Disease Emergence and Dynamics on Biologically Motivated Contact Networks

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

Simon Peter Johnstone-Robertson

BSc (Mathematics and Physics), Nelson Mandela Metropolitan University
BSc Honours (Physics), Nelson Mandela Metropolitan University
PDS (Mathematical Sciences), University of Stellenbosch
MSc (Mathematics), University of Stellenbosch

School of Science
College of Science, Engineering and Health
RMIT University

December 2017

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship. This research was also undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government, together with a National Science Foundation/National Institute of Health Ecology and Evolution of Infectious Diseases Program award NIH R01 GM105246-01, and an RMIT PhD Scholarship. Collaborators of the work presented in Chapter 2 were supported by a Department of Agriculture, Australian Government grant: WEDPP 2012/2014.

Simon Peter Johnstone-Robertson
June 8, 2018

Acknowledgements

Martijn Van der Merwe, Johannes Kotzerke, and Prof. Jared Cole for assistance with submitting MATLAB code to the supercomputer facilities of VPAC and NCI. The rabies simulations in Chapter 2 would never have finished without your help—so thank you!

Rita Nygaard (http://www.clipartqueen.com/) who generously granted permission to use the mouse silhouette she designed in the tick-host networks illustrated in Chapter 3.

Prof. Maria Diuk-Wasser for providing mouse tick-burden data to analyse and for so kindly hosting Dr. Stephen Davis and myself whilst we visited her research group in New York.

The SMGS Writing Group comprising Jacobien Carstens, Jess Liebig, Jess Dunn, Nigel Clay, David Ellison, amongst others. Your comments always improved the writing I was brave enough to show you.

Isaac Mwangari for helping me format several parts of this thesis.

Dr. Stephen Davis and Prof. Lewi Stone for their generosity, guidance, and friendship over the last four years, and for also improving my writing.

My dear friend Georgie Crowden for proofreading this thesis and for baking me many an apple pie to encourage me along the way.

My mum Penelope Johnstone-Robertson for teaching me the value of an education and for always believing in me.

My wife Heather Johnstone-Robertson for her unfailing love, encouragement, and sacrifice these last four years. Without you this work would not have been possible.

To Heather,
May our adventures with Jesus
continue forever.

It is the glory of God to conceal a matter;
to search out a matter is the glory of kings.
Proverbs 25:2 (NIV)

Abstract

Infectious disease transmission requires that epidemiologically relevant contact occurs between infectious and susceptible individuals. Thus, for mathematical models to accurately predict disease emergence and dynamics they must incorporate the contact patterns responsible for transmission. In this context, this thesis investigates how the level of contact detail included in an infectious disease model influences its predictions.

Three models are considered. The first investigates infections spreading through territorial populations, with potential canine rabies spread in Australian wild dogs a case study. Two factors governing wild dog contacts are considered: geographic distance and heterogeneous wild dog behaviour. Not including spatial constraints results in a model that overestimates the probability of an epidemic and that fails to generate the outcome 'rate of spread'. Conversely, not incorporating heterogeneous dog behaviour results in a model that underestimates the probability an epidemic will occur.

The second model investigates tick-borne pathogen spread between ticks and vertebrate hosts. Key features of tick feeding behaviour include: tick aggregation on hosts, co-aggregation of larval and nymphal ticks on the same hosts, and co-feeding. Co-aggregation increases R_0. Models failing to incorporate tick co-aggregation will therefore underestimate the likelihood of pathogen emergence, especially in geographic regions and seasons where larval burden is high and for pathogens mainly transmitted during co-feeding.

The third model investigates the effect of clustering (triangle and square contact patterns) on the spread of infection through social networks. Clustering reduces R_0 and the magnitude of the reduction increases with higher transmission probabilities. Models that fail to incorporate clustering will overestimate the likelihood of disease establishment, especially for highly transmissible diseases.

In conclusion, the three disease models collectively reveal model predictions are improved and additional outcomes are generated by the inclusion of realistic host contact patterns. These findings reinforce the value of incorporating biologically-faithful contact patterns into infectious disease models.

Contents

List of figures

3 Tick-borne pathogens

4 Clustered social networks

List of tables

Chapter 1

Introduction

1.1 The disease model spectrum

A requirement for the transmission of any infectious disease is that epidemiologically relevant contact occurs between infectious and susceptible hosts [1–4]. Understanding the contact patterns of hosts is therefore fundamental to understanding infection transmission as a whole. In particular, for mathematical models to accurately predict the disease trends and dynamics observed in populations it is imperative they capture the underlying contact patterns responsible for transmission. At the same time, this need is counterbalanced by the desire to keep models simple so that they and their results can be solved and interpreted [5]. Availability of data also plays a role in governing the level of complexity with which contact can be modelled [6]. In the case of rabies, for example, it is both unethical and logistically difficult to observe the contacts of free-roaming rabid dogs since such dogs pose an immediate threat to human lives. These factors have resulted in a spectrum of disease models (Figure 1.1) ranging from those that make simplistic assumptions about contact to those that model infection transmission on real, dynamic contact networks constructed from detailed, empirical host-interaction data [5, 7, 8].

The simplest approach to modelling contact is to assume the host population mixes randomly (so that every susceptible has equal risk of infection from any infectious case), the host population is homogeneous (so that each host has the same number of contacts), and the contact rate is either constant (frequency dependent) or directly proportional to host population density (density dependent) [9]. Compartmental models which typically incorporate these assumptions have been applied to a plethora of diseases including influenza, measles, rabies, HIV, and foot-and-mouth disease [10–15]. Empirical

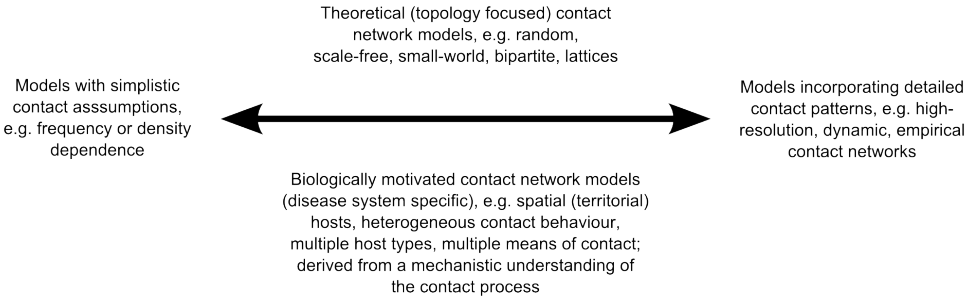

Figure 1.1 – The disease model spectrum. Models vary between those that make simplistic assumptions about host contact patterns to those that investigate the spread of infection on empirical contact networks. Theoretical studies that focus on the impact of contact network topology on the spread of infection lie somewhere between these two extremes. Mechanistic models allow the underlying biology of a disease system (i.e. those factors which govern what an epidemiologically-relevant contact is and how such contact comes about) to inform the type of contact network to be considered and what its properties are to be, e.g. whether the network is clustered, bipartite, directed/undirected, or scale-free/random, etc.

evidence supporting the use of density or frequency dependent contact rates, however, is generally lacking, and the evidence that does exist suggests a non-linear function may be more suitable [6, 16–18]. Model predictions of disease dynamics, intervention impact, and whether or not a threshold host population density exists are known to depend on the functional form assumed for the contact rate [16]. This, together with the lack of empirical support for the two traditional contact rate functions, has resulted in several alternatives having been proposed in an attempt to improve model accuracy [6, 16]. Nevertheless, these alternative functions generally vary asymptotically between density dependence at low host population density and frequency dependence (corresponding to a saturation of contacts) at higher densities [16, 17]. Fundamentally speaking then, there is no increase in the level of contact detail included in these models since a single function is deemed sufficient to capture the contact patterns of the entire host population.

At the other end of the infectious disease model spectrum are models that take contact patterns into account using high levels of detail. The most complex models incorporate high-resolution, dynamic, empirical contact networks where nodes represent hosts, an edge between two nodes represents the contact between them, the edges are weighted, and contacts occur chronologically (i.e. both the timing and order of the contacts are considered). Edges are typically weighted when the cumulative contact duration (or the Euclidean distance) between hosts is as important in determining the prob-

ability of transmission as whether or not hosts make contact at all [5, 7, 8]. The order of contacts is important because transmission can only occur between two hosts if they contact each other after either one becomes infectious. Similarly the timing of contacts needs to be considered because an infectious host can only infect its susceptible neighbours whilst it remains infectious (i.e. during its infectious period) [5, 19]. Depending on the disease and host population, edges in a network may also be directional, indicating which of a host's neighbours the host could infect were it to become infectious [20]. A concern associated with these complex models, however, is whether their results are only applicable to the population in which the empirical contact network was observed or whether they can be generalized [5, 7].

More widely applicable, but nevertheless highly detailed, contact networks have been inferred from stochastic, agent-based models [21, 22]. Here, the networks are assembled using population-level data, e.g. census, demographic, travel/migration, employment and school attendance data, to randomly assign hosts to locations they frequent in their communities during 12-hour periods (day/night). Each location is populated in such a way as to maintain spatial integrity (hosts generally go to the nearest schools and hospitals) as well as the known age-distributions and population sizes at each location. An important feature of these spatio-temporal contact networks is that nodes at multiple locations may represent the same host (i.e. there is one node for each location a host might visit, e.g. household, work, school, hospital, etc.). A host, however, may only occupy one of these locations during every 12-hour period. The stochasticity of the model therefore has to do with which location each host visits during each 12-hour period. This is affected by factors such as the host's age, their infection status, and whether they have to remain home to take care of a sick child or not. In public locations (e.g. schools and work) nodes that frequent the same location are typically grouped into cliques of finite size (in which every node is connected to every other node within the clique). Nodes that visit the same location but are not in the same clique are not permitted to contact one another in an attempt to reflect the reality that not every host contacts every other host in these types of locations. In general then, one can view the contact network for each 12-hour period as a sub-graph of the complete, but static, network that represents all potential contacts between hosts at all locations. In addition, the model is intrinsically dynamic since the contact network sub-graph is different for every 12-hour period, especially as the disease of interest progresses through the population.

A drawback common to both stochastic, agent-based models and those based on empirical contact networks is that they are resource and data intensive [5, 7, 19–22]. Furthermore, the level of contact detail required to

accurately model the spread of infectious diseases is unclear at present, with different studies having drawn contrasting conclusions [5, 19]. Machens et al. [5] propose "a contact matrix of probability distributions that takes into account the heterogeneity of contact durations between (and within) classes of individuals [i.e. hosts]" sufficiently captures host contact patterns. This in turn implies the timing and order of contacts, together with the exact structure of the observed contact network (i.e. 'who contacts who'), are not important. Conversely, Rocha et al. [19] concluded the timing of contacts is essential to understanding the spread of sexually transmitted diseases. Alternatively, spatially-dependent disease systems (such as the plague and foot-and-mouth disease) have lent themselves to the use of spatial kernels (i.e. distributions of flea movements between families of gerbils for plague and distances cattle are transported between farms for foot-and-mouth disease) for modelling contact between hosts and found this sufficient to accurately replicate observed disease trends and dynamics [17, 23–25]. Thus, the required level of contact detail appears to depend on the disease and host population being considered, as well as the underlying biology of the problem.

In spite of the uncertainty surrounding the level of contact detail required in disease models, the increasing interest in, and availability of, detailed contact patterns has led to the theoretical study of the influence of network topology on the spread of infection (Figure 1.1). This has also been motivated by the fact that broad classes of real networks (e.g. social, biological, communication, etc.) frequently have similar properties [26–28]. One such property is that real networks are often 'small-world' having relatively low average path length and a high clustering coefficient [20]. This phenomenon was first observed for social networks by Milgram [29]. Small-world networks are epidemiologically important because infections have been shown to spread substantially faster through them than those with other topologies [30, 31]. Random network models (e.g. Erdos-Renyi networks [32]) were initially used as theoretical representations of small-world networks, however they fail to capture the high levels of clustering observed in real networks. For this reason Watts and Strogatz [33] proposed a new network generative procedure that can be tuned to produce theoretical networks that fall anywhere between regular lattices (with high levels of clustering but large average path length) and random networks (with low to moderate levels of clustering and small average path length). This remains a popular procedure for generating small-world networks, and for these networks the threshold above which an epidemic will occur has been derived [30]. In addition, Kuperman and Abramson [31] have shown that the behaviour of the endemic state (i.e. the proportion of nodes infected) over time on these networks varies between small irregular fluctuations (for regular lattices) and large-amplitude oscilla-

tions (for random networks).

A subset of small-world networks that have been studied in detail are those with scale-free degree distributions [28]. The degree distribution of these networks is typically described by a power-law $P(k) \approx k^{-\alpha}$, where $P(k)$ denotes the proportion of nodes adjacent to k other nodes by means of an edge (i.e. of degree k) and where $\alpha \in (2, 3]$. Scale-free networks have attracted substantial attention for two main reasons. First, they capture the contact heterogeneity frequently observed in empirical networks [26–28, 34], whereby a handful of hosts account for the majority of contacts whilst the remaining hosts have relatively few contacts. Secondly, they were originally shown to lack an epidemic threshold [35]. A consequence of such a finding is that an epidemic can occur for any transmission probability, and upon its introduction an infection will persist thereafter indefinitely. Importantly, May and Lloyd [36] showed this to be true only for networks of infinite size by deriving an epidemic threshold for finite scale-free contact networks. A procedure for generating theoretical contact networks with power-law degree distributions was first described by Barabási and Albert [27]. In it they proposed the underlying mechanism generating scale-free networks was a combination of 'preferential attachment' and growth over time, where nodes being added to a network would preferentially connect with (i.e. contact) those nodes already in the network with high degree. More recently, however, mechanisms that generate scale-free degree distributions for static networks have been identified [37, 38]. Perhaps the best known of these is the good-get-richer model of Caldarelli et al. [37], where nodes are assigned a 'fitness' and an edge between any pair of nodes is generated with a probability proportional to the fitness of the two nodes involved.

Some other network topologies that have been studied theoretically include adaptive (or behavioural) networks [39–41], bipartite networks [42, 43], and lattices lacking long-range connectivity [44, 45]. These will not be discussed any further other than to say that adaptive networks permit a host to adjust its contact patterns in response to both its own infection status and that of its contacts, bipartite networks only permit contact (edges) between hosts (nodes) from distinct groups (e.g. cattle farms and the markets they sell livestock to or buy livestock from [20]), and lattices only permit contact between hosts that are nearest neighbours. It is safe to say then that the general success of theoretical studies into the influence of contact network topology on the spread of infection has been, at least in part, due to the broad range of network topologies that can, and have been, studied using this approach.

In spite of this success, one of the drawbacks to theoretical models is that they are often abstract in nature with parameters that have no obvious

biological interpretation. There is a complementary class of models, though, that relates the contact between hosts to measurable physical, biological, and ecological properties, namely mechanistic models. The differences between theoretical and mechanistic models come about because of the different purposes for which they are each constructed. Mechanistic models are constructed with the aim of incorporating biological reality and they accomplish this by being explicit about how the properties of the pathogen and host population(s) affect the contacts and transmission processes. Theoretical models, on the other hand, are often constructed to investigate the effect that just a single model parameter (i.e. topological feature) has on disease dynamics at a time. Another way to distinguish between these two classes of models is that theoretical models tend to focus on population-level features of contact whereas mechanistic models tend to focus on those factors that cause contact at the individual level.

These differences are perhaps best illustrated by comparing theoretical lattices and mechanistic models that generate spatial networks[1]. In both cases the influence of spatial constraints (i.e. distance) on infection transmission is of interest. With lattices, nodes are spatially arranged in a grid and the spatial constraints on which nodes come into contact is artificially determined by whether or not two nodes are nearest neighbours. Typically, therefore, an arbitrary fixed number of 4 or 6 nearest neighbour contacts is imposed on all nodes in the network. In contrast, mechanistic spatial networks assign locations to nodes such that they are randomly distributed over a landscape. Thereafter, the spatial constraints on which nodes come into contact is captured by a kernel that typically reflects a distribution of distances hosts have been observed to move in reality. Defining contact in

[1]The difference between theoretical and mechanistic models isn't simply limited to contact networks but also applies to other types of mathematical constructions as well. For example, consider metapopulation models in which the progression of an infection through multiple interacting sub-populations is modelled. As a starting point disease dynamics are typically modelled within each sub-population using compartmental ordinary differential equations (ODEs). The interaction (which leads to transmission) between sub-populations is then captured by a set of model parameters that link the ODEs for each sub-population. In theoretical models these parameters are referred to as coupling parameters; however, these parameters have no obvious biological interpretation (i.e. it isn't immediately clear how to obtain empirical estimates for these parameters from field measurements). In contrast, mechanistic models relate the transmission of infection between sub-populations to the rate of introduction or movement of hosts from one sub-population to another, both of which are measurable physical quantities. Thus, the difference between theoretical and mechanistic models boils down to the purpose for their construction; theoretical models focus on understanding the effect of a single (set of) model parameter(s) on disease dynamics, whereas mechanistic models aim to emulate biological reality with parameters that have a clear physical, biological, or ecological interpretation.

this way means that most contacts are formed between nearest neighbours, but unlike lattices there are still a few contacts that occur over longer distances (as is observed in many empirical contact networks). Another feature of mechanistic spatial networks that sets them apart from theoretical lattices is that the (individual-level) probability any two particular nodes come into contact is also defined in terms of measurable individual host properties, e.g. the rate at which a host traverses its home range (when a node represents an animal) or the number of animals resident on a farm (when a node represents a farm). Importantly, this allows the effect of individual host heterogeneity on the spread of infection to be investigated at the same time as that of spatial constraints.

In summary then, the complementary nature of theoretical and mechanistic models means that together they can be viewed as bridging the gap between models that make simplistic contact assumptions and those that incorporate high-resolution, empirical contact networks (Figure 1.1).

1.2 Research question

The central research question to be considered in this thesis is,

> How does the level of contact detail included in an infectious disease model influence its predictions in terms of disease emergence and dynamics?

This question is important since its answer determines the level of contact detail required to accurately predict the disease trends and dynamics observed in real populations. In this thesis, three infectious disease models will be constructed to investigate the research question. The first mechanistic model will describe the spread of infection through a territorial host population, specifically canine rabies in Australian wild dogs, whilst the second will describe the spread of tick-borne pathogens between ticks and their vertebrate hosts. The third model, a theoretical model, will investigate the influence of clustering (triangular and square contact patterns) on the spread of infection through social networks. The specific features of each model, and how it relates to modelling contact, are described below in the context of the thesis layout; a more detailed description can be found at the start of each chapter.

1.3 Thesis layout

In Chapter 2, the spread of infection through a territorial host population is considered, with the potential spread of canine rabies through the wild dog population of Australia as a case study. Australian wild dogs are territorial by nature but also occupy an expansive landscape. Consequently, to model the contact process, a function for the contact rate that takes both the distance between wild dogs and their individual inclination to make contact with other wild dogs into account is proposed. Thereafter, the probability rabies will spread spatially, how quickly it will do so, and how each of the model outcomes vary as a function of the input variables is investigated. A global sensitivity analysis is performed to identify the model input variables most influential in determining whether canine rabies will spread through the Australian wild dog population.

In Chapter 3, the spread of pathogens between ticks and their vertebrate hosts is investigated. The key features of contact relevant to transmission include aggregation (whereby the majority of ticks of a given life-stage feed on a minority of the hosts) and co-aggregation (whereby the majorities of the different life-stages of ticks feed on the same minority of hosts). Additional sources of biological complexity include the multiple transmission routes between ticks (i.e. both tick-to-host-to-tick and tick-to-tick transmission is possible) and that different life-stages of ticks quest at different times of the year. A novel contact network model is presented that captures the aforementioned contact properties, and in particular tick aggregation and co-aggregation. Analytic formulae for the epidemic threshold parameter, R_0, are derived for the proposed contact network model. The relative effect of co-aggregation on the establishment of tick-borne pathogens is quantified by comparing the value of R_0 when co-aggregation is present in tick-host networks to when it is absent. Last of all, simulations of Lyme disease transmission are used to visualize the relationship between R_0 and the extent to which ticks aggregate and co-aggregate on vertebrate hosts.

In Chapter 4, the effect of clustering on the spread of infection through social networks is investigated. A feature common to most social networks is that if one individual contacts two other individuals, these two other individuals are likely to be contacts of each other. In network terminology, this triangular contact pattern is referred to as clustering. Higher-order forms of clustering can also be observed, e.g. squares involving four hosts. At present, the literature is inconclusive with respect to the impact of clustering on the spread of infection. Some studies conclude that clustering raises the value of R_0, thereby increasing the probability of disease establishment, whilst others

suggest the opposite is true. This dilemma is resolved by deriving an exact formula for R_0 that takes all triangle and square clustering effects into account, and also competing interests resulting from multiple infectious individuals attempting to infect the same susceptible individuals. Simulations of infection transmission are performed on Erdos-Renyi random networks to verify the derived formula and illustrate the influence that clustering has on the early stages of an epidemic.

Finally, in Chapter 5 the results obtained for each infectious disease model in Chapters 2–4 are discussed in light of the research question posed in Section 1.2.

Chapter 2

Diseases in territorial host populations: canine rabies in Australian wild dogs as a case study

2.1 Introduction

In this chapter, the potential spread of canine rabies through the wild dog population of Australia is modelled as an illustrative example of an infection spreading through a territorial host population. For canine rabies the overwhelming means by which infection is passed from one dog to another is by bite. Thus an implicit requirement for an epidemiologically relevant contact to occur is that two wild dogs frequent the same location, at the same time. From this it follows that any factor that influences the locations a wild dog visits, as well as when and how often, could also influence where, when, and how often rabies transmission occurs, and ultimately how disease dynamics play out over space and time.

For the Australian wild dog population three such factors will be considered. The first is that wild dogs are territorial by nature, whilst the second is that wild dogs occupy an expansive landscape. These two factors together imply the spatial scale of movement of wild dogs is smaller than the size of the landscape they occupy, such that, in practice, some dogs never come into contact with each other as they never frequent the same location. That is to say, the territorial nature of wild dogs implies contacts tend to occur between dogs occupying neighbouring territories. The third factor is heterogeneous social behaviour, by which is meant that some dogs are shy and tend to avoid contact whilst other dogs are highly sociable. Heterogeneity influences how often epidemiologically-relevant contacts occur since greater levels of heterogeneity imply more highly sociable dogs, and more highly sociable dogs imply more contact with more dogs.

To mechanistically model the contact process at the individual (wild dog) level whilst taking each of the three factors above into account, a function for the rate at which any pair of wild dogs come into contact is proposed that depends on the geographic distance between the two dogs and their individual inclination towards making contact. Importantly, each parameter in the proposed contact rate function is shown to have a clear biological interpretation such that it has the potential to be measured in the field.

2.1.1 Chapter layout

The work presented in this chapter was conducted in collaboration with Prof. Michael Ward, Prof. Peter Fleming, and Dr. Stephen Davis (my senior PhD supervisor), and was published as a research article in *PLoS Neglected Tropical Diseases* [46]. The main research question posed was whether a rabies epidemic could occur in the Australian wild dog population given the

introduction of a rabid dog. This differs from the broad research question considered in this thesis (Section 1.2)—how the level of contact detail included in an infectious disease model influences predictions about disease emergence and dynamics. Nevertheless, one is able to infer the answer to the thesis research question by considering the results presented in the research article. For this reason, this chapter has been structured with a layout similar to that of the research article [46]. In particular, the context of the modelling is set out by discussing the global burden of rabies, what makes rabies pathogenesis unique from that of other diseases, and the risk of rabid dog incursion into Australia. Thereafter a function for the rate at which two wild dogs come into contact is proposed as the basis of a mechanistic percolation model used to simulate the spread of rabies through dynamic contact networks for the Australian wild dog population. The results of the model are presented next, which include estimates for the speed at which rabies is expected to spread geographically and the probability an epidemic will occur given the introduction of an infected dog. A global sensitivity analyses is also presented that identifies the model input variables most important in determining whether a rabies epidemic will occur. Last of all, the implications of these findings, as well as the underlying model assumptions, are discussed in the context of the published contact network and rabies literature. Discussion of the implications of these findings in terms of the broad thesis research question is reserved until Chapter 5.

2.2 Risk of rabies incursion into Australia

Rabies is a zoonotic disease caused by a virus of the genus Lyssavirus. Annually, it causes an estimated 61,000 human deaths in over 150 countries and territories where it is endemic [47–49]. The canine strain of the virus is the most widely distributed globally [50], with rabid dogs accounting for the vast majority (up to 99%) of human rabies infections and deaths [49, 51].

A key feature of rabies that assists its spread and persistence is the long incubation period [52, 53]; it can vary from 10 days to 6 months, but for most cases lasts between 2 weeks and 3 months [48]. The extended incubation period often allows rabies to remain undetected and enter new areas [54, 55], and is one reason the eventual incursion of canine rabies into Australia is considered to be likely [48, 56].

Australia is historically free of canine rabies, with only the bat strain (Australian bat lyssavirus) endemic on the continent [47, 48, 56, 57]. In South-East Asia, though, canine rabies is endemic and currently spreading eastward along the Indonesian archipelago such that it is now less than 300 km from the northern Australian border [48, 54–58]. Two locations have been identified as probable entry points, namely Arnhem Land in the Northern Territory and Cape York Peninsula in Queensland [48]. Canine rabies introduction is anticipated to occur via a sub-clinically infected dog illegally brought into the country by means of a fishing vessel, pleasure craft, or boat continuing cross-cultural traditions established centuries ago [59, 60].

Incursion alone, however, is insufficient for disease establishment. The index case will also have to contact, bite, and successfully transmit the virus to at least one other resident dog. Several factors make this first transmission event possible: (1) although much of northern Australia is largely uninhabited, many of the remote communities are on or close to the coast, (2) free-roaming dogs are common in these communities [61, 62], and (3) wild dogs, comprising mostly dingoes and their hybrids with free-roaming domestic dogs, are ubiquitous across northern Australia [63, 64].

Sparkes et al. [65] have noted that Australia stands to lose much should canine rabies become established in its wild dog population. The first and most significant loss could be that of human life, especially in peri-urban areas where wild and domestic dogs frequently coincide and the likelihood of human exposure is greatest. There would also be economic losses due to a reduction in ecotourism, rabies-related livestock losses, and the large-scale costs associated with mass vaccination of domestic dogs and wildlife, and the administration of post-exposure prophylaxis to humans. Perhaps harder to quantify, but nevertheless important, is the impact rabies would have

on the Australian nation's affinity for, interaction with, and conservation of wildlife. Australia is renowned for its unique fauna with which people frequently interact. This forms a key component to the national image as a whole. The presence of rabies in wild dogs would lead to greater restraint on wildlife interactions due to the fear of potential rabies exposure. This is especially true should other native wildlife besides wild dogs be at risk of canine rabies infection. Consequent to each of these foreseeable losses, risk assessment for the sustained transmission of canine rabies within the Australian wild dog population has been identified as a national research priority [56].

Spatial rabies transmission in small ($\leq 2.2\,\mathrm{km}^2$), remote communities of northern Australia with high densities of domestic dogs ($\geq 137\,\mathrm{dogs/km}^2$) has been modelled [66] but there are no existing models for wild dogs which are territorial by nature and occupy an expansive Australian landscape at much lower densities. A stochastic transmission network (percolation) model for the spatial spread of canine rabies through the Australian wild dog population is presented here that allows the probability that a canine rabies epidemic will occur to be estimated, given its introduction, along with the geographic rate at which rabies will spread. A global sensitivity analysis is also conducted to identify where knowledge is most critical for predicting model outcomes.

2.3 Wild dog contact networks

The spread of canine rabies through the Australian wild dog population was simulated by implementing the model of Davis [67]. Each simulation began by distributing nodes (representing the centroids of potential wild dog home ranges) uniformly at random across a landscape of width 250 km and height 125 km (Figure 2.1). The density of nodes distributed in each simulation was calculated using the formula derived in Section 2.3.1.

Because the incubation period of rabies is of the order of weeks to months, and because wild dogs inhabit an expansive landscape, the timescale over which rabies was expected to spread was of the order of months to years. For this reason wild dog demographic processes were included in the model by allowing home ranges to become vacant through natural mortality and thereafter become reoccupied through wild dog movement and recruitment. Each node was therefore either occupied by a wild dog or unoccupied at any given moment during a simulation.

When a node was occupied, its location represented the average position of the dog inhabiting the corresponding home range. Consequently, wild

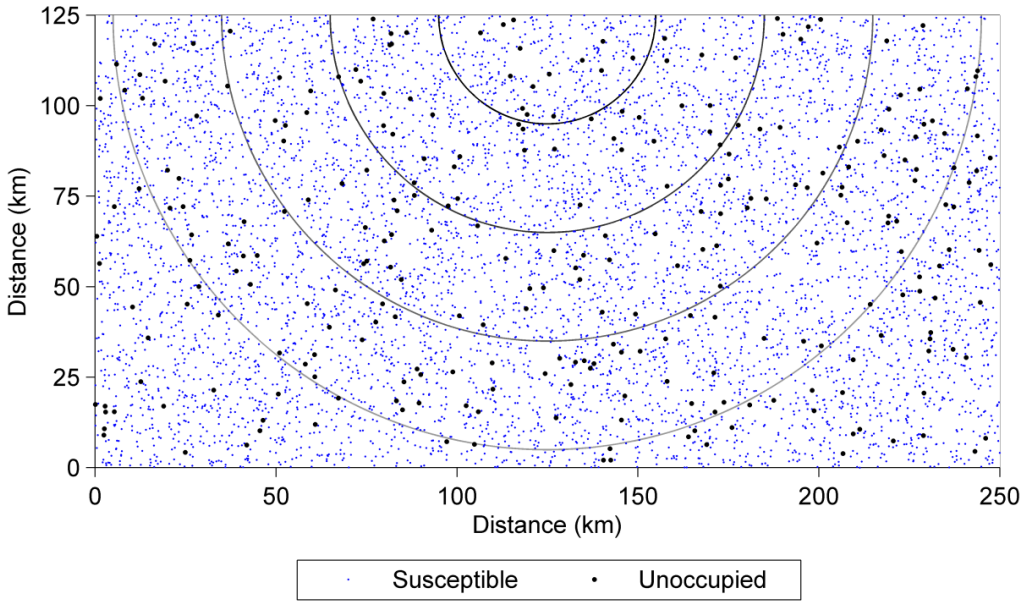

Figure 2.1 – The distribution of susceptible wild dogs (occupied nodes) prior to canine rabies introduction. Each node, representing the centroid of a potential wild dog home range, is either (blue) occupied by a susceptible wild dog or (bold black) unoccupied. An occupied node's location corresponds to the average position of the dog that occupies the home range. Only two demographic processes were assumed to be at work prior to the introduction of a sub-clinically infected (exposed) index case, namely death due to natural causes (rendering occupied nodes unoccupied) and recruitment (rendering unoccupied nodes re-occupied by susceptible wild dogs). After the introduction of a sub-clinically infected index case (near the centre of the upper boundary of the landscape) an epidemic was deemed to have occurred if the infection spread spatially beyond predetermined milestone distances of 30, 60, 90, and 120 km (dark to light grey semicircles respectively).

dogs were not considered stationary at these locations. Instead, each dog makes contact with other dogs over time by way of their normal everyday movement across the landscape when foraging, finding water, taking shelter, communicating, finding a mate, and raising young [48, 68, 69]. The rate at which two dogs i and j contact one another, k_{ij}, was proposed to be a function of the Euclidean distance, s_{ij}, between the two dogs' mean positions (nodes), as well as their individual inclination towards making contact with other dogs (sociability), denoted x_i and x_j respectively. Specifically, the contact rate was given by

$$k_{ij} = \lambda^2 e^{-\lambda s_{ij}} x_i x_j, \tag{2.1}$$

where λ (units: km^{-1}) is a spatial parameter that captures the scale over which distance between two dogs affects their rate of contact. For large values of λ the maximum contact rate ($s_{ij} = 0$) is high and the contact rate declines sharply with distance (Figure 2.2) such that dogs effectively only come into contact with their nearest neighbours. Conversely, for small values of λ the maximum contact rate is low but dogs come into contact with other dogs that are far away almost as often as their nearest neighbours. Equation (2.1) has the property that the average number of contacts a dog has with other dogs over some period of time is independent of λ [67]. Another appealing feature, revealed by dimensional analysis (see Section 2.3.2), is that a dog's sociability, x_i, can be related to the area of land it traverses per day.

2.3.1 Deriving the density of nodes to be distributed

Before the introduction of rabies, nodes, representing the centroids of potential wild dog home ranges, are set to be either occupied by susceptible wild dogs or unoccupied (Figure 2.1). If it is assumed that:

1. Only two demographic processes are at work prior to rabies introduction, namely death due to natural causes (rendering occupied nodes unoccupied) and replacement (rendering unoccupied nodes re-occupied by susceptible wild dogs),

2. The lifespan of wild dogs and the replacement period are both exponentially distributed (with constant mean values), and

3. The density of susceptible wild dogs (occupied nodes) prior to rabies introduction is at a non-trivial equilibrium,

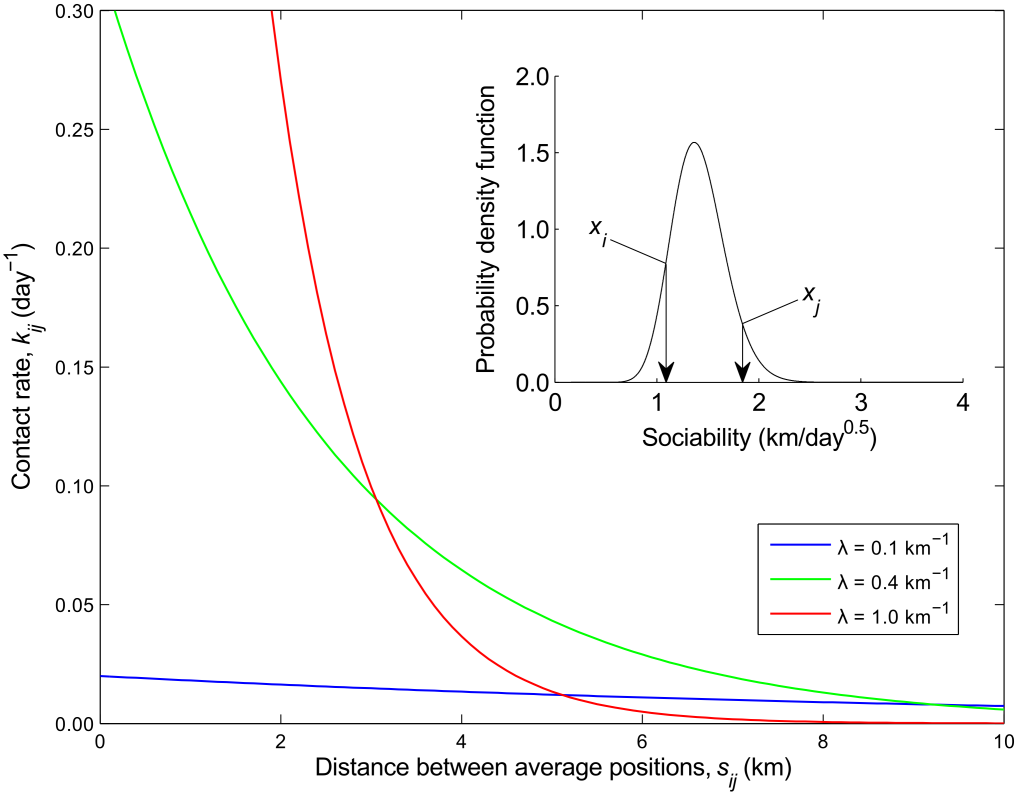

Figure 2.2 – The proposed wild dog contact rate as a function of geographic distance. The contact rate, k_{ij}, of two wild dogs i and j, with sociabilities $x_i = 1.09\,\mathrm{km/day}^{0.5}$ and $x_j = 1.84\,\mathrm{km/day}^{0.5}$, as a function of the distance between their average positions, s_{ij}, for three values of the spatial scale parameter, λ: (blue) $0.1\,\mathrm{km}^{-1}$, (green) $0.4\,\mathrm{km}^{-1}$, and (red) $1.0\,\mathrm{km}^{-1}$. The inverse of the contact rate is equal to the mean time between contact events such that a contact rate of $k_{ij} = 0.2\,\mathrm{day}^{-1}$ implies dogs i and j come into contact on average once every 5 days. (Inset) The wild dog sociability distribution (gamma distribution: shape parameter = 30, scale parameter = $\left(\frac{1}{30}\right)\sqrt{2}\,\mathrm{km/day}^{0.5}$) from which x_i and x_j were sampled and that was also used to generate the transmission network realization in Figure 2.3. The mean value of this distribution, given by the product of the shape and scale parameters, equates to wild dogs traversing $2\,\mathrm{km}^2/\mathrm{day}$ on average.

then an analytic relationship between the density of nodes and the density of susceptible wild dogs (occupied nodes) can be derived. To do this, let

$$\begin{aligned}
\Omega(t) &= \text{the density of occupied nodes (susceptible wild dogs) at time } t,\\
\Upsilon(t) &= \text{the density of unoccupied nodes at time } t,\\
N(t) &= \text{the density of nodes (occupied and unoccupied) at time } t\\
&= \Omega(t) + \Upsilon(t), \qquad (2.2)\\
\mu &= \text{the natural mortality rate}\\
&= 1/\left(\text{mean wild dog lifespan}\right), \text{ and}\\
\rho &= \text{the replacement rate}\\
&= 1/\left(\text{mean replacement period}\right).
\end{aligned}$$

From assumptions 1 and 2 it follows that the behaviour over time of the density of occupied nodes (for a landscape of fixed area) is described by

$$\frac{d\Omega}{dt} = \rho\Upsilon - \mu\Omega. \qquad (2.3)$$

Substituting Equation (2.2) into Equation (2.3) yields

$$\frac{d\Omega}{dt} = \rho N - (\rho + \mu)\Omega. \qquad (2.4)$$

At equilibrium (assumption 3) it holds that $d\Omega/dt = 0$, which implies

$$N = \left(\frac{\rho + \mu}{\rho}\right)\Omega. \qquad (2.5)$$

Equation (2.5) can be used to calculate the density of nodes to be distributed across the landscape at the start of a simulation (prior to rabies introduction) given fixed values of the mean wild dog lifespan, mean replacement period, and susceptible wild dog density. Node density is equal to the density of susceptible wild dogs multiplied by a factor $1/\phi$, where $\phi = \rho/\left(\rho + \mu\right)$ is the proportion of nodes occupied by susceptible wild dogs. Equation (2.5) was used to implement each simulation as follows:

1. Values for each of the model input variables (including mean wild dog lifespan, mean replacement period, and susceptible wild dog density) were sampled from their respective distributions (see global sensitivity analysis in Section 2.5),

2. The samples generated in 1 were used to calculate the number of nodes to be distributed across the simulated $250\,\text{km} \times 125\,\text{km}$ landscape using Equation (2.5),

3. Nodes were uniformly randomly distributed across the landscape until the node density calculated in 2 was reached,

4. Each node in 3 was occupied by a susceptible wild dog with probability ϕ and each dog assigned its own unique sociability,

5. The susceptible wild dog closest to the centre of the upper boundary of the landscape was infected with rabies (the index case) to emulate rabies incursion along the northern coast of Australia (see Section 2.4), and

6. Rabies transmission between wild dogs was simulated until either no dogs were infected or rabies had percolated beyond the fourth milestone distance (120 km from the index case).

