of B_2 is infected, and no members of $V \setminus (\mathrm{Out}(B_1) \cup \mathrm{Out}(B_2))$ are infected (this generalises to larger sets of SCCs). However, when defining a QSD for a set of SCCs we will assume that none of the SCCs can be reached from any of the others, for example, if B_1 can be reached from B_2 then conditioning on the infection always being in B_2 will have a similar effect to conditioning on the infection always being in B_1 and B_2, i.e. the QSDs corresponding to S_{B_2} and S_{B_1,B_2} will be in some sense similar.

If the infection is suspected to originate in B, or from individual(s) not contained in any SCCs but from which B can be reached, then the QSD corresponding to S_B is the obvious choice for representing the endemic situation. If the infection is suspected to originate in B_1 and B_2, or from individual(s) not contained in any SCCs but from

which B_1 and B_2 can be reached, then the QSD corresponding to S_{B_1,B_2} is the obvious choice for representing the endemic situation.

Note that the QSD corresponding to S_B and restricted to the states of individuals in B is precisely the unique QSD we get by first removing all individuals outside of B and then conditioning on non extinction. Similarly, the QSD corresponding to $S_{B_1,B_2,\ldots,B_n}$ and restricted to the states of individuals in B_i is precisely the unique QSD obtained by first removing all individuals outside of B_i and then conditioning on non extinction (because we have assumed that one SCC cannot be reached from another). Since, for any $D[B]$ which is an SCC, we have $P_{T,\Gamma}(\xi_t^B \cap B \neq \emptyset) = P_{T^T,\Gamma}(\xi_t^B \cap B \neq \emptyset)$ it is straightforward that the practical relevance of the QSD corresponding to S_B in the original network is the same as the practical relevance of the QSD corresponding to S_B in the transposed network. Similarly, if the QSDs corresponding to S_{B_1}, $S_{B_2}\ldots$ are practically relevant then so is the QSD corresponding to $S_{B_1,B_2,\ldots}$ in both the original network and the transposed network.

We are now in a position to generalise our definition of quasi-prevalence such that it is also applicable to weakly connected networks:

$$\begin{aligned} & P_{T,\Gamma}^{A,\{B_1,B_2,\ldots,B_n\}}(\text{quasi-prevalence}) \\ = \quad & \lim_{t\to\infty} P_{T,\Gamma}(\xi_t^C \cap A \neq \emptyset \mid \xi_t^C \cap B_i \neq \emptyset \quad \forall i \in \{1,2,\ldots,n\}) \\ = \quad & \lim_{t\to\infty} P_{T,\Gamma}(\xi_t^C \cap A \neq \emptyset \mid \xi_t^{B_i} \cap B_i \neq \emptyset \quad \forall i \in \{1,2,\ldots,n\}), \end{aligned} \tag{2.14}$$

where $D[B_i]$ is an SCC for all $i \in \{1,2,\ldots,n\}$ and $C = \cup_{i=1}^n B_i$, and $A \subset \text{Out}(C)$. Thus, $P_{T,\Gamma}^{A,\{B_1,B_2,\ldots,B_n\}}(\text{quasi-prevalence})$ is the probability that at least one member of A is infected in the unique QSD defined on the set of system configurations $S_{B_1,B_2,\ldots,B_n}$, i.e. given that the infection is endemic in B_1 and B_2 …and B_n, and the infection is not present in $V \setminus \text{Out}(C)$. The definition captures the value at which $P_{T,\Gamma}(\xi_t^C \cap A \neq \emptyset)$ initially plateaus before its slow decay to zero in exactly the same way in which the definition for strongly connected networks captures the initial plateau of $P_{T,\Gamma}(\xi_t^V \cap A \neq \emptyset)$ (assuming the QSD has practical relevance). In many specific examples, there will be just one SCC or just one that dominates the network. For the case of a single SCC the definition simplifies to

$$P_{T,\Gamma}^{A,\{B\}}(\text{quasi-prevalence}) \quad = \quad \lim_{t\to\infty} P_{T,\Gamma}(\xi_t^B \cap A \neq \emptyset \mid \xi_t^B \cap B \neq \emptyset),$$

where $D[B]$ is the SCC and $A \subset \text{Out}(B)$. Note that if the whole network is strongly connected then $B = V$ and we obtain our previous definition of quasi-prevalence for strongly connected networks as a special case.

We define quasi-invasion in this context:

$$\begin{aligned} & P_{T,\Gamma}^{A,\{B_1,B_2,\ldots,B_n\}}(\text{quasi-invasion}) \\ = \quad & \lim_{t\to\infty} P_{T,\Gamma}(\xi_t^A \cap C \neq \emptyset \mid \xi_t^{B_i} \cap B_i \neq \emptyset \quad \forall i \in \{1,2,\ldots,n\}), \end{aligned} \tag{2.15}$$

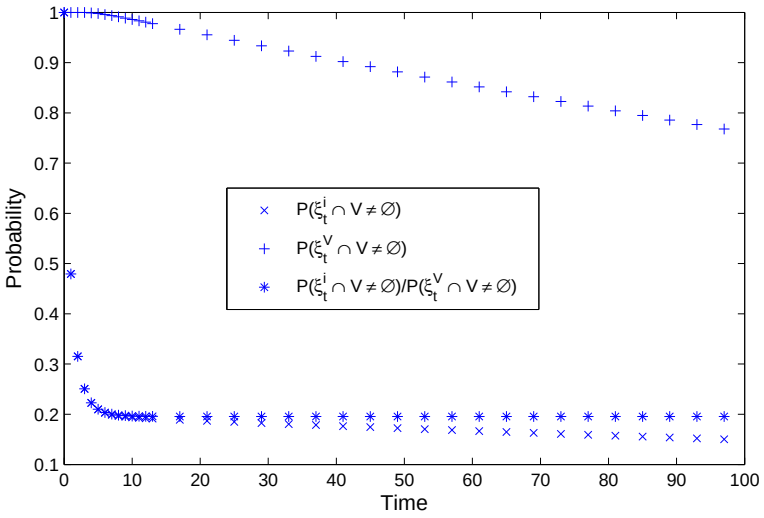

Figure 2.7: Here, we approximate P(quasi-invasion) for a single initial infected (i) outside of the strongly connected component. The network consisted of 26 individuals with the strongly connected component comprised of 25 individuals.

where here $A \subset \text{In}(C)$. Thus, if $B_i \subset A$ for some i then invasion from A is certain since the conditional probability in the definition is then equal to 1 for all t, and by monotonicity

$$P_{T,\Gamma}^{A,\{B_1,B_2,\ldots,B_n\}}(\text{quasi-invasion}) \geq P_{T,\Gamma}^{A',\{B_1,B_2,\ldots,B_n\}}(\text{quasi-invasion}) \qquad A' \subset A,$$

so 'adding' initial infecteds cannot decrease the probability of invasion.

In the case of a single SCC, $D[B]$, quasi-invasion simplifies to:

$$P_{T,\Gamma}^{A,\{B\}}(\text{quasi-invasion}) \quad = \quad \lim_{t\to\infty} P_{T,\Gamma}(\xi_t^A \cap B \neq \emptyset \mid \xi_t^B \cap B \neq \emptyset),$$

where, again, setting $B = V$ gives the previous definition of quasi-invasion for strongly connected networks.

We will next seek to justify $P_{T,\Gamma}^{A,\{B_1,B_2,\ldots,B_n\}}$(quasi-invasion) as the probability that at least one of the components $B_1 \ldots B_n$ is invaded when A is the set of initial infecteds. Firstly, we state the prevalence-invasion relationship for this more general context

Theorem 2.9.1.

$$P_{T,\Gamma}^{A,\{B_1,B_2,\ldots,B_n\}}(\textit{quasi-prevalence}) = P_{T^T,\Gamma}^{A,\{B_1,B_2,\ldots,B_n\}}(\textit{quasi-invasion}).$$

This theorem can be rewritten as

$$\begin{aligned} &\lim_{t\to\infty} P_{T,\Gamma}(\xi_t^C \cap A \neq \emptyset \mid \xi_t^{B_i} \cap B_i \neq \emptyset \quad \forall i \in \{1,2,\ldots,n\}) \\ = \quad &\lim_{t\to\infty} P_{T^T,\Gamma}(\xi_t^A \cap C \neq \emptyset \mid \xi_t^{B_i} \cap B_i \neq \emptyset \quad \forall i \in \{1,2,\ldots,n\}), \end{aligned} \tag{2.16}$$

which follows from the property of duality: reversing the direction of time produces a graphical representation for the transposed network, a path from 0 on i's time line to t on j's time line, where $i, j \in B_i$, is mapped to a path of the same type, and a path from 0 on i's time line to t on j's time line, where $i \in C, j \in A$, is mapped to a path from 0 on j's time line to t on i's.

Assuming the QSD corresponding to $S_{B_1,B_2,\ldots}$ is practically relevant (it is equally practically relevant for the original network and the transposed network) then $P_{T^T,\Gamma}(\xi_t^C \cap A \neq \emptyset) = P_{T,\Gamma}(\xi_t^A \cap C \neq \emptyset)$ plateaus before its slow decay to zero, and this plateau is captured by $P_{T^T,\Gamma}^{A,\{B_1,B_2,\ldots,B_n\}}(\text{quasi-prevalence}) = P_{T,\Gamma}^{A,\{B_1,B_2,\ldots,B_n\}}(\text{quasi-invasion})$. It is natural to interpret the value at which $P_{T,\Gamma}(\xi_t^A \cap C \neq \emptyset)$ plateaus as the probability that the infection establishes itself somewhere in $C = \cup_{i=1}^n B_i$ when the members of A are the initial infecteds, i.e. at least one of the SCCs is invaded.

2.10 Discussion

By considering the unique QSD associated with Markovian SIS dynamics on finite strongly connected networks, along with its implications under duality, we have provided meaningful mathematical definitions for both endemic prevalence (quasi-prevalence) and invasion probability (quasi-invasion). Utilising these definitions, we have provided a general statement of the exact relationship between invasion probability and endemic prevalence at the individual and population level, for any finite undirected network of arbitrary heterogeneity (including undirected networks with weighted links and individual-specific recovery parameters). The definitions also generalise to weakly connected networks, and the prevalence-invasion relationship, with slight modification, applies to arbitrary directed networks.

We note that for infinite homogeneous networks, invasion probability (in these cases, the probability of indefinite persistence) from a single initial infected has been shown to be equal to the fraction of the population infected in the upper invariant measure (Grimmett, 2010; Neal, 2008). Furthermore, the relationship between the probability of long-term persistence and quasi-stationary distributions has previously been investigated (see Chaterjee and Durrett (2009) and, for the related concept of 'metastability', see Schonmann (1985) and Simonis (1996)). However, although the prevalence-invasion relationship follows easily from a combination of the QSD and duality, to our knowledge this is the first general statement of this exact relationship in the context of arbitrary networks. We have thus related two fundamental epidemiological quantifiers in systems where they cannot usually be calculated analytically due to complexity.

It is generally easier to collect empirical data on endemic prevalence rather than directly on invasion risk. In the case of undirected networks, prevalence data can thus be utilised to inform invasion risk. This method echoes Anderson and May's (1991) estimation of the basic reproductive ratio of measles from the total number

of susceptible individuals in England and Wales (using data from Fine and Clarkson (1982)).

When other infectious agents exhibit qualitatively similar behaviour on the same undirected network, we can expect that the individuals carrying the greatest level of endemic infection are also those most likely to initiate new successful invasions. This lends support to the targeting of high-risk individuals in these systems as an effective strategy for the mitigation and control of emerging epidemics.

Chapter 3

Moment-closure for Markovian epidemic dynamics on networks

3.1 Introduction

In this chapter we focus on epidemic dynamics where individuals can only be infected once or not at all. Specifically, we consider a Markovian network-based model with a general compartmental structure described as susceptible-exposed-infected-recovered (SEIR) (see, for example, Keeling and Rohani (2007)). This model is the same as the Markovian network-based SIR model, defined in section 1.10, except that after a susceptible individual $i \in V$ receives an infectious contact it must first pass through an 'exposed' state, lasting for a period that is exponentially distributed with parameter α_i, before entering the infected state. While in the exposed state individuals do not make infectious contacts. Thus, with reference to the definition of the simpler SIR version in section 1.10, the Markovian network-based SEIR model can be described by a continuous time Markov chain $\{\sigma(t)\}$, where $\sigma(t)$ takes values in $\{S, E, I, R\}^V$, and with transition rates as in table 3.1.

Table 3.1: Transitions for the Markovian network-based SEIR model

from	to	at rate
$\sigma : \sigma_i = S$	$\sigma^{i\to E}$	$\sum_{j\in V} T_{ij}\mathbb{1}(\sigma_j = I)$
$\sigma : \sigma_i = E$	$\sigma^{i\to I}$	α_i
$\sigma : \sigma_i = I$	$\sigma^{i\to R}$	γ_i

For this model there are 4^N Kolmogorov forward equations which give a full description of the evolution of the system (given some initial distribution). However, moment closure methods allow us to write down much smaller systems of ordinary differential equations which attempt to capture the evolution of the expected number in each compartment. For example, the time derivative of the expected number of pairings of a susceptible and an infected individual depends on the expected numbers of connected triples in various states. By approximating the expected number of connected triples

of a given type, in terms of expected numbers of pairs and individuals, a small closed system of equations is obtained (Matsuda et al., 1992; Keeling, 1999; House and Keeling, 2010). Similarly, at the individual level, the time derivative of the probability of a connected pair being in a given state depends on the joint probabilities of the triples which it forms with its neighbours. By expressing the joint distribution for such triples in terms of pairs and individuals, a closed system is obtained (Sharkey, 2008).

We will first outline the construction of pair-based moment closure systems (at the population level), with a focus on finite and directed networks. We will adopt the systematic approach to construction, starting at the individual level and then making independence and homogeneity assumptions, given by Sharkey (2008). We will then go on to develop 'exact' moment closure systems for the case of tree networks, extending the work of Sharkey et al. (2013) and Kiss et al. (2014) from SIR to SEIR dynamics. We then propose an exact closure theorem, extending a result given by Kiss et al. (2014), which allows us to define exact systems for non-tree networks, and examine the relevance of network structure to the dimensionality of such systems. Finally, we will define hierarchies of approximating moment closure systems, which are non-decreasing in dimensionality, and which start with a pair-based system and end with an exact system (Sharkey and Wilkinson, 2015).

3.2 Pair-based systems at the population level

From the table of transition rates 3.1, the following exact time derivatives for the marginal probabilities of individuals and pairs of individuals being in certain states can be derived (for the notation, see definition 1.10.1):

$$\begin{aligned}
\langle \dot{S}_i \rangle &= -\sum_{j \in N_i} T_{ij} \langle S_i I_j \rangle \qquad i \in V, \\
\langle \dot{E}_i \rangle &= \sum_{j \in N_i} T_{ij} \langle S_i I_j \rangle - \alpha_i \langle E_i \rangle, \\
\langle \dot{I}_i \rangle &= \alpha_i \langle E_i \rangle - \gamma_i \langle I_i \rangle, \\
\langle \dot{S_i I_j} \rangle &= -(T_{ij} + \gamma_j) \langle S_i I_j \rangle + \alpha_j \langle S_i E_j \rangle - \sum_{k \in N_i \setminus j} T_{ik} \langle I_k S_i I_j \rangle \qquad i, j \in V : j \in N_i, \\
\langle \dot{S_i E_j} \rangle &= -\alpha_j \langle S_i E_j \rangle - \sum_{k \in N_i \setminus j} T_{ik} \langle I_k S_i E_j \rangle + \sum_{k \in N_j \setminus i} T_{jk} \langle S_i S_j I_k \rangle,
\end{aligned} \tag{3.1}$$

where N_i the set of i's in-neighbours and we use 'dot' notation for time derivatives. Also note that

$$\langle R_i \rangle = 1 - \langle S_i \rangle - \langle E_i \rangle - \langle I_i \rangle.$$

This individual level approach for describing the dynamics was developed by Sharkey (2008).

For simplicity of exposition, in the remainder of this section we will assume that the individual level rates are homogeneous, i.e. $T_{ij} \in \{0, \beta\}, \alpha_i = \alpha, \gamma_i = \gamma$. Summing over all $i \in V$ in the first three equations of 3.1 and then over all $i, j \in V : j \in N_i$ in the others, gives the time derivatives for expected population level quantities:

$$\begin{aligned}
[\dot{S}] &= -\beta[S\overleftarrow{I}], \\
[\dot{E}] &= \beta[S\overleftarrow{I}] - \alpha[E], \\
[\dot{I}] &= \alpha[E] - \gamma[I], \\
[\dot{S\overleftarrow{I}}] &= -(\beta+\gamma)[S\overleftarrow{I}] + \alpha[S\overleftarrow{E}] - \beta[\overrightarrow{I}S\overleftarrow{I}], \\
[\dot{S\overleftarrow{E}}] &= -\alpha[S\overleftarrow{E}] - \beta[\overrightarrow{I}S\overleftarrow{E}] + \beta[S\overleftarrow{S}\overleftarrow{I}],
\end{aligned} \tag{3.2}$$

where, for example, $[S]$ is the expected number of susceptible individuals at time t, and $[S\overleftarrow{I}]$ is the expected number of ordered pairings of individuals at time t where the first is susceptible and the second is infected, and where the second can directly contact the first (such population level quantities, defined explicitly for directed networks, were considered by Sharkey (2006)). These equations are exact but do not form a closed solvable system. To overcome this obstacle we can attempt to represent the triples, i.e. $[\overrightarrow{I}S\overleftarrow{I}], [\overrightarrow{I}S\overleftarrow{E}], [S\overleftarrow{S}\overleftarrow{I}]$, in terms of pairs and singlets, by making use of the closures (see section 1.10.1):

$$\langle A_i B_j C_k \rangle \approx \langle A_i B_j \rangle \frac{\langle B_j C_k \rangle}{\langle B_j \rangle} \qquad G[i,j,k] \text{ connected open triple, } j \text{ central}, \tag{3.3}$$

$$\langle A_i B_j C_k \rangle \approx \langle A_i B_j \rangle \frac{\langle B_j C_k \rangle}{\langle B_j \rangle} \frac{\langle A_i C_k \rangle}{\langle A_i \rangle \langle C_k \rangle} \qquad G[i,j,k] \text{ connected triangle}, \tag{3.4}$$

where $A, B, C \in \{S, E, I, R\}$ and G is the underlying graph. However, for the rest of this section we will assume that there are no triangles in G such that we only need the first of these two closures (when we do not make this assumption, the second closure is used for the triples which are also triangles in G).

Making use of closure (3.3), we now write:

$$\begin{aligned}
[\overrightarrow{A}B\overleftarrow{C}] &\approx \frac{1}{\beta^2} \sum_{i,j=1}^{N} T_{ji} \sum_{k \neq i}^{N} T_{jk} \frac{\langle A_i B_j \rangle \langle B_j C_k \rangle}{\langle B_j \rangle} \\
&\approx \frac{1}{N\bar{K_{in}}^2} \frac{[\overrightarrow{A}B][B\overleftarrow{C}]}{[B]} \frac{1}{\beta^2} \sum_{i,j=1}^{N} T_{ji} \sum_{k \neq i}^{N} T_{jk} \\
&= \mathcal{C}_{(\rightarrow\leftarrow)} \frac{[\overrightarrow{A}B][B\overleftarrow{C}]}{[B]},
\end{aligned} \tag{3.5}$$

where $\bar{K_{in}}$ is the average number of in-neighbours (= average number of out-neighbours) of an individual. We can view $\mathcal{C}_{(\rightarrow\leftarrow)}$ as a correction term satisfying:

$$\text{number of triples } (i \rightarrow j \leftarrow k) = N\bar{K_{in}}^2 \mathcal{C}_{(\rightarrow\leftarrow)}.$$

Following Sharkey (2008), the second approximation in equation (3.5) is based on homogeneity assumptions:, e.g. $\langle A_i \rangle = [A]/N \forall i$ and $\langle A_i B_j \rangle = [A\overleftarrow{B}]/(N\bar{K_{in}}) \forall i, j : j \in N_i$ etc. Note that, although $[S]/N$ may be a crude approximation for a given $\langle S_i \rangle$, it performs well when computing some quantity which is computed by summing over many such probabilities; especially if the errors are due to random 'noise' (and similarly for $[A\overleftarrow{B}]/(N\bar{K_{in}})$).

We note that

$$\begin{aligned} \mathcal{C}_{(\rightarrow\leftarrow)} &= \frac{\text{number of triples } (i \rightarrow j \leftarrow k)}{\bar{K_{in}} \times \text{ number of pairs } (i \rightarrow j)} \\ &= \frac{N\bar{K_{in}}\bar{K}_{in}^{(+1)}(\text{excess})}{N\bar{K_{in}}^2} \\ &= \frac{\bar{K}_{in}^{(+1)}(\text{excess})}{\bar{K_{in}}}, \end{aligned} \tag{3.6}$$

where $\bar{K}_{in}^{(+1)}(\text{excess})$ is the average in-degree of the 'head' of an arc, not counting the 'tail' of the arc, hence the word 'excess'. This reveals $\mathcal{C}_{(\rightarrow\leftarrow)}$ as the quotient of the average *excess* in-degree of an out-neighbour and the average in-degree (= average out-degree). Thus, the approximation represented in equation (3.5) can also be arrived at by the following argument which assumes the system is in a particular configuration/pure state: firstly, the expected number of in-neighbours in state C of a randomly selected individual in state B is $[B\overleftarrow{C}]/[B]$. Like the beyond-pair independence assumption, we now assume that the expected number of in-neighbours in state C, of a randomly selected individual in state B with in-degree k, is given by $k[B\overleftarrow{C}]/\bar{K_{in}}[B]$ (so that it scales linearly with k, which we would expect if the distribution of the states of an individual's neighbours were not correlated with its own state and degree). Therefore, the expected number of *excess* in-neighbours of the second individual in a randomly selected $\overrightarrow{A}B$ pair, that are in state C, is given by $\bar{K}_{in}^{(+1)}(\text{excess})[B\overleftarrow{C}]/\bar{K_{in}}[B] = \mathcal{C}_{(\rightarrow\leftarrow)}[B\overleftarrow{C}]/[B]$. The reader is directed to Rand (1999), and House and Keeling (2010) for more discussion of this type of reasoning in moment closure approximations.

An approximation for $[A\overleftarrow{B}\overleftarrow{C}]$ can be similarly derived:

$$\begin{aligned} [A\overleftarrow{B}\overleftarrow{C}] &\approx \frac{1}{\beta^2}\sum_{i,j=1}^{N} T_{ij} \sum_{k\neq i}^{N} T_{jk} \frac{\langle A_i B_j \rangle \langle B_j C_k \rangle}{\langle B_j \rangle} \\ &\approx \frac{1}{N\bar{K_{in}}^2} \frac{[A\overleftarrow{B}][B\overleftarrow{C}]}{[B]} \frac{1}{\beta^2} \sum_{i,j=1}^{N} T_{ji} \sum_{k\neq i}^{N} T_{jk} \\ &= \mathcal{C}_{(\leftarrow\leftarrow)} \frac{[A\overleftarrow{B}][B\overleftarrow{C}]}{[B]}, \end{aligned} \tag{3.7}$$

where $\mathcal{C}_{(\leftarrow\leftarrow)}$ satisfies:

$$\text{number of triples } (i \leftarrow j \leftarrow k) = N\bar{K_{in}}^2 \mathcal{C}_{(\leftarrow\leftarrow)}.$$

We note that

$$\begin{aligned}\mathcal{C}_{(\leftarrow\leftarrow)} &= \frac{\text{number of triples } (i \leftarrow j \leftarrow k)}{\bar{K_{in}} \times \text{ number of pairs } (i \rightarrow j)} \\ &= \frac{N\bar{K_{in}}\bar{K}^{(+1)}_{out}(\text{excess})}{N\bar{K_{in}}^2} \\ &= \frac{\bar{K}^{(+1)}_{out}(\text{excess})}{\bar{K_{in}}}, \end{aligned} \tag{3.8}$$

where $\bar{K}^{(+1)}_{out}(\text{excess})$ is the average out-degree of the 'head' of an arc, not counting the 'tail' of the arc. Hence, $\mathcal{C}_{(\leftarrow\leftarrow)}$ is the quotient of the average *excess* out-degree of an out-neighbour and the average out-degree (= average in-degree).

The situation is simpler if the network is undirected, in which case $\mathcal{C}_{(\rightarrow\leftarrow)} = \mathcal{C}_{(\leftarrow\leftarrow)}$. For a finite network, the needed constants can be computed from the network's (directed) adjacency matrix. For an Erdös-Rényi random graph, in the limit of large population size, we would expect $\mathcal{C}_{(\rightarrow\leftarrow)} = \mathcal{C}_{(\leftarrow\leftarrow)} = 1$, while for a K regular network (Bethe lattice) we would have $\mathcal{C}_{(\rightarrow\leftarrow)} = \mathcal{C}_{(\leftarrow\leftarrow)} = (K-1)/K$. For the standard configuration network we expect $\mathcal{C}_{(\rightarrow\leftarrow)} = \mathcal{C}_{(\leftarrow\leftarrow)} = E[K^2 - K]/(E[K])^2$ where K is the degree of a randomly selected individual (note that for random networks we must interpret the arithmetic averages in equations (3.6) and (3.8) as expected values). For these last three cases there are also (almost surely) no triangles.

If there are no triangles in G, then using closures 3.5 and 3.7 for the expected triples quantities in system (3.2) yields a closed approximating system of just six ordinary differential equations (ODEs); an additional one is required since the closure introduces the 'new' pair quantity $[S\overleftarrow{S}]$, the time derivative of which is constructed entirely analogously (a few extra variables/equations are required if there are triangles in G). This system was first constructed by Sharkey (2006).

In more heterogeneous networks, it is possible to reduce the 'strength' of the homogeneity assumptions by attempting to track the expected number $[S^a]$ of in-degree 'a' individuals which are susceptible, and the expected number $[S^a\overleftarrow{I}^b]$ of pairings of an in-degree 'a' individual with an in-degree 'b' individual where the first is susceptible and the second is infected and the second can directly contact the first (and similarly for all the various individual states, see Eames and Keeling (2002)). Thus we capture correlations between the state of an individual and it's status in relation to network structure. In an entirely analogous way to before, we can write down the following exact time derivatives for the expected population level quantities (again assuming

homogeneous rates for the individual level processes):

$$\begin{aligned}
[\dot{S^a}] &= -\beta \sum_b [S^a \overleftarrow{I}^b], \\
[\dot{E^a}] &= \beta \sum_b [S^a \overleftarrow{I}^b] - \alpha [E^a], \\
[\dot{I^a}] &= \alpha [E^a] - \gamma [I^a], \\
[S^a \overleftarrow{\dot{I}}^b] &= -(\beta+\gamma)[S^a \overleftarrow{I}^b] + \alpha [S^a \overleftarrow{E}^b] - \beta \sum_c [\overrightarrow{I}^c S^a \overleftarrow{I}^b], \\
[S^a \overleftarrow{\dot{E}}^b] &= -\beta \sum_c [\overrightarrow{I}^c S^a \overleftarrow{E}^b] - \alpha [S^a \overleftarrow{E}^b] + \beta \sum_c [S^a \overleftarrow{S}^b \overleftarrow{I}^c]. \qquad (3.9)
\end{aligned}$$

Applying the beyond-pair independence assumption, and (in-degree sensitive) homogeneity assumptions, leads to the following closures for the triples quantities (assuming at this point that there are no triangles in G):

$$\begin{aligned}
[\overrightarrow{A}^a B^b \overleftarrow{C}^c] &\approx \frac{1}{\beta^2} \sum_{i:|N_i|=a} \sum_{j:|N_j|=b} T_{ji} \sum_{k:|N_k|=c, k\neq i} T_{jk} \frac{\langle A_i B_j \rangle \langle B_j C_k \rangle}{\langle B_j \rangle} \\
&\approx \frac{[b]}{[\overrightarrow{a}\, b][b \overleftarrow{c}\,]} \frac{[\overrightarrow{A}^a B^b][B^b \overleftarrow{C}^c]}{[B^b]} \frac{1}{\beta^2} \sum_{i:|N_i|=a} \sum_{j:|N_j|=b} T_{ji} \sum_{k:|N_k|=c, k\neq i} T_{jk} \\
&= \mathcal{C}_{(\overrightarrow{a}\, b \overleftarrow{c}\,)} \frac{[\overrightarrow{A}^a B^b][B^b \overleftarrow{C}^c]}{[B^b]}, \qquad (3.10)
\end{aligned}$$

$$\begin{aligned}
[A^a \overleftarrow{B}^b \overleftarrow{C}^c] &\approx \frac{1}{\beta^2} \sum_{i:|N_i|=a} \sum_{j:|N_j|=b} T_{ij} \sum_{k:|N_k|=c, k\neq i} T_{jk} \frac{\langle A_i B_j \rangle \langle B_j C_k \rangle}{\langle B_j \rangle} \\
&\approx \frac{[b]}{[a \overleftarrow{\,b}\,][b \overleftarrow{c}\,]} \frac{[A^a \overleftarrow{B}^b][B^b \overleftarrow{C}^c]}{[B^b]} \frac{1}{\beta^2} \sum_{i:|N_i|=a} \sum_{j:|N_j|=b} T_{ji} \sum_{k:|N_k|=c, k\neq i} T_{jk} \\
&= \mathcal{C}_{(a \overleftarrow{\,b}\, \overleftarrow{c}\,)} \frac{[A^a \overleftarrow{B}^b][B^b \overleftarrow{C}^c]}{[B^b]}, \qquad (3.11)
\end{aligned}$$

where $[a]$ is the number of individuals of in-degree a, $[\overrightarrow{a}\, b]$ is the number of ordered pairings of individuals where the first has in-degree a, the second has in-degree b, and the first can directly contact the second. $\mathcal{C}_{(\overrightarrow{a}\, b \overleftarrow{c}\,)}$ and $\mathcal{C}_{(a \overleftarrow{\,b}\, \overleftarrow{c}\,)}$ can be viewed as correction terms satisfying:

$$[\overrightarrow{a}\, b \overleftarrow{c}\,] = \frac{[\overrightarrow{a}\, b][b \overleftarrow{c}\,]}{[b]} \mathcal{C}_{(\overrightarrow{a}\, b \overleftarrow{c}\,)}, \qquad (3.12)$$

$$[a \overleftarrow{\,b}\, \overleftarrow{c}\,] = \frac{[a \overleftarrow{\,b}\,][b \overleftarrow{c}\,]}{[b]} \mathcal{C}_{(a \overleftarrow{\,b}\, \overleftarrow{c}\,)}. \qquad (3.13)$$

$\mathcal{C}_{(\overrightarrow{a}\, b \overleftarrow{c}\,)}$ is the quotient of the average number of *excess* in-neighbours of in-degree c of the 'b' individual in an $\overrightarrow{a}\, b$ pair, and the average number of in-neighbours of in-degree

c of an in-degree b individual (and this is symmetric in a and c). $\mathcal{C}_{(a\overleftarrow{b}\overleftarrow{c})}$ is the quotient of the the average number of *excess* out-neighbours of in-degree a of the 'b' individual in a $b\overleftarrow{c}$ pair, and the average number of out-neighbours of in-degree a of an in-degree b individual. Note that this is equivalent to the quotient of the average number of *excess* in-neighbours of in-degree c of the 'b' individual in an $a\overleftarrow{b}$ pair, and the average number of in-neighbours of in-degree c of an in-degree b individual.

