Stochastic evolutionary and epidemic processes on networks

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

Christopher E. Overton

July 2020

Abstract

Population dynamics have traditionally focussed on large, homogeneous populations. However, real populations usually have some degree of structure that determines how individuals interact. This thesis is concerned with such structured population dynamics. Our first direction is to add further realism to network-structured evolutionary processes, by considering ecological dynamics and environmental variation. To accommodate realistic ecological dynamics we develop an individual-based model for evolution in network-structured population in which birth and death events are decoupled. From this, we derive the evolutionary graph theory model, gaining insight into the assumptions underpinning this model. These ecological dynamics are then used to explain why certain networks allow adaptive mutations to spread more successfully and to investigate the effect of clonal interference on networks. In network-structured populations, environmental variation has thus far not been considered. However, evolution in variable environments has been widely studied in well-mixed populations. This variation leads to evolutionary bet-hedging, where species hedge against these fluctuations to reduce their likelihood of extinction. We add such environmental variation to the evolutionary graph theory framework, to explore the evolution of bet-hedging strategies in network-structured populations. Variation can act either within or between generations. In large well-mixed populations, it has been shown that selection for bet-hedging against within-generational variation should not occur, contradicting empirical observations. We show that in network-structured populations within-generational variation can have a significant impact on the evolutionary process for any population size.

This realism adds further complexity to the evolutionary models, and therefore these analyses rely mostly on stochastic simulations. Analytical results are challenging, since network-structure causes the state space of the model to increase exponentially with population size. For some networks with sufficient symmetry, analytical results may be possible, however this does not facilitate systematic analysis of different network structures. Therefore, approximation methods can be powerful. By adapting methods from statistical physics, including moment-closure techniques, we develop node-level approximations to the standard evolutionary graph theory dynamics on arbitrarily complex networks. Such moment-closure methods are also commonly used to approximate network-based epidemic models. In this setting, these approximations are efficient to analyse and provide theoretical insights, such as the epidemic threshold. However, it can be hard to relate these

theoretical results to the underlying stochastic process, since these simplified models exhibit a stable endemic equilibrium that is not present in the stochastic model. Building on the existing methods, we develop a framework that is directly related to the underlying stochastic process by approximating the expected number of infected individuals in the quasi-stationary distribution. This describes the expected behaviour of the system given that the disease-free state has not been reached, providing a picture of endemic disease prevalence. These methods provide a toolkit for approximating network-structured population dynamics, and can facilitate approximation of the more realistic dynamics obtained through ecological dynamics and environmental variation.

Acknowledgements

There are many people who have helped throughout the course of this research. Firstly, I would like to thank my supervisor Kieran Sharkey for giving me this opportunity and supporting me throughout. His guidance and knowledge ensured the successful completion of this research. I look forward to continuing to work together in the future.

I am thankful to my numerous collaborators that have helped guide my research. Their knowledge and expertise has been invaluable in ensuring my research has gone smoothly, and I have learnt a lot through our discussions and interactions.

I am grateful for the welcoming scientific community that I have entered. Through various workshops, seminars and conferences I have had the opportunity to engage with experts from a wide range of scientific backgrounds, all of whom have furthered my development through fruitful discussions.

My colleagues and peers at the University of Liverpool, including the PostGrad Society and the maths 5-a-side football players, have played a vital role throughout my PhD, supporting me both academically and socially. This network has made the past few years both enjoyable and productive.

This research would not have been possible without the support of my parents and my partner, Charlotte. Their support through the decision to undertake a PhD and the following workload provided a bedrock on which I could comfortably build my research. Their faith and encouragement allowed me to persevere and complete this work.

Finally, I would like to thank the University of Liverpool and Department of Mathematical Sciences for funding my research and providing a supporting environment in which to work.

Contents

List of Figures

List of Tables

List of Notation

- $\Sigma(t)$ - Continuous time stochastic process
- Σ_t - Discrete time stochastic process
- $P(\Sigma(t) = j)$ - probability that the stochastic process is in state j
- Q - transition rate matrix
- $\mathbf{p}(t)$ - Probability distribution of a Markov process at time t
- $\mathcal{L}$ - infinitesimal generator
- $\rho(\Sigma(t) = j)$ - probability that the stochastic process is in state j conditioned on non-absorption
- $\mathcal{G}$ - network
- $\mathcal{V}$ - set of nodes in a network
- $\mathcal{E}$ - set of edges in a network
- G - adjacency matrix for a network $\mathcal{G}$
- f_A - fitness of type A individual
- f_A^i - fitness of type A individual in node i
- $\rho^A(i)$ - fixation probability of type A from a state i
- W - weighted adjacency matrix
- k_j - degree of node j
- $[S_k]$ - expected number of degree k nodes in state S
- $\langle S_i \rangle$ - probability that node i is in state S
- $M(U_i, w)$ - probability that a type U_i individuals gives birth to a type w offspring

- $b(i, x, \mathcal{I}, W)$ - rate of which individual i gives birth onto site x, given the population is in state $\mathcal{I}$ and has weighted adjacency matrix W
- $d(i, \mathcal{I}, W)$ - death rate of individual i given the population is in state $\mathcal{I}$ and has weighted adjacency matrix W
- $B(c, i, n, \mathcal{I}, W)$ - birth rate modified by negative feedback loop
- $D(c, i, \mathcal{I}, W)$ - death rate modified by negative feedback loop
- $h_{\mathcal{A}}(\mathcal{I})$ - hitting probability of state $\mathcal{A}$ starting from state $\mathcal{I}$
- X_C^t - the event where the set of nodes C is in state X at time t
- $\chi(\Omega_{j\to i}^t | S^t)$ - the rate at which individual j replaces individual i at time t given the process is in state S
- σ_α - ordered states of the stochastic process
- $\rho_\alpha(t) = \rho(\Sigma(t) = \sigma_\alpha)$ - vectorised conditional distribution
- $P_\alpha(t) = P(\Sigma(t) = \sigma_\alpha)$ - vectorised probability distribution
- $(QP)_\alpha$ - rate of change for state σ_α
- $|C_k|$ - number of degree k nodes
- $|C_{k,l}|$ - number of pairs involving a degree k and a degree l node

Chapter 1

Introduction

Evolution is a complex process in population biology, with applications ranging from healthcare [51, 98, 104, 176, 182] to wildlife [110]. Evolution describes how the characteristics of a population change over time, in order to become better suited to their environment or adapt to changing environments. This happens through a combination of mutation, which generates new characteristics, and competition, whereby the new characteristics eventually replace the old characteristics. The work of Charles Darwin in "On the Origin of species by Means of Natural Selection" [34] provides a ground on which most evolutionary theory has been built. Darwin suggests that natural selection leads to the evolution of populations; i.e. that nature will select for mutants with advantageous characteristics, and eventually these will replace the existing characteristics. Such evolution can act on any population, and the timescale depends on the rate of mutation, lifespan, and selective pressure for each mutation from the environment. The complexity of evolution has produced multiple modelling mechanisms, from the micro-scale, looking at rates of mutation [63, 150], to the macro scale, looking at competition between different types [21, 102, 109, 116].

The traditional approach to modelling evolution has been to use deterministic models under the assumption that the population is homogeneous. These assumptions are made since they facilitate efficient analysis; using models such as adaptive dynamics [37] and the replicator equation [28]. These models give insight into the evolutionary process, however evolution is highly stochastic, since both mutation and the population dynamics are random. To address this issue, stochastic evolutionary models have been developed, such as the Moran process [116] and the Wright-Fisher model [45, 196]. These stochastic

models assume that although a mutant may have a fitness advantage, it is not guaranteed to take over a resident population, due to the random nature of birth and death. A key difference between these models and the deterministic models is their finite population size. Deterministic models often consider frequencies of each species, which corresponds to the infinite population limit where stochasticity is assumed not to be important. However, in stochastic models it has been observed that when taking the limit of infinite populations, stochastic effects can still be present for a finite initial mutant population [116], and therefore even in large populations, ignoring the stochasticity may not be a realistic assumption.

Another limiting factor in traditional evolutionary models is the assumption of well-mixed populations, where all individuals interact with each other. Real populations usually exhibit some degree of structure, which can correspond to spatial structure or social structure, for example. Population structure can have a significant impact on evolution, such as amplifying the probability of an advantageous mutant taking over the population [102]. Therefore, structure needs to be considered when attempting to understand real systems.

A vital area for research into evolution today is in the healthcare environment. Many current healthcare challenges are characterised by an underlying evolutionary process, whether this is related to waning immunity due to evolving pathogens [41, 154], pathogens evolving resistance to standard treatment regimes [98, 176] or the development of cancer and its ability to defend against treatment [202]. Of particular focus in this thesis is the study of the evolution of pathogens, since the study of epidemic dynamics involves systems similar to those of evolutionary dynamics. This link is important because it facilitates the use of existing techniques into modelling pathogen competition and evolution.

Antimicrobial resistance, for example, is an adaptation (which can potentially arise through mutation or horizontal gene transfer) that enables pathogens to become more resistant to current treatment options. This can lead to unchecked spread of a pathogen and higher risk of death from infections. Addressing this requires understanding how these mutant strains interact and compete with the existing resident strains. Using evolutionary or epidemic theory alone may not be enough to understand this problem, and therefore studying the two in tandem may grant further insight into the interactions between these processes. This will also facilitate research into processes such as the spread of influenza and HIV. Influenza outbreaks occur seasonally, with people becoming infected multiple times. This is due to the wide range of influenza strains co-circulating, as well as mutation within the strains producing new types that our bodies can no longer recognise and fight [41]. This

complex interplay between mutation and competition presents a challenge to healthcare modellers when trying to produce the vaccine each winter, since they need to predict which strains will be most prevalent. Similarly, in HIV the rate at which the virus mutates within hosts makes it a challenging pathogen to fight, because it can evolve to evade the immune system [182].

Having a better understanding of the evolutionary process will allow us to combine the traditional epidemic modelling process with evolution. Within this thesis, we do not explore any particular examples of pathogens, but instead investigate similarities between epidemic and evolutionary models, and study important properties of these two systems, which may prove useful in understanding real healthcare challenges.

This chapter introduces the evolution and epidemic modelling frameworks that we use in this thesis. We first introduce stochastic processes, in particular discrete-time and continuous-time Markov chains, which form the basis of the models we use. After this we introduce networks, describing notation and some examples. Networks are used to represent the structure of the populations that we are modelling. The next step is the introduction of the Moran process, which we then link to evolutionary graph theory, a model for studying evolution in network-structured populations. Finally, the stochastic, network-based SIS epidemic model is introduced. This chapter ends with an outline of the research presented in this thesis.

1.1 Stochastic Processes

A stochastic process is a function that varies randomly [80, 130, 183]. This random variation means that knowing the value of the function at a given index (which can refer to discrete steps or continuous time) does not allow one to know the future behaviour of the function. Instead, it is only possible to know the probability of certain future realisations being obtained. In biological scenarios, this uncertainty can be due to the random nature of birth and death, or in the financial world, the uncertainty can be due to randomness in the value of the underlying product. Stochastic processes can be developed to describe these systems, granting some insight into the potential future outcomes.

Formally, a stochastic process is a function Σ of a random variable Z and an index variable t [130, 183]; i.e.

$$\Sigma(t) = f(Z, t).$$

$\Sigma(t)$ takes values $\sigma \in S$, Z takes values $z \in \Omega$, and the index variable t can correspond to time or some counting variable, for example. Each of the possible realisations of Σ corresponds to a value z of the random variable. Where the random variable Z selects a realised value z from its state space Ω, the stochastic process Σ selects a realised sample path σ from the space of all possible paths. In this thesis, we mainly focus our attention on stochastic processes that act on finite or countable state spaces, since in the systems we consider we are interested in discrete states.

1.1.1 Discrete-time Markov chain

Consider a stochastic process $\Sigma(t)$, whose state space is finite and whose index variable is a counting variable. That is, the process can take states from the finite set S, and the index variable t comes from $T = \{0, 1, 2, ...\}$. This is a discrete-time stochastic process, which we will denote by Σ_t. Such a process is called a discrete-time Markov chain if the likelihood of future events depends only on the state of the system at the given time and not the past behaviour (this is called the Markov property) [80, 183]. That is, given the system has reached a certain state, the route that it took to reach this state has no impact on the future dynamics. More formally, a stochastic process Σ_t is a Markov chain if and only if

$$P(\Sigma_t = j|, \Sigma_0 = \sigma_0, \Sigma_1 = \sigma_1, \Sigma_2 = \sigma_2, ..., \Sigma_{t-1} = i) = P(\Sigma_t = j|\Sigma_{t-1} = i).$$

This is the transition probability of the process, and describes the structure of the Markov chain. Using the transition probability, the probability distribution of the next time step can be calculated, based on the initial distribution.

A Markov process is called stationary if this transition probability is time independent [80, 183], so that the probability of moving between two states only depends on the time steps considered, and not the value of t. The models in this thesis follow stationary processes, and therefore we will restrict our attention to such cases. The transition probability therefore becomes

$$P_{ij} = P(\Sigma_t = j|\Sigma_{t-1} = i) = P(\Sigma_1 = j|\Sigma_0 = i),$$

which is independent of the current time t.

1.1.2 Transition probability matrix

A discrete-time Markov chain can be completely defined by a one-step transition probability matrix and the initial probability distribution across its state space. The one-step transition probability matrix, P, consists of entries P_{ij} that denote the transition probabilities from state i to state j. We can calculate the probability of being in a certain state at any future time given an initial state by constructing a matrix $P^{(n)}$ that describes the n-step transition probabilities of the Markov chain; i.e. each entry of $P^{(n)}$ describes the probability of moving from a state i to a state j in n steps. It has been shown [80] that the n-step transition probability matrix is the one step transition probability matrix to the power of n; i.e. $P^{(n)} = P^n$.

Denoting the current distribution by the row vector $\mathbf{p}_0$, such that each entry of $\mathbf{p}_0$ is the probability of the Markov chain being in a certain state, the distribution in n time-steps is given by

$$\mathbf{p}_n = \mathbf{p}_0 P^n. \tag{1.1}$$

1.1.3 Communicating classes

From a certain state of the Markov chain, it may be possible to reach either any other state or only a subset of the state space. If a state $j \in S$ can be reached from state $i \in S$, then j is said to be accessible from i. State j is therefore accessible from state i if and only if $P_{ij}^N > 0$ for some integer $N \geq 0$ [80], since P_{ij}^N is the probability of reaching state j from i in N steps.

If i is also accessible from j, these two states are said to communicate with each other. This communication relation forms an equivalence relation on the state space of the Markov Chain, whereby all individuals equivalent with respect to this relation communicate. We refer to a subset of individuals that communicate with each other as a communicating class, and the state space can be stratified into communicating classes. A property revealed through this stratification is that although it may be possible to exit one class and enter another, it will never be possible to re-enter the original class, since otherwise an individual in this class would communicate with the second class. When the entire state space reduces to a single communicating class, we say that the Markov Chain is irreducible [80].

If there exists a communicating class from which no other communicating class is accessible, this is an absorbing class of the system, since once the process enters this class it will never leave [80]. This is a particularly relevant concept in population dynamics, where

there often exists states of extinction from which it is impossible to enter a non-extinct class. If an absorbing class is always accessible from any non-absorbing class then we know, with probability one, that the system will eventually always enter one of the absorbing classes and never leave, so these form the possible final states of the system. Figure 1.1 illustrates how communicating classes are related to the transitions of the Markov process.

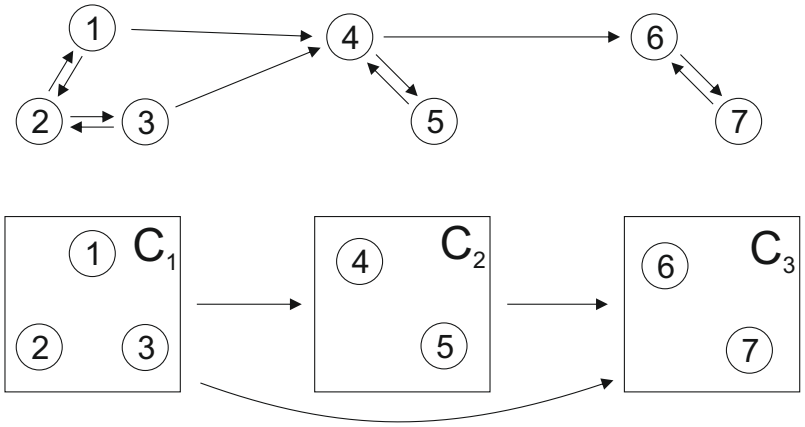

Figure 1.1: Illustration of communicating classes. The upper figure shows which states can transition into each other. The lower figure shows how these states can be grouped into communicating classes, and which classes are accessible from others. Note that C_3 is an absorbing class of the system.

1.1.4 Continuous-time Markov chain

Although discrete-time Markov chains can be useful, population dynamics often involve continuous-time systems, which can be represented by continuous-time Markov chains. A continuous-time Markov chain is a stochastic process on a finite or countable state space which satisfies the Markov property (Equation (1.1)), but now the index variable is continuous rather than discrete [130, 183], and represents time.

Therefore, we are interested in processes $Y(t)$ with $t \in [0, \infty)$, which satisfy

$$P(\Sigma(t) = j | \Sigma(t_1) = \sigma_1, \Sigma(t_2) = \sigma_2, ..., \Sigma(t_n) = i) = P(\Sigma(t) = j | \Sigma(t_n) = i),$$

whenever $t_1 < t_2 < ... < t_n < t$.

Since this process is on a continuous index variable, a one-step transition probability cannot be defined. Instead, the rate at which transitions occur is used to investigate the model.

The rate at which the process leaves a state i and moves to state j is given by

$$q_{ij} = \lim_{h \to 0} \frac{P_{ij}(h)}{h}, \tag{1.2}$$

where

$$P_{ij}(t) = P(\Sigma(t) = j | \Sigma(0) = i).$$

The total rate at which the process leaves a state i, denoted by q_i, is given by

$$q_i = \lim_{h \to 0} \frac{1 - P_{ii}(h)}{h}. \tag{1.3}$$

Since we know

$$P_{ii}(h) + \sum_{j=0:j \neq i}^{\infty} P_{ij}(h) = 1,$$

we must have

$$q_i = \sum_{j=0:j \neq i}^{\infty} q_{ij}. \tag{1.4}$$

The transition rates q_{ij} are bounded such that $0 \leq q_{ij} < \infty$ [80]. For an infinite state Markov chain it is possible that q_i can be infinite. For Markov chains with a finite state space however, q_i is finite since Equation (1.4) becomes

$$q_i = \sum_{j=0:j \neq i}^{N} q_{ij}.$$

Using these infinitesimal transition rates a matrix can be constructed called the transition rate matrix (or Q-matrix), with entries $Q_{ij} = q_{ji}$ if $j \neq i$ and $Q_{ii} = -q_i$ [125]. Note that in standard probability theory this matrix is usually defined as $Q_{ij} = q_{ij}$. We opt to use the former, since this is commonly used in statistical physics and epidemiology, which provides the basis to many of the methods that we develop. This matrix contains all the information about the transitions of the Markov chain, and can be used to reconstruct the

transition probabilities $P_{ij}(t)$. That is,

$$\begin{aligned} P_{ij}(h) &= Q_{ji}h + o(h) \\ P_{ii}(h) &= 1 + Q_{ii}h + o(h) \end{aligned} \tag{1.5}$$

The time evolution of the Markov chain can be described using a master equation

$$\frac{d\mathbf{p}(t)}{dt} = Q\mathbf{p}(t), \tag{1.6}$$

where $\mathbf{p}(t)$ is a column vector describing the probability that the Markov chain is in each state; i.e. $\mathbf{p}_i(t) = P(\Sigma(t) = i)$. This can be solved for any initial condition to describe the future distributions of the Markov chain.

In the case where the state space of the process is uncountable, we refer to this as a continuous-time Markov process rather than chain. For such a case, we cannot define a transition matrix since we do not have a countable set of states to transition to.

1.1.5 Poisson process

A homogeneous Poisson process [80] is a stochastic process that is often used in modelling population dynamics. The Poisson process is a continuous-time Markov chain $\Sigma(t)$, that counts the number of events up to time t, where in the time interval $(t, t+h)$ the probability of one event is $\lambda h + o(h)$, and the probability of more than one is $o(h)$.

Therefore the Poisson process has transition probability function

$$P(\Sigma(t+h) = j | \Sigma(t) = i) = \begin{cases} \lambda h + o(h) & \text{for } j = i+1 \\ o(h) & \text{for } j > i+1 \\ 0 & \text{for } j < i. \end{cases}$$

The elements of the Q-matrix can be constructed from Equations (1.2) and (1.3).

$$Q = \begin{pmatrix} -\lambda & 0 & 0 & \dots & \dots \\ \lambda & -\lambda & 0 & 0 & \dots \\ \vdots & \lambda & \ddots & \ddots & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

The Poisson process has many properties that are useful in modelling situations, since this process can capture random processes in which only one event can take place at a time. For such a process, we may wish to know the time between events. Under the Poisson process, the waiting times between subsequent events are independent and identically distributed according to an exponential distribution with parameter λ [80]. This grants insight into the expected time to the next event, as well as characterising the waiting time with a simple and well-known distribution.

1.1.6 Infinitesimal generators

Whilst the infinitesimal transition rates describe how the probability distribution of the Markov chain changes with time, sometimes other functions of the process may be relevant.

The infinitesimal generator describes how the expected values of functions of a Markov process change in infinitesimal time intervals. For a function f acting on the process $\Sigma(t)$, the infinitesimal generator, $\mathcal{L}$, is defined as [130]

$$\mathcal{L}f(i) = \frac{d}{dt}\mathbb{E}[f(\Sigma(t))|\Sigma(0) = i] = \lim_{t\to 0} \frac{\mathbb{E}[f(\Sigma(t))|\Sigma(0) = i] - f(i)}{t}.$$

As an example, this can be used to obtain the infinitesimal transition rates by taking the function $f(i) = \mathbb{1}_{\{i=j\}}$. In this case, we obtain

$$\mathcal{L}f(i) = \lim_{t\to 0} \frac{P(\Sigma(t) = j|\Sigma(0) = i)}{t} = q_{ij}(t)$$

if $i \neq j$ and

$$\mathcal{L}f(i) = \lim_{t\to 0} \frac{P(\Sigma(t) = j|\Sigma(0) = j) - 1}{t} = -q_j(t)$$

if $i = j$. These are equivalent to the infinitesimal transition rates from an initial distribution i. Therefore, for continuous-time Markov chains, the infinitesimal generator, $\mathcal{L}$, can be taken as an alternative definition of the transition rate matrix Q. An advantage to using $\mathcal{L}$ is that this can be used as an operator on various functions f, so can be used to find various properties of the process. Additionally, the generator can be particularly important in the case where the state space is uncountable, since here a matrix representation is not applicable.

Two commonly used functions in Markov processes are the hitting probability and expected hitting time. The hitting probability of a state $A \in S$ is the probability that

the process eventually reaches state A, given that it started in some state i. Defining T^A as the time when the process first enters state A, then the hitting probability, when the process is initiated in state i, is given by

$$h_A(i) = P(T^A < \infty | \Sigma(0) = i).$$

Applying the generator, we obtain

$$\mathcal{L}h_A(i) = \frac{d}{dt}\mathbb{E}[h_A(\Sigma(t))|\Sigma(0) = i].$$

Given that the Markov process starts in state i, the expected value of the hitting probability does not change with time, and therefore this derivative must be equal to zero, giving

$$\mathcal{L}h_A(i) = 0.$$

If our initial state $i = A$, then the hitting probability is equal to 1, so we have $h_A(A) = 1$. The hitting probability can be found as the minimal non-negative solution to

$$\begin{cases} h_A(i) = 1 & \text{if } i = A, \\ \mathcal{L}h_A(i) = 0 & \text{if } i \neq A. \end{cases}$$

The expected hitting time is the expected time until the Markov process reaches a state A. The expected hitting time, starting from state i, is defined as

$$k_A(i) = \mathbb{E}[T^A|\Sigma(0) = i].$$

Applying the generator to the expected hitting time, we obtain

$$\mathcal{L}k_A(i) = \frac{d}{dt}\mathbb{E}[k_A(\Sigma(t))|\Sigma(0) = i].$$

The derivative can be calculated by

$$\frac{d}{dt}\mathbb{E}[k_A(\Sigma(t))|\Sigma(0) = i] = \lim_{h\to 0} \frac{\mathbb{E}[k_A(\Sigma(t+h))|\Sigma(0) = i] - \mathbb{E}[k_A(\Sigma(t))|\Sigma(0) = i]}{h}. \quad (1.7)$$

Since both of the expectations on the right-hand side condition on $\Sigma(0) = i$, the expected hitting time from 0 must be equal, which we will assume is equal to τ. The expected time

from $t+h$ therefore has to be $\tau-(t+h)$ and from t has to be $\tau-t$. The difference between these two is h, so Equation (1.7) becomes

$$\frac{d}{dt}\mathbb{E}[k_A(\Sigma(t))|\Sigma(0)=i] = \lim_{h\to 0}\frac{-h}{h} = -1.$$

Therefore,

$$\mathcal{L}k_A(i) = -1.$$

If our initial state $i = A$, then the expected hitting time is equal to 0, so we have $k_A(A) = 0$. The expected hitting time can be found as the minimal non-negative solution to

$$\begin{cases} k_A(i) = 0 & \text{if } i = A, \\ \mathcal{L}k_A(i) = -1 & \text{if } i \neq A. \end{cases}$$

1.1.7 Embedded discrete-time processes

In the dynamics of biological populations, the main properties of interest are often related to the long-term behaviour of the system. This can vary from steady-state dynamics across the state space [21, 30, 109] to the probability of reaching certain absorbing states [19, 102, 129], for example. Within a continuous-time Markov chain there exists an embedded discrete-time Markov chain that has the same long-term behaviour [94]. Therefore, this behaviour can be found by analysing either the continuous-time process or the embedded discrete-time process.

Consider a continuous-time Markov chain $\Sigma(t)$. If we track the sequence of states that $\Sigma(t)$ enters, which we denote by Σ_n, then the process Σ_n follows the same trajectory as $\Sigma(t)$ and the long-term behaviour must be the same. Σ_n is a discrete-time process, simply tracking the transitions of the continuous-time process $\Sigma(t)$. Since $\Sigma(t)$ satisfies the Markov property, Σ_n must also satisfy the Markov property, and is therefore a discrete-time Markov chain that is embedded into $\Sigma(t)$.

To define this discrete-time Markov chain, we need to find the one-step transition probability matrix, $\bar{P}$. For $\Sigma(t)$, if the process starts in a state i, the probability that it moves to state j in the next event is given by [94]

$$\frac{Q_{ji}}{\sum_{l\in S} Q_{li}}. \tag{1.8}$$

Therefore the probability that Y_n moves from state i to state j has to be given by Equation (1.8), and we have the transition probability

$$\bar{P}_{ij} = \frac{Q_{ji}}{\sum\limits_{l \in S} Q_{li}}.$$

This defines the transition probability matrix for the embedded discrete-time Markov chain.

1.1.8 Gillespie algorithm

The Gillespie algorithm is an individual-based simulation algorithm that is widely used to obtain exact realisations of the dynamics described by the master equation (Equation (1.6)). This algorithm randomly samples the next state to which the system jumps, and the time to the next jump, giving realisations of the state space with the relevant probability by considering how these probabilities are obtained from the master equation. The expected dynamics of the system are calculated by averaging over a statistically significant number of simulations.

Following [42], we denote by $f(\Sigma(t), s)$ the probability that given $\Sigma(t) = \sigma$ at time t the next event happens in the interval $[t+s, t+s+\delta)$. This probability is the intersection of two events: no jump happening in the time interval $[t, t+s)$ and a jump occurring in $[t+s, t+s+\delta)$. Since the process is Markovian these events are independent, and therefore $f(\Sigma(t), s) = g(\Sigma(t), s)W_{t+s}(\delta)$, where $g(\Sigma(t), s)$ is the probability that no jump occurs in $[t, t+s)$ and $W_{t+s}(\delta)$ is the probability of a jump occurring in $[t+s, t+s+\delta)$.

The first term, $g(\Sigma(t), s)$, can be split into the product of the probability that there is no event in $[t, t+s-\epsilon)$ and that there is no event in $[t+s-\epsilon, t+s)$, since the process is Markovian. That is

$$g(\Sigma(t), s) = g(\Sigma(t), s-\epsilon)g(\Sigma(t+s-\epsilon), \epsilon).$$

Since no event has occurred, $\Sigma(t+s-\epsilon) = \Sigma(t)$. From the definition of the transition probabilities (Equation (1.5)), we have

$$g(\Sigma(t+s-\epsilon), \epsilon) = 1 - \epsilon \sum_{n \neq \sigma} Q_{n\sigma} = 1 - \epsilon H(\sigma),$$

where $H(\sigma) = \sum_{n \neq \sigma} Q_{n\sigma}$. Therefore,

$$g(\Sigma(t), s) = g(\Sigma(t), s - \epsilon)[1 - \epsilon H(\sigma)].$$

This rearranges to

$$\frac{g(\Sigma(t), s) - g(\Sigma(t), s - \epsilon)}{\epsilon} = -g(\Sigma(t), s - \epsilon)H(\sigma),$$

which if we take the limit $\epsilon \to 0$ becomes

$$\frac{dg(\Sigma(t), s)}{ds} = -g(\Sigma(t), s)H(\sigma).$$

The initial condition is $g(\Sigma(t), 0) = 1$, since no event can occur instantaneously, and therefore we have the solution

$$g(\Sigma(t), s) = e^{-sH(\sigma)}.$$

Now we consider the probability that an event occurs in the interval $[t + s, t + s + \delta)$, $W_{t+s}(\delta)$. This is given by $H(\sigma)\delta$. Therefore, we obtain

$$f(\Sigma(t), s) = \delta H(\sigma) e^{-sH(\sigma)}.$$

Alternatively, we can write

$$f(\Sigma(t), s) = \int_{t+s}^{t+s+\delta} \tilde{f}(\tau) d\tau,$$

where $\tilde{f}(\tau)$ is the probability density function for the time until the next event τ. To find $\tilde{f}(s)$, we need to take the derivative of $f(\Sigma(t), s)$, which gives

$$\tilde{f}(s) = \lim_{\delta \to 0} \frac{\int_{t+s}^{t+s+\delta} \tilde{f}(\tau) d\tau - \int_{t+s}^{t+s} \tilde{f}(\tau) d\tau}{\delta} = \lim_{\delta \to 0} \frac{\delta H(\sigma) e^{-sH(\sigma)}}{\delta} = H(\sigma) e^{-sH(\sigma)}.$$

Therefore, the time to the next event follows an exponential distribution with parameter $H(\sigma) = \sum_{n \neq \sigma} Q_{n\sigma}$. When a transition occurs, we need to determine which transition took place. If we condition on an event taking place at time s, then the probability that the

transition is to state m is given by

$$F_m(\sigma) = \frac{Q_{m\sigma}}{H(\sigma)}. \tag{1.9}$$

Therefore, to simulate the stochastic process we need to sample two events. Firstly, sample the time until the next transition by sampling from the probability distribution $\tilde{f}$. Secondly, determine which transition took place by sampling from distribution F. Upon deciding which transition takes place, we can update the system accordingly, and then repeat the simulation algorithm from the new initial state.

That is, we follow the algorithm [13, 84, 54]

1. Generate initial conditions: $t = t_0$, $\sigma = \sigma(0)$.
2. Calculate the Q-matrix.
3. Determine the time until the next transition, δt, by sampling from an exponential distribution with parameter $H(\sigma) = \sum_{n \neq \sigma} Q_{n\sigma}$.
4. Determine which state the transition takes the system to, m, by sampling from the distribution $F(\sigma)$ (Equation (1.9)).
5. Update the system to the new state, $\sigma = m$, and the new time, $t = t + \delta t$.
6. Repeat from step (2).

The probability that this algorithm returns a specific trajectory is equal to the probability of that trajectory in the Markov process. In particular, the probability of the Markov process being in a certain state at a given time is approximately

$$P(\Sigma(t) = j) \approx \frac{\text{number of simulations in which } \Sigma(t) = j}{\text{total number of simulations}}.$$

Taking the number of simulations to infinity, this converges to the true probability. For a discrete-time Markov process, we can use this algorithm by setting the time until the next transition in step (3) to be $\delta t = 1$.

1.1.9 Quasi-stationary distributions

If a Markov process has any absorbing states, then eventually the process will only be distributed between these states in the limit of infinite time, provided an absorbing state is

accessible from every state. This can be useful in some circumstances, such as in evolutionary models where it will give the likelihood of a species taking over a population. However, in epidemiology for example, the knowledge that eventually a disease is guaranteed to die out gives little insight into the impact it may have on a population. In such cases, it is possible to instead investigate the dynamics of absorbing Markov chains conditioned on not having entered any absorbing states.

For a Markov chain $\Sigma(t)$, the conditional distribution is defined as the probability of being in each state, given that the system is not in an absorbing state [33]. That is, the conditional probability

$$\rho(\Sigma(t) = j) = P(\Sigma(t) = j | \Sigma(t) \notin A),$$

where A is the set of all absorbing states. By the definition of conditional probabilities, we have

$$\rho(\Sigma(t) = j) = \begin{cases} \frac{P(\Sigma(t)=j)}{1-P(\Sigma(t)\in A)} & \text{for } j \notin A \\ 0 & \text{for } j \in A. \end{cases}$$

The long-term behaviour of this conditional distribution can be solved to find the steady state, which is called the quasi-stationary distribution (QSD) of the Markov chain. This distribution is invariant under the conditional distribution, and is therefore similar to a true stationary distribution of the Markov chain. This has been proven to exist and be unique when the transient states form a single irreducible communicating class [33]. The study of quasi-stationary distributions has been applied to various systems, to understand how they behave when absorbing events are rare [31, 73, 85, 93, 137, 139, 147, 161].