2.3.2 Wild dog contact rate dimensional analysis

To illustrate that a wild dog's sociability, x_i, is related to the area of land it traverses per day a dimensional analysis is performed of the wild dog contact rate, k_{ij}, proposed in Equation (2.1) and repeated here for ease of reference:

$$k_{ij} = \lambda^2 e^{-\lambda s_{ij}} x_i x_j. \tag{2.6}$$

A variable enclosed in square brackets, e.g. $[x_i]$, will be used to denote the variable's dimension, whilst T and D will be used to indicate the dimensions of time and distance respectively.

First, the fact that k_{ij} is a rate means it has dimension $[k_{ij}] = \mathrm{T}^{-1}$. Next, because s_{ij} is the Euclidean distance between the mean positions (nodes) of dogs i and j it must have dimension $[s_{ij}] = \mathrm{D}$. Also, the exponent on the right-hand side of Equation (2.6) must be dimensionless[1] which implies the units of λ and s_{ij} cancel out, i.e. $[\lambda] = 1/[s_{ij}] = \mathrm{D}^{-1}$. Rearranging Equation (2.6) and recognizing $[x_j] = [x_i]$ it follows that

$$[x_i]^2 = \frac{[k_{ij}]}{[\lambda]^2} = \frac{\mathrm{D}^2}{\mathrm{T}}, \tag{2.7}$$

which then gives

$$[x_i] = \sqrt{\frac{\mathrm{D}^2}{\mathrm{T}}}. \tag{2.8}$$

[1] If the exponent was not dimensionless, then the Maclaurin series for the exponential function would equate to the summation of terms with different dimensions, which is not permissible. Hence the exponent must be dimensionless.

The dimension of the radicand under the square root, D^2/T, represents area per unit time. Equation (2.8) therefore says that a wild dog which traverses a large area of land per unit time is one that is highly sociable. This is appealing because it is natural to expect a dog that covers a large area of land per unit time to encounter more dogs during some time interval than one that only covers a small area of land per unit time. An implication of this result worth noting is that just because a wild dog has a small *home range* does not mean it is not sociable. If, for example, it were to traverse its small home range quite frequently then it could still encounter other dogs on a regular basis, even if they happen to be the same dogs on multiple occasions. Thus, a wild dog's propensity for contacting other dogs (i.e. its sociability) can be related to the area of land it traverses per unit time rather than the size of its home range.

Lastly, for the particular units implemented in this model ($D \sim$ km, $T \sim$ day) the units of a wild dog's sociability are $\text{km/day}^{0.5}$.

2.4 Canine rabies transmission networks

To simulate the spread of rabies, long-range percolation [23, 67, 70] was implemented on dynamic wild dog contact networks. Rabies was introduced into the wild dog population by selecting the dog closest to the centre of the upper boundary of the landscape as the index case and infecting it. Thereafter a directed edge from an infectious dog i to a susceptible dog j, representing a successful transmission event, was generated with probability

$$p_{ij} = 1 - e^{-\tau k_{ij} \delta}, \tag{2.9}$$

where τ is the transmission probability given contact and $\delta = 1\,\text{day}$ is the model time step. Dogs exposed to infection were assigned a gamma distributed incubation period after which they were rabid (infectious) for an exponentially distributed period of time [71–73]. When a dog died (either due to natural mortality or disease) its corresponding node was rendered 'unoccupied' and remained so until a new susceptible dog (with its own unique sociability) arrived. The average length of time until an unoccupied home range was reoccupied will be referred to as the 'replacement period'.

In each simulation rabies was transmitted from dog to dog and eventually either died out or progressed 120 km from the index case (lightest grey semicircle, Figure 2.3). Because, in spatial disease systems, the percolation of infection beyond various milestone distances can be construed as alternative definitions of an epidemic [23, 67], four sensible and convenient milestone

distances were considered, namely 30, 60, 90, and 120 km (dark to light grey semicircles, Figure 2.3).

2.5 Global sensitivity analysis

For each transmission network realization 11 outcome variables of interest were recorded:

(1–4) binary outcomes indicating whether or not rabies percolated beyond each of the four milestone distances,

(5) the number of dogs infected by the index case during its infectious period,

(6–9) the time taken for rabies to percolate beyond each milestone distance,

(10) the total number of nodes within the first milestone distance, and

(11) the number of dogs (occupied nodes) within the first milestone distance before rabies was introduced.

If and when rabies percolated beyond the fourth milestone distance (120 km) a further 3 conditional outcomes were recorded:

(12) the reduction in wild dog density within the first milestone distance,

(13) the proportion of infected (exposed), but not yet infectious, dogs within the first milestone distance, and

(14) the proportion of infectious dogs within the first milestone distance.

To quantify the sensitivity of outcomes 1–5 to each model input variable a global sensitivity analysis was performed. In particular, Sobol's (sensitivity) indices were calculated because they describe the proportional contribution of each input variable's uncertainty to the variance of each respective outcome [74, 75]. In addition, Sobol's indices capture any interaction effects between the input variables and do not require or assume any particular model structure (e.g. linear or monotonic) [74].

To calculate Sobol's indices the Monte Carlo procedure proposed by Saltelli [75] was implemented, for which the technical details are described in Appendix A. Briefly, though, a distribution was defined for each input variable by specifying a range of values (see parameterization in Section 2.6) that captures either its natural variation (e.g. geographic, seasonal, etc.) or

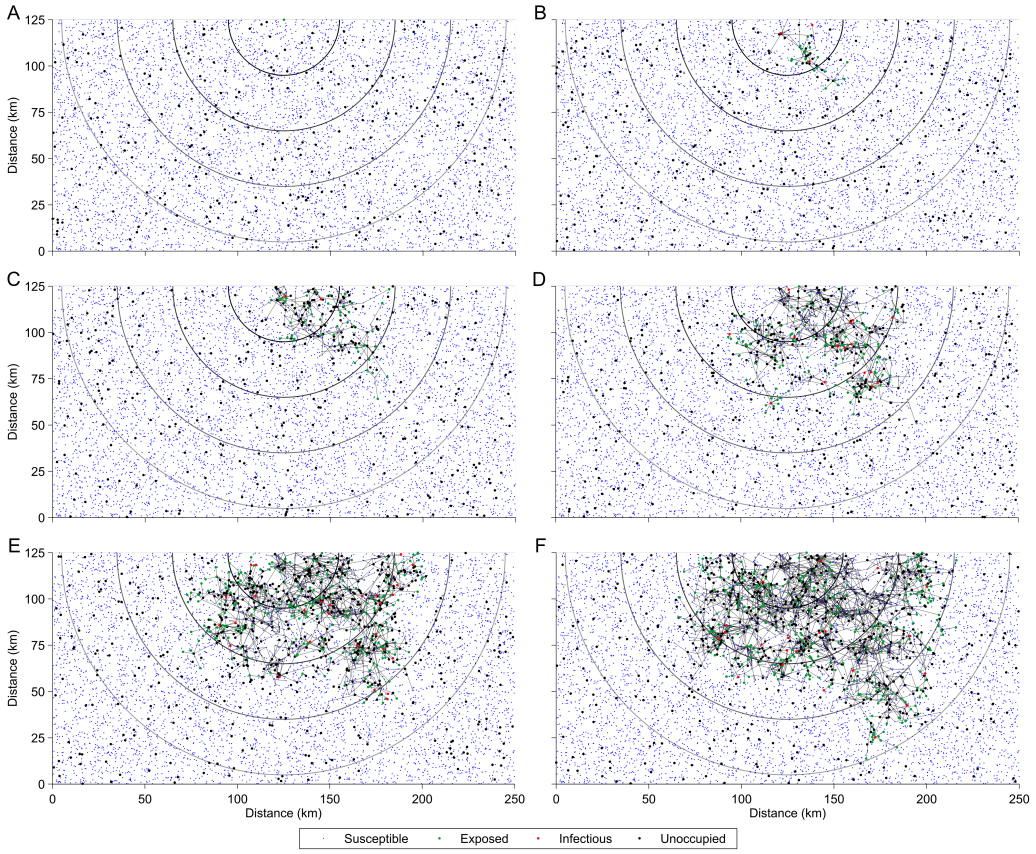

Figure 2.3 – A simulated canine rabies epidemic in the Australian wild dog population. The spread of canine rabies (A) 0, (B) 24, (C) 46, (D) 70, (E) 92, and (F) 116 weeks after the introduction of a single sub-clinically infected (exposed) index case into the Australian wild dog population. Each node, representing a potential wild dog home range centroid, is either occupied by a (blue) susceptible, (bold green) exposed, or (bold red) infectious dog, or alternatively is (bold black) unoccupied. An occupied node's location corresponds to the average position of the dog that occupies the home range. An edge between two nodes represents a transmission event that occurred sometime in the past between two dogs (one infectious, one susceptible) that either currently occupy or previously occupied the nodes (the infection (occupancy) status of each dog (node) may have changed in the time since the transmission event). Edges are shown for all transmission events since the start of the epidemic (not only recent transmission) such that a node may be adjacent to multiple edges by the time rabies percolates (116 weeks for this particular transmission network realization) beyond the fourth milestone distance (lightest grey semi-circle, 120 km). The parameter values used to generate this transmission network realization were: landscape width = 250 km and height = 125 km, wild dog density = 0.20 dogs/km^2, wild dog sociability shape parameter = 30 and scale parameter = $\left(\frac{1}{30}\right)\sqrt{2}$ km/day$^{0.5}$ (gamma distribution), probability of a bite given contact = 0.50, transmission probability given a bite = 0.50, spatial scale parameter $\lambda = 0.4\,\text{km}^{-1}$, mean infectious period = 3 days (exponential distribution), model time step $\delta = 1$ day, incubation period shape parameter = 7 and scale parameter = 4 days (gamma distribution), mean wild dog lifespan = 3 years (exponential distribution), and mean replacement period = 50 days (exponential distribution).

any other uncertainty in its value (e.g. no published estimates) and assuming a uniform distribution across this range. Next, each input variable distribution was sampled 50,000 times (assuming independent input variables) by employing (quasi-Monte Carlo) Sobol sequences, which are deterministic, uniformly distributed sequences that ensure more uniform sampling over the input parameter space than is achieved by traditional pseudo-random sampling. Doing so facilitates faster convergence of the sensitivity indices such that fewer samples are required [76]. The sample values were then stored in two matrices $\mathbf{M}$ and $\mathbf{M'}$ (half in each matrix) such that each row of the matrices denoted a sample vector from the model input parameter space and each column contained the sample values of a single input variable. For each input variable, e.g. y_j, a further two matrices $\mathbf{N}_j$ and $\mathbf{N}_{\neg j}$ were defined as a combination of the columns of $\mathbf{M}$ and $\mathbf{M'}$ (see Appendix A). Thus for this model with 6 input variables (5 biological, 1 random number generator seed—see Section 2.5.1 below) a total of 14 matrices of size (25,000 $\times$ 6) were generated. Next, a transmission network realization was generated for each row (sample vector) of each matrix. This resulted in 14 different, but closely related, distributions of values for each model outcome from which the sensitivity indices were subsequently calculated using Equations (A.10)–(A.13) in Appendix A.

Two sensitivity indices were calculated for each input and outcome variable pair, namely the first-order and total effects. The first-order effect, S_j, is the average proportional reduction in the variance of an outcome variable, z, when an input variable, y_j, has no uncertainty (i.e. it is assigned a fixed value). The greater the value of S_j the more influential is input variable y_j in determining the value of z. Conversely, the total effect, S_j^T, is defined as the average proportion of variance that remains when all input variables are assigned fixed values (i.e. are known) except for input variable y_j. The total effect may be interpreted as the sum of an input variable's first-order effect and any additional effects resulting from its interaction with other input variables.

Lastly, to estimate 95% confidence intervals for Sobol's indices the bootstrapping procedure prescribed in [77] was employed.

2.5.1 Random number generator seed

Because the canine rabies model is a stochastic simulation model, even if all biological input variables were assigned fixed values, model outcomes would still vary from one simulation to the next. That is to say, model outcome variance is not completely explained by biological or ecological input uncertainty. This is important because the additional variance affects the accuracy

of global sensitivity indices estimated using the procedure above. To address this, the random number generator seed (which explains all model variance over and above that accounted for by the biological input variables) was treated as an additional input variable, sampled from its own unique distribution, and assigned a value at the start of each simulation (as was done for the biological input variables). For the rabies model, the seed was sampled from a discrete uniform distribution with range $[0, 2^{32} - 1]$. This range was selected as it includes all permissible integer values for the random number generator seed (MATLAB R2012b), thereby ensuring no bias was inadvertently introduced into model outcome distributions by falsely limiting the range of values the seed could assume. An indication that sampling the random number generator seed has been done correctly is that the seed accounts for a non-negligible proportion of model outcome variance (at the very least this should be true for the total effect) and that when model outcomes are plotted against the seed value a horizontal trend line is observed (i.e. model outcomes should be independent of the seed).

2.6 Parameterization

Because many wild dogs are hybrids of dingoes and modern domestic dog breeds [78] canine rabies pathogenesis parameters for wild dogs were assumed to be similar in value to those reported in the literature for domestic dogs. Far more information is available for canine rabies pathogenesis parameters than those relating to wild dog ecology in northern Australia [48, 56, 71, 72, 79, 80]. Consequently, disease related parameters were assigned point estimates obtained from the literature whereas for ecological parameters a range of feasible values guided by a combination of expert opinion and the literature were considered (Table 2.1). If absolutely no information was available for a particular ecological parameter, a range of values spanning an order of magnitude was selected based on mechanistic reasoning.

At present, there are three published, field-data derived estimates for the mean incubation period of canine rabies. Hampson et al. [72] and Hampson et al. [71] report values of 25.5 days and 22.3 (95%CI: 20.0–25.0) days respectively, derived directly from Tanzanian rabid dog natural infection and contact tracing data. Coleman and Dye [81], on the other hand, obtain their estimate of 4.18 (SE: 0.27) weeks from a thesis [79] that calculates the mean incubation period from observed canine rabies cases in Zimbabwe. It was therefore assumed the incubation period was gamma distributed (as observed by [71]) with a mean of 28 days (fixed shape parameter = 7, fixed scale parameter = 4 days). Under these assumptions, 93.5% of simulated in-

Table 2.1 – Canine rabies model parameter point estimates and input variable distributions.

Parameter, Symbol (Units)	Point Estimate/Distribution	Literature
Model time step, δ (days)	1	
Random number generator seed[a,b]	$U\{0, 2^{32} - 1\}$	
Canine rabies pathogenesis parameters		
Incubation period shape parameter	7	[71, 72, 79, 81]
Incubation period scale parameter (days)	4	[71, 72, 79, 81]
Mean infectious period (days)	3	[71, 80, 81]
Probability a rabid dog bites another dog given contact	0.5	Assumed
Transmission probability given a bite	0.5	[71]
Wild dog ecological parameters		
Population density[a] (dogs/km^2)	$U(0.05, 0.38)$	[48, 82]
Wild dog sociability shape parameter	30	Assumed
Wild dog sociability scale parameter[a] (km/day$^{0.5}$)	$U\left(\frac{1}{30}\sqrt{0.5}, \frac{1}{30}\sqrt{5}\right)$	Assumed
Spatial scale parameter[a], λ (km^{-1})	$U(0.1, 1.0)$	Assumed
Mean lifespan[a] (years)	$U(2, 4)$	[56, 68, 69]
Mean replacement period[a] (days)	$U(1, 100)$	[82]

[a] For each of the six input variables 50,000 values were sampled from their respective uniform distributions.
[b] Values for the random number generator seed were sampled from the discrete uniform distribution.

cubation periods in the transmission network model lasted between 2 weeks and 3 months, with the remaining 6.5% shorter than 2 weeks.

The infectious period of canine rabies is much shorter than the incubation period. Tepsumethanon et al. [80] reported that naturally infected rabid dogs in Thailand survived a median of 4 (95%CI: 3.7–4.3) days after first displaying symptoms (abnormal behaviour or bite event). Two field-data derived mean infectious periods of 3.1 (95%CI: 2.9–3.4) days and 0.81 (range: 0.29–1.71) weeks have also been published by Hampson et al. [71] and Coleman and Dye [81] respectively. There is some evidence supporting a gamma distributed infectious period [71]. Nevertheless, the infectious period is often modelled using the closely related exponential distribution [72, 73]. This same approach was adopted here, assuming a fixed mean of 3 days such that 96.4% of simulated infectious periods were shorter than 10 days.

In Equation (2.9), τ is defined as the transmission probability given contact. But rabies transmission requires that an infectious dog inoculate a susceptible dog by biting it rather than by merely being in its vicinity. Thus, assuming independence, τ is equivalent to the probability a rabid dog bites another dog given contact (for which there is currently no published estimate and so a fixed intermediate value of 0.5 was assumed) multiplied by the probability of successful transmission given a rabid dog bites a susceptible dog (estimated by Hampson et al. [71] to be 0.49 (95%CI: 0.45–0.52) and which was taken as 0.5).

There are no published estimates for the population density of wild dogs in tropical northern Australia where a rabid dog incursion is most likely. To select a range of feasible values, wild dog population density estimates obtained for other regions around Australia (which have different terrain types, climates, and levels of anthropogenic impact) were considered. In general, wild dog density appears to depend largely on landscape carrying capacity, with lower densities observed in arid regions (0.08 dogs/km^2) and higher densities in higher-rainfall areas (0.14–0.3 dogs/km^2) [48]. It is also likely that dog densities have responded positively to agriculture since anthropogenic resources subsidize both food (sheep, cattle, goats, kangaroos, and rabbits) and water (nowhere in the cattle zone is further than 10 km from water) [82, 83]. Conversely, wild dog densities may be reduced by human-related control measures such as baiting. To account for the variation in wild dog densities resulting from each of these factors densities within the range 0.05–0.38 dogs/km^2 were considered.

The literature provides no estimates for the area of land a wild dog traverses per day, reporting only home ranges, which are not rates and therefore not equivalent [48, 56]. Nevertheless, because wild dog home ranges vary between 10 and 100 km^2 [56], and because a wild dog probably only covers a

small portion of its home range each day (although potentially frequenting some areas more than once each day), mean values for the area of land traversed per day were considered from a plausible range spanning an order of magnitude (0.5–5 km^2/day). Assuming a wild dog's sociability is equal to the square root of the area it traverses per day (Section 2.3.2) a range for the mean sociability of $\sqrt{0.5}$–$\sqrt{5}$ $\mathrm{km/day}^{0.5}$ was then calculated. To assign each wild dog a unique sociability, and thereby incorporate individual contact heterogeneity into the model, wild dog sociabilities were assumed to be gamma distributed. Specifically, the sociability shape parameter was assumed to take a fixed value (arbitrarily chosen to be 30) whilst the sociability scale parameter was sampled from the range $\left[\frac{1}{30}\sqrt{0.5}, \frac{1}{30}\sqrt{5}\right]$ $\mathrm{km/day}^{0.5}$ to ensure the desired range of values for the mean area of land traversed per day.

There are no recorded estimates in the literature for the introduced spatial scale parameter λ. Nor are there any appropriate data (e.g. proximity log or GIS tracking data) relating the contact rate of wild dogs to the distance between their mean positions that can be used to estimate λ. Wild dog contact distance distributions, however, are likely to be highly left-skewed [82]. From a mechanistic perspective, it seems reasonable to expect wild dog contact rates to decline substantially over distances between 1 km and 10 km. The input parameter λ was therefore sampled from a range of values spanning an order of magnitude (0.1–1.0 km^{-1}), where for $\lambda = 0.1\,\mathrm{km}^{-1}$ the wild dog contact rate declines by 50% every 7 km and for $\lambda = 1.0\,\mathrm{km}^{-1}$ it declines by 50% every 0.7 km.

The natural mortality rate of wild dogs is a complex parameter that is most likely age- (highest in juveniles < 18 months) and density-dependent; the precise relationship, however, has not yet been quantified. It has been noted, though, that wild dogs live up to 10 years, with most dying by 5–7 years [56, 68, 69]. Therefore, to maintain model transparency wild dog lifespan was assumed to be exponentially distributed and the mean value was sampled from a range of 2–4 years such that the probability a wild dog lived ≥ 10 years varied from 0.7 to 8.3%. The replacement period was also assumed to be exponentially distributed and its mean value was sampled from a range of 1–100 days based on expert opinion [82].

2.7 Results

2.7.1 Probability of a canine rabies epidemic

Of the 50,000 transmission network realizations obtained from evaluating the model on the rows of matrices $\mathbf{M}$ and $\mathbf{M'}$, canine rabies percolated 30,

60, 90, and 120 km a total of 11,306 (23%), 10,782 (22%), 10,695 (21%), and 10,672 (21%) times respectively. It is therefore estimated there is a 21% probability the introduction of a single sub-clinically infected dog into the wild dog population of Australia and subsequent transmission of rabies virus will result in a canine rabies epidemic. Furthermore, because there is little difference between the probabilities rabies percolates 30 km and 120 km, once an epidemic starts it is unlikely to stop naturally. For 99.4% of the simulations in which rabies percolated 120 km, rabies was still present within milestone 1 (30 km) at the time of percolation beyond milestone 4. This suggests the percolation of rabies 120 km or further is a reasonable indicator that canine rabies has become established in the wild dog population (at least in the region close to where the initial transmission event occurred).

2.7.2 Basic reproduction number

The basic reproduction number, R_0, was estimated to be 1.09 (95% CI: 1.07–1.10) from the mean number of wild dogs infected by the index case over all 50,000 transmission network realizations[2].

2.7.3 Rate of spread

Given the current absence of epidemiological and ecological data for canine rabies in Australian wild dogs, the time taken for rabies to percolate beyond the fourth milestone distance (120 km) was used to plot a baseline distribution of speeds the infection might travel (Figure 2.4). The distribution indicates a mean (median) rate of spread of 90 (67) km/year.

2.7.4 Relationship between outcome and input variables

The distributions of model outcomes for the 50,000 transmission network realizations were used to investigate the relationships between the outcomes and six input variables (Figures 2.5–2.12). To do this, each simulation was binned by input variable value and thereafter the median (or mean) value of each outcome in each bin calculated.

In Figure 2.5 the probability canine rabies percolates beyond each milestone distance is plotted as a function of each input variable. The closely overlapping curves confirm there is little difference between the probability

[2]Strictly speaking, this quantity is not equal to R_0 as the index case is not a typical infectious case. This is discussed at greater length in Chapter 4.

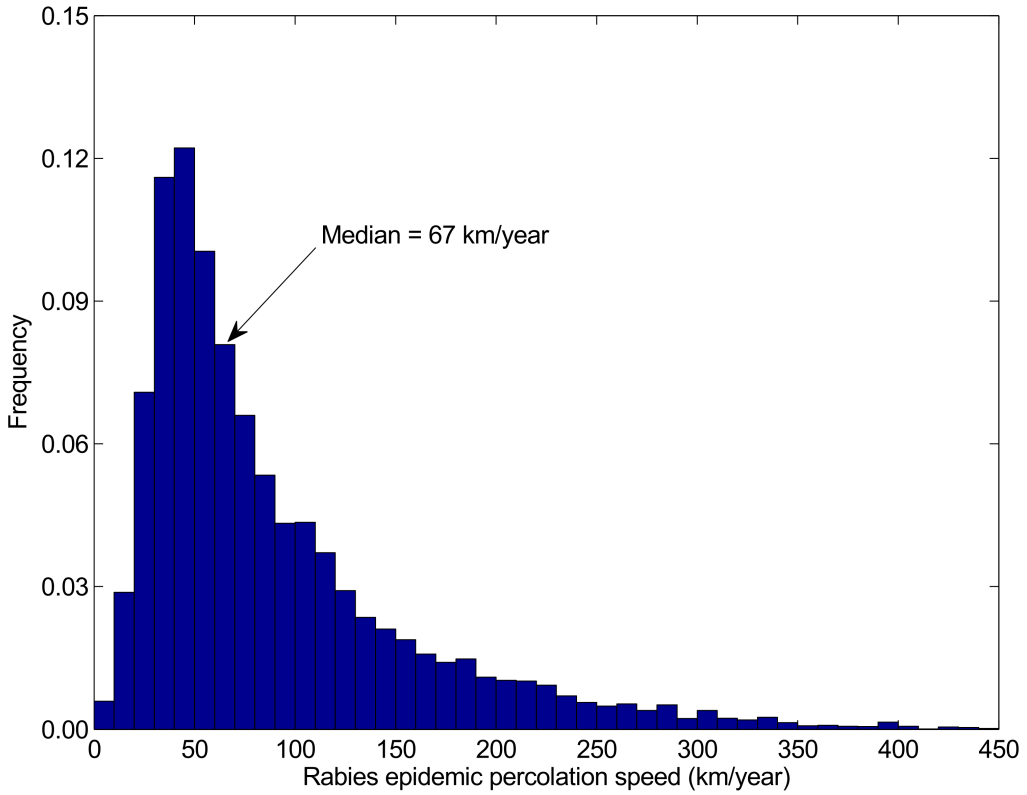

Figure 2.4 – The distribution of canine rabies percolation speeds. The distribution of speeds canine rabies might spread through the Australian wild dog population, estimated from the time taken for the infection to percolate beyond the fourth milestone distance (120 km) given that it percolates beyond the fourth milestone distance.

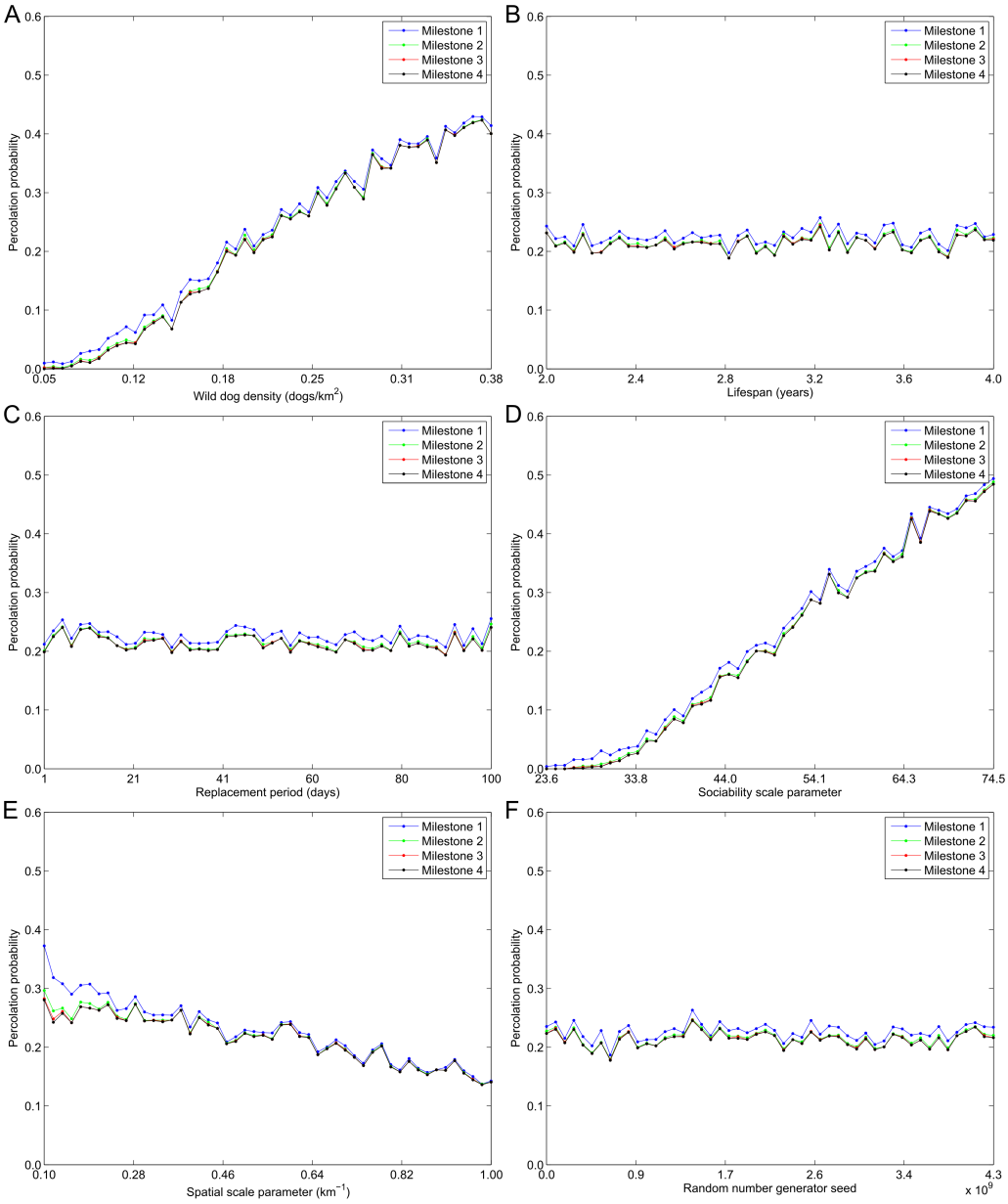

Figure 2.5 – The percolation probability versus model input parameters. The probability canine rabies percolates (blue, milestone 1) 30 km, (green, milestone 2) 60 km, (red, milestone 3) 90 km, and (black, milestone 4) 120 km as a function of (A) wild dog density, (B) mean lifespan, (C) mean replacement period, (D) sociability scale parameter, (E) spatial scale parameter, λ, and (F) random number generator seed.

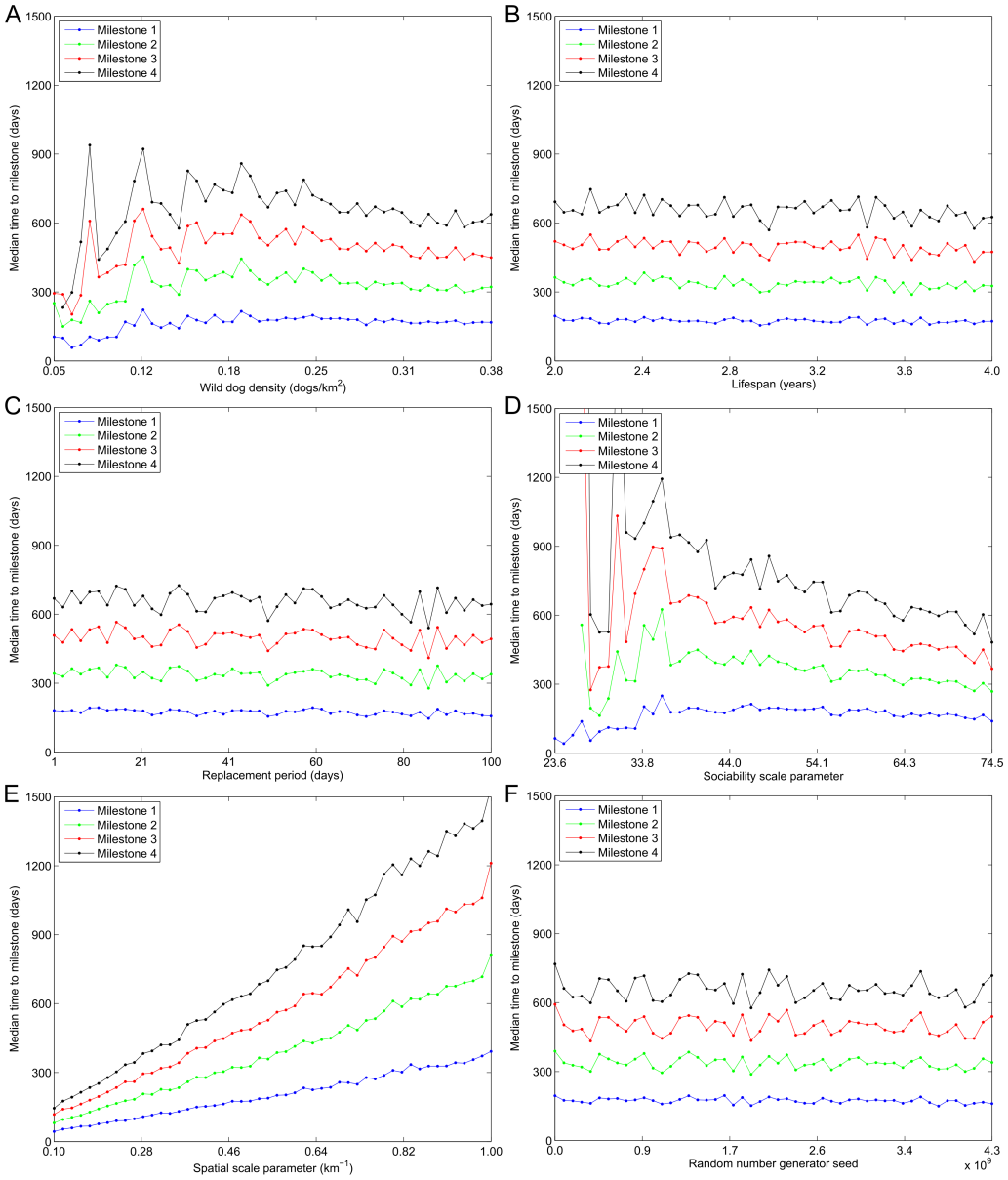

Figure 2.6 – The time to milestone versus model input parameters. The median time taken for canine rabies to percolate (blue, milestone 1) 30 km, (green, milestone 2) 60 km, (red, milestone 3) 90 km, and (black, milestone 4) 120 km as a function of (A) wild dog density, (B) mean lifespan, (C) mean replacement period, (D) sociability scale parameter, (E) spatial scale parameter, λ, and (F) random number generator seed, given rabies percolates beyond each respective milestone distance.