For example, if there are no $\overrightarrow{a}b\overleftarrow{c}$ triples then $\mathcal{C}_{(\overrightarrow{a}b\overleftarrow{c})} = [\overrightarrow{a}b\overleftarrow{c}] = 0$. For the standard configuration network, in which the degrees of individuals are not correlated, we would expect $\mathcal{C}_{(\overrightarrow{a}b\overleftarrow{c})} = \mathcal{C}_{(a\overleftarrow{b}\overleftarrow{c})} = (b-1)/b$. Even in a configuration network where the degrees of neighbouring individuals are correlated it could still take this value since, for a given individual, knowing the degree of one of its neighbours may not provide extra information about the degrees of its other neighbours. For a given finite (non-random) network, we can compute the needed constants from the (directed) adjacency matrix.

A slightly less general form, for the case of undirected networks, is given by Eames and Keeling (2002) who assume that $\mathcal{C}_{(\overrightarrow{a}b\overleftarrow{c})} = \mathcal{C}_{(a\overleftarrow{b}\overleftarrow{c})} = (b-1)/b$. This ignores the possibility of the degrees of an individual's neighbours being correlated.

The approximation represented by equation (3.10) can also be arrived at by the following argument: Note that the expected number of in-neighbours of in-degree c and state C, of a randomly selected in-degree b individual in state B, is given by $[B^b\overleftarrow{C^c}]/[B^b]$. Making the beyond-pair independence assumption, it is also reasonable to assume that, given the selected individual has k in-neighbours of in-degree c, then this expected number is then given by $k[b][B^b\overleftarrow{C^c}]/[b\overleftarrow{c}][B^b]$ (so it scales linearly with k in a reasonable way). Therefore, the expected number of *excess* in-neighbours of in-degree c and state C, of the 'b' individual in an $\overrightarrow{A^a}B^b$ pair, is given by $[\overrightarrow{a}b\overleftarrow{c}][b][B^b\overleftarrow{C^c}]/[\overrightarrow{a}b][b\overleftarrow{c}][B^b] = \mathcal{C}_{(\overrightarrow{a}b\overleftarrow{c})}[B^b\overleftarrow{C^c}]/[B^b]$. A similar argument exists for the approximation in 3.11.

Using closures 3.10 and 3.11 for the expected triples quantities in system (3.9) yields a closed approximating system where the number of ODEs is of the order of M^2, where M is the number of distinct degrees that the individuals in the network may have. This can be very large and so further assumptions can be made about both the network structure and the dynamics on the network in order to simplify the system and reduce the number of variables (see House and Keeling (2010)).

All of these pair-wise systems can be derived, as shown, by first making the beyond-pair-level independence assumptions, 3.3 and 3.4, and then further homogeneity assumptions. Removing just the homogeneity assumptions, we are left with a system where the number of variables is of the order of $N\bar{K_{in}}$ which, with modern computing power, can be solvable for non-trivial finite networks. Such systems, which embrace heterogeneity but make the beyond-pair independence assumption, are the focus of the remainder of this chapter. In particular, how accurate is this assumption in different

scenarios, and is it possible to lessen the strength of the assumption by making it at a higher level, e.g. a beyond-triple independence assumption? Also, what is the relevance of triangles and cycles in general?

3.3 Exact systems for tree networks

Definition 3.3.1. *We use the following notation to denote quantities relating to 'subsystems':*

- ψ_W *is a subsystem comprising of the set of individuals* $W \subset V(D)$.
- *Let* A *be a mapping from* W *to* $\{S, E, I, R\}$*, and let* A_i *be the image of* $i \in W$ *under* A*. Thus,* A *can be interpreted as a configuration for subsystem* ψ_W*, i.e. the configuration where, for all* $i \in W$*, individual* i *is in state* A_i.
- ψ_W^A *denotes the indicator random variable for the event that* $\sigma_i(t) = A_i, \forall i \in W$*. Hence we can write:*

$$\mathbb{P}_{T,\bar{\alpha},\bar{\gamma}}(\sigma_i(t) = A_i, \forall i \in W) = \mathbb{P}_{T,\bar{\alpha},\bar{\gamma}}(\psi_W = A) = \langle \psi_W^A \rangle_{T,\bar{\alpha},\bar{\gamma}}.$$

Remark. When referring to such indicator variables, the probability space on which they are defined will be clear from the context, i.e. the model under consideration. Unless otherwise stated, all the random elements in any equation or system of equations are defined on a single (arbitrary) probability space which is consistent with a 'run' of the model. Hence, reference to initial conditions and parametrisation will usually be dropped.

Definition 3.3.2. *We place set* $C \subset V(D)$ *into the 'cavity state' (Karrer and Newman, 2010) by removing the ability of all individuals in* C *to make infectious contacts to others, this being equivalent to removing all arcs emanating from all* $j \in C$*. We will denote the resulting network by* $D(cav\ C)$ *and the resulting contact parameter matrix by* $T_{\setminus C}$.

We note that susceptible individuals cannot make infectious contacts until they receive an infectious contact. Therefore, if a set $C \subset V$ are susceptible at time t then whether or not this set was placed into the cavity state has had no impact on the system up to time t.

We can now write:

$$\langle \psi_W^A \mid \sigma_i(t) = S, \forall i \in C \rangle_T = \langle \psi_W^A \mid \sigma_i(t) = S, \forall i \in C \rangle_{T_{\setminus C}}, \tag{3.14}$$

which holds wherever such conditional expectations are defined. We have used the standard vertical line to indicate an event being conditioned upon. Moreover, if $C \subset W$

and $A_j = S, \forall j \in C$, then:

$$\frac{\langle \psi_W^A \rangle_T}{\mathbb{P}_T(\sigma_j(t) = S, \forall j \in C)} = \frac{\langle \psi_W^A \rangle_{T_{\setminus C}}}{\mathbb{P}_{T_{\setminus C}}(\sigma_j(t) = S, \forall j \in C)}, \tag{3.15}$$

wherever the denominators are positive, since $\psi_W^A = 1$ implies $\sigma_j(t) = S, \forall j \in C$. Note that the numerators are zero if the denominators are zero. This implies that, for an arbitrary subsystem ψ_W, and arbitrary configuration A, we can write:

$$\langle \psi_W^A \rangle_T = \langle \psi_W^A \rangle_{T_{\setminus C}} \qquad \text{if } C \subset W, A_j = S, \forall j \in C. \tag{3.16}$$

Let us now note that the fates of two individuals i and k are statistically independent if their in-components are disjoint and there is no imposed correlation between the two in-components at $t = 0$. Indeed, it is intuitive that an individual i's fate is, in general, independent of the behaviour of individuals outside of its in-component, and that the in-components of individuals in i's in-component are subsets of i's in-component. In fact, it can be shown that S_i, for $i \in V$, is a function of the initial conditions and a set of random variables (independent of initial conditions) assigned to the arcs where both 'ends' are in i's in-component (Kenah and Robins, 2007), and these random variables are independent when assigned to arcs which emanate from different individuals (and so any dependence between S_i and S_j, where i and j have disjoint in-components, must derive from the initial conditions). The random variable in question, assigned to arc (i, j) say, is the time that it takes i to make its first infectious contact to j upon i's becoming exposed (or since $t = 0$ if i is initially infected), given that i does become exposed (or is initially exposed/infected). Once $j \in V$ has received its first infectious contact, or if j is not susceptible at $t = 0$, its fate is then completely independent of the states and behaviour of the rest of the population (see transition table 3.1).

Definition 3.3.3. *An undirected graph $G = (V, E)$ is a 'tree' iff for any two individuals/vertices $i, j \in V$ there is a unique path from i to j. An undirected graph $G = (V, E)$ is a 'forest' iff for any two individuals/vertices $i, j \in V$ there is either a unique path from i to j or no path at all, i.e. a forest is a collection of disjoint trees.*

Definition 3.3.4. *A network/digraph $D = (V, A)$ is a 'tree network' iff its underlying graph is a tree or forest. See figure 3.1 for an example tree network.*

Definition 3.3.5. *An undirected graph $G = (V, E)$ is 'biconnected' iff there are at least two vertex disjoint paths between every pair of vertices. For example, a tree is not biconnected since there is exactly one path between every pair of vertices in the tree.*

For any tree network $D = (V, A)$ (with underlying graph G), all neighbours of $j \in V$ in $D(\text{cav } j)$ have mutually disjoint in-components, and so the fates of these neighbours are statistically independent if there are no correlations between the states

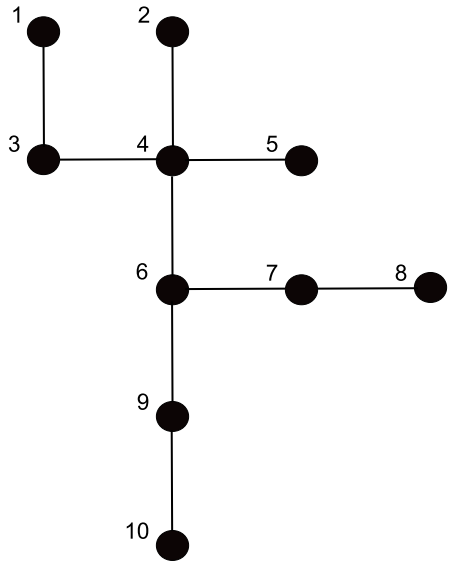

Figure 3.1: An undirected tree network

of any individuals at $t = 0$. For the rest of this chapter, unless otherwise stated, we will thus assume that the initial conditions imply that the states of individuals are statistically independent at $t = 0$ (a pure initial system state satisfies this but choosing a specific number of initial infecteds at random does not). In the case of a tree network we can now write:

$$\begin{aligned}
\langle X_i S_j Z_k \rangle_T &= \langle X_i S_j Z_k \rangle_{T_{\setminus j}} \qquad (i,j),(k,j) \in E(G) \\
&= \langle S_j \rangle_{T_{\setminus j}} \frac{\langle X_i S_j \rangle_{T_{\setminus j}}}{\langle S_j \rangle_{T_{\setminus j}}} \frac{\langle S_j Z_k \rangle_{T_{\setminus j}}}{\langle S_j \rangle_{T_{\setminus j}}} \\
&= \frac{\langle X_i S_j \rangle_{T_{\setminus j}} \langle S_j Z_k \rangle_{T_{\setminus j}}}{\langle S_j \rangle_{T_{\setminus j}}} \\
&= \frac{\langle X_i S_j \rangle_T \langle S_j Z_k \rangle_T}{\langle S_j \rangle_T}, \qquad (3.17)
\end{aligned}$$

where $X, Z \in \{S, E, I, R\}$ and the first and last steps follow from equation (3.16) (note that $\langle S_j \rangle = 0$ implies $\langle X_i S_j Z_k \rangle = 0$). Indeed, we can conclude that for a tree-network, given some individual is susceptible at time t and there are no correlations between individuals at $t = 0$, then the states of the neighbours of that individual are statistically independent at time t. This is intuitive since if j is susceptible at time t, in a tree network, then there has not been any 'communication' between the in-components of the neighbours of j before time t (since this would require the infection 'passing through' individual j).

Theorem 3.3.1. *For Markovian network-based SEIR dynamics on a tree network, if the states of individuals are statistically independent at $t = 0$ then the following system*

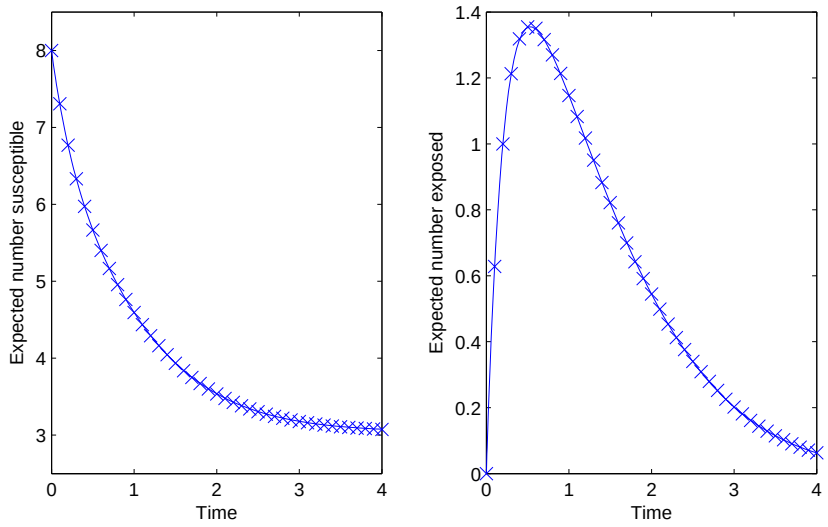

Figure 3.2: Comparison between the numerical solution of system (3.18) (curves) and simulations of the corresponding stochastic process (crosses). In this example, the network on which the disease spreads is that which is depicted in figure 3.1, with $T \in \{0, 2\}, \alpha_i = 2, \gamma_i = 1, \forall i$. Individuals 3 and 9 were initially infected, all others being initially susceptible.

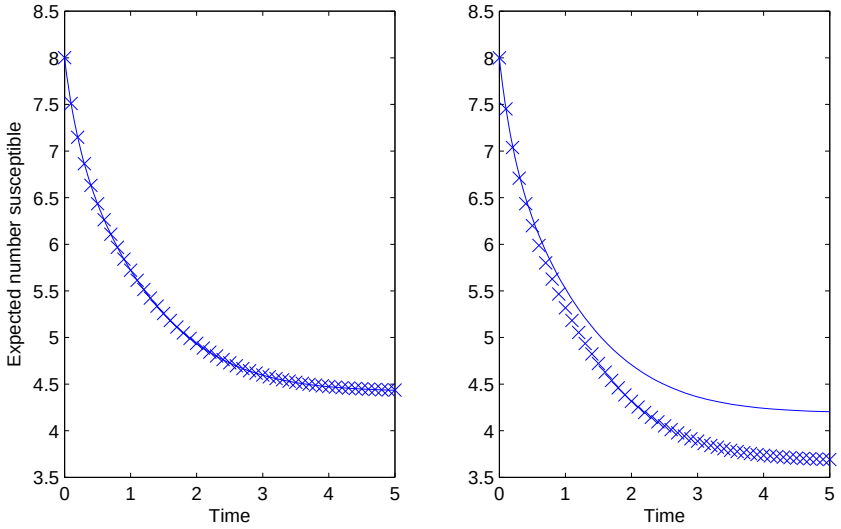

Figure 3.3: Comparison between the numerical solution of system (3.18) (curves) and simulations of the corresponding stochastic process (crosses). In this example, the network on which the disease spreads is that which is depicted in figure 3.1, with $T \in \{0, 2\}, \alpha_i = 2, \gamma_i = 1, \forall i$. For a) all individuals were (independently) initially infected with probability 0.2 and initially susceptible otherwise. For b) two initial infecteds were chosen uniformly at random (without replacement) and the rest were initially susceptible, this meaning that the states of individuals were not statistically independent at $t = 0$.

is exactly consistent with the underlying stochastic process:

$$\begin{aligned}
\langle \dot{S}_i \rangle &= -\sum_{j \in N_i} T_{ij} \langle S_i I_j \rangle, \qquad i \in V \\
\langle \dot{E}_i \rangle &= \sum_{j \in N_i} T_{ij} \langle S_i I_j \rangle - \alpha_i \langle E_i \rangle, \\
\langle \dot{I}_i \rangle &= \alpha_i \langle E_i \rangle - \gamma_i \langle I_i \rangle, \\
\langle \dot{S_i I_j} \rangle &= -(T_{ij} + \gamma_j) \langle S_i I_j \rangle + \alpha_j \langle S_i E_j \rangle - \sum_{k \in N_i \setminus j} T_{ik} \frac{\langle I_k S_i \rangle \langle S_i I_j \rangle}{\langle S_i \rangle}, \qquad i, j \in V : j \in N_i \\
\langle \dot{S_i E_j} \rangle &= -\alpha_j \langle S_i E_j \rangle - \sum_{k \in N_i \setminus j} T_{ik} \frac{\langle I_k S_i \rangle \langle S_i E_j \rangle}{\langle S_i \rangle} + \sum_{k \in N_j \setminus i} T_{jk} \frac{\langle S_i S_j \rangle \langle S_j I_k \rangle}{\langle S_j \rangle}, \\
\langle \dot{S_i S_j} \rangle &= -\sum_{k \in N_j \setminus i} T_{jk} \frac{\langle S_i S_j \rangle \langle S_j I_k \rangle}{\langle S_j \rangle} - \sum_{k \in N_i \setminus j} T_{ik} \frac{\langle I_k S_i \rangle \langle S_i S_j \rangle}{\langle S_i \rangle}.
\end{aligned} \tag{3.18}$$

Proof. The theorem is derived from the application of the closure in equation (3.17) to system (3.1) (and deriving an expression for $\langle \dot{S_i S_j} \rangle$ which also makes use of the closure). □

See figures 3.2 and 3.3 for numerical examples on the tree network depicted in figure 3.1. Although the system is in general inexact if the network is a non-tree network, the equations can still be applied to obtain approximate results (the system may become exact for certain initial conditions on non-tree networks - see Sharkey et al. (2013)).

3.4 Exact systems for non-tree networks

Here we prove a theorem which generalises the closure represented by equation (3.17). We then use this to derive a class of exact systems for Markovian SEIR dynamics on arbitrary networks. We illustrate this with some examples, and finally state a theorem specifying the maximum size of subsystem needed to exactly represent the dynamics on a given network (this maximum subsystem size gives us some idea of the total number of equations which will be needed).

3.4.1 Exact closure theorem

For a given directed graph $D = (V, A)$, we make the following definitions:

Definition 3.4.1. *Let $X, Y, Z \subset V$ be disjoint and non-empty. The set Z is 'dynamically partitioning' with respect to X and Y iff we have $f_E(X, Y, Z) = 1$ where:*

$$f_E(X, Y, Z) = \begin{cases} 1 & \text{if } IN(X) \cap IN(Y) = \emptyset \quad (\text{in } D(cav\ Z)) \\ 0 & \text{otherwise} \end{cases}. \tag{3.19}$$

Here, E is chosen to represent 'exact'; this is appropriate since we shall now see that $f_E(X, Y, Z) = 1$ implies the existence of an exact closure relation.

Remark. If the network is undirected then $f_E(X,Y,Z) = 1$ if and only if there is no path between X and Y in $D(\text{cav } Z)$.

Theorem 3.4.1. *(Exact closure theorem) Let $X,Y,Z \subset V(D)$ be disjoint and non-empty. If Z is dynamically partitioning with respect to X and Y, and the states of individuals are statistically independent at $t = 0$, then:*

$$\langle \psi_X^A \psi_Y^B \psi_Z^C \rangle = \frac{\langle \psi_X^A \psi_Z^C \rangle \langle \psi_Y^B \psi_Z^C \rangle}{\langle \psi_Z^C \rangle} \qquad \text{if } C_i = S, \forall i \in Z \tag{3.20}$$

holds for arbitrary subsystem configurations A, B, C, wherever $\langle \psi_Z^C \rangle \neq 0$. However, note that $\langle \psi_Z^C \rangle = 0$ implies $\langle \psi_X^A \psi_Y^B \psi_Z^C \rangle = 0$.

Proof. Since $C_i = S, \forall i \in Z$, we can write (see equation (3.16)):

$$\langle \psi_X^A \psi_Y^B \psi_Z^C \rangle_T = \langle \psi_X^A \psi_Y^B \psi_Z^C \rangle_{T_{\setminus Z}},$$

and since the in-components of X and Y are disjoint in $D(\text{cav } Z)$, which follows from the fact that Z is dynamically partitioning with respect to X and Y, then the states of individuals in X are independent of the states of individuals in Y in $D(\text{cav } Z)$ (given that the states of individuals are statistically independent at $t = 0$). We can now write:

$$\begin{aligned} \langle \psi_X^A \psi_Y^B \psi_Z^C \rangle_{T_{\setminus Z}} &= \frac{\langle \psi_X^A \psi_Z^C \rangle_{T_{\setminus Z}} \langle \psi_Y^B \psi_Z^C \rangle_{T_{\setminus Z}}}{\langle \psi_Z^C \rangle_{T_{\setminus Z}}} && (3.21) \\ &= \frac{\langle \psi_X^A \psi_Z^C \rangle_T \langle \psi_Y^B \psi_Z^C \rangle_T}{\langle \psi_Z^C \rangle_T}, && (3.22) \end{aligned}$$

where the last step follows from equation (3.16). □

Remark. It is not difficult to envisage that the exact closure will hold true for other non-Markovian models of SEIR dynamics on networks. The closure holds when an individual or set of individuals being susceptible at time t implies that two other sets of individuals have not interacted up to time t, and so no statistical dependence between these two sets can have emerged by time t. However, this is negated if statistical dependence between the two sets is built into the initial conditions.

The theorem is a generalisation of the main result of Kiss et al. (2014) which is stated in terms of single dynamically partitioning individuals on undirected networks, and SIR dynamics rather than SEIR. In that context they are referred to simply as partitioning individuals due to their correspondence to graph partitioning (Newman, 2010). Some examples of applying the exact closure theorem are shown in Figure 3.4. In this Figure and throughout the remainder of this thesis, network links without arrowheads denote undirected links whereas those with arrowheads denote directed links. We will next define an exact (closed) system which can take advantage of closure (3.20) such that the number of variables is to some extent curtailed.

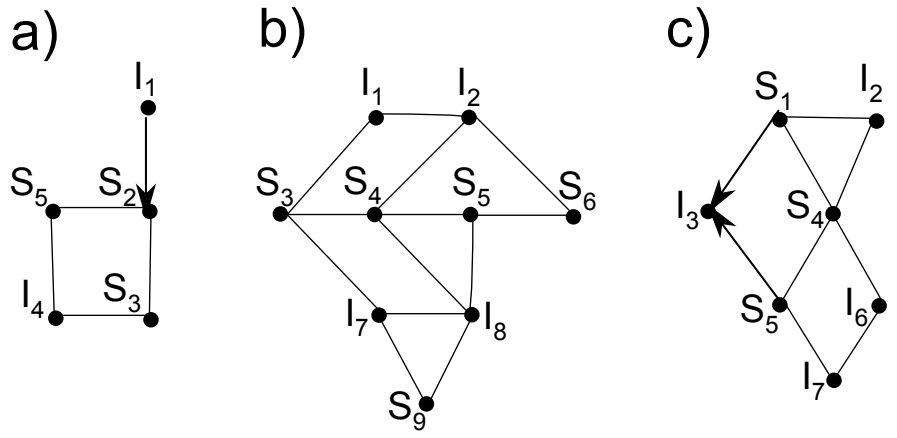

Figure 3.4: Three examples of applying the exact closure theorem. Here directed links have arrowheads and undirected links do not. a) This configuration is a typical example of when dynamical partitioning allows an exact closure. Applying Theorem 3.4.1, we see that there is dynamical partitioning about node 2, so we have $\langle I_1 S_2 S_3 I_4 S_5\rangle = \langle I_1 S_2\rangle\langle S_2 S_3 I_4 S_5\rangle/\langle S_2\rangle$. b) We can dynamically partition on this graph about a cluster of susceptible nodes. In fact there are two exact closures we can write down: $\langle I_1 I_2 S_3 S_4 S_5 S_6 I_7 I_8 S_9\rangle = \langle I_1 I_2 S_3 S_4 S_6\rangle\langle S_3 S_4 S_5 S_6 I_7 I_8 S_9\rangle/\langle S_3 S_4 S_6\rangle = \langle I_1 I_2 S_3 S_4 S_5 S_6\rangle\langle S_3 S_4 S_5 I_7 I_8 S_9\rangle/\langle S_3 S_4 S_5\rangle$. c) Here we can apply the exact closure theorem to obtain $\langle S_1 I_2 S_4 S_5 I_6 I_7\rangle = \langle S_1 I_2 S_4\rangle\langle S_4 S_5 I_6 I_7\rangle/\langle S_4\rangle$. Note that I_3 is not included in this closure.

Definition 3.4.2. *Let ψ_W^A be an indicator variable, as in definition 3.3.1, and let $k \in W$ and $X \in \{S, E, I, R\}$. Then*

$$h_k^X(\psi_W^A) \equiv \psi_W^B, \tag{3.23}$$

where

$$B_i = \begin{cases} A_i & \forall i \in W, i \neq k \\ X & i = k. \end{cases} \tag{3.24}$$

This means that h_k^X changes the configuration by putting individual k into state X. If $A_k = X$ then $h_k^X(\psi_W^A) \equiv \psi_W^A$.

Theorem 3.4.2. *For Markovian network-based SEIR dynamics, the time derivative for the probability of an arbitrary susbsystem configuration (expectation of the indicator*

variable for the event that it occurs) can be written:

$$\begin{aligned}\dot{\langle\psi_W^A\rangle} &= \sum_{k\in W}\mathbb{1}(A_k=S)\left[-\sum_{n\in V\setminus W}T_{kn}\langle\psi_W^A I_n\rangle-\sum_{n\in W}T_{kn}\mathbb{1}(A_n=I)\langle\psi_W^A\rangle\right]\\ &+\sum_{k\in W}\mathbb{1}(A_k=E)\left[\sum_{n\in V\setminus W}T_{kn}\langle h_k^S(\psi_W^A)I_n\rangle\right.\\ &\left.+\sum_{n\in W}T_{kn}\mathbb{1}(A_n=I)\langle h_k^S(\psi_W^A)\rangle-\alpha_k\langle\psi_W^A\rangle\right]\\ &+\sum_{k\in W}\mathbb{1}(A_k=I)\left[\alpha_k\langle h_k^E(\psi_W^A)\rangle-\gamma_k\langle\psi_W^A\rangle\right]\\ &+\sum_{k\in W}\mathbb{1}(A_k=R)\left[\gamma_k\langle h_k^I(\psi_W^A)\rangle\right]. \end{aligned}\tag{3.25}$$

Proof. The theorem is straightforward to prove by a careful consideration of the transition rates in the Markov chain (see table 3.1), and some tedious algebra (see Sharkey (2013) for the equivalent equation, and proof, in the context of SIR dynamics). □

The equations in system (3.1) can now be seen to be special cases of theorem 3.4.2. By using this theorem to write down time derivatives for $\langle S_i\rangle, \langle E_i\rangle$ and $\langle I_i\rangle, \forall i \in V$, and also for any 'new' variables which emerge, a closed and exact system M is obtained where the variables are a moment-induced set of probabilities of subsystem configurations. The number of variables in M is finite since both $|V|$ is finite and the number of states for an individual is finite. However, the number of equations will generally be very large for most systems, preventing numerical solution. Note that the number of subsystems of size $n \in \{1, 2, \ldots, N\}$, which appear in the variables of M, is of the order of the number of distinct subgraphs of D which consist of n individuals and where at least one individual is reachable from all others (in the subgraph). This follows from the fact that, with reference to equation (3.25), the subsystems on the right-hand-side either consist of the same set of individuals as on the left-hand-side, i.e. W, or of $W \cup j$ where there is at least one arc from $j \in V \setminus W$ to some individual in W.

To reduce the number of equations we need to curtail the generation of 'new' subsystems which occurs as a result of the repeated application of theorem 3.4.2. With reference to equation (3.25) in theorem 3.4.2, we note the following:

Corollary 3.4.1. *For subsystem ψ_W and configuration A, if $A_k = S$ where $k \in W$, and if $f_E(n, W \setminus k, k) = 1$ where $n \in V \setminus W$, then*

$$\langle\psi_W^A I_n\rangle = \frac{\langle\psi_W^A\rangle\langle S_k I_n\rangle}{\langle S_k\rangle}.\tag{3.26}$$

Proof. Follows from the exact closure theorem. □

Making use of corollary 3.4.1, which expresses the probability of a subsystem configuration consisting of $|W|+1$ individuals in terms of subsystem configurations consisting of $|W|$ or less individuals, in equation (3.25) of theorem 3.4.2 gives:

$$\begin{aligned}
\dot{\langle\psi_W^A\rangle} &= \sum_{k\in W}\mathbb{1}(A_k=S)\Bigg[-\sum_{n\in W}T_{kn}\mathbb{1}(A_n=I)\langle\psi_W^A\rangle\Bigg] \\
&\quad+\sum_{k\in W}\mathbb{1}(A_k=E)\Bigg[\sum_{n\in W}T_{kn}\mathbb{1}(A_n=I)\langle h_k^S(\psi_W^A)\rangle-\alpha_k\langle\psi_W^A\rangle\Bigg] \\
&\quad+\sum_{k\in W}\mathbb{1}(A_k=I)\Bigg[\alpha_k\langle h_k^E(\psi_W^A)\rangle-\gamma_k\langle\psi_W^A\rangle\Bigg] \\
&\quad+\sum_{k\in W}\mathbb{1}(A_k=R)\Bigg[\gamma_k\langle h_k^I(\psi_W^A)\rangle\Bigg] \\
&\quad-\sum_{k\in W}\mathbb{1}(A_k=S)\sum_{n\in V\setminus W}T_{kn}\Bigg[\Big(1-f_E(n,W\setminus k,k)\Big)\langle\psi_W^A I_n\rangle \\
&\quad+f_E(n,W\setminus k,k)\frac{\langle\psi_W^A\rangle\langle S_kI_n\rangle}{\langle S_k\rangle}\Bigg] \\
&\quad+\sum_{k\in W}\mathbb{1}(A_k=E)\sum_{n\in V\setminus W}T_{kn}\Bigg[\Big(1-f_E(n,W\setminus k,k)\Big)\langle h_k^S(\psi_W^A)I_n\rangle \\
&\quad+f_E(n,W\setminus k,k)\frac{\langle h_k^S(\psi_W^A)\rangle\langle S_kI_n\rangle}{\langle S_k\rangle}\Bigg]. \qquad (3.27)
\end{aligned}$$

Thus, by taking advantage of the closure (corollary 3.4.1), we have limited the number of subsystems of size greater than $|W|$ which appear on the right-hand-side of equation (3.25) of theorem 3.4.2.