1.1.10 Simulating the QSD

For Markov processes, we can simulate the dynamics of the master equation

$$\frac{d\mathbf{p}(t)}{dt} = Q\mathbf{p}(t),$$

using the Gillespie algorithm. Using the output of these stochastic simulations, it is possible to calculate the dynamics of the conditional distribution, ρ [58]. From the Gillespie algorithm,

$$P(\Sigma(t) = j) \approx \frac{|\{\Sigma(t) = j\}|}{|\{\Sigma(t) \in S\}|}$$

and

$$P(\Sigma(t) \in A) \approx \frac{|\{\Sigma(t) \in A\}|}{|\{\Sigma(t) \in S\}|},$$

where $\{\Sigma(t) \in X\}$ is the set of simulations in which the Markov process is in a state in the set X at time t, $|Z|$ is the cardinality of a set Z, and S is the full state space.

Therefore, the conditional probability that the Markov process is in state j at time t is approximated by

$$\rho(\Sigma(t) = j) \approx \frac{\frac{|\{\Sigma(t)=j\}|}{|\{\Sigma(t)\in S\}|}}{1 - \frac{|\{\Sigma(t)\in A\}|}{|\{\Sigma(t)\in S\}|}} = \frac{|\{\Sigma(t) = j\}|}{|\{\Sigma(t) \in S\}| - |\{\Sigma(t) \in A\}|}.$$

That is, to obtain the dynamics of the conditional distribution, only average over simulations that have not yet reached an absorbing state. Averaging over these simulations will give the conditional distribution, and we can continue the simulations in time until the conditional distribution attains a steady-state to find the QSD. An issue with such a method is that since this averages over simulations that are not extinct, as t increases the number of non-extinct simulations decreases. Therefore, fewer simulations are used as t increases, causing the results to become noisier. This means that to capture the QSD accurately a large number of simulations are sometimes required.

Therefore, alternative methods have been proposed to simulate the QSD. In the above case, we can run into issues when trying to find the QSD if the likelihood of reaching the ground state is high, since there are very few surviving simulations over which we can average. To account for this issue, instead of discounting any simulations that reach a ground state, one can instead return any simulations that reach the ground state into another state. This ensures that no simulation reaches the ground state, and therefore avoids the noise related issues. To capture the QSD accurately, the state the process is redistributed into has to be chosen carefully.

One approach is removing the probability of entering a ground state; i.e. if the system is in a state that can directly enter the ground state, remove the probability of this event [58]. This method redistributes any simulations that enter the ground state into their previous state, and has been found to provide a reasonable approximation to the QSD in some cases; however, it can lead to the states "next" to extinction being overly expressed. The method proposed by Oliveira and Dickman [35] instead redistributes the process into a randomly chosen state, with probability proportional to the frequency of that state over the history

of the simulation. Such a method captures the QSD exactly whilst avoiding the issue of noise dominating the results, and therefore can be a superior method when studying the QSD of stochastic processes that have a high probability of absorption.

1.2 Networks

Within this thesis, we focus on populations with an underlying structure that can be represented by a network (or graph). Many population dynamics models have focused on well-mixed populations, where all individuals are identical and have the same interactions, whereas in real populations there is often heterogeneity across the population, such as variety in which individuals interact with each other. This can be particularly important in evolution and epidemiology for example, since it means that not all individuals are competing with each other or able to infect each other. This type of structural heterogeneity can be captured using networks to describe which individuals are able to interact with each other [91, 102]. Here, the population dynamics considered determines the nature of the interaction.

When using well-mixed populations, it is natural to assume large population limits and use deterministic dynamics, since this can make the mathematics more tractable and yield closed-form solutions. However, as we discuss in the later sections of this introduction, taking such limits can miss important stochastic behaviour of the systems. Adding network structure makes the population finite, and therefore in this setting it is natural to study stochastic population dynamics.

We define population structure as a directed network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, ..., N\}$ is the set of all the nodes/vertices of the networks, which we can map via an arbitrary bijection to the set of all individuals (or sites) in the population, and $\mathcal{E}$ is the set of edges between nodes. If an edge $(i, j) \in \mathcal{E}$ exists, then node j can interact with node i. This interaction can either be static or change with time. For the work in this thesis we focus on static networks that are strongly connected; i.e. there exists a path between all nodes.

When considering connected networks, it is common to represent the network using an adjacency matrix G, which contains information on the connections between nodes. For a network $\mathcal{G}$, the adjacency matrix has dimension N, such that each row corresponds to a given node; i.e. row i describes the nodes that connect to node i. If there exists an edge from node j to node i, that is $(i, j) \in \mathcal{E}$, then $G_{ij} = 1$, otherwise $G_{ij} = 0$.

Work on network-structured populations often uses idealised networks, which have a

well-defined structure with a high level of symmetry. In this thesis, we primarily focus on random networks. Here, we define a few key examples of idealised networks and random network families that are used throughout the thesis. All example networks shown in the figures here are undirected, however directed variants of the illustrated networks can also be constructed.

1.2.1 Idealised networks

Idealised networks are popular because they have a high level of symmetry. This can facilitate efficient analysis of the population dynamics, which otherwise would not be possible due to the complexity of considering each node on the network independently. Within this thesis, we focus primarily on four types of idealised network: the complete network, the star, the circle, and the square lattice.

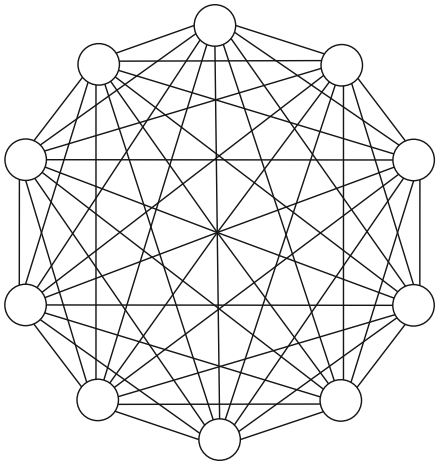

Figure 1.2: 10 node complete network

Complete network

The complete network (Figure 1.2) represents the classical well-mixed population. In this network, all nodes are connected to each other and interact. Since each node is identical with respect to its position on the network, this allows the dynamics to focus on the number

of nodes with certain properties and not be concerned with their position.

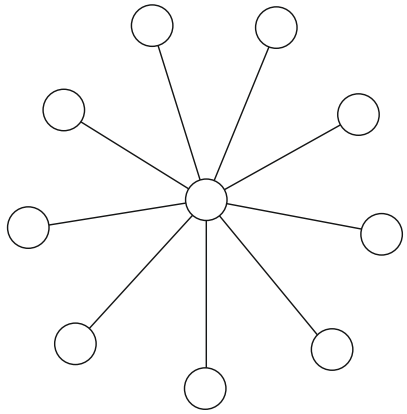

Figure 1.3: 10 node star

Star

On the star (Figure 1.3), there is one focal node which is connected to every other node in the population, and all the other nodes are connected only to the focal node. Throughout this thesis, we refer to the focal node as the central node, and the other nodes as leaf nodes. Since all leaf nodes are identical with respect to their position, we can simplify the dynamics by studying how the number of nodes with certain properties on leaf nodes change, in tandem with how the properties of the central node change.

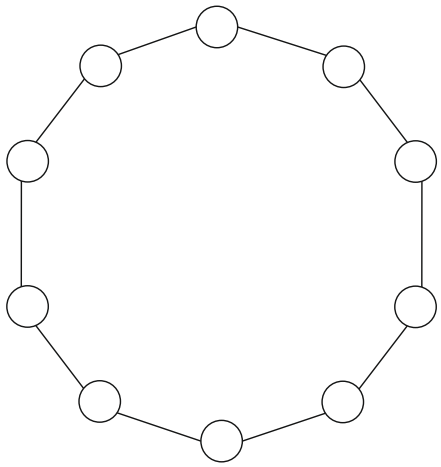

Figure 1.4: 10 node circle

Circle

In the circle (Figure 1.4), each node is connected to exactly 2 other nodes in such a way that the network forms a single closed loop. For certain dynamics, this network has sufficient symmetry to facilitate efficient analysis of the population dynamics. For example, in evolutionary graph theory (which we describe later) the dynamics ensure that on the circle all nodes with the same properties are grouped together, and updating can only take place at either end of this group, allowing the dynamics to be investigated by analysing how the size of this group changes.

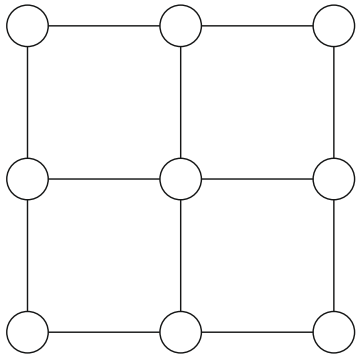

Figure 1.5: 9 node square lattice

Square lattice

On the square lattice (Figure 1.5), nodes occupy the grid points on a square grid, with the lines between grid points marking the edges between nodes. Two variations are considered. One variation is the wrapped lattice, in which the outer edges are connected to their parallel outer edges. This variant results in all nodes having exactly 4 neighbours. In this thesis, we focus on the fixed variant, where the outer edges are not connected to each other. This results in all interior nodes having 4 neighbours, the edge nodes having 3 neighbours and the corner nodes having 2 neighbours.

The square lattice is perhaps one of the most commonly considered population structures. This is because it highly structures the population, creating large distances between disjoint nodes on the network, whilst allowing for efficient analysis and approximation due to the high degree of symmetry.

1.2.2 Random networks

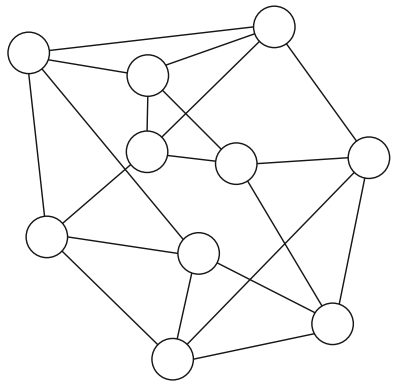

Figure 1.6: 10 node random k-regular network with average degree $k = 4$.

Random k-regular

Random k-regular networks (Figure 1.6) are designed such that edges are chosen at random but the finished network will have every node with degree equal to k. Therefore, all nodes behave similarly in terms of their interactions, but the structure will be random, since loops can be of varying sizes and there are varying levels of clustering across the structure. Allowing the population to be random can capture more realistic dynamics, but still facilitates efficient approximation of population dynamics, since each node can be assumed to behave reasonably similarly.

Erdős-Rényi

Erdős and Rényi (1959) [43] conducted a large volume of work on random networks. To construct a random network, they described the following algorithm:

1. Define a set $\mathcal{V}$ of N vertices.

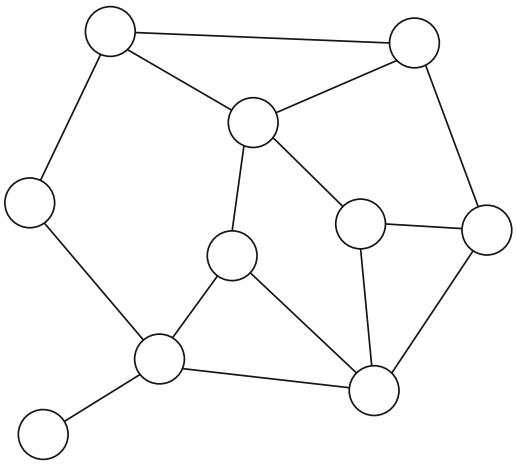

Figure 1.7: 10 node Erdős-Rényi random network with average degree 3.2.

2. For every pair of vertices assign a link between the two with probability q, independently for each pair.

Therefore, the expected number of edges for a randomly chosen vertex i is equal to $k = (N-1)q$. The resulting network is called the Erdős-Rényi random network (Figure 1.7). The degree distribution of the network is given by a binomial distribution with $n = N-1$ and $p = q$. Therefore, the probability that a node i has degree m is given by

$$P(deg(i) = m) = \binom{N-1}{m} q^m (1-q)^{n-1-m}.$$

The advantage of the Erdős-Rényi random network over the random k-regular network is that this type of population structure does not assume that all nodes have exactly the same number of interactions, whilst still ensuring reasonably similar behaviour. This can more accurately capture realistic populations, since although individuals are likely to behave similarly in the number of interactions they form, it is unlikely that there will be no heterogeneity.

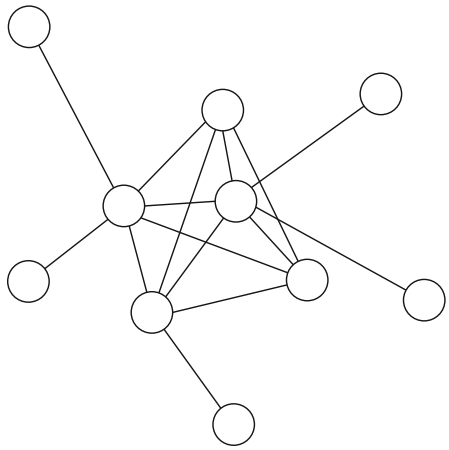

Figure 1.8: 10 node scale-free random network with average degree 4.

Scale-free

Although Erdős-Rényi random networks might capture desirable properties of some interaction networks, other networks, such as social interaction networks, have very different structure. In social networks there tends to be a small number of highly connected nodes and a large number of nodes with few connections. Such dynamics can be captured by a scale-free network (Figure 1.8), which has a degree distribution that tends asymptotically to a power law as the degree of interest increases. That is, for large k the fraction of nodes with degree k follows

$$P(k) \sim k^{-\gamma}.$$

There are multiple proposed algorithms for generating such a network. A common method is the Barabasi-Albert method [1]. This method uses preferential attachment to dictate how new links are made in the population; i.e. when a new node is added it is more likely to interact with high degree nodes than low degree nodes. The algorithm goes as follows.

1. Define an initial configuration of an m node complete network.

2. New nodes are than added to the population one at a time. These new nodes are then connected to $n < m$ existing nodes.

3. For each new link, the probability that the link goes to node i already in the network is given by $p_i = k_i / \sum_j k_j$, where k_i is the degree of node i. Here the sum is taken over nodes j that are already in the network.

This algorithm ensures that new links are most likely to form to highly connected nodes, and therefore we see a power law like distribution in the degree distribution of the network.

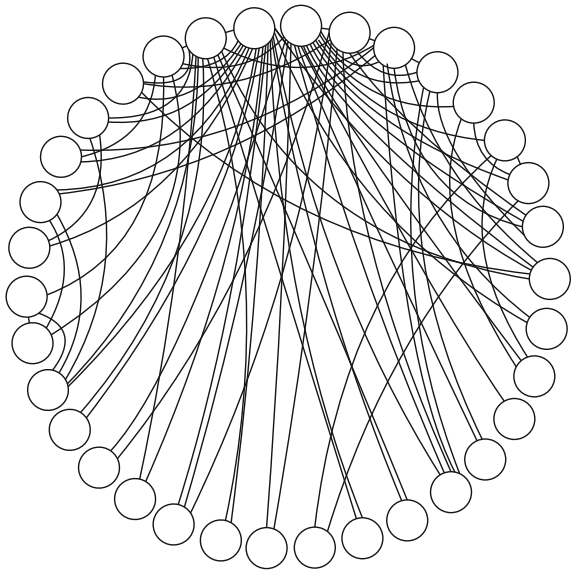

Figure 1.9: 34 node karate club network.

1.2.3 Zachary's karate club

It is important to consider how biological processes might spread in a real population. In this thesis, we consider a single real-world network when investigating our models. The network we consider is Zachary's karate club [201]. This network includes 34 individuals in a university karate club. This club was studied over a period of three years, and links between pairs who interacted outside the club were recorded. Using these links, an interaction network was produced, which is shown in Figure 1.9. This is an example of a real human social network, which might facilitate the spread of an infectious disease or evolution of social ideas.

In this network, it is observed that there are a few nodes who are highly connected, and then the number of connections rapidly drop off, with a large number of nodes who interact with only a few members of the network. Therefore, this network behaves similarly to a scale-free network (though is perhaps not as extreme as a true scale-free network). This agrees with the understanding that social networks tend to be scale-free in nature.

1.3 Stochastic evolutionary processes

An example where stochastic processes have been applied to population dynamics is in the study of evolution. Traditionally, evolutionary dynamics described how frequencies change in infinite populations using deterministic models; however, the methods used are not effective in finite populations. In such a population, the stochasticity of nature can have a significant impact on the dynamics, and therefore a stochastic approach can be used to capture this. In this section, we introduce the Moran process as an example of a stochastic evolutionary model. We then describe evolutionary graph theory, which is a generalisation of the Moran process to graph-structured populations.

1.3.1 The Moran process

The Moran process [116] was constructed to model stochastic fluctuations in population genetics through birth and death events. Since its conception, this model has been adapted and applied when studying the evolution of finite populations in which all individuals are assumed to have identical interactions, for example see [25]. This process assumes a finite homogeneous population consisting of two types of individuals, all of whom are equally likely to compete and interact. In each time step, a random individual is selected to produce an identical offspring and another random individual is selected to die. This maintains a constant population size, which we assume to be of size N, with one individual updating at each time step. This process assumes that the mutation rate is negligible on the timescale of the model, since reproduction produces identical offspring. If the two types have the same fitness, then all individuals are equally likely to be selected to reproduce, but in the more interesting case where one individual, say of type A, is fitter than the other, say of type B, then we assume that the fitter individual will be more likely to reproduce. In this case, an individual is selected to reproduce with probability proportional to its fitness, before selecting another individual to die at random. That is, the probability of selecting

an individual of type A to reproduce is

$$p_A = \frac{n_A f_A}{n_A f_A + (N - n_A) f_B},$$

where n_X is the number of type X individuals and f_X is the fitness of an individual of type X, for $X \in \{A, B\}$. Similarly the probability of selecting a type B individual to reproduce is

$$p_B = \frac{(N - n_A) f_B}{n_A f_A + (N - n_A) f_B} = 1 - p_B.$$

Defining the reproduction probability in such a manner ensures that the fitter type will be selected to reproduce with higher probability, and therefore is expected to be favoured in the evolutionary process. Upon reproducing, the offspring replaces an individual chosen at random. The dynamics of the process can be described by combining these two events. From a state with i type A and $N - i$ type B, it is possible to move to a state with $i + 1$ type A, $i - 1$ type A, or remain unchanged. To move from i to $i + 1$, a type A is selected for birth and a type B is selected for death. Denoting by $p_{i,i+1}$ the probability of moving from i to $i + 1$, we obtain

$$p_{i,i+1} = \frac{i f_A}{i f_A + (N - i) f_B} \frac{N - i}{N - 1}, \quad 1 \leq i \leq N - 1,$$

Similarly the probabilities of the other transitions are given by

$$\begin{aligned} p_{i,i-1} &= \frac{(N - i) f_B}{i f_A + (N - i) f_B} \frac{i}{N - 1}, \quad 1 \leq i \leq N - 1, \\ p_{i,i} &= 1 - p_{i,i+1} - p_{i,i-1}, \quad 1 \leq i \leq N - 1. \end{aligned}$$

This process is Markovian, since the updating only depends on the system state at a given time. Therefore, this process can be described by a discrete-time Markov chain with the transition probabilities described above. This process has two absorbing states, corresponding to the all A and all B states. Since the population is finite, the process will always reach one of the two absorbing states, which is called fixation of that type. When investigating this evolutionary process, there are two main avenues of interest; the probability that a certain type fixates and expected time to reach fixation. In evolution, we are interested in the probability that a rare mutant can arise and take over a resident population. This is given by the probability that, starting from a subset of mutant individuals, the process

eventually reaches the all mutant absorbing state. This is the hitting probability of the all mutant absorbing state and is called the fixation probability.

Definition 1. Fixation probability - Fixation probability is the probability that a given type eventually takes over the population. For a given type A, we refer to the fixation probability of type A as the probability that from an initial subset of the population, type A takes over the whole population.

Using the boundary conditions $p_{0,0} = p_{N,N} = 1$, Karlin and Taylor [80] derived expressions for the hitting probability for this type of Markov chain, which has subsequently been extended to finite populations [7, 179]. Using the formula of Karlin and Taylor [80], the fixation probability of $i \in \{1, ..., N\}$ initial type A individuals invading a type B population is given by

$$\rho^A(i) = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^{j} \frac{p_{k,k-1}}{p_{k,k+1}}}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \frac{p_{k,k-1}}{p_{k,k+1}}}. \tag{1.10}$$

Considering rare mutations, the fixation probability of a single initial type A is

$$\rho^A(1) = \frac{1}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \frac{p_{k,k-1}}{p_{k,k+1}}}. \tag{1.11}$$

In the original work of Moran, fitness was assumed to be constant (independent of the frequency of each type). We can therefore assume that the fitness of type A is $r > 1$ and the fitness of type B is normalised to 1 (this can be done arbitrarily by relabelling of type A and type B). In this case, the fixation probability of i initial mutant A individuals in a resident B population can be shown to be given by the Moran probability

$$\rho^A(i) = \frac{1 - r^{-i}}{1 - r^{-N}}. \tag{1.12}$$

What is particularly insightful about this simple equation, is that as the population size becomes infinite, $N \to \infty$, the fixation probability tends to $1 - r^{-i} < 1$. This shows that even in infinite populations, fixation of an advantageous mutant is not guaranteed. Therefore, a theoretically stronger individual may not succeed due to random fluctuations, such as potentially being selected for death before having the opportunity to reproduce.

This illustrates the stochastic nature of the evolutionary process, which is referred to as random drift. The effect of random drift can be particularly pronounced if the fitness of the type A individual is also equal to 1. Equation (1.12) only holds for $r \neq 1$. If $r = 1$ then we instead obtain [126]

$$\rho^A(1) = \frac{1}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \frac{1}{1}} = \frac{1}{N}.$$

This case is referred to as neutral drift, since the mutant individuals have no fitness advantage or disadvantage compared to the resident type, and therefore the only factor determining which type fixates is the random drift. The fixation probability of $1/N$ is intuitive, since all individuals are effectively identical, so each individual is equally likely to eventually take over the population and fixate.

In addition to fixation probability, we can consider the time to fixation. This is important because it can affect the validity of assuming the mutation rate is negligible. If time to fixation is fast, it might be reasonable to assume a subsequent mutation is unlikely to take place, but if fixation is very slow this assumption is unrealistic, and we are likely to encounter complications such as clonal interference [53]. This is particularly important when considering evolution in structured populations, and understanding fixation time in homogeneous populations is needed for comparison. Traulsen and Hauert [179], Antal et al. [6] and Broom et al. [19] have derived expressions for the time to fixation for arbitrary fitness functions for the well-mixed population.

1.3.2 Evolutionary graph theory

One of the limitations of traditional evolutionary models is the assumption of a well-mixed population. In real populations, it is not generally the case that all individuals interact and compete with each other, and most of the time individuals will only interact with a subset of the population. Therefore, there is some intrinsic population structure governing which individuals interact, and it is not natural to assume a well-mixed population. As an attempt to address this issue, Lieberman et al. [102] introduced evolutionary graph theory. This field is an adaptation of the Moran process to structured populations, where the relationship structure can be represented by a connected graph. Connections in the graph represent which individuals the offspring of a given individual can replace.

We assume a population that consists of a finite set $\mathcal{V}$ of individuals, which have been labelled via an arbitrary bijection to $\{1, 2, ..., N\}$, where $N = |\mathcal{V}|$. The interactions between individuals in the population can be described as a set of weighted, directed edges, $\mathcal{E}$, where $e_{ij} \in \mathcal{E}$ if and only if the offspring of individual j can replace individual i. The weight of an edge e_{ij} represents the probability that the offspring of j replaces i. We can therefore describe this population structure by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. This can be represented by a weighted matrix $W = (w_{ij})$. We can also define the adjacency matrix of the graph G, such that $G_{ij} = 1$ if $w_{ij} > 0$, and zero otherwise. This shows which individuals are able to interact with each other. The case we use throughout this thesis assumes that an offspring of a given individual is equally likely to replace any connected individual, $w_{ij} = a$ for all i with $e_{ij} \in \mathcal{E}$, and is guaranteed to replace an individual, such that $\sum_i w_{ij} = 1$. This implies that $w_{ij} = 1/k_j$, where k_j is the out degree of node j.

Evolutionary graph theory was introduced using a dynamics called the invasion process (or birth-death with selection on birth). This is similar to the Moran process, whereby first an individual is selected for birth proportional to its fitness and then the offspring replaces one of the connected individuals, proportional to the weight of these links (Figure 1.10). We consider two types of individuals, which are labelled A and B. After a mutant arises in the population, the mutation rate is assumed to be negligible on the time scale considered, so that no further mutations arise until either the mutants take over the whole population and "fixate", or the mutants die out and the residents remain.

When considering the complete graph (complete network), the invasion process is equivalent to the Moran process, since this graph represents the well-mixed population where all individuals interact. Therefore, the fixation probability of i mutants with constant fitness $r > 1$ is given by the Moran probability (Equation (1.12)).

In addition to showing that the complete graph yields the Moran probability, Lieberman et al. [102] proved a theorem, called the isothermal theorem, which finds a set of graphs that have fixation probability equivalent to the Moran probability. Under the invasion process, a graph has fixation probability equal to the Moran probability if and only if it is isothermal. An isothermal graph is a left stochastic graph, $\sum_{i=1}^{N} w_{ij} = 1$ for all $j \in \mathcal{V}$, which satisfies the property that the incoming weights are equivalent for each vertex, $\sum_{j=1}^{N} w_{ij} = 1$ for all $i \in \mathcal{V}$. This has been further generalised to graphs that are not left stochastic via the circulation theorem, which requires the total incoming weights to be equal to the total outgoing weights [102]. Since the conception of evolutionary graph theory, other evolutionary dynamics, which we call update rules, have been defined that

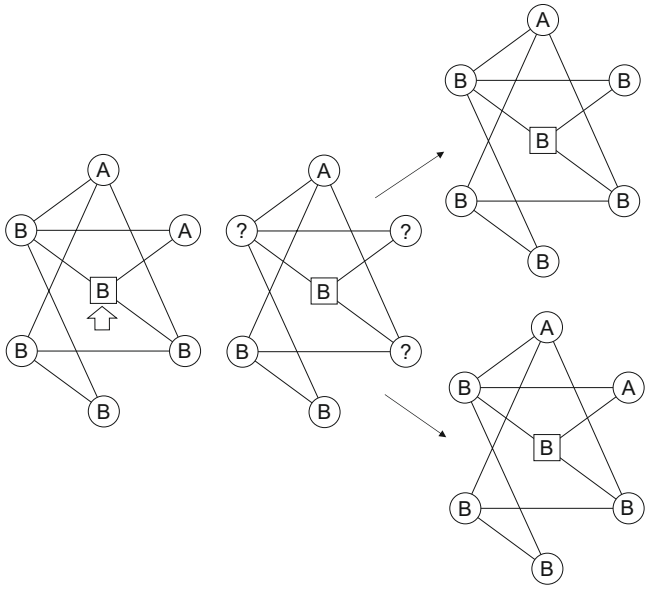

Figure 1.10: Update dynamics of the invasion process. A node is selected for birth proportional to its fitness, indicated by the white arrow. One of the neighbouring individuals, indicated by question marks is then replaced by an identical offspring of the selected individual. Therefore either one of the type B individuals is replaced by another type B, and the system is unchanged, or a type A is replaced by a type B, resulting in a new system state.

can be used within the evolutionary graph theory framework. For example, in death-birth with selection on birth dynamics [129], first death takes place at random before selecting a neighbouring individual of the dead node to reproduce, with probability proportional to their fitness. Within these other dynamics the circulation theorem does not generally hold, and Pattni et al. [140] have found conditions for a graph to be equivalent to Moran for each update rule.

Studying which graphs are equivalent to the complete graph (and the Moran process) has shown that it is a restrictive subset of all graphs. Therefore the question naturally arises, how do other graphs affect the fixation probability? Many graphs have been found that can amplify the fixation probability of advantageous mutants [102]. One example is the star graph, which is a strong amplifier of selection.

On the star graph, the symmetry allows the dynamics to be studied via four transition probabilities. That is, we consider the probability that a leaf node changes state, either to

type A or type B depending on the state of the central node, or the central node changes state, again to either type A or type B. Under the dynamics of the invasion process, these transition probabilities are given by [60]

$$\begin{aligned} p_{i,i+1}^{AA} &= \frac{f_A}{(i+1)f_A + (N-1-i)f_B}\frac{N-1-i}{N-1} \\ p_{i,i}^{AB} &= \frac{(N-1-i)f_B}{(i+1)f_A + (N-1-i)f_B} \\ p_{i,i-1}^{BB} &= \frac{f_B}{if_A + (N-i)f_B}\frac{N-1-i}{N-1} \\ p_{i,i}^{BA} &= \frac{if_A}{if_A + (N-i)f_B}. \end{aligned}$$

It can be shown that [60, 102], for the invasion process, the fixation probability of a single initial mutant for large N is approximately

$$\rho^A(i) \approx \frac{1-r^{-2i}}{1-r^{-2N}}. \tag{1.13}$$

Comparing this to the Moran probability, the relative fitness of an advantageous mutant on a star is amplified from r to r^2. This illustrates the significant impact that structure can have on selection. It has been discovered that under the invasion process, most random graphs will amplify selection. Under death-birth with selection on birth however, the star graph significantly suppresses selection, and it has been found that under these dynamics most random graphs will suppress selection [70].

In graphs like the star, which are considered to amplify selection, amplification can depend on the initial distribution mutants might take. If one assumes a uniform distribution, such that a mutant is equally like to arise in any node on the star, then the fixation probability is given by Equation (1.13). However, mutations generally arise from imperfect reproduction, when small genetic changes are introduced to the offspring. Therefore, mutation will appear proportionally to the number of times an individual is replaced. Under this assumption, since the central node of the star graph will be replaced the most often a mutant is most likely to arise here. However, under the dynamics of the invasion process the central node suppresses selection, since it is replaced if any of the leaf nodes are selected for birth. In this case, the star suppresses selection rather than amplifying. This has motivated research to find structures that, under the invasion process, can amplify

selection for various initial mutant distributions [141].

In addition to affecting the fixation probability, graph structure can also affect other properties of the evolutionary process, such as the time to fixation. Fixation time is an important quantity to study, since this could lead to clonal interference [53]. Also, understanding how long a mutant will take to spread through a population is useful in the study of cancer for example, where slowing down the spread of an advantageous mutant can give more time to treat the patients.

For the star graph, the mean fixation time has been studied in [60] for arbitrary evolutionary games under the invasion process dynamics. Other works have also investigated the time to fixation in the star graph and other complex graphs that amplify selection under the invasion process [68, 69, 177]. It has been observed that most graphs increase time to fixation relative to the complete graph, especially those that amplify selection, and it is hypothesised with strong circumstantial evidence that the complete graph has the shortest fixation time in all cases [114]. Therefore, although such population structures may amplify the strength of selection for advantageous mutants, the increase in time to fixation may adversely affect the mutant due to clonal interference.

1.3.3 Incorporating game theory

Another popular aspect of evolutionary graph theory is the incorporation of ideas from evolutionary game theory. In this case, fitness is no longer assumed constant, but instead depends on payoffs received from a series of interactions with connected individuals. The payoff received from interactions can be described using a payoff matrix

$$\begin{array}{c} \\ A \\ B \end{array} \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{array},$$

where a type A individual receives a from interacting with a type A and b from interacting with a type B, and a type B receives c from interacting with a type A and d from interacting with a type B. From this, we can derive functions for the payoff of each type. Denoting

by π_C^j the payoff of a strategy C individual in node j, we have

$$\pi_A^j = \frac{a\sum_{i=1}^{N} G_{ij}A_i + b\sum_{i=1}^{N} G_{ij}B_i}{\sum_{i=1}^{N} G_{ij}},$$
$$\pi_B^j = \frac{c\sum_{i=1}^{N} G_{ij}A_i + d\sum_{i=1}^{N} G_{ij}B_i}{\sum_{i=1}^{N} G_{ij}},$$

where $C_i = 1$ if the individual in node i is of type $C \in \{A, B\}$, and is zero otherwise. Recall that G is the adjacency matrix of the network, so $G_{ij} = 1$ if there is a link from j to i and zero otherwise. The fitness function of a type A individual occupying node j is defined as

$$f_A^j = 1 - w + w\pi_A^j, \tag{1.14}$$

and similarly the fitness of a B individual occupying node j is given by

$$f_B^j = 1 - w + w\pi_B^j, \tag{1.15}$$

where w is a parameter dictating the strength of the evolutionary game on selection. It has been shown that evolutionary games are highly sensitive to graph structure, and that, under certain dynamics, graph structure can promote the evolution of strategies that would lose in a well-mixed population [129]. For simple graphs, such as the star and complete graphs, we can analyse the dynamics by substituting the fitness functions (f_A^j and f_B^j) in place of f_A and f_B into the transition probabilities. However, due to the complexity of the games we do not arrive at simple closed form expressions like Equations (1.12) and (1.13). Instead we obtain an equation similar to Equation (1.10), which is relatively simple to solve numerically to calculate the fixation probability.

1.3.4 Approximating evolutionary graph theory

Within evolutionary graph theory, the classical examples involve highly idealised graphs, such as the complete graph, star, and circle. In general however, graphs do not have high levels of symmetries that allow these analytical expressions to be derived and we need to

study how each individual node in the population changes. Hindersin et al. [68] apply a numerical method to solve the discrete-time transition equation (1.1) for the evolutionary process. However, the complexity of this system scales exponentially and is not feasible for large populations. Therefore, finding other methods is important.