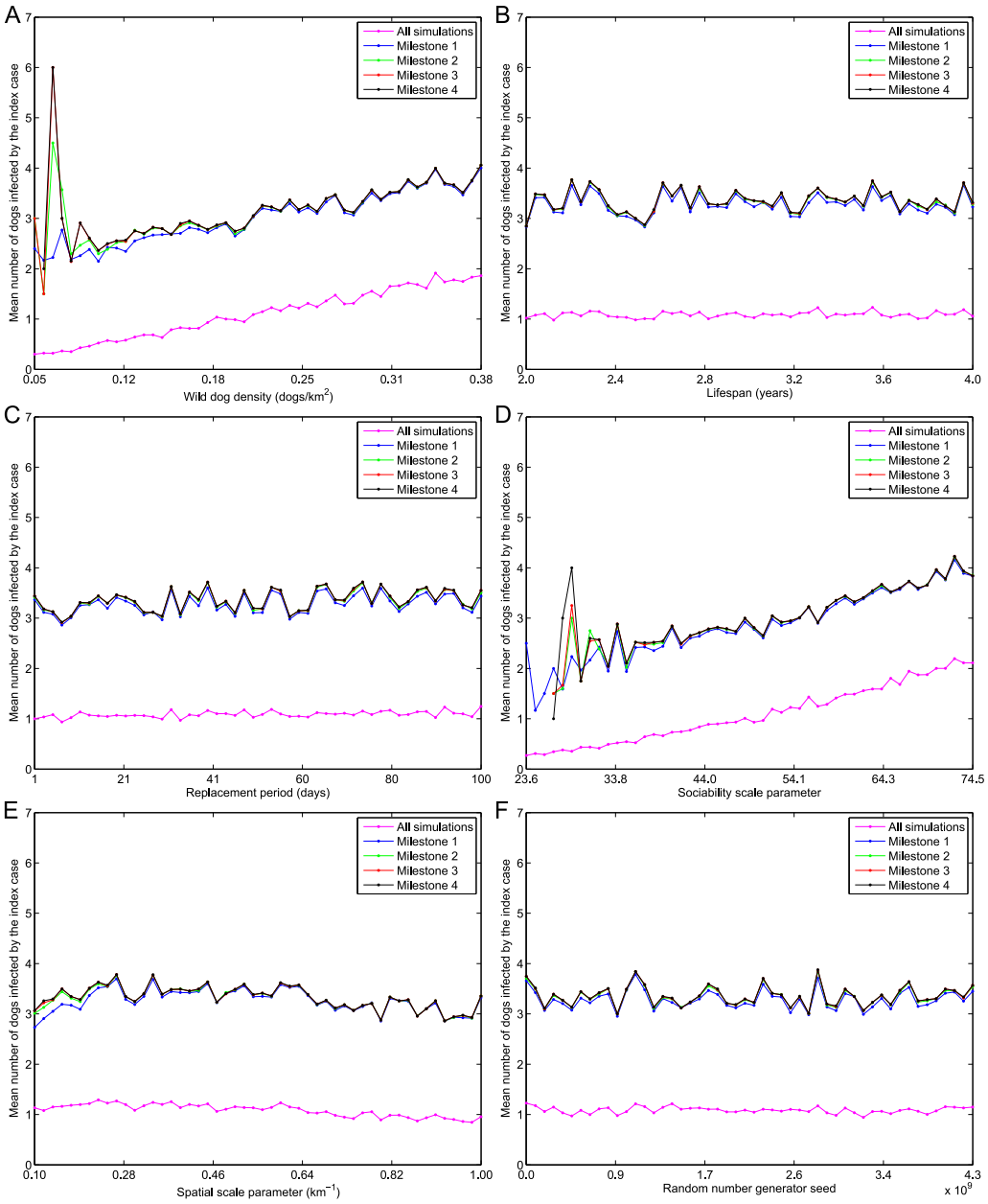

Figure 2.7 – Number of dogs infected by the index case versus model input parameters. The mean number of dogs infected by the index case as a function of (A) wild dog density, (B) mean lifespan, (C) mean replacement period, (D) sociability scale parameter, (E) spatial scale parameter, λ, and (F) random number generator seed, given rabies percolates (blue, milestone 1) 30 km, (green, milestone 2) 60 km, (red, milestone 3) 90 km, or (black, milestone 4) 120 km. The (magenta, all simulations) curve is the mean number of dogs infected by the index case irrespective of whether rabies percolates or not, equivalent to the basic reproduction number R_0 conditional on the corresponding model input variable taking a particular value.

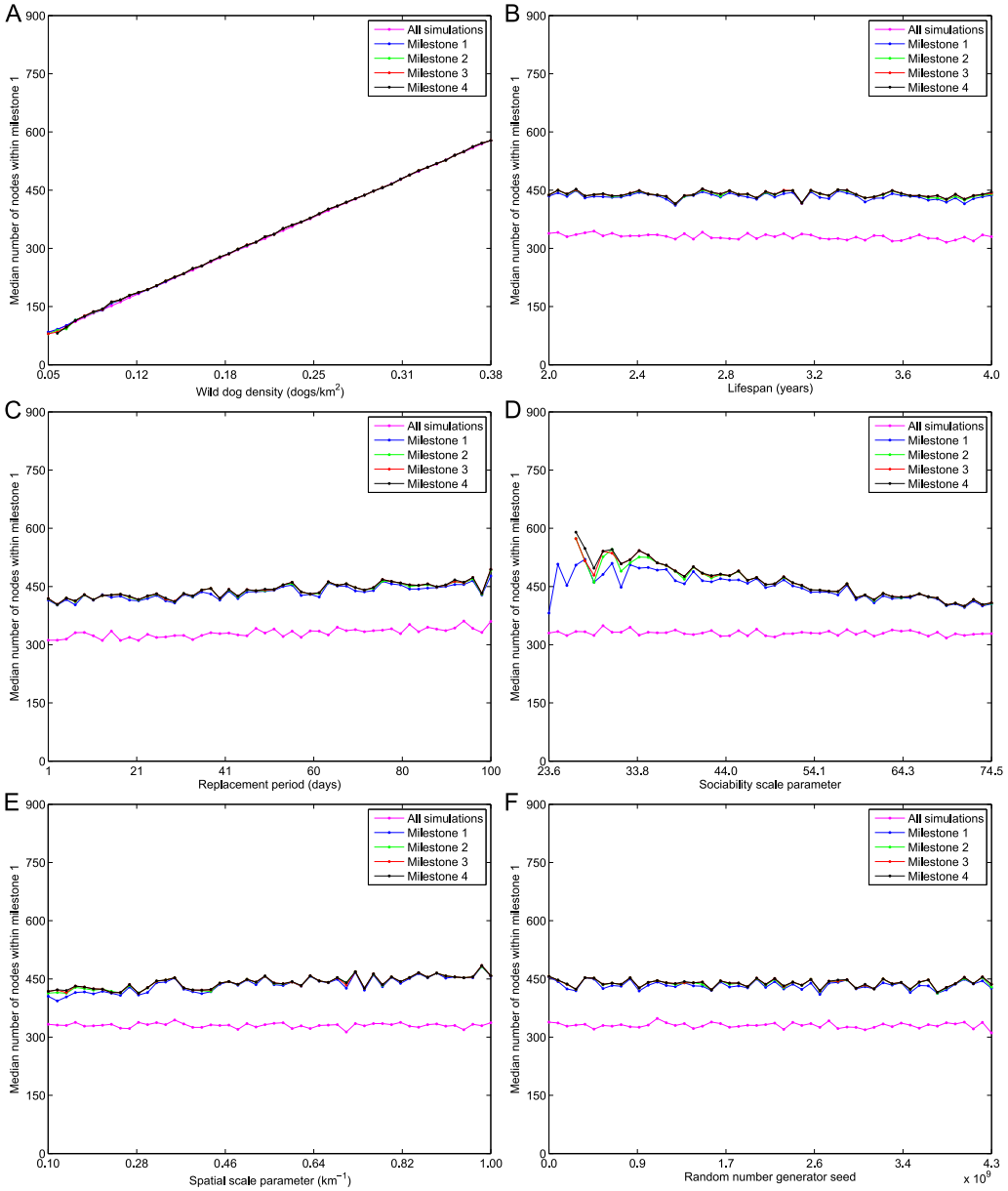

Figure 2.8 – Number of nodes within milestone 1 versus model input parameters. The median number of nodes within milestone 1 as a function of (A) wild dog density, (B) mean lifespan, (C) mean replacement period, (D) sociability scale parameter, (E) spatial scale parameter, λ, and (F) random number generator seed, given rabies percolates (blue, milestone 1) 30 km, (green, milestone 2) 60 km, (red, milestone 3) 90 km, or (black, milestone 4) 120 km. The (magenta, all simulations) curve is the median number of nodes within milestone 1 irrespective of whether rabies percolates or not.

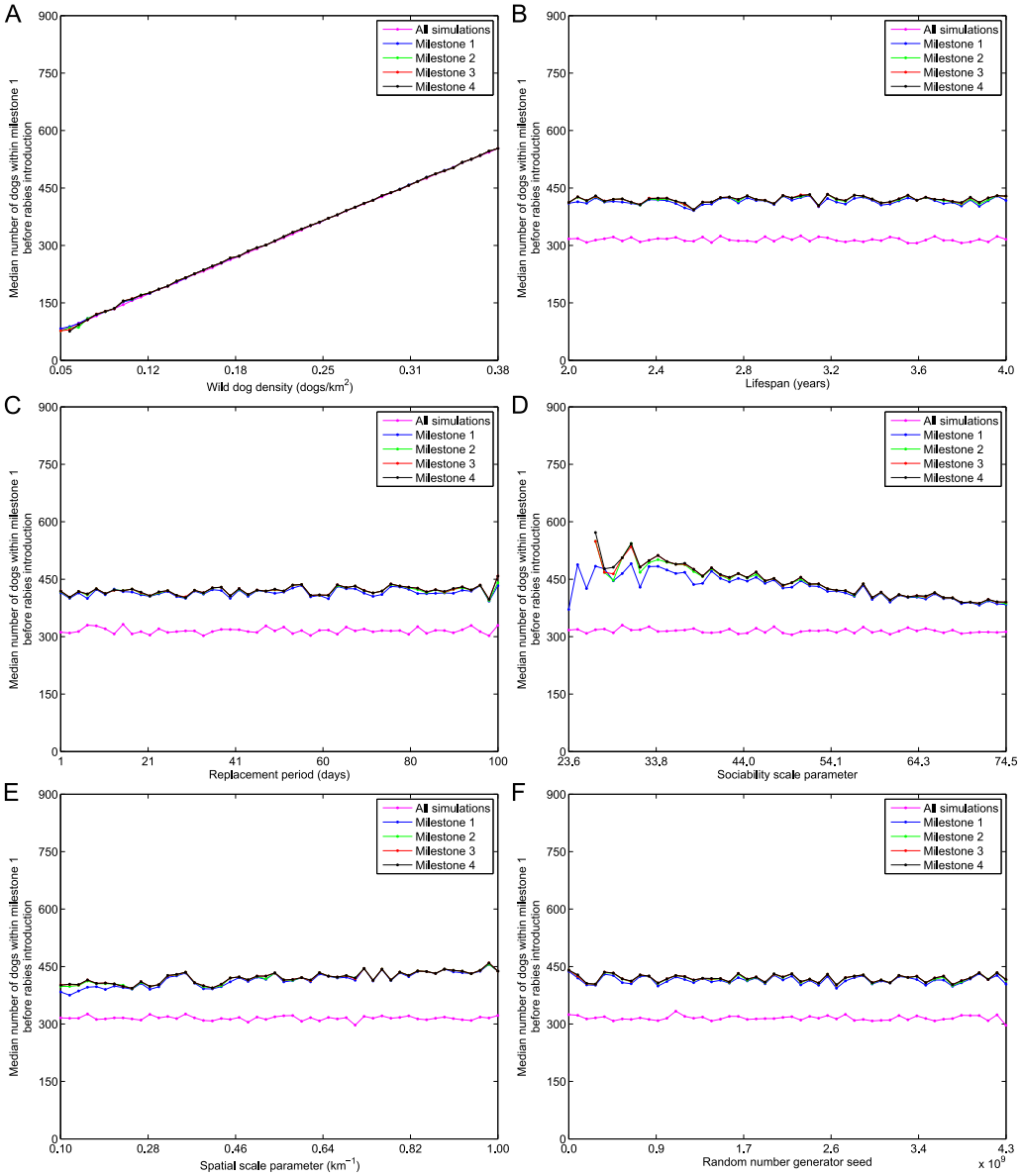

Figure 2.9 – Number of dogs within milestone 1 before rabies introduction versus model input parameters. The median number of dogs within milestone 1 before rabies introduction as a function of (A) wild dog density, (B) mean lifespan, (C) mean replacement period, (D) sociability scale parameter, (E) spatial scale parameter, λ, and (F) random number generator seed, given rabies percolates (blue, milestone 1) 30 km, (green, milestone 2) 60 km, (red, milestone 3) 90 km, or (black, milestone 4) 120 km. The (magenta, all simulations) curve is the median number of dogs within milestone 1 before rabies introduction irrespective of whether rabies percolates or not.

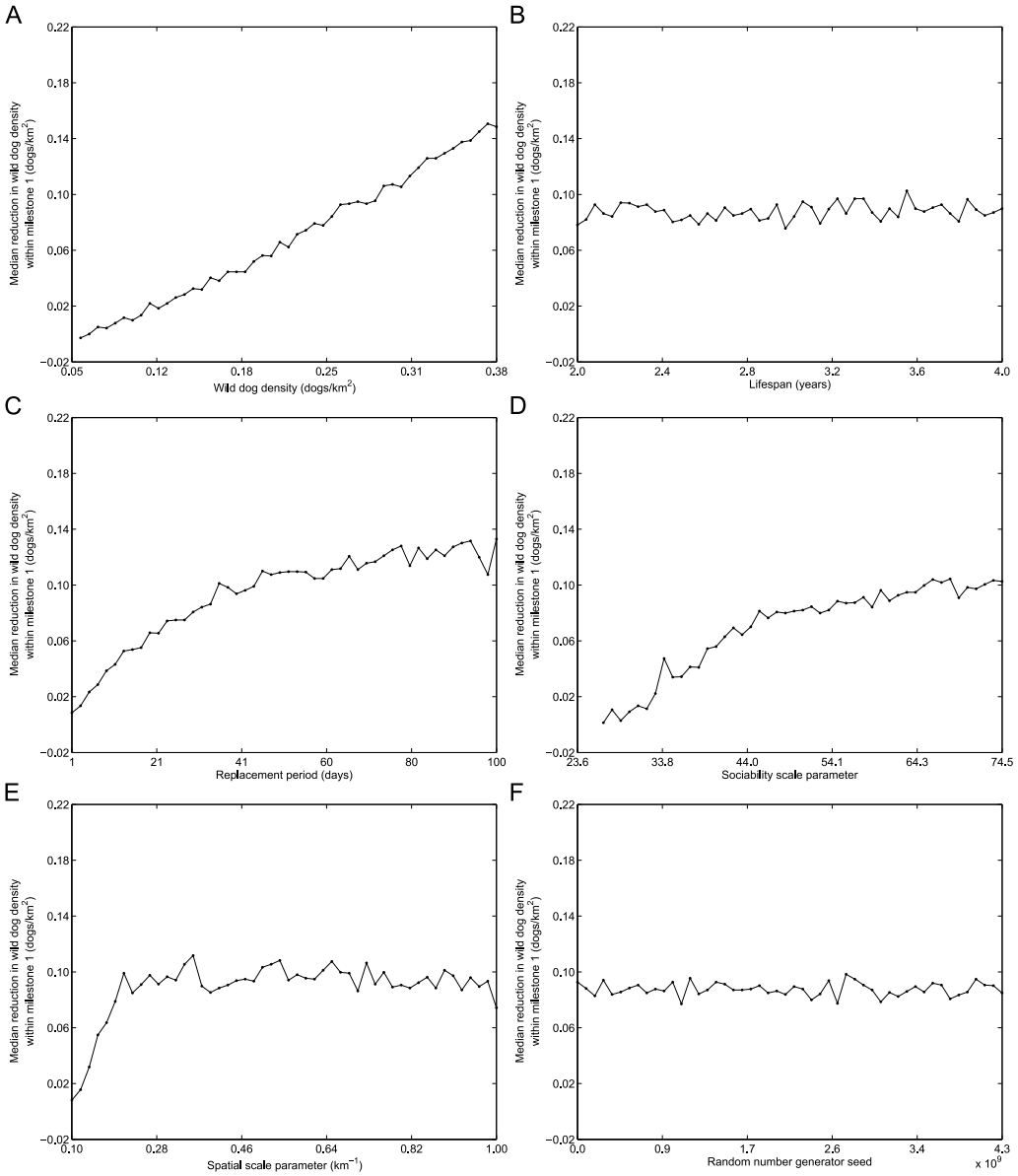

Figure 2.10 – Reduction in wild dog density within milestone 1 versus model input parameters. The median reduction in wild dog density within milestone 1 as a function of (A) wild dog density, (B) mean lifespan, (C) mean replacement period, (D) sociability scale parameter, (E) spatial scale parameter, λ, and (F) random number generator seed, given rabies percolates beyond milestone 4 (120 km).

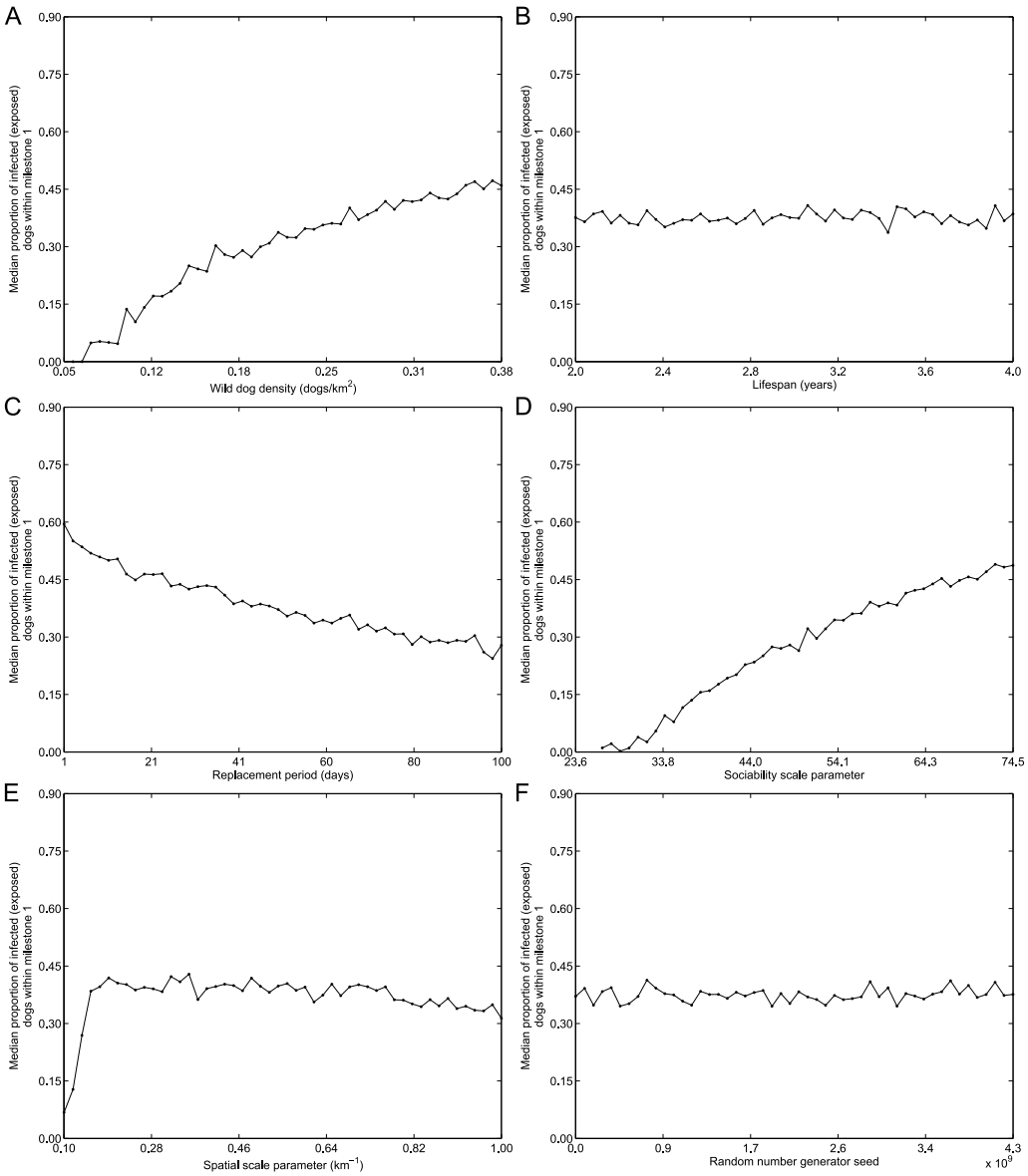

Figure 2.11 – Proportion of infected (exposed) dogs within milestone 1 versus model input parameters. The median proportion of infected (exposed) dogs within milestone 1 as a function of (A) wild dog density, (B) mean lifespan, (C) mean replacement period, (D) sociability scale parameter, (E) spatial scale parameter, λ, and (F) random number generator seed, given rabies percolates beyond milestone 4 (120 km).

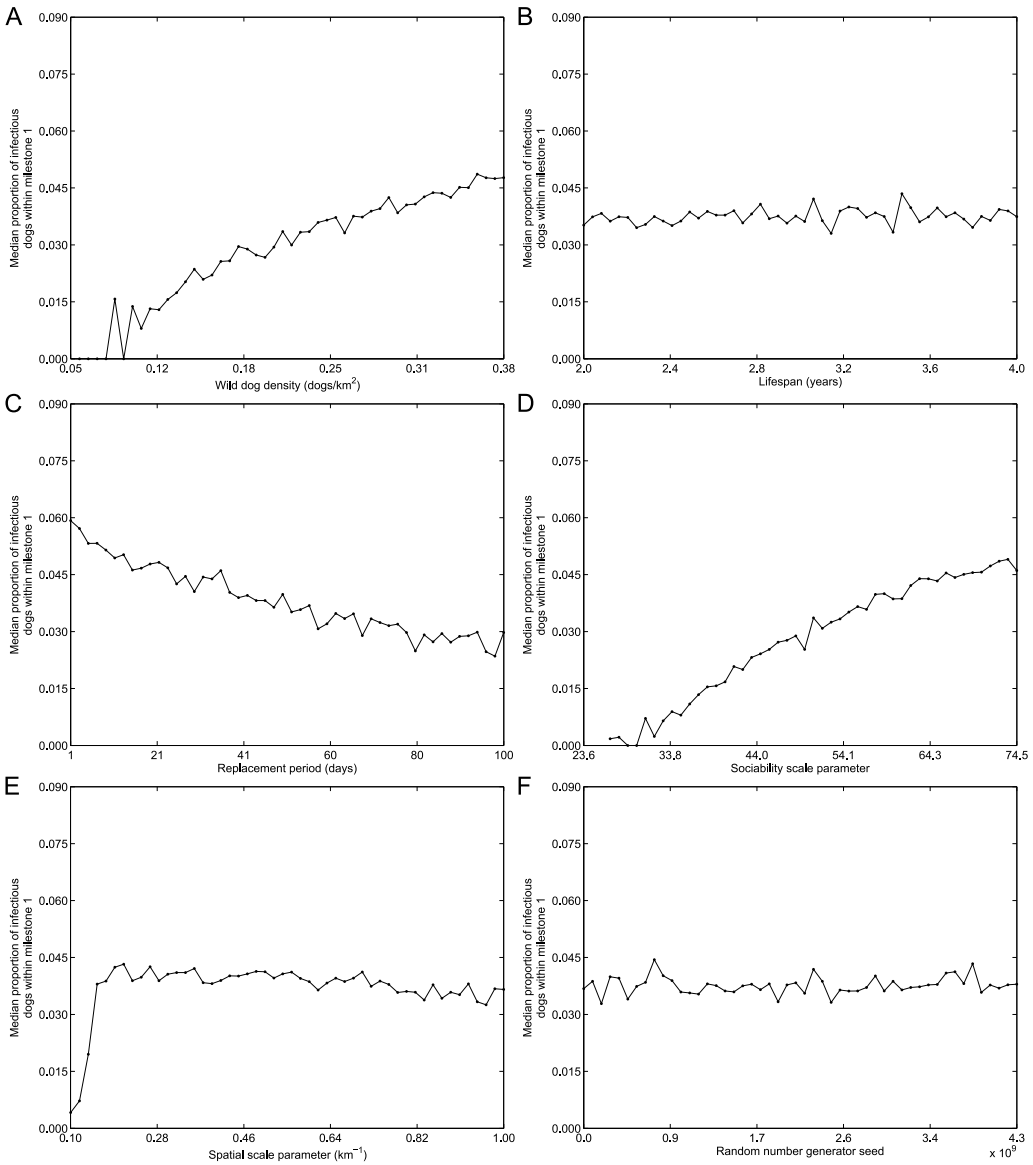

Figure 2.12 – Proportion of infectious dogs within milestone 1 versus model input parameters. The median proportion of infectious dogs within milestone 1 as a function of (A) wild dog density, (B) mean lifespan, (C) mean replacement period, (D) sociability scale parameter, (E) spatial scale parameter, λ, and (F) random number generator seed, given rabies percolates beyond milestone 4 (120 km).

rabies will spread 30 km (milestone 1) or 120 km (milestone 4). Also noticeable is that as values for wild dog density or the sociability scale parameter increase the probability of percolation increases rapidly. Specifically, for wild dog densities below 0.15 dogs/km^2 and a sociability scale parameter less than 36.0 km/day$^{0.5}$ the probability rabies will percolate beyond milestone 4 is less than 12% and 5% respectively. Once wild dog densities rise above 0.35 dogs/km^2 and the sociability scale parameter increases to 67.0 km/day$^{0.5}$ the probability of percolation exceeds 40%. Although not as strong a trend, the more territorial wild dogs are in their behaviour (i.e. as values for the spatial scale parameter, λ, get larger) the lower the probability of percolation. Last of all, as values for λ decrease the curve corresponding to milestone 1 (30 km) begins to separate from those for milestones 2, 3, and 4, indicating that some rabies outbreaks progress beyond milestone 1 but not the milestones further away. This occurs when wild dogs come into contact with other dogs far away, not only their nearest neighbours, but either wild dog density or the sociability scale parameter is too low to sustain the spatial transmission of rabies through the wild dog population (Figure 2.5, Panels A and D). This is important because it means that even though there may only be a few dogs infected during a short-lived epidemic, they may still be spread over a reasonably large geographic region with a radius of 30 km.

In Figure 2.6 the median time for rabies to percolate beyond each milestone distance, given rabies percolated beyond each respective milestone distance, is shown as a function of each input variable. Three features are immediately apparent, the first of which is that the curves are equidistant. This indicates the wave front speed was constant over time and with distance from the index case. The second feature is that the median time for percolation grows almost linearly as the spatial scale parameter, λ, increases. Thirdly, at low values for wild dog density and the sociability scale parameter the median time to percolation is highly variable. In this region of parameter space canine rabies very seldom percolated beyond the milestone distances (see Figure 2.5), and consequently only a small number of simulations contribute to the calculation of the median time to percolation in these bins resulting in greater variance between bins. Once wild dog density and the sociability scale parameter are larger than 0.15 dogs/km^2 and 40.0 km/day$^{0.5}$ respectively, such that the probability of percolation is greater than 10%, the estimates for the median time to percolation become less variable. To a smaller degree, a similar effect is observed for all other model outcomes for the same reason (see Figures 2.7–2.12).

The relationship between the mean number of dogs infected by the index case and each of the six model input variables is illustrated in Figure 2.7. The magenta (all simulations) curve illustrates the conditional dependence of

R_0 on each of the input variables, whereas the remaining curves illustrate the dependence of the mean number of dogs infected by the index case conditional on rabies having percolated beyond each respective milestone distance. The first result worth noting is that R_0 increases linearly with wild dog density and the sociability scale parameter. This is reasonable since these input variables increase the number of dogs the index case (and any other dog for that matter) comes into contact with, and therefore has opportunity to infect. A second, more subtle, result is that as the spatial scale parameter λ increases, the conditional estimate of R_0 marginally decreases. On face value this seems incorrect since, as mentioned in Section 2.3, it is known from [67] that the average contact rate $\mathrm{E}\big[k_i\big] = \mathrm{E}\big[\Sigma_j k_{ij}\big]$ is independent of λ. The key to understanding this result, however, is that although $\mathrm{E}\big[k_i\big]$ is independent of λ, the number of unique dogs contacted is not. When the value of λ is increased a dog compensates for its loss of contact with other dogs far away by contacting its nearest neighbours more often. In doing so the number of unique dogs contacted is reduced. Thus at higher values of λ the index case contacts fewer dogs, and consequently has less opportunity to transmit rabies, which is reflected in a reduced R_0. The third and final result is that in those simulations where a rabies epidemic occurred (i.e. where rabies percolated beyond the various milestones) the index case infected up to 3 times more dogs than one would expect on average (cf. corresponding magenta, all simulations curve). This suggests a useful, early indicator during an outbreak of whether or not rabies will continue spreading spatially through the wild dog population is the number of dogs infected by the index case.

Figures 2.8 and 2.9 serve more as checks that the model produces sensible results than being interesting results in and of themselves. Nevertheless, it is still useful to describe the dependence of the median number of nodes within milestone 1 (Figure 2.8) and the median number of dogs within milestone 1 before rabies introduction (Figure 2.9) on the model input variables, as doing so is helpful for interpreting the results obtained for the remaining model outcomes (Figures 2.10–2.12).

To understand Figure 2.8 recall that the number of nodes distributed across the landscape at the start of each simulation is calculated using Equation (2.5). Consequently, the number of nodes within milestone 1 is expected to be directly proportional to wild dog density, to grow linearly with mean replacement period, and to decline with mean wild dog lifespan. Although in Figure 2.8 the first two of these relationships is observed (Panels A and C), it appears as if the number of nodes is invariant with respect to mean wild dog lifespan (Panel B). This is because the range of values from which the mean lifespan is sampled is relatively small. Changing the mean wild dog lifes-

pan from 2 to 4 years only reduces the number of nodes within milestone 1 by a minimal 4%. Consequently, the number of nodes within milestone 1 does change with mean lifespan as expected, just not by a readily observable amount. Another interesting observation (Figure 2.8, Panels D and E) is that when the median number of nodes within milestone 1 is calculated from only those simulations in which rabies percolated beyond the various milestone distances a trend appears that is absent when all simulations (magenta curve) are considered. This occurs because when the sociability scale parameter and λ take low and high values respectively the probability rabies will spread spatially is low (see Figure 2.5). Thus, for there to be any chance of rabies percolating beyond the various milestones an exceptional set of circumstances is required. One such requirement is that there is a high density of nodes.

The number of dogs within milestone 1 before rabies introduction is directly proportional to wild dog density (Figure 2.9, Panel A), but, unlike the number of nodes, is independent of all other model input variables (this is tantamount to recognizing that wild dog density is proportional to wild dog density). Furthermore, as was the case for the number of nodes, there must be a large number of dogs within milestone 1 when the probability of percolation is low if there is to be any risk of rabies percolating through the wild dog population (Panels D and E).

When interpreting the results for the reduction in wild dog density, the proportion of infected (exposed) dogs, and the proportion of infectious dogs within milestone 1 (Figures 2.10–2.12) the most important factor to bear in mind is that of timescale. Consider that when the probability of percolation is low rabies must spread quickly through the wild dog population for there to be any chance of it percolating over large distances. But a fast spreading epidemic implies not many dogs within milestone 1 will be infected before rabies progresses beyond milestone 4. Consequently, for regions in the input parameter space where the probability of percolation is low the reduction in wild dog density, the proportion of infected dogs, and the proportion of infectious dogs within milestone 1 will be negligible. This is precisely what is observed at low values for wild dog density and the sociability scale parameter (Figures 2.10–2.12, Panels A and D).

The importance of timescale is further highlighted by the relationship between the spatial scale parameter λ and the time rabies takes to percolate beyond milestone 4 (Figure 2.6, Panel E). When λ is small the time to milestone 4 is short and as a result the three model outcomes are negligible once again (Figures 2.10–2.12, Panel E). However, when $\lambda \geq 0.50\,\mathrm{km}^{-1}$ such that the time to milestone 4 is of the order of years rather than months, then the reduction in wild dog density increases to a substantial $0.04\,\mathrm{dogs/km}^2$ whilst

up to a third of the dogs within milestone 1 are either infected or infectious.

Finally, given the established significance of timescale, it naturally follows that the three model outcomes are also influenced by the mean replacement period. Panel C in Figures 2.10–2.12 indicates that the replacement period has opposite effects on the reduction in wild dog density and the proportion of dogs infected or infectious with rabies. For example, when the mean replacement period is a relatively short 21 days the reduction in wild dog density is $< 0.03\,\text{dogs/km}^2$ whereas nearly 50% of the dogs within milestone 1 are infected or infectious. As the mean replacement period increases to 80 days, though, the reduction in density increases to $0.06\,\text{dogs/km}^2$ whilst the proportion of dogs infected or infectious declines to 33%. To understand these trends, consider that a short replacement period implies a fast influx of new susceptible dogs. But this in turn means that the total number of dogs decreases only marginally (since any dog that dies is quickly replaced) and that the transmission of rabies can continue unabated (with the result being that a high proportion of dogs are infected or infectious). When the replacement period is long, dogs that die are not replaced quickly and the total number of dogs is substantially reduced. This lower density of dogs results in fewer transmission events, which in turn is reflected in the smaller proportion of dogs infected or infectious with rabies.

2.7.5 Global sensitivity analysis

The results of the global sensitivity analysis are reported in Tables 2.2 and 2.3, and plotted in Figure 2.13. Some of the sensitivity indices are negative even though by definition they should fall within the range $[0, 1]$. This is because reported global sensitivity indices are estimates of the true sensitivities, having been calculated from a finite number of samples. Monte Carlo variability generates estimates that are marginally different from their true values, such that when the true values are close to 0 or 1 the sensitivity estimates may fall just below 0 or above 1 [77]. Although increasing sample size typically resolves this small source of error (causing the difference between sensitivity estimates and their true values to converge to zero), it is possible to obtain this error for any sample size if a true sensitivity index is precisely 0 or 1. In practice then, the key to selecting an appropriate sample size involves both ensuring the convergence of the sensitivity estimates and deciding what an acceptable (negligible) difference (error bound) is. In the results presented here estimates outside the range $[0, 1]$ by less than 0.01 were considered acceptable (assuming the true sensitivity values were 0 or 1 as appropriate).

For the binary outcomes indicating whether or not rabies percolated be-

Table 2.2 – First-order effect global sensitivity indices.

	Population density	**Mean lifespan**	**Mean replacement period**	**Sociability scale parameter**	**Spatial scale parameter** (λ)	**Random number generator seed**
Distribution	$U(0.05, 0.38)$	$U(2, 4)$	$U(1, 100)$	$U\left(\frac{1}{30}\sqrt{0.5}, \frac{1}{30}\sqrt{5}\right)$	$U(0.1, 1.0)$	$U\{0, 2^{32} - 1\}$
Milestone 4 (120 km)	0.116	0.010	−0.008	**0.153**	0.013	0.008
Milestone 3 (90 km)	0.115	0.009	−0.009	**0.152**	0.012	0.006
Milestone 2 (60 km)	0.113	0.010	−0.008	**0.151**	0.012	0.006
Milestone 1 (30 km)	0.110	0.010	−0.001	**0.145**	0.020	0.011
No. of dogs infected by the index case	0.070	0.012	0.005	**0.102**	0.005	0.003

Bold values indicate the input variable with highest sensitivity index for each respective outcome variable.

Table 2.3 – Total effect global sensitivity indices.

	Population density	**Mean lifespan**	**Mean replacement period**	**Sociability scale parameter**	**Spatial scale parameter** (λ)	**Random number generator seed**
Distribution	$U(0.05, 0.38)$	$U(2, 4)$	$U(1, 100)$	$U\left(\frac{1}{30}\sqrt{0.5}, \frac{1}{30}\sqrt{5}\right)$	$U(0.1, 1.0)$	$U\{0, 2^{32} - 1\}$
Milestone 4 (120 km)	**0.830**	0.650	0.663	0.565	0.297	0.668
Milestone 3 (90 km)	**0.831**	0.652	0.665	0.565	0.298	0.669
Milestone 2 (60 km)	**0.832**	0.655	0.668	0.561	0.303	0.673
Milestone 1 (30 km)	**0.839**	0.675	0.687	0.546	0.326	0.694
No. of dogs infected by the index case	**0.898**	0.806	0.824	0.384	0.196	0.814

Bold values indicate the input variable with highest sensitivity index for each respective outcome variable.