For an arbitrary network, by applying equation (3.27) to the indicator random variables S_i, E_i and I_i for all $i \in \{1, 2, ..., N\}$, and then reapplying it to every new subsystem configuration that emerges, a closed set of differential equations for the exact time-evolution of the probability of an individual being in a particular state is obtained for all individuals (we also must assume that f_E is zero when one of its arguments is the empty set).

Definition 3.4.3. *For a given network (parametrisation), a closed and exact system M_E is obtained by using equation (3.27) to write down time derivatives for $\langle S_i\rangle, \langle E_i\rangle$ and $\langle I_i\rangle, \forall i \in V$, and also for any 'new' variables which emerge (the size of this system is automatically curtailed, where possible, by making use of the exact closure theorem).*

Remark. It follows that $\langle S_i\rangle$, $\langle E_i\rangle$ and $\langle I_i\rangle$ ($\forall i \in V$) and $\langle S_iI_j\rangle$ ($\forall i, j \in V : T_{ij} > 0$) are variables in M_E for any network.

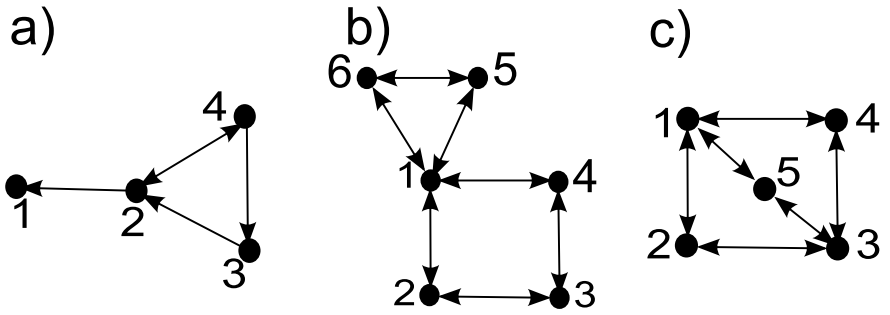

Figure 3.5: Some example networks. For dynamics on these networks we consider a homogeneous recovery rate γ, a homogeneous rate of becoming infected once exposed α, and a contact rate of 1 across each network link (arc).

3.4.2 Examples

Before determining the network structures under which dynamical partitioning occurs more generally, we consider some examples. For further examples in the context of undirected networks the reader is directed to Kiss et. al. (2014).

Example 1

Consider the network in Figure 3.5a. Let us suppose that all individuals recover at rate γ when infected, become infected at rate α when exposed, and that the contact rate across all network links is unity. For simplicity we shall also make this assumption through the remainder of the explicit examples in this chapter. We can apply the closure in corollary 3.4.1, which is a special case of the exact closure theorem, to build up the moment-induced system M_E. Let us consider the probability of individual 1 being exposed to see how this works. We have:

$$\langle \dot{E_1} \rangle = \langle S_1 I_2 \rangle - \alpha \langle E_1 \rangle. \tag{3.28}$$

Here and throughout this chapter we order individuals according to the numerical order of their labels; the relevant subsystem configurations need to be understood with reference to the full network. Now, individual 2 is dynamically partitioning with respect to individuals 1 and 3, and it is also dynamically partitioning with respect to individuals 1 and 4. Hence:

$$\langle \dot{S_1 I_2} \rangle \;\; = \;\; -(1+\gamma)\langle S_1 I_2 \rangle + \alpha \langle S_1 E_2 \rangle,$$

and for $\langle S_1 E_2 \rangle$:

$$\begin{aligned} \langle \dot{S_1 E_2} \rangle &= \langle S_1 S_2 I_3 \rangle + \langle S_1 S_2 I_4 \rangle - \alpha \langle S_1 E_2 \rangle \\ &= \frac{\langle S_1 S_2 \rangle \langle S_2 I_3 \rangle}{\langle S_2 \rangle} + \frac{\langle S_1 S_2 \rangle \langle S_2 I_4 \rangle}{\langle S_2 \rangle} - \alpha \langle S_1 E_2 \rangle. \end{aligned} \tag{3.29}$$

Rather than a complete analysis of all moment-induced subsystem configurations that arise, we take a single pair configuration $S_2 I_3$ from this equation as an example. Here,

individual 3 is not dynamically partitioning with respect to individuals 2 and 4 but individual 2 is dynamically partitioning with respect to 1 and 3 so:

$$\langle \dot{S_2 I_3} \rangle = -\frac{\langle I_1 S_2 \rangle \langle S_2 I_3 \rangle}{\langle S_2 \rangle} - \langle S_2 I_3 I_4 \rangle - (1+\gamma)\langle S_2 I_3 \rangle + \alpha \langle S_2 E_3 \rangle. \tag{3.30}$$

Then, for $\langle S_2 I_3 I_4 \rangle$, individual 2 is dynamically partitioning with respect to individual 1 and individuals 3 and 4 so:

$$\begin{aligned} \langle S_2 \dot{I_3} I_4 \rangle &= -\frac{\langle I_1 S_2 \rangle \langle S_2 I_3 I_4 \rangle}{\langle S_2 \rangle} - 2(1+\gamma)\langle S_2 I_3 I_4 \rangle \\ &\quad +\alpha \langle S_2 E_3 I_4 \rangle + \alpha \langle S_2 I_3 E_4 \rangle. \end{aligned} \tag{3.31}$$

We see that here, M_E represents a significant dimensional reduction compared to the full system M.

Example 2

For the undirected network in Figure 3.5b there is dynamical partitioning about individual 1. Starting with (for example) the probability of individual 1 being exposed, we have:

$$\langle \dot{E_1} \rangle = \langle S_1 I_2 \rangle + \langle S_1 I_4 \rangle + \langle S_1 I_5 \rangle + \langle S_1 I_6 \rangle - \alpha \langle E_1 \rangle. \tag{3.32}$$

Now, choosing the first of these pairs to develop one part of the moment-induced system M_E gives:

$$\begin{aligned} \langle \dot{S_1 I_2} \rangle &= -\langle S_1 I_2 I_4 \rangle - \frac{\langle S_1 I_2 \rangle \langle S_1 I_5 \rangle}{\langle S_1 \rangle} - \frac{\langle S_1 I_2 \rangle \langle S_1 I_6 \rangle}{\langle S_1 \rangle} \\ &\quad -(1+\gamma)\langle S_1 I_2 \rangle + \alpha \langle S_1 E_2 \rangle, \end{aligned} \tag{3.33}$$

and

$$\begin{aligned} \langle \dot{S_1 E_2} \rangle &= \langle S_1 S_2 I_3 \rangle - \langle S_1 E_2 I_4 \rangle - \frac{\langle S_1 E_2 \rangle \langle S_1 I_5 \rangle}{\langle S_1 \rangle} - \frac{\langle S_1 E_2 \rangle \langle S_1 I_6 \rangle}{\langle S_1 \rangle} \\ &\quad -\alpha \langle S_1 E_2 \rangle. \end{aligned} \tag{3.34}$$

Then, for the first of these triples:

$$\begin{aligned} \langle S_1 \dot{S_2 I_3} \rangle &= -\langle S_1 S_2 I_3 I_4 \rangle - \frac{\langle S_1 S_2 I_3 \rangle \langle S_1 I_5 \rangle}{\langle S_1 \rangle} - \frac{\langle S_1 S_2 I_3 \rangle \langle S_1 I_6 \rangle}{\langle S_1 \rangle} \\ &\quad -(1+\gamma)\langle S_1 S_2 I_3 \rangle + \alpha \langle S_1 S_2 E_3 \rangle, \end{aligned} \tag{3.35}$$

and for the quad we have:

$$\begin{aligned} \langle S_1 \dot{S_2 I_3 I_4} \rangle &= -\frac{\langle S_1 S_2 I_3 I_4 \rangle \langle S_1 I_5 \rangle}{\langle S_1 \rangle} - \frac{\langle S_1 S_2 I_3 I_4 \rangle \langle S_1 I_6 \rangle}{\langle S_1 \rangle} \\ &\quad -2(1+\gamma)\langle S_1 S_2 I_3 I_4 \rangle + \alpha \langle S_1 S_2 E_3 I_4 \rangle + \alpha \langle S_1 S_2 I_3 E_4 \rangle. \end{aligned} \tag{3.36}$$

Here, the maximum size of a subsystem configuration is four individuals. We note that this is equal to the size of the largest cycle and that this was also true for example 1. However, this is not always the case as shown by the next example.

Example 3

Figure 3.5c shows a network where the maximum simple cycle size is 4 but the maximum size of a subsystem configuration in M_E is 5. Starting with the the probability of individual 1 being exposed we have:

$$\dot{\langle E_1 \rangle} = \langle S_1 I_2 \rangle + \langle S_1 I_4 \rangle + \langle S_1 I_5 \rangle - \alpha \langle E_1 \rangle. \tag{3.37}$$

Then, taking just the subsystem configuration in the first term:

$$\dot{\langle S_1 I_2 \rangle} = -\langle S_1 I_2 I_4 \rangle - \langle S_1 I_2 I_5 \rangle - (1+\gamma)\langle S_1 I_2 \rangle + \alpha \langle S_1 E_2 \rangle, \tag{3.38}$$

and again taking the first term gives:

$$\begin{aligned} \dot{\langle S_1 I_2 I_4 \rangle} &= -\langle S_1 I_2 I_4 I_5 \rangle - 2(1+\gamma)\langle S_1 I_2 I_4 \rangle \\ &\quad + \alpha \langle S_1 E_2 I_4 \rangle + \alpha \langle S_1 I_2 E_4 \rangle, \end{aligned} \tag{3.39}$$

and for the quad:

$$\begin{aligned} \dot{\langle S_1 I_2 I_4 I_5 \rangle} &= -3(1+\gamma)\langle S_1 I_2 I_4 I_5 \rangle + \alpha \langle S_1 E_2 I_4 I_5 \rangle \\ &\quad + \alpha \langle S_1 I_2 E_4 I_5 \rangle + \alpha \langle S_1 I_2 I_4 E_5 \rangle. \end{aligned} \tag{3.40}$$

Finally, taking the last of these terms gives:

$$\begin{aligned} \dot{\langle S_1 I_2 I_4 E_5 \rangle} &= -(2+2\gamma+\alpha)\langle S_1 I_2 I_4 E_5 \rangle \\ &\quad + \langle S_1 I_2 I_3 I_4 S_5 \rangle + \alpha \langle S_1 E_2 I_4 E_5 \rangle + \alpha \langle S_1 I_2 E_4 E_5 \rangle. \end{aligned} \tag{3.41}$$

In this case we see that the maximum size of a subsystem configuration is at the size of the system (5 individuals) and is not constrained by the largest cycle (4 individuals). This leads to the question of what aspect of a network specifies the largest subsystem configuration that appears in M_E. We answer this question in the following section.

3.4.3 System size

Here we define the type of network structures that are amenable to dynamical partitioning. We start from single node subsystems and expand out, via equation (3.27), until the largest subsystem is reached incorporating that individual before dynamical partitioning prevents larger subsystems emerging. For the undirected case, the situation simplifies considerably (Kiss et al., 2013) since all dynamically partitioning individuals are also cut-vertices (individuals which, when removed, increase the number of connected components). It is then helpful to represent the network as a collection of blocks (maximal biconnected subgraphs) where the between-block structure is tree-like (see Figure 3.6a). This makes it straightforward to assess the feasibility of constructing a solvable exact system by making use of dynamical partitioning. Notice that it is

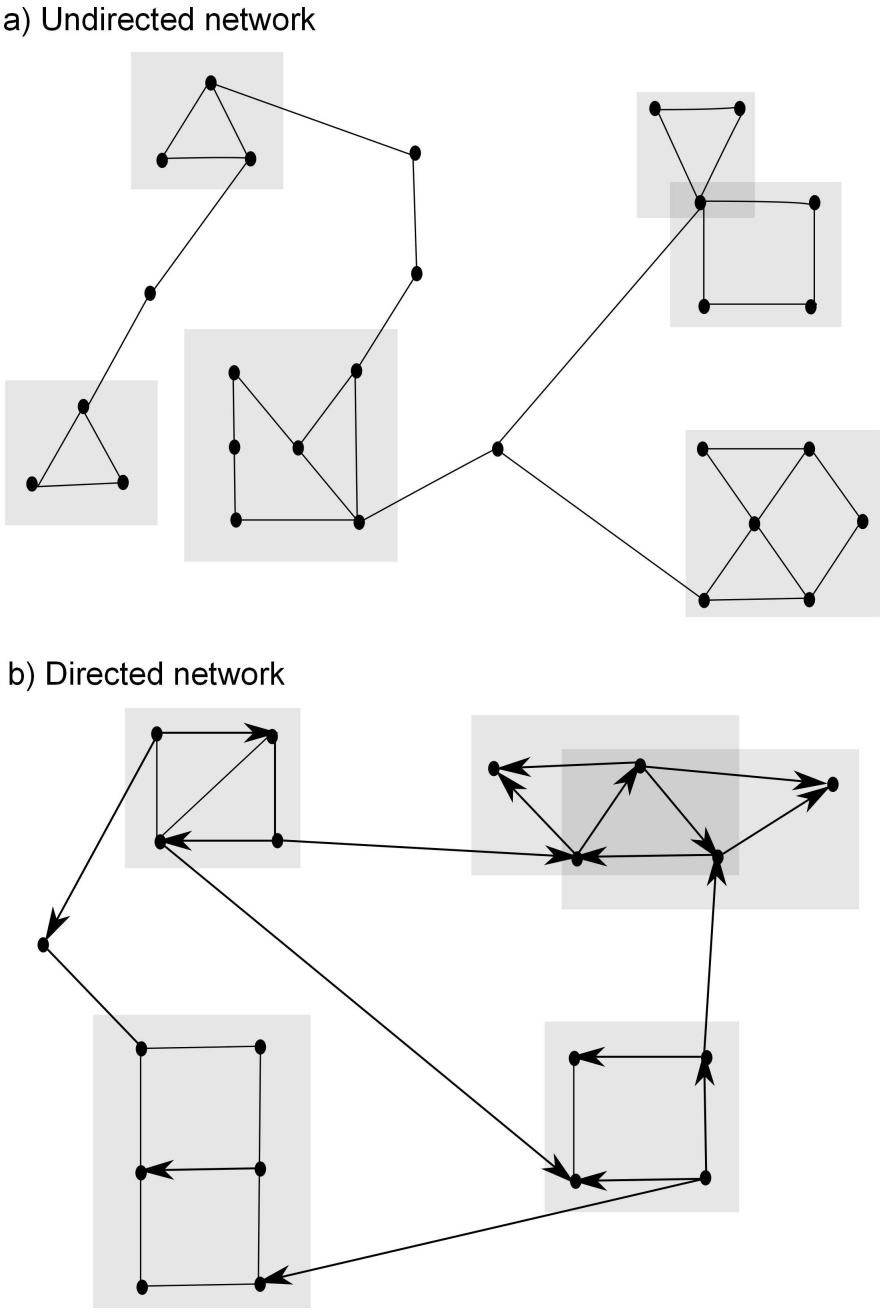

Figure 3.6: Examples of networks that decompose into transmission blocks. The transmission blocks are represented by the shaded rectangles. a) An undirected network where the effectiveness of dynamical partitioning is made clear by the number of blocks (which resemble structured households). b) A directed network where identifying the transmission blocks is more complicated.

possible for a node to belong to more than one block as in the top right of Figure 3.6a although the overlap between any two blocks can only be a single node. It is interesting that this representation of the network resembles the household model structure where analytic progress can also be made (Ball et al., 2010). For directed networks, the situation is more complicated. Here we define 'transmission blocks' to play a similar role to blocks. Indeed, blocks and transmission blocks will have equivalent definitions in the undirected case. We use the term transmission block rather than block since there are likely to be other useful extensions of the block concept for directed networks.

Definition 3.4.4. *The subgraph $D[W]$ is a 'directed sub-block' if and only if there is at least one node reachable from all others in $D[W]$ and its underlying graph is biconnected (recall from definition 3.3.5 that an undirected graph is biconnected iff there are at least two vertex disjoint paths between every pair of vertices).*

Remark. According to this definition, any block in an undirected network is also a directed sub-block. Hence, the blocks illustrated in Figure 3.6a are all directed sub-blocks.

Definition 3.4.5. *We will refer to a directed sub-block $D[W]$ as a 'transmission block' if and only if there does not exist $U \supset W$ such that $D[U]$ is also a directed sub-block.*

The shaded boxes in Figure 3.6 are examples of transmission blocks. Figure 3.6b gives an example of these on a directed graph. Notice that now it is possible for transmission blocks to overlap by more than one node (the darker shaded triangle is a directed sub-block which 'belongs' to two distinct transmission blocks). This can happen when a region of the network has paths to two or more other regions that do not have paths between each other. Figure 3.7 shows some more examples of these definitions for directed networks. Figure 3.7 a) and b) have underlying graphs that are biconnected. However, b) has a node (node 1) which is reachable from all others whereas a) does not, and so b) is a directed sub-block while a) is not. Figure 3.7 b) is also a transmission block since it is maximal. Additionally, neither have sub-graphs of the underlying graphs that are biconnected and so neither contain directed sub-blocks as subgraphs. Figure 3.7 c) is a transmission block (the underlying graph is biconnected and node 2 is reachable from all others). It also contains several directed sub-blocks (for example nodes 1,2 and 3). Figure 3.7 d) contains a transmission block as a subgraph (nodes 1,2,3,4) and contains several directed sub-blocks.

We can now state the main result of this section:

Theorem 3.4.3. *The largest subsystem configuration (not necessarily unique) in M_E consists of the same number of individuals as the largest transmission block, or it contains 2 individuals if there are no transmission blocks.*

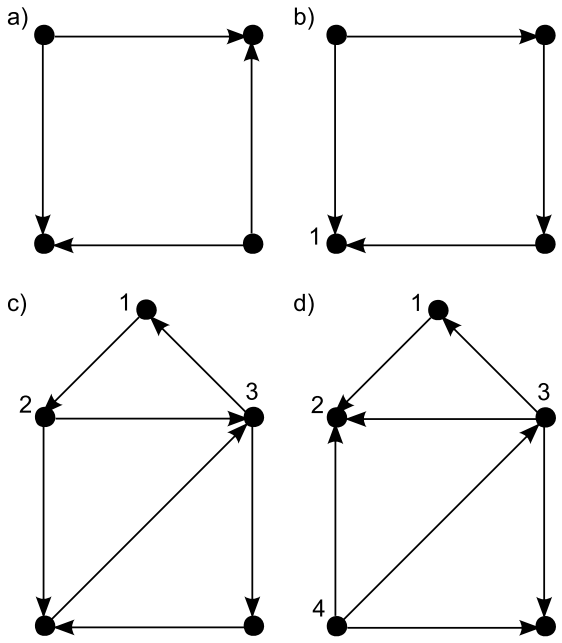

Figure 3.7: Four directed graphs/networks. Graph a) is not a transmission block whereas graphs b) and c) are transmission blocks. Graph d) contains a transmission block as a subgraph.

Proof. The Theorem follows from Corollary A.1.1 and Corollary A.1.2 (see Appendix 1). From Corollary A.1.1, the individuals contained in a subsystem larger than a pair appearing in M_E belong to some transmission block. From Corollary A.1.2, any transmission block appears as a subsystem in M_E. □

Definition 3.4.6. *Let $D = (V, A)$ be a directed graph. A vertex-induced subgraph $D[W]$, where $W \subset V$, is a 'k-motif' if and only if $|W| = k$ and there is at least one individual reachable from all others in $D[W]$.*

Remark. From the way in which the exact system M_E is built up, via equation (3.27) (see definition 3.4.3), the individuals of a subsystem configuration appearing in M_E form a single k-motif.

Theorem 3.4.4. *If a directed graph $D = (V, A)$ has n transmission blocks, $D[W_1], \ldots, D[W_n]$, then:*

$$\begin{aligned} \textit{dimensionality of } M_E &\leq 3|V| + 3|A| + \sum_{i=1}^{n} \sum_{k=3}^{|W_i|} 3^k \#\textit{distinct k-motifs in } D[W_i] \\ &\leq 3|V| + 3|A| + \sum_{i=1}^{n} \sum_{k=3}^{|W_i|} 3^k \binom{|W_i|}{k}. \end{aligned}$$

Proof. From the way in which the exact system M_E is built up, via equation (3.27) (see definition 3.4.3), it is immediate that there are $3N$ configurations for subsystems

of order 1 (individuals in states S, E and I), and at most $3|A|$ configurations for subsystems of order 2 (connected pairs in states SI, SE and SS). From Corollary A.1.1 (see Appendix 1), the individuals in a subsystem configuration larger than a pair are contained within some single transmission block and form a k-motif. Since there are three possible individual states, S, E, I, for individuals in the subsystem configurations of M_E, the theorem must hold. □

3.5 Hierarchies of approximate systems

The systems of equations in the previous section are exact, but limited in applicability because of the limited scope for dynamical partitioning in most networks. To suitably curtail the large number of equations, the networks need to have a structure which is roughly tree-like.

More typically, we want to trade off some exactness for systems which are numerically tractable and provide a good, rather than exact description of the underlying dynamics. The pair-level SEIR system (equation (3.18)) is exact for tree networks but is also a reasonably good approximation for Markovian SEIR dynamics on a wide range of networks. Higher-order systems will typically be more accurate, but will have considerably greater computational cost. Here we formally define hierarchies of approximate systems that can be applied to Markovian SEIR dynamics on any network.

We define 'pseudo-partitioning' according to different criteria. We define two hierarchies of systems via what we term 'cycle-partitioning' and 'size-partitioning'. We then also consider a 'hybrid-partitioning' hierarchy utilising both methods. Although these pseudo-partitionings can be defined more generally, as in the case of dynamical partitioning itself, we shall restrict our attention here to pseudo-partitioning with respect to single susceptible nodes.

Generalising from the case of dynamical partitioning, we define a function $f_p(X, Y, i)$ which acts on subsets of the network and which enables a systematic curtailing of the number of subsystem configurations necessary for a solvable system. We then use this function to determine some pseudo-partitioning of subsets X and Y with respect to

node i. By analogy with equation (3.27), we have:

$$\begin{aligned}
\langle \dot{\psi}^A_W \rangle \approx & \sum_{k\in W} \mathbb{1}(A_k = S)\left[- \sum_{n\in W} T_{kn}\mathbb{1}(A_n = I)\langle \psi^A_W\rangle\right] \\
& + \sum_{k\in W} \mathbb{1}(A_k = E)\left[\sum_{n\in W} T_{kn}\mathbb{1}(A_n = I)\langle h^S_k(\psi^A_W)\rangle - \alpha_k\langle \psi^A_W\rangle\right] \\
& + \sum_{k\in W} \mathbb{1}(A_k = I)\left[\alpha_k\langle h^E_k(\psi^A_W)\rangle - \gamma_k\langle \psi^A_W\rangle\right] \\
& + \sum_{k\in W} \mathbb{1}(A_k = R)\left[\gamma_k\langle h^I_k(\psi^A_W)\rangle\right] \\
& - \sum_{k\in W} \mathbb{1}(A_k = S) \sum_{n\in V\setminus W} T_{kn}\Bigg[\Big(1 - f_p(n, W\setminus k, k)\Big)\langle \psi^A_W I_n\rangle \\
& + f_p(n, W\setminus k, k)\frac{\langle \psi^A_W\rangle\langle S_k I_n\rangle}{\langle S_k\rangle}\Bigg] \\
& + \sum_{k\in W} \mathbb{1}(A_k = E) \sum_{n\in V\setminus W} T_{kn}\Bigg[\Big(1 - f_p(n, W\setminus k, k)\Big)\langle h^S_k(\psi^A_W) I_n\rangle \\
& + f_p(n, W\setminus k, k)\frac{\langle h^S_k(\psi^A_W)\rangle\langle S_k I_n\rangle}{\langle S_k\rangle}\Bigg]. \qquad (3.42)
\end{aligned}$$

So, when $f_p(X, Y, i) = 1$, we treat i as if it is dynamically partitioning with respect to X and Y and so the right-hand-side of the equation does not generate larger subsystems. The specific type of approximate system depends on how $f_p(X, Y, i)$ is defined and is formed by assuming equality between the left and right hand sides.

Note that equation (3.42) defines a solvable system that is based on the closure in corollary 3.4.1. However, other closures such as the Kirkwood-closure $\langle X_i Y_j Z_k\rangle \approx \langle X_i Y_j\rangle\langle Y_j Z_k\rangle\langle Z_k X_i\rangle/(\langle X_i\rangle\langle Y_j\rangle\langle Z_k\rangle)$, where $G[\{i, j, k\}]$ is a complete graph, fall outside of this scheme. It is, however, straightforward to define a solvable hierarchy of approximate systems that incorporates the standard Kirkwood closure as a special case.

By analogy with the closure in corollary 3.4.1, and the Kirkwood closure, we thus define the following closure for the expectation of the product of two indicator variables, ψ^A_W and I_n, where $n \notin W$:

$$\langle \psi^A_W I_n\rangle \approx \langle \psi^A_W\rangle\langle I_n\rangle \prod_{j\in W}\left[\frac{\langle I_n (A_j)_j\rangle}{\langle I_n\rangle\langle (A_j)_j\rangle}\right]^{G_{nj}}. \qquad (3.43)$$

For the case where $G[W \cup \{n\}]$ is a complete graph of three nodes, this is seen to reproduce the standard Kirkwood closure (recall that we use G to denote both the underlying graph and its adjacency matrix). The closure in equation (3.43) has the theoretical advantage of treating all of the connected pairs that n forms with the members of W in the same way. In the derivation of the previous closure (corollary 3.4.1),

this issue was non-existent since there was only one pair formed between n and a member of W, i.e. the pair it formed with $i \in W$ where i was dynamically partitioning relative to $W \setminus i$ and n. However, the disadvantage of this Kirkwood-based closure is that it is more likely to extend the dimensionality of the system by introducing new pair configurations, such as $\langle I_i I_j \rangle$, which wouldn't otherwise be needed for a closed system. We expect that, overall, the benefit of keeping the sizes of subsystems lower would outweigh this disadvantage. On the contrary, note that when using the original closure (corollary 3.4.1) to build up the system, the 'new' pair state that is introduced by the closure is always of the form $\langle S_i I_j \rangle$ where $(j,i) \in A(D)$ and such a pair configuration would be part of the system in any case.

With reference to equation (3.42), this Kirkwood-style closure can be utilised to generate a closed system via the time derivative:

$$\begin{aligned}
\dot{\langle \psi^A_W \rangle} \approx & \sum_{k\in W} \mathbb{1}(A_k = S)\left[-\sum_{n\in W} T_{kn}\mathbb{1}(A_n = I)\langle \psi^A_W \rangle \right] \\
& + \sum_{k\in W} \mathbb{1}(A_k = E)\left[\sum_{n\in W} T_{kn}\mathbb{1}(A_n = I)\langle h^S_k(\psi^A_W)\rangle - \alpha_k \langle \psi^A_W \rangle \right] \\
& + \sum_{k\in W} \mathbb{1}(A_k = I)\left[\alpha_k \langle h^E_k(\psi^A_W)\rangle - \gamma_k \langle \psi^A_W \rangle \right] \\
& + \sum_{k\in W} \mathbb{1}(A_k = R)\left[\gamma_k \langle h^I_k(\psi^A_W)\rangle \right] \\
& - \sum_{k\in W} \mathbb{1}(A_k = S) \sum_{n\in V\setminus W} T_{kn}\Bigg[\Big(1 - f_p(n, W\setminus k, k)\Big)\langle \psi^A_W I_n \rangle \\
& + f_p(n, W\setminus k, k)\langle \psi^A_W \rangle \langle I_n \rangle \prod_{j\in W}\left[\frac{\langle I_n (A_j)_j \rangle}{\langle I_n \rangle \langle (A_j)_j \rangle} \right]^{G_{nj}} \Bigg] \\
& + \sum_{k\in W} \mathbb{1}(A_k = E) \sum_{n\in V\setminus W} T_{kn}\Bigg[\Big(1 - f_p(n, W\setminus k, k)\Big)\langle h^S_k(\psi^A_W) I_n \rangle \\
& + f_p(n, W\setminus k, k)\frac{\langle h^S_k(\psi^A_W)\rangle \langle S_k I_n \rangle}{\langle S_k \rangle} \prod_{j\in W\setminus k}\left[\frac{\langle I_n (A_j)_j \rangle}{\langle I_n \rangle \langle (A_j)_j \rangle} \right]^{G_{nj}} \Bigg]. \qquad (3.44)
\end{aligned}$$

Either equation (3.42) or (3.44) can be used in conjunction with suitable definitions of $f_p(X, Y, i)$ to generate hierarchies of approximate systems. We shall mostly use equation (3.42) for explicit examples. However, for completeness, we shall briefly discuss equation (3.44) in a later section devoted to it.

It is worth noting that both of these closures are based on multiplying by a quotient of pair configurations and singlets. Other schemes with more complex closures should also be possible. For example, the exact closure theorem allows closures where we do not necessarily need to have only singlet states in the denominator (see Figure 3.4b).

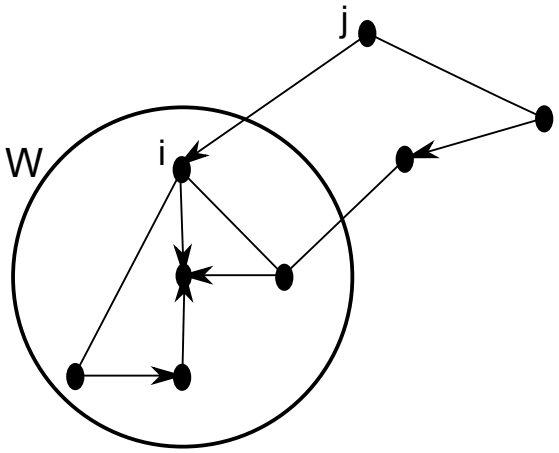

Figure 3.8: An example of a node $i \in W$ which is not dynamically partitioning with respect to node j and $W \setminus i$, but it is cycle-partitioning up to $x = 2$.

3.5.1 Cycle-partitioning

With reference to Figure 3.8, although node i is not dynamically partitioning with respect to $W \setminus i$ and node j, we might observe that it is in some sense 'approximately' dynamically partitioning because the path length between j and W is reasonably long when i is deleted. It seems sensible to define a type of pseudo-partitioning according to this path length, and this is the approch which we will adopt below. However, note that a more effective definition of pseudo-partitioning, at a perhaps higher computational cost, would be to also account for the 'strength' of the path between j and W. For example, although this path length is reasonably long when i is deleted, it may be that the rates of contacts between the interacting pairs in the path are so high that, assuming j becomes infected, W will also be infected with high probability.