One method is to use individual-based stochastic simulations. This method generates sample trajectories of the evolutionary process, and by averaging over sufficient simulations can accurately capture the expected behaviour, such as the fixation probability and time, for any graph and evolutionary game. The issue with such a method is that, since for many graphs the time to fixation can be very long, it is computationally expensive to run a sufficient number of simulations to accurately capture the expected dynamics. To address this issue, in the constant fitness case threshold numbers of advantageous mutants have been derived, above which fixation is almost guaranteed [12]. This allows simulations to be ended before fixation without significantly changing the results. This work improves the efficiency of stochastic simulations, but is still time consuming and can make systematic analysis of multiple graphs challenging.

Instead of studying the exact dynamics, results have been derived by approximating the evolutionary process. One method to determine the relative strength of strategies playing an evolutionary game on a graph is taking the weak selection limit [2, 129, 203]. This method allows us to determine which strategy is preferred for various graphs with a closed form expression.

The weak selection limit is close to neutral selection, and is a biologically relevant limit [29, 128]. We will use this limit to analyse the dynamics of an arbitrary evolutionary game on the complete graph. Although exact results can be found on the complete graph, these are not a closed form result. By using the weak selection limit, approximate closed form conditions can be found that describe which strategy will be favoured, giving useful insight into the evolutionary process. The use of the weak selection limit has been applied to complex graphs to obtain analytical results [2, 129], though here we focus on the complete graph to give a simple illustration.

On the complete graph, all individuals of the same type are identical, and therefore we can drop the index j from the fitness and payoff functions. Therefore, equations 1.14 and 1.15 become

$$\begin{aligned} f_A &= 1 - w + w\pi_A \\ f_B &= 1 - w + w\pi_B, \end{aligned}$$

Assuming the invasion process, the transition probabilities when there are j type X individuals are given by

$$\begin{aligned} p_{j,j+1} &= \frac{jf_A}{jf_A + (N-j)f_B}\frac{N-j}{N} \\ p_{j,j-1} &= \frac{(N-j)f_B}{jf_A + (N-j)f_B}\frac{j}{N}, \end{aligned}$$

and therefore the ratio of transition probabilities is

$$\gamma_j = \frac{1 - w + w\pi_B}{1 - w + w\pi_A}.$$

To take the weak selection limit we consider $w \ll 1$, which yields (via Taylor approximation about $w = 0$)

$$\gamma_j \approx 1 - w(\pi_A - \pi_B).$$

The fixation probability for a single initial type A individual is given by Equation (1.11). Since the ratio of transition probabilities on the denominator is given by γ_j, we can approximate the product in Equation (1.11) by

$$\prod_{j=1}^{k} \gamma_j \approx 1 - w\sum_{j=1}^{k}(\pi_A - \pi_B),$$

since all the other terms involve higher orders of w, which we assume to be negligible.

To calculate which strategy is more likely to invade another, we find the ratio of fixation probabilities, which is given by

$$\frac{\rho^B(1)}{\rho^A(1)} = \left(1 - \frac{1 + \sum_{j=1}^{N-2}\prod_{k=1}^{j}\frac{p_{k,k-1}}{p_{k,k+1}}}{1 + \sum_{j=1}^{N-1}\prod_{k=1}^{j}\frac{p_{k,k-1}}{p_{k,k+1}}}\right)\frac{1 + \sum_{j=1}^{N-1}\prod_{k=1}^{j}\frac{p_{k,k-1}}{p_{k,k+1}}}{1} = \prod_{j=1}^{N-1}\frac{p_{k,k-1}}{p_{k,k+1}} = \prod_{j=1}^{N-1}\gamma_k.$$

Under the weak selection limit, this can be shown [179] to be

$$\frac{\rho^B(1)}{\rho^A(1)} = 1 - \frac{w}{2}[(a-b-c+d)(N-1) - a - b - c + 3d + (2b-2d)N].$$

Therefore, for the type A individual to be favoured ($\rho^A(1) > \rho^B(1)$) we require the second

term to be positive. For large populations, this term reduces to $(a+b-c-d)w/2$, so the condition for type A to be favoured is

$$0 < a + b - c - d. \tag{1.16}$$

We can also explore the fixation probability using the weak selection limit. It can be shown [179] that the fixation probability of a single initial A individual is given by

$$\rho^A(1) = \frac{1}{N} + \frac{w}{4N}\left[(a-b-c+d)\frac{2N-1}{3} - a - b - c + 3d + (2b-2d)N\right].$$

Using this formula, conditions under which the type X individual has a higher probability to fixate than it would under neutral selection can be obtained; i.e. $\rho^A(1) > 1/N$. This holds if the second term is positive. For large populations, this term reduces to $[(a-b-c+d)/3 + b - d]w/2$, so the condition for the type A individual to be stronger than a neutral individual is

$$\frac{a-b-c+d}{3} + b - d > 0. \tag{1.17}$$

Comparing these two conditions in their current form is not intuitive, so we wish to rearrange these so that they impose a condition on the same relationship. In the case where $a-b-c+d > 0$ we can rearrange Equation (1.16) to

$$\frac{d-b}{a-b-c+d} < \frac{1}{2}$$

and Equation (1.17) to

$$\frac{d-b}{a-b-c+d} < \frac{1}{3}.$$

Therefore, the conditions for A to be better than neutral selection is a subset of the conditions under which A out-competes B. In the region where A outperforms B but is not better than neutral, although A is more likely to invade B than vice versa, the probability that A invades B is lower than the probability that the progeny of a specific B individual would take over the population, and therefore A is unlikely to invade a resident population of B.

If $a-b-c+d < 0$ we can rearrange Equation (1.16) to

$$\frac{d-b}{a-b-c+d} > \frac{1}{2}$$

and Equation (1.17) to

$$\frac{d-b}{a-b-c+d} > \frac{1}{3}.$$

In this case, the conditions for A to be better than neutral selection contains the conditions under which A out-competes B. In the region where A is better than neutral but does not outperform B, a population of resident A individuals is likely to be invaded by a mutant B individual, and a resident A population is even more likely to be invaded by a mutant B. By taking the weak selection limit, we have been able to derive closed form conditions describing which strategy will be favoured in the evolutionary process.

1.4 Stochastic epidemic processes

Another example of stochastic population dynamics is in epidemiology. The standard approach to modelling in epidemiology is to stratify the host population into a finite number of compartments, which represent the status of the individuals in the population. For example, the infected compartment would contain all individuals who are infected with the pathogen (and hence able to transmit the pathogen). The traditional epidemic models of Kermack and McKendrick [86] use a deterministic dynamical system to describe the flow of density between the different compartments. These approaches ignore any stochasticity of the epidemic and any underlying structure within the compartments, such as further stratification by age or contact structure.

In reality, the spread of infectious diseases is a stochastic process, and hence in this thesis we consider stochastic epidemic models. Such models require the individual hosts within each compartment to be treated explicitly. Therefore, we cannot simply use the densities of the compartments but need to know the set of individuals within each compartment to evolve the system through time. We still consider these stochastic models as being compartmental in nature, however instead of being interested in the density within each compartment we focus on which compartment each individual belongs to. Following the work with evolutionary models, we will refer to the compartment that an individual is in as the state of that individual. That is, if an individual is in the infected compartment, I, we will refer to this individual as being in state I.

Since infection spreads via contact between infected individuals and susceptible individuals, it is natural to think about epidemics occurring on a contact network. This network describes which individuals are able to make contacts with each other. If one of

the individuals is infected, it has the potential to make infectious contacts to its connected neighbours. By using a contact network, it is possible to explore how the properties of the network influence the epidemic process.

1.4.1 Markovian network-based SIS model

The Markovian network-based SIS model is a stochastic model describing how pathogens spread on a host contact network [17, 61, 91, 136, 138, 139, 155]. In the model, individuals can flip back and forth between just two states; susceptible and infected. When an individual is infected, its neighbours in the network (or graph) that are susceptible are directly at risk of becoming infected. Infected individuals eventually return to the susceptible state and are again at risk of becoming infected. When all individuals are susceptible, they remain so for all future time and the pathogen is said to have died out. The all-susceptible state is thus an absorbing state. The model is thought to be a good mathematical representation for the propagation of sexually transmitted diseases and computer viruses [40].

To formally construct this model, consider a finite set $\mathcal{V}$ of individuals, who are labelled via an arbitrary bijection to $\{1, 2, \ldots, |\mathcal{V}|\}$. Let $N = |\mathcal{V}| < \infty$. Individuals can be in one of two states: susceptible, denoted by S, or infected, denoted by I. An individual $j \in \mathcal{V}$, while infected, makes infectious contacts to an individual $i \in \mathcal{V} \setminus \{j\}$ according to a Poisson process with rate $T_{ij} \geq 0$. If a susceptible individual $k \in \mathcal{V}$ receives an infectious contact, it immediately becomes infected for an exponentially distributed time period with mean $1/\gamma_k$, after which it immediately becomes susceptible again. We assume that the transmission matrix T is strongly connected; i.e. every individual is at risk of infection if at least one individual is infected.

This model is described by a continuous-time Markov chain $\{\Sigma(t) : t \geq 0\}$ with finite state space $\{S, I\}^N$, parameterised by an irreducible square matrix T with non-negative entries and a vector γ with positive entries. Let $\sigma_\alpha \in \{S, I\}^N$ denote a state of the population. Let $\Sigma_i(t)$ denote the status of individual i at time t, and for a given state σ_α, let $\sigma_{\alpha i}$ denote the status of individual i in that state.

From a given state σ_α, we can transition to a new state in which one individual has changed state; from S to I or I to S. If the status of individual i is changing, we denote the new state by $\sigma_\alpha^{i \to X}$, where $X \in \{S, I\}$ is the new status of i. The transition rates for the Markov chain are given in Table 1.1, where δ is the Kronecker delta.

Table 1.1: Transitions for the Markovian network-based SIS model

from	to	at rate
$\sigma_\alpha : \sigma_{\alpha i} = S$	$\sigma_\alpha^{i \to I}$	$\sum_{j \in \mathcal{V}} T_{ij} \delta_{I \sigma_{\alpha j}}$
$\sigma_\alpha : \sigma_{\alpha i} = I$	$\sigma_\alpha^{i \to S}$	γ_i

The time evolution of the Markov chain is captured by the master equation,

$$\frac{d\mathbf{p}(t)}{dt} = Q\mathbf{p}(t),$$

where $P_\alpha(t) = P(\Sigma(t) = \sigma_\alpha)$ is the probability that the system is in state σ_α at time $t \geq 0$, and Q is a matrix of transition rates (obtained from Table 1.1). Although this can be solved to determine the future behaviour, in many cases this is infeasible since the matrix Q grows exponentially with N.

One of the key properties of the model is the presence of a unique absorbing state corresponding to the disease-free state. This is the only absorbing state of the system, and therefore, for finite population size and parameters the trajectories will always enter this state and remain here. This is demonstrated in Figure 1.11, where it can be seen that the expected behaviour of the model tends towards extinction.

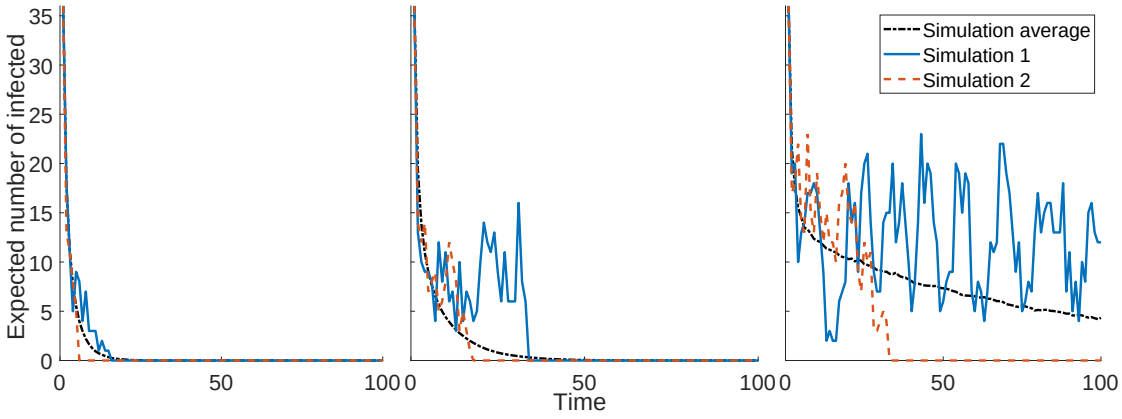

Figure 1.11: Output of stochastic simulations for the Markovian network-based SIS model on a 36 node square lattice. Simulation 1 is a simulation that took a relatively long time to reach extinction, simulation 2 is a simulation that went extinct relatively quickly, and the simulation average is the average across 10,000 simulations. The left-most plot has a low transmission rate, and this increases as the plots move from left to right.

This model has also been referred to as the contact process [103], where it was constructed as a type of interacting particle system. This is important, because evolutionary models such as the voter model have also been derived from interacting particle systems, suggesting that there are similarities between the two modelling frameworks.

1.4.2 Moment-closure approximations for network-based SIS

Due to the prohibitive computational cost of solving the master equation, approximation methods are useful. The heterogeneous mean-field and pair-approximation methods, which we briefly describe here, can be interpreted as approximating the expected behaviour of the stochastic model. For detailed derivations and analysis of these models see [91].

Under the heterogeneous mean-field model, we assume that: all individuals with the same degree can be treated identically, the status of neighbouring individuals are independent, the recovery rate is identical for all individuals ($\gamma_i = \gamma$ for all $i \in \mathcal{V}$), and whenever two nodes are connected the transmission rate is identical ($T_{ij} = \tau$ for all $i, j \in \mathcal{V}$ with $T_{ij} > 0$). Additionally, we assume that the network is undirected for simplicity, so that $T_{ij} = T_{ji}$. The rate of change in the expected number of susceptible and infected individ-

uals, stratified by the degree of the individual, is then approximated by [91]

$$\begin{aligned}[\dot{S_k}] &\approx -\tau \sum_{l\in\mathcal{M}} |C_{k,l}| \frac{[S_k]}{|C_k|}\frac{[I_l]}{|C_l|} + \gamma[I_k] \\ [\dot{I_k}] &\approx \tau \sum_{l\in\mathcal{M}} |C_{k,l}| \frac{[S_k]}{|C_k|}\frac{[I_l]}{|C_l|} - \gamma[I_k],\end{aligned}$$

where $[S_k]$ is the expected number of susceptible individuals of degree k at time t, $|C_k|$ is the number of degree k nodes, $|C_{k,l}|$ is the number of pairs involving a degree k node and a degree l node, and $\mathcal{M}$ is the set of unique degrees on the network. Above, and throughout, we use 'dot' notation for derivatives with respect to time. Whilst the assumption of neighbouring individuals being independent is unrealistic, the resulting model has low computational cost, and hence it is popular to study.

Instead of assuming statistical independence between individuals, models have been derived by writing down exact equations for the expected number of individuals and pairs:

$$\begin{aligned}[\dot{S_k}] =& \gamma[I_k] - \sum_{l\in\mathcal{M}} \tau[S_k I_l] \\ [\dot{I_k}] =& -\gamma[I_k] + \sum_{l\in\mathcal{M}} \tau[S_k I_l] \\ [\dot{S_k I_l}] =& \gamma([I_k I_l] - [S_k I_l]) + \tau(\sum_{m\in\mathcal{M}} [S_k S_l I_m] - \sum_{m\in\mathcal{M}} [I_m S_k I_l] - [S_k I_l]) \\ [\dot{S_k S_l}] =& \gamma([S_k I_l] + [I_k S_l]) - \tau(\sum_{m\in\mathcal{M}} [S_k S_l I_m] + \sum_{m\in\mathcal{M}} [I_m S_k S_l]) \\ [\dot{I_k I_l}] =& \tau([S_k I_l] + [I_k S_l] - 2\gamma[I_k I_l] + \tau(\sum_{m\in\mathcal{M}} [I_m S_k I_l] + \sum_{m\in\mathcal{M}} [I_k S_l I_m]),\end{aligned} \tag{1.18}$$

where $[A_k B_l]$ is the expected number of pairs at time t, between degree k and l individuals in states A and B respectively, and $[A_k B_l C_h]$ is the expected number of triples at time t, between degree k, l and h individuals, in states A, B and C respectively.

Solving this system exactly involves deriving a full hierarchy of equations describing triples and quads and so on [40], and therefore we wish to approximate this system by closing the hierarchy early. This can be done by expressing triples as some function of pairs and individuals. To approximate the triples, we analyse the number of edges starting from a susceptible node, following [40, 91]. The total number of SA edges (for $A \in \{S, I\}$) from a degree k node to a degree l node is $[S_k A_l]$. Since we have $[S_k]$ susceptible degree k

nodes, we have approximately $[S_k A_l]/(k[S_k])$ edges leading from a given susceptible degree k node to a given degree l node in state A. Therefore, for a chosen susceptible degree k node the probability that two neighbours, with degree l and m, are in states A and B is given by $[A_l S_k][S_k B_m]/k^2[S_k]^2$. We have $k(k-1)$ choices of the two neighbours, and $[S_k]$ choices of the susceptible node, and therefore we can approximate the expected number of triples $[A_l S_k B_m]$ as

$$[A_l S_k I_m] \approx \frac{k-1}{k} \frac{[A_l S_k][S_k I_m]}{[S_k]}.$$

This approximation makes the homogeneity assumption that the neighbours of susceptible degree k nodes are interchangeable and the states of pairs are independent. Using this expression, the system of equations (1.18) is closed at the level of pair terms, which allows the system to be solved with reasonably low computational cost.

These two models act at the population level, since they describe how the expected number of individuals with certain traits change. Following the motivation behind these models, node-level models have been developed that describe how the probability of individual nodes being infected change with time. Such models have been referred to as individual-based models [167, 168], node-level models [134], propagation models [91] or quenched mean-field [44, 107]. The advantage of such models over the population-level models is that we do not need to make any homogeneity assumptions about the underlying populations, and therefore properties such as clustering, directed edges and degree heterogeneity are naturally captured. The downside however is that the computational cost scales with at least the number of nodes.

Under Markovian network-based SIS, the dynamics of individual nodes are given by [167]

$$\begin{aligned} \langle \dot{S}_i \rangle &= -\sum_j T_{ij} \langle S_i I_j \rangle + \gamma_i \langle I_i \rangle, \\ \langle \dot{I}_i \rangle &= \sum_j T_{ij} \langle S_i I_j \rangle - \gamma_i \langle I_i \rangle, \end{aligned} \tag{1.19}$$

where $\langle A_i \rangle$ represents the probability $P(\Sigma_i(t) = A)$ with $A \in \{S, I\}$, and $\langle A_i B_j \rangle$ represents the probability $P(\Sigma_i(t) = A, \Sigma_j(t) = B)$ with $A, B \in \{S, I\}$.

This equation exactly describes the rate of change for individual nodes in terms of

pairs. Pairs of nodes are exactly described by

$$\begin{aligned}
\langle \dot{S_iI_j} \rangle &= \sum_k T_{jk}\langle S_iS_jI_k\rangle - \sum_k T_{ik}\langle I_kS_iI_j\rangle \\
&\quad -(T_{ij}+\gamma_j)\langle S_iI_j\rangle + \gamma_i\langle I_iI_j\rangle, \\
\langle \dot{S_iS_j} \rangle &= -\sum_k T_{jk}\langle S_iS_jI_k\rangle - \sum_k T_{ik}\langle I_kS_iS_j\rangle, \\
\langle \dot{I_iI_j} \rangle &= \sum_k T_{jk}\langle I_iS_jI_k\rangle + \sum_k T_{ik}\langle I_kS_iI_j\rangle - (\gamma_i+\gamma_j)\langle I_iI_j\rangle \\
&\quad +T_{ij}\langle S_iI_j\rangle + T_{ji}\langle I_iS_j\rangle,
\end{aligned} \tag{1.20}$$

where $\langle A_iB_jC_k\rangle$ represents the probability $P(\Sigma_i(t) = A, \Sigma_j(t) = B, \Sigma_k(t) = C)$ with $A, B, C \in \{S, I\}$. To solve this requires a hierarchy of equations up to full system size. Following similar logic to the population-level equations, this system can be approximated by making assumptions of statistical independence. Assuming that the states of individuals are independent, $\langle S_iI_j\rangle \approx \langle S_i\rangle\langle I_j\rangle$, we can close the hierarchy at the level of individuals. Alternatively, we can assume independence at the level of pairs. The natural assumption of statistical independence to apply to pairs is that, given three nodes in a line, if the state of the central node is known then the states of the outer two nodes are independent. For all triples in the system above, the central node in the configuration is always the centre node of a line between the two outer nodes. Therefore, if we consider the triple $\langle A_iB_jC_k\rangle$, this can be approximated as a function of lower order terms by using conditional probabilities and assuming statistical independence. By the definition of conditional probabilities, we obtain

$$\langle A_iB_jC_k\rangle = \langle A_iC_k|B_j\rangle\langle B_j\rangle.$$

Assuming that the states of nodes i and k are independent given the state of node j, this becomes

$$\langle A_iB_jC_k\rangle \approx \langle A_i|B_j\rangle\langle C_k|B_j\rangle\langle B_j\rangle = \frac{\langle A_iB_j\rangle\langle B_jC_k\rangle}{\langle B_j\rangle}, \tag{1.21}$$

which closes the hierarchy at the level of pairs. Other methods to approximate triples in terms of pairs and individuals have been proposed [82, 157, 167], however we do not consider them in this chapter.

The population-level methods described above can be derived rigorously from the node-

level methods [167]. In the exact case, we have

$$[A_k] = \sum_{j:k_j=k} \langle A_j \rangle$$

and

$$[A_k B_l] = \sum_{i:k_i=k} \sum_{j:k_j=l} \langle A_i B_j \rangle$$

where $A, B \in \{S, I\}$ and k_i is the degree of node i. Using this, the rate of change for the population-level terms can be derived. From this, we can also approximate the node-level quantities as

$$\langle A_i \rangle \approx \frac{[A_{k_i}]}{|C_{k_i}|}, \tag{1.22}$$

and

$$\langle A_i B_j \rangle \approx \frac{[A_{k_i} B_{k_j}]}{|C_{k_i,k_j}|}. \tag{1.23}$$

The models described here exhibit an epidemic threshold, above which the pathogen persists and below which the pathogen dies out (illustrated in Figure 1.12 for the node-level pair-based model). For the population-level models and individual-based node-level model, above these thresholds a unique, globally stable equilibrium exists [82, 83, 91, 96, 184]. For the node-level pair-based model, the disease-free solution has been shown to become unstable as the transmission rate increases [107], at which point we have shown that an endemic equilibrium solution exists (Appendix 5.A). Numerically, this endemic equilibrium appears to be unique and globally attracting, similar to the endemic solutions in the other models.

High above the epidemic threshold, the endemic equilibrium solutions of these models approximate the behaviour of the stochastic model for a long time, since the time to disease fade out is very long. However, as the transmission rate decreases, the endemic equilibrium does not capture the dynamics of the stochastic process. This is illustrated in Figure 1.12. Therefore, it is unclear how to relate these deterministic models to the stochastic process.

1.4.3 Other epidemic dynamics

The SIS model represents pathogens that do not grant host immunity upon recovery, hence infected individuals return to the susceptible compartment. Many pathogens however do

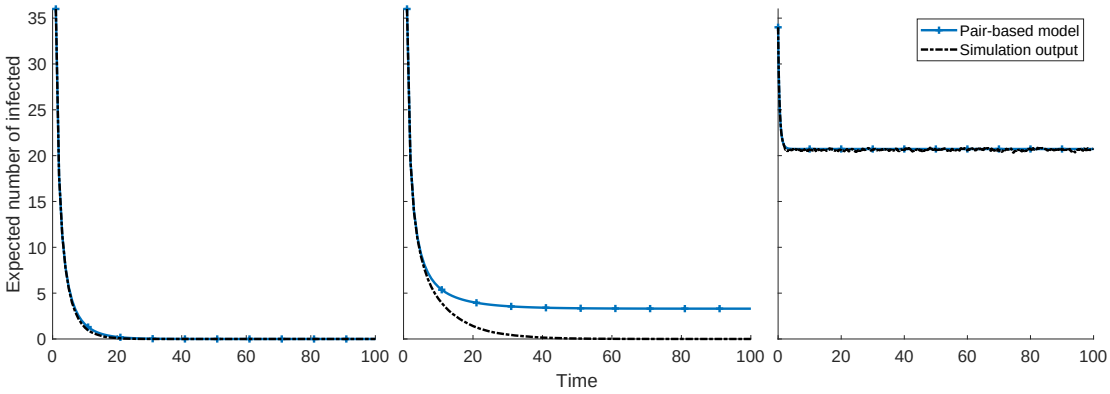

Figure 1.12: Comparing the standard pair-based model with the output of stochastic simulations on a 36 node square lattice. We plot the expected number of infected individuals against time for each of the methods. As the figures move from left to right the transmission rate increases. In the right-most figure, steady state-like behaviour is observed in the stochastic model, since the expected time to extinction is very long.

confer host immunity, which can either be lifelong or wane after some period of time. To capture such dynamics, other compartments can be introduced to represent immune individuals. The standard approach, based on the original epidemic model of Kermack and Mckendrick, is the SIR model [86]. In this model, there is a new compartment called removed, which contains individuals who have just left the infected compartment. These individuals are no longer susceptible to infection. To capture waning immunity, removed individuals can be allowed to transition back to the susceptible compartment, at which point they will be able to become infected again, resulting in the SIRS model [78].

The SIR model allows the complexities of immunity to be captured. Another important aspect of many pathogens is a latent period, where an infected individual may not be able to transmit the infection. This can be captured by adding another compartment called exposed, which contains individuals who have been exposed to the pathogen but are not yet infectious to other potential hosts. Adding this compartment to the SIR model results in the SEIR model [101].

1.4.4 Epidemiology of evolving pathogens

Understanding evolution is important for improving the future of healthcare, due to the presence of evolutionary processes in many healthcare challenges. For example, to understand cancer we need to understand how tumour cells mutate, compete and interact, which

is essentially an evolutionary process. Some insight has been gained into tackling cancer by using evolutionary game theory to predict how treatments may affect the tumour [202]. In this thesis however, we focus on infectious diseases.

With the discovery of antimicrobial compounds, many historically fatal diseases have become easy to treat in the last century. However, the future impact of such diseases is not so clear, since from the first widespread use of antimicrobials we have witnessed a steady rise of pathogens that are resistant to the treatments [195]. This is leading towards a tipping point, since the rate at which resistance is appearing vastly outweighs the rate at which new treatments are developed. Therefore, we are starting to see infections where the pathogen is resistant to all the recommended treatment options. This could lead to a public health crisis where we are no longer able to treat infections that we would previously consider simple. Understanding how to control the development of resistance is therefore of vital interest, and is a key area where the study of evolution overlaps with epidemiology.

Additionally, evolution plays a key role in the persistence of many viral infections, such as influenza and HIV. Influenza has multiple coexisting sub-types, which mutate frequently enabling them to become unrecognisable to the host immune system, and therefore immunity can wane, allowing the pathogen to persist. The presence of many sub-types also presents a huge issue in vaccine design, as modellers need to predict which strain will be most prevalent in order to design a vaccine each flu season. This involves studying the competition between the various coexisting strains, which again is a key aspect in evolutionary processes. This thesis explores both evolutionary and epidemic models and focusses on exploring the relations between the two types of model.

1.5 Outline

Here we provide an outline of the contents of this thesis. All the work presented is new and where work has been published or submitted for publication details are given.

In Chapter 2, we develop a biologically motivated general model for evolution in network structured populations. Evolutionary graph theory, despite being a popular model with mathematicians and theoreticians, has not been widely applied to real world populations, illustrating the need for a more biologically relevant model. Using the model that we develop, evolutionary graph theory can be derived as a special case, which illuminates the underpinning biological assumptions behind this framework. Additionally, our model can be used to obtain the SIS epidemic model as a special case, deriving a link between network-

based evolutionary processes and SIS. This work has been submitted for publication as a research article. For this chapter, I developed the research question, constructed and analysed the model, and wrote the manuscript, along with collaborators K. Pattni and K. Sharkey.

In Chapter 3, we investigate how bet-hedging strategies may evolve in network structured populations. Evolutionary bet-hedging describes how species adapt and change in variable environments, where the conditions experienced by individuals (and therefore their fitness) can change with time. In well-mixed populations, this is well understood, with various empirical and theoretical studies. However, the effect of population structure has not been thoroughly explored. Using the framework of evolutionary graph theory, we find evidence supporting the existence of within-generational bet-hedging. Such bet-hedging has been observed empirically, but widely dismissed by theoretical conclusions derived from well-mixed models. This work has been submitted for publication as a research article. For this chapter, I designed the project and, with the help of my supervisor K. Sharkey, performed the analysis and writing.

In Chapter 4, we approximate the dynamics of evolutionary graph theory, applying numerical approximation techniques from the epidemic and statistical physics literature. We first construct exact equations describing the behaviour of individual nodes in the population. From these node-level systems, we gain insight into existing approximation methods that have been shown to perform well in various scenarios. Additionally, we construct new methods that can be applied to study the transient dynamics of the evolutionary process as well as the fixation probabilities of different types. The proposed methods can be applied to a wide range of graphs, and grant insight into how the network properties affect the evolutionary dynamics from different initial conditions. The ability to explore different initial conditions is often not captured by similar approximations due to homogeneity assumptions, and can be crucial when studying which nodes on a graph present the largest invasion risk. This work has been published in the *Journal of Theoretical Biology* as a research article [134]. For this work, working with K. Sharkey I developed the research questions and the models. I analysed the models and wrote the manuscript with the help of my collaborators M. Broom and C. Hadjichrysanthou.

In Chapter 5, approximation methods are developed to capture the quasi-stationary distribution in the Markovian network-based SIS epidemic model. In epidemic models, the transient behaviour is the key area of interest, since knowing that eventually the disease will go extinct does not give any insight into the impact it may have on a population.

Therefore, it is important to study dynamics if extinction does not occur, which is given by the quasi-stationary distribution. Using moment-closure methods similar to Chapter 4, we develop an approximation method that can accurately capture the QSD for any network and parameters. This work forms a research article in preparation. For this chapter, I designed the project, conducted the analysis and wrote the manuscript, with assistance and feedback from collaborators R. Wilkinson, J. Miller, A. Loyinmi and K. Sharkey.

Chapter 2

Eco-Evolutionary Dynamics in Network Structured Populations

2.1 Introduction

The metapopulation model of Levins [99] was one of the first models to consider discrete spatial structure in the form of spatially separated sites that can be empty or occupied by a local population of infinite size whose individuals can migrate to other sites. This model has been extended in various ways, for example, a network of sites was considered in [64]. Metapopulation models are characterised by their extinction-colonisation dynamics, where local populations on occupied sites can go extinct and unoccupied sites become colonised by migrants. This means it is possible to have both occupied and unoccupied sites. These extinction-colonisation dynamics are ubiquitous for models based on similar discrete spatial structure. In structured epidemic models [66, 115], where sites are seen as hosts who can carry infectious disease, the susceptible-infected-susceptible (SIS) dynamics consist of colonisation events in the form of susceptible hosts getting infected and extinction events in the form of infected hosts recovering. In individual-based lattice models, such as the competing contact process [38], sites can accommodate at most one individual so extinction is a death event and colonisation is a birth event. A notably different model is the individual-based framework of evolutionary graph theory [102] (described in Section 1.3.2), where each site always has one individual present on it. Due to this restriction, this framework differs in terms of the dynamics used in the aforementioned models where empty sites are allowed. Dynamics that allow empty sites have been applied to biologically

relevant scenarios, for example, in the case of epidemic models: foot-and-mouth disease [83], sexually transmitted diseases [40] and influenza [166]. On the other hand, evolutionary graph theory is dominated by theoretical discussions about the importance of population structure on evolution [102, 20, 60]. To bridge the gap between these models, we need to study them within a single framework that will allow us to view their relationship in terms of the underlying biological assumptions made at the individual level.

In [22], evolution was described using the principals of birth, mutation, interaction and death acting at the level of individuals. Using this model it is possible to rigorously derive population-level models of evolution such as the classical models of [46] and [197], and the canonical equation of adaptive dynamics [36]. Furthermore, different biological insights can be obtained by changing the timescale of the individual-level processes, since the evolutionary dynamics predicted at the population level will be qualitatively different. In [23], continuous spatial structure was incorporated into the individual-based model of [22], such that individuals can move and interact within a certain spatial range. This was similarly used to provide biological insights into population-level models, for example, a large interaction range with weak migration and mutation results in a metapopulation [99], where individuals are organised in spatially isolated clusters.

In this chapter, we extend this individual-based model to a discrete spatial structure represented by a network, such that each node represents a site that can be occupied by multiple individuals. Building the framework in this way allows us to use the methods in [22] to consider different evolutionary models by changing the timescale of individual-level processes. In the case where mutation rates tend to zero, we are essentially considering only the ecological dynamics, and as the mutation rate increases we obtain eco-evolutionary dynamics. In the latter case, we can then consider where ecological and evolutionary processes happen at similar timescales which, for example, is the case in RNA viruses [57]. Our framework would be particularly useful in cases where structure plays an important role, where examples include the compartmental nature of lungs that allows antibiotic resistance to evolve in cystic fibrosis patients [11, 131] and tumour evolution where structure promotes diversity [124].