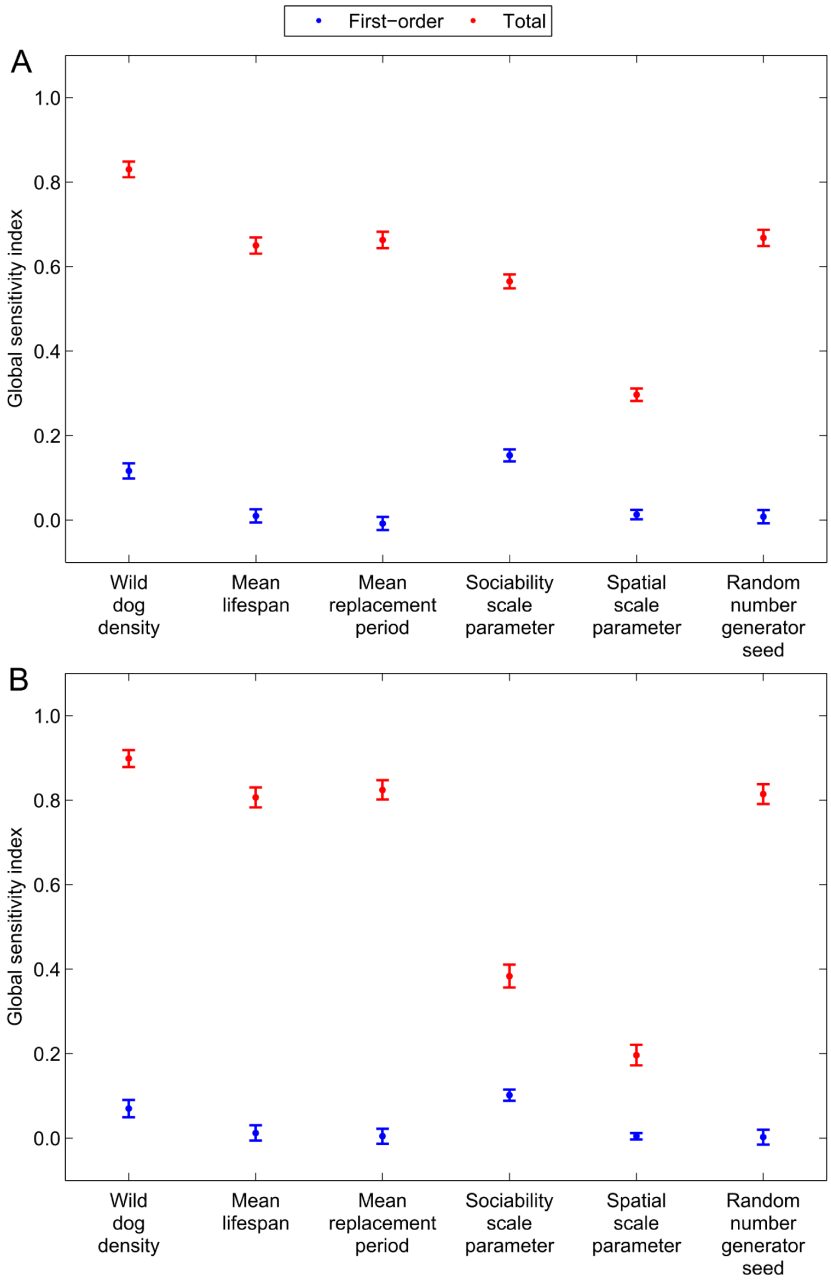

Figure 2.13 – Global sensitivity indices and bootstrap-estimated 95% confidence intervals. The (blue) first-order and (red) total effect global sensitivity indices with bootstrap-estimated 95% confidence intervals for the six model input variables (wild dog density, mean lifespan, mean replacement period, sociability scale parameter, spatial scale parameter, and random number generator seed) and two model outcomes: (A) the binary outcome indicating whether or not rabies percolates beyond the fourth milestone distance (120 km) and (B) the number of dogs infected by the index case during its infectious period.

yond each respective milestone distance, as well as the number of dogs infected by the index case, the wild dog sociability scale parameter and wild dog density were the first and second most influential input variables respectively (Table 2.2, Figure 2.13 Panels A and B). The sociability scale parameter accounted for 15% of the variance whilst wild dog density explained 11%. For the number of dogs infected by the index case these values declined slightly to 10% and 7% respectively. The importance of density, though, was more pronounced when input variable interactions were taken into account explaining over 83% of the variance in all outcomes (Table 2.3, Figure 2.13 Panels A and B).

2.8 Discussion

Canine rabies typically persists in the context of an urban transmission cycle in which stray dogs and unvaccinated, free-roaming owned dogs account for a substantial proportion of the population [48, 54, 58, 71, 72, 84–86]. In this chapter, the potentially unique scenario of canine rabies spreading in a sylvatic transmission cycle with the Australian wild dog as reservoir host has been investigated. It is estimated there is a 21% probability the incursion of a single sub-clinically infected dog and subsequent infection of one or more wild dogs will result in the sustained spatial transmission of rabies within the wild dog population, and on so doing spread at a mean (median) speed of 90 (67) km/year. Given current levels of uncertainty for wild dog ecology parameters, global sensitivity analysis indicates wild dog sociability and wild dog density are the first and second most influential parameters in determining whether a rabies epidemic will occur. Rate of spread, on the other hand, is governed by the spatial scale over which distance begins to affect wild dog contact rates.

In recent years, wildlife contact networks have attracted increasing levels of attention [87]. This is because individual contact behaviour (e.g. heterogeneity and territoriality) influences whether an epidemic can occur, how quickly an infection will spread, and the epidemic final size [88, 89]. Observation of real contact networks also facilitates the construction of contact network models with relevant properties, such that the spread of infection through populations can be simulated. These models can be used to identify factors driving transmission, predict patterns of disease spread, and design effective surveillance and intervention strategies [87]. Given the current absence of Australian wild dog contact network data, the long-range percolation model of Davis [67] was used to simulate the spread of rabies through Australia's wild dog population. Importantly, the model naturally

accommodates heterogeneity (by assigning each dog its own sociability) and territoriality (through the spatial scale parameter λ). This model may be applied in a similar manner to other wildlife populations for which there is either no contact network information or the observation thereof is difficult. This is true even when the host population is not territorial (and the corresponding contact network is small-world in nature) since a sufficiently small value for λ would render each host capable of contacting every other host in the population.

Global sensitivity analysis is an increasingly popular approach to performing sensitivity analysis of epidemiological models [74]. This is because it complements a model's ability to answer questions such as "What is the probability an epidemic will occur?" by identifying those factors most influential in determining its answer. One can therefore regard global sensitivity analysis as a tool that tells researchers 'where to look' and improves their understanding of the processes driving the spread of infection. That is to say, it highlights the parameters worth exploring further and measuring more accurately to improve model outcome predictions. This is particularly useful when several model parameters have wide ranges reflecting large uncertainty or natural variation over space and time (as was the case for nearly all the wild dog ecology parameters in the current study) since a wide range of values for a particular parameter does not imply model outcomes will be sensitive to it (e.g. the mean replacement period). Given the parameter identified as most influential in determining whether a rabies epidemic will occur (i.e. the sociability scale parameter) had a range defined as an order of magnitude (due to lack of published information) field data on wild dog movements and contact behaviour are glaringly missing. This knowledge gap has also been identified by Sparkes et al. [48]. It is therefore recommended that future research should focus on the observation of wild dog contact networks as this will go a long way to improving model accuracy and could at the same time serve to confirm, or generate new hypotheses for, the assumed wild dog contact rate functional form (Equation (2.1)).

The assumption that a wild dog's sociability is equal to the square root of the area of land it traverses per day is a consequence of the wild dog contact rate proposed in Equation (2.1) (see Section 2.3.2). Interestingly, the area of land traversed per day has the same units as those of the diffusion co-efficient D in a system of partial differential equations describing the spatial propagation of travelling infection waves. Furthermore, the velocity at which these travelling waves propagate is proportional to the square root of the diffusion co-efficient, $\sqrt{D}$ [90]. Thus the result that wild dog sociability is a key parameter in the percolation of rabies through the wild dog population beyond milestone distances is consistent with established theory describing

the spatial propagation of infection through a host population (e.g. fox rabies in Europe).

This work has some important caveats to bear in mind. The first is that rabies infection could modify wild dog behaviour in a manner significant for the sociability and spatial scale parameter, λ, values. Predicting exactly how these parameters, and possibly even the contact rate (Equation (2.1)), might differ for rabid dogs, though, is not easy, especially given rabies presents as furious rabies and dumb (paralytic) rabies [48]. Future work should investigate the sensitivity of model outcomes to these factors by comparing results generated for alternative contact rate functions, as well as when the values for sociability and λ are conditional on wild dog disease status (i.e. furious or dumb).

A second caveat is that home ranges (nodes) were assumed to remain unoccupied for an exponentially distributed period of time. In reality, the mean replacement period is likely to be a complex function of wild dog migration between home ranges (movements over and above normal everyday activity captured by the contact rate), seasonal birth rates, and wild dog density. Future research should include investigating whether incorporating more realistic wild dog demography (for which there is currently a paucity of knowledge, particularly for tropical ecosystems in northern Australia) will significantly alter the model outcomes and global sensitivity analysis results presented here.

A third caveat is that when parameterizing the model, the sociability shape parameter was fixed to an arbitrarily chosen value of 30. Choosing a larger value would have reduced the upper limit of the range of values from which the sociability scale parameter was sampled. This, in turn, would have reduced both the probability of percolation (Figure 2.5, Panel D) and the proportion of model outcome variance explained by the sociability scale parameter. If instead a smaller value for the sociability shape parameter had been chosen the opposite would be true. Consequently, accurately quantifying the parameters describing the distribution of wild dog sociabilities (from the area of land they traverse per day) will be important for future predictions of canine rabies spread in Australian wild dogs.

A requirement for rabies to become established is that each infectious dog must infect at least one other dog on average. When a dog in this model is sociable (i.e. when it covers a large area of land per unit time) and when wild dog density is high this requirement is satisfied, and therefore it is reasonable that these two input variables explain whether a rabies epidemic will occur. This interpretation agrees well with the finding that the number of dogs infected by the index case was also most sensitive to these two input variables.

The nearly linear relationship between the rate of spread (time to milestone) and spatial scale parameter, λ, arises from the territorial nature of wild dogs (represented in the model by λ). When an infectious disease spreads in a territorial population (where the traditional random mixing assumption does not apply) a localized depletion of susceptible hosts may occur [23, 67]. One way for the infection to continue spreading is long-range transmission which allows the infection to 'escape' the localized outbreak. The further transmission occurs the faster the infection spreads. In the model presented this occurs when wild dogs traverse large distances (i.e. when λ is small) which supports the hypothesis in [48] that rabies may spread faster in resource-poor areas, e.g. semi-arid or desert regions, where wild dogs traverse greater distances and have larger home ranges [91].

Rate of spread is an important outcome to consider since it determines the level of intervention required to contain an epidemic and bring about disease elimination once detected. The estimates for the epidemic rate of spread in this work are similar to, if not slightly above, the $20 - 80\,\mathrm{km/year}$ published in the literature for other sylvatic rabies virus strains and host populations [48, 92]. This may be because wild dogs are anatomically larger, with allometrically-scaled larger home ranges [93], and therefore genuinely more mobile (higher sociabilities, smaller λ) than raccoons, foxes, and badgers, but alternatively could just be a result of the wide parameter ranges that were considered. In either case, the preparedness of Australia could be greatly improved by targeted field studies that aim to better understand wild dog movement and behaviour.

Chapter 3

The spread of tick-borne pathogens between ticks and vertebrate hosts

3.1 Introduction

The type of contact required for the transmission of tick-borne pathogens is a tick attaching itself to a vertebrate host and taking a blood meal. Ticks need to take blood meals to progress through a number of immature life-stages and reach adulthood. In the case of the most important zoonotic pathogens, two immature life-stages of the tick vector, termed larvae and nymphs, maintain the pathogens and are also responsible for transmission to humans [94, 95].

In this chapter a mechanistic network model is presented for tick-borne pathogens that explicitly accounts for the tendency of feeding ticks to aggregate and for larvae and nymphs to co-aggregate on the same vertebrate hosts. Aggregation is very common with almost all macroparasites observing a distribution known as the 80-20 rule, whereby 80% of individual parasites are found on just 20% of hosts [96, 97]. Co-aggregation has empirical support, with studies of blacklegged ticks, *Ixodes scapularis*, on white-footed mice, *Peromyscus leucopus*, in the United States showing that mice with high nymphal counts also tend to have high larval burdens later in the season [98]. Similarly, studies of *Ixodes ricinus* in Europe came to the same conclusion [99, 100]. These tick species are respectively responsible for human cases of Lyme disease in the United States and Europe.

The original contribution of the work presented here are analytic expressions for R_0 that account for co-aggregation of ticks and coincident co-aggregation, also known as co-feeding. Co-feeding of nymphs and larvae allows transmission from an infected nymph to susceptible larvae feeding in close proximity and at the same time, but without the involvement of a systemic infection in the vertebrate host. The resulting equations for R_0 explicitly describe the relationship between the basic reproduction number and the strength of dependence between counts of larvae and counts of nymphs.

3.1.1 Chapter layout

This chapter begins by reviewing the relevant biological and ecological knowledge. A mechanistic network model for tick and vertebrate host contact patterns is then developed and analytic formulae for R_0 derived. Simulations of Lyme disease transmission on finite realizations of the contact network model are used to visualize the relationship between the basic reproduction number and the extent of co-aggregation. Last of all, the implications of the results for the spread of different tick-borne pathogens are briefly discussed, whilst a discussion of the results in the context of the thesis research question is reserved until Chapter 5.

3.2 Background biology and ecology

There are a multitude of tick-borne pathogens, some of which only infect animals, but others of which are zoonotic in nature and are increasingly infecting humans. Here the focus will be on the ticks, vertebrate hosts, and transmission pathways implicated in the spread and maintenance of the most common tick-borne zoonosis, Lyme disease, which is present in both North America and Europe.

3.2.1 Tick life cycle and phenology

The *Ixodes* tick life cycle comprises four life-stages: egg, larva, nymph, and adult (Figure 3.1). Every year eggs hatch into larvae which then quest for small vertebrate hosts (e.g. white-footed mice in the United States) to which they can attach and take a blood meal from. Having engorged itself, a fed larval tick will drop off its vertebrate host to moult over winter and emerge as a nymph the following year. Nymphs typically emerge slightly earlier in a season than do larvae. This enables them to obtain a blood meal from a small to medium sized vertebrate host and subsequently moult into adults in the same year. Female adult ticks proceed to take a third blood meal from larger vertebrate hosts (e.g. white-tailed deer in the United States) after which they mate and lay eggs in leaf litter. The eggs hatch into larvae the following year such that closure of the tick life cycle occurs after two years.

The seasonal questing behaviour of the different life-stages of ticks is referred to as tick phenology. The typical phenology of *I. scapularis* larvae and nymphs in Northeast United States is illustrated in Figure 3.2. The phenology of adults has not been shown as they are not important for Lyme disease transmission in this region [94]. This may also be true of other regions given that adult *I. scapularis* ticks in the United States and adult *I. ricinus* ticks in Europe generally feed on larger, and therefore different, hosts to larvae and nymphs [94, 102–104], although this is not always the case [102, 105].

The majority of nymphal *I. scapularis* ticks in Northeast United States quest in early spring. This is in contrast to larvae which have a bimodal questing pattern, with a smaller early peak towards the end of spring and a larger later peak approximately two months later. The questing behaviour of *I. scapularis* in Northeast United States is not the same as that observed in Upper Midwest United States. Nor is it the same as the questing behaviour of *I. ricinus* in Britain and Europe. In Upper Midwest United States larval phenology is also bimodal but has a larger first peak than second peak (i.e. the

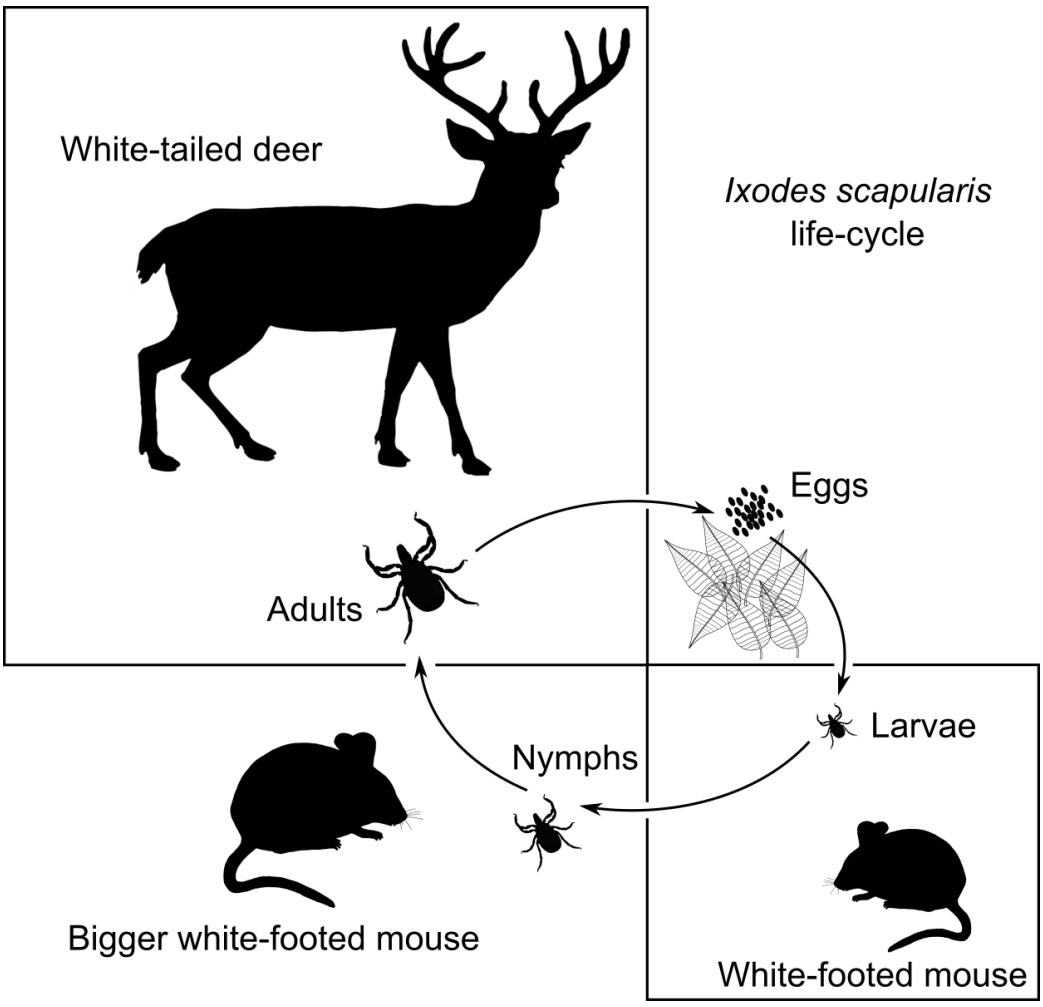

Figure 3.1 – The life-cycle of *Ixodes scapularis* ticks in Northeast United States. Eggs hatch into larvae which take a blood meal from small vertebrates, predominantly white-footed mice (*Peromyscus leucopus*). After overwintering, fed larvae molt to emerge as unfed nymphs which take a second blood meal from bigger vertebrates. Later that same season, fed nymphs molt to become adults where only females take a third blood meal from large vertebrates such as white-tailed deer (*Odocoileus virginianus*). Fed adult females then lay their eggs in leaf litter which become the larvae of the following season. Figure adapted from [101].

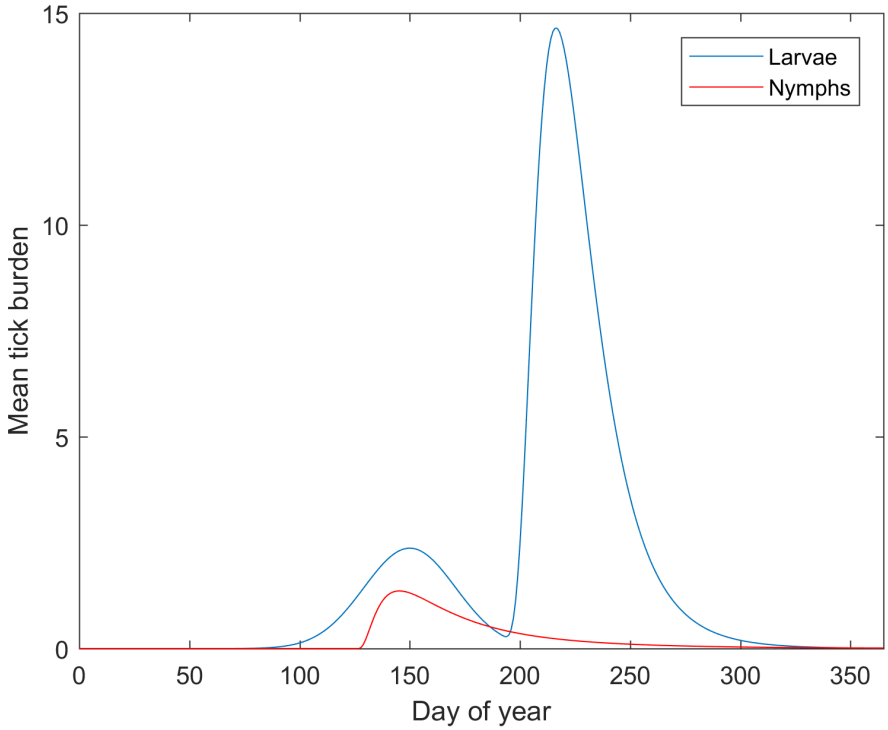

Figure 3.2 – Illustrative curves of the mean larval and nymphal burden on vertebrate hosts in Northeast United States. Day 0 has been assigned to 1 January. Larvae emerge in two pulses over the course of a year (blue curve), with the majority of larvae emerging in the second pulse during late spring around July. Nymphs emerge in a single pulse during early spring around May (red curve).

majority of larvae feed in spring) [94]. In Britain and Europe, the questing behaviour of *I. ricinus* varies dramatically with climate and habitat, even between localized regions within a single country. For example, the phenology of larvae and nymphs in Britain has been described as taking any one of three forms: a bimodal distribution with large early peak and small second peak, a wide unimodal distribution, or a thin unimodel distribution [100, 103, 106, 107].

Larval and nymphal tick phenology can, and has been, described mathematically. The approach implemented in [94, 101, 108] for *I. scapularis* ticks in Northeast United States, and which was initially proposed by Brunner and Ostfeld [98], is briefly described here. Random variables are defined for the number of larval and nymphal ticks feeding on a vertebrate host t days since the beginning of the year. These are denoted $Z_L(t)$ and $Z_N(t)$ respectively, and both are assumed to be negative binomially distributed. This is a reasonable assumption since, like many other parasites [96, 97], *Ixodes* ticks aggregate on their vertebrate hosts [97–100, 109]. The mean number of larvae and nymphs feeding on a host t days since the beginning of the year, $\bar{Z}_L(t)$ and $\bar{Z}_N(t)$ respectively, are then described by a bimodal curve comprising an early normal distribution followed by a later log-normal distribution (for larvae) and a unimodal log-normal distribution (for nymphs):

$$\bar{Z}_L(t) = \begin{cases} H_E e^{-\frac{1}{2}\left(\frac{t-\tau_E}{\mu_E}\right)^2} & \text{if } t \leq \tau_L, \\ H_E e^{-\frac{1}{2}\left(\frac{t-\tau_E}{\mu_E}\right)^2} + H_L e^{-\frac{1}{2}\left[\ln\left(\frac{t-\tau_L}{\mu_L}\right)/\sigma_L\right]^2} & \text{if } t > \tau_L, \end{cases} \tag{3.1}$$

$$\bar{Z}_N(t) = \begin{cases} 0 & \text{if } t \leq \tau_N, \\ H_N e^{-\frac{1}{2}\left[\ln\left(\frac{t-\tau_N}{\mu_N}\right)/\sigma_N\right]^2} & \text{if } t > \tau_N. \end{cases} \tag{3.2}$$

The parameters that appear in Equations (3.1) and (3.2) control the timing, height, and width of the phenology curves; the reader is referred to Table 3.1 for their definitions. Estimates for each of the parameters can be obtained by fitting the two equations to larval and nymphal tick count data using maximum likelihood techniques [101].

3.2.2 Transmission routes

There are three ways tick-borne pathogens can be transmitted between ticks: systemically, via co-feeding, and transovarially.

Table 3.1 – Parameter definitions for the mathematical formulae describing *I. scapularis* tick phenology in Northeast United States.

Parameter	Definition
Larval phenology	
H_E	Maximum height of the early larval peak.
τ_E	Timing of the early larval peak maximum.
μ_E	Width of the early larval peak.
H_L	Maximum height of the late larval peak.
τ_L	Start of the late larval peak.
μ_L	Timing of the late larval peak maximum.
σ_L	Width of the late larval peak.
Nymphal phenology	
H_N	Maximum height of the nymphal peak.
τ_N	Start of the nymphal peak.
μ_N	Timing of the nymphal peak maximum.
σ_N	Width of the nymphal peak.

Systemic transmission

In the systemic route of transmission, a pathogen is transmitted between ticks when an infected tick takes a blood meal from a susceptible vertebrate host, a systemic infection then develops in the host, and the pathogen is transmitted to any susceptible ticks that subsequently feed on the host after a short incubation period has elapsed. For several tick-borne pathogens (including *Borrelia burgdorferi*, the causative agent of Lyme disease) systemic transmission between the immature larval and nymphal tick life-stages is the predominant route by which the pathogen spreads [94, 110].

Co-feeding transmission

A second route for the transmission of pathogens between ticks is that of co-feeding. As for systemic transmission, this route requires that two ticks, one infectious, the other susceptible, feed on the same vertebrate host. Co-feeding transmission, though, additionally requires that the two ticks feed at the same time and in close proximity to one another since it does not involve systemic infection of the host. Co-feeding transmission is particularly important for pathogens, such as tick-borne encephalitis (TBE) virus in Europe, where systemic infection is cleared by a host's immune system within a couple of days, but it plays a smaller role in the emergence and maintenance of Lyme disease where host infection may be lifelong [103, 104, 111].

Transovarial transmission

Tick-borne pathogens can also be transmitted from an infected, engorged, adult female tick to her offspring (eggs). This transmission route, referred to as transovarial or vertical transmission, is generally inefficient [110, 112, 113], especially for *B. burgdorferi* [101, 114], but may still be numerically important because of the large number of eggs each adult female produces [103, 104]. In spite of this potential for amplification, Matser et al. [110] have shown that transovarial transmission only marginally contributes to the spread of several tick-borne pathogens, including *B. burgdorferi*. Consequently, this transmission route will not be considered in the network model that follows (see Section 3.3), and so the model and its results will only apply to those pathogens for which the contribution of this transmission route is negligible.

3.3 Directed tick-host contact and transmission networks

Tick feeding behaviour can be represented as a directed bipartite contact network where a node in the network represents either an immature tick or a vertebrate host, an edge represents a tick taking a blood meal from a host, and edge direction indicates the direction of *potential* pathogen transmission (Figure 3.3). In these networks adult ticks are not considered because they typically feed on larger, and therefore different, hosts to immature tick instars, and, secondly, because the transmission cycle of several tick-borne pathogens of public health concern, including Lyme disease, does not involve the adult tick life-stage [94]. An additional simplification is that each year the vertebrate host population is assumed to be replaced by a new generation of vertebrate hosts. This implies the population of hosts from which ticks feed as larvae in one season is different to the population of hosts they take a blood meal from the following season as nymphs. This is reasonable given the principal competent hosts of immature ticks are typically short-lived (≤ 12 months) small vertebrates (see Section 3.2.1).

In a tick-host contact network, the in- and out-degree of a node representing a vertebrate host correspond to the numbers of nymphs and larvae respectively that feed on the vertebrate host over the course of a year. The aggregation of nymphs can be incorporated by having the in-degree of vertebrate host nodes follow a negative binomial distribution, such that a disproportionate number have high in-degree. Similarly, a negative binomially distributed out-degree for vertebrate host nodes captures the aggregation of larvae. Co-aggregation of the immature tick life-stages can then be manifested as a positive correlation, representing dependence, between the in- and out-degrees of vertebrate hosts.

The transmission of a tick-borne pathogen through a tick-host contact network results in a transmission network. Illustrated in Figure 3.3 is a tick-host contact network together with one of many possible realizations of transmission through that network. In the resulting transmission network, nodes have the same interpretation as in the contact network; an edge, however, no longer represents a tick taking a blood meal from a vertebrate host, but rather the transmission of a tick-borne pathogen from tick-to-tick, tick-to-host, or host-to-tick depending on where it begins and ends. Edges between two ticks represent co-feeding transmission, whereas edges between a vertebrate host and a tick represent one step in the systemic transmission process. The inclusion of edges between two ticks means that, unlike contact networks, transmission networks are not bipartite. The relative importance of

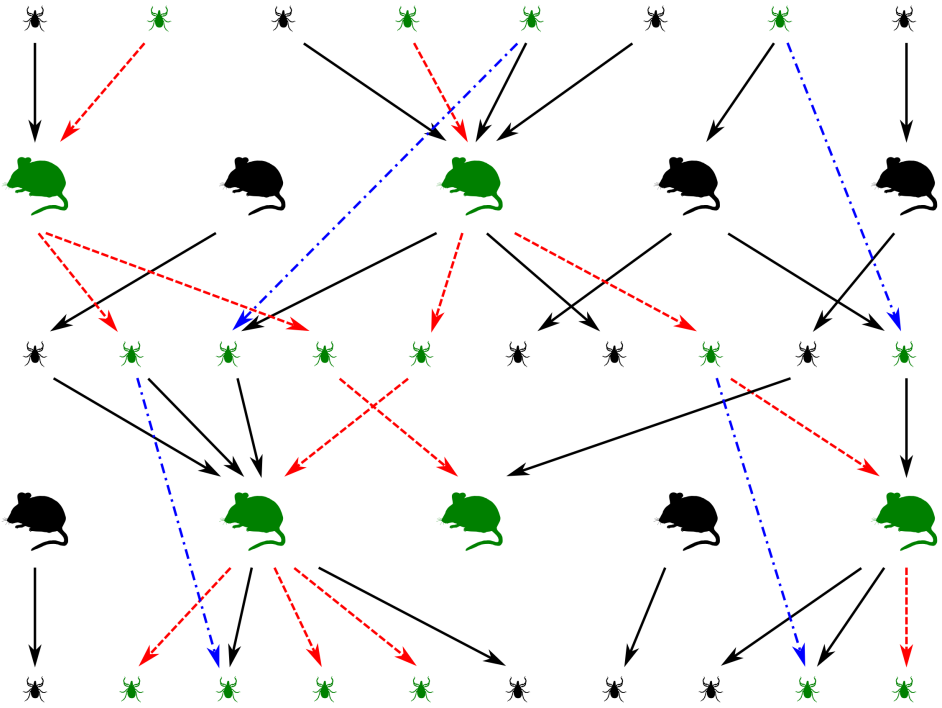

Figure 3.3 – A tick-borne pathogen transmission network superimposed on the underlying tick-host contact network. Each node in the network represents either a tick or a vertebrate host, with infected ticks and hosts indicated in green. Red (dashed) and blue (dot-dashed) edges denote tick-borne pathogen transmission—a red edge between a tick and a host represents systemic transmission, whereas a blue edge between two ticks represents co-feeding transmission. A black (solid) edge between a tick and a host represents a blood meal during which systemic transmission did not occur.

co-feeding transmission to tick-borne pathogen spread will vary by pathogen [110]. For Lyme disease, co-feeding transmission may be more important in some geographic regions (e.g. Europe) than others (e.g. Northeast United States) [94].

An edge in a tick-host contact network does not indicate the time of year the associated blood meal was taken by the tick from its vertebrate host. For both systemic and co-feeding transmission, though, the timing of when ticks take a blood meal relative to one another is important. For example, if the majority of larvae and nymphs were to quest at the same time of year then co-feeding would play a larger role in the spread of tick-borne pathogens than if their questing behaviour did not overlap. The opposite would be true for systemic transmission. To generate transmission networks from tick-

host contact networks the probability an edge appears in a transmission network needs to be related to the time interval between two ticks taking their respective blood meals. This is possible if a mathematical description of tick phenology is available (see Section 3.2.1) and a method for doing so is described in Appendix B.

3.4 Deriving R_0

In this section, formulae for R_0 will be derived for three networks. In the first, tick co-aggregation and co-feeding transmission will both be considered negligible. The second network will assume ticks co-aggregate on hosts. The third network will incorporate tick co-aggregation and co-feeding transmission.

3.4.1 R_0 without co-aggregation, without co-feeding

The parameter R_0 is defined for single-host infections as the expected number of individuals infected by a typical infectious case in an otherwise susceptible population. For vector-borne diseases and multi-host pathogens in general the definition of R_0 is less straightforward as it must average over the multiple host types involved in pathogen transmission. Diekmann et al. [115] introduced the next generation matrix approach as a solution to this problem, and Hartemink et al. [104] were the first to apply the methodology to tick-borne pathogens. Following their example, Matser et al. [110] used sensitivity analysis techniques to identify the respective contributions of the three transmission routes (systemic, co-feeding, and transovarial) to the spread of a variety of tick-borne pathogens. Davis and Bent [94] subsequently implemented the next generation matrix methodology along with loop analysis to identify the transmission loops (repeating chains of transmission) which sustain several tick-borne pathogens in Northeast United States. The pertinent result of Davis and Bent [94] was that some tick-borne pathogens (including *B. burgdorferi*) rely nearly exclusively on a single transmission loop wherein larvae are infected by vertebrate hosts that were themselves infected by nymphs. Consequent to this, Davis and Bent [94] proposed the following simplified next generation matrix could be used to model the transmission of these pathogens

$$\mathbf{K} = \begin{bmatrix} 0 & k_{12} \\ k_{21} & 0 \end{bmatrix}, \tag{3.3}$$

where host type 1 is a tick infected as a larva and host type 2 is a vertebrate host infected by a nymph; k_{ij} is the expected number of hosts of type i

infected by a typical infectious host of type j in an otherwise susceptible population.

The basic reproduction number associated with Equation (3.3) is the geometric mean of k_{12} and k_{21}, that is

$$R_0 = \sqrt{k_{12}k_{21}}. \tag{3.4}$$

The non-zero k_{ij} can be derived using epidemiological reasoning. In the context of tick-host contact networks, this amounts to relating the k_{ij} to the mean in- and out-degrees of the nodes representing ticks and vertebrate hosts.

To derive k_{12}, consider an infectious vertebrate host with out-degree k_{out}. On average such a host will infect $\nu_{\text{lh}}k_{\text{out}}$ larvae, where ν_{lh} is the host-to-larva transmission probability. If the probability a typical infectious host has out-degree k_{out} is denoted by $\mathrm{P}(k_{\text{out}})$, then from the expected value of a discrete random variable it follows that

$$k_{12} = \sum_{k_{\text{out}}} \nu_{\text{lh}}k_{\text{out}}\mathrm{P}(k_{\text{out}})\,. \tag{3.5}$$

There is a subtle difference between a 'typical' infectious host and a host selected uniformly at random that needs to be emphasized at this point. A typical infectious host is one where the risk of the host having been infected in the first place is proportional to the number of nymphs that fed on it, whereas a uniformly randomly selected host is one where all hosts are at equal risk of being infected. This is equivalent to the difference between a node reached by selecting an edge uniformly at random versus selecting a node uniformly at random, a topic that is returned to in Chapter 4. Here, in the context of tick-borne pathogens, this difference is captured by denoting the probability a uniformly randomly selected host has out-degree k_{out} by $p_{k_{\text{out}}}$. The equality of $\mathrm{P}(k_{\text{out}})$ and $p_{k_{\text{out}}}$ does not hold in general, and in particular not when co-aggregation occurs (see Section 3.4.2). For calculating k_{12} the relevant probabilities are those for a typical infectious host.

In the absence of tick co-aggregation, though, the in- and out-degree of a vertebrate host are independent which implies $\mathrm{P}(k_{\text{out}}) = p_{k_{\text{out}}}$. Equation (3.5) can therefore be expressed as

$$k_{12} = \nu_{\text{lh}}\langle k_{\text{out}}\rangle, \tag{3.6}$$

where $\langle k_{\text{out}}\rangle = \sum k_{\text{out}}p_{k_{\text{out}}}$ is the mean out-degree of a uniformly randomly selected vertebrate host.

The other non-zero entry of the next generation matrix is k_{21}, the probability a tick infected as a larva infects a vertebrate host as a nymph. This

probability is given by

$$k_{21} = \sigma \nu_{\text{hn}}, \tag{3.7}$$

where σ is the probability a fed larva survives, successfully molts, and then attaches and takes a blood meal from a competent vertebrate host as a nymph the following season, and ν_{hn} is the nymph-to-host transmission probability. Substituting both this and Equation (3.6) into Equation (3.4) then yields

$$R_0 = \sqrt{\sigma \nu_{\text{lh}} \nu_{\text{hn}} \langle k_{\text{out}} \rangle}, \tag{3.8}$$

and so R_0 is proportional to the square root of the mean annual larval burden.

3.4.2 R_0 with co-aggregation

To extend the results of Section 3.4.1, it will now be assumed that ticks co-aggregate on vertebrate hosts. The immediate implication is that a vertebrate host's in- and out-degree are no longer independent and $\mathrm{P}(k_{\text{out}}) \neq p_{k_{\text{out}}}$.

Consider vertebrate hosts with in-degree $k_{\text{in}} = 0$. With no infectious nymphs taking a blood meal these hosts cannot be infected and so they cannot be typical infectious hosts. This alludes to the general principal that the probability a vertebrate host with in-degree k_{in} is infected is proportional to k_{in}, such that a typical infectious host has in-degree k_{in} with probability

$$\mathrm{P}(k_{\text{in}}) = \frac{k_{\text{in}} p_{k_{\text{in}}}}{\langle k_{\text{in}} \rangle}, \tag{3.9}$$

where $p_{k_{\text{in}}}$ is the probability a uniformly randomly selected vertebrate host has in-degree k_{in} and $\langle k_{\text{in}} \rangle = \sum k_{\text{in}} p_{k_{\text{in}}}$ is the mean in-degree of such a host. The implication of Equation (3.9) is that a typical infectious host is more likely to have high in-degree than a host that has been uniformly randomly selected. The difference between $\mathrm{P}(k_{\text{in}})$ and $p_{k_{\text{in}}}$ is well appreciated in the complex network literature for infectious diseases [116–118] and is analogous to the difference between starting an epidemic by selecting an edge versus selecting a node uniformly at random.