Definition 3.5.1. *The set of individuals that can reach at least one member of $X \subset V$, by traversing $a \in \mathbb{N}$ arcs or less, is denoted $IN_a(X)$. Here, $\mathbb{N} = \{0, 1, 2, ...\}$.*

Definition 3.5.2. *Node $i \in V$ is 'cycle-partitioning' at order $x \in \mathbb{N}$ with respect to disjoint and non-empty subsets $X, Y \subset V$, where $i \notin X \cup Y$, if and only if we have $f_{C(x)}(X, Y, i) = 1$ where:*

$$f_{C(x)}(X, Y, i) = \begin{cases} 1 & \text{if } IN_a(X) \cap IN_b(Y) = \emptyset \quad \forall a, b : a + b = x \text{ (in } D(\text{cav } i)) \\ 0 & \text{otherwise,} \end{cases} \tag{3.45}$$

where $a, b \in \mathbb{N}$.

We make the following observations: i) If the network is undirected then $f_{C(x)}(X, Y, i) = 0$ if and only if there is at least one path of length x or less between some member of X and some member of Y when i is deleted. ii) An individual who is dynamically partitioning with respect to two subsets is also cycle-partitioning at all orders with

respect to those subsets. iii) In Figure 3.8, node i is cycle-partitioning with respect to $W \setminus i$ and j for $x = 0$, $x = 1$, and $x = 2$, but not $x > 2$. iv) Any individual $i \in V$ is always cycle-partitioning at order $x = 0$ with respect to any other two subsets.

Adapting Corollary 3.4.1 such that cycle-partitioning individuals of order $x \in \{0, 1, 2, \ldots\}$ are 'treated' as dynamically partitioning individuals, we substitute $f_p(n, W \setminus k, k) = f_{C(x)}(n, W \setminus k, k)$ into equation (3.42).

Remark. By using equation (3.42) to write down time derivatives for $\langle S_i \rangle$, $\langle E \rangle$ and $\langle I \rangle$, $\forall i \in V$, and for every new subsystem configuration which emerges, we obtain a closed set of differential equations which form the xth system in a hierarchy of approximating systems (note that the system corresponding to $x = 0$ is the pair-level system given by equation (3.18)). The associated system will be denoted by $M_{C(x)}$.

Examples

We can consider cycle-partitioning for the network in Figure 3.5b. If we cycle-partition at $x = 1$, then the first two terms of equation (3.34) are closed at the level of pairs. Specifically, for the first term, node 2 is cycle-partitioning with respect to nodes 1 and 3. For the second term, node 1 is cycle-partitioning with respect to nodes 2 and 4. This gives:

$$\begin{aligned} \langle \dot{S_1 E_2} \rangle \approx & \frac{\langle S_1 S_2 \rangle \langle S_2 I_3 \rangle}{\langle S_2 \rangle} - \frac{\langle S_1 E_2 \rangle \langle S_1 I_4 \rangle}{\langle S_1 \rangle} - \frac{\langle S_1 E_2 \rangle \langle S_1 I_5 \rangle}{\langle S_1 \rangle} - \frac{\langle S_1 E_2 \rangle \langle S_1 I_6 \rangle}{\langle S_1 \rangle} \\ & - \alpha \langle S_1 E_2 \rangle. \end{aligned} \tag{3.46}$$

Thus, triples within the square are no longer 'kept intact', and so, within the square, the model closes at the level of pairs. However, triples made up of the members of the triangle are kept intact. For example, we have:

$$\begin{aligned} \langle \dot{S_1 I_5} \rangle = & -\langle S_1 I_5 I_6 \rangle - (1 + \gamma) \langle S_1 I_5 \rangle + \alpha \langle S_1 E_5 \rangle \\ & - \frac{\langle S_1 I_5 \rangle \langle S_1 I_2 \rangle}{\langle S_1 \rangle} - \frac{\langle S_1 I_5 \rangle \langle S_1 I_4 \rangle}{\langle S_1 \rangle}. \end{aligned} \tag{3.47}$$

Figure 3.9 shows this hierarchy of systems. Here, the $x = 0$ system is the pair-level system (equation (3.18)). The $x = 1$ system is an improvement since it picks up the triangle. The $x = 2$ system picks up the square as well and is equivalent to the exact system M_E.

If we apply cycle-partitioning to Figure 3.5c instead, then the $x = 0$ system is the pair-level system as always. The $x = 1$ system is also the pair-level system and the $x = 2$ system is equivalent to the exact system M_E (equivalent to M in this case). Hence, cycle-partitioning does not necessarily lead to improved systems as x increases and it does not always lead to a reduction in system size with respect to the exact system M_E. The results from the $x = 0$ pair-level system and the exact system M_E

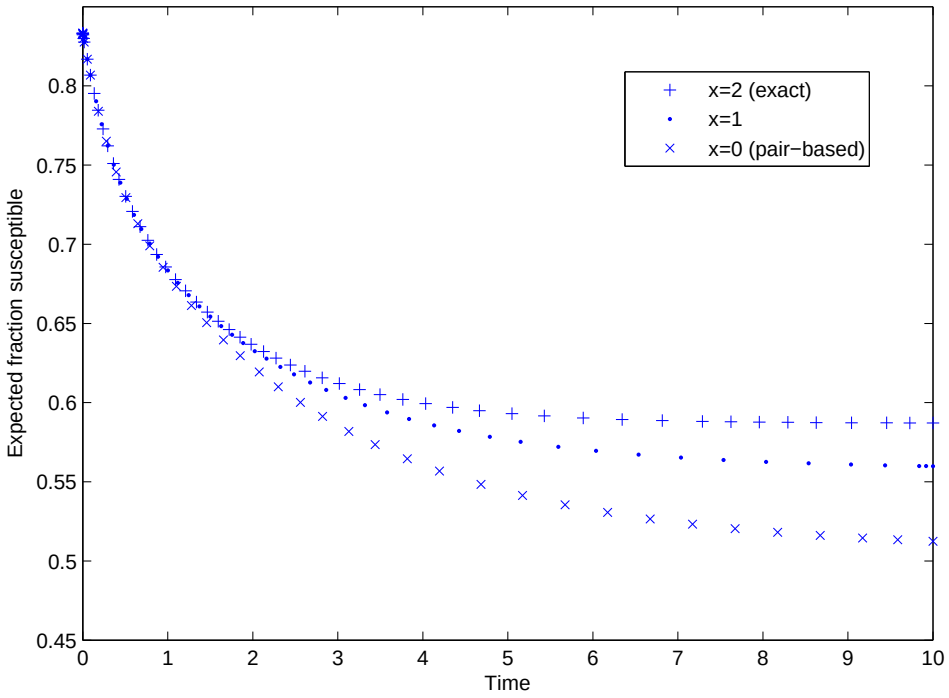

Figure 3.9: Cycle-partitioning applied to the scenario in Figure 3.5b with $x = 0$, which corresponds to the pair-level system, through $x = 1$ and finally $x = 2$ which is exact for this scenario. Here we assumed that all individuals were independently susceptible at time $t = 0$ with probability $5/6$ and infected otherwise (the states of individuals initially independent). We also assumed $\alpha = \gamma = 1$.

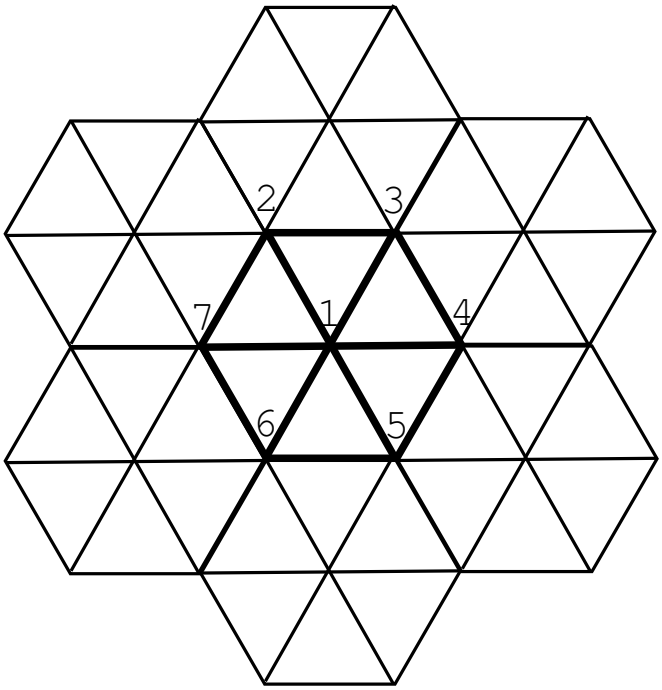

Figure 3.10: A triangle lattice - an extreme example where cycle-partitioning at order greater than $x = 0$ requires subsystem configurations which contain all individuals.

applied to Figure 3.5c can be seen in the section on size-partitioning below (Figure 3.11) and so are not reproduced here.

An extreme example of the failure of cycle-partitioning to produce a hierarchy of 'improving' systems is given by the triangular lattice shown in Figure 3.10. Here, the $x = 0$ system is the pair-level system. For $x = 1$, consider the triple $A_3S_1C_4$ ($A, C \in \{S, E, I\}$). Here we do not have cycle-partitioning since by deleting node 1, there is a path of length 1 between nodes 3 and 4. As we move to size 4 subsystems, (e.g. adding a node to the above triple either by the edge (1,2) or the edge (3,2)), it is readily seen that there will always exist subsystems which do not cycle-partition for $x = 1$, at all subsystem sizes. Hence for the triangle lattice, even for $x = 1$ cycle-partitioning, we obtain a system with subsystem configurations at the size of the full network. Some cycle-partitioning does occur however, so the resulting system is not exact. For example, for the triple $A_2S_1C_4$, deleting node 1 means that the shortest path from 2 to 4 is via node 3 and is of length 2. So we have cycle-partitioning here. We also have it for configuration $A_7S_1C_4$. This configuration is also cycle-partitioning at $x = 2$ (the path length from node 7 to node 4 after deletion of node 1 is 3) but we no longer cycle-partition $A_2S_1C_4$. Finally, at $x = 3$, no cycle-partitioning occurs anywhere and we have an exact system containing subsystem configurations containing all individuals ($M_{C(3)} = M_E = M$).

In general, if the largest transmission block in a network has n individuals, then any cycle-partitioning system of order $x \geq n - 2$ is exact (see Theorem A.2.1 in Appendix

2). This is illustrated by the network in Figure 3.5b where the largest transmission block is of size $n = 4$ and the $x = 2$ cycle-partitioning system is exact (Figure 3.9). This is also the case for the network in Figure 3.5c where $n = 5$ and the $x = 3$ system is exact (the $x = 2$ system also happens to be exact here as well). Another general result is that if the smallest directed sub-block consists of n individuals, then the cycle-partitioning systems of order $x < n - 2$ are all equivalent to the pair-level (x=0) system (see Theorem A.2.2 in Appendix 2). This is illustrated by the network in Figure 3.5c where the smallest directed sub-block is $n = 4$, and we found that the $x = 1$ cycle-partitioning system is the same as the pair-level system.

3.5.2 Size-partitioning

The issues arising in some networks such as Figure 3.5c, where even cycle-partitioning at $x = 2$ requires subsystem configurations containing all individuals, and the extreme example of the triangular lattice, motivate an alternative pseudo-partitioning approach whereby the sizes of subsystem configurations are more directly constrained.

Definition 3.5.3.

$$f_{S(x)}(X) = \begin{cases} 1 & \textit{if } |X| = x + 1 \\ 0 & \textit{otherwise}, \end{cases} \tag{3.48}$$

where $X \subset V$ *and* $x \in \mathbb{N}$.

Here we make the substitution $f_p(n, W \setminus k, k) = f_{S(x)}(W \setminus k)$ into equation (3.42).

Remark. As with previous pseudo-partitioning, a complete approximate system arises from the time derivatives for the individual-level probabilities and then repeatedly writing down the time derivatives for each subsystem configuration that emerges. As with cycle-partitioning, the $x = 0$ size-partitioning system corresponds to the pair-level system.

Examples

As an example, consider the $x = 1$ size-partitioning system for Figure 3.5c, where the cycle-partitioning hierarchy was redundant. Equation (3.39) now becomes:

$$\begin{aligned} \langle S_1 \dot{I}_2 I_4 \rangle &\approx \frac{\langle S_1 I_2 I_4 \rangle \langle S_1 I_5 \rangle}{\langle S_1 \rangle} - 2(1+\gamma)\langle S_1 I_2 I_4 \rangle \\ &\quad + \alpha \langle S_1 E_2 I_4 \rangle + \alpha \langle S_1 I_2 E_4 \rangle. \end{aligned}$$

For $x = 2$ size-partitioning, equation (3.39) is left untouched since the exact rate equation for a subsystem state of size 3 does not involve subsystem states larger than 4. However, equation (3.41) becomes:

$$\begin{aligned} \langle S_1 I_2 \dot{I}_4 E_5 \rangle &\approx -(2 + 2\gamma + \alpha)\langle S_1 I_2 I_4 E_5 \rangle + \frac{\langle S_1 I_2 I_4 S_5 \rangle \langle I_3 S_5 \rangle}{\langle S_5 \rangle} \\ &\quad + \alpha \langle S_1 E_2 I_4 E_5 \rangle + \alpha \langle S_1 I_2 E_4 E_5 \rangle. \end{aligned}$$

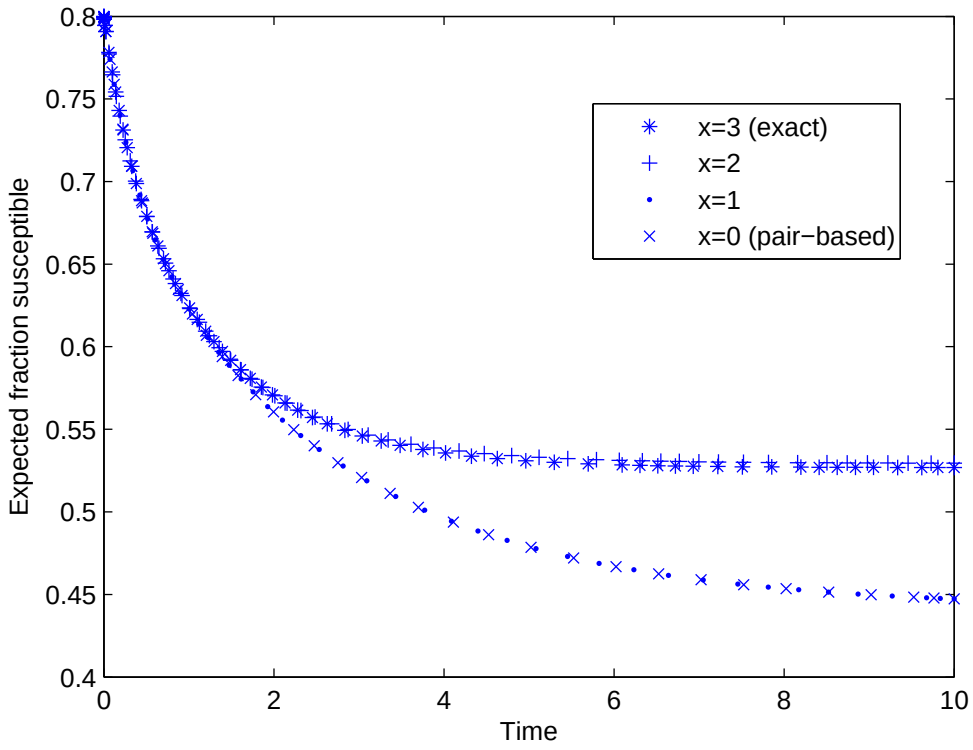

Figure 3.11: Size-partitioning applied to the scenario in Figure 3.5c with $x = 0$, which corresponds to the pair-level system, through $x = 1$, $x = 2$, and finally $x = 3$ which is exact for this scenario. Individuals were assumed to be initially susceptible with probability 4/5 and infected otherwise (the states of individuals initially independent). We also assumed $\alpha = \gamma = 1$.

In this way, we obtain three different approximate systems: $x = 0$, $x = 1$ and $x = 2$. For $x > 2$, the system is exact. Figure 3.11 shows results from the application of each of these three approximate systems and the exact $x = 3$ system to Markovian SEIR dynamics on the network depicted in Figure 3.5c. An interesting feature that should be noted for the $x = 2$ system is that it very slightly underestimates the rate of spread of the epidemic. Typically, experience shows that the closure of these equations leads to an over-estimate the rate of spread, but this provides a counter example.

While size-partitioning will resolve issues such as the triangular lattice, it has problems of its own. Specifically, we see from Figure 3.11 that since the smallest cycle size in Figure 3.5c is 4, the $x = 1$ size-partitioning system is almost identical to the $x = 0$ pair-level system. The $x = 3$ and $x = 4$ systems are also almost identical. Hence, the extra computation in evaluating at $x = 2$ and $x = 4$ is wasteful. In this sense, cycle-partitioning has an advantage by only picking up complete cycles in the network.

An additional problem with size-partitioning is that it ignores genuine dynamical partitioning. For example, for Figure 3.5b, we would require subsystem configurations

of size 6 ($x = 5$) to describe this exactly within the size-partitioning scheme. However, if we permit genuine dynamical partitioning, we only need subsystem configurations of size less than or equal to 4. This issue is readily resolved by considering the modified scheme:

$$f_{E,S(x)}(X, Y, i) = \begin{cases} 1 & \text{if } f_E(X, Y, i) = 1 \text{ or } f_{S(x)}(Y) = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{3.49}$$

which incorporates genuine dynamical partitioning into size-partitioning. With this rule, in Figure 3.5b, the genuine dynamical partitioning around node 1 is utilised wherever possible.

3.5.3 Hybrid-partitioning

Both cycle-partitioning and size-partitioning have their merits. Size-partitioning avoids unnecessarily large subsystem configurations where cycle-partitioning cannot be effectively implemented beyond an early stage, such as in the triangle lattice. On the other hand, cycle-partitioning picks out cycles in the network and closes at the pair level unless complete cycles can be incorporated, avoiding wasteful computation with minimal gain in accuracy.

We can construct a hybrid-partitioning scheme that captures the benefits of both cycle-partitioning and size-partitioning while avoiding the problems of both. We define this hybrid-partitioning as:

Definition 3.5.4.

$$f_{C(x)S(x)}(X, Y, i) = \begin{cases} 1 & \textit{if } f_{C(x)}(X, Y, i) = 1 \textit{ or } f_{S(x)}(Y) = 1 \\ 0 & \textit{otherwise} \end{cases}. \tag{3.50}$$

This leads to a hierarchy of systems defined by substituting $f_p(n, W \setminus k, k) = f_{C(x)S(x)}(n, W \setminus k, k)$ into equation (3.42). This also has the pair-level system for $x = 0$. We also note that alternative hierarchies could be designed with different values of x for the size-partitioning and the cycle-partitioning parts.

This definition of pseudo-partitioning benefits from the advantages of both cycle-partitioning and size-partitioning. Firstly, if there are only large cycles, the hierarchy is closed at a low order by cycle-partitioning. This is desirable since, as illustrated in Figure 3.11, continuing on generates little benefit unless we are able to continue to the size of the smallest cycle. However, if the system is not amenable to cycle-partitioning, as in the triangular lattice, then size-partitioning is required. A network illustrating the benefits of this is shown in Figure 3.12. For hybrid-partitioning with $x = 1$, let us start with the probability that node 1 is exposed:

$$\langle \dot{E}_1 \rangle = \langle S_1 I_2 \rangle + \langle S_1 I_5 \rangle + \langle S_1 I_6 \rangle + \langle S_1 I_7 \rangle - \alpha \langle E_1 \rangle. \tag{3.51}$$

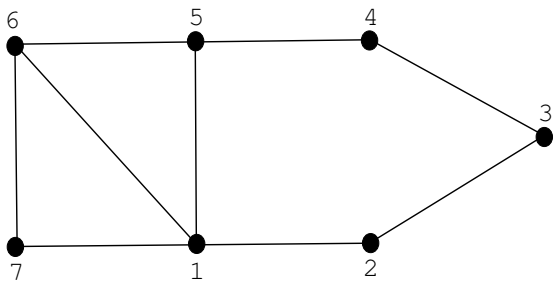

Figure 3.12: A network that illustrates the benefits of hybrid-partitioning. Expanding from node 1 with $x = 1$, we utilise both cycle-partitioning and size-partitioning, capturing the advantages of both.

Since we know that $\langle \dot{S_1 I_2} \rangle$ depends on $\langle S_1 E_2 \rangle$, we consider the differential equation for $\langle S_1 E_2 \rangle$:

$$\begin{aligned} \langle \dot{S_1 E_2} \rangle \approx & \frac{\langle S_1 S_2 \rangle \langle S_2 I_3 \rangle}{\langle S_2 \rangle} - \frac{\langle S_1 E_2 \rangle \langle S_1 I_5 \rangle}{\langle S_1 \rangle} - \frac{\langle S_1 E_2 \rangle \langle S_1 I_6 \rangle}{\langle S_1 \rangle} \\ & - \frac{\langle S_1 E_2 \rangle \langle S_1 I_7 \rangle}{\langle S_1 \rangle} - \alpha \langle S_1 E_2 \rangle, \end{aligned}$$

where we have employed $x = 1$ cycle-partitioning. Similarly, for $\langle S_1 E_5 \rangle$ in equation (3.51) we obtain:

$$\begin{aligned} \langle \dot{S_1 E_5} \rangle \approx & \frac{\langle S_1 S_5 \rangle \langle I_4 S_5 \rangle}{\langle S_5 \rangle} + \langle S_1 S_5 I_6 \rangle - \frac{\langle S_1 E_5 \rangle \langle S_1 I_2 \rangle}{\langle S_1 \rangle} \\ & - \langle S_1 E_5 I_6 \rangle - \frac{\langle S_1 E_5 \rangle \langle S_1 I_7 \rangle}{\langle S_1 \rangle} - \alpha \langle S_1 E_5 \rangle, \end{aligned}$$

where, again, $x = 1$ cycle-partitioning has been implemented where possible. Here, the time derivative of the second term must depend on $\langle S_1 S_5 E_6 \rangle$, for which we have:

$$\begin{aligned} \langle \dot{S_1 S_5 E_6} \rangle \approx & \frac{\langle S_1 S_5 S_6 \rangle \langle S_6 I_7 \rangle}{\langle S_6 \rangle} - \frac{\langle S_1 S_5 E_6 \rangle \langle S_1 I_7 \rangle}{\langle S_1 \rangle} - \alpha \langle S_1 S_5 E_6 \rangle \\ & - \frac{\langle S_1 S_5 E_6 \rangle \langle I_4 S_5 \rangle}{\langle S_5 \rangle} - \frac{\langle S_1 S_5 E_6 \rangle \langle S_1 I_2 \rangle}{\langle S_1 \rangle}. \end{aligned}$$

The closures on the first line are via $x = 1$ size-partitioning, whereas the closures on the second line are via meeting the criteria for both $x = 1$ size-partitioning and $x = 1$ cycle-partitioning.

So, this hybrid-partitioning obtains the best of both methodologies. Cycle-partitioning avoids unnecessarily including extra terms in the large cycle 1-2-3-4-5-1 which we have seen (Figure 3.11) generates minimal extra accuracy. Size-partitioning forces partitioning where the subsystem configurations get beyond a specified size, here constraining the maximum subsystem size to be 3.

3.5.4 Alternative closure

Before leaving this section on pseudo-partitioning, we include a brief aside on pseudo-partitioning using the alternative Kirkwood-style closure defined in equation (3.43). In

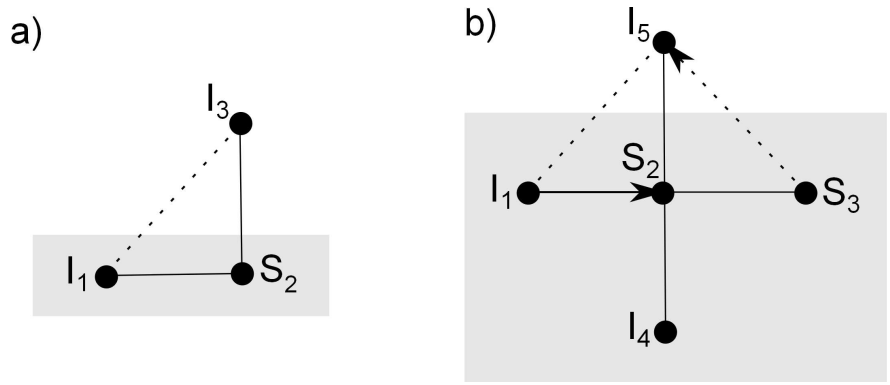

Figure 3.13: Two simple examples of applying the alternative Kirkwood-style closure as in equation (3.43). The shaded region specifies the given subsystem configuration, and there is a new IS link towards it in accordance with the way in which the induced state spaces are built up. The dashed lines represent additional links between the new node and the original subsystem (these would be ignored by the closure rule in equation (3.42).)

this case, we can still apply the cycle, size and hybrid methods, but we use equation (3.44) in place of equation (3.42). Two examples of applying this are illustrated in Figure 3.13. Here the shaded regions represent the existing subsystem states and the solid lines coming out of these regions represents the new infectious node being added on. The dashed lines represent other links between the new infectious nodes and the original subsystems. Supposing that the criteria for pseudo-partitioning is met at this stage (i.e. the relevant $f_p(.) = 1$), for Figure 3.13a we obtain

$$\langle I_1 S_2 I_3 \rangle \approx \frac{\langle I_1 S_2 \rangle \langle S_2 I_3 \rangle \langle I_1 I_3 \rangle}{\langle I_1 \rangle \langle S_2 \rangle \langle I_3 \rangle},$$

and for Figure 3.13b, we obtain

$$\langle I_1 S_2 S_3 I_4 I_5 \rangle \approx \frac{\langle I_1 S_2 S_3 I_4 \rangle \langle I_1 I_5 \rangle \langle S_2 I_5 \rangle \langle S_3 I_5 \rangle}{\langle I_1 \rangle \langle S_2 \rangle \langle S_3 \rangle \langle I_5 \rangle^2}.$$

We note that for cycle-partitioning with $x > 0$, the two closure methods result in the same system (building up the sytem using equation (3.44) becomes equivalent to building up the system via equation (3.42)) since the types of additional links drawn in Figure 3.13 could not then be present when the partitioning occurs.

Notice that when the closure of triples always occurs (e.g. $x = 0$ cycle-partitioning or $x = 0$ size-partitioning), the variant of the pair-level system introduced by Sharkey (2008, 2011) is obtained under this Kirkwood-style closure. This variant is expected to be able to handle networks with triangles more accurately than the variant considered by Sharkey et al. (2013) and Kiss et al. (2014) that follows from equation (3.42).

3.6 Discussion

Recently it has been possible to establish exact and practicable representations of stochastic epidemic dynamics on finite tree networks (Sharkey et al., 2013) using closure methodologies evaluated at the level of individuals (Sharkey, 2008, 2011). Message-passing also gives exact representations on trees (Karrer and Newman, 2010) and this can be shown, under some circumstances with Poisson transmission processes, to be equivalent to moment closure systems (see Wilkinson and Sharkey (2014) and chapter 4 of this thesis). Under suitable and very restrictive homogeneity assumptions, exact population-level versions of these closed systems (e.g. Keeling, 1999) can also be derived for idealised graphs with homogeneous initial conditions (Sharkey, 2008).

Within the individual-level closure construction, it is possible to go beyond trees and obtain exact representations of epidemic dynamics on non-tree networks using the idea of dynamical partitioning (referred to by Kiss et al. (2014), in the context of undirected networks, as partitioning). Here we extended the concept of dynamical partitioning to arbitrary directed networks, and then used this to define an exact system using our exact closure theorem. The extent to which these systems are computationally viable depends primarily on the underlying structure of the network.

More specifically, starting from the probabilities of the states of individual nodes in a given network, we uniquely defined the full set of exact induced moment equations by automatically implementing the exact closure theorem where applicable. We also defined transmission blocks as a natural decomposition of a network for the closure of the SEIR system. Transmission blocks represent a possible extension of the block concept in graph theory into directed networks. Using this concept, we proved a theorem stating that the size of the largest subsystem appearing in the set of exact moment-induced equations is equal to the size of the largest transmission block.

We also investigated hierarchies of approximate moment closure systems. In the epidemic literature, it is normally the case that moment closure models are constructed at the level of pairs, or occasionally at triples or quads (Matsuda et al., 1992; Bauch, 2005; House et al., 2009). This is often accompanied with an assertion that higher order models exist. However, to our knowledge, these higher order epidemic models have never been defined explicitly. This is understandable since these models rapidly become too complex to be of real practical relevance, but it does leave open the theoretical question of how these models can be defined (Sharkey, 2011). To address this, we introduced 'pseudo-partitioning' to construct complete hierarchies of approximate closed systems that are well-defined at all orders. In fact, we defined several hierarchies of closed systems; one in terms of subsystem size, one in terms of the size of cycles in the network, and a hybrid method taking the best of both of the previous methods. Undoubtedly other hierarchies can be defined as well. In addition, we investigated two

mechanisms of closure - one based on exact dynamical partitioning and the other which is more related to the Kirkwood closure.

The closure based directly around dynamical partitioning has the pair-based system considered by Sharkey et al. (2013) as its zeroth order variant (for all of the size, cycle and hybrid approaches). The hierarchies based around the Kirkwood-type closure all have the pair-based system introduced by Sharkey (2008, 2011) as their zeroth order variant (this is designed to handle networks with triangles in a more effective way). We also observed that the conditions for cycle-partitioning at orders greater than zero mean that both methods of closure become equivalent.

The hierarchies of systems generated some interesting observations concerning the convergence to exactness with order. For example, for size-partitioning, the systems converge to the exact solution with increasing order, but this convergence is not always monotonic (see Figure 3.11). In all of our previous numerical work, we had found that moment closure models over-exaggerate the spread of an epidemic, but here we observed a counter example (see also Sharkey (2011), where possible explanations of overestimation/underestimation are discussed). An unanswered question is whether the approximate systems always increase in accuracy as the order of the hierarchy increases. Intuitively we would expect that they do, and this is validated by the examples so far investigated.

Chapter 4

The message passing approach to general epidemics on networks

4.1 Introduction

In this chapter we will consider a generalised version of the Markovian network-based S(E)IR model, as considered by Karrer and Newman (2010). The generalised model will assume the same underlying dynamics but the individual level processes will no longer be resticted to Poisson processes. This is an important generalisation since, as discussed in their paper, real-world infectious periods are not well described by the exponential distribution, and the effect of moving to a non-exponential infectious period can greatly affect the evolution of the epidemic. For this model, Karrer and Newman (2010) developed a system of equations, in the language of 'message passing' algorithms, which exactly captures the probability of a given individual being in a given state at time t (for all individuals) provided that the network is a tree network and the states of individuals are statistically independent at $t = 0$. They also proved that, in the case of non-tree networks and the same restriction on initial conditions, their equations provide a rigorous lower bound on $\langle S_i \rangle \forall i \in V$.