This chapter is structured as follows. Section 2.2 describes our framework to model evolution in a network-structured population with eco-evolutionary dynamics based on individual-level processes as in [22]. Section 2.3 gives a result showing that ecological dynamics can be suppressed in our framework's eco-evolutionary dynamics by using a negative ecological feedback loop. In Section 2.4, we use our framework to construct a spatial

birth and death model with ecologically motivated birth and death rates, which includes the SIS epidemic model [66] as a special case. We then apply the result in Section 2.3 to this model to derive evolutionary graph theory dynamics. Section 2.5 studies the long-term behaviour of a mutant invading a resident population in the spatial birth and death model in the cases without and with clonal interference. In particular, we look at the fixation probability. Without clonal interference, it is shown that adding intra-site competition and removing natural death increases the fixation probability of a mutant with an advantageous birth rate. With clonal interference, we consider the fixation probability of a mutant before another advantageous mutant arises. In this case, increasing the mutation rate reduces the amount of time available for a mutant to fixate before interference and, therefore, the circulation theorem from evolutionary graph theory no longer holds.

2.2 Evolution modelling framework with network structure and eco-evolutionary dynamics

We consider a population in which individuals are distributed over a finite number of connected sites. Individuals and sites represent different things depending on the modelling context. Examples can be found in the metapopulation and epidemiology literature such as the fragmented habitat of fritillary butterflies [65] and farms housing livestock infected with foot and mouth disease [108]. The sites are assumed to be arranged on a network such that individuals can spread to a connected site only. Examples of natural and artificial networks where the spread of individuals is restricted to nearest neighbours include email networks spreading computer viruses [123] and livestock movement networks [90].

The framework describes a birth and death process. This means that the population is updated in continuous time through either a birth or death event that respectively increases or decreases the population size by one. In [22] these events are described at the individual level incorporating interaction and mutation. We do this but also take into account the network structure. For birth events, individuals are assumed to reproduce asexually giving rise to an offspring that is of identical type when there is no mutation or of a different type when there is mutation. Individuals spread upon birth such that offspring can be placed onto a connected site where they mature immediately and remain until death. Examples of where this type of spreading dynamics can be used include modelling dispersal in plants [47], spread of social behaviour like alcoholism [159] and spread of infectious disease in

epidemics. For death events, it is assumed that individuals free up any space that they previously occupied. Deaths and births are assumed to be independent events allowing the population size and density to fluctuate.

Mutation allows the introduction of a continuous number of new types into the population. In this case, when the evolution of a population is studied over a large period of time multiple different types can appear that could potentially result in clonal interference [53], where two or more adaptive mutations are in competition with one another. This also allows consideration of a richer adaptive landscape. Interaction between individuals on the same site and potentially different sites as well can affect birth and/or death. In particular, interaction allows the consideration of frequency dependent selection over the adaptive landscape through the use of evolutionary game theory [109].

The mathematical description of the framework is as follows. We consider N distinct sites that are connected to each other in a network. If two sites are connected, then individuals can spread from one to the other. This network is represented by a connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of sites and $\mathcal{E}$ is the set of weighted edges between sites. If there exists an edge between two sites, then individuals can place their offspring in the connected site. This can also be represented by a matrix W with entries $W_{mn} \geq 0$, such that site n is connected to site m (and therefore individuals can spread offspring from n to m) if $W_{mn} > 0$. Each site can be occupied by multiple individuals. Each individual i is characterised by a separate position and trait such that $i = (U_i, X_i)$ where the phenotype is given by $U_i \in \mathcal{U} \subset \mathbb{R}^l$, and the position is given by $X_i \in \mathcal{X} = \{1, 2, \ldots, N\}$. This way of characterising individuals is taken from [23], but here $\mathcal{X}$ is a discrete set. The state of the population at a given point in time is given by the multiset $\mathcal{I}$ containing elements $i = (U_i, X_i)$. Since $\mathcal{I}$ is a multiset, if both $i = (u, x)$ and $j = (u, x)$, and $i, j \in \mathcal{I}$, then there are at least two copies of (u, x) in $\mathcal{I}$. We define the multiset $\mathcal{I}_n = \{i \in \mathcal{I} : X_i = n\}$ to represent the individuals present in site n; it therefore follows that $\mathcal{I}_n \subseteq \mathcal{I}$. As in [22], the individual-level processes follow a Poisson process, but in our case the network structure W can have an impact. The death rate of individual $i \in \mathcal{I}$ is given by $d(i, \mathcal{I}, W)$. The birth rate of individual $i \in \mathcal{I}$ when their offspring is spread to site x is given by $b(i, x, \mathcal{I}, W)$. The probability that an offspring of individual i carries a mutation is $\mu(i)$. The probability that individual i gives birth to an offspring with trait w is given by $M(U_i, w)$ such that $M(U_i, w) = 0$ if $w \notin \mathcal{U}$.

Putting this together gives a model of population evolution described by a continuous-time Markov process, which we will denote by $\Sigma(t)$. Let $\mathcal{I}$ be the state of the population at

time t. The infinitesimal dynamics of the population after t is described by the generator $\mathcal{L}$ that acts on real bounded functions ϕ as follows

$$\begin{aligned}\mathcal{L}\phi(\mathcal{I}) = &\sum_{i\in\mathcal{I}}\sum_{n\in\mathcal{X}}[1-\mu(i)]b(i,n,\mathcal{I},W)[\phi(\mathcal{I}\cup\{(U_i,n)\})-\phi(\mathcal{I})]\\ &+\sum_{i\in\mathcal{I}}\sum_{n\in\mathcal{X}}\mu(i)b(i,n,\mathcal{I},W)\int_{\mathbb{R}^l}[\phi(\mathcal{I}\cup\{(w,n)\}-\phi(\mathcal{I})]M(U_i,w)dw\\ &+\sum_{i\in\mathcal{I}}d(i,\mathcal{I},W)[\phi(\mathcal{I}\setminus\{i\})-\phi(\mathcal{I})].\end{aligned} \tag{2.1}$$

The event described by the first line is an offspring born with no mutation, the second line is an offspring born with a mutation and the last line is an individual dies.

When studying the evolution of a population, we are interested in eventually reaching some population state $\mathcal{A}$ from an initial state $\mathcal{I}$. If we define $T^{\mathcal{A}}$ as the first time the Markov process reaches $\mathcal{A}$, then we can use this to investigate whether the process reaches $\mathcal{A}$ or not. The hitting probability of $\mathcal{A}$ from a state $\mathcal{I}$ is the probability that $T^{\mathcal{A}}$ is finite, given that the Markov process starts in state $\mathcal{I}$; i.e. $P(T^{\mathcal{A}}<\infty|\Sigma(0)=\mathcal{I})$. The hitting time is the expected time to reach $\mathcal{A}$ from $\mathcal{I}$; i.e. $\mathbb{E}[T^{\mathcal{A}}|\Sigma(0)=\mathcal{I}]$. From the infinitesimal generator, the hitting probability, denoted $h_{\mathcal{A}}(\mathcal{I})$, is given by solving

$$\begin{cases}\mathcal{L}h_{\mathcal{A}}(\mathcal{I})=0,\\ h_{\mathcal{A}}(\mathcal{A})=1.\end{cases} \tag{2.2}$$

The generator can also be used to find the hitting time, denoted $k_{\mathcal{A}}(\mathcal{I})$, by solving

$$\begin{cases}\mathcal{L}k_{\mathcal{A}}(\mathcal{I})=-1,\\ k_{\mathcal{A}}(\mathcal{A})=0.\end{cases}$$

The derivation of the hitting probability and time are given in Section 1.1.6.

2.3 Suppressing ecological dynamics in eco-evolutionary dynamics

Here we show that we can suppress ecological dynamics in the eco-evolutionary dynamics proposed, leaving evolutionary dynamics that are based on ecologically motivated assump-

tions. In models that only consider evolutionary dynamics, such as the Moran process [116] and evolutionary graph theory [102], these underlying ecological assumptions are lost. This is because their evolutionary dynamics are directly defined from the assumption of fixed population size and density, rather than treating it as a consequence of suppressing ecological dynamics.

In ecology, the carrying capacity describes the maximum population size that can be sustained in a given environment. This depends on the composition of the population, since different types may have different carrying capacities. When ecological dynamics are suppressed however, the carrying capacity does not depend upon the composition of the population. To achieve this behaviour, we will create a negative ecological feedback loop that balances out opposing ecological forces pushing the system toward an equilibrium. For example, there is negative feedback between predators and their prey where an increase in predators leads to decrease in prey and vice versa [15]. In our framework, the ecological forces that result in a birth oppose those that result in a death. We therefore balance out these forces such that a population converges to a given size regardless of its composition.

The equilibrium state we consider is a population of size N with each site having density 1. A negative ecological feedback loop around this state is created by modifying the birth and death rates. The Heaviside step function

$$H_m[n] = \begin{cases} 0 & n < m, \\ 1 & n \geq m \end{cases}$$

is used to identify the conditions required for this negative ecological feedback loop to act. The modified death rate of individual i amplifies its death rate by c if present on a site with multiple occupancy but otherwise has no effect; that is,

$$D(c, i, \mathcal{I}, W) = c^{H_2[|\mathcal{I}_{X_i}|]} d(i, \mathcal{I}, W) \quad c \geq 1.$$

Similarly, the modified birth rate of individual i amplifies its birth rate onto site n by c if site n is empty but otherwise has no effect; that is,

$$B(c, i, n, \mathcal{I}, W) = c^{H_0[-|\mathcal{I}_n|]} b(i, n, \mathcal{I}, W) \quad c \geq 1.$$

The infinitesimal generator for the modified birth and death rates, denoted $\mathcal{L}_c$, is given by Equation (2.1) but with b replaced by B and d replaced by D. The parameter c controls

the strength of the negative ecological feedback loop's effect. For $c = 1$, there is no effect. For $c > 1$, there is an effect making it more likely that individuals sharing a site will die and that offspring are placed onto empty sites. In the limit $c \to \infty$, there is maximum effect. In this case, ecological dynamics are suppressed resulting in fixed population size and density.

When ecological dynamics are suppressed, the system updates through a replacement event where a birth and a death are coupled. This is formally shown by considering the hitting probability. Using the generator $\mathcal{L}_c$, the hitting probability in the limit as $c \to \infty$ of the eco-evolutionary dynamics can be shown (Appendix 2.A) to reduce to

$$\begin{aligned} h_{\mathcal{A}}(\mathcal{I}) =& \frac{1}{\lambda_{\mathcal{I}}} \sum_{i \in \mathcal{I}} \Bigg[\sum_{j \in \mathcal{I}} r(i, j, U_i, \mathcal{I}, W)[1 - \mu(i)] h_{\mathcal{A}}(\mathcal{I} \cup \{(U_i, X_j)\} \setminus \{j\}) \\ &+ \int_{\mathbb{R}^l} r(i, j, w, \mathcal{I}, W) \mu(i) h_{\mathcal{A}}(\mathcal{I} \cup \{(w, X_j)\} \setminus \{j\}) M(U_i, w) dw \Bigg] \end{aligned} \tag{2.3}$$

where $\lambda_{\mathcal{I}}$ is the rate of leaving state $\mathcal{I}$ and $r(i, j, u, \mathcal{I}, W)$ is the rate at which individual i's offspring of type u replaces individual j in state $\mathcal{I}$. This shows that in the limiting dynamics we have derived, the population is updated through replacement events that happen with rate r. Dropping W for brevity, it is shown in Appendix 2.A that the replacement rate r for Equation (2.3) is given by

$$r(i, j, u, \mathcal{I}) = b(i, X_j, \mathcal{I}) \frac{d(j, \mathcal{K}(u, j))}{\sum_{k \in \mathcal{K}(u,j)_{X_j}} d(k, \mathcal{K}(u, j))} + d(j, \mathcal{I}) \frac{b(i, X_j, \mathcal{I} \setminus \{j\})}{\sum_{k \in \mathcal{I} \setminus \{j\}} b(k, X_i, \mathcal{I} \setminus \{j\})} \tag{2.4}$$

where $\mathcal{K}(u, j) = \mathcal{I} \cup \{(u, X_j)\}$, and $\mathcal{K}(u, j)_{X_j}$ represents the individuals on site X_j in this state. We can see that there is both a birth-death (BD) and death-birth (DB) component in r. The first term is a BD component where individual i first gives birth to an offspring that is placed onto site X_j who then replaces individual j. The second term is a DB component where individual j dies first and then individual i gives birth to an offspring that is placed onto site X_j, hence replacing individual j.

2.4 Framework Application I: Deriving evolutionary graph theory dynamics from a spatial birth and death process

In this section, we construct a model from our framework by using ecologically motivated birth and death rates, which we will refer to as the spatial birth and death model (SBD). This model contains the SIS epidemic model [66] and competing contact process [38] as special cases. By applying the result from Section 2.3, this model gives evolutionary dynamics based on birth and death rates, which we refer to as SBD evolutionary dynamics. These evolutionary dynamics will be compared to those of evolutionary graph theory that are based on fitness, a measure of reproductive success. If they are equivalent, fitness can be interpreted in terms of birth and death rates, uncovering hidden assumptions and providing biological insight into evolutionary graph theory dynamics.

The ecological dynamics for SBD use density-dependent regulation of population size based on [77]. Individuals on the same site compete for survival through pairwise interactions resulting in the death of an individual. This competition has negative feedback such that increasing population size results in increased competition and vice versa. For individual i, let δ_{U_i} be the natural death rate and γ_{U_i,U_j} be the death rate due to competition with individual j. It is specified in [77] that the inverse of γ can be interpreted as the payoff in terms of evolutionary games [109]. That is, a larger payoff is received when γ is lower. The death rate is then given by

$$d(i,\mathcal{I},W) = \delta_{U_i} + \sum_{j\in\mathcal{I}_{X_i}\setminus\{i\}} \gamma_{U_i,U_j},$$

where self-interactions have been discounted. It is assumed that $\gamma_{u,v} > 0\ \forall u,v \in \mathcal{U}$ to ensure negative feedback. The birth rate is given by

$$b(i,n,\mathcal{I},W) = s^{|\mathcal{I}_n|}\beta_{U_i}W_{nX_i} \qquad s\in[0,1].$$

The birth rate of individual i is β_{U_i}. It is weighted by W_{nX_i} to capture the network effect of individual i's position when placing its offspring in site n. We added $s^{|\mathcal{I}_n|}$ to capture the ability of an offspring to survive when invading site n depending on its occupancy. For $0 < s < 1$, there is negative feedback such that survival decreases as occupancy increases and vice versa. For $s = 0$, the convention that $0^0 = 1$ is used implying that offspring cannot invade and only survive on vacant sites. For $s = 1$, offspring always survive when

invading.

SBD forms a basis for the susceptible-infected-susceptible (SIS) epidemic model [66] and its various extensions. The SIS model captures the ecological dynamics of an infection as it spreads between hosts. A host can be infected (I) or susceptible (S) and is represented by a node in a network. Infection can only spread from an infected to a susceptible. Becoming infected is therefore proportional to the number of infected neighbours. Infected individuals recover and become susceptible independent of their neighbours. SIS dynamics generally consider a single pathogen type, but multi-species SIS-type dynamics are obtained from SBD as follows. Each site is a host, with a vacant site representing S and an occupied site representing I. The presence of individual i on a site indicates having infection U_i, i.e. the trait of individual i. The death rate represents recovery from infection and can be defined as follows

$$d(i, \mathcal{I}, W) = \delta_{U_i}.$$

We have set $\gamma_{u,v} = 0$ for all $u, v \in \mathcal{U}$. Recovery from infection U_i happens with rate δ_{U_i}. The birth rate represents spread of infection and can be defined as follows

$$b(i, n, \mathcal{I}, W) = 0^{|\mathcal{I}_n|}\beta_{U_i} W_{nX_i}. \tag{2.5}$$

We have set $s = 0$ to restrict spread of infection to S (vacant sites) only. Infection U_i spreads with rate β_{U_i} and is weighted by W_{nX_i} to capture the effect of network structure. Constructing the SIS model using SBD allows us to consider extensions that have eco-evolutionary dynamics. A straightforward extension is the competing contact process [38]. This uses SIS dynamics to study inter-host competition where two different infections are competing to occupy hosts. Other extensions, such as [16], allow hosts to carry more than one infection. Here, there is intra-host competition where infections compete within a host. In our setting, this can be achieved when $s > 0$. Therefore, SBD allows us to consider a combination of inter and intra-host competition between infections.

In evolutionary dynamics, a model of interest is evolutionary graph theory. We wish to investigate whether such a model can be obtained from the SBD eco-evolutionary dynamics. To do this, we first apply the result from Section 2.3 to obtain SBD evolutionary dynamics.

In this case, the replacement rate (Equation 2.4) is given by

$$r(i,j,u,\mathcal{I},W) = s\beta_{U_i}W_{X_jX_i}\frac{\delta_{U_j}+\gamma_{U_j,u}}{\delta_{U_j}+\gamma_{U_j,u}+\delta_u+\gamma_{u,U_j}} + \delta_{U_j}\frac{\beta_{U_i}W_{X_jX_i}}{\sum\limits_{k\in\mathcal{I}\setminus\{j\}}\beta_{U_k}W_{X_jX_k}}.$$

The exponent of s is 1 in the BD component as every site has one individual in this case. The hitting probability in SBD, denoted $h_{\mathcal{A}}^{\mathrm{SBD}}(\mathcal{I})$, is given by substituting this replacement rate into Equation (2.3). On the other hand, the hitting probability in evolutionary graph theory, denoted $h_{\mathcal{A}}^{\mathrm{EGT}}(\mathcal{I})$, is obtained by solving Equation (2.2) using an infinitesimal generator for evolutionary graph theory that we define as follows

$$\begin{aligned}\mathcal{L}^{\mathrm{EGT}}\phi(\mathcal{I}) =& \sum_{i\in\mathcal{I}}\Bigg[\sum_{j\in\mathcal{I}}[1-\mu(i)]R(i,j,U_i,\mathcal{I},W)[\phi(\mathcal{I}\cup\{(U_i,X_j)\}\setminus\{j\})-\phi(\mathcal{I})] \\ &+\mu(i)\int_{\mathbb{R}^l}R(i,j,w,\mathcal{I},W)[\phi(\mathcal{I}\cup\{(w,X_j)\}\setminus\{j\})-\phi(\mathcal{I})]M(U_i,w)dw\Bigg]\end{aligned}$$

where R is the replacement rate in evolutionary graph theory dynamics. This generator with continuous mutations has not been considered before but it allows direct comparisons between h^{SBD} and h^{EGT}. In particular, the hitting probability in evolutionary graph theory is given by

$$\begin{aligned}h_{\mathcal{A}}^{\mathrm{EGT}}(\mathcal{I}) =& \frac{1}{\lambda_{\mathcal{I}}}\sum_{i\in\mathcal{I}}\Bigg[\sum_{j\in\mathcal{I}}R(i,j,U_i,\mathcal{I},W)[1-\mu(i)]h_{\mathcal{A}}^{\mathrm{EGT}}(\mathcal{I}\cup\{(U_i,X_j)\}\setminus\{j\}) \\ &+\int_{\mathbb{R}^l}R(i,j,w,\mathcal{I},W)\mu(i)h_{\mathcal{A}}^{\mathrm{EGT}}(\mathcal{I}\cup\{(w,X_j)\}\setminus\{j\})M(U_i,w)dw\Bigg], \qquad (2.6)\end{aligned}$$

whose form is similar to that of h^{SBD}. Therefore, for equivalence between SBD evolutionary dynamics and evolutionary graph theory dynamics, we check whether they have the same hitting probabilities, that is, $h_{\mathcal{A}}^{\mathrm{SBD}}(\mathcal{I}) = h_{\mathcal{A}}^{\mathrm{EGT}}(\mathcal{I})$. For the comparisons we make, we consider standard and other definitions of the replacement rate R.

Standard evolutionary graph theory dynamics

In evolutionary graph theory, three families of dynamics are generally considered [163]; link (L), death-birth (DB), and birth-death (BD) dynamics. In link dynamics, a link in

the network is selected, then the offspring of the individual at the start of the link replaces the individual at the end of the link. In death-birth (birth-death), an individual is first selected for death (birth) before a neighbouring individual is selected for birth (death). Each of these families have two distinct cases, given in [140], where individuals are selected for either birth or for death. If selection is on birth we append birth to the end of the dynamics (e.g. birth-death with selection on birth, BDB), and similarly for death (e.g. link with selection on death, LD). Selection is dependent on the fitness of the individuals. In evolutionary game theory [109], fitness is the average payoff received by an individual. Payoffs depend upon the strategy played in a game specifying the rules of interactions between individuals. When fitness is constant it does not depend upon the interactions with other individuals. The fitness of individual i will be denoted f_{U_i} and is assumed to be independent of its site. The replacement rates for the standard evolutionary graph theory dynamics are given in Table 2.1 and only hold for those states where each site has density 1; i.e. for $\mathcal{I}$ such that $|\mathcal{I}|_x = 1 \ \forall x \in \mathcal{X}$.

The conditions required to obtain the standard evolutionary graph theory dynamics from SBD evolutionary dynamics are summarised in Table 2.2, excluding BDD dynamics which could not be obtained. Details are given in Appendix 2.B. The conditions specify whether s, β, δ, γ are suppressed, identical for all traits, proportional to fitness and subject to other requirements. With the exception of LD dynamics, these conditions extend to the case where fitness is not constant such that it could depend upon the system state and not just the trait of an individual. The following insights are obtained from deriving the dynamics in this way:

- The standard evolutionary graph theory dynamics use only one component of SBD evolutionary dynamics. Those using the BD component are obtained by suppressing the natural death rate by setting $\delta_u = 0 \ \forall u \in \mathcal{U}$. This can be viewed as a biological scenario where individuals rarely die naturally but undergo intense intra-site competition with invaders. Those that can successfully invade are therefore more likely to spread. Fitness is interpreted as the birth rate when it acts on birth. The inverse fitness is interpreted as the death rate due to competition when it acts on death. On the other hand, those using the DB component are obtained when offspring cannot survive on occupied sites ($s = 0$). Biologically, this can be viewed as invasion being difficult, hence those types that can outlive their competitors are more likely to spread. Inverse fitness is interpreted as the natural death rate when it acts on death.

Table 2.1: Standard evolutionary graph theory dynamics

Dynamics	Description	$R(i,j,u,\mathcal{I},W)$
Death-Birth-Death (DBD)/ Voter Model	Individual j dies inversely proportional to its fitness and is replaced by neighbour i with probability proportional to $W_{X_jX_i}$.	$\frac{1/f_{U_j}}{\sum_{n\in\mathcal{I}} 1/f_{U_n}} \frac{W_{X_jX_i}}{\sum_{k\in\mathcal{I}\setminus\{j\}} W_{X_jX_k}}$
Death-birth-birth (DBB)	Individual j dies randomly, with probability $1/N$, and is then replaced by neighbour i with probability proportional to $f_{U_i}W_{X_jX_i}$.	$\frac{1}{N}\frac{f_{U_i}W_{X_jX_i}}{\sum_{k\in\mathcal{I}\setminus\{j\}} f_{U_k}W_{X_jX_k}}$
Link-birth (LB)	Individual i replaces j with probability proportional to $f_{U_i}W_{X_jX_i}$.	$\frac{f_{U_i}W_{X_jX_i}}{\sum_{n,k\in\mathcal{I}} f_{U_n}W_{X_kX_n}}$
Link-death (LD)	Individual i replaces j with probability proportional to $W_{X_jX_i}/f_{U_j}$.	$\frac{W_{X_jX_i}/f_{U_j}}{\sum_{n,k\in\mathcal{I}} W_{X_kX_n}/f_{U_k}}$
Birth-death-birth (BDB)/ Invasion Process	Individual i is chosen proportional to fitness, then replaces a neighbour j with probability proportional to $W_{X_jX_i}$.	$\frac{f_{U_i}}{\sum_{n\in\mathcal{I}} f_{U_n}}\frac{W_{X_jX_i}}{\sum_{k\in\mathcal{I}} W_{X_kX_i}}$
Birth-death-death (BDD)	Individual i is selected randomly who then replaces neighbour j proportional to $W_{X_jX_i}/f_{U_j}$.	$\frac{1}{N}\frac{W_{X_jX_i}/f_{U_j}}{\sum_{k\in\mathcal{I}} W_{X_kX_i}/f_{U_k}}$

Table 2.2: Assumptions required for all $u, v \in \mathcal{U}$ to obtain standard evolutionary graph theory dynamics from the SBD evolutionary dynamics. BDD dynamics are not listed as they could not be obtained.

Dynamics	Suppressed	Identical	Proportional to Fitness	Other
LB	$\delta_u = 0$	$\gamma_{u,v} = \gamma_{v,u}$	$\beta_u = f_u$	$s > 0$
LD	$\delta_u = 0$	$\beta_u = \beta_v$	$\gamma_{u,v} = 1/f_u,\ u \neq v$	$s > 0,\ \lvert\mathcal{U}\rvert = 2,\ \mu(i) = 0$
BDB	$\delta_u = 0$	$\gamma_{u,v} = \gamma_{v,u}$	$\beta_u = f_u$	$s > 0$, W is left stochastic
DBD	$s = 0$	$\beta_u = \beta_v$	$\delta_u = 1/f_u$	–
DBB	$s = 0$	$\delta_u = \delta_v$	$\beta_u = f_u$	–

Fitness is interpreted as the birth rate when it acts on birth.

- Link dynamics is a type of BD dynamics. In their definitions in Table 2.1, the order of birth and death is ambiguous. They are therefore classified separately from BD and DB dynamics. This hides intra-site competition within these dynamics. In LB dynamics intra-site competition is random with both individuals equally likely to die. In LD dynamics an individual dies inversely proportional to fitness due to intra-site competition.

- BDD dynamics cannot be obtained from SBD evolutionary dynamics. It requires birth and movement to be separate, which have been combined in our framework. This is evident in its definition (Table 2.1), where the term representing birth does not specify where an offspring is placed. To be able to obtain it we have to combine movement with death. In this case, an offspring is placed on a neighbouring site depending upon how likely the incumbent individual dies. Offspring placement is independent of neighbours in BDB so can be obtained from SBD evolutionary dynamics. This means that BDB and LB are related as they share the same spreading mechanism. In fact, [140] shows BDB is obtained from LB when W is left stochastic. Therefore, they have a different effect on birth rate. In LB position affects birth rate, which is $\beta_{U_i} \sum_{x \in \mathcal{X}} W_{xX_i}$ for i. In BDB the birth rate of i is β_{U_i} regardless of position.

- DBD and DBB can be obtained from SIS-type epidemic dynamics. This is because

they share the same spreading mechanism and do not have intra-site competition. They are obtained by setting $\beta_u \to \infty \ \forall u \in \mathcal{U}$ in Equation (2.5), resulting in vacant sites immediately being occupied by offspring. This is how [39] uses the contact process to obtain the voter model [72], which has identical dynamics to DBD. This illustrates that pathogen evolution, at least at the between host level, is likely to behave similarly to death-birth evolutionary dynamics rather than birth-death.

- There is no self-replacement in DBD and DBB. This means an individual cannot be replaced by its own offspring. Deriving them from SBD evolutionary dynamics specifies that death happens first followed by birth, preventing self-replacement. This information is lost in the standard definitions (Table 2.1), which allow self-replacement. This implies that self-replacement is only possible in BD type dynamics. In fact, DBD can be obtained from our derivation of LD when W is right stochastic. In this case self-replacement is allowed as LD is a type of BD dynamics. However, using this definition is limited due to the restrictions on LD (see Table 2.2).

Other evolutionary graph theory dynamics

BDB and DBB are combined in [204] using a parameter to allow a smooth transition between the two. Setting the parameter to 1 gives BDB, 0 gives DBB and a value in the 0 to 1 range gives a combination of them. SBD evolutionary dynamics are a viable and biologically motivated alternative to such a kind of dynamics.

The DB dynamics in [81] are obtained from SBD evolutionary dynamics by setting $s = 0$. In this case the birth and death events depend upon an individual's trait. Their dynamics are not based on fitness but parameters similar to δ and β. As their system is constructed in discrete time, these parameters are weights giving the likelihood of death and birth.

2.5 Framework Application II: Long-term behaviour of a spatial birth and death process

In certain circumstances, the SBD model behaves like existing models. However, these may not be realistic and therefore it is important to understand the behaviour of the SBD model outside of these circumstances. This biologically motivated framework can provide useful insights into the evolutionary process, and illuminate when conclusions from existing

models, such as evolutionary graph theory, may and may not be relevant. We will analyse the long-term behaviour of the SBD model for eco-evolutionary and evolutionary dynamics, both with and without clonal interference.

2.5.1 No clonal interference

Here we assume that adaptive mutations arise in succession, as in the classic model of [118]. This means a mutant either fixates or goes extinct in a resident population before another mutation arises. This behaviour is obtained in our framework in the rare mutation limit, $\mu(i) \to 0 \ \forall i$. This is confirmed by [22] who derive adaptive dynamics [36, 113] in this limit. However, we could also consider no mutation, $\mu(i) = 0 \ \forall i$, which is consistent with evolutionary graph theory [3]. We assume no mutation because the results are identical for the evolutionary scenario considered.

We consider an evolutionary scenario that is typically the case in evolutionary graph theory. There are two types; type B (trait 0) and type A (trait 1), so $\mathcal{U} = \{0, 1\}$. A mutant type A is introduced into a resident B population by randomly replacing a resident. The two types compete and eventually one fixates; i.e. all individuals are type A or type B. We are interested in the probability of type A fixating.

The fixation probability is formally defined in our framework as follows. Let $\mathcal{R} = \{\mathcal{I} : U_i = 0 \ \forall i \in \mathcal{I}\}$ be the set of states where type B fixate; i.e. there is at least one B but no A. Similarly, let $\mathcal{M} = \{\mathcal{I} : U_i = 1 \ \forall i \in \mathcal{I}\}$ be the set of states where type A fixate. The probability that type A fixate from state $\mathcal{I}$ is given by, $\rho^A(\mathcal{I}) = \lim_{t\to\infty} P(\Sigma(t) \in \mathcal{M} | \Sigma(0) = \mathcal{I})$. Since we are only interested in when the process reaches $\mathcal{R}$ or $\mathcal{M}$, these are absorbing states of the Markov process, and therefore this is equivalent to the hitting probability of the set $\mathcal{M}$. Therefore, the type A fixation probability is calculated by solving the equation

$$\begin{cases} \mathcal{L}_c \rho^A(\mathcal{I}) = 0 & \mathcal{I} \notin \mathcal{M} \cup \mathcal{R}, \\ \rho^A(\mathcal{I}) = 0 & \mathcal{I} \in \mathcal{R}, \\ \rho^A(\mathcal{I}) = 1 & \mathcal{I} \in \mathcal{M}. \end{cases} \tag{2.7}$$

For the fixation probability, an initial state is generally assumed to have 1 type A and $N - 1$ type B, with each site occupied by one individual only. This allows us to make comparisons with evolutionary graph theory where these are the possible initial states. Let $\mathcal{I}_0 = \{(0, x) : x = 1, \ldots, N\}$ be the state where there is one B on each site. The

average fixation probability of a randomly placed initial type A is calculated as follows,

$$\bar{\rho}^A = \frac{1}{N} \sum_{i \in \mathcal{I}_0} \rho^A(\mathcal{I}_0 \setminus \{i\} \cup \{(1, X_i)\}).$$

A type A mutant is assumed to be equally likely to appear on any given site. For convenience, we will denote the average fixation probability for a given value of c as $\bar{\rho}^A_c$.

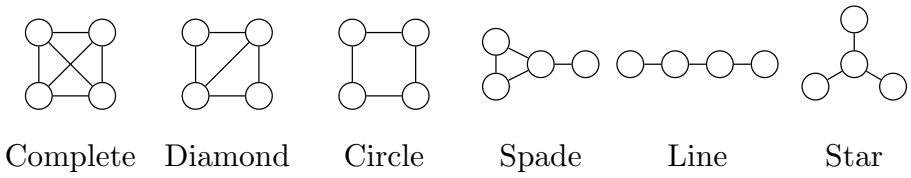

Figure 2.1: All undirected 4 node networks.

The SBD model will be used to compare the average type A fixation probability for different dynamics. Type A individuals are assumed to have identical death rates to type B ($\delta_0 = \delta_1$ and $\gamma_{u,v} = \gamma_{v,u}\ \forall u, v \in \{0, 1\}$) but an advantageous birth rate ($\beta_1 > \beta_0$). The following cases are considered:

1. We start with SIS-type dynamics ($\delta_0 = \delta_1 > 0$ and $s = 0$).
2. We then allow invasion ($\delta_0 = \delta_1 > 0$ and $s = 1$).
3. We then disallow natural death ($\delta_0 = \delta_1 = 0$ and $s = 1$).

For numerical evaluation of $\bar{\rho}^A$ in these cases, we set $\beta_0 = 3, \beta_1 = 10, \gamma_{u,v} = 5\ \forall u, v \in \{0, 1\}$ and $\delta_0 = \delta_1 = 1$ when greater than 0. This is plotted against the negative ecological feedback loop amplifier, c. We start from $c = 1$ with eco-evolutionary dynamics and gradually increase c to converge to evolutionary dynamics. This is done for all undirected networks with 4 sites as shown in Figure 2.1. 4 sites are chosen since this produces a small set of structurally distinct networks to investigate. For each network, W is left stochastic with self-loops included; i.e. all outgoing weights from a site with k neighbours are $1/(k+1)$. This implies that a type u individual has birth rate β_u regardless of position. The different cases are all shown in Figure 2.2. From these cases we observe that

$$\bar{\rho}^A \text{ in (i)} < \bar{\rho}^A \text{ in (ii)} < \bar{\rho}^A \text{ in (iii)} \tag{2.8}$$

for all networks. SIS-type dynamics are therefore the least beneficial for the advantageous mutant we have considered. Moving from (i) to (ii), we observe that allowing invasion is beneficial, since $\bar{\rho}^A$ shifts higher with the networks maintaining their order. As we move from (ii) to (iii), disallowing natural death provides a further benefit, since again $\bar{\rho}^A$ shifts higher. However, the networks now reverse their order, showing that the total effect of allowing invasion and disallowing natural death is largest in the star and smallest in the complete network. This is investigated further by analytically calculating $\bar{\rho}^A$ in the star and complete networks for evolutionary dynamics.