The co-aggregation of ticks on vertebrate hosts means that the in- and out-degree of vertebrate host nodes are positively correlated. This, together with Equation (3.9), implies that a typical infectious host is more likely to have high out-degree than a uniformly randomly selected host. Tick co-aggregation can be mathematically accounted for by writing

$$\mathrm{P}(k_{\text{out}}) = \sum_{k_{\text{in}}} \mathrm{P}(k_{\text{out}} | k_{\text{in}}) \, \mathrm{P}(k_{\text{in}}) \, , \tag{3.10}$$

where $P(k_{\text{out}}|k_{\text{in}})$ is the conditional probability a vertebrate host (referring to the entire population of hosts) with in-degree k_{in} has out-degree k_{out}.

Substituting Equations (3.9) and (3.10) into Equation (3.5) yields

$$\begin{aligned} k_{12} &= \sum_{k_{\text{in}}} \sum_{k_{\text{out}}} \nu_{\text{lh}} k_{\text{out}} \text{P}(k_{\text{out}}|k_{\text{in}}) \, \frac{k_{\text{in}} p_{k_{\text{in}}}}{\langle k_{\text{in}} \rangle} \\ &= \frac{\nu_{\text{lh}}}{\langle k_{\text{in}} \rangle} \sum_{k_{\text{in}}} \sum_{k_{\text{out}}} k_{\text{out}} k_{\text{in}} \text{P}(k_{\text{out}}|k_{\text{in}}) \, p_{k_{\text{in}}}. \end{aligned} \tag{3.11}$$

This expression can be simplified because the product of $\text{P}(k_{\text{out}}|k_{\text{in}})$ and $p_{k_{\text{in}}}$ is equivalent to the joint probability distribution that a uniformly randomly selected vertebrate host has in-degree k_{in} and out-degree k_{out}, denoted $p_{k_{\text{in}},k_{\text{out}}}$. Specifically,

$$k_{12} = \frac{\nu_{\text{lh}}}{\langle k_{\text{in}} \rangle} \sum_{k_{\text{in}}} \sum_{k_{\text{out}}} k_{\text{out}} k_{\text{in}} p_{k_{\text{in}},k_{\text{out}}},$$

and now the double summation reduces to

$$k_{12} = \nu_{\text{lh}} \frac{\langle k_{\text{in}} k_{\text{out}} \rangle}{\langle k_{\text{in}} \rangle} \tag{3.12}$$

because $\langle k_{\text{in}} k_{\text{out}} \rangle$ is the mean of the product of the in- and out-degree of vertebrate hosts. Because tick co-aggregation does not affect k_{21} it follows from Equation (3.4) that

$$R_0 = \sqrt{\sigma \nu_{\text{lh}} \nu_{\text{hn}} \frac{\langle k_{\text{in}} k_{\text{out}} \rangle}{\langle k_{\text{in}} \rangle}}. \tag{3.13}$$

The increase in $\langle k_{\text{in}} k_{\text{out}} \rangle$ with co-aggregation, relative to its value when there is no co-aggregation, is a measure of the in- and out-degree correlation of vertebrate hosts (although not a formal measure such as the Pearson correlation co-efficient or the covariance) [119]. This can be illustrated with a simple example involving the larval and nymphal burdens of two hypothetical mice. Suppose the first mouse is host to 5 larvae and 1 nymph whereas the second mouse is host to 1 larva and 5 nymphs. In this scenario the product of the larval and nymphal burden of each mouse is 5 such that $\langle k_{\text{in}} k_{\text{out}} \rangle$, the mean of these two products, is also equal to 5. If instead the tick burdens on these two mice are correlated, such that the mouse with the highest larval burden is also the mouse with the highest nymphal burden, then the product of the larval and nymphal burden is 25 for the one mouse and 1 for the other, which implies $\langle k_{\text{in}} k_{\text{out}} \rangle = 13$. Although a simple example, the effect of tick co-aggregation is clear: it increases the mean product of larval and nymphal burdens on vertebrate hosts, and in so doing R_0 as well.

3.4.3 R_0 with co-aggregation and co-feeding

In the next generation matrix proposed by Davis and Bent [94], i.e. Equation (3.3), co-feeding transmission was explicitly ignored (since $k_{11} = 0$). This next generation matrix, though, was based on the analysis of tick-borne pathogens (including *B. burgdorferi*) spreading in Northeast United States. In parts of Britain and Europe, however, the contribution of co-feeding transmission to the spread of Lyme disease is understood to be important [120]. Furthermore, States et al. [121] have recently shown co-feeding transmission of *B. burgdorferi* strains, which would otherwise be rapidly cleared by the immune response of the predominant host in Northeast United States, namely *P. leucopus*, facilitates co-existence of multiple strains. This could be especially important for regions where synchronous questing behaviour of immature *I. scapularis* ticks occurs (e.g. Upper Midwest United States). Lastly, for pathogens not considered by Davis and Bent [94], e.g. TBE virus in Europe, co-feeding transmission is known to play a critical role in the spread of infection [104]. Thus to incorporate co-feeding transmission would render the next generation matrix model applicable to a greater range of geographic regions and tick-borne pathogens.

For tick-host networks that incorporate both tick co-aggregation and co-feeding transmission the corresponding next generation matrix is given by

$$\mathbf{K} = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & 0 \end{bmatrix}, \tag{3.14}$$

where the subscripts have the same interpretation as before, and the basic reproduction number is

$$R_0 = \frac{1}{2}\left(k_{11} + \sqrt{k_{11}^2 + 4k_{12}k_{21}}\right). \tag{3.15}$$

The inclusion of co-feeding transmission has no effect on the formulae for k_{12} and k_{21}. The additional non-zero next generation matrix element, k_{11}, is the expected number of larvae infected by a tick that was itself infected whilst feeding as a larva. A tick infected whilst feeding as a larva can only transmit the infection if it survives to take a second blood meal from a competent vertebrate host the following season as a nymph. This occurs with probability σ. When an infected nymph takes a blood meal from a vertebrate host with out-degree k_{out} it will infect $\nu_{\text{ln}}k_{\text{out}}$ larvae on average, where ν_{ln} is the nymph-to-larva co-feeding transmission probability. From the expected value of a discrete random variable it follows that

$$k_{11} = \sigma \sum_{k_{\text{out}}} \nu_{\text{ln}} k_{\text{out}} \mathrm{P}(k_{\text{out}}), \tag{3.16}$$

where $\mathrm{P}(k_{\mathrm{out}})$ is the probability a nymph that takes a blood meal does so from a vertebrate host with out-degree k_{out}.

As in Section 3.4.2, tick co-aggregation means that the in- and out-degree (of vertebrate hosts) are correlated. As before, Equations (3.10) and (3.9) are substituted into Equation (3.16) to account for this which yields

$$k_{11} = \sigma\nu_{\mathrm{ln}}\frac{\langle k_{\mathrm{in}}k_{\mathrm{out}}\rangle}{\langle k_{\mathrm{in}}\rangle} \tag{3.17}$$

such that the basic reproduction number is given by

$$R_0 = \frac{1}{2}\left(\sigma\nu_{\mathrm{ln}}\frac{\langle k_{\mathrm{in}}k_{\mathrm{out}}\rangle}{\langle k_{\mathrm{in}}\rangle} + \sqrt{\left(\sigma\nu_{\mathrm{ln}}\frac{\langle k_{\mathrm{in}}k_{\mathrm{out}}\rangle}{\langle k_{\mathrm{in}}\rangle}\right)^2 + 4\sigma\nu_{\mathrm{lh}}\nu_{\mathrm{hn}}\frac{\langle k_{\mathrm{in}}k_{\mathrm{out}}\rangle}{\langle k_{\mathrm{in}}\rangle}}\right). \tag{3.18}$$

From Equation (3.18) it is straight-forward to see that in the absence of systemic transmission (i.e. when $\nu_{\mathrm{lh}}\nu_{\mathrm{hn}} = 0$) R_0 simplifies to just k_{11}.

3.4.4 Relative effect

A useful way to quantify the relative effect of co-aggregation is to take the ratio of R_0 when co-aggregation is present to when co-aggregation is absent. To do this a formula for R_0 is required for when tick co-aggregation does not occur but where co-feeding transmission does. This will be denoted $R_{0,\mathrm{nca}}$. Using arguments similar to those presented in Sections 3.4.1–3.4.3 it is not hard to show that

$$R_{0,\mathrm{nca}} = \frac{1}{2}\left(\sigma\nu_{\mathrm{ln}}\langle k_{\mathrm{out}}\rangle + \sqrt{\left(\sigma\nu_{\mathrm{ln}}\langle k_{\mathrm{out}}\rangle\right)^2 + 4\sigma\nu_{\mathrm{lh}}\nu_{\mathrm{hn}}\langle k_{\mathrm{out}}\rangle}\right). \tag{3.19}$$

The relative effect of tick co-aggregation on R_0 is then obtained by dividing Equation (3.18) by Equation (3.19) to obtain

$$\begin{aligned}\epsilon &= \frac{R_0}{R_{0,\mathrm{nca}}} \\ &= \frac{\sigma\nu_{\mathrm{ln}}\langle k_{\mathrm{in}}k_{\mathrm{out}}\rangle + \sqrt{\left(\sigma\nu_{\mathrm{ln}}\langle k_{\mathrm{in}}k_{\mathrm{out}}\rangle\right)^2 + 4\sigma\nu_{\mathrm{lh}}\nu_{\mathrm{hn}}\langle k_{\mathrm{in}}\rangle\langle k_{\mathrm{in}}k_{\mathrm{out}}\rangle}}{\sigma\nu_{\mathrm{ln}}\langle k_{\mathrm{in}}\rangle\langle k_{\mathrm{out}}\rangle + \sqrt{\left(\sigma\nu_{\mathrm{ln}}\langle k_{\mathrm{in}}\rangle\langle k_{\mathrm{out}}\rangle\right)^2 + 4\sigma\nu_{\mathrm{lh}}\nu_{\mathrm{hn}}\langle k_{\mathrm{in}}\rangle^2\langle k_{\mathrm{out}}\rangle}}.\end{aligned} \tag{3.20}$$

If ticks co-aggregate on vertebrate hosts then $\epsilon > 1$. When ticks fail to co-aggregate, the independence of the in- and out-degree of vertebrate hosts implies $\langle k_{\mathrm{in}}k_{\mathrm{out}}\rangle = \langle k_{\mathrm{in}}\rangle\langle k_{\mathrm{out}}\rangle$ such that $\epsilon = 1$. In the unlikely event the majority of larvae feed on a different subset of hosts to the majority of nymphs,

such that the in- and out-degree of vertebrate hosts are negatively correlated, then $0 < \epsilon < 1$. If co-feeding transmission is negligible (i.e. when $\nu_{\text{ln}} = 0$) the relative effect of tick co-aggregation simplifies to

$$\epsilon = \sqrt{\frac{\langle k_{\text{in}} k_{\text{out}} \rangle}{\langle k_{\text{in}} \rangle \langle k_{\text{out}} \rangle}}. \tag{3.21}$$

The equivalent expression for tick-borne pathogens where co-feeding transmission is the predominant route of transmission (i.e. when $\nu_{\text{lh}}\nu_{\text{hn}} = 0$) is

$$\epsilon = \frac{\langle k_{\text{in}} k_{\text{out}} \rangle}{\langle k_{\text{in}} \rangle \langle k_{\text{out}} \rangle}. \tag{3.22}$$

3.5 Simulating R_0

To investigate the relationship between R_0 and the level of aggregation and co-aggregation in tick-host contact networks the transmission of Lyme disease was simulated on mechanistic tick-mouse contact networks.

3.5.1 Tick-mouse contact networks

Directed, acyclic, bipartite tick-mouse contact networks, similar to the one shown in Figure 3.3, were constructed as follows. First, the number of seasons, $s = 2$, the number of mice per season, M, and the mean number of larval ticks per mouse each season, $\langle k_{\text{out}} \rangle$, were set (see Table 3.2). To ensure both seasons had nymphal ticks, the number of nymphs in the first season was set to equal the number of larvae (note, however, that in Figure 3.3 only those nymphs in the first season that successfully took a blood meal are shown). The values of s, M, and $\langle k_{\text{out}} \rangle$ therefore determined the number of nodes in a network.

Edges representing ticks taking blood meals from mice were generated one season at a time. An 'attractiveness' score, a_m, generated from a negative binomial distribution with mean value $\langle k_{\text{out}} \rangle$ and aggregation parameter α, was assigned to each mouse, $m \in \{1, 2, 3, \ldots, M\}$. These scores were then converted into in-degree probability weights, $p_m = a_m / \sum_i a_i$. The process was repeated (using the same mean and aggregation parameter values) such that each mouse was also assigned an out-degree probability weight as a measure of its 'attractiveness' to larval ticks. Because the in- and out-degree probability weights were generated from negative binomial distributions they captured the required level of nymphal and larval tick aggregation on mice respectively.

Table 3.2 – Tick-mouse contact and Lyme disease transmission network parameters.

Parameter, Symbol	Point Estimate / Range (Step Size)
Tick-mouse contact networks	
No. of seasons, s	2
No. mice per season, M	200
Mean annual larval tick burden per mouse, $\langle k_{\text{out}} \rangle$	100–300 (100)
Aggregation parameter, α	0.2–4.7 (0.5)
Target rank correlation coefficient, ρ_{target}	0.0–0.4 (0.05)
Acceptable error bound, δ	0.01
Probability a fed larva survives and feeds as a nymph from a mouse the following season, σ	0.10
No. of contact networks[a]	50
Lyme disease transmission networks	
No. of transmission networks[b]	1,000
Nymph-to-larva transmission probability, ν_{ln}	2.4×10^{-3}
Mouse-to-larva transmission probability, ν_{lh}	7.3×10^{-2}
Nymph-to-mouse transmission probability, ν_{hn}	0.83

[a] Per aggregation and co-aggregation parameter combination.
[b] Per contact network.

To ensure the desired level of co-aggregation, ρ_{target} (see below), Spearman's rank correlation coefficient, ρ, for the in- and out-degree probability weights of the mice was calculated. If ρ was not within some acceptable error bound, δ, of ρ_{target}, then the out-degree probability weights of two uniformly randomly chosen mice were swapped. If this reduced the absolute difference between ρ and ρ_{target} the change was accepted, otherwise it was rejected 99% of the time. This out-degree probability weight swapping procedure was continued until either $|\rho - \rho_{\text{target}}| \leq \delta$ or 100,000 iterations had been performed, whichever came first[1].

To generate edges between nymphs and mice, the in-degree probability weights of the mice were converted to cumulative in-degree probability weights, $P_m = \sum_{i \leq m} p_i$. A Bernoulli experiment was then conducted for each fed larval tick from the previous season to determine which of them survived to take a blood meal as a nymph in the current season (which occurred with probability σ). For each tick deemed to have taken a blood meal as a nymph, a uniformly distributed random number, r, between 0 and 1 was generated to determine which mouse it took a blood meal from; the mouse, m, was determined as the one with cumulative in-degree probability weight satisfying $P_m \geq r > P_{m-1}$. Having determined the mouse, a directed edge from the nymph to the mouse was generated by setting the corresponding element of the network adjacency matrix equal to 1. The same process was used to generate edges from mice to larvae, the only difference being that all larvae took a blood meal from a mouse (since those that fail to do so play no role in the transmission of Lyme disease [94] and consequently need not be included in the network). To complete the tick-host network, the process of generating edges between ticks and mice was repeated for the second season.

Three parameters were varied whilst generating tick-mouse contact networks, namely $\langle k_{\text{out}} \rangle$, α, and ρ_{target} (see Table 3.2). The mean number of larvae per mouse each season, $\langle k_{\text{out}} \rangle$, was assigned the values of 100, 200, or 300 to allow investigation into the relationship between R_0 and the mean annual larval burden of mice. A total of 50 tick-mouse contact networks were generated for every pair of values for the aggregation and co-aggregation parameters. The aggregation parameter, α, was varied from 0.2 to 4.7 in step sizes of 0.5, whilst the target level of co-aggregation, ρ_{target}, was varied from 0.0 to 0.4 in step sizes of 0.05. Consequently, 4,500 tick-host contact networks

[1]Whether or not ρ_{target} is achievable depends on the number of mice each season and the value of the aggregation parameter, α, for the negative binomial distribution used to generate tick 'attractiveness' scores for mice. In general, it is expected that the more mice there are each season the more likely ρ_{target} can be achieved for a given negative binomial aggregation parameter value, although more mice may also mean more iterations are required for convergence.

were generated for each value of $\langle k_{\text{out}} \rangle$.

3.5.2 Lyme disease transmission networks

To calculate R_0 using Equation (3.18) values for k_{11}, k_{12}, and k_{21} are required. These were obtained by counting transmission events of the three types on simulated transmission networks. For each tick-mouse contact network a total of 1,000 transmission simulations were conducted[2], such that values for k_{11}, k_{12}, and k_{21} were calculated from 50,000 transmission network realizations for every combination of $\langle k_{\text{out}} \rangle$, α, and ρ_{target}.

For k_{11} and k_{21}, a typical infectious tick was selected by uniformly randomly selecting from among the nymphs of both seasons in the tick-mouse contact network. This is an acceptable approach since all ticks take a single blood meal as larvae and hence are at equal risk of being infected. Next, the index tick was infected and the number of larvae and mice subsequently infected by it (with probability ν_{ln} and ν_{hn} respectively—see Table 3.2) recorded. Values for k_{11} and k_{21} were obtained by calculating the average number of larvae and the average number of mice infected by the index tick over all simulations respectively.

For k_{12}, a typical infectious mouse was selected by uniformly randomly selecting an edge representing a nymph taking a blood meal in either of the two seasons, and then moving along the edge to the mouse from which it was taken. The mouse was infected and subsequently allowed to infect (with probability ν_{lh}—see Table 3.2) any ticks that took a blood meal from it as larvae. The value of k_{12} was calculated as the average number of larvae infected by the index mouse over all simulations.

Finally, point estimates for transmission probabilities ν_{lh} and ν_{ln} were calculated using Equations (B.7) and (B.9) respectively in Appendix B. As the tick phenology curves used to calculate these probabilities were for the questing behaviour of *I. scapularis* ticks in Northeast United States, the results generated by the network model are specific to Lyme disease spread in this geographic region and may not apply to others where tick phenology is significantly different, e.g. Upper Midwest United States and Europe.

3.5.3 Visualizing R_0

The relationship between the basic reproduction number for Lyme disease and the aggregation and co-aggregation parameters, α and ρ_{target} respectively, is shown in Figures 3.4 and 3.5. Co-feeding transmission (Panels B,

[2]Technically, 2,000 transmission simulations were conducted: 1,000 to determine k_{11} and k_{21}, and a further 1,000 to determine k_{12}.

D, and F) only marginally increases the value of R_0 compared to when only systemic transmission occurs (Panels A, C, and E). In contrast, the value of R_0 is substantially lifted by increases in the mean larval tick burden from 100 (Panels A and B), to 200 (Panels C and D), and then 300 (Panels E and F). This is reasonable since, when only systemic transmission occurs, Equation (3.13) predicts that increasing the mean larval burden by a factor of θ will cause R_0 to go up by a factor of $\sqrt{\theta}$. Similarly, when co-feeding transmission occurs, Equation (3.18) predicts R_0 will increase by a factor that lies between $\sqrt{\theta}$ and θ; the precise value is determined by the relative contributions of co-feeding and systemic transmission. Figures 3.4 and 3.5 also reveal that increasing tick co-aggregation always leads to greater values for R_0, whereas higher levels of tick aggregation only increases the value of R_0 when larvae and nymphs co-aggregate on mice. Lastly, aggregation and co-aggregation have a synergistic effect on R_0 such that their combined effect is greater than the sum of their individual effects.

The influence of aggregation and co-aggregation on R_0 in terms of the relative effect of co-aggregation versus no co-aggregation is shown in Figures 3.6 and 3.7. The left-hand columns of these two figures highlight the invariance of ϵ to changes in the mean larval burden, as expected from Equation (3.21). The right-hand columns also show that changes in mean larval burden make very little difference to the relative effect when there is a mix of systemic transmission and co-feeding transmission. Importantly, the invariance of ϵ with respect to mean larval burden means that aggregation and co-aggregation have a larger *absolute* effect on R_0 when the mean larval burden is higher. This is confirmed by comparing the coloured shading between the rows of panels in Figures 3.4 and 3.5.

3.6 Discussion

The principle contribution of this chapter has been to derive simple equations for the dependence of R_0 for tick-borne pathogens on the levels of tick aggregation and co-aggregation, as well as the effect of coincident co-aggregation when pathogens are transmitted via co-feeding. Equation (3.13) describes how R_0 is affected by the interaction between aggregation and co-aggregation and may be written as

$$R_0 = c_1 \sqrt{\frac{\langle k_{\text{in}} k_{\text{out}} \rangle}{\langle k_{\text{in}} \rangle}} \tag{3.23}$$

so that R_0^2 is proportional to the mean product of annual larval and nymphal burdens, scaled by the mean annual nymphal burden. When the spread

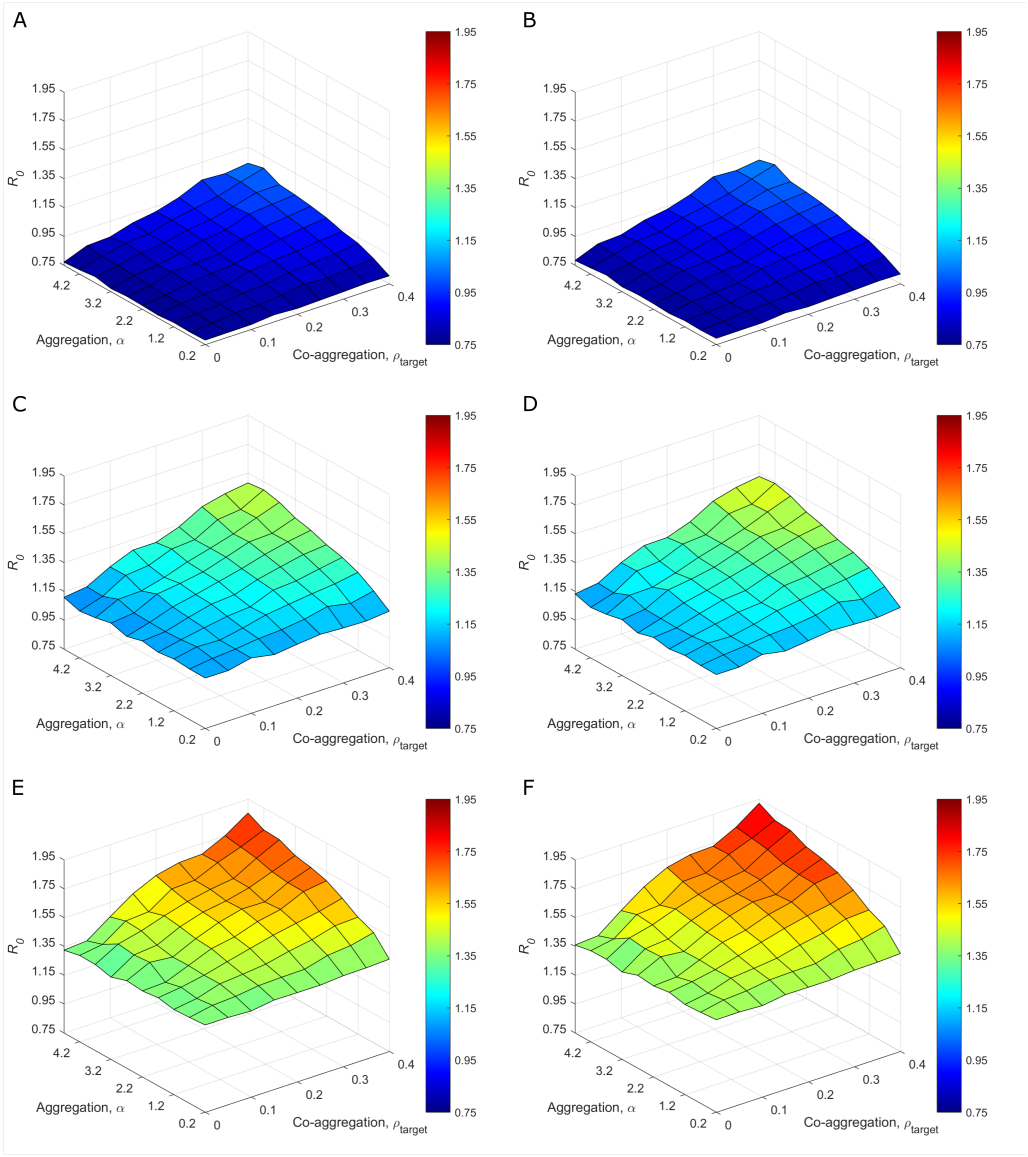

Figure 3.4 – The basic reproduction number for Lyme disease as a function of tick aggregation and co-aggregation. The relationship between R_0 and the aggregation and co-aggregation parameters, α and ρ_{target} respectively, is shown for mean larval tick burdens of (A, B) 100, (C, D) 200, and (E, F) 300. In Panels A, C, and E only systemic transmission occurs, whereas in Panels B, D, and F a small amount of co-feeding transmission is present.

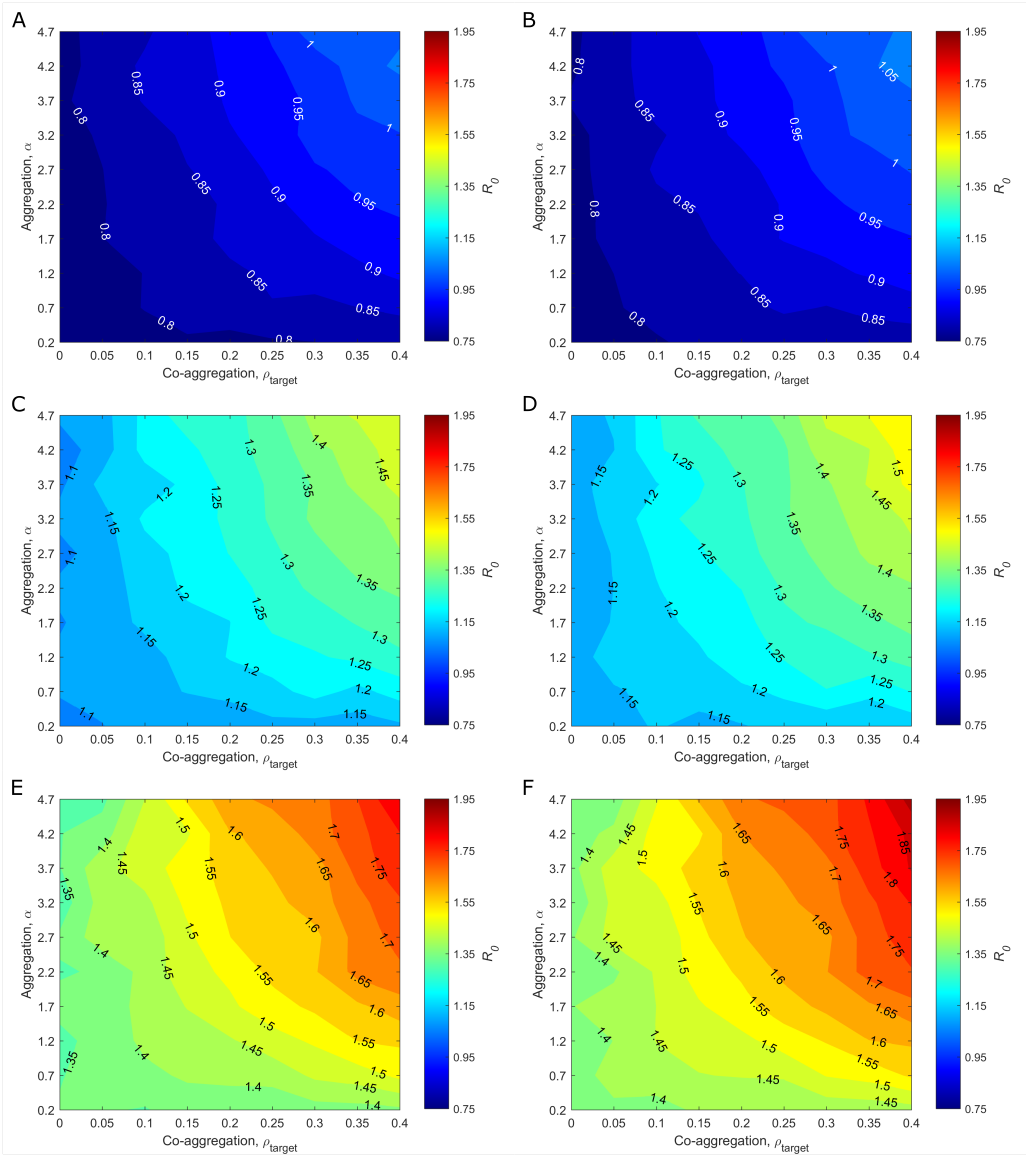

Figure 3.5 – The basic reproduction number for Lyme disease as a function of tick aggregation and co-aggregation. Contour plots for the three dimensional surfaces in Figure 3.4. The relationship between R_0 and the aggregation and co-aggregation parameters, α and ρ_{target} respectively, is shown for mean larval tick burdens of (A, B) 100, (C, D) 200, and (E, F) 300. In Panels A, C, and E only systemic transmission occurs, whereas in Panels B, D, and F a small amount of co-feeding transmission is present.

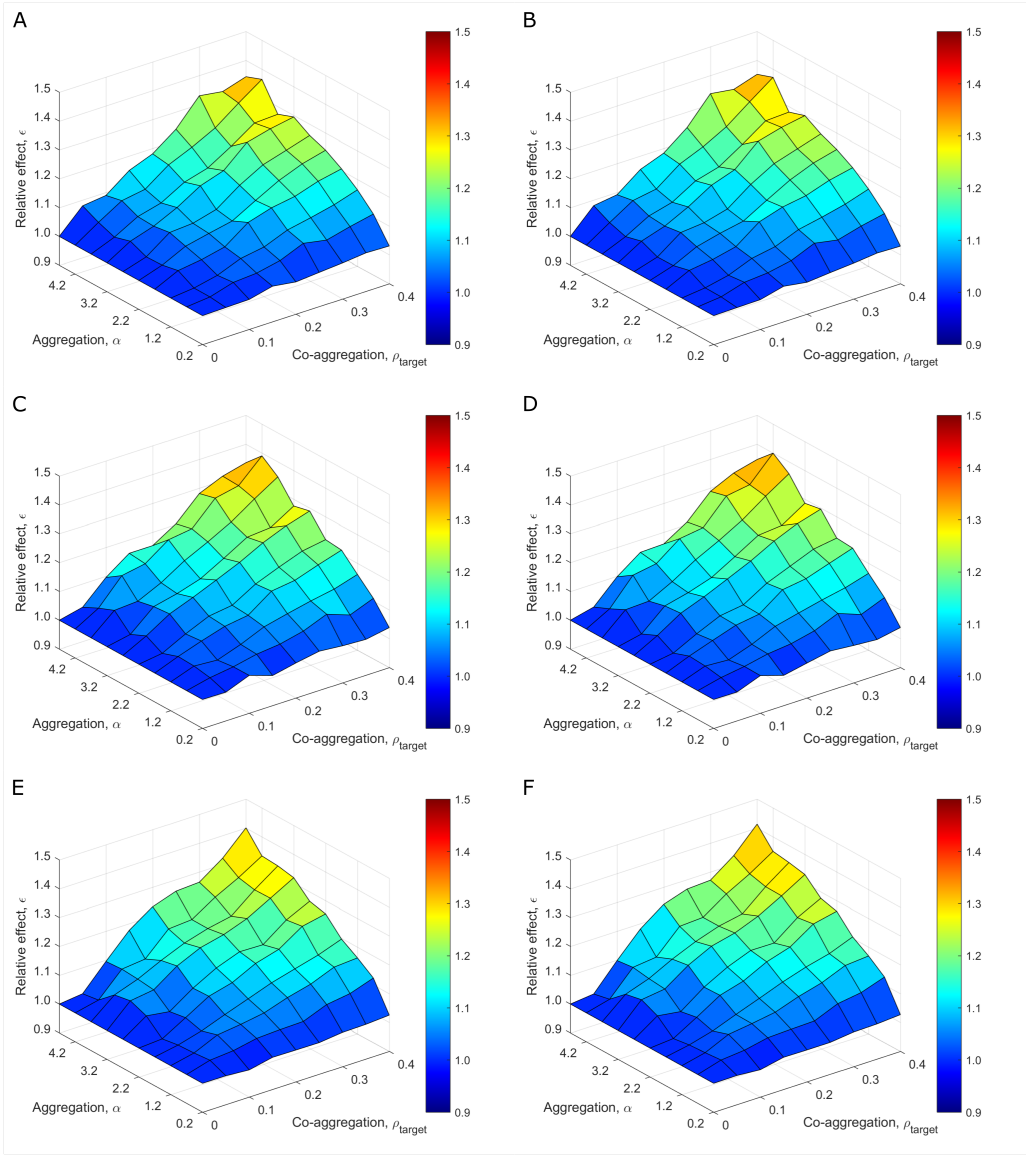

Figure 3.6 – The relative effect of tick co-aggregation on the basic reproduction number for Lyme disease. The relationship between the relative effect parameter, ϵ, and the aggregation and co-aggregation parameters, α and ρ_{target} respectively, is shown for mean larval tick burdens of (A, B) 100, (C, D) 200, and (E, F) 300. In Panels A, C, and E only systemic transmission occurs, whereas in Panels B, D, and F a small amount of co-feeding transmission is present.

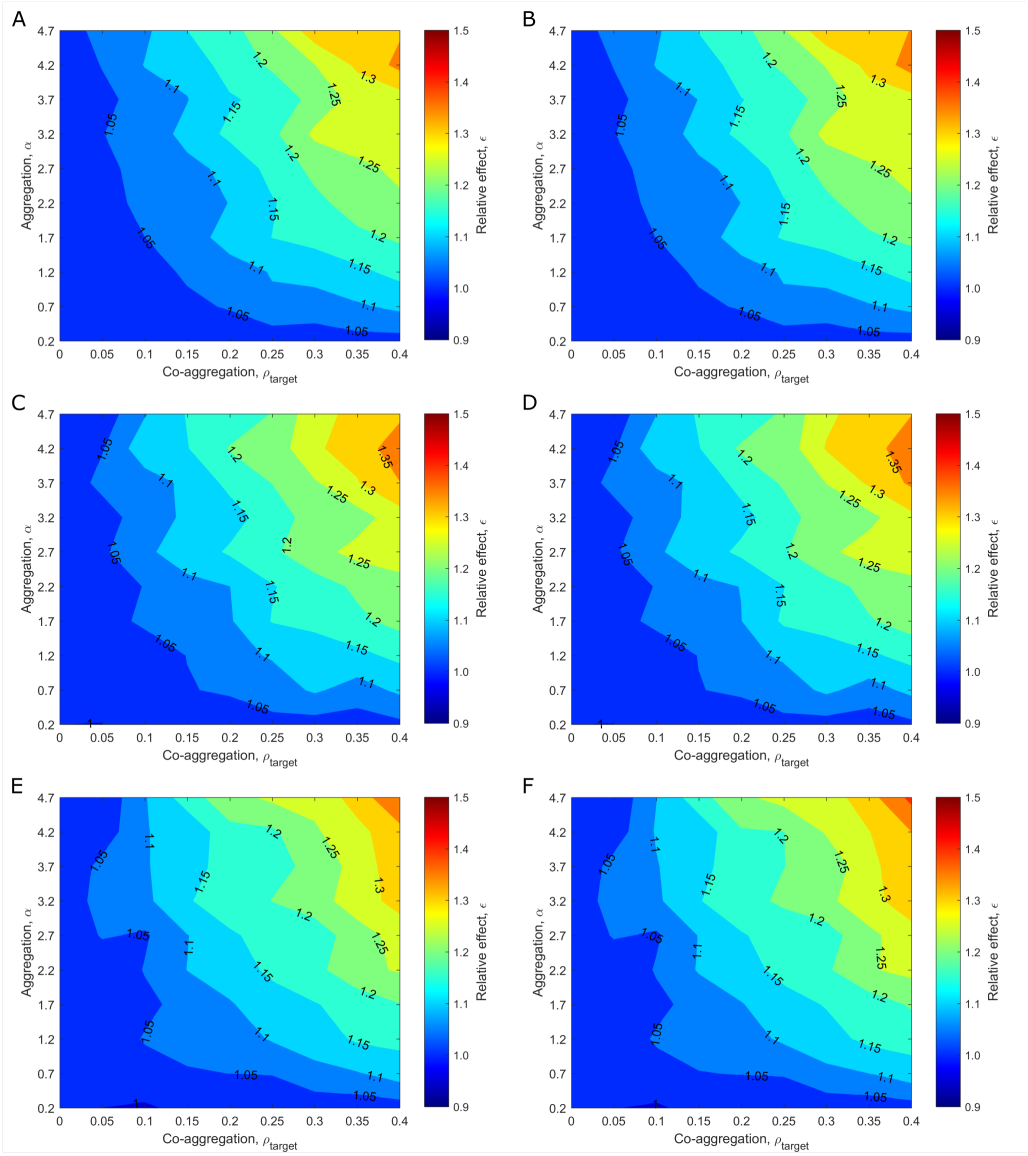

Figure 3.7 – The relative effect of tick co-aggregation on the basic reproduction number for Lyme disease. Contour plots for the three dimensional surfaces in Figure 3.6. The relationship between the relative effect parameter, ϵ, and the aggregation and co-aggregation parameters, α and ρ_{target} respectively, is shown for mean larval tick burdens of (A, B) 100, (C, D) 200, and (E, F) 300. In Panels A, C, and E only systemic transmission occurs, whereas in Panels B, D, and F a small amount of co-feeding transmission is present.

of a pathogen is dominated by the co-feeding route of transmission then Equation (3.18) simplifies to

$$R_0 = c_2 \frac{\langle k_{\text{in}} k_{\text{out}} \rangle}{\langle k_{\text{in}} \rangle}, \tag{3.24}$$

which states that R_0, rather than R_0^2, is proportional to the mean product of annual larval and nymphal burdens, scaled by the mean annual nymphal burden. In epidemiological terms the mean product of annual larval and nymphal burdens scaled by the mean annual nymphal burden is actually the mean annual larval burden of a typical infectious host. The biological interpretation of this term is that the stronger the correlation between larval burden and nymphal burden the greater the difference between a typical infectious host and one selected uniformly at random. The difference between Equations (3.23) and (3.24) implies that co-aggregation will have a larger relative effect on the magnitude of R_0 for pathogens such as TBE virus in Europe that are sustained by co-feeding transmission [104, 111] than it will for pathogens that rely on systemic infections such as Lyme disease [94, 104]. For both pathogens though, the effect of co-aggregation can be to lift R_0 above the threshold value of 1 and so lead to persistence.