We will further generalise the model considered by Karrer and Newman (2010), allowing more heterogeneity in the individual level processes and initial conditions. Following their approach, we will then derive a system of 'message passing' equations which can give exact results in the case of tree networks, and a lower bound on $\langle S_i \rangle \forall i \in V$ otherwise. We will then show that generalised pair-based moment closure systems (allowing arbitrary infectious/exposed periods) can be derived from these message passing equations, indicating that the pair-based systems can give exact results for tree networks and a rigorous lower bound on $\langle S_i \rangle \forall i \in V$ for non-tree networks (the pair based system defined in theorem 3.3.1 of chapter 3, for the Markovian network-based SEIR model, is seen to be a special case).

At the end of the chapter we will discuss the scenario where the states of individuals are not statistically independent at $t = 0$. In particular, we will consider the relation-

ship between the case where every individual is initially susceptible with probability z (independently), and initially infected otherwise, and the case where zN initially susceptible individuals are chosen uniformly at random (without replacement), and the rest are initially infected (assuming zN is an integer). Indeed, a single individual chosen uniformly at random to initiate the epidemic is more common in the literature than the assumption of initial independence between individuals (see, for example, Ball et al. (2010) and Meyers et al. (2006)). In the limit of large population size, this is the most interesting case since it (generally) gives positive probability to a very small outbreak (non-invasion), allowing stochastic fluctuations near the start of the process to be crucial for the epidemic outcome (as, presumably, in the real world).

4.2 The model

Following Karrer and Newman (2010), while generalising to allow more heterogeneity (Wilkinson and Sharkey, 2014), the individual level processes are specified as follows (for SIR, SEIR will be addressed later): the probability that individual $j \in V$ will make at least one *infectious* contact to individual $i \in V$ within time period t of having first entered the infected state is given by $\int_0^t f_{ij}(\tau)d\tau$ where $f_{ij} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is defined such that $0 \leq \int_0^\infty f_{ij}(\tau)d\tau \leq 1$. The time that it takes for individual i to enter the recovered/removed state, measured from the moment it enters the infected state, is described by a probability density function $r_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. Note that r_j has implications for f_{ij} since for an infectious contact to be made from j to i it must be made while j is still in the infected state. For consistency we can relate the two functions as follows:

$$f_{ij}(\tau) = h_{ij}(\tau) \int_\tau^\infty r_j(\tau')\mathrm{d}\tau',$$

where $h_{ij} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a probability density function for the time it takes for j to make its first contact to i after j having first entered the infected state (this contact may be non-infectious). Note that it is only the first infectious contact from one individual to another which may be relevant to the dynamics since as soon as a susceptible individual receives an infectious contact it becomes infected, and from that point on further infectious contacts have no effect (this is, in general, the case for all S(E)IR models).

For initial conditions we will assume that at $t = 0$ the states of individuals are statistically independent and that the probability that individual i is susceptible, infected or removed at $t = 0$ is given by z_i, y_i and x_i respectively ($z_i + y_i + x_i = 1$). We will equate an individual being in the removed state at $t = 0$ with that individual having been vaccinated against the disease.

The network/digraph $D = (V, A)$ on which the process takes place is implied by the functions $f_{ij}(\tau)$ since we assume that there is an arc from j to i iff $\int_0^\infty f_{ij}(\tau)d\tau > 0$.

Similarly, the network/digraph restricts these functions such that $\int_0^\infty f_{ij}(\tau)d\tau > 0$ iff there is an arc from j to i.

Given that an individual has received an infectious contact, or is *not* susceptible at $t = 0$, then its 'fate' is decoupled from the behaviour and states of all other individuals. This makes it straightforward to write down equations for $\langle X_i \rangle$, where $X \in \{I, R\}$, as follows:

$$\begin{aligned} \langle I_i \rangle &= 1 - \langle S_i \rangle - \langle R_i \rangle, \\ \langle R_i \rangle &= x_i + \int_0^t r_i(\tau)\Big[1 - x_i - \langle S_i(t-\tau)\rangle\Big]\mathrm{d}\tau. \end{aligned} \tag{4.1}$$

Thus, it is possible to compute these quantities once we have $\langle S_i \rangle$. As we shall see, the message passing formalism provides a way forwards.

4.3 Exact systems for tree networks

In this section we shall restrict our attention to the case where the underlying (undirected) graph of the network is a tree or forest. The fundamental quantity in the message passing formalism for tree networks is $H^{i \leftarrow j}(t)$ (also defined for non-tree networks), where $(j, i) \in A(D)$, which is the probability that i has not received an infectious contact from j by time t given that i is in the cavity state, i.e. in $D(\text{cav } i)$. Note that placing an individual in the cavity state does not affect its fate since its ability to contact others only comes into play if it is infected. We can now write (Karrer and Newman, 2010):

$$\langle S_i \rangle = z_i \prod_{j \in N_i} H^{i \leftarrow j}(t). \tag{4.2}$$

This follows from the fact that when i is in the cavity state, and the underlying graph is a tree or forest, the in-components of each of i's neighbours are mutually disjoint. Therefore, if the states of all individuals are statistically independent at $t = 0$ then the states of i's neighbours are statistically independent at all times.

Given its definition, the message passing equation can be expressed (Karrer and Newman, 2010; Wilkinson and Sharkey, 2014):

$$H^{i \leftarrow j}(t) = 1 - \int_0^t f_{ij}(\tau)\Big[y_j + z_j - z_j \Phi_i^j(t-\tau)\Big]\mathrm{d}\tau, \tag{4.3}$$

where $\Phi_i^j(t)$ is the probability that j has not received an infectious contact by time t given that i and j are both in the cavity state, i.e. in $D(\text{cav } \{i, j\})$. The above message passing equation can be understood by considering that $z_j(1 - \Phi_i^j(t-\tau))f_{ij}(\tau)\Delta$ is the probability that j is initially susceptible and receives its first infectious contact *before* time $t - \tau$, at time t_1 say, and then makes its first infectious contact to i between time $t_1 + \tau$ and time $t_1 + \tau + \Delta$, for $\Delta \to 0^+$, given that i is in the cavity state. For the tree

networks considered in this section we can substitute in equation (4.3):

$$\Phi_i^j(t) = \prod_{k \in N_j \setminus i} H^{j \leftarrow k}(t). \tag{4.4}$$

If $N_j \setminus i = \emptyset$ then we define the right-hand-side of equation (4.4) to be equal to 1. Here, $\Phi_i^j(t)$ can be expressed as a product of probabilities of independent events because of the simple structure of tree networks, i.e. there is no more than one path from any individual to any other individual and no cycles. This is discussed in more detail in the subsequent section on non-tree networks.

To understand how to numerically 'solve' such equations, let us consider the case where the network is undirected and $j \in V$ has only one neighbour i (j is a leaf). We can write:

$$H^{i \leftarrow j}(t) = 1 - \int_0^t y_j f_{ij}(\tau) \mathrm{d}\tau, \tag{4.5}$$

which is directly solvable since the function and constant in the integrand are known a priori. Next, if j is the only individual that can directly infect i, then

$$H^{k \leftarrow i}(t) = 1 - \int_0^t f_{ij}(\tau) \Big[y_j + z_j - z_j H^{i \leftarrow j}(t - \tau) \Big] \mathrm{d}\tau, \tag{4.6}$$

and this too is now directly solvable. In other words, by working 'inwards' from the leaves, one can compute $H^{i \leftarrow j}(t) \forall (i,j) : i \in V, j \in N_i$. The (marginal) probability distribution for the state of a given individual at time t can then be obtained from equations (4.1) and (4.2). For a directed tree network, we work 'inwards' from the set of arcs $\{(i,j) \in A(D) : (k,i) \notin A(D), \forall k \in V \setminus j\}$.

By setting $z_i, y_i \in \{0,1\}\ \forall i \in V$, we can consider any pure initial system state (note that we assume initially infected individuals 'become' infected at $t = 0$). In this case, we could reduce the number of our equations by removing from the network those individuals that are vaccinated. However, we can also consider mixed (probabilistic) initial system states - as long as the states of individuals are initially statistically independent. For instance, we might consider the case where every individual is independently vaccinated with probability x.

We obtain the specific form in the Karrer and Newman (2010) paper by setting $z_i = z, x_i = 0\ \forall i$. The solution of this system represents a measure of an 'average epidemic' but we note that the initial distribution in the total number of infecteds is binomial and allows the event of no initial infecteds. Typically, as previously discussed, we are more interested in the expected outcome when a single initial infected individual is seeded uniformly at random in a susceptible population, and this will be addressed in section 4.6.

To start to link the message passing method with the pair-based systems, we express some relevant probabilities for connected pairs.

4.3.1 Exact message passing equations for pair-states

We can express the probabilities of some configurations for connected pairs in terms of the message passing formalism as follows:

$$\begin{aligned}\langle S_i S_j\rangle &= z_i z_j \Phi_i^j(t)\Phi_j^i(t),\\ \langle S_i I_j\rangle &= z_i \Phi_j^i(t)\left[- z_j \int_0^t g_{ij}(\tau)\dot{\Phi}_i^j(t-\tau)\mathrm{d}\tau + y_j g_{ij}(t)\right],\end{aligned} \tag{4.7}$$

where

$$g_{ij}(t) = \int_t^\infty h_{ij}(\tau)\mathrm{d}\tau \int_t^\infty r_j(\tau)\mathrm{d}\tau \tag{4.8}$$

is the probability that j does not recover by time t and does not make an infectious contact to i by time t, where time is measured from the moment j enters the infected state. The integral in equation (4.7) can be understood by observing that $-z_j g_{ij}(\tau)\dot{\Phi}_i^j(t-\tau)\Delta$ is the probability that j is initially susceptible and receives its first infectious contact between time $t-\tau$ and time $t-\tau+\Delta$, for $\Delta \to 0^+$, and then does not recover, or make an infectious contact to i, within time period τ, given that i is in the cavity state.

4.3.2 The case of Poisson contact processes; novel pair-based systems emerge

In this section we will consider the case where individuals contact their neighbours via independent Poisson processes, i.e.

$$\begin{aligned} h_{ij}(\tau) &= T_{ij} e^{-T_{ij}\tau},\\ f_{ij}(\tau) &= T_{ij} e^{-T_{ij}\tau}\int_\tau^\infty r_j(\tau')\mathrm{d}\tau'\\ &= T_{ij}\int_\tau^\infty h_{ij}(\tau')\mathrm{d}\tau' \int_\tau^\infty r_j(\tau')\mathrm{d}\tau'\\ &= T_{ij} g_{ij}(\tau),\end{aligned} \tag{4.9}$$

where T_{ij} is the 'rate' at which j makes infectious contacts to i when j is in the infected state.

Let us now consider the time derivative of the message passing equation (4.3) for this special case (using Leibniz's integral rule):

$$\begin{aligned} \dot{H^{i\leftarrow j}}(t) &= z_j \int_0^t f_{ij}(\tau)\dot{\Phi}_i^j(t-\tau)\mathrm{d}\tau - y_j f_{ij}(t)\\ &= z_j T_{ij}\int_0^t g_{ij}(\tau)\dot{\Phi}_i^j(t-\tau)\mathrm{d}\tau - y_j T_{ij} g_{ij}(t)\\ &= -T_{ij}\frac{\langle S_i I_j\rangle}{z_i \Phi_j^i(t)}\\ &= -T_{ij} H^{i\leftarrow j}(t)\frac{\langle S_i I_j\rangle}{\langle S_i\rangle} \qquad \langle S_i\rangle \neq 0,\end{aligned} \tag{4.10}$$

and we can also write:

$$\begin{aligned}
\dot{\Phi}^i_j(t) &= \sum_{k\in N_i\setminus j} H^{i\leftarrow\dot{k}}(t) \prod_{l\in N_i\setminus\{j,k\}} H^{i\leftarrow l}(t) \\
&= -\sum_{k\in N_i\setminus j} T_{ik} H^{i\leftarrow k}(t) \frac{\langle S_i I_k\rangle}{\langle S_i\rangle} \prod_{l\in N_i\setminus\{j,k\}} H^{i\leftarrow l}(t) \\
&= \frac{-1}{z_i H^{i\leftarrow j}(t)} \sum_{k\in N_i\setminus j} T_{ik}\langle S_i I_k\rangle \\
&= -\sum_{k\in N_i\setminus j} T_{ik} \frac{\langle S_i I_k\rangle}{\langle S_i\rangle} \Phi^i_j(t) \qquad \langle S_i\rangle \neq 0. \qquad (4.11)
\end{aligned}$$

We are now in a position to transform the message passing system into a pair-based system (for the present case of Poisson contact processes); using equations (4.10) and (4.11) to find the time derivatives of the probabilities expressed in equations (4.2) and (4.7):

$$\langle \dot{S}_i\rangle = -\sum_{j\in N_i} T_{ij}\langle S_i I_j\rangle, \qquad (4.12)$$

$$\begin{aligned}
\langle \dot{S_i I_j}\rangle &= -\sum_{k\in N_i\setminus j} T_{ik} \frac{\langle I_k S_i\rangle\langle S_i I_j\rangle}{\langle S_i\rangle} + \sum_{k\in N_j\setminus i} T_{jk} \frac{\langle S_i S_j\rangle\langle S_j I_k\rangle}{\langle S_j\rangle} - T_{ij}\langle S_i I_j\rangle \\
&\quad - \int_0^t r_j(t-\tau) e^{-T_{ij}(t-\tau)} \sum_{k\in N_j\setminus i} T_{jk} \frac{\langle S_i S_j(\tau)\rangle\langle S_j I_k(\tau)\rangle}{\langle S_j(\tau)\rangle} \frac{\Phi^i_j(t)}{\Phi^i_j(\tau)} \mathrm{d}\tau \\
&\quad - z_i y_j \Phi^i_j(t) e^{-T_{ij}t} r_j(t), \qquad (4.13)
\end{aligned}$$

$$\langle \dot{S_i S_j}\rangle = -\sum_{k\in N_j\setminus i} T_{jk} \frac{\langle S_i S_j\rangle\langle S_j I_k\rangle}{\langle S_j\rangle} - \sum_{k\in N_i\setminus j} T_{ik} \frac{\langle I_k S_i\rangle\langle S_i S_j\rangle}{\langle S_i\rangle}, \qquad (4.14)$$

where $\langle X_i Y_j(\tau)\rangle$ is the probability that i is in state X and j is in state Y at time τ. Above, equations (4.11) to (4.14) form a pair-based (closed) system of integro-differential equations, the solution of which (in principle) allows us to find the (marginal) probability distribution for the state of any given individual at any given time via equation (4.1). To remove any trace of the message-passing formalism from this new pair-based system we can re-express the variable $\Phi^i_j(t)$ as:

$$\Phi^i_j(t) = \exp\left(-\int_0^t \sum_{k\in N_i\setminus j} T_{ik} \frac{\langle S_i I_k(\tau)\rangle}{\langle S_i(\tau)\rangle} \mathrm{d}\tau \right). \qquad (4.15)$$

We recognise this pair-based system as a generalisation (to arbitrary infectious periods) of the pair-based moment-closure equations (discussed in section 3.3 and Sharkey et al. (2013)) which assume the following approximation for the probability of the state of a connected open triple: $\langle A_i B_j C_k\rangle = \langle A_i B_j\rangle\langle B_j C_k\rangle / \langle B_j\rangle$, where $\langle B_j\rangle \neq 0$. We know that this approximation is exact when $B = S$ (susceptible) for the case of tree networks

(see section 3.3). It is clear that the number of equations in this new pair-based system is of the order of the number of arcs in the network, i.e. $|A(D)|$.

The message passing system, and the pair-based system derived from it (for Poisson contact processes), are exact for any tree network. More specifically, they are exact for any directed graph where there is no more than one simple directed path from any individual to any other individual and no directed cycles. The reason for this will be made clear in section 4.4. Certain initial conditions may also ensure that the systems are exact on some non-tree networks, as discussed by Sharkey et al. (2013) in relation to the pair-based equations for Markovian network-based SIR dynamics.

The message passing system and the pair-based system, for Poisson contact processes, are equivalent in the sense that they produce the same time series for the probabilities of the states of individuals. Next, we will consider both systems for exponential and fixed (non-random) infectious periods, where the equations become simpler.

Exponential infectious periods

For the case of exponential infectious periods we have $r_i(\tau) = \gamma_i e^{-\gamma_i \tau}$, where γ_i is the rate at which i recovers when infected, and so:

$$\begin{aligned} f_{ij}(\tau) &= T_{ij} e^{-(T_{ij}+\gamma_j)\tau}, \\ \dot{f_{ij}}(\tau) &= -(T_{ij}+\gamma_j) f_{ij}(\tau), \\ g_{ij}(\tau) &= e^{-(T_{ij}+\gamma_j)\tau}. \end{aligned} \tag{4.16}$$

The message passing system is now conveniently constructed as a system of ODEs:

$$\begin{aligned} \dot{H^{i \leftarrow j}}(t) &= -\int_0^t \dot{f_{ij}}(t-\tau)\Big[y_j + z_j - z_j \Phi_i^j(\tau)\Big]\mathrm{d}\tau - f_{ij}(0)\Big[y_j + z_j - z_j \Phi_i^j(t)\Big] \\ &= (T_{ij}+\gamma_j)(1 - H^{i \leftarrow j}) - T_{ij}(y_j + z_j - z_j \Phi_i^j(t)), \\ \langle S_i \rangle &= z_i \prod_{j \in N_i} H^{i \leftarrow j}(t), \\ \langle I_i \rangle &= 1 - \langle S_i \rangle - \langle R_i \rangle, \\ \langle \dot{R_i} \rangle &= \gamma_i \langle I_i \rangle. \end{aligned} \tag{4.17}$$

The corresponding pair-based system is derived by substituting the exponential infectious period into the equation for $\langle \dot{S_i I_j} \rangle$ (equation (4.13)). In particular, for the

last two terms of equation (4.13), we can write:

$$\begin{aligned}
& -\int_0^t r_j(t-\tau)e^{-T_{ij}(t-\tau)} \sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j(\tau)\rangle\langle S_jI_k(\tau)\rangle}{\langle S_j(\tau)\rangle}\frac{\Phi_j^i(t)}{\Phi_j^i(\tau)}\mathrm{d}\tau \\
& -z_iy_j\Phi_j^i(t)e^{-T_{ij}t}r_j(t) \\
= \; & -\gamma_j\Phi_j^i(t)\int_0^t e^{-(T_{ij}+\gamma_j)(t-\tau)}\frac{\langle S_iS_j(\tau)\rangle\dot{\Phi_i^j}(\tau)}{\Phi_i^j(\tau)\Phi_j^i(\tau)}\mathrm{d}\tau - \gamma_jz_iy_j\Phi_j^i(t)e^{-(T_{ij}+\gamma_j)t} \\
= \; & -\gamma_jz_iz_j\Phi_j^i(t)\int_0^t g_{ij}(t-\tau)\dot{\Phi_i^j}(\tau)\mathrm{d}\tau - \gamma_jz_iy_j\Phi_j^i(t)g_{ij}(t) \\
= \; & -\gamma_j\langle S_iI_j\rangle, \qquad (4.18)
\end{aligned}$$

and so the closed pair-based system of ODEs is:

$$\begin{aligned}
\langle\dot{S_i}\rangle &= -\sum_{j\in N_i} T_{ij}\langle S_iI_j\rangle, \\
\langle\dot{S_iI_j}\rangle &= -\sum_{k\in N_i\setminus j} T_{ik}\frac{\langle I_kS_i\rangle\langle S_iI_j\rangle}{\langle S_i\rangle} + \sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j\rangle\langle S_jI_k\rangle}{\langle S_j\rangle} \\
&\quad -T_{ij}\langle S_iI_j\rangle - \gamma_j\langle S_iI_j\rangle, \\
\langle\dot{S_iS_j}\rangle &= -\sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j\rangle\langle S_jI_k\rangle}{\langle S_j\rangle} - \sum_{k\in N_i\setminus j} T_{ik}\frac{\langle I_kS_i\rangle\langle S_iS_j\rangle}{\langle S_i\rangle}, \\
\langle\dot{R_i}\rangle &= \gamma_i\langle I_i\rangle, \\
\langle I_i\rangle &= 1-\langle S_i\rangle-\langle R_i\rangle. \qquad (4.19)
\end{aligned}$$

Fixed (non-random) infectious periods

For fixed infectious periods (or fixed time to recovery) we have $r_i(\tau)=\delta(\tau-\omega_i)$ where $\delta(\tau)$ is the Dirac delta function and ω_i is the time it takes i to recover once it has been infected. We now have:

$$\begin{aligned}
f_{ij}(\tau) &= T_{ij}e^{-T_{ij}\tau}\theta(\omega_j-\tau), \\
\dot{f_{ij}}(\tau) &= -T_{ij}f_{ij}(\tau) - T_{ij}e^{-T_{ij}\tau}\delta(\tau-\omega_j), \\
g_{ij}(\tau) &= e^{-T_{ij}\tau}\theta(\omega_j-\tau), \qquad (4.20)
\end{aligned}$$

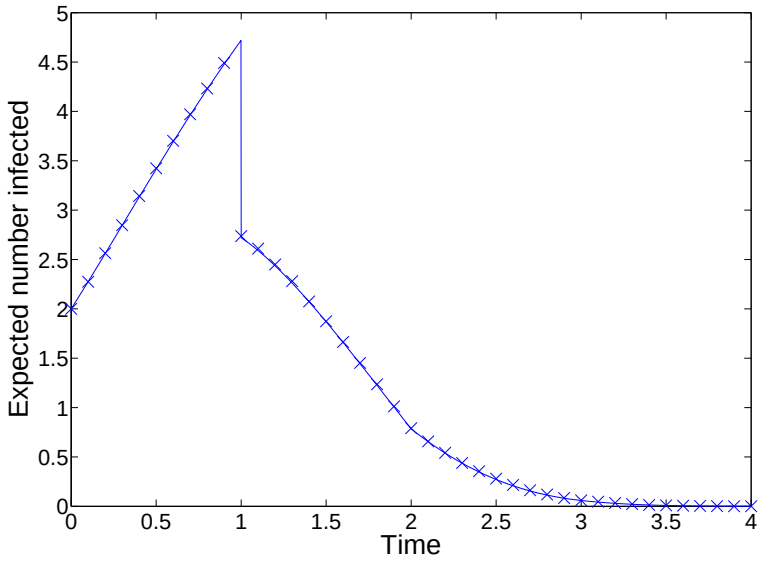

Figure 4.1: Here we investigate SIR dynamics (Poisson contact, fixed recovery) on an undirected tree network of 10 individuals. We set the infectious period to unity for all individuals and $T_{ij} = 1 \ \forall i, j \in V : j \in N_i$. Two non-adjacent index-individuals were selected to be initial infecteds, whilst the rest were vaccinated with probability $1/10$ but susceptible otherwise (at $t = 0$). The line represents the output from our representation (equation (4.21)) whilst the crosses indicate corresponding numerical results from 10,000 full stochastic simulations.

where $\theta(t)$ is the Heaviside step function. The message passing system is now constructed as a system of delay differential equations (DDEs):

$$\begin{aligned}
H^{i\dot{\leftarrow}j}(t) &= -\int_0^t \dot{f}_{ij}(t-\tau)\Big[y_j + z_j - z_j\Phi_i^j(\tau)\Big]\mathrm{d}\tau - f_{ij}(0)\Big[y_j + z_j - z_j\Phi_i^j(t)\Big] \\
&= T_{ij}(1 - H^{i\leftarrow j}(t)) + T_{ij}e^{-T_{ij}\omega_j}\theta(t-\omega_j)\Big[y_j + z_j - z_j\Phi_i^j(t-\omega_j)\Big] \\
&\quad -T_{ij}\Big[y_j + z_j - z_j\Phi_i^j(t)\Big], \\
\langle S_i\rangle &= z_i\prod_{j\in N_i} H^{i\leftarrow j}(t), \\
\langle R_i\rangle &= x_i + \theta(t-\omega_i)\Big[1 - x_i - \langle S_i(t-\omega_i)\rangle\Big], \\
\langle I_i\rangle &= 1 - \langle S_i\rangle - \langle R_i\rangle.
\end{aligned} \tag{4.21}$$

For a numerical example, see figure 4.1.

Again, a corresponding pair-based system is derived by substituting the fixed infectious period into the equation for $\langle S_i\dot{I}_j\rangle$ (equation (4.13)). In particular, for the last two terms of equation (4.13), we can write:

$$\begin{aligned}
&-\int_0^t r_j(t-\tau)e^{-T_{ij}(t-\tau)}\sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j(\tau)\rangle\langle S_jI_k(\tau)\rangle}{\langle S_j(\tau)\rangle}\frac{\Phi_j^i(t)}{\Phi_j^i(\tau)}\mathrm{d}\tau \\
&-z_iy_j\Phi_j^i(t)e^{-T_{ij}t}r_j(t) \\
=\ &-\theta(t-\omega_j)e^{-T_{ij}\omega_j}\sum_{k\in N_j\setminus i}\frac{\langle S_iS_j(t-\omega_j)\rangle\langle S_jI_k(t-\omega_j)\rangle}{\langle S_j(t-\omega_j)\rangle}\frac{\Phi_j^i(t)}{\Phi_j^i(t-\omega_j)} \\
&-z_iy_j\Phi_j^i(t)e^{-T_{ij}t}\delta(t-\omega_j),
\end{aligned} \tag{4.22}$$

and so the closed pair-based system of DDEs is:

$$\begin{aligned}
\langle\dot{S}_i\rangle &= -\sum_{j\in N_i} T_{ij}\langle S_iI_j\rangle, \\
\langle S_i\dot{I}_j\rangle &= -\sum_{k\in N_i\setminus j} T_{ik}\frac{\langle I_kS_i\rangle\langle S_iI_j\rangle}{\langle S_i\rangle} + \sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j\rangle\langle S_jI_k\rangle}{\langle S_j\rangle} - T_{ij}\langle S_iI_j\rangle \\
&\quad -\theta(t-\omega_j)e^{-T_{ij}\omega_j}\sum_{k\in N_j\setminus i}\frac{\langle S_iS_j(t-\omega_j)\rangle\langle S_jI_k(t-\omega_j)\rangle}{\langle S_j(t-\omega_j)\rangle}\frac{\Phi_j^i(t)}{\Phi_j^i(t-\omega_j)} \\
&\quad -z_iy_j\Phi_j^i(t)e^{-T_{ij}t}\delta(t-\omega_j), \\
\langle S_i\dot{S}_j\rangle &= -\sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j\rangle\langle S_jI_k\rangle}{\langle S_j\rangle} - \sum_{k\in N_i\setminus j} T_{ik}\frac{\langle I_kS_i\rangle\langle S_iS_j\rangle}{\langle S_i\rangle}, \\
\langle R_i\rangle &= x_i + \theta(t-\omega_i)\Big[1 - x_i - \langle S_i(t-\omega_i)\rangle\Big], \\
\langle I_i\rangle &= 1 - \langle S_i\rangle - \langle R_i\rangle.
\end{aligned} \tag{4.23}$$

4.4 Approximate systems and a rigorous bound for non-tree networks

Karrer and Newman (2010) proved that their message passing formalism, when applied to non-tree networks, provides a rigorous lower bound on $\langle S_i \rangle \ \forall i \in V$. Here we repeat their analysis in order to confirm that this bound is still obtained in our slightly more general setting, and to show that the derived pair-based systems consequently provide the same bound on $\langle S_i \rangle \ \forall i \in V$.

Following Karrer and Newman (2010), we represent general SIR dynamics on an arbitrary digraph $D = (V, A)$ by randomly weighting/removing the arcs as follows: 1) assign an infectious period τ_i to every individual $i \in V$, sampling from density functions r_i. 2) weight every arc $(j, i) \in A$ with a contact time ω_{ij}, sampling from density functions h_{ij}. 3) for every arc $(j, i) \in A$, if its weighting $\omega_{ij} > \tau_j$ then completely remove it from the digraph. 4) for every individual $i \in V$, with probability x_i, completely remove every arc emanating from i.

The resulting weighted digraph is denoted D'. $n_{iB}(D')$, where $B \subset N_i$, denotes the set of individuals from which i can be reached by a simple directed path of total weighting less than t, such that a member of B is the penultimate individual, given that i is in the cavity state (all arcs emanating from i are removed).

Let $i \leftarrow B$, where $B \subset N_i$, denote the event that i (in the cavity state) does not receive any infectious contacts from any of the members of B by time t. Let $|N_i| = M$ and let us label each of these neighbours as $N_i^{(1)}, N_i^{(2)}, \ldots, N_i^{(m)}, \ldots, N_i^{(M)}$ where the ordering is arbitrary. We can now express $\langle S_i \rangle$ as a product of conditional probabilities:

$$\begin{aligned}
\langle S_i \rangle &= z_i P(i \leftarrow \cup_{p=1}^{M} N_i^{(p)}) \\
&= z_i P(i \leftarrow N_i^{(1)}) \\
&\quad \times P(i \leftarrow N_i^{(2)} \mid i \leftarrow N_i^{(1)}) \\
&\quad \times \ldots \times P(i \leftarrow N_i^{(m)} \mid i \leftarrow \cup_{p=1}^{m-1} N_i^{(p)}) \\
&\quad \times \ldots \times P(i \leftarrow N_i^{(M)} \mid i \leftarrow \cup_{p=1}^{M-1} N_i^{(p)}).
\end{aligned} \tag{4.24}$$

The particular way in which D' is constructed means that, for any $j \in N_i$, we have:

$$P(i \leftarrow j) = E\left[\prod_{k \in n_{ij}} \frac{z_k}{1 - x_k}\right], \tag{4.25}$$

where the expectation operator is here applied to a function of the random variable $n_{ij}(D')$, and this function is assumed to be equal to 1 when $n_{ij} = \emptyset$. Equation (4.25) follows from the fact that all members of $n_{ij}(D')$ must be initially susceptible if D' is to represent the event that i (in the cavity state) does not receive an infectious contact from j by time t. $z_k/(1 - x_k)$ is the probability that k is initially susceptible given that it is not vaccinated (we excluded the possibility of a member of $n_{ij}(D')$ being

vaccinated in step 4 of the construction of D'). Similarly, for any $B \subset N_i : j \notin B$, we can write:

$$P(i \leftarrow j \mid i \leftarrow B) \approx E\left[\prod_{k \in n_{ij} \setminus n_{iB}} \frac{z_k}{1 - x_k}\right]. \tag{4.26}$$

Now, assuming that approximation (4.26) is actually exact (which it is for tree networks), and since $n_{ij} \setminus n_{iB} \subset n_{ij}$, with set equality occurring for tree networks, we have:

$$P(i \leftarrow j) \leq P(i \leftarrow j \mid i \leftarrow B), \tag{4.27}$$

with equality occurring for tree networks. In fact, $n_{ij}(D')$ and $n_{iB}(D')$ are necessarily disjoint sets if there is no more than one path from any individual to any other individual in D. A rigorous proof of inequality (4.27) relies on the fact that $\mathbb{1}_{i \leftarrow j}$ and $\mathbb{1}_{i \leftarrow B}$ can both be expressed as increasing functions of the same set of 'associated' (as defined by Esary et al. (1967)) random variables.