Average Fixation probability for a complete network with SBD evolutionary dynamics

Consider a complete network with arbitrary weights

$$W_{ij} = w > 0 \qquad \forall i, j \in \{1, \dots, N\}. \tag{2.9}$$

For evolutionary dynamics, we only need to consider population states with one individual on each site. The position of type B and type A individuals does not matter in these states due to site homogeneity. Therefore, states with the same number of type A individuals, k (which means there are $N-k$ residents), are lumped together and referred to by this number. We are interested in the rate at which the system transitions from some state k to a state with an additional trait u individual; i.e. with $k-(-1)^u$ type A. The replacement rate for such a transition is denoted $r_{k,u}$. For SBD evolutionary dynamics, this is given by

$$r_{k,u} = k(N-k)\left(s\beta_u w\frac{\delta_{1-u}+\gamma_{1-u,u}}{\delta_0+\delta_1+\gamma_{0,1}+\gamma_{1,0}} + \delta_{1-u}\frac{\beta_u}{\beta_0(N-k-u)+\beta_1(k-1+u)}\right). \tag{2.10}$$

The average fixation probability of a single initial type A on the complete network (Equation (2.9)) is given by [80]

$$\bar{\rho}^A_{\text{comp}} = \frac{1}{1+\sum_{m=1}^{N-1}\prod_{n=1}^{m}\frac{r_{n,0}}{r_{n,1}}}. \tag{2.11}$$

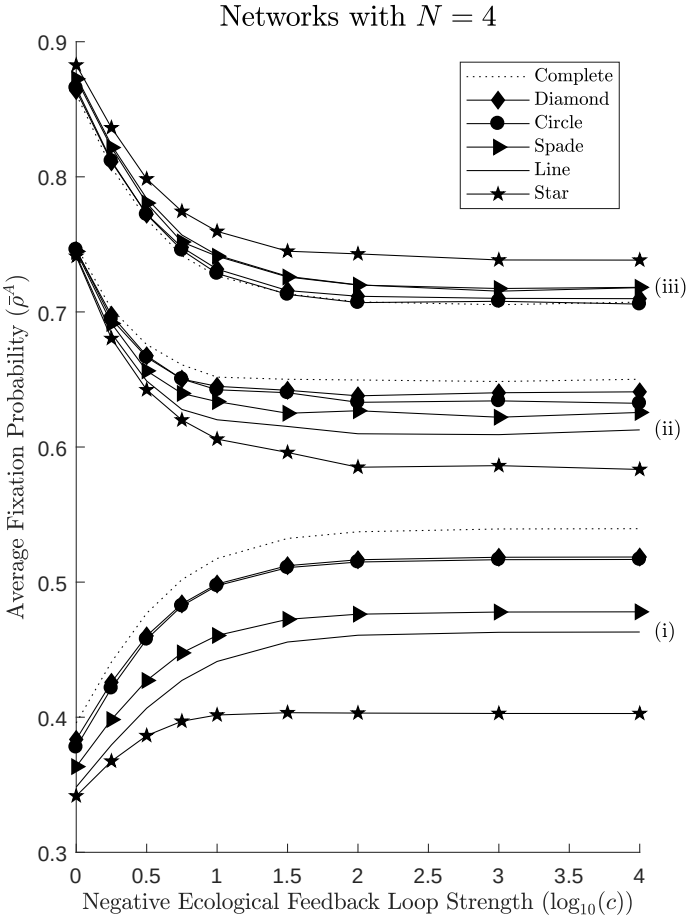

Figure 2.2: Average fixation probability ($\bar{\rho}^A$) of a type A in all undirected networks with 4 sites. Three different cases are considered, labelled (i)-(iii) on the right-hand side. For all cases, we set $\beta_0 = 3,\ \beta_1 = 10$, and $\gamma_{u,v} = 5\ \forall u, v \in \{0, 1\}$. For each case individually we have (i) $s = 0$ and $\delta_u = 1$, (ii) $s = 1$ and $\delta_u = 1$, and (iii) $s = 1$ and $\delta_u = 0$ for all $u \in \{0, 1\}$. In case (i), $\bar{\rho}^A$ is calculated analytically solving Equation (2.7). In cases (ii) and (iii), $\bar{\rho}^A$ is calculated by running 10^5 simulations.

The term $r_{n,0}/r_{n,1}$ is the rate of a type B increasing divided by the rate of a type A increasing. It is known as the backward bias of A or forward bias of B, and follows from Equation (2.10):

$$\frac{r_{k,0}}{r_{k,1}} = \frac{s\beta_0 w \frac{(\delta_1+\gamma_{1,0})}{(\delta_0+\delta_1+\gamma_{0,1}+\gamma_{1,0})} + \delta_1 \frac{\beta_0}{(N-k)\beta_0+(k-1)\beta_1}}{s\beta_1 w \frac{(\delta_0+\gamma_{0,1})}{(\delta_0+\delta_1+\gamma_{0,1}+\gamma_{1,0})} + \delta_0 \frac{\beta_1}{(N-k-1)\beta_0+k\beta_1}}.$$

The bias and average fixation probability are shown in Table 2.3 for the cases previously considered (page 65). The average fixation probabilities shown are a closed-form version of Equation (2.11) such that: (i) is obtained from [70]; (ii) is not shown due to its complexity; (iii) is obtained from the Moran probability [116] as the bias is constant. Comparing the biases in Table 2.3 reveals

$$\text{Bias (i)} > \text{Bias (ii)} > \text{Bias (iii)}. \tag{2.12}$$

Details are given in Appendix 2.C. In particular, the key requirement for Equation (2.12) to hold is that $\beta_1 > \beta_0$, which is already assumed to be the case. Equation (2.12) holding implies that Equation (2.8) holds for all $N > 1$, since a larger bias gives a lower fixation probability as seen in Equation (2.11). As $N \to \infty$ the difference between these cases diminishes because their biases all converge to β_0/β_1.

In Figure 2.3, $\bar{\rho}^A$ in the complete network (Equation (2.9)) is numerically evaluated for different values of c. It shows that $\bar{\rho}^A$ converges to $\bar{\rho}^A_{\text{comp}}$ as c gets larger. This confirms that the negative ecological feedback loop functions as desired. These plots are generated by setting the competition rates to $\gamma_{u,v} = 5\ \forall u, v \in \{0, 1\}$, which is relatively high compared to the birth rate of residents ($\beta_0 = 3$) and mutants ($\beta_1 \lessapprox 5$). This means that the number of individuals on a site would be close to one due to the intense competition between individuals sharing a site. The average fixation probability is therefore similar for all values of c. However, as the birth rate of type A increases ($\beta_1 \gtrapprox 5$) a clear separation of the fixation probability occurs for different values of c, since now the birth rate can overcome the competition in a site allowing more than one individual on a site. This allows more mutants to coexist on a site enabling them to easily overcome the type B, thus explaining why the average fixation probability is larger for $c = 1$ than $c \to \infty$ for a higher type A birth rate.

Average fixation probability for a star with SBD evolutionary dynamics

Consider the star with arbitrary weights

$$W_{11} = W_{i1} = w_c > 0, \quad W_{ii} = W_{1i} = w_l > 0 \quad \forall i \in \{2, \dots, N\}. \tag{2.13}$$

Site 1 is called the centre site as it is connected to all other $N - 1$ sites. All other sites are called the leaves as they are only connected to the centre site. As before, we only

Table 2.3: Bias and fixation probability for the complete network for SBD evolutionary dynamics. Cases considered assume advantageous type A with $\beta_1 > \beta_0$ (further details on page 65).

Case/ EGT Dynamics	Bias $\left(\frac{r_{k,0}}{r_{k,1}},\ k = 1, \ldots, N-1\right)$	Fixation probability ($\bar{\rho}^A_{\text{comp}}$)
(i) SIS-type dynamics ($\delta_0 = \delta_1 > 0,\ s = 0$)/ DBB	$\frac{\beta_0}{\beta_1}\frac{k\beta_1 + (N-k-1)\beta_0}{(N-k)\beta_0 + (k-1)\beta_1}$	$\frac{N-k}{N}\frac{1-(\beta_0/\beta_1)^k}{1-(\beta_0/\beta_1)^{N-1}}$
(ii) then allow invasion ($\delta_0 = \delta_1 = \delta > 0,\ s = 1$)/ None	$\frac{s\beta_0 w\frac{1}{2} + \delta\frac{\beta_0}{(N-k)\beta_0+(k-1)\beta_1}}{s\beta_1 w\frac{1}{2} + \delta\frac{\beta_1}{(N-k-1)\beta_0+k\beta_1}}$	Not shown, see text.
(iii) then disallow natural death ($\delta_0 = \delta_1 = 0,\ s = 1$)/ LB, BDB	β_0/β_1	$\frac{1-(\beta_0/\beta_1)}{1-(\beta_0/\beta_1)^N}$

consider population states with one individual on each site due to evolutionary dynamics being used. In such states, the position of type B and type A individuals present on leaves does not matter, since the leaves are identical. The population state is then given by (u, k) where u is the centre individual's type and k is the number of type A on leaves ($N-1-k$ is the number of residents on leaves). Since $u \in \{0, 1\}$, in state (u, k) there are $1-u \leq k \leq N-1-u$ type A on the leaves provided that there is at least one A and one B in the population. Let $r(u, k, u', k')$ be the rate of transitioning from state (u, k) to (u', k'). We only need to consider the two transitions where a change in state occurs. First, for SBD evolutionary dynamics a type u centre can replace a type $1-u$ leaf with rate

$$r(u, k, u, k-(-1)^u) = k^{1-u}(N-1-k)^u\left(sw_c\beta_u\frac{\delta_{1-u}+\gamma_{1-u,u}}{\delta_0+\delta_1+\gamma_{0,1}+\gamma_{1,0}} + \delta_{1-u}\right). \quad (2.14)$$

Second, a type u centre is replaced by a type $1-u$ leaf with rate

$$r(u, k, 1-u, k) = k^{1-u}(N-1-k)^u\left(sw_l\beta_{1-u}\frac{\delta_u+\gamma_{u,1-u}}{\delta_0+\delta_1+\gamma_{0,1}+\gamma_{1,0}} + \delta_u\frac{\beta_{1-u}}{(N-1-k)\beta_0+k\beta_1}\right). \quad (2.15)$$

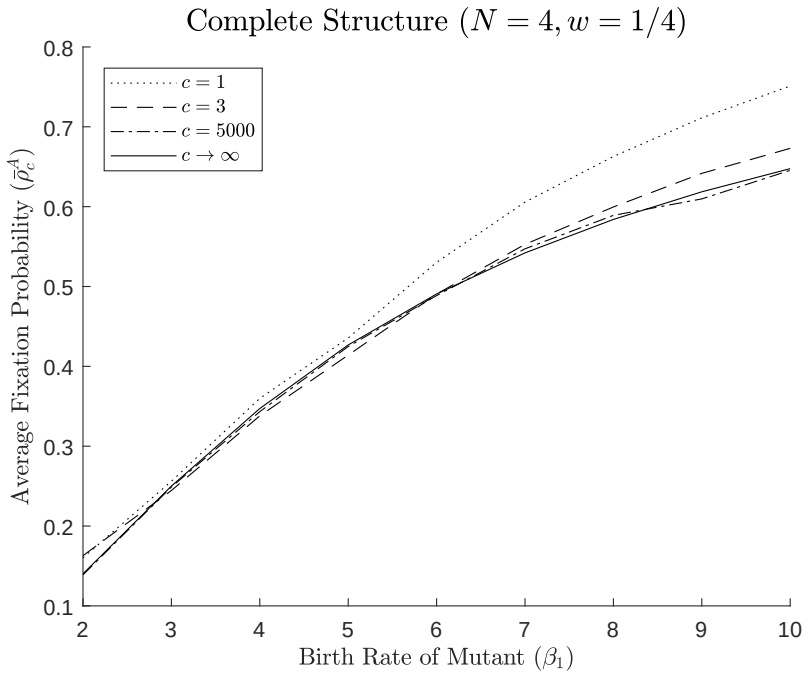

Figure 2.3: Average fixation probability ($\bar{\rho}^A$) of a type A in the complete network when using the SBD evolutionary dynamics, for various values of c, with $\beta_0 = 3$, $\delta_u = 1$ and $\gamma_{u,v} = 5\ \forall u, v \in \{0, 1\}$. For $c \to \infty$, $\bar{\rho}^A$ is given by $\bar{\rho}^A_{\text{comp}}$. For other values of c, $\bar{\rho}^A$ is calculated by running 10^4 simulations. As c gets larger we can see that $\bar{\rho}^A$ converges to $\bar{\rho}^A_{\text{comp}}$.

With these rates, the average fixation probability in the star, denoted $\bar{\rho}^A_{\text{star}}$, is obtained using the formula of [60] (see Appendix 2.D). For the three cases considered (page 65), $\bar{\rho}^A$ is given in Appendix 2.D. For $N \to \infty$, it is shown in Appendix 2.E that Equation (2.8) holds for the star (Equation (2.13)).

To investigate the interplay between the BD and DB components of the SBD evolutionary dynamics, we consider the case where we set $w_c = 1/N$ and $w_l = 1/2$ so that the birth rate is exactly $\beta_u\ \forall u \in \{0, 1\}$. As N gets larger the rate of replacing the centre individual is dominated by the BD component and the rate at which a leaf individual is replaced is dominated by the DB component. This means that the less connected leaf individuals are more reliant on BD to spread their offspring whereas the highly connected centre individual is more reliant on DB to spread its offspring.

Figure 2.4 shows the average fixation probability, $\bar{\rho}^A$, for different values of c in the star. Its qualitative properties are similar to that of the complete network and we once

again see that the negative ecological feedback loop functions as desired.

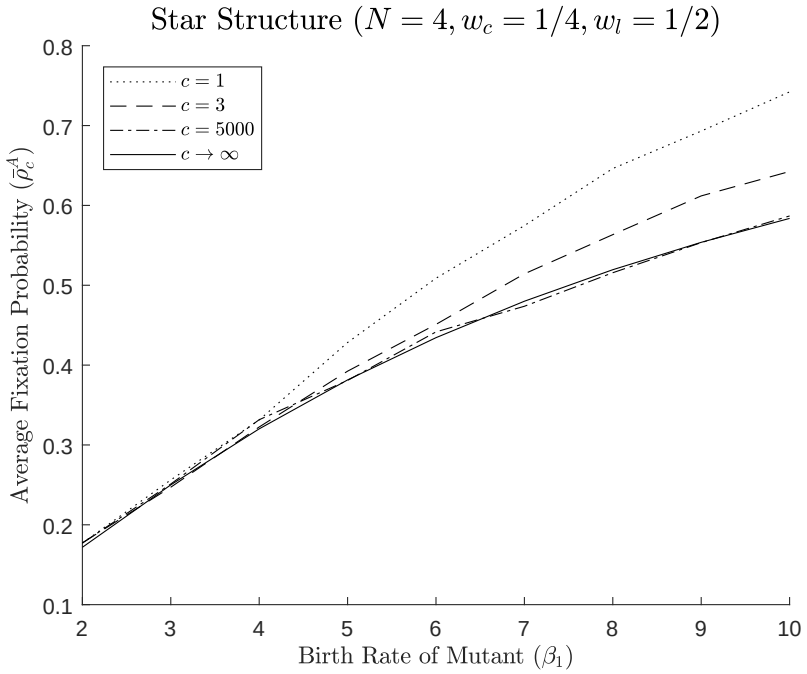

Figure 2.4: Average fixation probability ($\bar{\rho}^A$) of a mutant in the star when using the SBD evolutionary dynamics, for various values of c, with $\beta_0 = 3,\ \delta_u = 1$ and $\gamma_{u,v} = 5\ \forall u, v \in \{0, 1\}$. For $c \to \infty$, $\bar{\rho}^A$ is given by $\bar{\rho}^A_{\text{star}}$. For other values of c, $\bar{\rho}^A$ is calculated by running 10^4 simulations. As c gets larger we can see that $\bar{\rho}^A$ converges to $\bar{\rho}^A_{\text{star}}$.

Comparison of average fixation probabilities for complete network and star

In [102] it is shown that when using BDB dynamics the star amplifies the average fixation probability when compared to the complete network; i.e. $\bar{\rho}^A_{\text{star}} > \bar{\rho}^A_{\text{comp}}$. This is consistent with what is observed in the earlier plots. By using SBD evolutionary dynamics, we can explain why and show when the star is no longer an amplifier.

In source-sink metapopulation dynamics [149], a source is a site that is a net exporter of individuals whereas a sink is a site that is a net importer of individuals. This means that a source site is advantageous in comparison to a sink site, since more offspring are produced. In case (iii), the replacement rates (equations (2.14) and (2.15)) show that a leaf site behaves like a source when $w_l > w_c$ and a sink when $w_l < w_c$. This is because

the birth rate from a leaf to the centre is proportional to w_l, whereas from the centre to a leaf is proportional to w_c. The star is therefore an amplifier in case (iii) when BDB is in operation; i.e. when $w_c = 1/N$ and $w_l = 1/2$. This is verified in Figure 2.5 which illustrates that $\bar{\rho}^A_{\text{star}} > \bar{\rho}^A_{\text{comp}}$ when $w_l > w_c$, $\bar{\rho}^A_{\text{star}} < \bar{\rho}^A_{\text{comp}}$ when $w_l < w_c$, and $w_l = w_c$ is the boundary between amplification and suppression where $\bar{\rho}^A_{\text{star}} = \bar{\rho}^A_{\text{comp}}$.

The natural death rate plays a fundamental role since it can prevent a leaf site from being a source. This is shown in case (ii) where leaf individuals are adversely affected compared to the centre individual, as seen in the replacement rates (equations (2.14) and (2.15)). In particular, a leaf individual has to compete with other leaf individuals when the centre individual dies but this is not the case for the centre individual when a leaf individual dies. Another way to look at this is that a natural death rate limits the amount of time a leaf individual has to spread its offspring before it dies. This is verified in Figure 2.6, where increasing the death rate causes $\bar{\rho}^A_{\text{star}} - \bar{\rho}^A_{\text{comp}}$ to decrease, such that the star is no longer an amplifier of selection. This is consistent with [60], which shows that the star is not an amplifier under DBD and DBB dynamics.

2.5.2 With clonal interference

In this section, we no longer assume that adaptations are successive and take into account the effect of clonal interference, which has been demonstrated in a range of asexual organisms [79]. For clonal interference in unstructured populations, it was shown that the fixation probability of a beneficial mutation decreases as the population size and mutation rate increases [53]. In addition to this, fixation is a rare event where a beneficial mutation beats clonal interference by having a large selective advantage. The inclusion of clonal interference will therefore provide a better understanding of the impact that population structure has on the fixation probability of an adaptive mutation.

To study the effect of clonal interference, we consider the case where a type B (trait 0) resident population can be invaded by two kinds of mutant, types A (trait 1) and C (trait 2); i.e. $\mathcal{U} = \{0, 1, 2\}$. We will assume that there is a constant mutation probability, $\mu(i) = \mu \ \forall i$, and the mutation function in Equation (2.1) is defined as follows

$$M(u, v) = \begin{cases} \frac{1}{2} & u \neq v, \\ 0 & u = v \end{cases} \quad \forall u, v \in \{0, 1, 2\}.$$

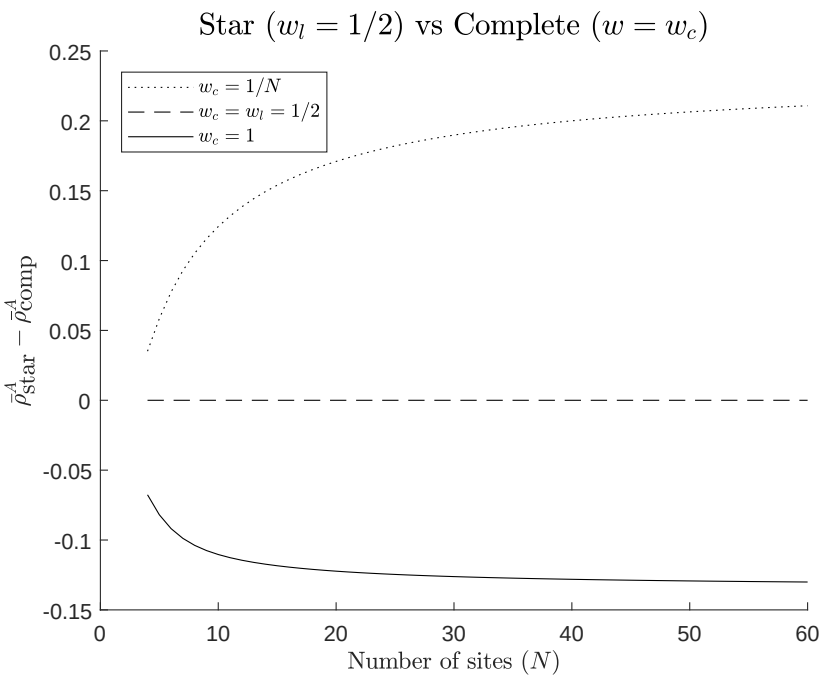

Figure 2.5: Plot of $\bar{\rho}^A_{\text{star}} - \bar{\rho}^A_{\text{comp}}$ against the number of sites (N) when using SBD evolutionary dynamics with $\beta_0 = 3,\ \beta_1 = 4,\ \gamma_{u,v} = 5$ and $\delta_u = 0\ \forall u, v \in \{0, 1\}$. It shows that the star is no longer an amplifier when $w_c > w_l$.

This means that the probability that a type u produces a type $v \neq u$ offspring is $\mu/2$. Note that in this case the integral in Equation (2.1) is changed to a summation because of the discrete number of mutations. The fixation probability before clonal interference is then defined as a type A mutant fixating in a population of type B residents before a type C mutant fixates in the population. This measures how successful a type A mutant is against the type B residents when there is a possibility of type C mutants appearing and interfering. This fixation probability can be calculated as follows. Let $\mathcal{R} = \{\mathcal{I} : U_i = 0\ \forall i \in \mathcal{I}\}$ be the set of states where type B fixate; i.e. there is at least one B but no A or C. Similarly, let $\mathcal{M}_1 = \{\mathcal{I} : U_i = 1\ \forall i \in \mathcal{I}\}$ be the set of states where type A fixate and $\mathcal{M}_2 = \{\mathcal{I} : U_i = 2\ \forall i \in \mathcal{I}\}$ be the set of states where type C fixate. The probability that type A fixate from state $\mathcal{I}$ is given by, $\rho^A(\mathcal{I}) = \lim_{t\to\infty} P(\Sigma(t) \in \mathcal{M}_1 | \Sigma(0) = \mathcal{I})$. Since we are only interested in when the process reaches $\mathcal{R}$, $\mathcal{M}_1$ or $\mathcal{M}_2$, these are absorbing states of the Markov process, and therefore this is equivalent to the hitting probability of the set

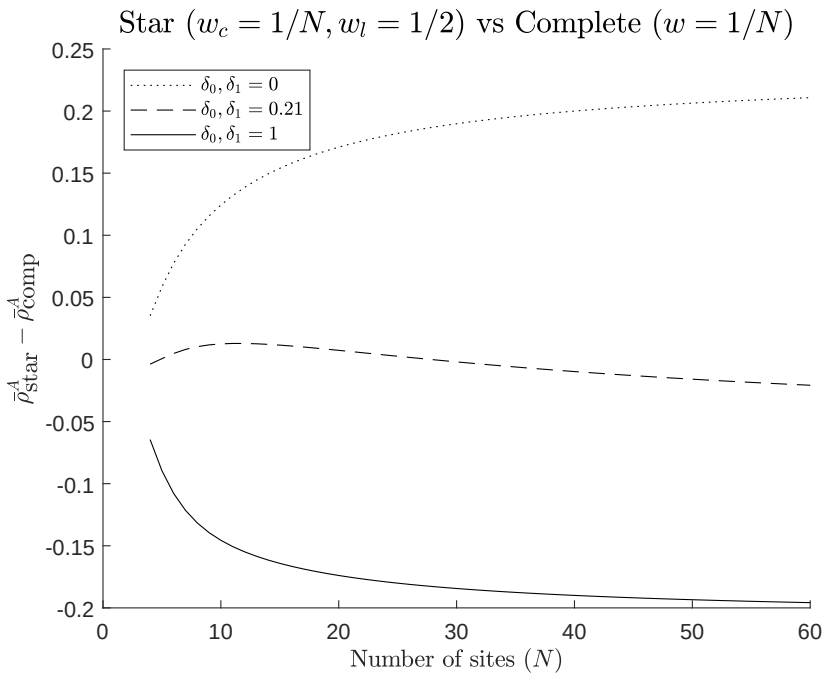

Figure 2.6: Plot of $\bar{\rho}^A_{\text{star}} - \bar{\rho}^A_{\text{comp}}$ against the number of sites (N) when using SBD evolutionary dynamics with $\beta_0 = 3,\ \beta_1 = 8,\ \gamma_{u,v} = 5$ and $\delta_u = 0, 0.21, 1\ \forall u, v \in \{0, 1\}$. It shows that the star is no longer an amplifier as the natural death rate increases.

$\mathcal{M}_1$. Therefore, the type A fixation probability is calculated by solving

$$\begin{cases} \mathcal{L}_c \rho^A(\mathcal{I}) = 0 & \mathcal{I} = \{(u, x) \notin \mathcal{M}_1 : \exists u = 1\} \\ \rho^A(\mathcal{I}) = 0 & \mathcal{I} \in \mathcal{R} \cup \mathcal{M}_2, \\ \rho^A(\mathcal{I}) = 1 & \mathcal{I} \in \mathcal{M}_1, \end{cases}$$

These equations specify that we are solving for the probability of hitting a state in $\mathcal{M}_1$ before reaching a state in $\mathcal{R} \cup \mathcal{M}_2$. With $\mathcal{I}_0$ defined as before, the average fixation probability with clonal interference is defined as follows

$$\bar{\rho}^A = \frac{1}{N} \sum_{i \in \mathcal{I}_0} \rho^A(\mathcal{I}_0 \setminus \{i\} \cup \{(1, X_i)\}),$$

where we are assuming that a mutant is equally likely to appear on any given site in order

to allow comparisons with results in evolutionary graph theory.

Clonal interference reduces the amount of time that a mutant of type A has to fixate, since the longer it takes the more likely a mutant of type C will appear. Without clonal interference, the complete network has the lowest fixation time whereas, for example, the star is substantially higher [52, 177**?**]. We should therefore expect the complete network to be least affected as the mutation rate increases in comparison to the star and other networks. Furthermore, the circulation theorem [102] in evolutionary graph theory that identifies structures whose fixation probability is equal to the Moran probability is based on the assumption that there is an infinite amount of time to fixate. Since these networks do not have equal fixation time, we should therefore expect this theorem to fail when there is clonal interference. To show that this is indeed the case, we plot the average fixation probability with clonal interference for different mutation rates in Figure 2.7 for all networks with 4 sites when using SBD evolutionary dynamics with parameters $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 3$, $\delta_u = 0$, $\gamma_{u,v} = 1$ for all $u, v \in \{0, 1, 2\}$. For the networks, W is defined to be left stochastic as it was before. This means that the dynamics are equivalent to BDB. When $\mu = 0$ we see that the fixation probability is identical for complete and circle structures because the circulation theorem holds, in particular, this is true because the incoming weight is equal to the outgoing weight for each site. However, this is no longer the case when $\mu > 0$. The average fixation probability under clonal interference decreases in all structures, with the complete structure being the least affected.

2.6 Discussion

In this chapter, we have developed a biologically motivated model for evolution in network-structured populations. The model we built is individual-based and accommodates eco-evolutionary dynamics. These dynamics allow the population size, density and composition to change as opposed to evolutionary dynamics that only allow the composition to change. We showed that using a negative ecological feedback loop allows suppression of ecological dynamics in our framework's eco-evolutionary dynamics, leaving us with evolutionary dynamics only. By changing the strength of this feedback loop we can move between eco-evolutionary and evolutionary dynamics. For example, death-birth evolutionary graph theory dynamics can be recovered from the multi-species SIS epidemic model by strengthening this feedback loop. Since death-birth can be obtained from epidemic-type dynamics, this suggests that pathogen evolution is more likely to follow death-birth rather than

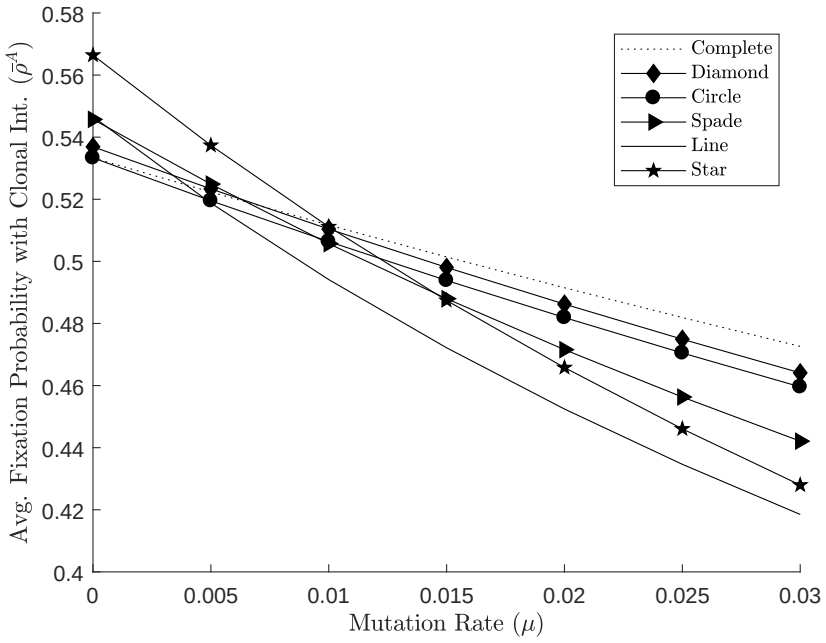

Figure 2.7: Comparison of average fixation probability with clonal interference ($\bar{\rho}^A$) when using SBD evolutionary dynamics. The parameters are set as follows $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 3$, $\delta_u = 0$, $\gamma_{u,v} = 1$ for all $u, v \in \{0, 1, 2\}$.

birth-death evolutionary dynamics. This is important to understand evolution in many healthcare challenges, such as antimicrobial resistance. Moving away from models with evolutionary dynamics only is important because fixed population size and density are an exception [27, 112] and this allows us to capture biological processes more accurately. For example, [151] compared evolutionary graph theory to a model with eco-evolutionary dynamics. This model showed that the decoupling of births and deaths promotes cooperation of cells in the epithelium, since it allows cooperators to cluster more effectively.

Using our framework, we defined a birth and death model where the ecological dynamics specify that the carrying capacity is dependent upon the population composition. In these ecological dynamics the intra-site competition is taken from [77], which provides a way to consider evolutionary games [109]. This birth and death model contains the SIS epidemic model and its variations, such that the birth rate is seen as the infection rate and the death rate as the recovery rate. The evolutionary dynamics of this model is used to

obtain evolutionary graph theory dynamics, which are based on the fitness of individuals; a measure of their reproductive success. For all except one of the standard evolutionary graph theory dynamics, we were able to relate fitness to birth and death rates. We found that, depending upon the dynamics, fitness is either proportional to the birth rate or inversely proportional to the death rate, which is in line with what is expected. By numerically investigating the shift from eco-evolutionary dynamics to evolutionary dynamics, we identified when the assumptions behind evolutionary dynamics may be realistic. As the relative strength of the mutant type increases, the evolutionary dynamics diverge from the eco-evolutionary dynamics. For the complete network, the negative ecological feedback loop does not need to be too strong to make the eco-evolutionary dynamics approach the evolutionary dynamics. For more extreme networks like the star however, this feedback has to be very strong to see agreement between the models, suggesting that the assumptions behind evolutionary dynamics may be more realistic for some population structures than others.

Long term population evolution was studied using the birth and death model (SBD), in the cases where clonal interference is and is not allowed. When there is no clonal interference, it is assumed that a mutant type invades a resident population and either fixates or goes extinct. The time to fixation depends upon the size and network structure of the population [52]. When there is no interference an infinite amount of time is allowed for the fixation of a mutant. The circulation theorem [102] in evolutionary graph theory is based on this assumption. However, when clonal interference is allowed a mutant has limited time to fixate before another beneficial mutation appears. In particular, the higher the mutation rate the more likely it is another beneficial mutation will appear and interfere with the current mutant trying to fixate, so time to fixation without interference is reduced. In this case, the circulation theorem does not hold as the infinite time to fixation assumption no longer holds.

To conclude, we have produced a biologically justified model that can be applied to realistic populations. By deriving existing models we have given biological insight into their underlying assumptions. This has opened the avenue for further investigations.