In addition to the derived analytic equations for R_0, simulations of Lyme disease transmission on mechanistic tick-mouse contact networks were used to visualize the relationship between R_0 and the level of tick aggregation and co-aggregation. The results of the simulations revealed that co-feeding transmission makes minimal difference to the value of R_0 for Lyme disease in Northeast United States. This is consistent with that which has previously been reported in the literature [94]. Furthermore, it is largely due to the small co-feeding transmission probability ν_{ln} relative to the systemic transmission probabilities ν_{hn} and ν_{lh} (see Table 3.2), which is a consequence of larval and nymphal ticks questing at different times of the year in this geographic region (see Figure 3.2). For other regions, e.g. Upper Midwest United States, where there is a significantly greater overlap in the questing behaviour of larval and nymphal ticks [94], one would expect ν_{ln} to be higher, ν_{lh} to be lower, and the effect of co-feeding transmission on R_0 to consequently be greater (assuming all other parameter values remain unchanged). Similarly, for TBE virus in Europe, where the fast clearance of systemic infections in vertebrate hosts within a couple of days [111] would render ν_{lh} relatively small, one would expect the value of R_0 to be significantly raised by the contribution of co-feeding transmission compared to if only systemic transmission occurred.

A caveat of the simulation results is that they are predicated on two assumptions, namely that the larval and nymphal burdens on hosts both follow a negative binomial distribution and that there is a monotonic relationship

(representing dependence) between the counts of larvae and nymphs on individual hosts. Whilst these two assumptions are reasonable in light of the trends typically observed for vertebrate host larval and nymphal burdens obtained from the field [98–100], should they not apply then the visualized relationship between R_0 and the level of tick aggregation and co-aggregation in Figures 3.4–3.7 would also no longer hold. Importantly, this would not render the derived analytic relationship in Equation (3.18) inaccurate since this is a more general result that—whilst capturing both tick aggregation and co-aggregation—does not make any assumptions about the larval and nymphal distributions on vertebrate hosts or the relationship between them.

Chapter 4

Clustered social networks

4.1 Introduction

A feature common to most social networks is that if one individual contacts two other individuals, these two other individuals are often also contacts of each other. This triangular contact pattern is referred to as clustering. Often, higher-order forms of clustering also occur, e.g. squares involving four individuals. At present, the literature is inconclusive with respect to the impact of clustering on the spread of infection through social networks. Some initial studies [122–124] found that clustering raises the value of R_0, or equivalently lowers the epidemic threshold, thereby increasing the probability of disease establishment. More recent evidence, however, suggests the opposite is true [117, 125–127]. One reason for these differing conclusions is that it is difficult to isolate the effect of clustering from those of other network features, e.g. degree heterogeneity and assortativity [125, 128]. Additionally, the results are often only approximations due to the simplifying assumptions made when deriving them, e.g. that squares can be ignored, triangles never share an edge, or that no two infectious individuals run the risk of infecting the same susceptible individual [117, 124–126].

In this chapter, an exact formula for R_0 is derived for the spread of infection through a clustered social network. A brute-force probabilistic approach is used to calculate the expected number of secondary infections caused by a typical infectious individual in the SIS disease framework. The formula for R_0 incorporates effects resulting from triangular and square clustering, as well as competition due to multiple infectious nodes attempting to infect the same susceptible neighbour nodes. Simulations on clustered Erdos-Renyi (ER) networks validate the analytic formula derived for R_0. Clustering is found to have a protective effect, lowering the value of R_0 compared to a network in which there is minimal clustering. In extreme circumstances (when the triangle and square clustering co-efficients are both ≥ 0.35) the value of R_0 may even start to decline at high transmission probabilities.

4.1.1 Chapter layout

This chapter begins by defining a typical infectious individual. A scenario is then sketched out that illustrates the various ways clustering can impact the number of individuals such an individual can infect. An analytic formula for R_0 that takes these factors into account is derived using a brute-force probabilistic approach, after which transmission simulations are conducted on ER networks to verify the derived formula. Lastly, the implication of the results are discussed in light of the published clustering literature.

4.2 A glossary of terms

For the sake of clarity several terms are defined here that are fundamental to the work presented in this chapter.

Index case. The first node infected during an infectious disease outbreak, chosen uniformly at random from amongst all nodes. Also referred to here as the zeroth generation of infection, with any nodes infected by the index case referred to as the first generation of infection, etc.

Adjacent. Two nodes are said to be adjacent when they are connected by an edge.

Neighbour. Two nodes are said to be neighbours if they are adjacent to one another.

Degree. The degree of a node is the number of neighbours it is adjacent to.

Ancestor. The node responsible for infecting another node. Each infected node has only one ancestor and that ancestor is from the previous generation of infection.

Descendants. The nodes infected by an infectious node. The descendants of a node refer to only those in the next generation of infection.

Excess degree neighbour. Any neighbour of a node infected along an edge other than its ancestor. For the remainder of this chapter this term will be abbreviated EDN.

Excess degree. The excess degree of a node infected along an edge is the number of EDNs it is adjacent to and is equal to its degree minus one.

4.3 A typical infectious individual

The basic reproduction number, R_0, is the expected number of secondary infections caused by a typical infectious individual in an otherwise susceptible population. Understanding what type of individual constitutes a 'typical' infectious individual is therefore fundamental to deriving an accurate formula for R_0.

Consider a random network in which the degrees of neighbours are uncorrelated. In such a network an infectious node is precisely k times more likely to be connected to, and therefore k times more likely to infect, a node of degree k than a node of degree 1. This means that at any stage during an outbreak the proportion of infected nodes that have high degree will be greater than the proportion of nodes that make up the network that have high degree. From this it follows that a typical infectious individual is not a node uniformly randomly selected from amongst all nodes, but instead is a node selected after weighting each node by its degree. That is to say, R_0 should be estimated from the mean number of neighbours infected by a node that itself was infected along an edge (e.g. a neighbour of the index case)[1].

4.4 Clustering and competing interests

In a networks setting, R_0 can be calculated from the mean number of neighbours infected by a neighbour of the index case. Before deriving R_0 analytically it is instructive to first sketch out a scenario that demonstrates how clustering might influence the number of neighbours such an infectious individual is able to infect. To reduce complexity, the spread of an infectious agent is considered whereby individuals become infected and remain infectious for only a single time step before returning to a state of susceptibility.

Consider Figure 4.1. Depicted is one of many possible realizations that could eventuate after a single time step has elapsed following the introduction of a disease into a population through a uniformly randomly selected index case i. During the first time step the index case infected neighbours k, m, n, o, and s (filled nodes), after which it recovered to become susceptible again itself (unfilled).

To estimate R_0 the average number of nodes one of the infectious neighbours of i, say node n, will infect during the second time step needs to be calculated. Because of clustering within the network, some of the nodes infected by the index case are also neighbours of node n (e.g. nodes o and s). Consequently, node n has fewer neighbours to infect during the second time step than it would have had had no triangles been present in the network.

Three other effects, collectively referred to as 'competing interests', also

[1]It should also be noted that the phrase 'in an otherwise susceptible population' in the definition of R_0 is redundant in the context of networks since the index case, for whom this phrase can only ever be correct, is not a typical infectious individual. In practice, the phrase should be interpreted as meaning 'in the initial stages of an outbreak' [129, 130], or more specifically 'in the first generation of infection' for clustered networks. This is discussed further in Section 4.8.

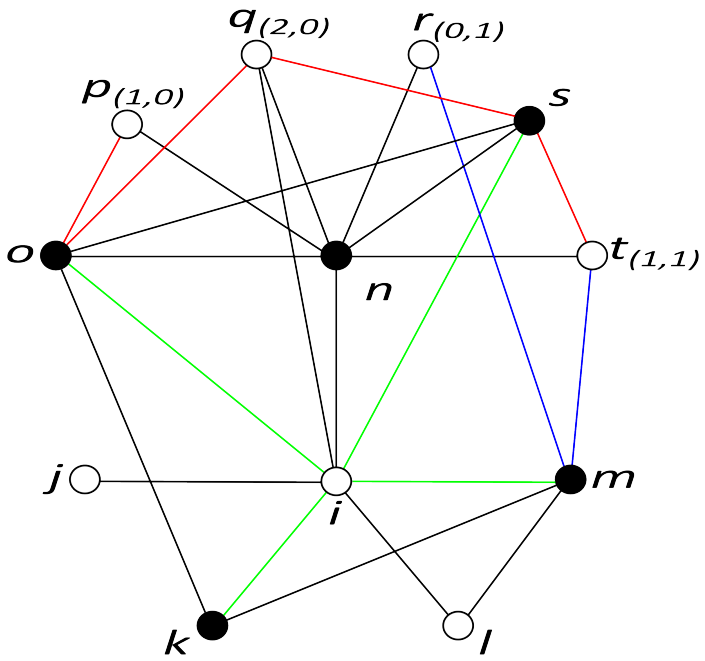

Figure 4.1 – Several factors that affect R_0 for infections spreading through clustered social networks. The sub-network of a large clustered social network, comprising the index case (node i) and all its neighbours, all the neighbours of a typical infectious case (node n), and all the edges between each of these nodes. In this particular realization, one time step after the introduction of a directly transmitted disease, the index case has infected five of its neighbours (filled nodes) before recovering itself (unfilled), two of which are also neighbours of n due to triangle clustering (nodes o and s). This reduces the number of neighbours n can infect during the second time step, as do the three forms of 'competing interests' indicated by the green, red, and blue edges. To account for the red and blue forms of competing interests, each susceptible excess degree neighbour of n (nodes p, q, r, and t) was assigned a 2-tuple 'type' denoted by (α, β). For any particular susceptible excess degree neighbour of n, the values of α and β indicate the number of infectious nodes other than n that could infect that node as a result of triangle (red edges) and square clustering (blue edges) respectively.

impact the number of nodes n may infect and therefore need to be taken into account. Each form of competing interests is identified in Figure 4.1 by edge colour (green, red, and blue)[2]. Green edges indicate nodes, infected by the index case during the first time step, that 'compete' with node n in that they may reinfect the index case during the second time step. This effect is not due to clustering. Red edges indicate neighbours of node n, infected by the index case during the first time step, that could potentially infect some of the susceptible EDNs of n during the second time step. This effect is due to the triangular form of clustering. Lastly, blue edges indicate nodes, infected by the index case during the first time step, that are neither node n nor neighbours of node n, which might also infect some of the susceptible EDNs of node n during the second time step. This effect is due to the square form of clustering.

The competing interests that are introduced by the presence of triangles and squares are accounted for in Figure 4.1 by allocating nodes a 2-tuple subscript (α, β), where α is the number of infected neighbours of n that n is competing with and β is the number of infected neighbours of i (that are neither n nor neighbours i and n have in common) that n is competing with. The success of n in infecting a neighbour of type (α, β) is determined by the total number of competitors, i.e. $\alpha + \beta$, but α and β have different probabilities of occurring, respectively $C_{\triangle}$ and $C_{\square}$, and must therefore be accounted for separately.

Thus there are four effects in total, three due to clustering, which need to be taken into account when deriving an analytic formula for R_0.

4.5 Deriving R_0

An analytic formula for R_0 is now derived from the expected value of a discrete random variable. It will be assumed that (1) the degrees of neighbour nodes are uncorrelated, (2) clustering is degree-independent, and (3) each infectious individual transmits the infection to a susceptible neighbour with probability T. To begin, several discrete random variables are defined, where the mention of any particular node refers to the corresponding node

[2]Central to the discussion of 'competing interests' is the assumption that the infection of a susceptible individual can only be attributed to a single infectious neighbour. To assume otherwise would result in an overestimate of R_0 because the number of *unique* susceptible individuals infected would be overestimated. Since it is most biologically sensible to attribute an infection to the first infectious individual to transmit the infection to the susceptible individual, R_0 should be estimated from the mean number of neighbours n infects during the second time step *before* any other infectious nodes infect them.

in Figure 4.1. Let

- K be a random variable for the degree of node i,
- X be a random variable for the number of infectious neighbours of i, besides node n, at $t = 1$ (green edges, Figure 4.1),
- K' be a random variable for the excess degree of n,
- Ψ be a random variable for the number of neighbours i and n have in common,
- Θ be a random variable for the number of neighbours i and n have in common that are infectious at $t = 1$,
- A be a random variable for the number of infectious EDNs of n that are adjacent to one of its susceptible EDNs at $t = 1$ (red edges, Figure 4.1),
- B be a random variable for the number of infectious neighbours of i, besides n and any neighbours i and n have in common, that are adjacent to one of node n's susceptible EDNs at $t = 1$ (blue edges, Figure 4.1),
- $\Omega_{\alpha,\beta}$ be a random variable for the number of susceptible EDNs of n at $t = 1$ of type $(A = \alpha, B = \beta)$[3], and
- $Z_{\alpha,\beta}$ be a random variable for the number of susceptible EDNs of n of type (α, β) that n infects during the second time step.

The following shorthand notation will also be adopted

$$\sum_{\alpha,\beta} = \sum_{\alpha} \sum_{\beta},$$

$$\mathrm{P}(\chi|\kappa) = \mathrm{P}(X = \chi|K = \kappa)\,,$$

and

$$f_{\mathrm{B}}\left(\theta; \psi, T\right) = \binom{\psi}{\theta} (T)^{\theta} (1 - T)^{\psi - \theta}\,,$$

where f_{B} denotes the probability mass function of a binomial distribution.

[3]That is, the *type* of a susceptible EDN of n at $t = 1$ is defined as the 2-tuple (α, β), where α is the number of infectious EDNs of n and β is the number of infectious neighbours of i, besides n and any neighbours i and n have in common, that are adjacent to the susceptible EDN as a result of triangle and square clustering respectively.

Having introduced the required random variables and notation, it is now noted that

$$\begin{aligned} R_0 &= \mathrm{E}\left[\text{no. of nodes infected by } n\right] \\ &= \mathrm{E}\left[\text{no. of ancestors reinfected by } n\right] \\ &\qquad + \\ &\quad \mathrm{E}\left[\text{no. of descendants infected by } n\right], \end{aligned} \tag{4.1}$$

since any node infected by n is either its ancestor or descendant, and the number of ancestors and descendants infected by n are independent random variables.

4.5.1 The number of ancestors n reinfects

Because every infectious individual has only one ancestor the expected number of ancestors n will reinfect is equal to the probability it reinfects i [118]:

$$\mathrm{E}\left[\text{no. of ancestors reinfected by } n\right] = \mathrm{P}(n \text{ reinfects } i)\,. \tag{4.2}$$

This probability depends on how many other infectious nodes n has to compete against (green edges, Figure 4.1), therefore

$$\mathrm{P}(n \text{ reinfects } i) = \sum_{\kappa,\chi} \mathrm{P}(n \text{ reinfects } i|\chi,\kappa)\,\mathrm{P}(\chi|\kappa)\,\mathrm{P}(\kappa)\,, \tag{4.3}$$

where $\mathrm{P}(\kappa)$ is the probability i has degree $K = \kappa$, $\mathrm{P}(\chi|\kappa)$ is the conditional probability i infected $X = \chi$ other nodes besides n during the first time step given it has degree κ, and $\mathrm{P}(n \text{ reinfects } i|\chi,\kappa)$ is the conditional probability n reinfects i during the second time step given it competes against χ other infectious neighbours of i.

Since R_0 is the mean number of nodes infected by a neighbour of the index case i, it holds that i has at least one neighbour, namely n. This implies the probability i has degree κ is not simply p_κ (i.e. the probability a uniformly randomly selected node has degree κ) but rather

$$\mathrm{P}(\kappa) = \frac{p_\kappa}{1 - p_0}, \tag{4.4}$$

since it cannot have a degree of zero[4]. It is noted here that Equation (4.4) is not the same as the excess degree distribution (which will appear later

[4]Strictly speaking the uniformly randomly selected index case i can have degree zero. However, in this trivial scenario an epidemic cannot occur (since the index case has no neighbours to infect) and it is then meaningless to calculate R_0 from a typical infectious case. For this reason, only scenarios in which the index case has degree $K \geq 1$ and where it has infected at least one of its neighbours, namely n, during the first time step are considered.

in Section 4.5.2) since the index case i is chosen uniformly at random from amongst the nodes, not edges, of a network.

The probability i infects χ other neighbours besides n given it has κ neighbours in total is given by

$$\begin{aligned} \mathrm{P}(\chi|\kappa) &= \binom{\kappa-1}{\chi} (T)^{\chi} (1-T)^{\kappa-1-\chi} \\ &= f_{\mathrm{B}}(\chi; \kappa-1, T) \end{aligned} \tag{4.5}$$

since there are at most $\kappa - 1$ nodes besides n that i can infect.

The key to deriving the conditional probability n reinfects i is to realize that all nodes infected by the index case during the first time step are equivalent and there is nothing unique about node n[5]. This implies the probability any one of them reinfects the index case is exactly the same, and moreover is equal to the probability i is reinfected altogether, divided by the number of infectious neighbours attempting to reinfect it. Given the index case infected χ other neighbours besides n, the probability n reinfects i is[6]

$$\mathrm{P}(n \text{ reinfects } i|\chi, \kappa) = \frac{1}{\chi+1}\left(1-(1-T)^{\chi+1}\right). \tag{4.6}$$

Substituting Equations (4.3)–(4.6) into Equation (4.2) yields

$$\begin{aligned} &\mathrm{E}\left[\text{no. of ancestors reinfected by } n\right] \\ &= \sum_{\kappa=1}^{M} \sum_{\chi=0}^{\kappa-1} \frac{1}{\chi+1}\left(1-(1-T)^{\chi+1}\right) f_{\mathrm{B}}(\chi; \kappa-1, T) \left(\frac{p_\kappa}{1-p_0}\right), \end{aligned} \tag{4.7}$$

where M is the maximum degree of the network.

[5]In the epidemic realization illustrated in Figure 4.1, the index case i is at risk of being reinfected by five of its neighbours during the second time step. From the perspective of the index case, which five of its neighbours happen to be infectious does not alter the probability it will be reinfected (Figure 4.2). Therefore, all infectious neighbours of i can be considered equivalent. It should be noted, however, that this is only true with respect to reinfection of the index case since each infectious neighbour of i is likely to have a different number of EDNs.

[6]The reader should note the subtle difference between the reinfection probability, Equation (4.6), and the transmission probability T. The reinfection probability is the probability node n transmits the infection to i *before* any of i's other infectious neighbours do, whereas T simply captures whether or not n transmits the infection to i. In the event n is the only infectious neighbour of i, i.e. $\chi = 0$, the reinfection probability is equal to T.

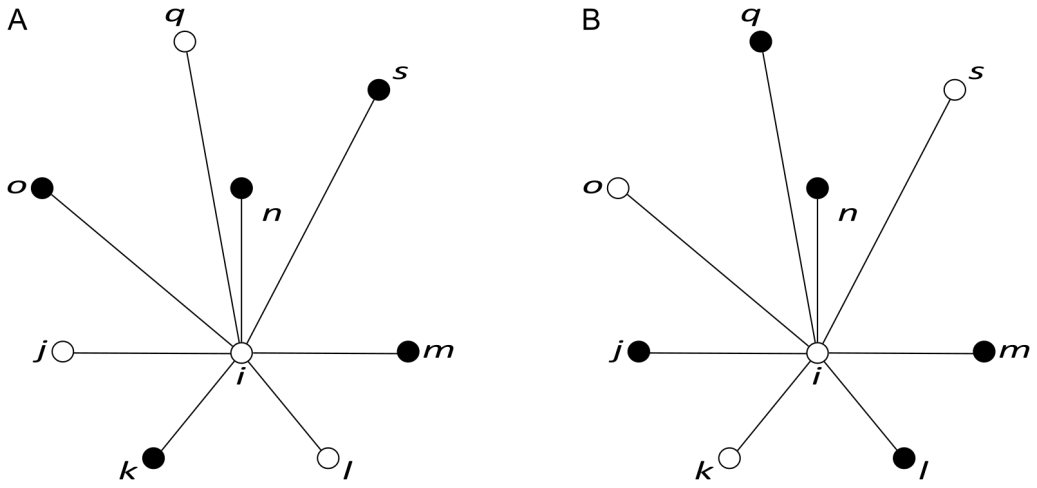

Figure 4.2 – The reinfection probability only depends on the number of infectious neighbours of the index case. The probability i is reinfected during the second time step is the same in both epidemic realizations (A) and (B) since the reinfection probability depends only on the number of i's neighbours that are infectious and not which of them are infectious.

4.5.2 The number of descendants n infects

Next, the number of descendants n is expected to infect during the second time step is derived (second term, Equation (4.1)). This quantity is clearly limited by the number of EDNs node n has, but it also depends on how many of these neighbours were also neighbours of i and infected by i. Knowing how many EDNs of n are susceptible at $t = 1$, however, is not all that is required. Rather, how many of them are of type (α, β) is what is needed. This is because the probability n infects any particular susceptible EDN depends on its type, i.e. the number of other infectious nodes n has to compete against due to triangle and square clustering (red and blue edges respectively, Figure 4.1). Once the number of susceptible EDNs of type (α, β) is known it is straightforward to calculate how many of these neighbours n is expected to infect. Thereafter, the total number of EDNs n is expected to infect can be obtained by summing over all possible values for α and β. This means that

$$\begin{aligned}&\mathrm{E}\left[\text{no. of descendants infected by } n\right]\\&= \sum_{\alpha=0}^{M-2} \sum_{\beta=0}^{M-1} \mathrm{E}\left[\text{no. of type } (\alpha,\beta) \text{ EDNs infected by } n\right], \end{aligned} \tag{4.8}$$

where the upper limits for the two summations are explained as follows. Suppose n (a node infected along an edge) has the maximum degree M of

any node in the network; then its excess degree is $M-1$. But this means that for any one susceptible EDN of n the maximum number of infectious EDNs of n it can be adjacent to is $M-2$, which is the upper limit for α. Similarly, suppose the index case i has degree M. Because one of i's neighbours is n, node i can have a maximum of $M-1$ infectious neighbours that are neither n nor neighbours i and n have in common[7]. Hence the upper limit for β is $M-1$.

From the definition of an expected value, the summand in Equation (4.8) may be written

$$\mathrm{E}\left[\text{no. of type } (\alpha,\beta) \text{ EDNs infected by } n\right] = \sum_{\zeta_{\alpha,\beta}} \zeta_{\alpha,\beta} \mathrm{P}(\zeta_{\alpha,\beta})\,, \tag{4.9}$$

where $\mathrm{P}(\zeta_{\alpha,\beta})$ is the probability n infects $Z_{\alpha,\beta} = \zeta_{\alpha,\beta}$ susceptible EDNs of type (α,β) during the second time step. Because this probability depends on the number of EDNs of n that are susceptible at $t=1$, as well as the number of infectious nodes n has to compete against due to triangle and square clustering,

$$\mathrm{P}(\zeta_{\alpha,\beta}) = \sum_{\kappa',\psi,\theta,\omega_{\alpha,\beta}} \mathrm{P}(\zeta_{\alpha,\beta}|\omega_{\alpha,\beta},\theta,\psi,\kappa')\,\mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa')\,\mathrm{P}(\theta|\psi,\kappa')\,\mathrm{P}(\psi|\kappa')\,\mathrm{P}(\kappa')\,, \tag{4.10}$$

where $\mathrm{P}(\kappa')$ is the probability n has excess degree $K'=\kappa'$, $\mathrm{P}(\psi|\kappa')$ is the conditional probability i and n have $\Psi=\psi$ common neighbours, $\mathrm{P}(\theta|\psi,\kappa')$ is the conditional probability i infected $\Theta=\theta$ of the ψ neighbours i and n have in common, $\mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa')$ is the conditional probability $\Omega_{\alpha,\beta}=\omega_{\alpha,\beta}$ of n's susceptible EDNs at $t=1$ are of type (α,β) given i infected θ of n's EDNs, and $\mathrm{P}(\zeta_{\alpha,\beta}|\omega_{\alpha,\beta},\theta,\psi,\kappa')$ is the conditional probability n infects $\zeta_{\alpha,\beta}$ susceptible EDNs of type (α,β) during the second time step given it has $\omega_{\alpha,\beta}$ such neighbours at $t=1$.

As the degrees of neighbour nodes are uncorrelated (Assumption 1, Section 4.5), the probability n has excess degree κ' is equal to the probability a node, reached by moving along a uniformly randomly selected edge, has degree $(\kappa'+1)$:

$$\mathrm{P}(\kappa') = \frac{(\kappa'+1)\,p_{\kappa'+1}}{\langle\kappa\rangle}, \tag{4.11}$$

where $\langle\kappa\rangle$ is the mean degree of a uniformly randomly selected node. Equation (4.11) is referred to as the excess degree distribution in the complex network literature [116–118].

[7]This occurs precisely when i and n have no neighbours in common and i infects all M of its neighbours during the first time step.

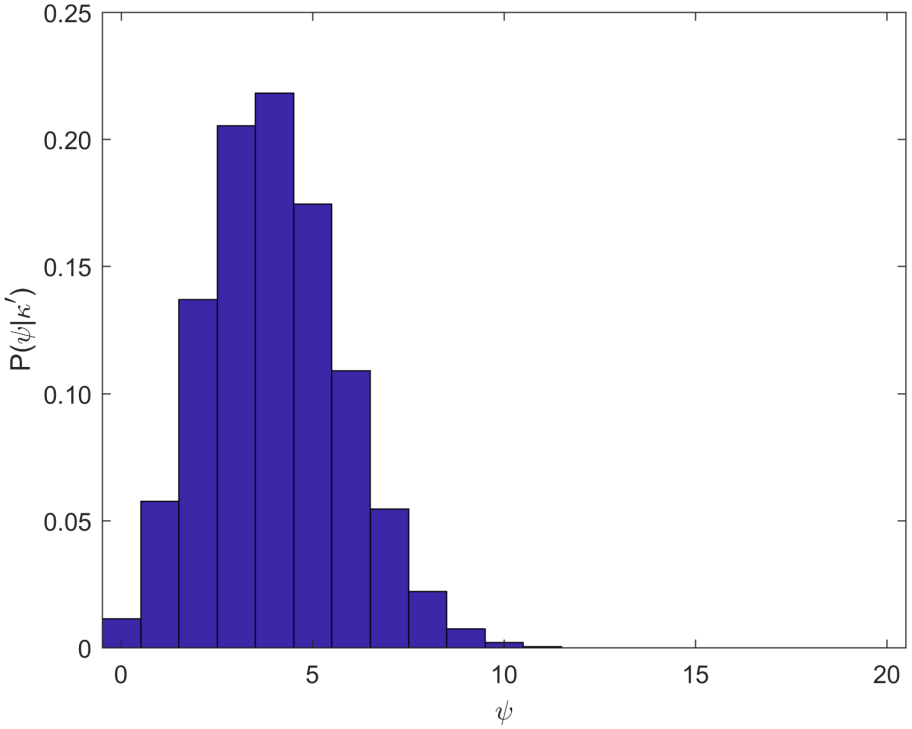

Figure 4.3 – The number of neighbours nodes i and n have in common. Given node n has $K' = 20$ excess degree neighbours and the triangle clustering coefficient is $C_\triangle = 0.2$, the number of neighbours i and n have in common, ψ, is binomially distributed with a mean value of 4.

Given n has excess degree κ', the number of neighbours i and n have in common, ψ, is binomially distributed (Figure 4.3):

$$\begin{aligned} \mathrm{P}(\psi|\kappa') &= \binom{\kappa'}{\psi} (C_\triangle)^\psi (1 - C_\triangle)^{\kappa'-\psi} \\ &= f_\mathrm{B}(\psi; \kappa', C_\triangle), \end{aligned} \tag{4.12}$$

where $C_\triangle$ is the triangle clustering coefficient[8].

Similarly, the number of EDNs of n infected by i during the first time step, given i and n have ψ neighbours in common, is binomially distributed,

[8]The reader is reminded here that clustering, be it triangular or square clustering, was assumed to be degree-independent.

such that

$$\begin{aligned} \mathrm{P}(\theta|\psi,\kappa') &= \binom{\psi}{\theta} (T)^{\theta} (1-T)^{\psi-\theta} \\ &= f_{\mathrm{B}}(\theta;\psi,T), \end{aligned} \tag{4.13}$$

where T is the transmission probability given contact as before. Equations (4.12) and (4.13) together capture the first effect of clustering on R_0 described in Section 4.4[9].

With the index case i having infected θ of node n's κ' EDNs, n has at most $(\kappa'-\theta)$ susceptible EDNs of type (α,β) at $t=1$. The probability n has $\omega_{\alpha,\beta}$ such neighbours is

$$\begin{aligned} \mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa') &= \binom{\kappa'-\theta}{\omega_{\alpha,\beta}} (a_{\alpha,\beta})^{\omega_{\alpha,\beta}} (1-a_{\alpha,\beta})^{\kappa'-\theta-\omega_{\alpha,\beta}} \\ &= f_{\mathrm{B}}(\omega_{\alpha,\beta};\kappa'-\theta,a_{\alpha,\beta}), \end{aligned} \tag{4.14}$$

where $a_{\alpha,\beta}$ is the probability any single susceptible EDN of n at $t=1$ is of type (α,β). The probability $a_{\alpha,\beta}$ is challenging to derive as it depends on both θ and the number of infectious neighbours of i, besides n and any neighbours i and n have in common (which in turn depends on the degree of i, the value of ψ, and the square clustering coefficient $C_{\square}$). For this reason the derivation of $a_{\alpha,\beta}$ is deferred to Appendix C.

The last probability in Equation (4.10) that needs to be defined is $\mathrm{P}(\zeta_{\alpha,\beta}|\omega_{\alpha,\beta},\theta,\psi,\kappa')$. Given n has $\omega_{\alpha,\beta}$ susceptible EDNs of type (α,β) at $t=1$, the probability n infects $\zeta_{\alpha,\beta}$ of these neighbours during the second time step is

$$\begin{aligned} \mathrm{P}(\zeta_{\alpha,\beta}|\omega_{\alpha,\beta},\theta,\psi,\kappa') &= \binom{\omega_{\alpha,\beta}}{\zeta_{\alpha,\beta}} (r_{\alpha,\beta})^{\zeta_{\alpha,\beta}} (1-r_{\alpha,\beta})^{\omega_{\alpha,\beta}-\zeta_{\alpha,\beta}} \\ &= f_{\mathrm{B}}(\zeta_{\alpha,\beta};\omega_{\alpha,\beta},r_{\alpha,\beta}), \end{aligned} \tag{4.15}$$

where

$$r_{\alpha,\beta} = \frac{1}{\alpha+\beta+1}\left(1-(1-T)^{\alpha+\beta+1}\right) \tag{4.16}$$

is the probability n infects any single susceptible EDN of type (α,β) given it competes against $(\alpha+\beta)$ other infectious nodes attempting to do the same[10].

[9] Equations (4.12) and (4.13) are kept separate because the number of neighbours i and n have in common, ψ, is needed in the next step to capture the effect of square clustering.

[10] As before, $r_{\alpha,\beta}$ is the probability n infects a susceptible EDN of type (α,β) *before* any of the other $(\alpha+\beta)$ infectious nodes it competes against manage to.