Inequality (4.27) implies that the conditioning in each term of the product in equation (4.24) can only serve to increase the total probability. Therefore

$$\langle S_i \rangle \geq z_i \prod_{j \in N_i} P(i \leftarrow j) = z_i \prod_{j \in N_i} H^{i \leftarrow j}(t). \tag{4.28}$$

Inequality (4.27) also implies that

$$\Phi_i^j(t) \geq \prod_{\substack{k \in N_j \\ k \neq i}} P(j \leftarrow k \mid i \text{ in cavity}), \tag{4.29}$$

where we have ignored $P(j \leftarrow i \mid i \text{ in cavity})$ since it is necessarily equal to 1 ($\Phi_i^j(t)$ is the probability that j has not received an infectious contact by time t given that i and j are both in the cavity state, where $(j, i) \in A(D)$ - see equation (4.3)). Now, taking i out of the cavity state, we only increase (or leave the same) the probability of infectious contact across any arc, and so

$$\prod_{\substack{k \in N_j \\ k \neq i}} P(j \leftarrow k \mid i \text{ in cavity}) \geq \prod_{\substack{k \in N_j \\ k \neq i}} H^{j \leftarrow k}(t), \tag{4.30}$$

with equality occurring for tree networks. Notice that this, in conjunction with equality in 4.27, implies equation (4.4) for tree networks. However, we also get equality in 4.30 whenever there are no cycles in the digraph D. Therefore, sufficient requirements for the exactness of these systems are: 1) there is no more than one path from any individual to any other individual in D and 2) there are no cycles in D. Equations (4.3), (4.29) and (4.30) imply that

$$H^{i \leftarrow j}(t) \geq 1 - \int_0^t f_{ij}(\tau)\Big[y_j + z_j - z_j \prod_{\substack{k \in N_j \\ k \neq i}} H^{j \leftarrow k}(t - \tau)\Big] \mathrm{d}\tau. \tag{4.31}$$

Following Karrer and Newman (2010), we define the function:

$$F^{i\leftarrow j}(t) = 1 - \int_0^t f_{ij}(\tau)\Big[y_j + z_j - z_j \prod_{\substack{k\in N_j \\ k\neq i}} F^{j\leftarrow k}(t-\tau)\Big]\mathrm{d}\tau, \tag{4.32}$$

and note that it corresponds to the way in which $H^{i\leftarrow j}(t)$ can be expressed for tree networks (equations ()4.3) and (4.4)).

Let $F_0^{i\leftarrow j}(t) = H^{i\leftarrow j}(t) \in [0,1]$ $(\forall i,j : j \in N_i)$, and define an iterative process (as in Karrer and Newman (2010)):

$$F_{n+1}^{i\leftarrow j}(t) = 1 - \int_0^t f_{ij}(\tau)\Big[y_j + z_j - z_j \prod_{\substack{k\in N_j \\ k\neq i}} F_n^{j\leftarrow k}(t-\tau)\Big]\mathrm{d}\tau. \tag{4.33}$$

From equation (4.31) we have:

$$F_1^{i\leftarrow j}(t) \leq F_0^{i\leftarrow j}(t), \tag{4.34}$$

and since this is true in all cases, we have:

$$F_1^{i\leftarrow j}(t) \geq 1 - \int_0^t f_{ij}(\tau)\Big[y_j + z_j - z_j \prod_{\substack{k\in N_j \\ k\neq i}} F_1^{j\leftarrow k}(t-\tau)\Big]\mathrm{d}\tau, \tag{4.35}$$

and so in general

$$F_{n+1}^{i\leftarrow j}(t) \leq F_n^{i\leftarrow j}(t). \tag{4.36}$$

Since $F_n^{i\leftarrow j}(t)$ is bounded below by $1 - \int_0^t f_{ij}(\tau)\mathrm{d}\tau$ (Karrer and Newman, 2010), this iterative procedure must converge from above such that

$$F_n^{i\leftarrow j}(t) \to F^{i\leftarrow j}(t) \leq H^{i\leftarrow j}(t) \qquad \text{as } n\to\infty, \tag{4.37}$$

and this allows us to write:

$$\langle S_i\rangle \;\; \geq \;\; z_i \prod_{j\in N_i} H^{i\leftarrow j}(t) \geq z_i \prod_{j\in N_i} F^{i\leftarrow j}(t). \tag{4.38}$$

For Poisson contact processes, the (approximate) dynamics can be cast as systems of differential equations in both formalisms (all occurrences of H and S, I and R, in equations (4.1) to (4.14), are changed respectively to F and X, Y and Z - indicating inexactness). Since they are implied by the message passing formalism, the solution of the pair-based systems on non-tree networks, i.e. arbitrary digraphs, provide a rigorous lower bound on $\langle S_i\rangle$ and approximations for $\langle I_i\rangle, \langle R_i\rangle$ $\forall i \in V$.

Figure 4.2 illustrates the application of the message passing approach to SIR dynamics with Poisson contact processes and fixed recovery processes on a non-tree network (we use equations (4.4) and (4.21), changing all occurrences of H and S, I and R, to F and X, Y and Z, respectively).

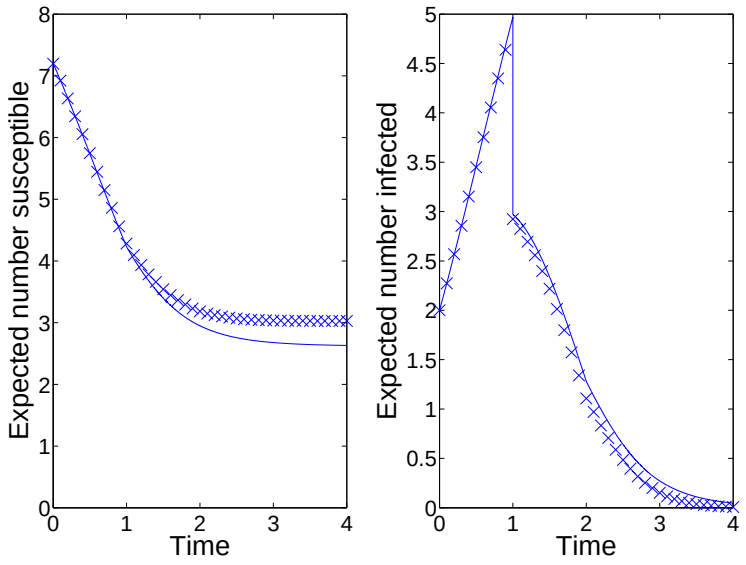

Figure 4.2: The same scenario as in figure 4.1 except that two extra undirected connections, i.e. four arcs, have been added to the network/digraph, creating multiple cycles. The lines represent the output from our representation (equations (4.21)) whilst the crosses indicate corresponding numerical results from 10,000 full stochastic simulations.

4.5 Generalising to SEIR dynamics

For SEIR dynamics, the time that it takes individual $i \in V$ to move from the exposed state into the infected state is described by a probability density function $q_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. We will also make use of a function $f^*_{ij} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ where $\int_0^t f^*_{ij}(\tau)d\tau$ is the probability that $j \in V$ will make at least one *infectious* contact to individual $i \in V$ within time period t of having first entered the exposed state. As such, it must satisfy:

$$f^*_{ij}(\tau) = \int_0^\tau q_j(\tau')f_{ij}(\tau - \tau')\mathrm{d}\tau'. \tag{4.39}$$

In all other respects, the notation that we use for SEIR dynamics can be understood from its previous usage for SIR dynamics, and the derivation of equations can be understood from the corresponding derivations for the SIR case. For ease of exposition and little loss in generality we will assume that no individuals start in the exposed state.

Similarly to the SIR version, it is straightforward to write down equations for the probability of an individual being in a given state, excluding susceptible, in terms of the probabilities of it being in the other states (this is because, given an individual is not susceptible, its fate does not then depend on the states or behaviour of other

individuals):

$$\begin{aligned}
\langle E_i\rangle &= z_i - \langle S_i\rangle - \int_0^t q_i(\tau)\Big[z_i - \langle S_i(t-\tau)\rangle\Big]\mathrm{d}\tau,\\
\langle I_i\rangle &= y_i + z_i - \langle S_i\rangle - \langle E_i\rangle\\
&\quad - \int_0^t r_i(\tau)\Big[y_i + z_i - \langle S_i(t-\tau)\rangle - \langle E_i(t-\tau)\rangle\Big]\mathrm{d}\tau,\\
\langle R_i\rangle &= 1 - \langle S_i\rangle - \langle E_i\rangle - \langle I_i\rangle.
\end{aligned} \tag{4.40}$$

The message passing equation for SEIR dynamics is:

$$H^{i\leftarrow j}(t) = 1 - \int_0^t f_{ij}^*(\tau)\Big[z_j - z_j\Phi_i^j(t-\tau)\Big] + y_j f_{ij}(\tau) \quad \mathrm{d}\tau, \tag{4.41}$$

where for tree networks, we can again substitute:

$$\Phi_i^j(t) = \prod_{k\in N_j\setminus i} H^{j\leftarrow k}(t) \tag{4.42}$$

and, also for tree networks:

$$\langle S_i\rangle = z_i \prod_{j\in N_i} H^{i\leftarrow j}(t). \tag{4.43}$$

The equations for the states of connected pairs in terms of the message passing formalism are:

$$\begin{aligned}
\langle S_iS_j\rangle &= z_iz_j\Phi_i^j(t)\Phi_j^i(t),\\
\langle S_iE_j\rangle &= z_i\Phi_j^i(t)\left[- z_j\int_0^t\int_\tau^\infty q_j(\tau')\dot{\Phi}_i^j(t-\tau)\mathrm{d}\tau'\mathrm{d}\tau\right],\\
\langle S_iI_j\rangle &= z_i\Phi_j^i(t)\left[- z_j\int_0^t\int_0^\tau q_j(\tau')g_{ij}(\tau-\tau')\dot{\Phi}_i^j(t-\tau)\mathrm{d}\tau'\mathrm{d}\tau + y_jg_{ij}(t)\right].
\end{aligned} \tag{4.44}$$

If the individual level contact processes are Poisson, we can derive the time derivative of $H^{i\leftarrow j}(t)$ as follows:

$$\begin{aligned}
\dot{H^{i\leftarrow j}}(t) &= z_j\int_0^t f_{ij}^*(\tau)\dot{\Phi}_i^j(t-\tau)\mathrm{d}\tau - y_jf_{ij}(t)\\
&= z_j\int_0^t\int_0^\tau\int_{\tau-\tau'}^\infty q_j(\tau')T_{ij}e^{-T_{ij}(\tau-\tau')}r_j(\tau'')\dot{\Phi}_i^j(t-\tau)\mathrm{d}\tau''\mathrm{d}\tau'\mathrm{d}\tau\\
&\quad -y_jT_{ij}e^{-T_{ij}t}\int_t^\infty r(\tau)\mathrm{d}\tau\\
&= z_jT_{ij}\int_0^t\int_0^\tau\int_{\tau-\tau'}^\infty\int_{\tau-\tau'}^\infty q_j(\tau')h_{ij}(\tau''')r_j(\tau'')\dot{\Phi}_i^j(t-\tau)\mathrm{d}\tau'''\mathrm{d}\tau''\mathrm{d}\tau'\mathrm{d}\tau\\
&\quad -y_jT_{ij}\int_t^\infty h_{ij}(\tau)\mathrm{d}\tau\int_t^\infty r_j(\tau)\mathrm{d}\tau\\
&= z_jT_{ij}\int_0^t\int_0^\tau q_j(\tau')g_{ij}(\tau-\tau')\dot{\Phi}_i^j(t-\tau)\mathrm{d}\tau'\mathrm{d}\tau\\
&\quad -y_jT_{ij}g_{ij}(t)\\
&= -T_{ij}H^{i\leftarrow j}(t)\frac{\langle S_iI_j\rangle}{\langle S_i\rangle} \qquad \langle S_i\rangle \neq 0,
\end{aligned} \tag{4.45}$$

and this allows us to write:

$$\dot{\Phi}^i_j(t) = -\sum_{k\in N_i\setminus j} T_{ik}\frac{\langle S_iI_k\rangle}{\langle S_i\rangle}\Phi^i_j(t) \qquad \langle S_i\rangle \neq 0.$$

The general pair-based system is:

$$\begin{aligned}
\langle\dot{S_i}\rangle &= -\sum_{j\in N_i} T_{ij}\langle S_iI_j\rangle,\\
\langle\dot{S_iI_j}\rangle &= -\sum_{k\in N_i\setminus j} T_{ik}\frac{\langle I_kS_i\rangle\langle S_iI_j\rangle}{\langle S_i\rangle} - T_{ij}\langle S_iI_j\rangle\\
&\quad -\int_0^t\int_0^{t-\tau} e^{-T_{ij}\tau'}r_j(\tau')q_j(t-\tau-\tau')\sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j(\tau)\rangle\langle S_jI_k(\tau)\rangle}{\langle S_j(\tau)\rangle}\frac{\Phi^i_j(t)}{\Phi^i_j(\tau)}\mathrm{d}\tau'\mathrm{d}\tau\\
&\quad +\int_0^t q_j(t-\tau)\sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j(\tau)\rangle\langle S_jI_k(\tau)\rangle}{\langle S_j(\tau)\rangle}\frac{\Phi^i_j(t)}{\Phi^i_j(\tau)}\mathrm{d}\tau\\
&\quad -z_iy_j\Phi^i_j(t)e^{-T_{ij}t}r_j(t),\\
\langle\dot{S_iS_j}\rangle &= -\sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j\rangle\langle S_jI_k\rangle}{\langle S_j\rangle} - \sum_{k\in N_i\setminus j} T_{ik}\frac{\langle I_kS_i\rangle\langle S_iS_j\rangle}{\langle S_i\rangle},\\
\langle\dot{S_iE_j}\rangle &= \sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j\rangle\langle S_jI_k\rangle}{\langle S_j\rangle} - \sum_{k\in N_i\setminus j} T_{ik}\frac{\langle I_kS_i\rangle\langle S_iE_j\rangle}{\langle S_i\rangle}\\
&\quad -\int_0^t q_j(t-\tau)\sum_{k\in N_j\setminus i} T_{jk}\frac{\langle S_iS_j(\tau)\rangle\langle S_jI_k(\tau)\rangle}{\langle S_j(\tau)\rangle}\frac{\Phi^i_j(t)}{\Phi^i_j(\tau)}\mathrm{d}\tau,
\end{aligned} \tag{4.46}$$

where

$$\Phi^i_j(t) = \exp\left(-\int_0^t\sum_{k\in N_i\setminus j} T_{ik}\frac{\langle S_iI_k(\tau)\rangle}{\langle S_i(\tau)\rangle}\mathrm{d}\tau\right). \tag{4.47}$$

If the exposed and infectious periods are all exponential, then this system reduces to the familiar pair-based system for the Markovian network-based SEIR model (see section 3.3).

It is straightforward that, in the case of non-tree networks, the message passing equations again provide a lower bound on $\langle S_i\rangle\forall i$. This can be shown in the same way as for SIR dynamics, since it is again possible to represent the model as a random weighted directed graph (the probability of the infection passing across a given arc, given that it arrives at the tail, is known a priori (see the next section (4.6) for elucidation)). Therefore, the pair-based system for the Markovian network-based SEIR model must also provide the same bound(s).

4.6 A note on initial conditions

For general network-based S(E)IR dynamics on tree networks, the expected outcome when a single initial infected is seeded uniformly at random, and the rest are susceptible, can be computed via the following methods: 1) Solve the message passing (or

pair-based) system N times, with each individual in turn as the single initial infected, and then average. This would be an exact but relatively time-consuming approach 2) Increase z towards 1 such that the ratio between the probability of there being one initial infected to the probability of there being more than one becomes large. We are then left with a sum of two terms, one corresponding to zero initial infecteds (contributing nothing to the time series) and the other to a single initial infected seeded uniformly at random. Thus, dividing the resulting time series (expected number infected) by the probability of having at least one initial infected, i.e. $1 - z^N$, approximates the desired result (for the expected number susceptible, we must first subtract Nz^N, and then divide by $1 - z^N$). We have achieved considerable success with this second approach in our numerical computations (see figure 4.3), but have found that the same method, in the case of non-tree networks, produces nonsensical results (stemming from the fact that the equations for non-tree networks are inexact to begin with).

In the previous sections of this chapter, we have assumed that the states of individuals are statistically independent at $t = 0$. However, when a specific number of initial infecteds are chosen at random, this assumption is violated. We will thus compare two families of initial conditions for general network-based S(E)IR dynamics: (1) each individual is initially susceptible with probability z and initially infected otherwise, and the states of individuals are statistically independent at $t = 0$. (2) a specific number $I_0 = (1 - z)N$ are chosen uniformly at random (without replacement) to be the initial infecteds and the remaining $S_0 = zN$ are initially susceptible, and so the states of individuals are not independent at $t = 0$. Notice that the expect number of individuals in each compartment at $t = 0$ is the same in both cases, and we have assumed that z is chosen such that I_0 and S_0 are positive integers less than N.

To proceed, we will need to represent the dynamics, and couple together the processes resulting from all the different possible initial configurations, as a random weighted directed graph, as in section 4.4. To do this, we define the random graph in exactly the same way except that we also assign an exposed period τ_i' to every individual $i \in V$, sampling from density functions q_i. We then define the total weighting of a path as the sum of all the weightings of the associated arcs *and* all of the exposed periods of the associated individuals, except the first and last. For this graph, n_i denotes the set of individuals from which $i \in V$ can be reached by some path of total weighting less than t.

For initial conditions (1) we can now write:

$$\langle S_i \rangle^{(1)} = \mathbb{E}\left[z^X\right], \tag{4.48}$$

where $X = |n_i| + 1$. This follows from the fact that the probability that i is susceptible at time t is the same as the probability that there is no initially infected individual in $n_i \cup \{i\}$, and each individual in this set is not initially infected, i.e. susceptible, with

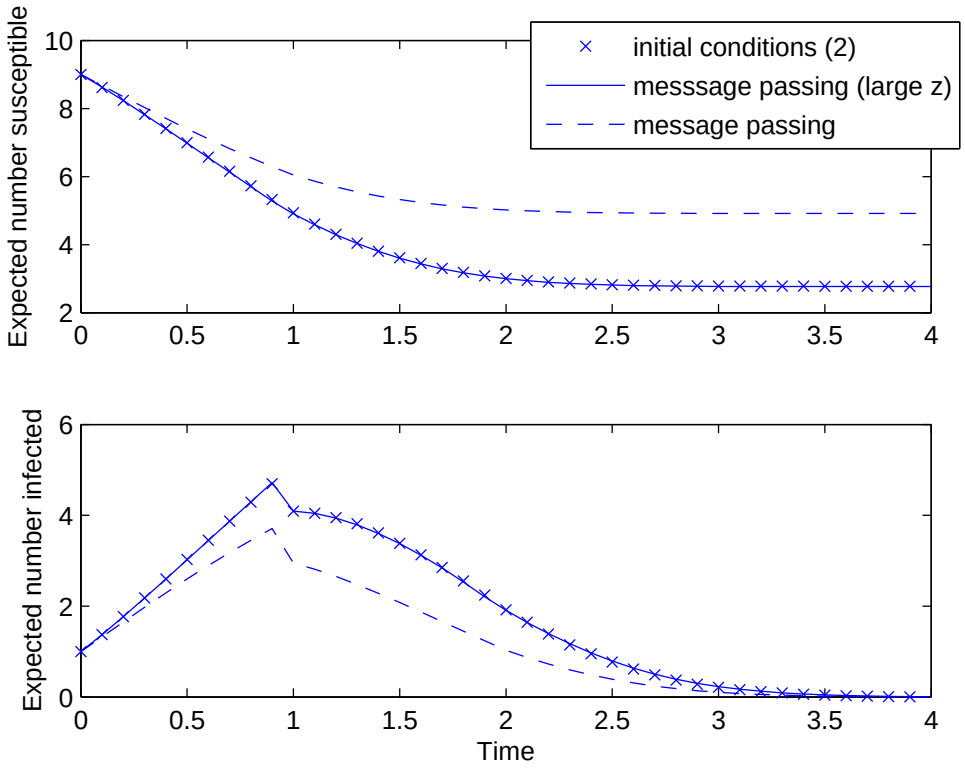

Figure 4.3: Here, we are investigating SIR dynamics (Poisson contact, fixed recovery) on an undirected tree network of 10 individuals. We set the infectious period to unity for all individuals and $T_{ij} \in \{0, 2\}$. The crosses represent the average of 10,000 stochastic simulations where a single initial infected was chosen uniformly at random, and the rest were initially susceptible. The dashed line represents the output from the corresponding message passing system with $z_i = z = 9/10$, $y_i = y = 1/10$ and $x_i = x = 0$. The solid line is the output from the message passing system with $z_i = z = 1 - 10^{-10}$, $y_i = y = 10^{-10}$ and $x_i = x = 0$ and then compensating by conditioning on their being at least one initial infected (see text).

probability z (independently). For initial conditions (2) we have:

$$\begin{aligned}
\langle S_i \rangle^{(2)} &= \mathbb{E}\left[\mathbb{1}_{X \leq S_0} \prod_{j=0}^{I_0-1} \left(1 - \frac{X}{N-j}\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{X \leq S_0} \frac{(N-X)!(zN)!}{(zN-X)!N!}\right] \\
&= \mathbb{E}\left[\mathbb{1}_{X \leq S_0} \frac{S_0(S_0-1)(S_0-2)\ldots(S_0-X+1)}{N(N-1)(N-2)\ldots(N-X+1)}\right] \\
&= \mathbb{E}\left[\mathbb{1}_{X \leq S_0} z^X \frac{(1-1/S_0)(1-2/S_0)\ldots(1-(X-1)/S_0)}{(1-1/N)(1-2/N)\ldots(1-(X-1)/N)}\right],
\end{aligned} \tag{4.49}$$

and this follows from the fact that the probability of there being an initially infected individual in $n_i \cup \{i\}$, when I_0 initial infecteds are chosen uniformly at random (without replacement), is now precisely the quantity in the square brackets (for a given realisation of n_i).

Thus in a single realisation, i.e. for an appropriately constructed D', the random variable, of which we take the expectation in equation (4.48), takes a value which is greater than or equal to the value taken by the corresponding random variable in equation (4.49). This follows since the indicator variable takes values in $\{0, 1\}$ and since $S_0 < N$. We can now write:

$$\langle S_i \rangle^{(1)} \geq \langle S_i \rangle^{(2)} \qquad \forall i \in V. \tag{4.50}$$

This indicates that: a) for tree networks, the message passing (or pair-based) system provides a rigorous *upper* bound on $\langle S_i \rangle^{(2)} \forall i$ since it computes $\langle S_i \rangle^{(1)} \forall i$ exactly. b) for non-tree networks, the message passing (or pair-based) system, by providing a lower bound on $\langle S_i \rangle^{(1)} \forall i$, may provide a better approximation of $\langle S_i \rangle^{(2)} \forall i$ than of $\langle S_i \rangle^{(1)} \forall i$, as in figure 4.4.

4.7 Discussion

For Poisson contact processes, the message passing formalism can be cast as a system of integro-differential equations, which conveniently simplify to ODEs/DDEs for exponential/fixed recovery processes. However, we note that for certain other biologically feasible sets of functions $\{f_{ij}(\tau) : i, j \in V, j \in N_i\}$, which do not correspond to Poisson contact processes, the message passing formalism may still allow the dynamics to be obtained via systems of ODEs/DDEs. See, for example, the 'top hat' function discussed by Karrer and Newman (2010). This is a clear advantage of the message passing formalism over the moment-closure formalism, the latter seeming to require the contact processes to be Poisson. In fact, for arbitrary contact and recovery processes, the message passing formalism is theoretically solvable as a system of integral equations.

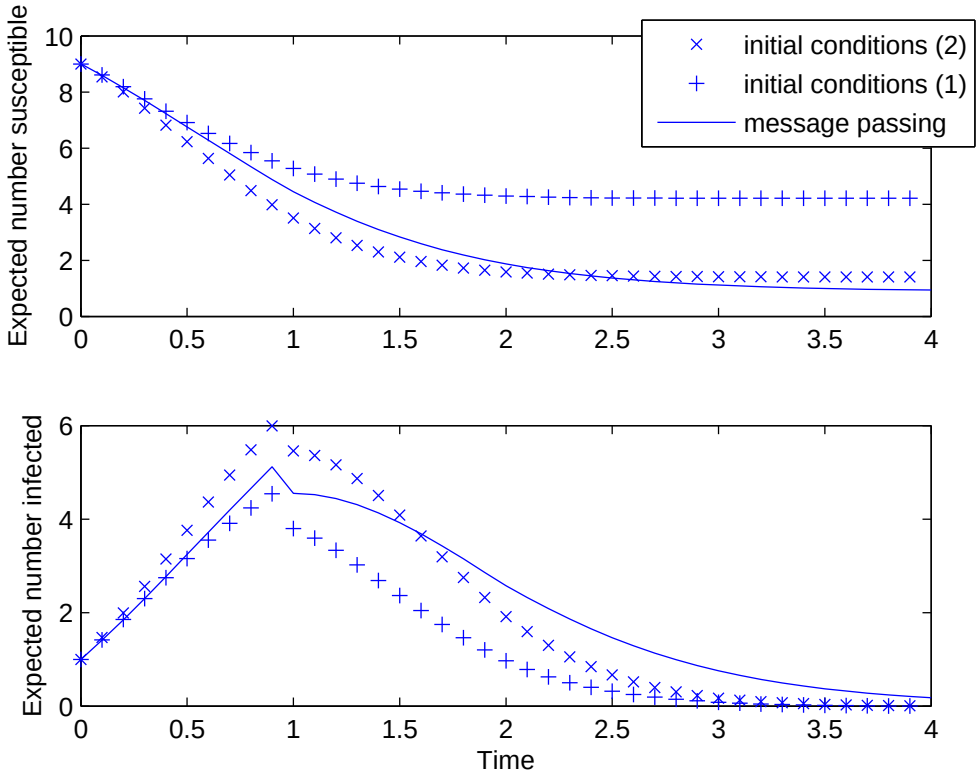

Figure 4.4: Here we are investigating the same scenario as in figure 4.3 except that two extra (undirected) connections have been added to the network, creating multiple loops. The 'x's represent the average of 10,000 stochastic simulations where a single initial infected was chosen uniformly at random, and the rest were initially susceptible. The '+'s represent the equivalent quantity where each individual was initially susceptible with probability 9/10 and initially infected otherwise (independently). The solid line represents the output from the corresponding message passing system, with $z_i = z = 9/10$, $y_i = y = 1/10$ and $x_i = x = 0$.

Other advantages of the message passing approach are its applicability in the domain of random graph ensembles and, by considering $H^{i\leftarrow j}(t)$ (or $F^{i\leftarrow j}(t)$) $\forall i,j : j \in N_i$ in the limit as $t \to \infty$, its connection to percolation-based theory for final outcome statistics (Karrer and Newman, 2010).

The pair-based formalism is a special case of the message passing approach in the sense that it seems to only apply to Poisson contact processes. In this case, the message passing system is more efficient than the pair-based system in terms of the number of equations. However, it is not immediately obvious how to extend the applicability of the message passing equations. For example, to generate exact equations for non-tree networks, or to susceptible-infected-susceptible dynamics. Conversely, the pair-based (moment closure) representation can allow both of these extensions in a straightforward way (Kiss et al., 2014; Nagy and Simon, 2013). Since the physical meaning of each term in the pair-based system is clear, it is also straightforward to make this system applicable to multiple competing diseases on the same network - the number of equations then grows linearly with the number of diseases. However, we note that in the context of configuration network ensembles and Poisson contact and recovery processes, Miller (2013) has shown that the dynamics for competing diseases can be solved via a low-dimensional message-type system (see also Karrer and Newman (2011)).

In our endeavour to understand the relationship between these two formalisms, we have shown that the pair-based moment closure formalism is applicable to arbitrary exposed and infectious periods and, for non-tree networks, provides a lower bound on $\langle S_i \rangle$ $\forall i$ - for tree networks the representation is exact. On the other hand, we have shown that the message passing formalism is applicable to arbitrary finite networks, where the contact/recovery processes are pair-specific/individual-specific, and can incorporate any pure initial system state including vaccinated individuals - or any mixed initial system state where the states of individuals are uncorrelated.

Chapter 5

Using the message passing formalism to prove new results for classic models

5.1 The deterministic SIR model provides a rigorous bound on the Markovian standard SIR model

In this section we will be concerned with the Markovian standard SIR model. We will assume a pure initial state. In doing so, we define $S_0 \subset V$ and $I_0 = V \setminus S_0$ to be the set of initial susceptibles and initial infecteds respectively. In order to avoid triviality we will also assume that $|I_0| > 0$, $|S_0| > 1$.

It was proved by Ball and Donnelly (1987) that the Markovian standard SI model is underestimated, in terms of the expected number susceptible at time t, by its deterministic counterpart. For the Markovian standard SIS model it is also straightforward to show that its deterministic counterpart underestimates the expected number susceptible at time t (see, for example, Allen (2008)). We here provide a proof of the corresponding bound for the Markovian standard SIR model.

We will present a sequence of three systems, all previously defined, which approximate the Markovian standard SIR model, and show that each one increasingly underestimates the expected number of susceptibles $E[X(t)]$ at any positive time point.

The first system is formed from the message passing equations of Karrer and Newman (2010), where we have made use of the generalisation of initial conditions provided by Wilkinson and Sharkey (2014), and the large amount of symmetry in the stochastic process.

The second system is derived by assuming, at the appropriate stage in the construction of the system, that the states of individuals are statistically independent (Sharkey, 2008). We refer to this system as the individual-based system.

The third system, which we refer to as the deterministic SIR system, is the well-known SIR model of Kermack and McKendrick (1927), in the special case where the

individual-level processes are assumed to be Poisson.