Appendix 2.A Hitting probability for modified dynamics

Dropping W from the birth and death rates for brevity, we have that

$$\begin{aligned}
0 &= \lim_{c\to\infty} \mathcal{L}_c h_{\mathcal{A}}(\mathcal{I}) \iff \\
0 &= \lim_{c\to\infty} \sum_{i\in\mathcal{I}} \sum_{x\in\mathcal{X}} [1-\mu(i)] B(c,i,x,\mathcal{I})[h_{\mathcal{A}}(\mathcal{I}\cup\{(U_i,x)\}) - h_{\mathcal{A}}(\mathcal{I})] \\
&\quad + \sum_{i\in\mathcal{I}} \sum_{x\in\mathcal{X}} \mu(u) B(c,i,x,\mathcal{I}) \int_{\mathbb{R}^l} [h_{\mathcal{A}}(\mathcal{I}\cup\{(w,x)\} - h_{\mathcal{A}}(\mathcal{I})] M(U_i,w)dw \\
&\quad + \sum_{i\in\mathcal{I}} D(c,i,\mathcal{I})[h_{\mathcal{A}}(\mathcal{I}\setminus\{i\}) - h_{\mathcal{A}}(\mathcal{I})].
\end{aligned}$$

Let

$$\lambda_{\mathcal{I}} = \lim_{c\to\infty} \sum_{i\in\mathcal{I}} \sum_{x\in\mathcal{X}} B(c,i,x,\mathcal{I}) + D(c,i,\mathcal{I}).$$

then rearranging gives

$$\begin{aligned}
h_{\mathcal{A}}(\mathcal{I}) = \lim_{c\to\infty} \frac{1}{\lambda_{\mathcal{I}}} \sum_{i\in\mathcal{I}} \Bigg[\sum_{x\in\mathcal{X}} B(c,i,x,\mathcal{I}) \Bigg(& [1-\mu(i)] h_{\mathcal{A}}(\mathcal{I}\cup\{(U_i,x)\}) \\
& + \mu(i) \int_{\mathbb{R}^l} h_{\mathcal{A}}(\mathcal{I}\cup\{(w,x)\}) M(U_i,w)dw \Bigg) \\
& + D(c,i,\mathcal{I}) h_{\mathcal{A}}(\mathcal{I}\setminus\{i\}) \Bigg].
\end{aligned}$$

We assume that the population starts in a state $\mathcal{I}$ where $|\mathcal{I}|_x = 1 \ \forall x \in \mathcal{X}$, we then have that

$$\begin{aligned}
h_{\mathcal{A}}(\mathcal{I}) = & \frac{1}{\lambda_{\mathcal{I}}} \sum_{i\in\mathcal{I}} \Bigg[\sum_{x\in\mathcal{X}} b(i,x,\mathcal{I}) \Bigg([1-\mu(i)] h_{\mathcal{A}}(\mathcal{I}\cup\{(U_i,x)\}) \\
& + \mu(i) \int_{\mathbb{R}^l} h_{\mathcal{A}}(\mathcal{I}\cup\{(w,x)\}) M(U_i,w)dw \Bigg) \\
& + d(i,\mathcal{I}) h_{\mathcal{A}}(\mathcal{I}\setminus\{i\}) \Bigg]
\end{aligned}$$

since all sites have density 1. From state $\mathcal{I}$ we consider the following different states $\mathcal{J}$ that the population can transition to.

1. For $\mathcal{J} = \mathcal{I} \cup \{(u, x)\}$ we have that

$$h_{\mathcal{A}}(\mathcal{J}) = \lim_{c\to\infty} \frac{1}{\lambda_{\mathcal{J}}} \sum_{j\in\mathcal{J}} \Bigg[\sum_{y\in\mathcal{X}} \Bigg([1-\mu(j)] B(c,j,y,\mathcal{J}) h_{\mathcal{A}}(\mathcal{J} \cup \{(U_j, y)\}) + \mu(j) B(c,j,y,\mathcal{J}) \int_{\mathbb{R}^l} h_{\mathcal{A}}(\mathcal{J} \cup \{(w,y)\} M(U_j, w) dw \Bigg) + D(c,j,\mathcal{J}) h_{\mathcal{A}}(\mathcal{J} \setminus \{j\}) \Bigg].$$

The birth rate in this case is given by

$$\lim_{c\to\infty} B(c,j,y,\mathcal{J}) = \lim_{c\to\infty} c^{H_0[-|\mathcal{J}_y|]} b(j,y,\mathcal{J}) = b(j,y,\mathcal{J})$$

as $H_0[-|\mathcal{J}|_y] = 0 \ \forall y \in \mathcal{X}$ as there are no empty sites. Similarly, the death rate in this case is given by

$$\lim_{c\to\infty} D(c,j,\mathcal{J}) = \lim_{c\to\infty} c^{H_2[|\mathcal{I}_{X_j}|]} d(j,\mathcal{J}) = \lim_{c\to\infty} c^{\delta_{X_j,x}} d(j,\mathcal{J})$$

as site x is the only site with two individuals and $\delta_{m,n}$ is the Kronecker delta function. This means that

$$\frac{\lim_{c\to\infty} B(c,j,y,\mathcal{J})}{\lambda_{\mathcal{J}}} = 0 \quad \forall j \in \mathcal{J}, y \in \mathcal{X}$$

and

$$\frac{\lim_{c\to\infty} D(c,j,\mathcal{J})}{\lambda_{\mathcal{J}}} = \frac{\lim_{c\to\infty} c^{\delta_{X_j,x}} d(j,\mathcal{J})}{\sum_{k\in\mathcal{J}} \lim_{c\to\infty} c^{\delta_{X_k,x}} d(k,\mathcal{J})} = \frac{\delta_{X_j,x} d(j,\mathcal{J})}{\sum_{k\in\mathcal{J}_x} d(k,\mathcal{J})}.$$

The hitting probability from state $\mathcal{J}$ is then given by

$$h_{\mathcal{A}}(\mathcal{J}) = \sum_{j\in\mathcal{J}_x} \frac{d(j,\mathcal{J}) h_{\mathcal{A}}(\mathcal{J} \setminus \{j\})}{\sum_{k\in\mathcal{J}_x} d(k,\mathcal{J})}.$$

2. For $\mathcal{J} = \mathcal{I} \setminus \{i\}$ such that $i \in \mathcal{I}$, by following a similar set of arguments as we have for case 1 we obtain the hitting probability from state $\mathcal{J}$ as follows

$$h_{\mathcal{A}}(\mathcal{J}) = \sum_{j \in \mathcal{J}} \frac{b(j, X_i, \mathcal{J})}{\sum_{k \in \mathcal{J}} b(k, X_i, \mathcal{J})} \Bigg([1 - \mu(j)] h_{\mathcal{A}}(\mathcal{J} \cup \{(U_j, X_i)\}) + \mu(j) \int_{\mathbb{R}^l} h_{\mathcal{A}}(\mathcal{J} \cup \{(w, X_i)\}) M(U_j, w) dw \Bigg).$$

Substituting the hitting probability from $\mathcal{J}$ for these two cases into the hitting probability from $\mathcal{I}$ gives

$$\begin{aligned} h_{\mathcal{A}}(\mathcal{I}) = & \frac{1}{\lambda_{\mathcal{I}}} \sum_{i \in \mathcal{I}} \Bigg[\sum_{x \in \mathcal{X}} b(i, x, \mathcal{I}) \\ & \Bigg([1 - \mu(i)] \frac{\sum_{j \in \mathcal{I}_x \cup \{(U_i, x)\}} d(j, \mathcal{I} \cup \{(U_i, x)\}) h_{\mathcal{A}}(\mathcal{I} \cup \{(U_i, x)\} \setminus \{j\})}{\sum_{j \in \mathcal{I}_x \cup \{(U_i, x)\}} d(j, \mathcal{I} \cup \{(U_i, x)\})} \\ & + \mu(i) \int_{\mathbb{R}^l} \frac{\sum_{j \in \mathcal{I}_x \cup \{(w, x)\}} d(j, \mathcal{I} \cup \{(w, x)\}) h_{\mathcal{A}}(\mathcal{I} \cup \{(w, x)\} \setminus \{j\})}{\sum_{j \in \mathcal{I}_x \cup \{(w, x)\}} d(j, \mathcal{I} \cup \{(w, x)\})} M(U_i, w) dw \Bigg) \\ & + d(i, \mathcal{I}) \sum_{j \in \mathcal{I} \setminus \{i\}} \frac{b(j, X_i, \mathcal{I} \setminus \{i\})}{\sum_{j \in \mathcal{I} \setminus \{i\}} b(j, X_i, \mathcal{I} \setminus \{i\})} \\ & \Bigg([1 - \mu(j)] h_{\mathcal{A}}(\mathcal{I} \setminus \{i\} \cup \{(U_j, X_i)\}) \\ & + \mu(j) \int_{\mathbb{R}^l} h_{\mathcal{A}}(\mathcal{I} \setminus \{i\} \cup \{(w, X_i)\}) M(U_j, w) dw \Bigg) \Bigg] \end{aligned}$$

This can be rewritten as follows

$$h_{\mathcal{A}}(\mathcal{I}) = \frac{1}{\lambda_{\mathcal{I}}} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} \Bigg[\Bigg(b(i, X_j, \mathcal{I}) \frac{d(j, \mathcal{J}(i, j))}{\sum_{k \in \mathcal{J}(i,j)_{X_j}} d(k, \mathcal{J}(i, j))}$$

$$+ d(j,\mathcal{I})\frac{b(i,X_j,\mathcal{I}\setminus\{j\})}{\sum_{k\in\mathcal{I}\setminus\{j\}} b(k,X_i,\mathcal{I}\setminus\{j\})}\Big)[1-\mu(i)]h_{\mathcal{A}}(\mathcal{J}(i,j)\setminus\{j\})$$
$$+ \int_{\mathbb{R}^l}\Bigg(b(i,X_j,\mathcal{I})\frac{d(j,\mathcal{K}(w,j))}{\sum_{k\in\mathcal{K}(w,j)_{X_j}} d(k,\mathcal{K}(w,j))}$$
$$+ d(j,\mathcal{I})\frac{b(i,X_j,\mathcal{I}\setminus\{j\})}{\sum_{k\in\mathcal{I}\setminus\{j\}} b(k,X_i,\mathcal{I}\setminus\{j\})}\Big)\mu(i)h_{\mathcal{A}}(\mathcal{K}(w,j)\setminus\{j\})M(U_i,w)dw\Bigg]$$

where $\mathcal{J}(i,j) = \mathcal{I}\cup\{(U_i,X_j)\}$ and $\mathcal{K}(w,j) = \mathcal{I}\cup\{(w,X_j)\}$. This can then be further simplified by writing

$$h_{\mathcal{A}}(\mathcal{I}) = \frac{1}{\lambda_{\mathcal{I}}}\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}}\Bigg(r(i,j,U_i,\mathcal{I})[1-\mu(i)]h_{\mathcal{A}}(\mathcal{J}(i,j)\setminus\{j\})$$
$$+\int_{\mathbb{R}^l} r(i,j,w,\mathcal{I})\mu(i)h_{\mathcal{A}}(\mathcal{K}(w,j)\setminus\{j\})M(U_i,w)dw\Bigg)$$

where $r(i,j,u,\mathcal{I})$ is the rate at which the offspring of individual i replaces j given that the offspring has trait u.

Appendix 2.B Deriving standard evolutionary graph theory dynamics

We need to show that the hitting probability for SBD evolutionary dynamics (Equation 2.3) is equivalent to that of evolutionary graph theory dynamics (Equation 2.6). We start by observing that for all the standard evolutionary graph theory dynamics, the replacement rate satisfies

$$R(i,j,u,\mathcal{I}) = R(i,j,v,\mathcal{I})\ \forall u,v\in\mathcal{U} \tag{2.16}$$

and therefore

$$\lambda_{\mathcal{I}} = \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} R(i,j,U_i,\mathcal{I})[1-\mu(i)] + \int_{\mathbb{R}^l} R(i,j,w,\mathcal{I})\mu(i)M(U_i,w)dw$$

$$= \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} R(i, j, \mathcal{I}),$$

where $R(i, j, \mathcal{I})$ is the replacement rate with the type of the offspring dropped. Furthermore, for all the standard evolutionary graph theory dynamics the following also holds

$$\lambda_{\mathcal{I}} = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} R(i, j, \mathcal{I}) = 1$$

since the replacement rates are defined as probabilities. We then require that the replacement rate r for SBD evolutionary dynamics have the same property as in Equation (2.16), that is,

$$r(i, j, u, \mathcal{I}) = r(i, j, v, \mathcal{I}) \ \forall u, v \in \mathcal{U}, \tag{2.17}$$

we can therefore use $r(i, j, \mathcal{I})$ as the offspring type can be dropped, and

$$R(i, j, \mathcal{I}) = \frac{r(i, j, \mathcal{I})}{\sum_{n \in \mathcal{I}} \sum_{k \in \mathcal{I}} r(n, k, \mathcal{I})}.$$

This ensures that the hitting probability is identical for both types of dynamics. Recall that the replacement rate for SBD evolutionary dynamics is given by

$$r(i, j, u, \mathcal{I}, W) = s\beta_{U_i} W_{X_j X_i} \frac{\delta_{U_j} + \gamma_{U_j, u}}{\delta_{U_j} + \gamma_{U_j, u} + \delta_u + \gamma_{u, U_j}} + \delta_{U_j} \frac{\beta_{U_i} W_{X_j X_i}}{\sum_{k \in \mathcal{I} \setminus \{j\}} \beta_{U_k} W_{X_j X_k}}.$$

We can now consider which of the standard evolutionary graph theory dynamics we can obtain from these dynamics.

LB dynamics Setting $\delta_u = 0$ and $\gamma_{u,v} = \gamma_{v,u} \ \forall u, v \in \mathcal{U}$ satisfies Equation (2.17) and gives

$$\frac{r(i, j, \mathcal{I})}{\sum_{n \in \mathcal{I}} \sum_{k \in \mathcal{I}} r(n, k, \mathcal{I})} = \frac{\beta_{U_i} W_{X_j X_i}}{\sum_{n \in \mathcal{I}} \sum_{k \in \mathcal{I}} \beta_{U_n} W_{X_k X_n}}$$

which is identical to the LB dynamics when $\beta_u = f_u \ \forall u \in \mathcal{U}$.

BDB dynamics Doing the same as with the derivation of LB dynamics, but setting W to be left stochastic gives

$$\frac{r(i,j,\mathcal{I})}{\sum_{n\in\mathcal{I}}\sum_{k\in\mathcal{I}} r(n,k,\mathcal{I})} = \frac{\beta_{U_i}}{\sum_{n\in\mathcal{I}}\beta_{U_n}} W_{X_jX_i}$$

which is identical to the BDB dynamics when $\beta_u = f_u \ \forall u \in \mathcal{U}$.

DBD dynamics Setting $s = 0$ and $\beta_u = \beta_v \ \forall u, v \in \mathcal{U}$ satisfies Equation (2.17) and gives

$$\frac{r(i,j,\mathcal{I})}{\sum_{n\in\mathcal{I}}\sum_{k\in\mathcal{I}} r(n,k,\mathcal{I})} = \frac{\delta_{U_j}\dfrac{W_{X_jX_i}}{\sum_{k\in\mathcal{I}\setminus\{j\}} W_{X_jX_k}}}{\sum_{n\in\mathcal{I}}\sum_{k\in\mathcal{I}\setminus\{n\}}\delta_{U_n}\dfrac{W_{X_nX_k}}{\sum_{m\in\mathcal{I}\setminus\{n\}} W_{X_nX_m}}} = \frac{\delta_{U_j}}{\sum_{n\in\mathcal{I}}\delta_{U_n}}\frac{W_{X_jX_i}}{\sum_{k\in\mathcal{I}\setminus\{j\}} W_{X_jX_k}}$$

which is identical to DBD dynamics when $\delta_u = 1/f_u \ \forall u \in \mathcal{U}$.

DBB Dynamics Setting $s = 0$ and $\delta_u = \delta_v \ \forall u, v \in \mathcal{U}$ satisfies Equation (2.17) and gives

$$\frac{r(i,j,\mathcal{I})}{\sum_{n\in\mathcal{I}}\sum_{k\in\mathcal{I}} r(n,k,\mathcal{I})} = \frac{\delta_{U_j}\dfrac{\beta_{U_i}W_{X_jX_i}}{\sum_{k\in\mathcal{I}\setminus\{j\}} \beta_{U_k}W_{X_jX_k}}}{\sum_{n\in\mathcal{I}}\sum_{k\in\mathcal{I}\setminus\{n\}}\delta_{U_n}\dfrac{\beta_{U_k}W_{X_nX_k}}{\sum_{m\in\mathcal{I}\setminus\{n\}} \beta_{U_m}W_{X_nX_m}}} = \frac{1}{N}\frac{\beta_{U_i}W_{X_jX_i}}{\sum_{k\in\mathcal{I}\setminus\{j\}} \beta_{U_k}W_{X_jX_k}}$$

which is identical to DBB dynamics when $\beta_u = f_u \ \forall u \in \mathcal{U}$.

LD Dynamics To obtain LD dynamics, we have to assume there is no mutation and that there are only two types; i.e. $|\mathcal{U}| = 2$. Note that excluding transitions to the same state will not affect $h_{\mathcal{A}}$ so if we discount transitions to the same state, we would require

that

$$\frac{R(i,j,\mathcal{I})}{\sum_{n\in\mathcal{I}}\sum_{\substack{k\in\mathcal{I}\\U_k\neq U_n}} R(n,k,\mathcal{I})} = \frac{r(i,j,\mathcal{I})}{\sum_{n\in\mathcal{I}}\sum_{\substack{k\in\mathcal{I}\\U_k\neq U_n}} r(n,k,\mathcal{I})} \quad \text{for } U_j \neq U_i. \tag{2.18}$$

Setting $\delta_u = 0$ and $\beta_u = \beta_v \ \forall u \in \mathcal{U}$ simplifies the RHS of Equation (2.18) to

$$\frac{W_{X_jX_i}\frac{\gamma_{U_j,U_i}}{\gamma_{U_j,U_i}+\gamma_{U_i,U_j}}}{\sum_{n\in\mathcal{I}}\sum_{\substack{k\in\mathcal{I}\\U_k\neq U_n}} W_{X_kX_n}\frac{\gamma_{U_k,U_n}}{\gamma_{U_k,U_n}+\gamma_{U_n,U_k}}} = \frac{W_{X_jX_i}\gamma_{U_j,U_i}}{\sum_{n\in\mathcal{I}}\sum_{\substack{k\in\mathcal{I}\\U_k\neq U_n}} W_{X_kX_n}\gamma_{U_k,U_n}}$$

which for $\gamma_{u,v} = 1/f_u$ when $u \neq v \ \forall u, v \in \mathcal{U}$ is equivalent to the LHS of Equation (2.18) when using LD dynamics.

Appendix 2.C Showing strict order in bias for the complete network

We want to show that Equation (2.12) holds. For simplicity, we can write Bias(ii) as follows

$$\frac{x+a}{y+b} = \frac{\delta\frac{\beta_0}{(N-k)\beta_0+(k-1)\beta_1} + s\beta_0 w\frac{1}{2}}{\delta\frac{\beta_1}{(N-k-1)\beta_0+k\beta_1} + s\beta_1 w\frac{1}{2}}.$$

Equation (2.12) can be then be rewritten in the following way

$$\frac{x}{y} > \frac{x+a}{y+b} > \frac{a}{b},$$

which implies that

$$0 > ay - bx.$$

Expanding this equation gives

$$0 > ay - bx$$

$$
\begin{aligned}
0 &> s\beta_0 w\frac{1}{2}\delta\frac{\beta_1}{(N-k-1)\beta_0+k\beta_1} - s\beta_1 w\frac{1}{2}\delta\frac{\beta_0}{(N-k)\beta_0+(k-1)\beta_1}\\
0 &> \frac{1}{(N-k-1)\beta_0+k\beta_1} - \frac{1}{(N-k)\beta_0+(k-1)\beta_1}\\
0 &> (N-k)\beta_0+(k-1)\beta_1 - [(N-k-1)\beta_0+k\beta_1]\\
0 &> \beta_0-\beta_1
\end{aligned}
$$

which is indeed as the case as it is assumed that $\beta_1 > \beta_0$.

Appendix 2.D Average fixation probability of the star network

In [60], the average fixation probability is given by

$$\bar{\rho}^A_{\text{star}} = \frac{\rho^A_{\text{star—centre}} + (N-1)\rho^A_{\text{star—leaf}}}{N}.$$

Here, $\rho^A_{\text{star—centre}}$ is the fixation probability of a type A mutant starting in the centre and $\rho^A_{\text{star—leaf}}$ is the fixation probability of a mutant starting in a leaf. They are given by

$$\rho^A_{\text{star—centre}} = \frac{p(1,0,1,1)}{A(1,N-1)} \quad \text{and} \quad \rho^A_{\text{star—leaf}} = \frac{p(0,1,1,1)}{A(1,N-1)}$$

where

$$A(l,m) = 1 + \sum_{j=l}^{m-1} p(1,j,0,j)\prod_{k=l}^{j}\frac{p(0,k,0,k-1)}{p(1,k,1,k+1)}$$

and

$$
\begin{aligned}
p(u,k,u,k-(-1)^u) &= \frac{r(u,k,u,k-(-1)^u)}{r(u,k,u,k-(-1)^u)+r(u,k,1-u,k)},\\
p(u,k,1-u,k) &= \frac{r(u,k,1-u,k)}{r(u,k,u,k-(-1)^u)+r(u,k,1-u,k)}.
\end{aligned}
$$

For each of the cases considered, $\bar{\rho}^A_{\text{star}}$ is given by:

$$\bar{\rho}^A_{\text{star}}((\text{i})) = \frac{\frac{1}{N}\frac{N-1}{N} + \frac{N-1}{N}\frac{\beta_1}{(N-2)\beta_0+2\beta_1}}{1+\sum_{j=1}^{N-2}\frac{\beta_0}{(N-j)\beta_0+j\beta_1}\prod_{k=1}^{j}\frac{(N-k)\beta_0+k\beta_1}{(N-1-k)\beta_0+(k+1)\beta_1}},$$

$$\bar{\rho}^A_{\text{star}}((\text{ii})) = \frac{\frac{1}{N}\frac{sw_c\beta_1/2+\delta}{s(w_c\beta_1+w_l\beta_0)/2+\delta\frac{N}{N-1}} + \frac{N-1}{N}\frac{sw_l\beta_1/2+\delta\frac{\beta_1}{(N-2)\beta_0+\beta_1}}{s(w_l\beta_1+w_c\beta_0)/2+\delta\frac{(N-2)\beta_0+2\beta_1}{(N-2)\beta_0+\beta_1}}}{1+\sum_{j=1}^{N-2}\frac{sw_l\beta_0/2+\delta\frac{\beta_0}{(N-1-j)\beta_0+j\beta_1}}{s(w_l\beta_0+w_c\beta_1)/2+\delta\frac{(N-j)\beta_0+j\beta_1}{(N-1-j)\beta_0+j\beta_1}}\prod_{k=1}^{j}\frac{sw_c\beta_0/2+\delta}{sw_c\beta_1/2+\delta}\frac{s(w_c\beta_1+w_l\beta_0)/2+\delta\frac{(N-k)\beta_0+k\beta_1}{(N-1-k)\beta_0+k\beta_1}}{s(w_c\beta_0+w_l\beta_1)/2+\delta\frac{(N-1-k)\beta_0+(k+1)\beta_1}{(N-1-k)\beta_0+k\beta_1}}},$$

$$\bar{\rho}^A_{\text{star}}((\text{iii})) = \frac{\frac{1}{N}\frac{w_c\beta_1}{w_c\beta_1+w_l\beta_0} + \frac{N-1}{N}\frac{w_l\beta_1}{w_l\beta_1+w_c\beta_0}}{1+\sum_{j=1}^{N-2}\frac{w_l\beta_0}{w_l\beta_0+w_c\beta_1}\left(\frac{\beta_0}{\beta_1}\frac{w_c\beta_1+w_l\beta_0}{w_c\beta_0+w_l\beta_1}\right)^j}.$$

Appendix 2.E Proof for the star network

We want to show that Equation (2.8) holds for the star when $N \to \infty$. In case (i) we have that

$$\lim_{N\to\infty}\bar{\rho}^A_{\text{star}}((\text{i})) = 0.$$

In case (ii) we have that

$$\lim_{N\to\infty}\bar{\rho}^A_{\text{star}}((\text{ii})) = \frac{\frac{w_l\beta_1}{w_l\beta_1+w_c\beta_0+2\delta}}{1+\frac{w_l\beta_0}{w_l\beta_0+w_c\beta_1+2\delta}\sum_{j=1}^{\infty}\left(\frac{w_c\beta_0+2\delta}{w_c\beta_1+2\delta}\frac{w_c\beta_1+w_l\beta_0+2\delta}{w_c\beta_0+w_l\beta_1+2\delta}\right)^j}.$$

The denominator in this case converges to

$$1+\frac{ar}{1-r}$$

where

$$a = \frac{w_l\beta_0}{w_l\beta_0 + w_c\beta_1 + 2\delta},$$
$$r = \frac{w_c\beta_0 + 2\delta}{w_c\beta_1 + 2\delta}\frac{w_c\beta_1 + w_l\beta_0 + 2\delta}{w_c\beta_0 + w_l\beta_1 + 2\delta}.$$

Let

$$x = \frac{w_l\beta_1}{w_l\beta_1 + w_c\beta_0 + 2\delta},$$

we therefore have that

$$\lim_{N\to\infty} \bar{\rho}^A_{\text{star}}((\text{ii})) = \frac{x(1-r)}{1 + r(a-1)}.$$

In case (iii) we have that

$$\lim_{N\to\infty} \bar{\rho}^A_{\text{star}}((\text{iii})) = \frac{x_{\delta=0}(1 - r_{\delta=0})}{1 + r_{\delta=0}(a_{\delta=0} - 1)}$$

where $x_{\delta=0}, a_{\delta=0}, r_{\delta=0}$ are x, a, r with $\delta = 0$. We have that

$$\lim_{N\to\infty} \bar{\rho}^A_{\text{star}}((\text{ii})) < \lim_{N\to\infty} \bar{\rho}^A_{\text{star}}((\text{iii}))$$

if $r > r_{\delta=0}, a < a_{\delta=0}, x < x_{\delta=0}$, which is indeed the case since $\beta_1 > \beta_0$ and $\delta > 0$ (in case (ii)). This therefore gives

$$\lim_{N\to\infty} \bar{\rho}^A_{\text{star}}((\text{i})) < \lim_{N\to\infty} \bar{\rho}^A_{\text{star}}((\text{ii})) < \lim_{N\to\infty} \bar{\rho}^A_{\text{star}}((\text{iii}))$$

as required.

Chapter 3

Evolutionary bet hedging in structured populations

Traditional evolutionary models consider constant environments, where the fitness of individuals does not change in time. However, many real populations feature fluctuating environments, which may alter the fitness of different individuals. Many traits in biological populations have been explained by selection for risk-spreading to safeguard against environmental variation, known as evolutionary bet-hedging [14, 100, 160, 162, 173, 174, 186]. In a stochastically varying environment, a species that maximises its mean reproductive rate, or mean fitness, is not necessarily the strongest, since this could coincide with increased sensitivity to fluctuations in the environment. A bet-hedger is defined as a strategy that has lower mean fitness than its rival, but is selected over the rival since it has reduced variation in its fitness, due to being less sensitive to these fluctuations. For example, consider a simple habitat that fluctuates between a short wet season and a long dry season. Mean fitness would be maximised by adapting to the dry season. However, such an adaptation may result in terrible performance during the wet season. A generalist, who is equally adapted to both seasons, will have lower mean fitness, but is protected from the environmental fluctuations and therefore has reduced variation in its fitness across seasons.

An ecological example of bet-hedging can be observed in the delayed germination strategy in desert annuals [24, 145, 186]. These plants release multiple seeds, most of whom germinate in the next season with a fraction remaining dormant until future seasons. Such a strategy reduces mean fitness, since some dormant seeds may be lost before germination. However, this strategy also reduces the variation in fitness, because it ensures that not

all offspring will die if next season is bad. Bet-hedging adaptations have also evolved in microbial communities under turbulent environments [14, 100]. These examples consider between-generational variation [55], whereby all individuals experience the same conditions at any time. Mathematically, adaption to counter this type of variation is easily described and understood assuming evenly mixed populations of species [55, 74, 173].

Environmental variation can also act locally on individuals, causing within-generational (or demographic) variation. In this context, the fitness of an individual can be different from that of another individual of the same type at a given time, but both will have fitness drawn from the same distribution. One example is where predation levels across the habitat are variable. Assuming a cost of spreading offspring across numerous sites, the strategy to maximise fitness corresponds to choosing a single nesting location. However, since this site could be predated, a bet-hedger could evolve that spreads offspring across numerous sites to reduce the predation risk. An ecological example of potential bet-hedging against within-generational variation has been observed in female sierra dome spiders [189]. These females exhibit a multiple paternity strategy, whereby the primary mate is the victor of a fight among potential suitors, and secondary mates are selected at random. Mean fitness would be maximised by only selecting the primary mate, but random secondary mating hedges against the fight only taking place between weak suitors. Many other examples are similar and focus on multiple-paternity as a bet-hedging strategy [48, 160, 189, 198]. Other work has identified strategies in Cabbage Butterflies [158] and Aphids [188] that potentially evolved as bet-hedgers against within-generational variation.

Despite the ecological observations, current mathematical models in well-mixed populations lead to the conclusion that such variation does not drive evolutionary adaption unless the population is unrealistically small [55, 74, 75], contradicting and challenging the ecological observations [26, 74, 75]. Such challenges have potentially led to the lack of examples of bet-hedging against within-generational variation in recent literature, apart from cases restricting themselves to small population sizes [160].

Real populations are often not well-mixed and typically exist within some defined population structure, such as spatial or social structure. Some population structures consist of distinct groups of individuals within patches (or demes). Bet-hedging in such populations has been investigated using metapopulation models and deme-structured models [97, 170, 171, 199]. These cases have demonstrated that within-generational bet-hedging strategies can evolve in metapopulations, provided the deme (or patch) contains sufficiently few individuals. Here we generalise the study of bet-hedging to arbitrarily structured pop-

ulations, which can capture interaction/competition structure as well as deme-structure. Mathematically, evolution in structured populations is described by evolutionary graph theory [102, 129], which we build upon to incorporate variation. By analysing the evolutionary process in structured populations, we show that the selective pressure for bet-hedging strategies can only increase with both within and between generational variation. This supports the conclusions from metapopulations in providing an explanation for the ecological observations. Importantly, we show that population structure can facilitate within-generational bet-hedging regardless of population size and we discuss how different types of structure impact its selection.

3.1 Methods

3.1.1 Evolutionary model

To determine the impact of population structure on the evolution of bet-hedging strategies we model the dynamics of the process. We are interested in when a bet-hedging strategy has a competitive advantage over the non bet-hedging strategy, which we call the normal-type. Evolution in structured populations (generally in a non-variable environment) is described by evolutionary graph theory [6, 19, 20, 71, 102, 151]. Here, population structure is represented by an undirected left stochastic graph.

We consider a population with two types (or strategies) of individuals, the bet-hedging strategy A and the normal-type strategy B, either of which can play the role of the resident. The population structure is defined by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of edges between these nodes.

Following [8, 156, 192], the fitness of an individual is proportional to its birth rate. Therefore, an individual is first selected to die at random, resulting in a vacant node in the population, which the neighbouring (connected) individuals of this node compete to replace with an identical offspring, with probability proportional to their fitness (Figure 3.1). Since offspring are identical, there is no further mutation until one strategy eliminates the other. Following [129], we refer to these dynamics as DBB. Whilst we focus on these dynamics, we discuss other dynamics, such as the invasion process, in Section 3.2.3.

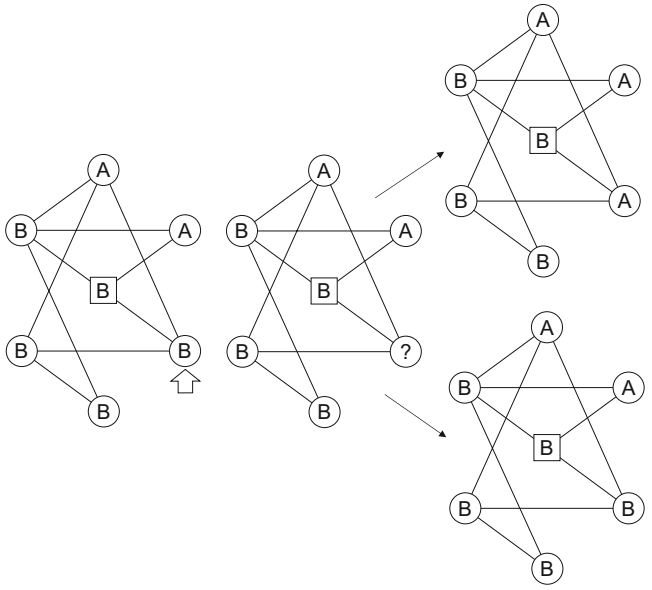

Figure 3.1: The update dynamics of the evolutionary process. First, an individual is randomly selected for death, indicated by the white arrow, resulting in a vacant node. This becomes the type of one of its neighbours, with probabilities proportional to their fitness. Either the selected node becomes occupied by a bet-hedger A or a normal-type B.

Traditional evolutionary graph theory dynamics do not capture variation in fitness, which is present in many real-world populations. To incorporate this, we treat fitness as a random variable that changes value with the environmental state. This general definition of fitness can account for the potential between and/or within-generational variation discussed above, and captures the many special cases currently studied [14, 56, 132, 160, 180, 186, 198, 199, 200].