4.5.3 The basic reproduction number

It is possible now to write down, at least in principle, a complete and analytic expression for R_0 that captures its dependence on $C_\triangle$ and $C_\square$. Substituting Equation (4.10) into Equation (4.9) gives

$$\begin{aligned}
&\mathrm{E}\left[\text{no. of type } (\alpha,\beta) \text{ EDNs infected by } n\right] \\
&= \sum_{\zeta_{\alpha,\beta}} \zeta_{\alpha,\beta} \sum_{\kappa',\psi,\theta,\omega_{\alpha,\beta}} \mathrm{P}(\zeta_{\alpha,\beta}|\omega_{\alpha,\beta},\theta,\psi,\kappa')\,\mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa')\,\mathrm{P}(\theta|\psi,\kappa')\,\mathrm{P}(\psi|\kappa')\,\mathrm{P}(\kappa') \\
&= \sum_{\kappa',\psi,\theta,\omega_{\alpha,\beta},\zeta_{\alpha,\beta}} \zeta_{\alpha,\beta}\,\mathrm{P}(\zeta_{\alpha,\beta}|\omega_{\alpha,\beta},\theta,\psi,\kappa')\,\mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa')\,\mathrm{P}(\theta|\psi,\kappa')\,\mathrm{P}(\psi|\kappa')\,\mathrm{P}(\kappa')\,.
\end{aligned} \tag{4.17}$$

Next, Equation (4.15) is substituted into Equation (4.17) and the fact that $\mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa')$, $\mathrm{P}(\theta|\psi,\kappa')$, $\mathrm{P}(\psi|\kappa')$, and $\mathrm{P}(\kappa')$ are all independent of $\zeta_{\alpha,\beta}$ is exploited to obtain

$$\begin{aligned}
&\mathrm{E}\left[\text{no. of type } (\alpha,\beta) \text{ EDNs infected by } n\right] \\
&= \sum_{\kappa',\psi,\theta,\omega_{\alpha,\beta},\zeta_{\alpha,\beta}} \zeta_{\alpha,\beta} f_{\mathrm{B}}\left(\zeta_{\alpha,\beta};\omega_{\alpha,\beta},r_{\alpha,\beta}\right)\mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa')\,\mathrm{P}(\theta|\psi,\kappa')\,\mathrm{P}(\psi|\kappa')\,\mathrm{P}(\kappa') \\
&= \sum_{\kappa',\psi,\theta,\omega_{\alpha,\beta}} r_{\alpha,\beta}\omega_{\alpha,\beta}\mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa')\,\mathrm{P}(\theta|\psi,\kappa')\,\mathrm{P}(\psi|\kappa')\,\mathrm{P}(\kappa')\,,
\end{aligned} \tag{4.18}$$

where the last line holds because

$$\sum_{\zeta_{\alpha,\beta}} \zeta_{\alpha,\beta} f_{\mathrm{B}}\left(\zeta_{\alpha,\beta};\omega_{\alpha,\beta},r_{\alpha,\beta}\right) = r_{\alpha,\beta}\omega_{\alpha,\beta}$$

is the mean of a binomial distribution. Repeating this process, by substituting Equation (4.14) into Equation (4.18), making use of the fact that $r_{\alpha,\beta}$, $\mathrm{P}(\theta|\psi,\kappa')$, $\mathrm{P}(\psi|\kappa')$, and $\mathrm{P}(\kappa')$ are all independent of $\omega_{\alpha,\beta}$, and substituting the mean value of a binomial distribution, the expression can be simplified further as follows

$$\begin{aligned}
&\mathrm{E}\left[\text{no. of type } (\alpha,\beta) \text{ EDNs infected by } n\right] \\
&= \sum_{\kappa',\psi,\theta,\omega_{\alpha,\beta}} r_{\alpha,\beta}\omega_{\alpha,\beta} f_{\mathrm{B}}\left(\omega_{\alpha,\beta};\kappa'-\theta,a_{\alpha,\beta}\right)\mathrm{P}(\theta|\psi,\kappa')\,\mathrm{P}(\psi|\kappa')\,\mathrm{P}(\kappa') \\
&= \sum_{\kappa',\psi,\theta} r_{\alpha,\beta} a_{\alpha,\beta}\left(\kappa'-\theta\right)\mathrm{P}(\theta|\psi,\kappa')\,\mathrm{P}(\psi|\kappa')\,\mathrm{P}(\kappa')\,.
\end{aligned} \tag{4.19}$$

Since $a_{\alpha,\beta}$ is a non-linear function of both θ and ψ (see Appendix C), Equation (4.19) cannot be simplified any further using the mean value of a

binomial distribution. Equations (4.11)–(4.13) are therefore all substituted into Equation (4.19) to obtain

$$\begin{aligned}&\mathrm{E}\left[\text{no. of type } (\alpha,\beta) \text{ EDNs infected by } n\right]\\&= \sum_{\kappa'\geq\psi\geq\theta\geq 0}^{M-1} r_{\alpha,\beta}a_{\alpha,\beta}\left(\kappa'-\theta\right) f_{\mathrm{B}}\left(\theta;\psi,T\right) f_{\mathrm{B}}\left(\psi;\kappa',C_{\triangle}\right)\left(\frac{\left(\kappa'+1\right)p_{\kappa'+1}}{\langle\kappa\rangle}\right),\end{aligned} \tag{4.20}$$

where the upper and lower bounds for the summation indices, as well as the relationship between them as dictated by the two binomial distributions, have also been included. Substituting Equation (4.20) into Equation (4.8) then finally yields the number of descendants n is expected to infect:

$$\sum_{\alpha=0}^{M-2}\sum_{\beta=0}^{M-1}\sum_{\kappa'\geq\psi\geq\theta\geq 0}^{M-1} r_{\alpha,\beta}a_{\alpha,\beta}\left(\kappa'-\theta\right) f_{\mathrm{B}}\left(\theta;\psi,T\right) f_{\mathrm{B}}\left(\psi;\kappa',C_{\triangle}\right)\left(\frac{\left(\kappa'+1\right)p_{\kappa'+1}}{\langle\kappa\rangle}\right). \tag{4.21}$$

In conclusion, the number of individuals a typical infectious case is expected to infect in an otherwise susceptible population whose social network is clustered is

$$\begin{aligned}R_0 = &\sum_{\kappa=1}^{M}\sum_{\chi=0}^{\kappa-1}\frac{1}{\chi+1}\left(1-\left(1-T\right)^{\chi+1}\right) f_{\mathrm{B}}\left(\chi;\kappa-1,T\right)\left(\frac{p_{\kappa}}{1-p_0}\right)\\&+\sum_{\alpha=0}^{M-2}\sum_{\beta=0}^{M-1}\sum_{\kappa'\geq\psi\geq\theta\geq 0}^{M-1} r_{\alpha,\beta}a_{\alpha,\beta}\left(\kappa'-\theta\right) f_{\mathrm{B}}\left(\theta;\psi,T\right) f_{\mathrm{B}}\left(\psi;\kappa',C_{\triangle}\right)\left(\frac{\left(\kappa'+1\right)p_{\kappa'+1}}{\langle\kappa\rangle}\right),\end{aligned} \tag{4.22}$$

where

$$r_{\alpha,\beta} = \frac{1}{\alpha+\beta+1}\left(1-\left(1-T\right)^{\alpha+\beta+1}\right) \tag{4.23}$$

and

$$a_{\alpha,\beta} = f_{\mathrm{B}}\left(\alpha;\theta,C_{\triangle}\right)\sum_{\kappa=\psi+1}^{M} f_{\mathrm{B}}\left(\beta;\kappa-\left(\psi+1\right),C_{\square}T\right)\left(\frac{p_{\kappa}}{1-\sum_{\kappa=0}^{\psi}p_{\kappa}}\right). \tag{4.24}$$

4.5.4 Negligible clustering

In the absence of square clustering ($C_\square = 0$), Equations (4.22)–(4.24) reduce to

$$R_0 = \sum_{\kappa=1}^{M} \sum_{\chi=0}^{\kappa-1} \frac{1}{\chi+1} \left(1-(1-T)^{\chi+1}\right) f_{\mathrm{B}}\left(\chi; \kappa-1, T\right) \left(\frac{p_\kappa}{1-p_0}\right) + \sum_{\alpha=0}^{M-2} \sum_{\kappa'\geq\theta\geq 0}^{M-1} r_\alpha a_\alpha \left(\kappa'-\theta\right) f_{\mathrm{B}}\left(\theta; \kappa', TC_\triangle\right) \left(\frac{\left(\kappa'+1\right) p_{\kappa'+1}}{\langle\kappa\rangle}\right), \tag{4.25}$$

where

$$r_\alpha = \frac{1}{\alpha+1}\left(1-(1-T)^{\alpha+1}\right) \tag{4.26}$$

and

$$a_\alpha = f_{\mathrm{B}}\left(\alpha; \theta, C_\triangle\right). \tag{4.27}$$

The summation over β and the subscript β from both $r_{\alpha,\beta}$ and $a_{\alpha,\beta}$ were dropped because in the absence of square clustering it is always true that $\beta = 0$ (see Equation (4.24)). In addition, the summation over ψ was dropped by merging the binomial distributions for ψ and θ. This is possible because a_α does not depend on ψ like $a_{\alpha,\beta}$ does when square clustering is non-negligible (c.f. Equations (4.24) and (4.27)). Thus the expression for R_0 is substantially simpler when square clustering is negligible.

It is also interesting to consider the formula for R_0 when the triangular form of clustering is negligible as well. When $C_\triangle = 0$ the index case i does not share any neighbours with node n, and so θ, which is effectively the number of EDNs of n infected by i, must be zero. The fact that $\theta = 0$, however, also means that $\alpha = 0$ (see Equation (4.27)). From this it follows that both the summation over α and the subscript α from r_α can be dropped, as was done previously for β. Furthermore, $\alpha = 0$ also implies that $r = T$ (see Equation (4.26)). Thus, in the absence of both triangle and square clustering the formula for R_0 simplifies to

$$R_0 = \sum_{\kappa=1}^{M} \sum_{\chi=0}^{\kappa-1} \frac{1}{\chi+1} \left(1-(1-T)^{\chi+1}\right) f_{\mathrm{B}}\left(\chi; \kappa-1, T\right) \left(\frac{p_\kappa}{1-p_0}\right) + \sum_{\kappa'=0}^{M-1} T\kappa' \left(\frac{\left(\kappa'+1\right) p_{\kappa'+1}}{\langle\kappa\rangle}\right), \tag{4.28}$$

which, with some simple algebraic manipulation of the second term, can be

expressed as

$$R_0 = \sum_{\kappa=1}^{M} \sum_{\chi=0}^{\kappa-1} \frac{1}{\chi+1} \left(1-(1-T)^{\chi+1}\right) f_{\mathrm{B}}\left(\chi;\kappa-1,T\right) \left(\frac{p_\kappa}{1-p_0}\right) + T\left(\frac{\langle\kappa^2\rangle - \langle\kappa\rangle}{\langle\kappa\rangle}\right), \tag{4.29}$$

where $\langle\kappa\rangle = \sum (\kappa'+1)\, p_{\kappa'+1}$ and $\langle\kappa^2\rangle = \sum (\kappa'+1)^2\, p_{\kappa'+1}$ are the first and second moments of the degree distribution of a uniformly randomly selected node respectively. Importantly, the second term in Equation (4.29) is the well known expression for the mean number of EDNs infected by a typical infectious case in a network that is unclustered [117, 126, 131].

4.6 Estimating R_0

4.6.1 Simulations

To verify the impact of clustering on R_0 as predicted by Equations (4.22)–(4.24), the spread of infection on clustered social networks was simulated. To do this, ER networks, with network size N and edge probability p, were generated as theoretical proxies for real social networks. The motivation for using ER networks was twofold. First, the degrees of nodes in ER networks are relatively homogeneous, following a Poisson distribution in the limit of large network size [131], such that degree assortativity is minimal. Second, because the probability an edge occurs between two nodes, p, is the same for all pairs of nodes in a network, the triangle and square clustering co-efficients are both equal to p and moreover are both degree-independent. Thus, the first two assumptions made when deriving the analytic formula for R_0 (see Section 4.5) are both satisfied by ER networks.

Simulation of infection transmission on each network began by selecting an index case, i, uniformly at random from amongst the nodes with non-zero degree and infecting it. Next, a typical infectious case, n, was chosen by uniformly randomly selecting one neighbour of the index case and infecting it as well[11]. Thereafter the index case was allowed to infect each of its remaining neighbours with probability T, after which it recovered to become susceptible again. Each node infected by i during the first time step was allowed to infect its susceptible neighbours (including the index case) with

[11] This approach is only acceptable in the limit of large network size, when N is on the order of several hundred nodes. If a network is smaller than this then the probability n has excess degree κ' would not be given by Equation (4.11).

probability T during the second time step. The number of nodes infected by n was recorded, including whether it reinfected the index case i. An estimate for R_0 was obtained by adding the proportion of times n reinfected i (over all transmission simulations for a particular combination of values for p and T—see below) to the average number of other nodes n also managed to infect (as is consistent with Equation (4.1)).

When generating ER networks, the number of nodes was set to a fixed value of $N = 300$. Two values for p were considered, namely 0.05 and 0.35. A total of 50 networks were generated for each value of p. To investigate the dependence of R_0 on the transmission probability, T was varied from 0 to 1 with a step size of 0.05. A total of 100 transmission simulations were conducted on each ER network at each value of T. Thus, for each combination of p and T the results of 5000 transmission simulations were used to calculate the corresponding estimate of R_0.

4.6.2 Analytic predictions

The analytic formula for R_0 in Equations (4.22)–(4.24) is a function of the degree distribution of a network, p_k. Thus, for a given value of T, the analytic estimate of R_0 varies between networks, even ER networks generated for the same value of p. To obtain an aggregate analytic estimate for R_0 against which the simulation results could be compared, the average value of the network-specific estimates was calculated, where the average was taken over all networks generated for the same value of p.

4.7 Visualizing R_0

In Figure 4.4 the basic reproduction number R_0 is plotted against the transmission probability T for two values of the ER network edge probability: (Panel A) $p = 0.05$ and (Panel B) $p = 0.35$. The solid black curve in Panel A denotes the prediction of the analytic formula when both $C_\triangle$ and $C_\square$ are equal to p. A similar curve was not generated for Panel B as the high maximum degree M of the corresponding ER networks made the formula for R_0 in Equations (4.22)–(4.24) computationally prohibitive[12]. The dash-dot red curves in both panels represent the analytic predictions when $C_\triangle = p$ and

[12]The network-specific analytic estimate of R_0 for just one of the highly clustered ($p = 0.35$) ER networks and for only a single value of T took 8 CPU's, running in parallel, and with access to 8GB of RAM, over a day to compute. Therefore, to have generated a solid black curve in Panel B corresponding to $C_\triangle = C_\square = 0.35$ would have taken nearly 3 years with today's standard desktop PC.

$C_\square = 0$ as in Equations (4.25)–(4.27), whilst the dashed green curves correspond to when the two clustering co-efficients are both set equal to zero as in Equation (4.29). The scatter plots illustrate the simulation estimates for R_0.

The first result worth noting is that there is good agreement between the analytic and simulation estimates of R_0 for clustered ER networks (compare the solid black curve and scatter plot in Panel A) which supports the validity of the derived analytic formula. Secondly, from both Panels A and B it is clear that as T increases, so too does the magnitude of the reduction in the value of R_0 due to clustering (compare the dashed green curve and scatter plot in each panel). Thirdly, the effect of triangle clustering in reducing the value of R_0 relative to the effect of square clustering varies with p (compare the dashed green curve, dash-dot red curve, and scatter plot in each panel); triangle clustering accounts for a greater proportion of the reduction in the value of R_0 when p is high, which is to say when the mean degree of a network is high. A fourth, unexpected, result is that for the ER networks generated with $p = 0.35$ the value of R_0 has a maximum at approximately $T = 0.20$ (see scatter plot in Panel B). This suggests that in highly clustered networks with large mean degree, less infectious diseases will spread more readily during the early stages of an epidemic than infections that are highly contagious. The fact that such a maximum really does occur is confirmed by the analytic predictions for R_0 when $C_\triangle = 0.35$ and $C_\square = 0$ (see dash-dot red curve in Panel B).

4.8 Should R_0 be estimated from the first generation of infectious individuals?

The formula for R_0 in Equations (4.22)–(4.24) was derived from the mean number of neighbours infected by a node in the first generation of infection. However, for the SIS disease framework, R_0 is traditionally derived from a node in the second generation of infection [118]. This is because an infectious individual in the second, and any subsequent, generation has an ancestor that was infected along an edge, whereas the ancestor of a node in the first generation was infected uniformly at random. That is to say, every infectious individual from the second generation onwards is equivalent in terms of its opportunity to infect other nodes (i.e. its transmission potential).

Whilst this is true for unclustered networks, it does not hold for clustered networks. This is because every generation of infection is different in terms of transmission potential when loops occur. To see this, consider that

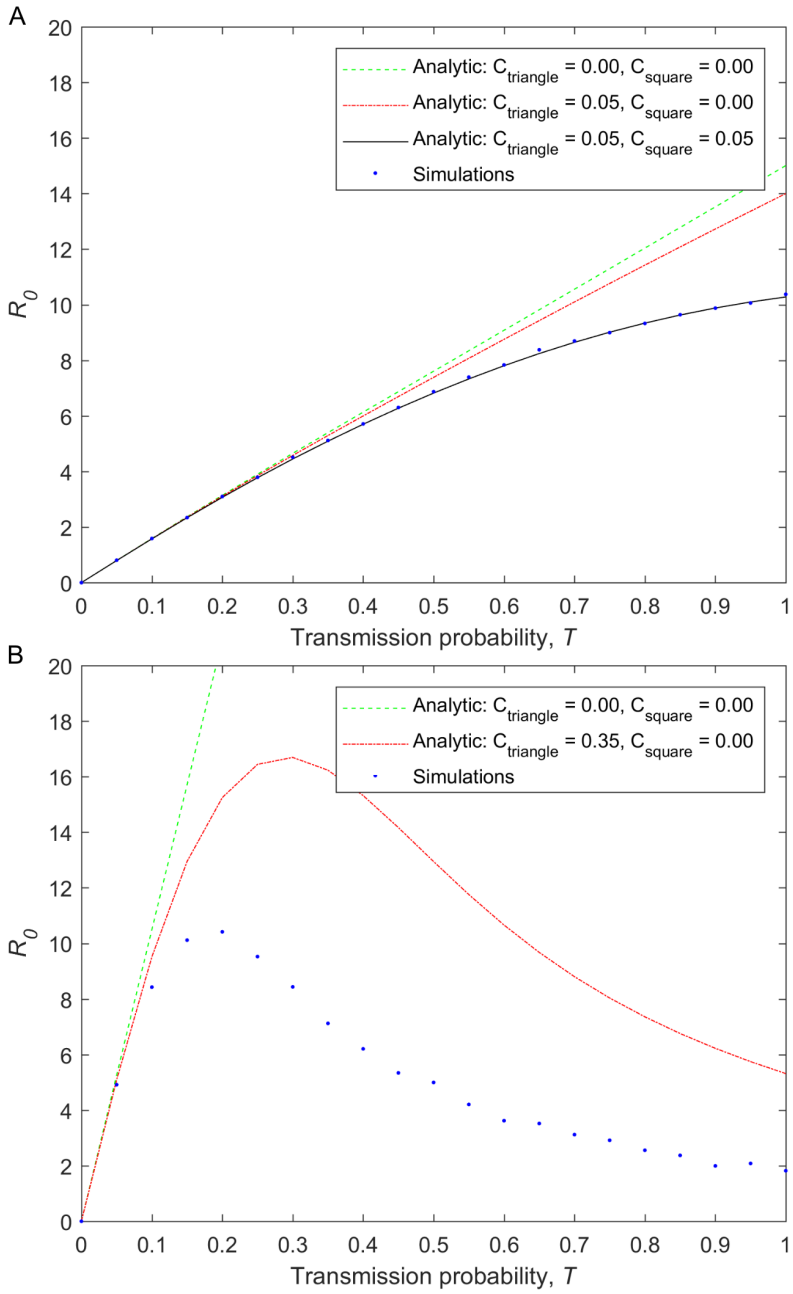

Figure 4.4 – The effect of clustering on R_0 as a function of the transmission probability T. The basic reproduction number is shown for ER networks of size $N = 300$ and edge probability (A) $p = 0.05$ and (B) $p = 0.35$. The solid black curve denotes the analytic prediction when both $C_{\triangle}$ and $C_{\square}$ are equal to p (Panel A only), the dash-dot red curves represent the analytic predictions when $C_{\triangle} = p$ and $C_{\square} = 0$ (Panels A and B), and the dashed green curves correspond to when the two clustering co-efficients are both set equal to zero (Panels A and B). The scatter plots illustrate the simulation results.

when deriving the mean number of neighbours infected by a node in the first generation, triangle and square clustering had to be accounted for. If instead the mean number of neighbours infected by an individual in the second generation was derived, pentagons would also have needed to be taken into account. Furthermore, for later generations even higher-order loops would have needed to be considered. Thus, for clustered networks the mean number of neighbours infected by an infectious individual in each successive generation decreases monotonically. Importantly, because a node in the first generation is more likely to reinfect its ancestor than a node in any subsequent generation is to infect its own, this monotonic decline in the number of nodes infected by an infectious individual begins from the first generation of infection.

This insight is useful. It says that if the mean number of neighbours infected by a node from a particular generation of infection is below unity then so too is the mean number of neighbours infected by an infectious individual in any subsequent generation of infection. Therefore, to ascertain whether an epidemic is possible, all that is required is an expression for the mean number of neighbours infected by a node in one of the early generations of infection. This, together with the fact that only triangles and squares needed to be taken into account, was the reason that an expression for R_0 was derived from the mean number of neighbours infected by a node in the first generation. An additional benefit of having taken this approach is that the expression for R_0 is more analytically tractable and computationally efficient than it would have been had the mean number of neighbours infected by a node in the second generation of infection been derived instead.

4.9 Discussion

In this chapter the basic reproduction number, R_0, was derived for the spread of an infectious disease through a clustered social contact network. Specifically, a brute-force probabilistic approach was used to calculate the expected number of secondary infections caused by a typical infectious individual for the SIS disease framework. The formula for R_0 incorporates effects resulting from both triangular and square clustering, as well as any additional competing interests due to multiple infectious nodes attempting to infect the same susceptible neighbour nodes. Simulations on clustered ER networks were performed to validate the analytic formula derived for R_0. Clustering in social networks was found to have a protective effect, lowering the value of R_0 compared to a network in which there is minimal clustering, thereby making it harder for an infection to invade the host population. Further-

more, in extreme circumstances (when $C_{\triangle}$ and $C_{\square} \geq 0.35$) the value of R_0 may even begin to decline at high transmission probabilities.

The analytic formula derived for R_0 is exact with all forms of clustering relevant to the first generation of infection taken into account. Importantly, this allows definitive conclusions to be drawn about the impact of clustering on infection spread during the early stages of an epidemic. The results presented here confirm those of recent studies which find that R_0 decreases with increased clustering and that the margin by which R_0 decreases depends on the transmission probability T [117, 126]. In contrast to the findings of Molina and Stone [117], though, it was found that square clustering can play a substantial role in reducing R_0. In fact, for the ER networks generated with an edge probability of $p = 0.05$ the effect of square clustering was even greater than that of triangle clustering (Figure 4.4, Panel A) which calls into question the assumption that squares can be ignored.

Perhaps the most surprising result was that, for networks with high levels of clustering and large mean degree, R_0 may be lower for highly infectious diseases than less transmissible ones (Figure 4.4, Panel B). This finding may be relevant for settings such as schools where the social contact networks of children and adolescents are highly interconnected and clustered [7]. To understand how this counter-intuitive result comes about, one must ask the question, "Why, when $T < 0.20$, does R_0 increase as T increases, whereas, when $T > 0.20$, R_0 decreases as T increases?" In both instances as the value of T goes up the index case infects more of its neighbours, reducing the number of susceptible excess degree neighbours available for a typical infectious case to subsequently infect. So, this cannot be the sole reason. The key to understanding this result lies in recognizing that the increased number of nodes infected by the index case means that the values of α and β (i.e. the number of infectious nodes a typical infectious individual must compete against due to triangle and square clustering respectively) also go up. Thus, the probability a typical infectious individual infects any one of its susceptible excess degree neighbours ($r_{\alpha,\beta}$ in Equation 4.23) does not increase monotonically with T, and in fact reaches a maximum value at some critical value $T = T_c \approx 0.20$. Once the value of T goes above T_c, the probability $r_{\alpha,\beta}$ starts to decline, causing R_0 to decline along with it.

There are three caveats for the results presented in this chapter. First, the analytic formula for R_0 is only strictly applicable to large networks. This is because the excess degree distribution in Equation (4.11), which describes the probability a typical infectious individual has excess degree κ', is only precise in the limit of large network size [116]. Second, when implementing the model simulations, a typical infectious case was chosen by uniformly randomly selecting from amongst the neighbours of a uniformly randomly

selected index case. As for the first caveat, this approach is only acceptable in the limit of large network size since only then is the degree distribution of individuals selected in this manner equal to the excess degree distribution. An alternative approach, previously implemented by Miller [126], and which may avoid this limitation, is to take the ratio of the number of infectious individuals in the second generation to the number in the first generation instead of selecting just a single typical infectious case per simulation from which to estimate R_0. The third, and final, caveat is that when calculating R_0 from the analytic formula it was assumed the triangle and square clustering co-efficients, $C_{\triangle}$ and $C_{\square}$ respectively, were equal to the probability an edge occurs between any pair of nodes, p. Any difference between the analytic and simulation-based estimates of R_0 may therefore be due to stochastic effects resulting in $C_{\triangle}$ and $C_{\square}$ differing from p for some of the ER networks generated. The overall effect of these differences, however, should vanish as the total number of networks generated is increased.

Chapter 5

Discussion

The focus of this thesis was to investigate how the level of contact detail included in an infectious disease model influences its predictions regarding disease emergence and dynamics. To do this, two mechanistic models and one theoretical model were studied. The first mechanistic model described the potential spread of canine rabies through the Australian wild dog population as an illustrative example of an infection spreading through a territorial population. The second described the spread of tick-borne pathogens between ticks and their vertebrate hosts. In the third model, the influence of clustering on the spread of infection through social contact networks was investigated. The analysis of each model ranged from a straightforward investigation of the relationships between model input and outcome variables to the derivation of analytic formulas for the basic reproduction number, R_0, as a function of model parameters. In this chapter, the implication of each model's results is discussed in terms of the overarching research question.

5.1 Diseases in territorial host populations: canine rabies in Australian wild dogs as a case study

In a wild dog population it is dog behaviour that determines where, when, and how often any pair of dogs frequent the same location at the same time, and therefore where, when, and how often rabies transmission might occur if one of them were to be rabid. Because wild dogs are territorial by nature, and also because they occupy an expansive landscape, a large geographical distance between the territories of two dogs means that they are unlikely to come into contact. Contacts between dogs therefore occur predominantly

between those that occupy neighbouring or nearby territories. This implies that the possible spread of canine rabies in Australia's currently naive wild dog population would be a spatial phenomenon. Wild dogs also vary a great deal in their social behaviour: some dogs traverse large areas of land per unit time and therefore frequently come into contact with neighbours, whilst others roam less and tend to shy away from interactions with other dogs. To incorporate both spatial constraints and heterogeneous social behaviour into the rabies model a contact rate for any pair of dogs was proposed that is a function of the distance between the two dogs as well as their individual inclination to be social.

Plotting the rabies model's outcomes against the values of its input variables facilitated investigation into the relationship between them (Figures 2.5–2.12). The more territorial wild dogs are in their behaviour (i.e. the greater the value of the spatial scale parameter λ) the less likely rabies is to spread (Figure 2.5, Panel E). Consequently, to ignore the effect of geographic distance on contact, for example by assuming random mixing, would result in a model that overestimates the probability rabies will spread through the wild dog population. From Panel D of Figure 2.5 the sociability scale parameter was observed to have the opposite effect: the greater the level of heterogeneity (one of the side effects of increasing the sociability scale parameter) the greater the probability rabies will spread. Thus to ignore heterogeneity in wild dog movements and behaviour would result in a model that underestimates the probability rabies will spread. Both of these findings are consistent with that which has been reported in the literature for infections spreading through populations occupying large landscapes and through populations comprising individuals with heterogeneous social behaviour [23, 67].

Incorporating greater realism into the rabies model, by taking spatial constraints into account, also meant that the speed at which rabies is expected to spread was a model outcome. This would not have been the case had the effects of distance simply been ignored, e.g. by assuming random mixing. Outcomes such as the rate of spread are useful to public health agencies in terms of improving preparedness for a rabies incursion and determining the feasibility of containment and other management strategies. From Panel E of Figure 2.6 it is clear that the more territorial wild dogs are (i.e. the larger the value of λ) the slower the rate at which rabies will spread geographically (since the time to each milestone is longer). Consequently, a model that ignores the effect of spatial constraints on wild dog contacts would overestimate the rate of geographic spread.

In conclusion, for canine rabies spreading through the wild dog population of Australia, considering geographic distance and heterogeneous wild dog movements and behaviour is critical for predicting the likelihood of a

rabies epidemic and how quickly the front of such an epidemic would spread southwards from the north coast of Australia. More generally, for infections that spread through territorial populations, the rabies model has generated deeper insight: parameters that quantify the movements of individuals across the landscape are key to determining disease dynamics. The rabies model quantified (complex) dog behaviour as variation in the area of land traversed per unit time combined with a spatial scale parameter that determined how free the movements were (whether the land being traversed was faithful to the individual's own territory). The territorial nature of other species, such as prides of lions in the Serengeti or badgers in the United Kingdom, might then be described using closely related parameters (i.e. with similar units) to the spatial scale parameter and wild dog sociability (area of land traversed per unit time) as an approach to capturing the contact patterns that drive disease transmission.

5.2 The spread of tick-borne pathogens between ticks and vertebrate hosts

The type of contact required for the transmission of the most important zoonotic tick-borne pathogens is a tick attaching itself to a vertebrate host and taking a blood meal. Key features of tick feeding behaviour are the aggregation of ticks on hosts, the co-aggregation of larval and nymphal ticks on the same hosts, and for some tick populations coincident co-aggregation, also known as co-feeding. Co-feeding facilitates the transmission of pathogens that tend not to cause systemic infection in the vertebrate host, but instead rely on direct transmission from nymphs to larvae feeding on the same vertebrate hosts, at the same time, and in close proximity. Each of these three contact features were incorporated into a network model that was used to evaluate the influence of tick co-aggregation on whether or not a tick-borne pathogen could become established.

The principal result of the proposed tick-borne pathogen transmission network model was an analytic formula for the basic reproduction number, R_0, that captures the effect of all three features of tick feeding behaviour (see Equation (3.18)). The equation predicts the effect of larval and nymphal co-aggregation is to always increase R_0, thereby making disease establishment more likely. Simulations of Lyme disease transmission on model tick-host contact networks confirmed this and also revealed that aggregation only raises the value of R_0 when ticks also co-aggregate. The implications of these findings are twofold: (1) for tick populations where larvae and nymphs do not

co-aggregate on hosts, a model need not incorporate tick aggregation either—since the predictions of the model will be sufficiently accurate—and (2) for tick populations that do show evidence of larval and nymphal co-aggregation, any model that fails to incorporate both aggregation and co-aggregation will underestimate the likelihood of tick-borne pathogen emergence.

Two additional, unexpected findings of the network model were that the absolute effect of co-aggregation on R_0 is greater when the mean annual larval burden is higher and that the relative effect of co-aggregation on R_0 is greater for tick-borne pathogens transmitted predominantly via co-feeding transmission compared to those transmitted systemically. The implication of these findings is that if a model does not include co-aggregation (which is usually the case), its predictions will underestimate R_0 more severely for geographic regions and seasons in which larval burden is high, especially for tick-borne pathogens transmitted predominantly between co-feeding ticks. Thus, in a broad sense, the results of the network model reveal that any previous exclusion of co-aggregation from tick-borne pathogen models could account for differences between the geographic regions they predicted disease emergence is possible and the regions where pathogens actually occur.

5.3 Clustered social networks

The relational nature of humans (and animals) means that individuals are much more likely to make contact with friends of friends (or close relations) than they are complete strangers. This behaviour is evidenced by the frequent occurrence of geometric structures such as triangles and squares in social networks, commonly referred to as clustering in the complex network literature. Clustering has the potential to influence the number of susceptible individuals an infectious individual comes into contact with and can therefore pass the infection on to.

To investigate the effect of clustering on disease emergence, an analytic formula for R_0 was derived for an infection spreading through a clustered social network. Clustering was found to reduce the value of R_0, and the magnitude of the reduction was shown to increase with higher transmission probabilities. Together these two results imply that a model that fails to incorporate clustering will overestimate the likelihood of disease establishment, and the severity of this error will be worse for infections that are highly transmissible. A third result was that for networks with low mean degree squares can reduce the value of R_0 more than triangles. Consequently, even if disease models incorporate triangle clustering they may still overestimate the likelihood of disease emergence if they neglect the effect of squares.

Lastly, for networks with high levels of clustering and high mean degree, it was also found that R_0 has a non-monotonic relationship with the transmission probability T, reaching a maximum at some intermediate value and declining thereafter. This non-intuitive finding comes about because high levels of clustering means that increasing numbers of infectious nodes compete to infect the same number of susceptible nodes as T is increased. That is to say, the non-monotonic relationship is effectively a local saturation effect that is typically associated with spatial networks (such as the wild dog contact networks of Chapter 2). The clustering results, however, indicate that competition and local saturation can occur in a non-spatial setting and as early as the second generation of infected individuals. This means the initial dynamics of an infectious disease may not be well characterized by R_0 for highly clustered populations (since a higher R_0 in these populations does not imply more transmission events since the start of an epidemic).

5.4 Conclusion

In conclusion, the focus of this thesis was to investigate how the level of contact detail included in an infectious disease model influences its predictions regarding disease emergence and dynamics. The three disease models studied collectively reveal that model predictions are improved and that additional outcomes are generated by the inclusion of realistic host contact patterns. These findings reinforce the value of incorporating biologically faithful contact patterns into infectious disease models.

Appendix A

Global sensitivity analysis: technical details

Introduction

The method used to quantify the sensitivity of the rabies model's outcomes to each of its input variable's uncertainty (Chapter 2) was to calculate Sobol's indices by implementing the Monte Carlo procedure proposed by Saltelli [75]. In this appendix the key steps involved are described after briefly reviewing the theory behind Sobol's indices and in particular the formulas estimated during the Monte Carlo procedure. The motivation behind including these technical details in this thesis is twofold. First, there are different approaches for performing a global sensitivity analysis. By declaring the method implemented in Chapter 2, it ensures the results reported in that chapter are reproducible. The second reason is that of clarity. Although a growing body of global sensitivity analysis literature exists, much of it is written without the non-specialist in mind. This acts as a barrier to the widespread application of this most useful theory. In this appendix the most accessible, and concise, description to date of how to implement the methodology proposed by Saltelli [75] has been written. This description was included as supplementary material in the research article published in *PLoS Neglected Tropical Diseases* [46].

A.1 Global sensitivity analysis theory

Consider a generic model $z = f(y_1, y_2, \ldots, y_k)$ with k independent random input variables that have joint probability density function $p\,(y_1, y_2, \ldots, y_k) = \prod_{i=1}^{k} p_i(y_i)$. Also, consider $\boldsymbol{v}$, a subset of the input variables, and $\boldsymbol{u}$ its complement such that $\{\boldsymbol{v}, \boldsymbol{u}\} = \{y_1, y_2, \ldots, y_k\}$. The sensitivity of outcome variable z with respect to subset $\boldsymbol{v}$ is well described by two sensitivity indices, namely the first-order effect and total effect, $S_{\boldsymbol{v}}$ and $S_{\boldsymbol{v}}^T$ respectively. Of particular interest to us is when $\boldsymbol{v}$ includes only a single input variable, e.g. $\boldsymbol{v} = \{y_j\}$. In this case $S_{\boldsymbol{v}}$ is said to be the first-order effect of input variable y_j and write

$$S_{\boldsymbol{v}} = S_j = \left(\frac{U_j - E^2(z)}{V(z)} \right), \tag{A.1}$$

where

$$E^2(z) = \left\{ \int \cdots \int f(y_1, y_2, \ldots, y_k) \prod_{i=1}^{k} p_i(y_i)\, dy_i \right\}^2 \tag{A.2}$$

is the square of the mean value of z,

$$V(z) = \int \cdots \int f^2(y_1, y_2, \ldots, y_k) \prod_{i=1}^{k} p_i(y_i)\, dy_i - E^2(z) \tag{A.3}$$

is the variance of z, and

$$U_j = \int \cdots \int f(y_1, y_2, \ldots, y_j, \ldots, y_k) f(y_1', y_2', \ldots, y_j, \ldots, y_k') \prod_{i=1}^{k} p_i(y_i)\, dy_i \prod_{\substack{i=1 \\ i \neq j}}^{k} p_i(y_i')\, dy_i' \tag{A.4}$$

is the mean value of a function F of $(2k-1)$ variables defined as

$$\begin{aligned} F(y_1, y_2, \ldots, y_j, \ldots, y_k, y_1', y_2', \ldots, y_{j-1}', y_{j+1}', \ldots, y_k') \\ = f(y_1, y_2, \ldots, y_j, \ldots, y_k) f(y_1', y_2', \ldots, y_j, \ldots, y_k'). \end{aligned} \tag{A.5}$$

Important to note is that in Equations (A.4) and (A.5):

1. The primed random input variables have the same distributions as their non-primed counterparts, i.e. y_i and y_i' have the same distribution p_i, but their values are independent, and

2. There is no random variable y_j' (or equivalently one can think of y_j' as always taking the same value as random variable y_j).

Similar to the first-order effect, when $\boldsymbol{v} = \{y_j\}$, $S_{\boldsymbol{v}}^T$ is said to be the total effect of input variable y_j and write

$$S_{\boldsymbol{v}}^T = S_j^T = (1 - S_{\neg j}) = (1 - S_{\boldsymbol{u}}), \tag{A.6}$$

where

$$S_{\neg j} = S_{\boldsymbol{u}} = \left(\frac{U_{\boldsymbol{u}} - E^2(z)}{V(z)} \right) \tag{A.7}$$

is the first-order effect for the set of all input variables excluding y_j, namely subset $\boldsymbol{u}$, and

$$U_{\boldsymbol{u}} = \int \cdots \int f(y_1, y_2, \ldots, y_j, \ldots, y_k) f(y_1, y_2, \ldots, y'_j, \ldots, y_k) \prod_{i=1}^{k} p_i(y_i)\, dy_i\, p_j(y'_j)\, dy'_j \tag{A.8}$$

is the mean value of a function G of $(k+1)$ input variables defined as

$$G(y_1, y_2, \ldots, y_j, \ldots, y_k, y'_j) = f(y_1, y_2, \ldots, y_j, \ldots, y_k) f(y_1, y_2, \ldots, y'_j, \ldots, y_k). \tag{A.9}$$

A.2 Monte Carlo procedure

From Equations (A.1)–(A.9) it follows that to calculate the first-order and total effects of an input variable y_j on an outcome variable z one must evaluate four integrals (Equations (A.2)–(A.4) and (A.8)) that are the mean values of closely related functions, e.g. F and G. In a Monte Carlo framework this is easily done by implementing the following procedure proposed by Saltelli [75]. First, distributions for each of the k input variables are defined. Next, two sample matrices

$$\mathbf{M} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{pmatrix} \text{ and } \mathbf{M'} = \begin{pmatrix} y'_{11} & y'_{12} & \cdots & y'_{1k} \\ y'_{21} & y'_{22} & \cdots & y'_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ y'_{n1} & y'_{n2} & \cdots & y'_{nk} \end{pmatrix}$$

are generated by sampling each input variable distribution a total of $2n$ times (n times for each matrix). Thereafter, both $E^2(z)$ and $V(z)$ are calculated from the products of the values of z computed on the rows (sample vectors) of $\mathbf{M}$ and $\mathbf{M'}$ as follows

$$\hat{E}^2(z) = \frac{1}{n} \sum_{r=1}^{n} f(y_{r1}, y_{r2}, \ldots, y_{rk}) f(y'_{r1}, y'_{r2}, \ldots, y'_{rk}) \tag{A.10}$$

$$\hat{V}(z) = \frac{1}{n}\sum_{r=1}^{n} f(y_{r1}, y_{r2}, \ldots, y_{rk}) f(y_{r1}, y_{r2}, \ldots, y_{rk}) - \hat{E}^2(z), \tag{A.11}$$

where n is the sample size of the Monte Carlo estimates. Following this, Monte Carlo estimates for U_j and $U_{\boldsymbol{u}}$ are obtained from

$$\hat{U}_j = \frac{1}{n}\sum_{r=1}^{n} f(y_{r1}, y_{r2}, \ldots, y_{rk}) f(y'_{r1}, y'_{r2}, \ldots, y'_{r(j-1)}, y_{rj}, y'_{r(j+1)}, \ldots, y'_{rk}) \tag{A.12}$$

$$\hat{U}_{\boldsymbol{u}} = \frac{1}{n}\sum_{r=1}^{n} f(y_{r1}, y_{r2}, \ldots, y_{rk}) f(y_{r1}, y_{r2}, \ldots, y_{r(j-1)}, y'_{rj}, y_{r(j+1)}, \ldots, y_{rk}), \tag{A.13}$$

where the inputs for the second factor in the products of the two equations are the rows (sample vectors) of two additional sample matrices,

$$\mathbf{N}_j = \begin{pmatrix} y'_{11} & y'_{12} & \cdots & y'_{1(j-1)} & y_{1j} & y'_{1(j+1)} & \cdots & y'_{1k} \\ y'_{21} & y'_{22} & \cdots & y'_{2(j-1)} & y_{2j} & y'_{2(j+1)} & \cdots & y'_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y'_{n1} & y'_{n2} & \cdots & y'_{n(j-1)} & y_{nj} & y'_{n(j+1)} & \cdots & y'_{nk} \end{pmatrix}$$

and

$$\mathbf{N}_{\boldsymbol{u}} = \mathbf{N}_{\neg j} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1(j-1)} & y'_{1j} & y_{1(j+1)} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2(j-1)} & y'_{2j} & y_{2(j+1)} & \cdots & y_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{n(j-1)} & y'_{nj} & y_{n(j+1)} & \cdots & y_{nk} \end{pmatrix}$$

respectively, constructed from combinations of the columns of $\mathbf{M}$ and $\mathbf{M'}$.