The sequence of inequalities which we seek to prove are:

$$\begin{aligned} E[X(t)] &\geq S(t)(\text{message passing}) \quad \forall t, & (5.1)\\ S(t)(\text{message passing}) &\geq S(t)(\text{individual-based}) \quad \forall t, & (5.2)\\ S(t)(\text{individual-based}) &\geq S(t)(\text{deterministic SIR}) \quad \forall t, & (5.3) \end{aligned}$$

where $S(t)(.)$ is the number of susceptible individuals at time t according to the given system, and where the parametrisation and initial conditions are kept the same. However, we will usually drop the system name when it is obvious from the context or section in which it appears. These inequalities indicate that the systems increasingly overestimate the impact of the infection relative to the Markovian standard SIR model. Throughout this section, the different systems will be treated as attempts to approximate the Markovian standard SIR model.

The message passing system

We have seen that the message passing equations can be applied to general S(E)IR dynamics on finite graphs/networks. If the underlying graph is a tree or forest then the equations give results which match the underlying stochastic process exactly; otherwise they give a lower bound on the probability of any given individual being susceptible at time t, and so a lower bound on the expected number susceptible at time t. The Markovian standard SIR model is equivalent to SIR dynamics on a fully connected finite network, where all individual level processes are Poisson. Thus, the message passing equations for the Markovian standard SIR model give:

$$S(t) = \sum_{i \in V} z_i \prod_{j \neq i} F^{i \leftarrow j} \tag{5.4}$$

where

$$F^{i \leftarrow j}(t) = 1 - \int_0^t \beta e^{-(\beta+\gamma)\tau} \left[1 - z_j \prod_{k \neq i,j} F^{j \leftarrow k}(t-\tau) \right] \mathrm{d}\tau. \tag{5.5}$$

However, the pure initial system state and the individual-level symmetry in the stochastic process allow us to simplify equation (5.4) to:

$$S(t) = |S_0|[F_1(t)]^{|S_0|-1}[F_2(t)]^{|I_0|}, \tag{5.6}$$

where

$$\begin{aligned} F_1(t) &= 1 - \int_0^t \beta e^{-(\beta+\gamma)\tau} \left[1 - [F_1(t-\tau)]^{|S_0|-2}[F_2(t-\tau)]^{|I_0|} \right] \mathrm{d}\tau. & (5.7)\\ F_2(t) &= \frac{\beta e^{-(\beta+\gamma)t} + \gamma}{\beta + \gamma}. & (5.8) \end{aligned}$$

The message passing system is then completed by:

$$\begin{aligned} I(t) &= N - S(t) - R(t), \\ \dot{R(t)} &= \gamma I(t). \end{aligned} \tag{5.9}$$

From the results in Karrer and Newman (2010), and Wilkinson and Sharkey (2014), we have inequality (5.1):

$$E[X(t)] \geq S(t)(\text{message passing}). \tag{5.10}$$

In order to compare the message passing system with our next system, the individual-based system, it is necessary to define:

$$Q(t) = 1 - \int_0^t \beta e^{-(\beta+\gamma)\tau} \left[1 - [Q(t-\tau)]^{|S_0|}[F_2(t-\tau)]^{|I_0|}\right] d\tau, \tag{5.11}$$

where $|S_0| - 2$ in the definition for $F_1(t)$ has been changed to $|S_0|$. It is straightforward to show that:

$$Q(t) \leq F_1(t). \tag{5.12}$$

For later reference, we will also need the time derivatives:

$$\dot{F_2(t)} = \gamma(1 - F_2(t)) - \beta F_2(t), \tag{5.13}$$

$$\dot{Q(t)} = \gamma(1 - Q(t)) - \beta Q(t) + \beta[Q(t)]^{|S_0|}[F_2(t)]^{|I_0|}, \tag{5.14}$$

which, since $F_2(0), Q(0) = 1$, imply that $F_2(t), Q(t) \in (0, 1]$.

The individual-based system

By assuming the states of individuals to be statistically independent, at the appropriate stage of construction, the following individual-based system can be derived (Sharkey, 2008):

$$\begin{aligned} \langle \dot{S} \rangle &= -\beta \langle S \rangle \left[(|S_0| - 1)\langle I \rangle + |I_0| e^{-\gamma t}\right], \\ \langle \dot{I} \rangle &= \beta \langle S \rangle \left[(|S_0| - 1)\langle I \rangle + |I_0| e^{-\gamma t}\right] - \gamma \langle I \rangle, \\ \langle \dot{R} \rangle &= \gamma \langle I \rangle, \end{aligned} \tag{5.15}$$

where $\langle S \rangle$, $\langle I \rangle$ and $\langle R \rangle$ represent the fraction of initially susceptible individuals still susceptible at time t, infected at time t and recovered at time t respectively. $e^{-\gamma t}$ is (exactly) the expected fraction of initially infected individuals still infected at time t. Thus, we define

$$S(t) = |S_0| \langle S \rangle \tag{5.16}$$

and note that it is straightforward to see that $\langle S \rangle, \langle I \rangle, \langle R \rangle \in [0, 1]$ (for feasible initial conditions).

We now reformulate the above system in terms of two quantities, $S_1(t)$ and $S_2(t)$, which are defined such that:

$$\langle S\rangle = [S_1(t)]^{|S_0|-1}[S_2(t)]^{|I_0|}, \tag{5.17}$$

$$\begin{aligned} \dot{S_1(t)} &= -\beta S_1(t)\langle I\rangle, \\ S_1(0) &= 1, \end{aligned} \tag{5.18}$$

$$\begin{aligned} \dot{S_2(t)} &= -\beta S_2(t)e^{-\gamma t}, \\ S_2(0) &= 1, \end{aligned} \tag{5.19}$$

and note that if $F_2(t) \geq S_2(t)$ and $F_1(t) \geq S_1(t)$ then inequality (5.2) must hold (compare equations (5.16) and (5.17) with equation (5.6)).

Immediately we can solve for $S_2(t)$:

$$S_2(t) = e^{\frac{\beta}{\gamma}(e^{-\gamma t}-1)}, \tag{5.20}$$

and this allows us to express its time derivative:

$$\dot{S_2(t)} = -\gamma S_2(t)\ln S_2(t) - \beta S_2(t). \tag{5.21}$$

Thus, since $0 \geq x\ln x \geq x-1$ for $x \in (0,1)$, and by comparison of equations (5.19) and (5.21) with equations (5.8) and (5.13), we have that

$$F_2(t) \geq S_2(t). \tag{5.22}$$

Defining $\langle I\rangle^*$ to be equal to the quantity which appears in square brackets in equation (5.15):

$$\langle I\rangle^* = (|S_0|-1)\langle I\rangle + |I_0|e^{-\gamma t}, \tag{5.23}$$

it is then possible, by a separation of variables, to derive:

$$\langle I\rangle^* = N - |S_0|\langle S\rangle - \left(1-\langle S\rangle\right) + \frac{\gamma}{\beta}\ln\langle S\rangle, \tag{5.24}$$

which allows us to express the time derivative of $S(t)$ as:

$$\dot{S(t)} = -\beta S(t)\langle I\rangle^* \tag{5.25}$$

$$= -\beta S(t)\left[N - S(t) - \left(1-\frac{S(t)}{|S_0|}\right) + \frac{\gamma}{\beta}\ln\frac{S(t)}{|S_0|}\right]. \tag{5.26}$$

Expressing $S(t)$ in terms of $S_1(t)$ and $S_2(t)$, and after some straightforward but lengthy algebra, we can also write:

$$\dot{S_1(t)} = -\gamma S_1(t)\ln S_1(t) - \beta S_1(t) + \beta S_1(t)^{|S_0|}S_2(t)^{|I_0|}. \tag{5.27}$$

Again, since $0 \geq x \ln x \geq x - 1$ for $x \in (0, 1)$, and since $F_2(t) \geq S_2(t)$, we have $Q(t) \geq S_1(t)$ (by comparison of equations (5.18) and (5.27) with equations (5.11) and (5.14)). This means that:

$$F_1(t) \geq S_1(t), \tag{5.28}$$

and indeed inequality (5.2) must hold:

$$S(t)(\text{message passing}) \geq S(t)(\text{individual-based}). \tag{5.29}$$

The deterministic SIR system

The deterministic SIR model was first defined by Kermack and McKendrick (1927) and can be represented as a system of ordinary differential equations (When Poisson individual level processes are assumed):

$$\begin{aligned} \dot{S(t)} &= -\beta S(t) I(t), \\ \dot{I(t)} &= \beta S(t) I(t) - \gamma I(t), \\ \dot{R(t)} &= \gamma I(t), \end{aligned} \tag{5.30}$$

where it is straightforward to show that $S(t), I(t), R(t) \in [0, N]$ (for feasible initial conditions). By a separation of variables it is also straightforward to show that:

$$\dot{S(t)} = -\beta S(t) \left[N - S(t) + \frac{\gamma}{\beta} \ln \frac{S(t)}{|S_0|} \right], \tag{5.31}$$

which, by comparison with equation (5.26), gives inequality (5.3):

$$S(t)(\text{individual-based}) \geq S(t)(\text{deterministic SIR}). \tag{5.32}$$

5.2 General epidemics on homogeneous graphs

Let us consider general epidemic dynamics, as defined by Karrer and Newman (2010), on a homogeneous graph $G = (V, E)$, e.g. a finite fully connected (complete) graph, finite ring lattice, infinite square lattice, Bethe lattice etc., where each individual has a finite number n of neighbours by which it can be directly infected. Note that, in general, it now matters how the initial infecteds are 'placed', e.g. if all the initial infecteds are neighbours on a ring lattice, then the number of susceptibles will decrease less rapidly than if the initial infecteds are 'spread out'. Therefore, to increase the symmetry of the stochastic model we will assume that each individual is initially susceptible with probability z and initially infected otherwise, and that the states of individuals are statistically independent at $t = 0$.

For these more general dynamics we define a function $f(\tau)$ such that $\int_0^t f(\tau) d\tau$ is the probability that a given individual, conditional on its getting infected, makes an *infectious* contact to a given neighbour within time period t, and $r(\tau)$ such that

$\int_0^t r(\tau)d\tau$ is the probability that a given individual, given that it gets infected, recovers within time period t. Therefore f, in general, depends on r.

Due to the homogeneity of the graph and the initial conditions, the message passing system is simply:

$$\begin{aligned} S(t) &= zN[F(t)]^n, \\ I(t) &= N - S(t) - R(t), \\ R(t) &= \int_0^t r(\tau)[N - S(t-\tau)]\mathrm{d}\tau, \end{aligned} \tag{5.33}$$

where

$$F(t) = 1 - \int_0^t f(\tau)\left[1 - z[F(t-\tau)]^{n-1}\right]\mathrm{d}\tau. \tag{5.34}$$

From the arguments given by Karrer and Newman, this system underestimates the expected number susceptible at time t; except where the network is a tree, or is locally tree-like, where the system is exact. This implies that across all undirected homogeneous graphs, for a given n, the infection will have the largest impact in the Bethe lattice (since the system is the same in each case but is only exact for the Bethe lattice). Indeed, clustering and the presence of other cycles in the network is known to, in general, slow down and/or limit the spread of the infection (see Miller (2009) and references therein).

Let us now consider the individual-level processes to be Poisson such that $f(\tau) = \beta e^{-(\beta+\gamma)\tau}, r(\tau) = \gamma e^{-\gamma\tau}$ (the Markovian version). By expressing the time derivative of $S(t)$ as a function of its 'current' value, in system (5.33), it is then straightforward to show that the following individual-based system:

$$\begin{aligned} \dot{S(t)} &= -\beta\frac{n}{N}S(t)I(t), \\ \dot{I(t)} &= \beta\frac{n}{N}S(t)I(t) - \gamma I(t), \\ \dot{R(t)} &= \gamma I(t), \end{aligned} \tag{5.35}$$

where $S(0) = zN, I(0) = (1-z)N$ and $R(0) = 0$, gives an even lower number of susceptibles at time t, i.e. $S(t)$ computed from system (5.33) is greater than $S(t)$ computed from system (5.35) (for the Markovian case).

Note that in this section we assumed a mixed initial system state and, for this reason, the proofs for the arguments in this section are not sufficient for proving inequalities (5.1) to (5.3) where a pure initial system state was assumed.

5.3 The general epidemic on an infinite complete graph (mean-field case)

Since a complete graph is homogeneous, system (5.33) of the previous section must still apply. However, since we here consider the case where $N \to \infty$, we will convert the

variables of the system into fractions (divide through by N):

$$\begin{aligned} S(t) &= z[F(t)]^n, \\ I(t) &= 1 - S(t) - R(t), \\ R(t) &= \int_0^t r(\tau)[1 - S(t-\tau)]\mathrm{d}\tau, \end{aligned} \tag{5.36}$$

where $S(t), I(t)$ and $R(t)$ are now the fractions of the population that are susceptible, infected and recovered respectively ($S(t) + I(t) + R(t) = 1$). However, the occurrence of n is now problematic since it also tends to infinity. To try for a different expression of $S(t)$, we first write:

$$\dot{S(t)} = \frac{S(t)}{F(t)} \left[\frac{n-1}{n} \int_0^t n f(\tau) \frac{\dot{S}(t-\tau)}{F(t-\tau)} d\tau - n f(t)(1-z) \right], \tag{5.37}$$

where we have made use of Leibniz's integral rule. Now, because an individual is able to make contacts to an infinite number of neighbours, it is sensible to impose the following (as $n \to \infty$):

$$\int_0^t f(\tau) d\tau \to 0, \tag{5.38}$$

$$n \int_0^t f(\tau) d\tau \to c(t), \tag{5.39}$$

where $c(t)$ is monotonically increasing from zero and converges to some finite number. The first imposition implies that the probability of a given infected individual making an infectious contact to a different given individual tends to zero. This is sensible since the number of individuals to which it can make contacts is infinite. The second imposition implies that the expected total number of infectious contacts that an infected individual will make is positive, but finite. In the limit as $n \to \infty$, and $t \to \infty$, $c(t)$ is the expected number of infectious contacts (to different individuals chosen uniformly at random) made by a given individual during its entire infectious period (assuming it gets infected). Note that the first imposition also implies that $F(t) \to 1$. Therefore, in the limit as $n \to \infty$, equation (5.37) becomes:

$$\dot{S(t)} = S(t) \left[\int_0^t f^*(\tau) \dot{S}(t-\tau) d\tau - f^*(t)(1-z) \right], \tag{5.40}$$

where $f^*(t)\Delta = n f(t)\Delta$ is the expected number of infectious contacts, to different individuals chosen uniformly at random, made by a given individual between time t and time $t + \Delta$ (for $\Delta \to 0$), where time is measured from the moment the individual first becomes infected (assuming that it does). $f^*(\tau)$ is related to $r(\tau)$ in the following way:

$$f^*(\tau) = s^*(\tau) \int_\tau^\infty r(t') dt',$$

where $s^*(t)\Delta$ is the expected number of contacts (that can be non-infectious), to different individuals chosen uniformly at random, made by a given individual between time t and time $t+\Delta$ (for $\Delta \to 0$), where time is measured from the moment the individual first becomes infected (assuming that it does).

Our general mean-field system can now be expressed:

$$\begin{aligned} \dot{S(t)} &= S(t)\left[\int_0^t f^*(\tau)\dot{S}(t-\tau)d\tau - I(0)f^*(t)\right], \\ I(t) &= 1 - S(t) - R(t), \\ R(t) &= \int_0^t r(\tau)[1-S(t-\tau)]\mathrm{d}\tau. \end{aligned} \tag{5.41}$$

For the case of Poisson contact and recovery processes we have $f^*(\tau) = \beta n e^{-(\beta+\gamma)\tau}$ and $r(\tau) = \gamma e^{-\gamma\tau}$. Imposing that $\beta n \to \beta'$ as $n \to \infty$, where β' is then the constant (exponential) rate at which an infected individual makes infectious contacts at random to the rest of the population, and then plugging this into the above system, generates the (Poisson) Kermack and McKendrick deterministic SIR system (system (5.30) after dividing through by N). This system is known to exactly capture Markovian SIR dynamics on a complete graph in the limit of large population size (Ethier and Kurtz, 1986; Andersson and Britton, 2000).

We conjecture that the above general mean-field system is an 'exact' representation of the general epidemic on a fully connected (complete) graph, in the limit of large population size (with some modest restrictions on the underlying stochastic process, as discussed by Barbour and Reinert (2013)). This conjecture is based on the fact that, due to the infinite population size and the finite expected number of infectious contacts per infected individual, the path of the infection is locally tree-like. This means that the independence assumption (Karrer and Newman, 2010; Wilkinson and Sharkey, 2014) in the message passing equations is valid.

5.4 General epidemics on configuration networks with two levels of mixing (superimposed even mixing)

Following Kiss et al. (2006) and Ball and Neal (2008), we consider SIR dynamics on a configuration network where the infection may be transmitted across network links or via a superimposed even mixing process. However, we will construct a system to capture the dynamics using the message passing formalism, and we will not (initially) assume that the individual level processes are Poisson. Thus, we define $f_L(\tau)$ such that $\int_0^t f_L(\tau)d\tau$ is the probability that an individual makes a local infectious contact to a given neighbour within time period t of having become infected, and $f_G(\tau)$ such that $\int_0^t f_G(\tau)d\tau$ is the probability that an individual makes a global infectious contact to another given individual within time period t of having become infected. Similarly to

the previous subsection, we will impose that

$$(N-1)\int_0^t f_G(\tau)d\tau \to c(t) \qquad \text{as } N \to \infty, \tag{5.42}$$

where $c(t)$ is a non-decreasing function which converges to some finite value. We will make the assumption that individuals are susceptible at $t = 0$ with probability z and infected otherwise, and that the states of all individuals are statistically independent at $t = 0$.

Let us first consider the probability that a randomly selected individual is susceptible at time t, this being equivalent to the expected fraction susceptible:

$$S(t) = zG_0(H(t))F(t)^{N-1}, \tag{5.43}$$

where G_0 is the generating function of the degree distribution, and satisfies:

$$G_0(H(t)) = \sum_{k=0}^{\infty} p_k H(t)^k, \tag{5.44}$$

where p_k is the fraction of the population which is of degree k. $H(t)$ is the probability that a random individual, after having been placed in the cavity state, does not receive a local infectious contact from a given neighbour by time t (averaged over the network ensemble). $F(t)$ is the probability that a random individual, after having been placed in the cavity state, does not receive a global infectious contact from another given individual by time t. Thus, $G_0(H(t))$ is the probability that a random individual, in the cavity state, does not receive any local infectious contacts by time t (averaged over the network ensemble), while $F(t)^{N-1}$ is the probability that it receives no global infectious contacts by time t.

We now write:

$$H(t) \;=\; 1 - \int_0^t f_L(\tau)\Big[1 - zG_1(H(t-\tau))F(t-\tau)^{N-2}\Big]d\tau, \tag{5.45}$$

where G_1 is the generating function of the excess degree of a neighbour of a random individual. It can be computed as (see, for example, Newman (2002)):

$$G_1(H(t)) = \frac{G_0'(H(t))}{G_0'(1)}. \tag{5.46}$$

For $F(t)$, we have:

$$F(t) \;=\; 1 - \int_0^t f_G(\tau)\Big[1 - zG_0(H(t-\tau))F(t-\tau)^{N-2}\Big]d\tau, \tag{5.47}$$

where here we use G_0 instead of G_1 since this is a global contact. The reader is advised to consult Karrer and Newman (2010), and their application of the message passing formalism to a configuration network scenario, in order to better understand the form of our equations for $S(t)$, $H(t)$ and $F(t)$.

The time derivative of $S(t)$ can be written:

$$\dot{S(t)} = S(t)\frac{G_0'(H(t))}{G_0(H(t))}\dot{H(t)} + S(t)(N-1)\frac{\dot{F(t)}}{F(t)}. \tag{5.48}$$

Now, assuming that in the limit as $N \to \infty$ the quantity in the square brackets of equation (5.47) is equal to $1 - S(t-\tau)$ (compare with equation (5.43)), then we can write:

$$\dot{F(t)} = \int_0^t f_G(\tau)S(\dot{t-\tau})d\tau - (1-z)f_G(t), \tag{5.49}$$

and so the second term in equation (5.48) can be expressed:

$$S(t)\int_0^t f_G^*(\tau)S(\dot{t-\tau})d\tau - S(t)(1-z)f_G^*(t), \tag{5.50}$$

where $f_G^*(t) = (N-1)f_G(t)$, and $\int_0^t f_G^*(\tau)d\tau = c(t)$ (as $N \to \infty$) is the expected number of global infectious contacts that an individual will make, to others chosen uniformly at random, within time period t of having been infected (note that we have also here assumed $S(t)/F(t) = S(t)$ in the limit as $N \to \infty$). Thus:

$$\dot{S(t)} = S(t)\left[\frac{G_0'(H(t))}{G_0(H(t))}\dot{H(t)} + \int_0^t f_G^*(\tau)S(\dot{t-\tau})d\tau - (1-z)f_G^*(t)\right]. \tag{5.51}$$

For the time derivative of $H(t)$, we can write:

$$\dot{H(t)} = \int_0^t f_L(\tau)S_2(\dot{t-\tau})d\tau - (1-z)f_L(t), \tag{5.52}$$

where we have defined:

$$S_2(t) \equiv zG_1(H(t))F(t)^{N-1}, \tag{5.53}$$

and due to form equivalence:

$$\dot{S_2(t)} = S_2(t)\left[\frac{G_1'(H(t))}{G_1(H(t))}\dot{H(t)} + \int_0^t f_G^*(\tau)S(\dot{t-\tau})d\tau - (1-z)f_G^*(t)\right]. \tag{5.54}$$

We now define the full system:

$$\begin{aligned}
\dot{S(t)} &= S(t)\left[\frac{G_0'(H(t))}{G_0(H(t))}\dot{H(t)} + \int_0^t f_G^*(\tau)S(\dot{t-\tau})d\tau - (1-z)f_G^*(t)\right], \\
I(t) &= 1 - S(t) - R(t), \\
R(t) &= \int_0^t r(\tau)[1-S(t-\tau)]d\tau,
\end{aligned} \tag{5.55}$$

where

$$\begin{aligned}
\dot{H(t)} &= \int_0^t f_L(\tau)S_2(\dot{t-\tau})d\tau - (1-z)f_L(t), \\
\dot{S_2(t)} &= S_2(t)\left[\frac{G_1'(H(t))}{G_1(H(t))}\dot{H(t)} + \int_0^t f_G^*(\tau)S(\dot{t-\tau})d\tau - (1-z)f_G^*(t)\right].
\end{aligned} \tag{5.56}$$

For the case of Poisson individual level processes, i.e. $f_L(\tau) = \beta_L e^{-(\beta_L+\gamma)\tau}$, $f_G(\tau) = \beta_G e^{-(\beta_G+\gamma)\tau}$, $r(\tau) = \gamma e^{-\gamma\tau}$, the equations simplify to a system of ODEs:

$$\begin{aligned} \dot{S(t)} &= -\beta_G^* S(t) I(t) + S(t) \frac{G_0'(H(t))}{G_0(H(t))} \dot{H(t)}, \\ I(t) &= 1 - S(t) - R(t), \\ \dot{R(t)} &= \gamma I(t), \end{aligned} \tag{5.57}$$

where $\beta_G^* = (N-1)\beta_G$ and

$$\begin{aligned} \dot{H(t)} &= (\beta_L + \gamma)(1 - H(t)) - \beta_L(1 - S_2(t)), \\ \dot{S_2(t)} &= -\beta_G^* S_2(t) I(t) + S_2(t) \frac{G_1'(H(t))}{G_1(H(t))} \dot{H(t)}. \end{aligned} \tag{5.58}$$

This last system, of just 5 equations, represents a large improvement in efficiency relative to the system given by Ball and Neal (2008). Indeed, the size of their system is of the order of the number of distinct degrees in the degree distribution. Further, we conjecture that our system is 'exact' (as is theirs) in the limit as $N \to \infty$, and that a rigorous proof of this would be possible along the same lines as the proof provided by Decreusefond et al. (2012) for Volz's (2008) system (which is equivalent to Miller's (2010) edge-based system for Markovian SIR dynamics on configuration networks, and to Karrer and Newman's (2010) message passing approach when applied to Poisson individual level processes and configuration networks). The derivation of the above system followed quite simply from the principles established by Karrer and Newman (2010).

Chapter 6

Final summary and discussion

In chapter 2, we investigated the Markovian network-based SIS model, this being a generalised version of the well-known Contact Process. Our main contribution here was to apply important results issuing from highly technical probabilists working on the Contact Process, such as Harris (1974) and Liggett (1999), in combination with the concept of the quasi-stationary distribution (Daroch and Seneta, 1967), to produce a general and useful result for modellers in epidemiology - The Prevalence-Invasion Relationship. Indeed, the very definitions of invasion and invasion probability, and to a lesser extent endemic prevalence, were previously problematic for this finite-network-based model. We have now provided precise mathematical definitions of such quantifiers, i.e. $\mathbb{P}^A_{T,\Gamma}$(quasi-invasion) and $\mathbb{P}^A_{T,\Gamma}$(quasi-prevalence), which are straightforward to approximate numerically, and proved the exact relationship between them (Wilkinson and Sharkey, 2013). Moreover, these definitions are based on the existence of a quasi-stationary distribution which is independent, or largely independent, of initial conditions; as such, they can be generalised to a large class of stochastic models which meet this requirement. In the limit of large population size, in cases where such a limit makes sense, our definitions are equivalent to existing definitions/measures such as: the probability of indefinite persistence in a branching process and the upper invariant measure of the Contact Process.

The work in chapter 2 also provided insight into the equality between invasion probability from a single initial infected in the Markovian standard SIR/SIS models, computed from the corresponding branching process, and the fraction of the population infected in the endemic equilibrium of the deterministic SIS model. The equality follows from the 'exactness' of these two computations in the limit of large population size, and the property of duality. In essence, we extended this result, and the equivalent result for the Contact Process on an infinite homogeneous network (see, for example, Grimmett (2010) and Neal (2008)), to a more general and heterogeneous finite-network setting.

There are some immediate practical implications of the Prevalence-Invasion Rela-

tionship when the network on which the disease spreads is undirected (assuming the model corresponds to a real world process): 1) The individuals which spend the most time in the infected state are the same individuals most capable of initiating large scale outbreaks, and should be targeted for intervention. 2) When measuring invasion probability or endemic prevalence, through data collection or mathematical models, the two are interchangeable.

For strongly connected directed networks, invasion probability is interchangeable with endemic prevalence after reversing the directions of network links. Such an interchange, via the reversal of network links, may be desirable and possible in technological networks. A similar relationship was seen to hold for the case of arbitrary directed networks with strongly connected components.

In chapter 3, we considered Markovian network-based SEIR dynamics. In particular, we adopted the technique of moment closure in order to capture the dynamics via systems of ordinary differential equations. In the introduction, pair-based systems at the population level were constructed for the case of finite directed networks, and the assumptions behind the 'closures' were made clear (following Sharkey (2008)). In the literature, directed networks and finite networks are often overlooked.

For the case of tree networks, a particular individual level pair-based system was seen to exactly capture the dynamics of the expected compartment sizes, assuming that the states of all individuals are independent at $t = 0$ (this encompasses all pure initial states) (for the SIR case, see Sharkey et al. (2013) and Kiss et al. (2014)). The exact closure which enabled this was then generalised such that exact systems for non-tree networks, which go beyond the pair-level, could be 'written down'; see the Exact Closure Theorem and 'dynamical partitioning' (Sharkey and Wilkinson, 2015). These systems, although potentially very large, are sometimes considerably less than the 4^N Kolmogorov forward equations needed to fully capture the evolution of the system. In addition, the dependence of the applicability of this approach on network structure, and particular sub-structures which we termed 'transmission blocks', was discussed. These transmission blocks are a particular subset of the biconnected sub-graphs of the underlying undirected graph/network.

For an arbitrary network, we then defined several hierarchies of approximating moment closure-type systems, where the first system of the hierarchy was always the pair-based system and the last system was always exact. This was made possible by proposing different criteria under which the Exact Closure Theorem is to be assumed to hold true and thereby employed in the construction of the system (the precise conditions under which the theorem holds true are ignored). Moment closure systems which go beyond the pair-level are rare in the literature. This is understandable since the systems quickly become unwieldy.

In chapter 4, the message passing approach (Karrer and Newman, 2010) for general

SEIR dynamics on arbitrary finite networks was described, and slightly generalised (Wilkinson and Sharkey, 2014). It was seen that in the case of tree networks, and if the states of individuals are independent at $t = 0$, then the message passing equations exactly capture the probability of a given individual being in a given state. For non-tree networks, and the same restriction on initial conditions, the equations provide a rigorous lower bound on the probability of a given individual being susceptible.

When the individual level contact processes were Poisson, it was shown that the message passing equations could be used to derive a pair-based system of integro-differential equations which allowed arbitrary (in distribution) exposed and infectious periods; the derivation involved taking time derivatives and renaming variables in such a way that the two systems of equations give exactly the same output. If, in addition, the exposed and infectious periods were exponentially distributed, then the pair-based system reduced to the familiar pair-based system of ODEs presented in chapter 3. This being so, it could be concluded that the pair-based system of ODEs also provides a rigorous lower bound on the probability of a given individual being susceptible, assuming that the states of individuals are initially independent.

The issue of initial conditions was then addressed such that the situation where a specific number of initial infecteds are chosen at random (without replacement), and the rest are susceptible, could be considered. In particular, for these initial conditions and a tree network, it was shown that the message passing equations provide an *upper* bound on the probability of a given individual being susceptible. For these initial conditions and a non-tree network, the bound was no longer assured but it was seen that the over-estimation of the spread, caused by the presence of cycles in the network, compensated in the 'right direction'; the output could provide an even better approximation for these initial conditions than for the type which were originally assumed.

For many networks of interest, the systems defined in chapters 3 and 4 will be extremely large, often prohibitively. In many cases, it will be much more computationally efficient to run, and analyse, large numbers of simulations. However, in more practicable low-dimensional systems, which make use of mean field type approximations to construct population level variables, the assumptions/approximations underlying their construction, and the kinds of errors they induce, are not obvious. It is likely that improved low-dimensional systems, with more clearly stated assumptions/approximations, and a better understanding of their effects, could be developed from our individual level systems. Indeed, this is something which we will pursue in future research.