To describe how this stochastic process changes, we define the probability of moving from one state to another, where a state represents which nodes are occupied by A and which are occupied by B. From any state S that is not all A or all B, we can move to a state S^+ with more type A individuals or S^- with more type B individuals, or remain in state S. To move from S to S^+ we require a B individual to die followed by selecting an A individual for reproduction. For a given node j, the probability of death is $1/N$. After the individual in node j is selected for death the neighbouring individuals, which we will refer to as the selection group (Figure 3.2), compete to replace j. We will refer to

the probability of selecting a certain type (given that an individual has been selected for death) as the selection probability of that type. Therefore the probability of moving to a state S^+ is given by the sum of probabilities that a type B is selected for death multiplied by the type A selection probability. That is,

$$P(S \to S^+) = \sum_{j=1}^{N} \frac{1}{N} B_j^S P(\text{type } A \text{ selected to reproduce} \mid \text{j dies} \cap \text{ state S}) \tag{3.1}$$

and similarly the probability of moving to a state with more B individuals is given by

$$P(S \to S^-) = \sum_{j=1}^{N} \frac{1}{N} A_j^S P(\text{type } B \text{ selected to reproduce} \mid \text{j dies} \cap \text{ state S}), \tag{3.2}$$

where $A_j^S = 1$ if and only if the individual in node j is type A in state S, and zero otherwise. These two probabilities dictate how the system evolves at each time step, and therefore provide a measure of the relative strength between the competing strategies.

Definition 2. Evolutionary bet-hedging - Evolutionary bet-hedging describes a strategy with a lower mean fitness than their competitors but also a lower variation in fitness. Such strategies can be favoured in the evolutionary process if their fitness variation is sufficiently low.

3.1.2 Impact of fitness variation

An obvious measure of variation is the variance, which is convenient since it is easily calculated. However, it has limitations; for example, if two distributions have equal mean and variance it gives no insight into which is more varied. This information is captured in the higher order moments of the distributions, such as the skew and kurtosis. A more comprehensive representation of variation is given by the convex order [165, 194], such that if one distribution is greater than another in convex order then it is more variable. Convex order describes variability by ordering the expected values of convex functions, which are sensitive to the variation. Convex order has been used to describe variability in population models of infectious diseases [194] and for investigating the effect of variability in group size on the evolution of collective action [143]. In these contexts, convex order is used as a precise measure of variability.

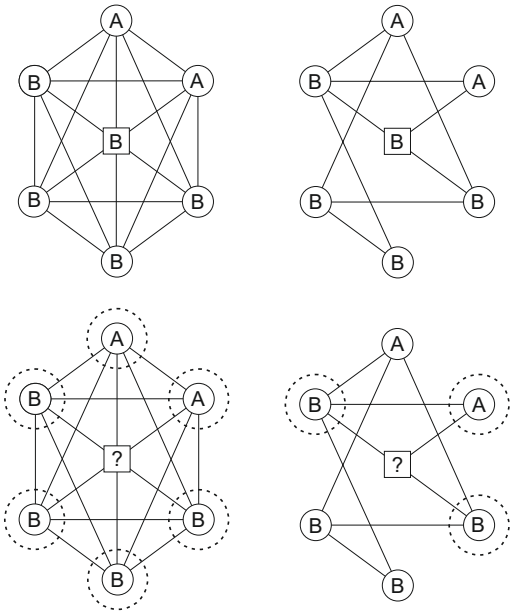

Figure 3.2: The selection neighbourhood. After an individual is selected for death (square node), the connected individuals (in dashed circles) compete to replace this individual. It can be seen that the different structures can influence the number of nodes in the selection neighbourhood and which nodes are included.

For two random variables X and Y, we say that X is less than Y in convex order (and therefore less variable than Y), denoted $X \leq_{cx} Y$, if and only if $\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)]$ for all convex functions ϕ. A useful result that can be obtained from convex ordering is [165]

$$X \leq_{cx} Y \implies \mathbb{E}[X] = \mathbb{E}[Y], \mathrm{Var}(X) \leq \mathrm{Var}(Y),$$

so if one random variable is less than another in convex order then its variance cannot be larger than the other. Establishing convex order can be difficult, but there are methods for doing this and under certain circumstances, this ordering of distributions reduces to an ordering of the variance of the distributions [165]. However, for our purposes, we only need to use this as a precise ordering of variability between any two distributions.

In the evolutionary process, selective pressure is governed by the selection probabilities on the right-hand side of Equations (3.1) and (3.2). For any given replacement event, the selection group consists of m bet-hedgers and n normal-types, so the selection probability

depends on m and n. The bet-hedger selection probability can be shown (Appendix 3.A) to reduce to

$$P(A \text{ reproduces } | m \text{ type } A \text{ and } n \text{ type } B) = \mathbb{E}\left[\frac{\sum_{i=1}^{m} f_A^i}{\sum_{i=1}^{m} f_A^i + \sum_{j=1}^{n} f_B^j}\right], \tag{3.3}$$

where f_A^i is the fitness of a bet-hedger i and f_B^j is the fitness of a normal-type j. Here, the bet-hedgers in the selection group (immediate neighbours of the individual selected for death) are labelled from 1 to m and the normal-types are labelled from 1 to n, so that $m+n = k$ where k is the size of the selection group. Noting that the selection probability of a normal-type among m bet-hedgers and n normal-types is 1 minus the selection probability of a bet-hedger among m bet-hedgers and n normal-types, the strength of selection can be represented solely by the bet-hedger selection probability. Equation (3.3) is the expected value of a convex function of normal-type fitness. Therefore, by the definition of convex order, increasing the variation of normal-type fitness in convex order can only increase this function, and therefore can only increase the selection probability of the bet-hedger.

To obtain the selection probability (Equation 3.3), the bet-hedger fitness is averaged over the total fitness of the surrounding individuals. For within-generational variation, each individual can sample a different value at the same time. Therefore, as the selection group size increases, whilst maintaining a similar proportion γ of bet-hedgers to normal-types (such that $m/n \approx \gamma$), the sum over the normal-types becomes less sensitive to the variation in normal-type fitness. This can be seen by transforming the summations on the right-hand side of Equation (3.3) into sample averages and applying the law of large numbers. Consequently, the selection probability becomes less sensitive to this variation as we increase selection group size, and for large selection groups, selection for reduced within-generational variation is diminished. This explains the result in large well-mixed populations that within-generational bet-hedging should not evolve [55]. However, since the selection group depends on the degree of the node in which an individual is chosen for death (see Figure 3.2 for an illustration), if the degree of this node is low then within-generational variation can have a large impact on the selection probability, regardless of population size.

By taking small clusters of fully connected individuals, inter-connected with a sparse

number of edges, we can create metapopulations. In this case, if the cluster size is small, the selection group will be small and within-generational bet-hedging can evolve, agreeing with predictions from metapopulation models [97, 170, 171, 199].

For between-generational variation, the fitness distributions within each type always sample the same value. Therefore, there is no diminishing effect due to large selection groups, and there will be selection for between-generational bet-hedging for all graphs and population sizes. Since there is no significant impact of population structure on selection, and evolution of between-generational bet-hedging has been widely explored mathematically [55, 74, 173], we focus on within-generational bet-hedging for the remainder of this chapter, and hence will drop the prefixed within-generational in what follows.

3.1.3 Approximate result for the variance

For evolutionary bet-hedging models, approximate results are often derived using a second order Taylor approximation [49, 55, 152, 153, 170, 171, 173]. For this analysis, we assume that bet-hedgers have constant fitness, so that the selection probability is a function of normal-type fitness. This analysis however can be applied to variable bet-hedger fitness. Applying a second order Taylor expansion about the normal-type mean fitness yields (Appendix 3.B.1)

$$P(A \text{ reproduces} \mid m \text{ type } A \text{ and } n \text{ type } B) \approx \frac{mc}{(mc+nd)} + \frac{mnc}{(mc+nd)^3}\sigma^2 \tag{3.4}$$

where c is the fitness of the bet-hedging type, and d and σ^2 are the expected value and the variance of the normal-type fitness, respectively. This suggests that mean and variance are key measures controlling the selection probability. Therefore, increasing the variation through the variance is likely to increase the selection probability of the bet hedger. We again observe that selection for bet-hedging depends on the size of the selection group rather than population size, since the second term in Equation (3.4) diminishes with m and n.

For the bet-hedger to be favoured in a given replacement event, we require the selection probability of the bet-hedger in this event to be larger than the selection probability of the normal-type in the opposite event. That is, we require the bet-hedger selection probability from m bet-hedgers against n normal-types to be larger than the normal-type selection probability from m normal-types against n bet-hedgers. Using Equation (3.4),

the variance at which these are equal, which we call the critical variance, is approximated by (Appendix 3.B.2)

$$\sigma^2_{k,m} \approx \frac{(dm+nc)^2(dn+mc)^2(d-c)}{((d^2m^3+d^2n^3-dm^3c+3dm^2nc+3dmn^2c-dn^3c+m^3c^2+n^3c^2)c)}. \tag{3.5}$$

From the derivative of Equation (3.5) with respect to m (Appendix 3.B.2), we see that increasing m from 1 to $k/2$ causes the critical variance to increase. This implies that the critical variance is largest in the evenly-mixed scenario, $m = n = k/2$, where

$$\sigma^2_{k,k/2} \approx \frac{(d-c)(c+d)^2k}{(4c)}.$$

In this scenario, a linear increase in the critical variance is required as k increases, showing that critical variance quickly grows with selection group size.

3.1.4 Fixation

The results in sections 3.1.2 and 3.1.3 focus on the relative strength of each strategy for a given selection event. However, the evolutionary process consists of multiple selection events with different selection groups. Therefore, the fixation probability determines the overall strength of each strategy. Since increasing variation in normal-type fitness increases the relative strength of bet-hedgers, we assume that the bet-hedger fixation probability will also increase.

Assuming that fixation probability is increasing with normal-type variation, it is useful to find the level of variation at which the bet-hedger becomes favoured in the evolutionary process; i.e. the fixation probability of a bet-hedger invading a normal-type population is higher than the normal-type invading bet-hedgers. We call this the overall critical variation. Subject to the assumptions of the Taylor approximation, there is a level of variation in the normal-type fitness at which the bet-hedger is favoured for a given selection event. We assume that there is also a critical variation for arbitrary fitness distributions. For certain distributions, such as the gamma distribution, this can be calculated using numerical methods (e.g. Appendix 3.C.1). Alternatively, this critical variation can be approximated by setting the variance of the distribution to be given by Equation (3.5). Since there is a level of variation at which, for a given selection group, the bet-hedger will be favoured, eventually the bet-hedger will be favoured in every selection group. Therefore, the overall

critical variation must exist.

On random k-regular graphs, the conclusions from the selection probability can easily be applied to fixation probability, since each selection group is of size k. Here, if the bet-hedger is favoured in every scenario for a size k selection group, they must be favoured overall. From the Taylor approximation, if the bet-hedger is favoured in the evenly-mixed scenario then the bet-hedger is favoured in every scenario, so we assume that the critical variation for the evenly-mixed scenario provides an upper bound for the overall critical variation. Extending the conclusions to arbitrary graphs is less clear, since there can be variability in node-degree. In such cases, it is not obvious how the cumulative effects of different selection group sizes will affect the relative strength of bet-hedgers. To investigate this, and to confirm that our assumptions hold, we investigate the fixation probability numerically in the next section.

3.2 Numerical results

Here we investigate the fixation probability numerically using stochastic simulations, in order to test the hypotheses from Section 3.1.4. Bet-hedgers have constant fitness and normal-type fitness is given by a gamma distribution. The gamma distribution is bounded below by zero and here convex ordering reduces to ordering the variance. Therefore, this makes sense as a fitness distribution and variation is easily controlled.

3.2.1 Regular graphs

To investigate the effect of variation on fixation, consider four 50 node graphs: a complete graph, and three random k-regular random graphs, with degrees 16, 8 and 4. To determine which strategy is stronger, the ratio of the bet-hedger fixation probability to the normal-type fixation probability is calculated. The overall critical value for each graph is given by the variation at which this ratio is equal to 1. We compare this to the upper bounds predicted by the evenly-mixed critical value, calculated exactly (see Appendix 3.C.1) and using the Taylor approximation.

It can be seen in Figure 3.3 that increasing the variation increases the bet-hedger fixation probability and decreases the normal-type fixation probability for all graphs, supporting the assumption from Section 3.1.4. Since the selection probability has large sensitivity to the degree of the graph (Equation (3.3)), for graphs with high degree, such as the com-

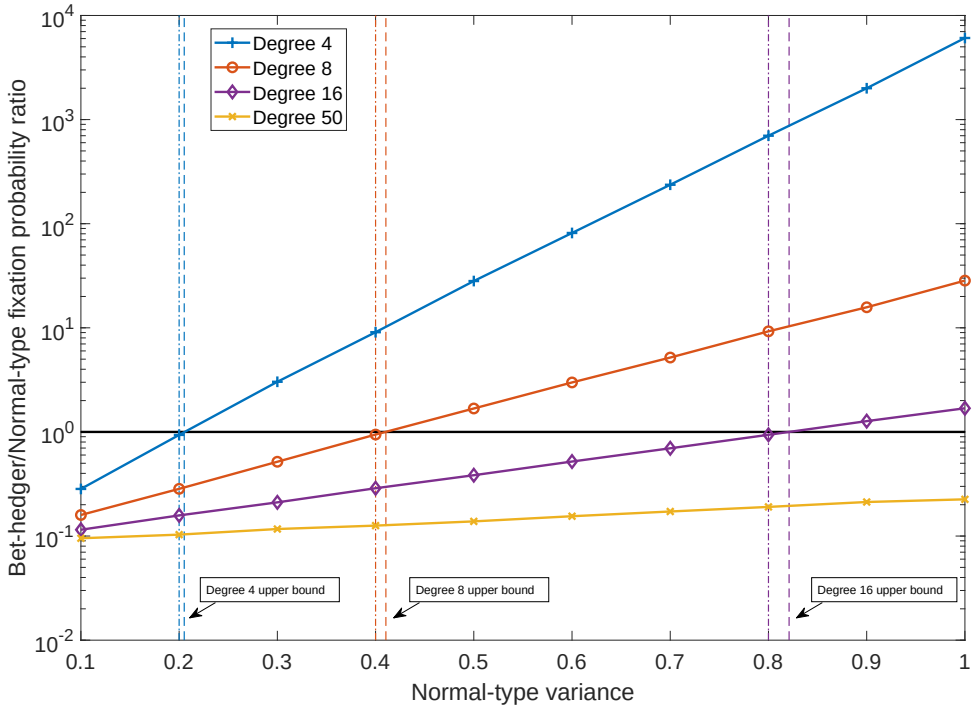

Figure 3.3: The impact of changing the within-generational variation in normal-type fitness on the overall relative strength of bet-hedgers on four 50 nodes graphs. Bet-hedgers have constant fitness with mean equal to 0.95, whereas normal-type fitness is fitness drawn from a gamma distribution, with mean equal to 1. Starting with a randomly placed initial bet-hedger, 1,000,000 simulations are run until fixation of either type. From this, the probability that the bet-hedger type takes over the population is calculated, giving the bet-hedger fixation probability. We repeat with a single normal-type invading a bet-hedger, and plot the ratio of bet-hedger fixation probability to the normal-type fixation probability (the solid lines with markers). If this ratio is below 1 (indicated by the horizontal line) the normal-type is favoured, otherwise the bet-hedger is favoured. The overall critical variance is the x-coordinate when each ratio crosses 1. The exact upper bounds are indicated by the dashed vertical lines and the approximate upper bounds are marked by the dash-dotted vertical lines.

plete graph, variation has little impact on the selection probability. Therefore, here the normal-type variance only has a slight impact on the fixation probability. This agrees with the traditional results obtained for well-mixed populations [55, 162]. However, reducing the average degree of the graph significantly increases the impact of variation on the fixation probability. Therefore, on such graphs, variation can play a key part in the evolutionary process.

The evenly-mixed critical value provides an upper bound on the overall critical varia-

tion, as assumed in Section 3.1.4. The upper bound provides a good approximation to the overall critical value for each graph tested, which we can use to gain insight into how much variation is required for the bet-hedger to be favoured in random k-regular graphs. Under the gamma distribution, the evenly-mixed critical value linearly increases with selection group size (Appendix 3.C.2), showing that increasing selection group size can quickly suppress selection for bet-hedging strategies. Comparing the Taylor approximation to the true upper bound, we observe that this provides a rough approximation to the true bound when the variance required for the bet-hedger to be favoured is low. However, as the variance required increases (in this case by increasing selection group size) the discrepancy between the two increases.

3.2.2 Impact of degree heterogeneity

Changing the average degree of the graph changes the strength of selection for reduced variation on random k-regular random graphs, which have no variability in the degree of different nodes. However, it is important to consider how degree variability will affect the evolutionary process.

To investigate the effect of degree variability, first consider the star graph and the circle, which have the same average degree in the limit of large population size. On the star graph, there is one focal individual who is connected to every other individual, who are only connected to the focal individual. This graph has high degree variability, with one node having degree $N-1$ and all others having degree 1, where N is the population size. On the circle, individuals are connected in a loop, so all nodes have degree 2 and there is no variability.

On the star graph selection only takes place when the individual in the central node is replaced (Appendix 3.D.1), and the size of the selection group is $N-1$. When the leaf node individuals are replaced, only the central node can be chosen, so no selection takes place. Therefore, the bet-hedger can only be favoured if it is favourable over the normal-type with selection group size $N-1$. In this case, selection for reduced variation quickly diminishes with population size. On the circle the selection group size for every replacement event is 2 (Appendix 3.D.2). Therefore the bet-hedger only needs to be favoured in a size 2 selection group, and there is no diminishing selection with population size.

Although an extreme case, this shows that increasing the variability in the degree distribution of the graph may reduce the selection for bet-hedging. We investigate this nu-

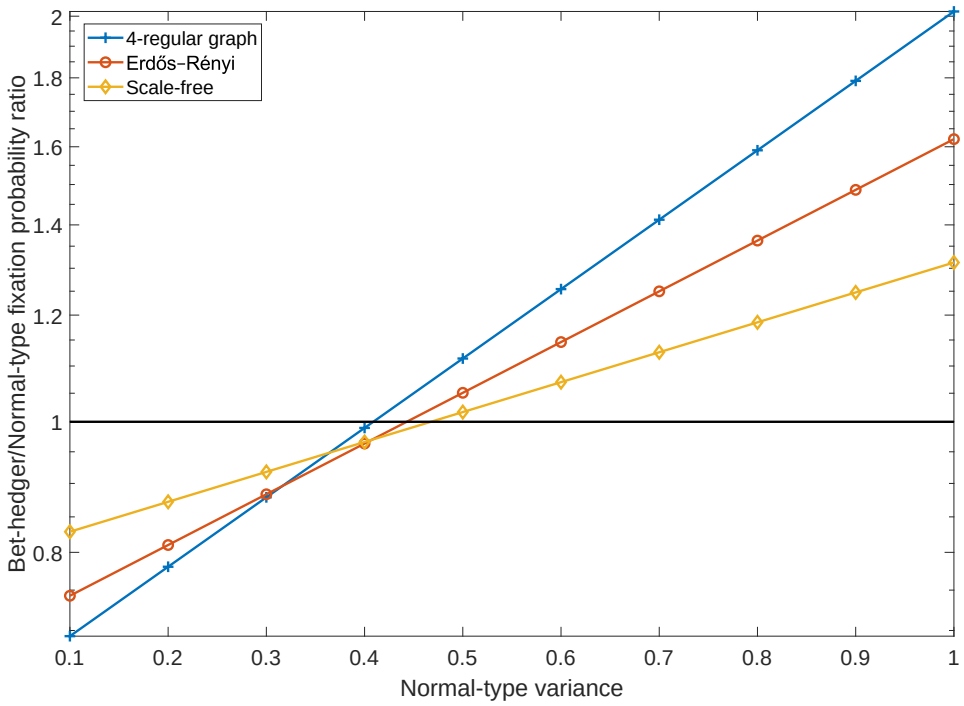

Figure 3.4: The impact of changing the variation in the degree distribution of the graph, whilst keeping the average degree unchanged, on selection for within-generational bet-hedging. Bet-hedgers have constant fitness with mean equal to 0.99, whereas normal-type fitness is drawn from a gamma distribution, with mean equal to 1. Starting with a randomly placed initial bet-hedger, 1,000,000 simulations are run until fixation of either type. From this, the probability that the bet-hedger type takes over the population is calculated, giving the bet-hedger fixation probability. We repeat with a single normal-type invading a bet-hedger, and plot the ratio of bet-hedger fixation probability to the normal-type fixation probability (the solid lines with markers). If this ratio is below 1 (indicated by the horizontal line) the normal-type is favoured, otherwise the bet-hedger is favoured. The overall critical variance is the x-coordinate when each ratio crosses 1.

merically using three different classes of random graph: random k-regular, Erdős-Rényi [43] and scale-free. Keeping the average degree constant, degree variability will be lowest for the random k-regular and largest for the scale-free. Figure 3.4 shows that increasing the degree variability increases the overall critical variation, hence reducing selection for bet-hedgers.

3.2.3 General evolutionary dynamics

This chapter has focused on death-birth with selection on birth dynamics. However, other evolutionary dynamics have been suggested for evolution in structured populations, such as birth-death with selection on death [6]. Under these dynamics, if fitness is taken to be a measure of survivability rather than birth rate, then adding population structure allows bet-hedging against within-generational variation to take place. This is because here the selection groups in Equation (3.3) only depend on local competition between the immediate neighbours. There are also dynamics that have global rather than local competition, such as the invasion process (birth-death with selection on birth) [102] and death-birth with selection on death [106]. Within such global update mechanisms, evolution does not select for bet-hedging against within-generational variation. This is because global competition results in the selection probabilities always involving every individual within the population, so the effect is diminished by the law of large numbers. Therefore, it can be seen that within-generational bet-hedging is facilitated by local competition between subsets of the population.

3.3 Discussion

Evolutionary bet-hedging explores how variation in fitness can affect the evolutionary process, showing that individuals with lower mean fitness can be preferred when their fitness is sufficiently less sensitive to these variations. A key area of discussion within evolutionary bet-hedging is the existence of strategies that potentially bet-hedge against within-generational variation; i.e. variations that affect individuals of the same type differently within each generation. Such strategies have been observed [48, 158, 160, 188, 189, 198], however mathematical theory has widely challenged their existence, instead suggesting that evolution should not select for this type of variation.

Traditional work has been limited to well-mixed populations. Real-populations however often exhibit some degree of structure, and this structure can have a significant impact on the evolutionary process. These impacts include amplifying the probability of advantageous mutants taking over a population [102] and facilitating the evolution of cooperative strategies in social dilemmas [129]. Using metapopulation structure has demonstrated that within-generational variation can be important when the patches are sufficiently small [97, 170, 171, 199]. By analysing bet-hedging strategies in generally structured pop-

ulations, we have shown that within-generational variation can be a key factor in selection, and strategies that bet-hedge against such variation can be favoured in the evolutionary process, regardless of population size.

Whether bet-hedging is favoured depends strongly on the structure of the population. We have shown that as long as the average degree of the graph is reasonably low and degree variability is not too high, selection for within-generational bet-hedging is strong. In such populations, bet-hedging strategies are likely to evolve, underpinning the results of some ecologists who have used bet-hedging against within-generational variation to explain observed strategies [158, 160, 188, 189]. Many real-world population structures will have these properties, such as parasitic waps [190] and plants [76], since individuals compete with a subset of the whole population and there will not be huge variability in the size of competing groups. Since our result only depends on the local competition aspect of the dynamics, it is reasonable to extend our conclusions to real-world evolutionary processes in which competition happens between small subsets of the population at any given time. Therefore, within-generational variation is important in empirical systems and within-generational bet-hedging is likely to be observed. This justifies the existing observations and can motivate further empirical research to identify within-generational bet-hedging species, which have not been fully explored, perhaps due to the existing theoretical conclusions [97, 160].

Appendix 3.A Selection probability

Given that an individual has been selected for death, one of the neighbours of this individual must be selected for birth, with probability one. We are therefore interested in deriving the probability that the selected individual is a bet-hedging type. We can assume without loss of generality that there are m bet-hedgers and n normal-type neighbours around the focal individual. We can label the bet-hedgers arbitrarily from 1 to m and the normal-types arbitrarily from 1 to n, such that bet-hedger i has fitness distribution f_A^i and normal j has fitness distribution f_B^j. The individuals in the neighbourhood compete and a random individual is selected with probability proportional to their fitness to reproduce.

Once we have sampled the fitness values for each individual, the total fitness is given by $\sum_{i=1}^{m} f_A^i + \sum_{j=1}^{n} f_B^j$, and therefore the probability of selecting bet-hedger i is $f_A^i / (\sum_{i=1}^{m} f_A^i + \sum_{j=1}^{n} f_B^j)$,

and normal j is $f_B^j/(\sum_{i=1}^{m} f_A^i + \sum_{j=1}^{n} f_B^j)$. Therefore the probability of selecting any bet-hedger is given by

$$\frac{\sum_{i=1}^{m} f_A^i}{\sum_{i=1}^{m} f_A^i + \sum_{j=1}^{n} f_B^j}. \tag{3.6}$$

When selecting which individual reproduces, we draw a random number from the uniform distribution between 0 and 1. If this number is smaller than Equation (3.6) then we select a bet-hedger, otherwise we select a normal. Therefore, the selection probability of a bet-hedger, before sampling any of the fitness values, is given by

$$P(A \text{ reproduces } | m \text{ type } A \text{ and } n \text{ type } B) = P\left(N < \frac{\sum_{i=1}^{m} f_A^i}{\sum_{i=1}^{m} f_A^i + \sum_{j=1}^{n} f_B^j} \right),$$

where $N \sim U(0,1)$. Defining $Y = \sum_{j=1}^{n} f_B^j$ and $X = \sum_{i=1}^{m} f_A^i$, we have

$$\begin{aligned}
P\left(N < \frac{X}{X+Y}\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\frac{x}{x+y}} f_{N,X,Y}(n,x,y)\,dn\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{0}^{\frac{x}{x+y}} f_{N|X,Y}(n|x,y)\,dn\right) f_{X,Y}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x}{x+y}\right) f_{X,Y}(x,y)\,dx\,dy \\
&= \mathbb{E}\left[\frac{X}{X+Y}\right].
\end{aligned}$$

Appendix 3.B Taylor approximation

3.B.1 Selection probability approximation

Let X be any random variable and f be an infinitely differentiable function of X. We can approximate the expected value of $f(X)$ by performing a Taylor expansion:

$$\begin{aligned}\mathbb{E}\left[f(X)\right] &= \mathbb{E}\left[f(\mu_X + (X-\mu_X)\right] \\ &\approx \mathbb{E}\left[f(\mu_X) + f'(\mu_X)(X-\mu_X) + \frac{1}{2}f''(\mu_X)(X-\mu_X)^2\right],\end{aligned}$$

where μ_X is the expected value of X. Since $\mathbb{E}\left[(X-\mu_X)\right] = 0$, this simplifies to

$$\mathbb{E}\left[f(\mu_X)\right] \approx f(\mu_X) + \frac{1}{2}f''(\mu_X)\sigma_X^2,$$

where σ_X^2 is the variance of X.

If we define $Y = \sum_{j=1}^{n} f_B^j$ and $X = \sum_{i=1}^{m} f_A^i$, and assume that X is constant then the selection probability is simply the expected value of a function of Y, $f(Y) = \frac{X}{X+Y}$, and we can approximate this as

$$\mathbb{E}\left[\frac{X}{X+Y}\right] \approx \frac{\mu_X}{\mu_X+\mu_Y} + \frac{1}{2}\frac{\mu_X}{(\mu_X+\mu_Y)^3}\sigma_Y^2.$$

Since $Y = \sum_{j=1}^{n} f_B^j$ and $X = \sum_{i=1}^{m} f_A^i$, $\mu_X = mc$, $\mu_Y = nd$ and $\sigma_Y^2 = n\sigma_{f_B^i}^2$.

3.B.2 Critical variance

Assume we have m bet-hedgers in the selection group. We want to know when a bet-hedger would have a higher relative strength in this scenario, and therefore we want to compare the selection probability $P(A|m)$ to the selection probability of a normal-type when there are m normal-types in the selection group, $P(B|k-m) = 1 - P(A|k-m)$. When these two probabilities are equal, the two types are equally favoured in such a scenario. Finding the variance for which these two are equal therefore gives the critical variance at which the bet-hedger becomes stronger in this scenario. Setting $P(A|m) = P(B|k-m)$ under the

Taylor approximation we get

$$\frac{mc}{mc+nd}+\frac{mcn\sigma^2}{(mc+nd)^3}=1-\frac{nc}{nc+md}+\frac{ncm\sigma^2}{(nc+md)^3},$$

which we can solve for σ^2 to obtain

$$\sigma^2_{k,m}=\frac{(dm+nc)^2(dn+mc)^2(d-c)}{(d^2m^3+d^2n^3-dm^3c+3dm^2nc+3dmn^2c-dn^3c+m^3c^2+n^3c^2)c}.$$

This gives the critical variance at which the bet-hedging type is stronger, when there are A bet-hedgers competing among k individuals.

In this expression, $n=k-m$, so this can be written in terms of k and m only. We can treat k as fixed since this is the selection group size, and therefore we can investigate how changing m affects the critical variance. We first note that we are only interested in $m\in\{1,...,k/2\}$, since for $m>k/2$ these scenarios are already considered to obtain the critical variance, i.e. for $m>k/2$ the critical variance is the same as for $m=k-k/2<k/2$.

Taking the derivative with respect to m of the critical variance, and assuming k is constant, we obtain

$$f(m)=\frac{\partial\sigma^2_{k,m}}{\partial m}=\frac{g(m)h(m)}{z(m)}$$

where

$$\begin{aligned}
g(m) &= \frac{3(d-c)^2m^2}{2}-\frac{3k(d-c)^2m}{2}+k^2(d^2+\frac{1}{2}dc+c^2),\\
h(m) &= 2((-d+c)m+dk)(d-c)^3((d-c)m+ck)(k-2m),\\
z(m) &= c(3(d-c)^2m^2-3k(d-c)^2m+k^2(d^2-dc+c^2))^2k.
\end{aligned}$$

Both $h(m)$ and $z(m)$ are strictly positive for all m (when $d>c$) so to investigate how changing m alters the sign of $f(m)$ we only need to analyse $g(m)$. Taking the derivative of $g(m)$ with respect to m of this we obtain

$$g'(m)=3(d-c)^2m-\frac{3k(d-c)^2}{2}$$

which is negative for $m\leq k/2$. Therefore the minimum value of $g(m)$ for $m\in\{1,...,k/2\}$

occurs at $k/2$ and is equal to

$$g(\frac{k}{2}) = \frac{5k^2(c+d)^2}{8}$$

which is positive. Since this is the minimum value, $g(m)$ must be positive for all $m \in \{1, ..., k/2\}$, and therefore $f(m) > 0$ for all $m \in \{1, ..., k/2\}$. This implies that $\sigma^2_{k.m}$ is an increasing function of m for $m \in \{1, ..., k/2\}$ and therefore the maximal value of $\sigma^2_{k,m}$ occurs at $m = k/2$. Therefore if the bet-hedger is favoured when the selection group consists of $k/2$ bet-hedgers and $k/2$ normal-types then the bet-hedger is favoured in every scenario.

Appendix 3.C Gamma distribution properties

3.C.1 Calculating the upper bound

To calculated the upper bound exactly we wish the find the variation (which is equivalent to variance for the gamma distribution) at which the probability of selecting a bet-hedger to reproduce in the evenly-mixed scenario; i.e. $k/2$ bet-hedger versus $k/2$ normal-types, is equal to $1/2$.

If we assume that the bet-hedger has constant fitness c and that the normal-type fitness is drawn from a gamma distribution, $f^j_B \sim \Gamma(d/\theta, \theta)$, then we need to find the value of θ which satisfies

$$\int_0^\infty \frac{mc}{mc+z} T \frac{1}{\Gamma(\frac{md}{\theta})\theta^{\frac{md}{\theta}}} z^{\frac{md}{\theta}-1} e^{-\frac{z}{\theta}} dz = \frac{1}{2}.$$

To solve this we define

$$f(\theta) = \int_0^\infty \frac{mc}{mc+z} T \frac{1}{\Gamma(\frac{md}{\theta})\theta^{\frac{md}{\theta}}} z^{\frac{md}{\theta}-1} e^{-\frac{z}{\theta}} dz - \frac{1}{2}.$$

Solving $f(\theta) = 0$ analytically is challenging due to the complex nature of the integral. However, for given values of θ it is easy to solve the integral numerically, since $m = k/2$, c and d are all known. Therefore we can construct a minimisation problem, where we aim to minimise the function $|f(\theta)|$. Since the integral is an increasing function of θ, which we know from the definition of convex order, $f(\theta)$ is an increasing function and therefore $|f(\theta)|$ does not have local minima, and therefore minimising this function can be efficiently implemented in Matlab (or other language) to find the true critical value for θ.