In closing worth noting is that, because matrices $\mathbf{N}_j$ and $\mathbf{N}_{\neg j}$ are specific to input variable y_j, for a model with k input variables a total of $2\,(k+1)$ sample matrices of size $(n \times k)$ need to be constructed to obtain a full set of first-order and total effect sensitivity indices. This, in turn, requires the implementation of $2n\,(k+1)$ simulations, one for each row of each matrix.

Appendix B

Relating tick-borne pathogen transmission probabilities to tick phenology

Introduction

To generate tick-borne pathogen transmission networks from tick-host contact networks the probability an edge appears in a transmission network needs to be related to the time interval between an infectious nymph and a susceptible larva taking their respective blood meals from the same vertebrate host. A method for doing so that makes use of mathematical descriptions of larval and nymphal tick phenology is presented here.

B.1 Relating tick-borne pathogen transmission to tick phenology

Recall from Equations (3.1) and (3.2) in Section 3.2.1 that the phenology of larval and nymphal ticks can be described mathematically, with $\bar{Z}_L(t)$ and $\bar{Z}_N(t)$ denoting the mean number of larvae and nymphs respectively that feed on a single host t days since the start of the year. If the average number of days larvae and nymphs remain attached to hosts whilst taking a blood meal are denoted d_L and d_N respectively, then the mean number of *unique* larvae that feed on a single host over the course of an entire year is given by

$$\langle k_{\text{out}} \rangle = \int_{t'=0}^{365} \frac{\bar{Z}_L(t')}{d_L}\, dt', \tag{B.1}$$

and the mean number of *unique* nymphs is

$$\langle k_{\text{in}} \rangle = \int_{t'=0}^{365} \frac{\bar{Z}_N(t')}{d_N}\, dt'. \tag{B.2}$$

Given that a vertebrate host is infected on day t, it follows from Equation (B.1) that the probability any larval tick that feeds on this host does so after day t is equal to

$$\frac{\int_{t'=t}^{365} \bar{Z}_L(t')\, dt'}{\int_{t'=0}^{365} \bar{Z}_L(t')\, dt'}. \tag{B.3}$$

The unconditional probability a larval tick feeds on a vertebrate host after the host has been infected by a nymph is obtained by integrating out the dependence of Equation (B.3) on t as follows:

$$\int_{t=0}^{365} a_N(t) \frac{\int_{t'=t}^{365} \bar{Z}_L(t')\, dt'}{\int_{t'=0}^{365} \bar{Z}_L(t')\, dt'}\, dt, \tag{B.4}$$

where the probability density function $a_N(t)$ is a measure of the risk that a vertebrate host will be infected on day t. Formally, this function is defined as

$$a_N(t) = \frac{\bar{Z}_N(t)}{\int_{t=0}^{365} \bar{Z}_N(t)\, dt}. \tag{B.5}$$

B.2 Vertebrate host-to-larva transmission probability, ν_{lh}

The host-to-larva transmission probability ν_{lh} is conditional on a larval tick having taken a blood meal from an infectious vertebrate host. A blood meal

alone, though, is not all that is required for transmission to be possible. The larval tick must also take its blood meal after systemic infection has developed in the host (i.e. after the latent period t_L has elapsed) and before the vertebrate host's infectious period t_I comes to an end (if the infection is not lifelong). Taking these considerations into account the host-to-larva transmission probability is given by

$$\nu_{\text{lh}} = \nu^*_{\text{lh}} \int_{t=0}^{365} a_N(t) \frac{\int_{t'=t+t_L}^{t+t_L+t_I} \bar{Z}_L(t')\, dt'}{\int_{t'=0}^{365} \bar{Z}_L(t')\, dt'}\, dt, \tag{B.6}$$

where ν^*_{lh} is the average probability a vertebrate host infects a larval tick given that it takes a blood meal during the vertebrate host's infectious period. Substituting Equations (B.1), (B.2), and (B.5) into Equation (B.6) means the host-to-larva transmission probability can also be written as follows

$$\nu_{\text{lh}} = \frac{\nu^*_{\text{lh}}}{d_L d_N \langle k_{\text{in}} \rangle \langle k_{\text{out}} \rangle} \int_{t=0}^{365} \bar{Z}_N(t) \int_{t'=t+t_L}^{t+t_L+t_I} \bar{Z}_L(t')\, dt' dt. \tag{B.7}$$

B.3 Nymph-to-larva transmission probability, ν_{ln}

Similar to systemic transmission, a larval tick taking a blood meal from the same vertebrate host as an infectious nymph is not all that is required for co-feeding transmission to occur. For nymph-to-larva transmission there are two additional conditions: the larval tick must feed at the same time as the infectious nymph and it must also feed in close proximity to the nymph. If co-feeding transmission is assumed to occur only whilst the infectious nymph is feeding, then the nymph-to-larva transmission probability is given by

$$\nu_{\text{ln}} = c\, \nu^*_{\text{ln}} \int_{t=0}^{365} a_N(t) \frac{\int_{t'=t}^{t+d_N} \bar{Z}_L(t')\, dt'}{\int_{t'=0}^{365} \bar{Z}_L(t')\, dt'}\, dt \tag{B.8}$$

$$= \frac{c\, \nu^*_{\text{ln}}}{d_L d_N \langle k_{\text{in}} \rangle \langle k_{\text{out}} \rangle} \int_{t=0}^{365} \bar{Z}_N(t) \int_{t'=t}^{t+d_N} \bar{Z}_L(t')\, dt' dt, \tag{B.9}$$

where c is the probability a larval tick feeds near enough to a nymph such that co-feeding transmission is possible and ν^*_{ln} is the nymph-to-larva transmission probability given the temporal and spatial requirements for co-feeding transmission have been satisfied.

Appendix C

Deriving the probability a susceptible excess degree neighbour is of type (α, β)

Introduction

Whilst deriving the basic reproduction number, R_0, for an infectious disease spreading through a clustered social network (Chapter 4), the probability any susceptible excess degree neighbour (EDN) of a typical infectious case (node n, Figure 4.1) is of type (α, β) was left undefined. To complete the derivation, this probability, denoted $a_{\alpha,\beta}$, is derived here.

C.1 Deriving $a_{\alpha,\beta}$

The probability $a_{\alpha,\beta}$ first appeared in Equation (4.14), which is the equation for the probability that n has $\omega_{\alpha,\beta}$ neighbours of type (α,β), and is repeated below for ease of reference:

$$\begin{aligned}\mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa') &= \binom{\kappa'-\theta}{\omega_{\alpha,\beta}} (a_{\alpha,\beta})^{\omega_{\alpha,\beta}} (1-a_{\alpha,\beta})^{\kappa'-\theta-\omega_{\alpha,\beta}} \\ &= f_{\mathrm{B}}(\omega_{\alpha,\beta};\kappa'-\theta,a_{\alpha,\beta})\,.\end{aligned}$$

Because $\mathrm{P}(\omega_{\alpha,\beta}|\theta,\psi,\kappa')$ is conditional on θ, ψ, and κ' so too is $a_{\alpha,\beta}$:

$$a_{\alpha,\beta} = \mathrm{P}(\alpha,\beta|\theta,\psi,\kappa')\,. \tag{C.1}$$

The three conditional random variables alone, however, are insufficient to derive an expression for $a_{\alpha,\beta}$. This is because for a susceptible EDN of n to be of type (α,β) it must not only be adjacent to α of the θ infectious EDNs of n but also β infectious neighbours of i, that are neither n nor neighbours that i and n have in common. Thus to calculate the probability a susceptible EDN of n is of type (α,β) one also has to know the degree of i. This implies Equation (C.1) needs to be expanded as follows

$$a_{\alpha,\beta} = \sum_{\kappa} \mathrm{P}(\alpha,\beta|\kappa,\theta,\psi,\kappa')\,\mathrm{P}(\kappa|\theta,\psi,\kappa')\,, \tag{C.2}$$

where $\mathrm{P}(\kappa|\theta,\psi,\kappa')$ is the conditional probability the index case i has degree κ given it shares ψ neighbours with node n and $\mathrm{P}(\alpha,\beta|\kappa,\theta,\psi,\kappa')$ is the conditional probability a susceptible EDN of n is of type (α,β) given node i has degree κ and also that it infected θ of the ψ neighbours it shares with n during the first time step.

Given nodes i and n have ψ neighbours in common the degree of i must be greater than or equal to $\psi+1$[1]. This implies

$$\mathrm{P}(\kappa|\theta,\psi,\kappa') = \frac{p_\kappa}{1-\sum_{\kappa=0}^{\psi} p_\kappa}. \tag{C.3}$$

The key to deriving an expression for $\mathrm{P}(\alpha,\beta|\kappa,\theta,\psi,\kappa')$ is to recognize that the number of infectious EDNs of n that connect to one of its susceptible EDNs is independent of the number of infectious neighbours of i, besides n and any neighbours i and n have in common, that connect to it (given i and n have ψ neighbours in common). This means that

$$\mathrm{P}(\alpha,\beta|\kappa,\theta,\psi,\kappa') = \mathrm{P}(\alpha|\kappa,\theta,\psi,\kappa')\,\mathrm{P}(\beta|\kappa,\theta,\psi,\kappa')\,. \tag{C.4}$$

[1]One is added because i is also a neighbour of n.

The first of the two conditional probabilities on the right hand side of Equation (C.4) is

$$\begin{aligned} \mathrm{P}(\alpha|\kappa,\theta,\psi,\kappa') &= \binom{\theta}{\alpha} (C_\triangle)^\alpha (1-C_\triangle)^{\theta-\alpha} \\ &= f_\mathrm{B}(\alpha;\theta,C_\triangle) \end{aligned} \tag{C.5}$$

since the probability an infectious EDN of n is adjacent to one of n's susceptible EDNs is $C_\triangle$. Similarly, the second conditional probability is given by

$$\begin{aligned} \mathrm{P}(\beta|\kappa,\theta,\psi,\kappa') &= \binom{\kappa-(\psi+1)}{\beta} (C_\square T)^\beta (1-C_\square T)^{\kappa-(\psi+1)-\beta} \\ &= f_\mathrm{B}(\beta;\kappa-(\psi+1),C_\square T) \end{aligned} \tag{C.6}$$

since i has $\kappa-(\psi+1)$ neighbours besides n and any neighbours they have in common, and $C_\square T$ is the probability any one of these neighbours was infected by i during the first time step and is also adjacent to a susceptible EDN of n.

Substituting Equations (C.3)–(C.6) into Equation (C.2) then yields

$$\begin{aligned} a_{\alpha,\beta} &= \sum_{\kappa=\psi+1}^{M} f_\mathrm{B}(\alpha;\theta,C_\triangle) f_\mathrm{B}(\beta;\kappa-(\psi+1),C_\square T) \left(\frac{p_\kappa}{1-\sum_{\kappa=0}^{\psi} p_\kappa}\right) \\ &= f_\mathrm{B}(\alpha;\theta,C_\triangle) \sum_{\kappa=\psi+1}^{M} f_\mathrm{B}(\beta;\kappa-(\psi+1),C_\square T) \left(\frac{p_\kappa}{1-\sum_{\kappa=0}^{\psi} p_\kappa}\right), \end{aligned} \tag{C.7}$$

which is the required result.

Bibliography

[1] O. Diekmann, H. Heesterbeek, and T. Britton. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, 2012.

[2] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G.S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W.J. Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, 5(3):381–391, 2008.

[3] S.P. Johnstone-Robertson, D. Mark, C. Morrow, K. Middelkoop, M. Chiswell, L.D.H. Aquino, L.-G. Bekker, and R. Wood. Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. *American Journal of Epidemiology*, 174:1246–1255, 2011.

[4] D.A. Grear, L.T. Luong, and P.J. Hudson. Network transmission inference: host behavior and parasite life cycle make social networks meaningful in disease ecology. *Ecological Applications*, 23(8):1906–1914, 2013.

[5] A. Machens, F. Gesualdo, C. Rizzo, A.E. Tozzi, A. Barrat, and C. Cattuto. An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *BMC Infectious Diseases*, 13(1):185, 2013.

[6] M.J. Smith, S. Telfer, E.R. Kallio, S. Burthe, A.R. Cook, X. Lambin, and M. Begon. Host–pathogen time series data in wildlife support a transmission function between density and frequency dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 106(19):7905–7909, 2009.

[7] M. Salathé, M. Kazandjieva, J.W. Lee, P. Levis, M.W. Feldman, and J.H. Jones. A high-resolution human contact network for infectious

disease transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51):22020–22025, 2010.

[8] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck, and P. Vanhems. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, 9(1):87, 2011.

[9] M. Begon, M. Bennett, R.G. Bowers, N.P. French, S.M. Hazel, and J. Turner. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiology and Infection*, 129 (1):147–153, 2002.

[10] S. Edlund, J. Kaufman, J. Lessler, J. Douglas, M. Bromberg, Z. Kaufman, R. Bassal, G. Chodick, R. Marom, V. Shalev, Y. Mesika, R. Ram, and A. Leventhal. Comparing three basic models for seasonal influenza. *Epidemics*, 3(3–4):135–142, 2011.

[11] C. Simon and N. Yosinao. A mathematical model to distinguish sociological and biological susceptibility factors in disease transmission in the context of H1N1/09 influenza. *Journal of Theoretical Biology*, 286: 50–56, 2011.

[12] J. Mossong and C.P. Muller. Modelling measles re-emergence as a result of waning of immunity in vaccinated populations. *Vaccine*, 21 (31):4597–4603, 2003.

[13] R.M. Anderson, H.C. Jackson, R.M. May, and A.M. Smith. Population dynamics of fox rabies in Europe. *Nature*, 289(5800):765–771, 1981.

[14] M.J. Tildesley, N.J. Savill, D.J. Shaw, R. Deardon, S.P. Brooks, M.E.J. Woolhouse, B.T. Grenfell, and M.J. Keeling. Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature*, 440(7080):83–86, 2006.

[15] R.M. May and R.M. Anderson. Transmission dynamics of HIV infection. *Nature*, 326(6109):137–142, 1987.

[16] H. McCallum, N. Barlow, and J. Hone. How should pathogen transmission be modelled? *Trends in Ecology & Evolution*, 16(6):295–300, 2001.

[17] H. Hu, K. Nigmatulina, and P. Eckhoff. The scaling of contact rates with population density for the infectious disease models. *Mathematical Biosciences*, 244(2):125–134, 2013.

[18] S. Davis, B. Abbasi, S. Shah, S. Telfer, and M. Begon. Spatial analyses of wildlife contact networks. *Journal of the Royal Society Interface*, 12 (102):20141004, 2015.

[19] L.E.C. Rocha, F. Liljeros, and P. Holme. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Computational Biology*, 7(3):1–9, 2011.

[20] I.Z. Kiss, D.M. Green, and R.R. Kao. The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread. *Journal of the Royal Society Interface*, 3(10): 669–677, 2006.

[21] T.C. Germann, K. Kadau, I.M. Longini, and C.A. Macken. Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (15):5935–5940, 2006.

[22] G.J. Milne, J.K. Kelso, H.A. Kelly, S.T. Huband, and J. McVernon. A small community model for the transmission of infectious diseases: comparison of school closure as an intervention in individual-based models of an influenza pandemic. *PLoS One*, 3(12):1–7, 2008.

[23] S. Davis, P. Trapman, H. Leirs, M. Begon, and J.A.P. Heesterbeek. The abundance threshold for plague as a critical percolation phenomenon. *Nature*, 454(7204):634–637, 2008.

[24] J. Reijniers, S. Davis, M. Begon, J.A.P. Heesterbeek, V.S. Ageyev, and H. Leirs. A curve of thresholds governs plague epizootics in Central Asia. *Ecology Letters*, 15(6):554–560, 2012.

[25] M.J. Keeling, M.E.J. Woolhouse, D.J. Shaw, L. Matthews, M. Chase-Topping, D.T. Haydon, S.J. Cornell, J. Kappey, J. Wilesmith, and B.T. Grenfell. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294(5543): 813–817, 2001.

[26] R. Albert, H. Jeong, and A.-L. Barabási. Internet: diameter of the World-Wide Web. *Nature*, 401(6749):130–131, 1999.

[27] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[28] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, and Y. Åberg. The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001.

[29] S. Milgram. The Small World Problem. *Psychology Today*, 2(1):60–67, 1967.

[30] C. Moore and M.E.J. Newman. Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678–5682, 2000.

[31] M. Kuperman and G. Abramson. Small world effect in an epidemiological model. *Physical Review Letters*, 86(13):2909–2912, 2001.

[32] P. Erdos and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[33] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[34] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[35] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, 2001.

[36] R.M. May and A.L. Lloyd. Infection dynamics on scale-free networks. *Physical Review E*, 64(6):066112, 2001.

[37] G. Caldarelli, A. Capocci, P. De Los Rios, and M.A. Munoz. Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters*, 89(25):258702, 2002.

[38] Y.-B. Xie, T. Zhou, and B.-H. Wang. Scale-free networks without growth. *Physica A: Statistical Mechanics and its Applications*, 387(7): 1683–1688, 2008.

[39] T. Gross, C.J.D. D'Lima, and B. Blasius. Epidemic dynamics on an adaptive network. *Physical Review Letters*, 96(20):208701, 2006.

[40] S. Funk, M. Salathé, and V.A.A. Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7(50):1247–1256, 2010.

[41] S. Funk, E. Gilad, C. Watkins, and V.A.A. Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (16):6872–6877, 2009.

[42] M.E.J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, 2002.

[43] J. Gómez-Gardeñes, V. Latora, Y. Moreno, and E. Profumo. Spreading of sexually transmitted diseases in heterosexual populations. *Proceedings of the National Academy of Sciences of the United States of America*, 105(5):1399–1404, 2008.

[44] C.J. Rhodes and R.M. Anderson. Dynamics in a lattice epidemic model. *Physics Letters A*, 210(3):183–188, 1996.

[45] C.J. Rhodes and R.M. Anderson. Epidemic thresholds and vaccination in a lattice model of disease spread. *Theoretical Population Biology*, 52 (2):101–118, 1997.

[46] S.P. Johnstone-Robertson, P.J.S. Fleming, M.P. Ward, and S.A. Davis. Predicted spatial spread of canine rabies in Australia. *PLoS Neglected Tropical Diseases*, 11(1):e0005312, 2017.

[47] World Health Organization. WHO expert consultation on rabies: second report. Technical Report 982, World Health Organization Technical Report Series, 2013.

[48] J. Sparkes, P.J.S. Fleming, G. Ballard, H. Scott-Orr, S. Durr, and M.P. Ward. Canine rabies in Australia: a review of preparedness and research needs. *Zoonoses and Public Health*, 62(4):237–253, 2015.

[49] World Health Organization. Rabies vaccines: WHO position paper. *Weekly Epidemiological Record*, 85(32):309–320, 2010.

[50] J.T. Paweska, L.H. Blumberg, C. Liebenberg, R.H. Hewlett, A.A. Grobbelaar, P.A. Leman, J.E. Croft, L.H. Nel, L. Nutt, and R. Swanepoel. Fatal human infection with rabies-related Duvenhage virus, South Africa. *Emerging Infectious Diseases*, 12(12):1965–1967, 2006.

[51] World Health Organization. Rabies fact sheet no 99. `http://www.who.int/mediacentre/factsheets/fs099/en/`. Accessed: 6 June 2015.

[52] J.S. Smith. New aspects of rabies with emphasis on epidemiology, diagnosis, and prevention of the disease in the United States. *Clinical Microbiology Reviews*, 9(2):166–176, 1996.

[53] A.I. Wandeler, J. Bingham, and F.X. Meslin. Dogs and rabies. In C.N.L. Macpherson, F.X. Meslin, and A.I. Wandeler, editors, *Dogs, Zoonoses and Public Health*, chapter 4, pages 43–66. CABI, Wallingford, UK, 2013.

[54] T. Hutabarat, M. Geong, A. Newsome, A. Ruben, and S. Cutter. Rabies and dog ecology in Flores. In *Urban Animal Management Conference Proceedings*, pages 7–14, Australia, 2003. ACIAR.

[55] A.A.G. Putra, K. Hampson, J. Girardi, E. Hiby, D. Knobel, I.W. Mardiana, S. Townsend, and H. Scott-Orr. Response to a rabies epidemic, Bali, Indonesia, 2008–2011. *Emerging Infectious Diseases*, 19(4):648–651, 2013.

[56] P.J.S. Fleming, B.L. Allen, G. Ballard, and L.R. Allen. Wild dog ecology, impacts and management in northern Australian cattle enterprises: a review with recommendations for RD&E investments. Technical report, Meat and Livestock Australia Limited, North Sydney, 2012.

[57] Tenzin and M.P. Ward. Review of rabies epidemiology and control in South, South East and East Asia: past, present and prospects for elimination. *Zoonoses and Public Health*, 59(7):451–467, 2012.

[58] A.A.G Putra, I. Gunata, D. Dharma, and H. Scott-Orr. Rabies on the move in Indonesia: incursion into Bali and response. In *Proceedings of the 12th Symposium of the International Society for Veterinary Epidemiology and Economics*, Durban, South Africa, 2009.

[59] M.P. Ward and M. Hernández-Jover. A generic rabies risk assessment tool to support surveillance. *Preventive Veterinary Medicine*, 120(1): 4–11, 2015.

[60] V.J. Brookes and M.P. Ward. Expert opinion to identify high-risk entry routes of canine rabies into Papua New Guinea. *Zoonoses and Public Health*, 2016.

[61] S. Dürr and M.P. Ward. Roaming behaviour and home range estimation of domestic dogs in Aboriginal and Torres Strait Islander communities in northern Australia using four different methods. *Preventive Veterinary Medicine*, 117(2):340–357, 2014.

[62] J. Sparkes, G. Körtner, G. Ballard, P.J.S. Fleming, and W.Y. Brown. Effects of sex and reproductive state on interactions between free-roaming domestic dogs. *PLoS One*, 9(12):1–13, 2014.

[63] B.L. Allen and P. West. Influence of dingoes on sheep distribution in Australia. *Australian Veterinary Journal*, 91(7):261–267, 2013.

[64] P.J.S. Fleming, B.L. Allen, L.R. Allen, G.A. Ballard, A.J. Bengsen, M.N. Gentle, L.J. McLeod, P.D. Meek, and G.R. Saunders. Management of wild canids in Australia: free-ranging dogs and red foxes. In A.S. Glen and C.R. Dickman, editors, *Carnivores of Australia: Past, Present and Future*, chapter 6, pages 107–152. CSIRO Publishing, Collingwood, Melbourne, 2014.

[65] J. Sparkes, G. Ballard, P. Fleming, and W. Brown. Social, conservation and economic implications of rabies in Australia. *Australian Zoologist*, 2014.

[66] S. Dürr and M.P. Ward. Development of a novel rabies simulation model for application in a non-endemic environment. *PLoS Neglected Tropical Diseases*, 9(6):1–22, 2015.

[67] S.A. Davis. Percolation on a spatial network with individual heterogeneity as a model for disease spread among animal host populations. In *19th International Congress on Modelling and Simulation*, pages 905–911, Perth, Australia, 2011.

[68] P.J.S. Fleming, L. Corbett, R. Harden, and P. Thomson. *Managing the Impacts of Dingoes and Other Wild Dogs*. Bureau of Rural Sciences, Canberra, 2001.

[69] L.K. Corbett. *The Dingo in Australia and Asia*. JB Books, Marleston, South Australia, 2nd edition, 2001.

[70] G. Grimmett. *Percolation*. Springer-Verlag, Berlin, 2nd edition, 1999.

[71] K. Hampson, J. Dushoff, S. Cleaveland, D.T. Haydon, M. Kaare, C. Packer, and A. Dobson. Transmission dynamics and prospects for the elimination of canine rabies. *PLoS Biology*, 7(3):462–471, 2009.

[72] K. Hampson, J. Dushoff, J. Bingham, G. Brückner, Y.H. Ali, and A. Dobson. Synchronous cycles of domestic dog rabies in sub-Saharan Africa and the impact of control efforts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18):7717–7722, 2007.

[73] M.J. Carroll, A. Singer, G.C. Smith, D.P. Cowan, and G. Massei. The use of immunocontraception to improve rabies eradication in urban dog populations. *Wildlife Research*, 37(8):676–687, 2010.

[74] J. Wu, R. Dhingra, M. Gambhir, and J.V. Remais. Sensitivity analysis of infectious disease models: methods, advances and their application. *Journal of The Royal Society Interface*, 10(86):1–14, 2013.

[75] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, 2002.

[76] F. Cannavó. Sensitivity analysis for volcanic source modeling quality assessment and model selection. *Computers & Geosciences*, 44:52–59, 2012.

[77] G.E.B. Archer, A. Saltelli, and I.M. Sobol. Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2):99–120, 1997.

[78] D. Stephens, A.N. Wilton, P.J.S. Fleming, and O. Berry. Death by sex in an Australian icon: a continent-wide survey reveals extensive hybridization between dingoes and domestic dogs. *Molecular Ecology*, 24(22):5643–5656, 2015.

[79] C.M. Foggin. *Rabies and Rabies-Related Viruses in Zimbabwe: Historical, Virological and Ecological Aspects*. PhD thesis, University of Zimbabwe, Harare, 1988.

[80] V. Tepsumethanon, B. Lumlertdacha, C. Mitmoonpitak, V. Sitprija, F.X. Meslin, and H. Wilde. Survival of naturally infected rabid dogs and cats. *Clinical Infectious Diseases*, 39(2):278–280, 2004.

[81] P.G. Coleman and C. Dye. Immunization coverage required to prevent outbreaks of dog rabies. *Vaccine*, 14(3):185–186, 1996.

[82] P.J.S. Fleming, G. Ballard, and T.M. Newsome. Unpublished data.

[83] C.D. James, J. Landsberg, and S.R. Morton. Provision of watering points in the Australian arid zone: a review of effects on biota. *Journal of Arid Environments*, 41(1):87–121, 1999.

[84] T.R. Eng, D.B. Fishbein, H.E. Talamante, D.B. Hall, G.F. Chavez, J.G. Dobbins, F.J. Muro, J.L. Bustos, M. de los Angeles Ricardy, A. Munguia, J. Carrasco, A.R. Robles, and G.M. Baer. Urban epizootic

of rabies in Mexico: epidemiology and impact of animal bite injuries. *Bulletin of the World Health Organization*, 71(5):615–624, 1993.

[85] M.K. Morters, O. Restif, K. Hampson, S. Cleaveland, J.L.N. Wood, and A.J.K. Conlan. Evidence-based control of canine rabies: a critical review of population density reduction. *Journal of Animal Ecology*, 82 (1):6–14, 2013.

[86] K. Hampson, A. Dobson, M. Kaare, J. Dushoff, M. Magoto, E. Sindoya, and S. Cleaveland. Rabies exposures, post-exposure prophylaxis and deaths in a region of endemic canine rabies. *PLoS Neglected Tropical Diseases*, 2(11):e339, 2008.

[87] M.E. Craft. Infectious disease transmission and contact networks in wildlife and livestock. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1669), 2015.

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

[89] L. Danon, A.P. Ford, T. House, C.P. Jewell, M.J. Keeling, G.O. Roberts, J.V. Ross, and M.C. Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, 2011:284909, 2011.

[90] V.G. Panjeti and L.A. Real. Mathematical models for rabies. In A.C. Jackson, editor, *Advances in Virus Research: Research Advances in Rabies*, volume 79, pages 377–395. Academic Press, 2011.

[91] T.M. Newsome, G.-A. Ballard, C.R. Dickman, P.J.S. Fleming, and R. van de Ven. Home range, activity and sociality of a top predator, the dingo: a test of the Resource Dispersion Hypothesis. *Ecography*, 36 (8):914–925, 2013.

[92] P. Lemey, A. Rambaut, J.J. Welch, and M.A. Suchard. Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, 27(8):1877–1885, 2010.

[93] A.J. Hendriks, B.J.C. Willers, H.J.R. Lenders, and R.S.E.W. Leuven. Towards a coherent allometric framework for individual home ranges, key population patches and geographic ranges. *Ecography*, 32(6):929–942, 2009.

[94] S. Davis and S.J. Bent. Loop analysis for pathogens: niche partitioning in the transmission graph for pathogens of the North American tick *Ixodes scapularis*. *Journal of Theoretical Biology*, 269(1):96–103, 2011.

[95] J.L. Brunner, L. Cheney, F. Keesing, M. Killilea, K. Logiudice, A. Previtali, and R.S. Ostfeld. Molting success of *Ixodes scapularis* varies among individual blood meal hosts and species. *Journal of Medical Entomology*, 48(4):860–866, 2011.

[96] M.E.J. Woolhouse, C. Dye, J.-F. Etard, T. Smith, J.D. Charlwood, G.P. Garnett, P. Hagan, J.L.K. Hii, P.D. Ndhlovu, R.J. Quinnell, C.H. Watts, S.K. Chandiwana, and R.M. Anderson. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences*, 94(1): 338–342, 1997.

[97] D.J. Shaw, B.T. Grenfell, and A.P. Dobson. Patterns of macroparasite aggregation in wildlife host populations. *Parasitology*, 117(6):597–610, 1998.

[98] J.L. Brunner and R.S. Ostfeld. Multiple causes of variable tick burdens on small-mammal hosts. *Ecology*, 89(8):2259–2272, 2008.

[99] S.E. Randolph, D. Miklisova, J. Lysy, D.J. Rogers, and M. Labuda. Incidence from coincidence: patterns of tick infestations on rodents facilitate transmission of tick-borne encephalitis virus. *Parasitology*, 118(2):177–186, 1999.

[100] N.G. Craine, S.E. Randolph, and P.A. Nuttall. Seasonal variation in the role of grey squirrels as hosts of *Ixodes ricinus*, the tick vector of the Lyme disease spirochaete, in a British woodland. *Folia Parasitologica*, 42:73–80, 1995.

[101] J.M. Dunn. *The Mathematical Epidemiology of Human Babesiosis in the North-Eastern United States*. PhD thesis, RMIT University, 2014.

[102] A. Milne. The ecology of the sheep tick, *Ixodes ricinus* L. Host relationships of the tick: part 2. Observations on hill and moorland grazings in northern England. *Parasitology*, 39(3-4):173–197, 1949.

[103] S.E. Randolph and N.G. Craine. General framework for comparative quantitative studies on transmission of tick-borne diseases using Lyme borreliosis in Europe as an example. *Journal of Medical Entomology*, 32(6):765–777, 1995.

[104] N.A. Hartemink, S.E. Randolph, S.A. Davis, and J.A.P. Heesterbeek. The basic reproduction number for complex disease systems: defining R_0 for tick-borne infections. *The American Naturalist*, 171(6):743–754, 2008.

[105] L. Tälleklint and T.G.T. Jaenson. Maintenance by hares of European *Borrelia burgdorferi* in ecosystems without rodents. *Journal of Medical Entomology*, 30(1):273–276, 1993.

[106] J.S. Gray. The development and seasonal activity of the tick *Ixodes ricinus*: a vector of Lyme borreliosis. *Review of Medical and Veterinary Entomology*, 79(6):323–333, 1991.

[107] G.M. Steele and S.E. Randolph. An experimental evaluation of conventional control measures against the sheep tick, *Ixodes ricinus* (L.) (Acari: Ixodidae). I. A unimodal seasonal activity pattern. *Bulletin of Entomological Research*, 75(3):489–500, 1985.

[108] J.M. Dunn, S. Davis, A. Stacey, and M.A. Diuk-Wasser. A simple model for the establishment of tick-borne pathogens of *Ixodes scapularis*: a global sensitivity analysis of R_0. *Journal of Theoretical Biology*, 335:213–221, 2013.

[109] R.S. Ostfeld, M.C. Miller, and K.R. Hazler. Causes and consequences of tick (*Ixodes scapularis*) burdens on white-footed mice (*Peromyscus leucopus*). *Journal of Mammalogy*, 77(1):266–273, 1996.

[110] A. Matser, N. Hartemink, H. Heesterbeek, A. Galvani, and S. Davis. Elasticity analysis in epidemiology: an application to tick-borne infections. *Ecology Letters*, 12(12):1298–1305, 2009.

[111] S.E. Randolph, L. Gern, and P.A. Nuttall. Co-feeding ticks: epidemiological significance for tick-borne pathogen transmission. *Parasitology Today*, 12(12):472–479, 1996.

[112] F.-R. Matuschka, T.W. Schinkel, B. Klug, A. Spielman, and D. Richter. Failure of Ixodes ticks to inherit *Borrelia afzelii* infection. *Applied and Environmental Microbiology*, 64(8):3089–3091, 1998.

[113] V. Danielová, J. Holubová, M. Pejcoch, and M. Daniel. Potential significance of transovarial transmission in the circulation of tick-borne encephalitis virus. *Folia Parasitologica*, 49(4):323–325, 2002.

[114] J. Piesman, J.G. Donahue, T.N. Mather, and A. Spielman. Transovarially acquired Lyme disease spirochetes (*Borrelia burgdorferi*) in field-collected larval *Ixodes dammini* (Acari: Ixodidae). *Journal of Medical Entomology*, 23(2):219–219, 1986.

[115] O. Diekmann, J.A.P. Heesterbeek, and J.A.J. Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.

[116] M. Newman. *Networks: An Introduction*. Oxford University Press Inc., New York, 2010. ISBN 978-0-19-920665-0.

[117] C. Molina and L. Stone. Modelling the spread of diseases in clustered networks. *Journal of Theoretical Biology*, 315:110–118, 2012.

[118] R. Parshani, S. Carmi, and S. Havlin. Epidemic threshold for the susceptible-infectious-susceptible model on random networks. *Physical Review Letters*, 104:258701, 2010.

[119] E. Shtilerman and L. Stone. The effects of connectivity on metapopulation persistence: network symmetry and degree correlations. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1806), 2015.

[120] M.J. Voordouw. Co-feeding transmission in Lyme disease pathogens. *Parasitology*, 142(2):290–302, 2015.

[121] S.L. States, C.I. Huang, S. Davis, D.M. Tufts, and M.A. Diuk-Wasser. Co-feeding transmission facilitates strain coexistence in *Borrelia burgdorferi*, the Lyme disease agent. *Epidemics*, 19:33–42, 2017.

[122] M.E.J. Newman. Properties of highly clustered networks. *Physical Review E*, 68(2):026121, 2003.

[123] T. Britton, M. Deijfen, A.N. Lagerås, and M. Lindholm. Epidemics on random graphs with tunable clustering. *Journal of Applied Probability*, 45(3):743–756, 2008.

[124] M.E.J. Newman. Random graphs with clustering. *Physical Review Letters*, 103(5):058701, 2009.

[125] J.C. Miller. Percolation and epidemics in random clustered networks. *Physical Review E*, 80(2):020901, 2009.

[126] J.C. Miller. Spread of infectious disease through clustered populations. *Journal of the Royal Society Interface*, 6(41):1121–1134, 2009.

[127] M. Keeling. The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, 67(1):1–8, 2005.

[128] I.Z. Kiss and D.M. Green. Comment on "Properties of highly clustered networks". *Physical Review E*, 78(4):048101, 2008.

[129] O. Diekmann and J.A.P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, volume 5. John Wiley & Sons, 2000.

[130] J.P. Aparicio and M. Pascual. Building epidemiological models from r_0: an implicit treatment of transmission in networks. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1609):505–512, 2007.

[131] M.E.J. Newman. Random graphs as models of networks. *arXiv preprint cond-mat/0202208*, 2002.