In chapter 5, the message passing formulation was applied to some classic models in mathematical epidemiology, and some new results obtained. It was seen that, relative to the Markovian standard SIR model (with a pure initial state), the deterministic SIR model overestimates the spread of the infection, providing a rigorous lower bound on the expected number susceptible. Similarly, a model which is formally the same as the

deterministic SIR model, but naively accounts for the finite number of neighbours by which an individual can be directly infected, overestimates the spread of the infection relative to Markovian SIR dynamics on a homogeneous network where each individual has precisely that number of neighbours. It was also shown how the original deterministic model of Kermack and McKendrick (1927), which does not assume Poisson individual level processes, could be derived from message passing equations.

The message passing approach of Karrer and Newman (2010) was finally used to construct a system for the case of general SIR dynamics in a configuration network setting, with superimposed even mixing (considered by Kiss et al. (2006) and Ball and Neal (2008)). To our knowledge, all systems in the literature which address such a scenario are at least as large as the number of distinct degrees in the degree distribution, and assume the individual level processes to be Poisson. Our system consists of just 5 equations (ODEs in the case of Poisson individual level processes) and we conjecture that it is 'exact' in the limit of large population size.

Appendix

A.1 Proof of the underpinning results for Theorem 3.4.3

Let us consider Markovian SEIR dynamics on a network $D = (V, A)$ for which M_E is the associated system derived by making use of the exact closure theorem wherever possible. Theorem 3.4.3 follows from Corollary A.1.1 and Corollary A.1.2 below.

Definition A.1.1. *A set $W_n \subset V$ of size $|W_n| = n$ can be 'generated' from a set $W_m \subset V$ of size $|W_m| = m$ where $2 \leq m < n$ if and only if a sequence of sets $W_m, \ldots, W_i, \ldots, W_n$ exist where $W_{i+1} = W_i \cup \{k\}$, where k is a single node in $V \setminus W_i$, and there exists an arc from k towards some individual $j \in W_i$ which is not dynamically partitioning relative to k and $W_i \setminus \{j\}$.*

Remark. The above definition is constructed such that $\langle \psi_W^A \rangle$, where $|W| > 2$, is a variable in M_E, for some A, if and only if W can be generated from some connected pair. This follows from the definition of M_E via equation (3.27).

Lemma A.1.1. *If a set W can be generated from some connected pair, then there exists $X \supset W$ such that $D[X]$ is a directed sub-block. There also exists some node $i \in W$ that it is reachable from all other nodes in both $D[W]$ and $D[X]$.*

Proof. The proof follows by induction. Lemma A.1.2 proves the statement for the case $|W| = 3$ while Lemma A.1.3 establishes the inductive step. □

Corollary A.1.1. *If $\langle \psi_W^A \rangle$, where $|W| > 2$, is a variable in M_E, then there exists $X \supset W$ such that $D[X]$ is a directed sub-block.*

Proof. This follows directly from Lemma A.1.1 and Definition A.1.1. □

Lemma A.1.2. *If a set W where $|W| = 3$ can be generated from some connected pair, then there exists $X \supset W$ such that $D[X]$ is a directed sub-block, and some $i \in W$ is reachable from all others in both $D[W]$ and $D[X]$.*

Proof. With reference to Figure 1, if a set $W_3 = \{i, j, k\}$ can be generated from the pair $W_2 = \{i, j\}$, with j connected towards i, then there is a link from k to either i or j. Further, from the definition of dynamical partitioning and the generating rule, there are two possibilities: 1) there exists two vertex disjoint paths P_1, P_2 from some

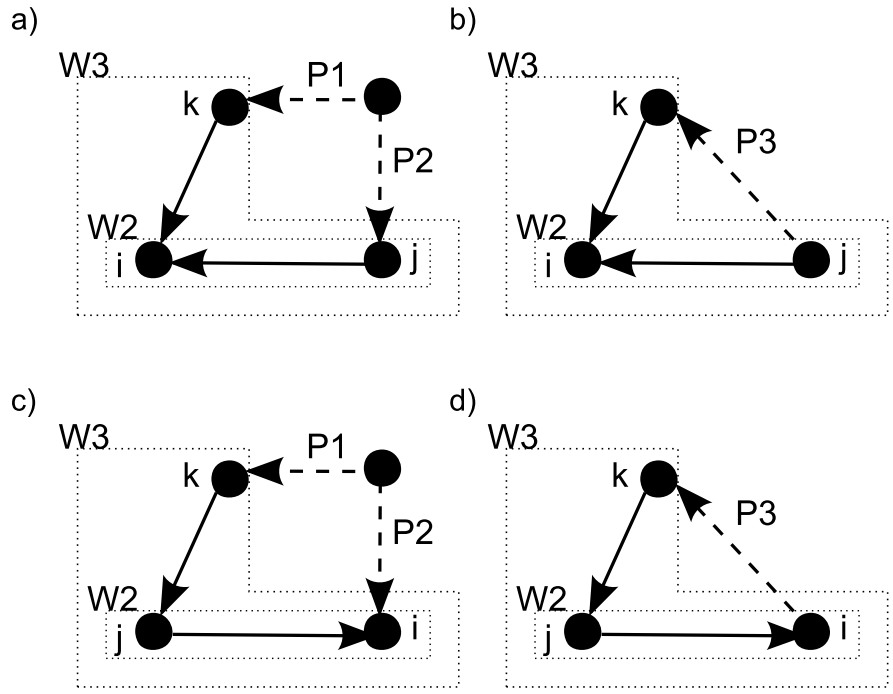

Figure 1: Demonstration for Lemma A.1.2: 'ways' in which a set $W_3 = \{i, j, k\}$ can be generated from the pair $\{i, j\}$, where j is connected towards i. Note that W_3 is always a subset of some directed sub-block, and i is reachable from all others in both $D[W_3]$ and the directed sub-block. The dashed arrows represent paths which may consist of any number of vertices.

individual (which could be k) to both members of W_2, and where k is the penultimate individual in one of these paths (see Figure 1a&c), or 2) there exists a path P_3 from one member of W_2 to the other, and k is the penultimate individual in this path (see Figure 1b&d). Note that in all cases depicted in Figure 1, W_3 is a subset of some directed sub-block in which i is reachable from all others (and i is reachable from all others in $D[W_3]$).

□

Lemma A.1.3. *If the statement made in Lemma A.1.1 is true for the case where $|W| = n$, then it is also true when $|W| = n + 1$.*

Proof. Firstly, note that W_{n+1}, where $|W_{n+1}| = n+1$, can be generated from some connected pair if and only if it can be generated from some set W_n, where $|W_n| = n$, which can itself be generated from some connected pair. Now suppose that Lemma A.1.1 is true for the case where $|W| = n$, and let W_n be a set of size n that can be generated from some connected pair. Then we have a set $X \supset W_n$ such that $D[X]$ is a directed sub-block where, without loss of generality, $i \in W_n \subset X$ is reachable from all others in both $D[W_n]$ and $D[X]$. With reference to Figure 2, and again focusing only on directed links, if a set $W_{n+1} = W_n \cup \{k\}$ $(k \notin W_n)$ can be generated from W_n, then either there

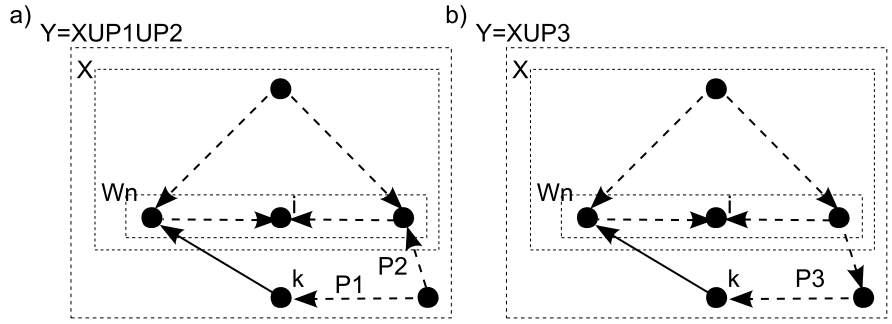

Figure 2: Demonstration for Lemma A.1.3. Here, the single node in $X \backslash W_n$ is illustrative of the nodes in this set which must be connected by at least one path leading to node i, and where the underlying graph $G[X]$ is biconnected. We have placed node k outside of X, but $k \in X \setminus W_n$ is also permitted. a) shows k belonging to one of two vertex disjoint paths from some node to W_n and b) shows k as the penultimate individual in a path from a node in W_n to a different node in W_n. In either case, $W_n \cup \{k\}$ is seen to always be a subset of some $Y \supset X$ where $D[Y]$ is a directed sub-block in which i is reachable from all others (and i is reachable from all others in $D[W_n \cup k]$).

exist two vertex disjoint paths P_1, P_2 from some individual to two different members of W_n and k is the penultimate individual in one of these paths (Figure 2a), or there exists a path P_3 from one member of W_n to a different member of W_n and k is the penultimate individual in this path (Figure 2b). This follows from the generating rule and the definition of dynamical partitioning. Note that if P_1, P_2 exist then $D[X \cup P_1 \cup P_2]$ is a directed sub-block in which i is reachable from all others (and i is reachable from all others in $D[W_{n+1}]$). Similarly, if P_3 exists then $D[X \cup P_3]$ is a directed sub-block in which i is reachable from all others (and i is reachable from all others in $D[W_{n+1}]$). □

Lemma A.1.4. *Let $D[X]$ be a directed sub-block and let $i \in W \subset X$, where $|W| \geq 2$, be reachable from all others in both $D[W]$ and $D[X]$. In this case, some set $W \cup \{k\}$, where $k \in X \setminus W$, can be generated from W, and i is reachable from all others in $D[W \cup \{k\}]$.*

Proof. From Figure 2, but with $k \in X$, we note that some set $W \cup \{k\}$, where $k \in X \backslash W$, can be generated from W if and only if there exist two vertex disjoint paths P_1, P_2 from some individual to two different members of W and where k is the penultimate individual in one of these paths, or there exists a path P_3 from one member of W to a different member of W and k is the penultimate individual in this path (note that we are referring to W_n in figure 2 as W). Our proof is by contradiction. We shall assume that neither of these scenarios hold and show that this contradicts the assumption that $D[X]$ is a directed sub-block.

Every individual in $X \setminus W$ is at the start of a path to i in $D[X]$. Figure 3 shows the ways in which $k \in X \setminus W$ may be connected to a node of W, in $D[X]$. Firstly, the underlying graph in Figure 3a is not biconnected so here $D[X]$ is not a directed

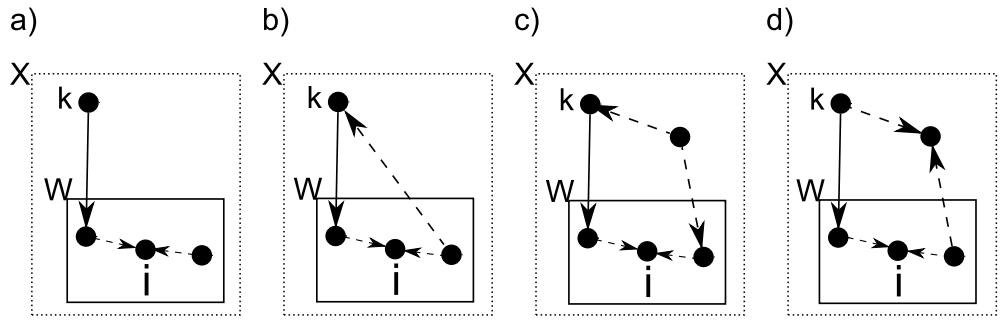

Figure 3: Demonstration for Lemma A.1.4: shows the ways in which $k \in X \setminus W$ may be connected to a node of W, in $D[X]$. We have cases a) the underlying graph of $D[X]$ is not biconnected, b) Existence of path $P3$, c) Existence of paths $P1$ and $P2$ and d) Existence of a node from which W is unreachable. Cases a) and d) imply $D[X]$ is not a directed sub-block and so the existence of paths $P1$ and $P2$, or of path $P3$, is established.

sub-block. Secondly, Figures 3b and c correspond to the existence of path $P3$ and the existence of paths $P1, P2$ respectively and hence $W \cup \{k\}$ is generated. Finally, Figure 3d has a node from which W is unreachable and so $D[X]$ cannot be a directed sub-block. Other more complicated variants of this path will also contain such nodes from which W is unreachable. Hence, if paths $P1$ and $P2$ do not exist, and path $P3$ does not exist, then $D[X]$ is not a directed sub-block.

□

Corollary A.1.2. *If there exists $X \subset V$ such that $D[X]$ is a directed sub-block, then there exists $\langle \psi_X^A \rangle$, for some A, as a variable in M_E.*

Proof. If $D[X]$ is a directed sub-block in which $i \in X$ is reachable from all others, then there exists at least one arc (j, i) in $D[X]$. The corollary then follows from lemma A.1.4 which proved that, for such a case, X can be generated from $\{i, j\}$ (and from the remark under definition A.1.1). □

A.2 Proof of general results on cycle-partitioning

The main results of this appendix are stated as Theorem A.2.1 and Theorem A.2.2.

Lemma A.2.1. *If $\langle \psi_W^A \rangle$ is a variable in M_{C_x}, then there is at least one individual reachable from all others in $D[W]$.*

Proof. Follows from the way in which $M_{C(x)}$ is constructed via equation (3.42) (or equation (3.44)). □

Theorem A.2.1. *If the largest transmission block in a network consists of n individuals, then any cycle-partitioning system of order $x \geq n - 2$ is exact.*

Proof. For any $W \subset V$ where at least one individual is reachable from all others in $D[W]$, if any $i \in W$ is cycle-partitioning at order $x \geq n-2$ with respect to some $j \notin W$ and $W \setminus i$, where (j,i) is an arc, then i is also dynamically partitioning with respect to j and $W \setminus i$. This follows because if i is not dynamically partitioning, but is cycle-partitioning at order $x > n-2$, then this implies the existence of a directed sub-block containing j, i and at least one other member of W, and which consists of more than n individuals. Therefore, by Lemma A.2.1, $M_{C(x)}$ only utilises genuine dynamical partitioning and we have $M_{C(x)} = M_E$ $(x \geq n-2)$.

□

Theorem A.2.2. *If the smallest directed sub-block consists of n individuals, then all cycle-partitioning systems of order $x < n-2$ are equivalent to the pair-level system.*

Proof. For any connected pair $W \subset V$ $(|W| = 2)$, if $i \in W$ is not cycle-partitioning at order $x < n-2$ with respect to $j \notin W$ and $W \setminus i$, where (j,i) is an arc, then there exists a directed sub-block containing $W \cup j$, and which consists of less than n individuals. Therefore, no such j can exist. From the way in which $M_{C(x)}$ is constructed, this means that no subsystem states larger than connected pairs emerge and we have the pair-level system, i.e. $M_{C(x)} = M_{C(0)}$ $(x < n-2)$.

□

Remark. Together, Theorems A.2.1 and A.2.2 imply that the difference in size between the largest directed sub-block (or largest transmission block) and smallest directed sub-block gives an upper bound on the number of distinct systems that the cycle-partitioning approach can provide. If all directed sub-blocks are the same size then no systems that are distinct from the pair-level system and the exact dynamical partitioning system M_E emerge. However, even when this difference is large the number of distinct systems may sometimes be small, as was shown to be the case for the triangle lattice (where the difference is $N-3$).

Bibliography

Allen, L.J.S. (2008) *An introduction to stochastic epidemic models*, Lecture Notes in Mathematics, vol. 1945, Springer, Berlin, pp. 81-130.

Anderson, R.M. and May, R.M. (1986), 'The invasion, persistence and spread of infectious diseases within animal and plant communities', *Phil Trans R Soc Lond B* **314**, 533-570.

Anderson, R.M. and May, R.M. (1991), *Infectious disease of humans. Dynamics and control*, New York: Oxford Univ Press.

Andersson, H. and Britton, T. (2000) *Stochastic epidemic models and their statistical analysis*, Springer Lecture Notes in Statistics, New York: Springer Verlag.

Bailey, N.T.J. (1975) *The Mathematical Theory of Infectious Diseases*, London: Griffin.

Ball, F. (1999) 'Stochastic and deterministic models for SIS epidemics among a population partitioned into households', *Mathematical Biosciences* **156**, 41-67.

Ball, F. and Donnelly, P. (1995) 'Strong approximations for epidemic models', *Stochastic Processes and their Applications* **55**, 1-21.

Ball, F. and Neal, P. (2008) 'Network epidemic models with two levels of mixing', *Mathematical Biosciences* **212**, No. 1, p. 69-87.

Ball, F., Sirl, D. and Trapman, P. (2010), 'Analysis of a stochastic SIR epidemic on a random network incorporating household structure', *Math. Biosci.* **224**, 53-73.

Balthrop, J., Forest, S., Newman, M.E.J. and Williamson, M.W. (2004) 'Technological networks and the spread of computer viruses', *Science* **304**, 5670:527-529.

Barbour, A.D. and Reinert, G. (2013), 'Approximating the epidemic curve', *Electron J Probab* **18**(54), 1-30.

Bartlett, M.S. (1956) 'Deterministic and stochastic models for recurrent epidemics', *Proceedings of the Third Berkley Symposium on Mathematical Statistics and Probability* **4**, 81-108.

Bauch, C. (2005) 'The spread of infectious diseases in spatially structured populations: an invasory pair approximation', *Math. Biosci.* **198**, 217-237.

Bender, E. A. and Canfield, E. R. (1978) 'The asymptotic number of labelled graphs with given degree sequences', *Journal of Combinatorial Theory A* **24**, 296-307.

Bollobás, B. (1985) *Random graphs*, Academic Press.

Chaterjee, S. and Durrett, R. (2009) 'Contact processes on random graphs with power law degree distributions have critical value 0', *Ann Probab* **37**, 6:2332-2356.

Clancy, D. and Mendy, S.T. (2011) 'Approximating the quasi-stationary distribution of the SIS model for endemic infection', *Methodol Comput Appl Probab* **13**:603-618.

Danon, L., Ford, A.P., House, T., Jewell, C.P., Keeling, M.J., Roberts, G.O., Ross, J.V. and Vernon, M.C. (2011) 'Networks and the epidemiology of infectious disease', *Interdisciplinary Perspectives on Infectious Diseases*, Article ID 284909, 28 pages.

Darroch, J.N. and Seneta, E. (1965) 'On quasi-stationary distributions in absorbing discrete-time finite Markov chains', *J Appl Probab* **2**, 1:88-100.

Darroch, J.N. and Seneta, E. (1967) 'On quasi-stationary distributions in absorbing continuous time finite Markov chains', *J Appl Probab* **4**, 192-196.

Decreusefond, L., Dhersin, J., Moyal, P. and Tran, V. C. 'Large graph limit for an SIR process in random network with heterogeneous connectivity', *Ann. Appl. Probab.* **22**, no. 2, 541–575.

Diekmann, O., Heesterbeek, H. and Britton, T. (2012) *Mathematical tools for understanding infectious disease dynamics*, Princeton University Press.

Dorogovtsev, S.N. (2010) *Lectures on complex networks*, Oxford Univ Press.

Durrett, R. (2007) *Random graph dynamics*, Cambridge: Cambridge Univ Press.

Durrett, R. and Levin, S.A. (1994) 'Stochastic spatial models: a user's guide to ecological applications', *Philos T Roy Soc B* **343**, 1305:329-350.

Eames, K.T.D. and Keeling, M.J. (2002) 'Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases', *Proc Natl Acad Sci USA* **99**, 20:13330-13335.

Erdös, P. and Rényi, A. (1959) 'On random graphs I', *Publicationes Mathematicae* 6, 290-297.

Esary, J.D., Proschan, F. and Walkup, D.W. (1967) 'Association of random variables with applications' *Ann Math Stat* **38**, 1466-1474.

Ethier, S.N. and Kurtz, T.G. (1986) *Markov processes: characterisation and convergence*, Wiley.

Feld, S.L. (1991) 'Why your friends have more friends than you do', *AJS* **96**(6), 1464-77.

Fine, P.E.M. and Clarkson, J.A. (1982) 'Measles in England and Wales. II. The impact of the measles vaccination programme on the distribution of immunity in the population', *Int J Epidemiol* **11**, 15-25.

Frisch, H.L. and Hammersley, J.M. (1963) 'Percolation and related topics', *Journal of the Society for Industrial and Applied Mathematics* **11**(4), 894918.

Garnett, G.P. and Anderson, R.M. (1996) 'Sexually transmitted diseases and sexual behaviour: insights from mathematical models', *J Infect Dis* **174**, S150-S161.

Gilligan, C.A. and van den Bosch, F. (2008) 'Epidemiological models for invasion and persistence of pathogens', *Annu Rev Phytopathol* **46**, 385-418.

Gillespie, D.T. (1976) 'A general method for numerically simulating the stochastic time evolution of coupled chemical reactions', *J. Comp. Phys.* **22**, 403-434.

Grassberger, P. (1983) 'On the critical behavior of the general epidemic process and dynamical percolation', *Math Biosci* **63**, 2:157-172.

Griffeath, D. (1981) 'The basic contact processes', *Stochastic Processes and their Applications* **11**, 151-185.

Grimmett, G. (2010) *Probability on graphs*, Cambridge: Cambridge Univ Press.

Grimmett, G.R. and Stirzaker, D.R. (1982) *Probability and random processes*, Oxford Univ Press.

Harris, T.E. (1974) 'Contact interactions on a lattice', *Ann Probab* **2**, 6:969-988.

Harris, T.E. (1976) 'On a class of set-valued markov processes', *Ann Probab* **4**, 2:175-194.

Harris, T.E. (1978) 'Additive set-valued markov processes and graphical methods', *Ann Probab* **6**, 3:355-378.

Hethcote, H.W. and Yorke, J.A. (1984) *Gonorrhoea. Transmission dynamics and control* Springer Lecture Notes in Biomathematics. Berlin: Springer.

Hethcote, H.W. (2000) 'Mathematics of infectious diseases', *SIAM Rev* **42**, 4:599-653.

Holley, A. and Liggett, T. (1975) 'Ergodic theorems for weakly interacting infinite systems and the voter model', *Ann Probab* **3**, 4:643-663.

House, T., Davies, G., Danon, L. and Keeling, M.J. (2009) 'A motif-based approach to network epidemics', *B. Math. Biol.* **71**, 1693-1706.

House, T. and Keeling, M.J. (2008) 'Deterministic epidemic models with explicit household structure', *Math. Biosci.* **213**, 29-39.

House, T. and Keeling, M.J. (2010) 'Insights from unifying modern approximations to infections on networks', *J. R. Soc. Interface* **8**, 54, 67-73.

Jacquez, J.A. and Simon, C.P. (1993) 'The stochastic SI model with recruitment and deaths I. Comparison with the closed SIS model', *Mathematical biosciences* **117** (1), 77-125.

Jonkers, A.R.T., Sharkey, K.J. and Christley, R.M. (2010) 'Preventable H5N1 avian influenza in the British poultry industry network exhibit characteristic scales', *J R Soc Interface* **7**, 45:695-701.

Karrer, B. and Newman, M.E.J. (2010) 'A message passing approach for general epidemic models', *Phys. Rev. E.* **82**, 016101.

Karrer, B. and Newman, M.E.J. (2011) 'Competing epidemics on complex networks', *Physical Review E* **84**, 3, 036106.

Keeling, M.J. (1999) 'The effects of local spatial structure on epidemiological invasions', *Proc R Soc Lond B* **266**, 859-867.

Keeling, M.J. and Eames, K.T.D. (2005) 'Networks and epidemic models', *J R Soc Interface* **2**, 295-307.

Keeling, M.J. and Rohani, P. (2008) *Modeling infectious diseases in humans and animals*, Princeton University Press.

Keeling, M.J., Rohani, P. and Grenfell, B.T. (2000) 'Seasonally forced disease dynamics explored as switching between attractors', *Physica D* **148**, 317-335.

Kenah, E. and Robins, J.M. (2007) 'Second look at the spread of epidemics on networks', *Phys Rev E* **76**, 3:036113.

Kephart, J.O. and White, S.R. (1991) 'Directed-graph epidemiological models of computer viruses', *Proc IEEE Symp Security and Privacy*, 343-359.

Kephart, J.O. and White, S.R. (1993) 'Measuring and modeling computer virus prevalence', *Proc IEEE Symp Security and Privacy*, 2-15.

Kermack, W.O. and McKendrick, A.G. (1927) 'A contribution to the mathematical theory of epidemics', *Poc R Soc Lond A* **115**, 700-721.

Kirkwood, J. G. (1935) 'Statistical mechanics of fluid mixtures', *J. Chem. Phys.* **3**, 300.

Kiss, I.Z., Green, D. and Kao, R. (2006) 'The effect of contact heterogeniety and multiple routes of transmission on final epidemic size', *Math. Biosci.* **203**, 124.

Kiss, I.Z., Morris, C.G., Sélley, F., Simon, P.L. and Wilkinson, R.R. (2014) 'Exact deterministic representation of Markovian SIR dynamics on networks with and without loops', *J. Math. Biol.* DOI: 10.1007/s00285-014-0772-0.

Liggett, T. M. (1996) 'Multiple transition points for the contact process on the binary tree', *Ann Probab* **24**, No. 4, pp. 1675-1710.

Liggett, T.M. (1999) *Stochastic interacting systems: contact, voter and exclusion processes*, Berlin: Springer.

Matsuda, H., Ogita, N., Sasaki, A. and Sato, K. (1992) 'Statistical mechanics of populations: the lattice Lotka-Volterra model', *Prog. Theor. Phys.* **88**, 1035-1049.

May, R. M. and Anderson, R. M. (1988) 'The transmission dynamics of human immunodeficiency virus', *Proc. Roy. Soc. Lond. B* **321**(1207):565-607.

McKendrick, A. G. (1926) 'Applications of mathematics to medical problems', *Proc. Edinburgh Math. Soc.* **44**:98-130.

Meyers, L. A., Newman, M. E. J. and Pourbohloul, B. (2006) 'Predicting epidemics on directed contact networks', *Journal of theoretical biology* **240**(3), 400-418.

Miller, J.C. (2009) 'The spread of infectious disease through clustered populations', *J. R. Soc. Interface*, doi:10.1098/rsif.2008.0524.

Miller, J.C. (2011) 'A note on a paper by Erik Volz: SIR dynamics in random networks', *Journal of Mathematical Biology* **62**, 349-358.

Miller, J.C. (2013) 'Cocirculation of infectious diseases on networks', *Physical Review E* **87**, 6, 060801.

Miller, J.C., Slim, A.C. and Volz, E.M. (2011) 'Edge-based compartmental modelling for infectious disease spread', *Journal of The Royal Society Interface* **9**(70), 890906.

Molloy, M. and Reed, B. (1995) 'A critical point for random graphs with a given degree sequence', *Random Struct Algor*, **6**(2-3), 161-180.

Nagy, N. and Simon, P.L. (2013) 'Monte Carlo simulation and analytic approximation of epidemic processes on large networks', *Cent. Eur. J. Math.* **11**, 4, pp 800-815.

Nåsell, I. (1995) 'The threshold concept in stochastic epidemic and endemic models', *Epidemic Models: Their Structure and Relation to Data*, ed Mollison D. Publ Newton Inst. Cambridge: Cambridge Univ Press. pp. 71-83.

Nåsell, I. (1996) 'The quasi-stationary distribution of the closed endemic SIS model', *Adv Appl Prob* **28**, 895-932.

Neal, P.J. (2008) 'The SIS great circle epidemic model', *J Appl Prob* **45**, 513-530.

Newman, M.E.J. (2002) 'Spread of epidemic disease on networks', *Phys Rev E* **66**, 016128.

Newman, M.E.J. (2010) *Networks, an introduction*, Oxford University Press, Oxford, UK.

Newman, M.E.J., Strogatz, S.H. and Watts, D.J. (2001) 'Random graphs with arbitrary degree distributions and their applications', *Phys. Rev. E* **64**, 026118.

Pastor-Satorras, R. and Vespignani, A. (2001a) 'Epidemic dynamics and endemic states in complex networks', *Phys Rev E* **63**, 066117.

Pastor-Satorras, R. and Vespignani, A. (2001b) 'Epidemic spreading in scale-free networks' *Phys Rev Lett* **86**, 3200-3203.

Rand, D. A. (1999) 'Correlation equations and pair approximations for spatial ecologies', *Advanced ecological theory: principles and applications*, 100.

Ross, R. (1911) 'Some quantitative studies in epidemiology', *Nature* **87**:466-467.

Schonmann, R.H. (1985) 'Metastability for the contact process', *J Statist Phys* **41**, 445-464.

Sharkey, K.J. (2008) 'Deterministic epidemiological models at the individual level' *J. Math. Biol.* **57**, 311-331.

Sharkey, K.J. (2011) 'Deterministic epidemic models on contact networks: correlations and unbiological terms', *Theor. Popul. Biol.* **79**, 115-129.

Sharkey, K.J., Bowers, R.G., Morgan, K.L., Robinson, S.E. and Christley, R.M. (2008) 'Epidemiological consequences of an incursion of highly pathogenic H5N1 avian influenza into the British poultry flock', *Proc R Soc B* **275**, 1630:19-28.

Sharkey, K. J., Fernandez, C., Morgan, K.L., Peeler, E., Thrush, M., Turnbull, J.F. and Bowers, R.G. (2006) 'Pair-level approximations to the spatio-temporal dynamics of epidemics on asymmetric contact networks', *J Math Biol* **53**(1), pp. 61-85, ISSN 0303-6812.

Sharkey, K.J., Kiss, I.Z., Wilkinson, R.R. and Simon, P.L. (2013) 'Exact equations for SIR epidemics on tree graphs', *Bulletin Mathem. Biol.* DOI: 10.1007/s11538-013-9923-5.

Sharkey, K.J. and Wilkinson, R.R. (2015) 'Complete hierarchies of SIR models on arbitrary networks with exact and approximate moment closure', arXiv:1501.06353 [q-bio.PE].

Simonis, A. (1996) 'Metastability for the d-dimensional contact process', *J Statist Phys* **83**, 1225-1239.

Volz, E. (2008) 'SIR dynamics in random networks with heterogeneous connectivity', *Journal of Mathematical Biology* **56**, 3:293-310.

Weiss, G.H. and Dishon, M. (1971) 'On the asymptotic behaviour of the stochastic and deterministic models of an epidemic', *Math Biosci* **11**, 261-265.

Wilkinson, R.R. and Sharkey, K.J. (2014) 'Message passing and moment closure for susceptible-infected-recovered epidemics on finite networks', *Phys. Rev. E* **89**, 022808.

Wilkinson, R.R. and Sharkey, K.J. (2013) 'An exact relationship between invasion probability and endemic prevalence for Markovian SIS dynamics on networks', *PLoS ONE* **8**, e69028.

Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Liu, W.C., Mellorand, D.J. and Thomas, M.R. (2005) 'Epidemiological implications of the contact network structure for cattle farms and the 20-80 rule', *Biol. Lett.* **1**, 350-352.