3.C.2 Linear upper bound

Our results from the Taylor approximation suggest that the upper bound on the overall critical variance required for the bet-hedger to be favoured is given by the evenly mixed scenario; i.e. $m = n = k/2$. Here we explore how changing selection group size impacts the critical variance for this scenario (and therefore the upper bound) using the exact selection probability, when the normal-type fitness is drawn from a gamma distribution. For the gamma distribution convex order reduces to ordering the variance of the distributions, so we can represent the critical variation using the critical variance. We assume that bet-hedgers have constant fitness equal to c and normal-types have fitness $f_B^j = Y_j \sim \Gamma(d/\theta, \theta)$, such that the normal-type have mean fitness d and variance in fitness given by $d\theta$. The selection probability of a bet-hedger is given by

$$P(A|m) = \mathbb{E}\left[\frac{mc}{mc + \sum_{j=1}^{m} Y_j}\right].$$

Since each Y_j is independent and identically distributed we can write $Z = \sum_{j=1}^{m} Y_j \sim \Gamma(md/\theta, \theta)$. Therefore we can now write the selection probability as

$$P(A|m) = \mathbb{E}\left[\frac{mc}{mc+Z}\right] = \int_0^\infty \frac{mc}{mc+z} \frac{1}{\Gamma(\frac{md}{\theta})\theta^{\frac{md}{\theta}}} z^{\frac{md}{\theta}-1} e^{-\frac{z}{\theta}} dz.$$

The critical variance for the evenly mixed scenario, $\sigma^2_{k,m} = d\theta_{k,m}$, needs to satisfy

$$\int_0^\infty \frac{mc}{mc+z} \frac{1}{\Gamma(\frac{md}{\theta_{k,m}})\theta_{k,m}^{\frac{md}{\theta_{k,m}}}} z^{\frac{md}{\theta_{k,m}}-1} e^{-\frac{z}{\theta_{k,m}}} dz = \frac{1}{2},$$

where $m = n = k/2$. Changing the selection group size to hk, we are interested in the selection probability with $x = hm$ bet-hedgers. The critical variance $d\theta_{hk,hm}$ needs to satisfy

$$\int_0^\infty \frac{hmc}{hmc+z} \frac{1}{\Gamma(\frac{hmd}{\theta_{hk,hm}})\theta_{hk,hm}^{\frac{hmd}{\theta_{hk,hm}}}} z^{\frac{hmd}{\theta_{hk,hm}}-1} e^{-\frac{z}{\theta_{hk,hm}}} dz = \frac{1}{2}.$$

Using a change of variable $z' = z/h$ this becomes

$$\int_0^\infty \frac{hmc}{hmc + hz'} \frac{1}{\Gamma(\frac{hmd}{\theta_{hk,hm}})\theta_{hk,hm}^{\frac{hmd}{\theta_{hk,hm}}}} (hz')^{\frac{hmd}{\theta_{hk,hm}}-1} e^{-\frac{hz'}{\theta_{hk,hm}}} h dz' = \frac{1}{2}.$$

Now if we set $\sigma^2_{k,hm} = d\theta_{k,hm} = dh\theta_{k,m} = h\sigma^2_{k,m}$ then this reduces to

$$\int_0^\infty \frac{mc}{mc + z'} \frac{1}{\Gamma(\frac{md}{\theta_{k,m}})\theta_{k,m}^{\frac{md}{\theta_{k,m}}}} (z')^{\frac{md}{\theta_{k,m}}-1} e^{-\frac{z'}{\theta_{k,m}}} dz' = \frac{1}{2},$$

which we know holds, and therefore for the gamma distribution the critical variance in the evenly mixed scenario increases linearly with the size of the selection group.

Appendix 3.D Idealised graphs transition probabilities

3.D.1 Star transition probabilities

On the star there are a few possible distinct transitions that can occur. If a bet-hedger is in the central node, the changes that can happen to the system are either the individual in this node dying and being replaced by a normal-type from the leaf nodes, or a normal-type individuals on a leaf node can die and be replaced by a bet-hedger from the central node. Similarly, if the central node is a normal-type either this individual can die and be replaced by a bet-hedger from a leaf node, or a bet-hedger leaf node individual can die and be replaced by a normal-type from the central node. Due to the symmetry of the leaf nodes, we can group all of these events together so that we only have these four possible state transitions to consider. Let $p^{AA}_{i,i+1}$ denote the probability that we go from a state with a bet-hedger in the central node and i bet-hedgers on the leaves to a state with A on the central node and $i+1$ bet-hedgers on the leaves. Let $p^{AB}_{i,i}$ denote the probability that we go from a state with a bet-hedger in the central node and i bet-hedgers on the leaves to a state with a normal-type in the central node and i bet-hedgers on the leaves. The system transitions is described by the following system of equations

$$p^{AA}_{i,i+1} = \frac{n-i}{n+1},$$

$$
\begin{aligned}
p_{i,i}^{AB} =& \frac{1}{n+1}\mathbb{E}\left[\frac{\sum_{j=1}^{n-i} f_B^j}{\sum_{j=1}^{i} f_A^j + \sum_{j=1}^{n-i} f_B^j}\right], \\
p_{i,i-1}^{BB} =& \frac{i}{n+1}, \\
p_{i,i}^{BA} =& \frac{1}{n+1}\mathbb{E}\left[\frac{\sum_{j=1}^{i} f_A^j}{\sum_{j=1}^{i} f_A^j + \sum_{j=1}^{n-i} f_B^j}\right].
\end{aligned}
$$

Selection is only taking place when an individual on a leaf node replaces the central node. This is because when a leaf node individual dies only the central node can replace this, so there is no competition based selection taking place. When we look at the selection probability for the central node being replaced it is clear that the selection group has size n, which is equal to $N-1$ for population size N. Therefore, if we consider a sufficiently large population the impact of within-generational variation on the selection probability rapidly diminishes due to the law of large numbers, and therefore there is no selection for within-generational bet-hedging on a large star graph.

3.D.2 Circle transition probabilities

On the circle, if we assume that we start with a cluster of connected bet-hedger individuals with no normal-type individuals between them, the symmetries of the graph allow us to consider the number of bet-hedger individuals rather than their locations, since the group of bet-hedgers can only change at the two boundaries where they meet normal-types. Denoting $p_{i,i+1}$ as the probability of moving from a state with i bet-hedgers to $i+1$ bet-hedgers, we can describe the transition probabilities between different states with the following equations

$$
\begin{aligned}
p_{1,0} =& \frac{1}{N}, \\
p_{i,i+1} =& \frac{2}{N}\mathbb{E}\left[\frac{f_B^j}{f_A^j + f_B^j}\right], i < N-1
\end{aligned}
\tag{3.7}
$$

$$p_{i,i-1} = \frac{2}{N}\mathbb{E}\left[\frac{f_B^j}{f_A^j + f_B^j}\right], i > 1$$
$$p_{N-1,N} = \frac{1}{N},$$

where j is arbitrary since the fitness distributions for a given type are independent and identically distributed. From this we can see that on the circle there will always be selection for a reduction in within-generational variation, potentially paving the way for a within-generational bet-hedging strategy to evolve.

Chapter 4

Methods for approximating stochastic evolutionary dynamics on graphs

The dynamics of evolutionary graph theory can be considered analytically and precise results can be derived for a number of simple graphs, such as the circle, star and complete graphs [19, 20, 102], mainly due to their symmetry. Analytic approaches for investigating evolutionary dynamics on complex graphs have also been proposed. However, such methods are usually limited by assumptions such as large populations [127, 129] or are specifically designed for investigating evolutionary processes under weak selection [2, 203], where the evolutionary game has only a small effect on reproductive success.

Important quantities of interest such as the exact fixation probability and time can, in principle, be obtained by solving the discrete-time difference equations of the underlying stochastic model [68], although this is only feasible for very small populations unless there are simplifying symmetries. Individual-based stochastic simulations [12, 105] provide numerically accurate representations of the evolutionary process on arbitrary graphs but have limited scope for generating conceptual insights into the dynamics on their own. They can also be computationally expensive on larger graphs, but as a precise representation of the underlying stochastic model, they allow us to evaluate the accuracy of approximate models by comparison.

In this chapter, we develop approximations to the stochastic model by using insights

from methods in statistical physics that have also been used extensively for epidemic modelling [18, 83, 88, 142, 166, 168]. Such methods have been applied to develop pair approximations for evolutionary processes on graphs which satisfy the homogeneity assumption that all individuals can be considered identical and interchangeable [59, 67, 117, 144, 175]. However, the underlying assumptions linking these models to the stochastic dynamics are not always clear. One contribution of this work is to derive these models explicitly by identifying the required assumptions. The starting point for all of our approximations is to derive an equation to describe the time-evolution of the state of any given individual node. From this equation, various routes to approximation become apparent by applying different assumptions. We then investigate the applicability and accuracy of the resulting approximation methods.

Evolutionary graph theory is traditionally explored as a discrete-time stochastic model. While it is possible to work with these dynamics, it is easier to work with a continuous-time approximation to the process. The continuous-time system is represented by a master equation describing how the probability of being in each system state changes. From the master equation we obtain exact equations (with respect to the continuous-time process) for the probabilities of the states of individual nodes (Theorem 4.1.1). These equations can then be approximated by adopting moment-closure methods. We focus on evaluating the probability that at the end of the evolutionary process, an initial subset of mutants placed on the graph will take over the whole population and 'fixate'. Using this continuous-time system is justified because the fixation probability and expected time to fixation are identical to those of the original discrete-time process. Within this framework we study when accurate approximations can be derived.

In Sections 4.1.1-4.1.3 we recap the stochastic evolutionary dynamics and the master equation, and derive a description of how node-level quantities change in the master equation. We then discuss and develop various techniques that can be used to approximate these systems of equations in Section 4.2. Within these approximation frameworks, we derive the pair approximation models used in the literature, which we will call the homogenised pair approximation, and the exact neutral drift model, and build new node-level approximation methods. In Section 4.3 we demonstrate how the different methods can be used to approximate the dynamics of the original discrete-time process. Section 4.3.1 studies how the methods perform when approximating the fixation probability of a single initial mutant placed on idealised and on complex graphs. Section 4.3.2 then shows how the methods perform when studying the evolutionary game dynamics in a Hawk-Dove game.

In Section 4.4 we discuss the results obtained from the methods developed and the insights these can give.

4.1 The stochastic model

4.1.1 Stochastic evolutionary dynamics

In this Chapter we consider the stochastic dynamics of evolutionary graph theory, as described in Section 1.3.2. Here, we consider a population whose relationship structure is represented by a strongly connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{1, 2, ..., N\}$ is the set of nodes and $\mathcal{E}$ denotes the set of edges. This can be represented by an adjacency matrix G, where $G_{ij} = 1$ if j is connected to i, and $G_{ij} = 0$ otherwise, with $G_{ii} = 0$ for all $i \in \mathcal{V}$. We consider populations consisting of two types of individuals, type A and type B, either of which can be in the role of invading mutant in a resident population. Each node is occupied by either an A or a B individual. Therefore we can let $A_i = 1$ if and only if node i is occupied by an A individual and $A_i = 0$ otherwise and let B_i denote the same for individuals of type B. Since $B_i = 1 - A_i$, the state of the system can be represented by the values of A_i at any given time.

To compare the relative strength of individuals, we need to define their fitness. This can be done as in Section 1.3.3, where the fitness of a type A in node j is given by

$$f_A^j = 1 - w + w\frac{a\sum_{i=1}^{N} G_{ij}A_i + b\sum_{i=1}^{N} G_{ij}B_i}{\sum_{i=1}^{N} G_{ij}}, \tag{4.1}$$

and of a type B in node j is given by

$$f_B^j = 1 - w + w\frac{c\sum_{i=1}^{N} G_{ij}A_i + d\sum_{i=1}^{N} G_{ij}B_i}{\sum_{i=1}^{N} G_{ij}}, \tag{4.2}$$

The special case of constant fitness is given by setting $w = 1$ and by setting $a = b = r$ and $c = d = 1$.

Traditional evolutionary graph theory considers a discrete-time Markovian evolutionary

process in which only one event can happen at each time step. When an event occurs, one individual reproduces and a connected individual dies, with the offspring replacing it. We refer to the mechanism by which this takes place as an update mechanism or rule. The probability of a certain event taking place depends upon this update mechanism. Some of the most commonly considered update mechanisms are birth-death with selection on birth (invasion process) [102], death-birth with selection on birth [106], birth-death with selection on death [6] and death-birth with selection on death (voter model) [129]. The methods developed in this chapter will be presented in the general case, and can be applied to any of the above update rules, but we shall focus on the invasion process when generating specific examples. In the invasion process, we select an individual to reproduce in proportion to their fitness (selection on birth) and then the offspring replaces a connected individual selected uniformly at random for death (birth then death).

4.1.2 The master equation

To approximate the discrete-time evolutionary process we first construct a continuous-time process that has the discrete-time process embedded (Section 1.1.7). To do this we model each (replacement) event using a Poisson process. The rate at which each event happens is equal to the probability of that event in the discrete-time model. Therefore, the total event pressure will be the sum of all such probabilities, which is equal to one, so that the time until the next event follows a Poisson process with rate parameter one. We then determine which event takes place using the relevant probability. Under this continuous-time system the fixation probability and expected time to fixation will be identical to those of the discrete-time system, since we use the same probabilities whenever an event occurs and the expected time between events is constant. This is important because these are the main quantities of interest in evolutionary dynamics.

We will use this system to build approximation methods to study the original discrete-time process. We choose to use continuous-time because it enables us to build a system of ordinary differential equations to approximate the dynamics, which allows us to make use of efficient numerical solvers and enables us to derive some analytic results.

Since this evolutionary process is a continuous-time Markov chain, we can construct a master equation (see Section 1.1.4) to describe the dynamics. Let $S_i = (s_1, s_2, ..., s_N)$ be a state of the system, where $i \in \{1, ..., 2^N\}$ and where $s_j = 1$ if node j is a type A individual and $s_j = 0$ otherwise. We define $S_1 = (0, 0, ..., 0)$ and $S_{2^N} = (1, 1, ..., 1)$ to be the states

consisting of only B individuals and only A individuals, respectively.

We introduce a vector $\mathbf{p}(t)$ which represents the probabilities of each system state at time t. That is, the ith entry of $\mathbf{p}(t)$, $p_i(t)$, is the probability that the system is in state S_i at time t. This Markovian evolutionary process has 2^N possible states and the transitions between them are governed by a $2^N \times 2^N$ transition rate matrix Q whose entries depend upon the graph and update mechanism we consider.

We write the rate of change in the state probabilities using the master equation of the Markov process:

$$\frac{d\mathbf{p}}{dt} = Q\mathbf{p}. \tag{4.3}$$

Such an equation can be constructed for any graph under a Markovian update mechanism. The absorbing states correspond to the all type B or all type A states, S_1 and S_{2^N}, so are given by p_1 and p_{2^N}.

Since we consider a strongly connected adjacency matrix G, provided we have at least one type A and one type B it is possible to get to either of the absorbing states and therefore from any mixed initial condition the system will always end up distributed between these two states. We define the fixation probability $\rho^A(S_i)$ of type A from an initial state S_i to be the probability of being in the all A absorbing state, that is

$$\rho^A(S_i) = \lim_{t\to\infty}(p_{2^N}(t)|p_i(0) = 1),$$

where $p_i(0)$ is the probability of being in the state S_i at time $t = 0$. Similarly we define the fixation probability of type B as

$$\rho^B(S_i) = \lim_{t\to\infty}(p_1(t)|p_i(0) = 1).$$

The computational cost of implementing system (4.3) increases exponentially with N [68], and thus the computation of the fixation probability becomes infeasible as the population size increases. Therefore it is of interest to build approximation methods. Pair approximations of the master equation have been developed under the homogeneity assumption that all nodes on the underlying graph are identical and interchangeable [67, 175], which can give interesting insight into the evolutionary dynamics. However the homogeneity assumptions made in these approximations result in the loss of insight into graph and node-specific dynamics, so we aim to develop approximations of the master equation which can capture this information.

4.1.3 Node-level equations

We approximate the master equation by approximating the dynamics of the state probabilities of individual nodes in the population. This is motivated by approaches in statistical physics and epidemic modelling [18, 88, 166, 168], and first requires exact equations describing how the probability of each node being occupied by a certain type changes with time, which can be derived from the master equation (4.3).

Definition 4.1.1. *Let $\chi(\Omega^t_{j\to i}|S^t)$ denote the rate at which the individual in node j replaces the individual in node i at time t given that the system is in state S at time t; we refer to this as the replacement rate.*

Definition 4.1.2. *X^t_C denotes the event that the set of nodes C is in state X at time t; for example $A^t_{\{i\}}$ is the event that node i is in the type A state at time t.*

Throughout this chapter we shall use the shorthand $B^t_{\{i\}}A^t_{\{j\}}X^t_C$ to represent the intersection of events $B^t_{\{i\}} \cap A^t_{\{j\}} \cap X^t_C$.

Theorem 4.1.1. *Under any Markovian update mechanism, for a structured population represented by the adjacency matrix G, the rate of change of the probability that the individual in node i is an A individual is*

$$\begin{aligned}\frac{dP(A^t_{\{i\}})}{dt} =& \sum_{j=1}^{N}\sum_{X_{V\setminus\{i,j\}}} G_{ij}P(B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j\}})\chi(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j\}}) \\ &- \sum_{j=1}^{N}\sum_{X_{V\setminus\{i,j\}}} G_{ij}P(A^t_{\{i\}}B^t_{\{j\}}X^t_{V\setminus\{i,j\}})\chi(\Omega^t_{j\to i}|A^t_{\{i\}}B^t_{\{j\}}X^t_{V\setminus\{i,j\}}),\end{aligned} \tag{4.4}$$

where the sum over $X_{V\setminus\{i,j\}}$ is over all possible states of the nodes $V\setminus\{i,j\}$.

Proof. See Appendix 4.A.

This theorem can be applied to any update mechanism by choosing an appropriate definition for the replacement rate, $\chi(\Omega^t_{j\to i})$, which we shall define for the invasion process as an example.

Example 4.1.1 (Invasion process)**.** *The invasion process (described in Section 1.3.2 and Figure 1.10) is an adaptation of the Moran process [116] to structured populations. Each*

event is determined by selecting an individual to reproduce with probability proportional to its fitness. It produces an identical offspring which replaces one of the connected individuals which is chosen uniformly at random. Therefore the rate at which the individual in node j replaces the individual in node i at time t under the invasion process rules is given by

$$\chi(\Omega^t_{j\to i}|S) = \frac{f^t_j|S}{F^t|S}\frac{1}{k_j}, \tag{4.5}$$

where f^t_j is the fitness of the individual occupying node j at time t, $F^t = \sum_{m=1}^{N} f^t_m$ is the total fitness of the population, and k_j denotes the degree of node j. Here, the factor f^t_j/F^t is the rate at which node j is selected to reproduce, and $1/k_j$ is the probability of replacing the neighbouring individual i which is selected uniformly at random.

When calculating $\chi(\Omega^t_{j\to i})$ in Equation (4.4), we will use the following expression for the fitness of the individual at a given node j at time t,

$$\begin{aligned} f^t_j = 1 - w + wP(A^t_{\{j\}})\frac{a\sum_{i=1}^{N} G_{ij}P(A^t_{\{i\}}) + b\sum_{i=1}^{N} G_{ij}P(B^t_{\{i\}})}{\sum_{i=1}^{N} G_{ij}} \\ + wP(B^t_{\{j\}})\frac{c\sum_{i=1}^{N} G_{ij}P(A^t_{\{i\}}) + d\sum_{i=1}^{N} G_{ij}P(B^t_{\{i\}})}{\sum_{i=1}^{N} G_{ij}}, \end{aligned} \tag{4.6}$$

which is a sum of equations (1.14) and (1.15) weighted by the node probabilities. We use this definition because when we evaluate Equation (4.6) given that the system is in a particular state S, as required by Equation (4.4), the values of $P(A^t_{\{k\}})$ and $P(B^t_{\{k\}})$ are either 1 or 0, which leads to the fitness of node j in that particular system state (Equations (1) and (2)). However, by defining fitness in terms of the node probabilities, this allows us to have a description of fitness which we can approximate (see Sections 4.2.2 and 4.2.3).

4.2 Approximating the stochastic model

In other fields, such as epidemiology, the construction of node-level equations such as Equation (4.4) can lead to a hierarchy of moment equations whereby these equations are

written in terms of pair probabilities, pairs are written in terms of triples and so on, until the full system state size is reached and the hierarchy is closed (see Section 1.4.2). This is useful when we can find appropriate closure approximations to close this hierarchy at a low order. However, we see that such an approach cannot be used here because we condition against the full system state in Equation (4.4) which means that the full system size appears even at the first order. We therefore attempt to find other methods to simplify this system of equations.

In this section, we will describe three different techniques to derive approximations for this system. The first technique yields a system of equations that becomes computationally infeasible in some circumstances, but by applying homogeneity assumptions to the underlying graph, we can derive the existing pair approximation models currently used in the literature [59, 67, 117, 144, 175] (Section 4.2.1). To reduce computation costs, we then develop methods based on restricting the number of states that we condition against in the replacement rate. We first obtain a method whose computational complexity scales linearly with the population size N and, after an appropriate scaling, approximates the fixation probability well on a wide range of graphs (Section 4.2.2). Then, in Section 4.2.3, we obtain a method which, although it scales with N^2, provides a good approximation to the evolutionary dynamics over the whole time series for various graphs, and in particular provides a very accurate approximation to the initial dynamics of the evolutionary process on all graphs.

4.2.1 Deriving the homogenised pair approximation model

One way of simplifying (4.4) is to assume that the fitness f_j^t does not need to be normalised by the total fitness F^t in the replacement rate (e.g. as in Equation (4.5) for the invasion process). This approximation is justified because it does not change the final value to which the exact node-level equations converge (and therefore the fixation probability), and will only transform the time series until fixation. Making this assumption, the node level equations simplify so that we only sum over the neighbours of the individual that we selected based on fitness. That is, when looking at the event where node j replaces node i, if we are selecting on death we need to condition against the state of all neighbours of i, and if selecting on birth we need to condition against the state of all neighbours of j. As an example, we shall assume here that selection occurs on birth so that we require conditioning on the neighbourhood of node j, however we can also make similar arguments

when selecting on death. Using $\bar{\chi}$ to represent this modification of χ in (4.4) and Q to represent the new probability distribution with the modified time series we obtain

$$\begin{aligned}\frac{dQ(A^t_{\{i\}})}{dt} =& \sum_{j=1}^{N} \sum_{X_{\mathcal{N}_j\setminus\{i\}}} G_{ij} Q(B^t_{\{i\}} A^t_{\{j\}} X^t_{\mathcal{N}_j\setminus\{i\}}) \bar{\chi}(\Omega^t_{j\to i} | B^t_{\{i\}} A^t_{\{j\}} X^t_{\mathcal{N}_j\setminus\{i\}}) \\ & - \sum_{j=1}^{N} \sum_{X_{\mathcal{N}_j\setminus\{i\}}} G_{ij} Q(A^t_{\{i\}} B^t_{\{j\}} X^t_{\mathcal{N}_j\setminus\{i\}}) \bar{\chi}(\Omega^t_{j\to i} | A^t_{\{i\}} B^t_{\{j\}} X^t_{\mathcal{N}_j\setminus\{i\}}), \end{aligned} \tag{4.7}$$

where $\mathcal{N}_j$ is the neighbourhood of node j; i.e. all nodes that are connected to j. To solve this system exactly requires the development of equations describing how the probability of each possible neighbourhood of nodes changes. This in turn would lead to a hierarchy of equations which is computationally similar to the master equation. However it is possible to develop approximation methods by assuming independence at the level of lower-order terms, such as individuals or pairs of nodes, and approximating the neighbourhood probabilities as a function of these.

For example, we can make a pair approximation by using conditional probabilities and assuming statistical independence at the level of pairs to rewrite the neighbourhood probability in terms of pair probabilities. From the definition of conditional probabilities, the right hand side of Equation (4.7) can be written as

$$\begin{aligned}\frac{dQ(A^t_{\{i\}})}{dt} =& \sum_{j=1}^{N} \sum_{X_{\mathcal{N}_j\setminus\{i\}}} G_{ij} Q(A^t_{\{j\}}) Q(B^t_{\{i\}} X^t_{\mathcal{N}_j\setminus\{i\}} | A^t_{\{j\}}) \bar{\chi}(\Omega^t_{j\to i} | B^t_{\{i\}} A^t_{\{j\}} X^t_{\mathcal{N}_j\setminus\{i\}}) \\ & - \sum_{j=1}^{N} \sum_{X_{\mathcal{N}_j\setminus\{i\}}} G_{ij} Q(B^t_{\{j\}}) Q(A^t_{\{i\}} X^t_{\mathcal{N}_j\setminus\{i\}} | B^t_{\{j\}}) \bar{\chi}(\Omega^t_{j\to i} | A^t_{\{i\}} B^t_{\{j\}} X^t_{\mathcal{N}_j\setminus\{i\}}). \end{aligned} \tag{4.8}$$

If we assume statistical independence of all nodes in the neighbourhood of j, given the state of j, we can rewrite the neighbourhood probability $Q(A^t_{\{j\}})Q(B^t_{\{i\}} X^t_{\mathcal{N}_j\setminus\{i\}} | A^t_{\{j\}})$ as

$$Q(A^t_{\{j\}})Q(B^t_{\{i\}} X^t_{\mathcal{N}_j\setminus\{i\}} | A^t_{\{j\}}) \approx Q(A^t_{\{j\}})Q(B^t_{\{i\}} | A^t_{\{j\}}) \prod_{l\in\mathcal{N}_j\setminus\{i\}} Q(X^t_{\{l\}} | A^t_{\{j\}}),$$

where $X^t_{\{l\}}$ is event where node l is in the same state as it is in the event $X^t_{\mathcal{N}_j\setminus\{i\}}$. Substi-

tuting this into Equation (4.8) gives

$$\begin{aligned}\frac{dQ(A^t_{\{i\}})}{dt} \approx &\sum_{j=1}^{N}\sum_{X_{\mathcal{N}_j\setminus\{i\}}} G_{ij}Q(A^t_{\{j\}})Q(B^t_{\{i\}}|A^t_{\{j\}})\prod_{l\in\mathcal{N}_j\setminus\{i\}} Q(X^t_{\{l\}}|A^t_{\{j\}})\bar{\chi}(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}}X^t_{\mathcal{N}_j\setminus\{i\}})\\ &-\sum_{j=1}^{N}\sum_{X_{\mathcal{N}_j\setminus\{i\}}} G_{ij}Q(B^t_{\{j\}})Q(A_{\{i\}}|B^t_{\{j\}})\prod_{l\in\mathcal{N}_j\setminus\{i\}} Q(X^t_{\{l\}}|B^t_{\{j\}})\bar{\chi}(\Omega^t_{j\to i}|A^t_{\{i\}}B^t_{\{j\}}X^t_{\mathcal{N}_j\setminus\{i\}}).\end{aligned}$$

Since $Q(B^t_{\{i\}}|A^t_{\{j\}}) = Q(B^t_{\{i\}}A^t_{\{j\}})/Q(A^t_{\{j\}})$, in order to evaluate these equations we require additional equations describing how pair probabilities change with time or some appropriate closure of pairs in terms of single node probabilities. From the master equation we can derive exact equations describing pairs. For the probability $P(B^t_{\{i\}}A^t_{\{j\}})$ we obtain

$$\begin{aligned}\frac{dP(B^t_{\{i\}}A^t_{\{j\}})}{dt} =&\sum_{k=1}^{N}\sum_{X_{V\setminus\{i,j,k\}}} G_{jk}P(B^t_{\{i\}}B^t_{\{j\}}A^t_{\{k\}}X^t_{V\setminus\{i,j,k\}})\chi(\Omega^t_{k\to j}|B^t_{\{i\}}B^t_{\{j\}}A^t_{\{k\}}X^t_{V\setminus\{i,j,k\}})\\ &-\sum_{k=1}^{N}\sum_{X_{V\setminus\{i,j,k\}}} G_{jk}P(B^t_{\{i\}}A^t_{\{j\}}B^t_{\{k\}}X^t_{V\setminus\{i,j,k\}})\chi(\Omega^t_{k\to j}|B^t_{\{i\}}A^t_{\{j\}}B^t_{\{k\}}X^t_{V\setminus\{i,j,k\}})\\ &+\sum_{k=1}^{N}\sum_{X_{V\setminus\{i,j,k\}}} G_{ik}P(B^t_{\{k\}}A^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j,k\}})\chi(\Omega^t_{k\to i}|B^t_{\{k\}}A^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j,k\}})\\ &-\sum_{k=1}^{N}\sum_{X_{V\setminus\{i,j,k\}}} G_{ik}P(A^t_{\{k\}}B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j,k\}})\chi(\Omega^t_{k\to i}|A^t_{\{k\}}B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j,k\}}).\end{aligned} \tag{4.9}$$

We can now apply the same assumption regarding total fitness that we used for the single node probabilities so that

$$\begin{aligned}\frac{dQ(B^t_{\{i\}}A^t_{\{j\}})}{dt} =&\sum_{k=1}^{N}\sum_{X_{\mathcal{N}_k\setminus\{i,j\}}} G_{jk}Q(B^t_{\{i\}}B^t_{\{j\}}A^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})\bar{\chi}(\Omega^t_{k\to j}|B^t_{\{i\}}B^t_{\{j\}}A^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})\\ &-\sum_{k=1}^{N}\sum_{X_{\mathcal{N}_k\setminus\{i,j\}}} G_{jk}Q(B^t_{\{i\}}A^t_{\{j\}}B^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})\bar{\chi}(\Omega^t_{k\to j}|B^t_{\{i\}}A^t_{\{j\}}B^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})\end{aligned}$$

$$+\sum_{k=1}^{N}\sum_{X_{\mathcal{N}_k\setminus\{i,j\}}} G_{ik}Q(B^t_{\{k\}}A^t_{\{i\}}A^t_{\{j\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})\bar{\chi}(\Omega^t_{k\to i}|B^t_{\{k\}}A^t_{\{i\}}A^t_{\{j\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})$$
$$-\sum_{k=1}^{N}\sum_{X_{\mathcal{N}_k\setminus\{i,j\}}} G_{ik}Q(A^t_{\{k\}}B^t_{\{i\}}A^t_{\{j\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})\bar{\chi}(\Omega^t_{k\to i}|A^t_{\{k\}}B^t_{\{i\}}A^t_{\{j\}}X^t_{\mathcal{N}_k\setminus\{i,j\}}). \tag{4.10}$$

Using conditional probabilities, the neighbourhood probability $Q(B^t_{\{i\}}B^t_{\{j\}}A^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})$ can be written as

$$Q(B^t_{\{i\}}B^t_{\{j\}}A^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}}) = Q(B^t_{\{j\}}A^t_{\{k\}})Q(B^t_{\{i\}}X^t_{\mathcal{N}_k\setminus\{i,j\}}|B^t_{\{j\}}A^t_{\{k\}})$$

We can now assume statistical independence of the remaining nodes given the state of j and k so that

$$Q(B^t_{\{i\}}B^t_{\{j\}}A^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}}) \approx Q(B^t_{\{j\}}A^t_{\{k\}})Q(B^t_{\{i\}}|B^t_{\{j\}}A^t_{\{k\}})\prod_{l\in\mathcal{N}_k\setminus\{i,j\}} Q(X^t_{\{l\}}|B^t_{\{j\}}A^t_{\{k\}}).$$

Since we know that node i is connected to node j we can assume that given the state of node j, the state of node i is independent of node k, and similarly the state of any node in the neighbourhood of k is independent of node j, which gives us

$$Q(B^t_{\{i\}}B^t_{\{j\}}A^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}}) \approx Q(B^t_{\{j\}}A^t_{\{k\}})Q(B^t_{\{i\}}|B^t_{\{j\}})\prod_{l\in\mathcal{N}_k\setminus\{i,j\}} Q(X^t_{\{l\}}|A^t_{\{k\}}).$$

Substituting this into Equation (4.10) gives

$$\frac{dQ(B^t_{\{i\}}A^t_{\{j\}})}{dt} \approx \tag{4.11}$$
$$\sum_{k=1}^{N}\sum_{X_{\mathcal{N}_k\setminus\{i,j\}}} G_{jk}Q(B^t_{\{j\}}A^t_{\{k\}})Q(B^t_{\{i\}}|B^t_{\{j\}})\prod_{l\in\mathcal{N}_k\setminus\{i,j\}} Q(X^t_{\{l\}}|A^t_{\{k\}})\bar{\chi}(\Omega^t_{k\to j}|B^t_{\{i\}}B^t_{\{j\}}A^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})$$
$$-\sum_{k=1}^{N}\sum_{X_{\mathcal{N}_k\setminus\{i,j\}}} G_{jk}Q(A^t_{\{j\}}B^t_{\{k\}})Q(B^t_{\{i\}}|A^t_{\{j\}})\prod_{l\in\mathcal{N}_k\setminus\{i,j\}} Q(X^t_{\{l\}}|B^t_{\{k\}})\bar{\chi}(\Omega^t_{k\to j}|B^t_{\{i\}}A^t_{\{j\}}B^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})$$

$$+\sum_{k=1}^{N}\sum_{X_{\mathcal{N}_k\setminus\{i,j\}}} G_{ik}Q(A^t_{\{i\}}B^t_{\{k\}})Q(A^t_{\{j\}}|A^t_{\{i\}})\prod_{l\in\mathcal{N}_k\setminus\{i,j\}} Q(X^t_{\{l\}}|B^t_{\{k\}})\bar{\chi}(\Omega^t_{k\to i}|A^t_{\{i\}}A^t_{\{j\}}B^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}})$$

$$-\sum_{k=1}^{N}\sum_{X_{\mathcal{N}_k\setminus\{i,j\}}} G_{ik}Q(B^t_{\{i\}}A^t_{\{k\}})Q(A^t_{\{j\}}|B^t_{\{i\}})\prod_{l\in\mathcal{N}_k\setminus\{i,j\}} Q(X^t_{\{l\}}|A^t_{\{k\}})\bar{\chi}(\Omega^t_{k\to i}|B^t_{\{i\}}A^t_{\{j\}}A^t_{\{k\}}X^t_{\mathcal{N}_k\setminus\{i,j\}}).$$

While this system is closed, its computational complexity increases exponentially with the maximum node degree of the graph, so it is not numerically feasible for graphs with highly connected nodes. While this could potentially be addressed by introducing approximations for nodes with high degree and this may lead to accurate models, here we continue towards a simplified model. To do this, we follow the same process as in epidemic models and make a homogeneity assumption by assuming that any pair is equally likely to be in any given state [91, 166]; i.e. $Q(X^t_{\{i\}}|Y^t_{\{j\}}) = Q(X^t|Y^t)$ for all pairs (i,j). This leads to

$$\begin{aligned}\frac{dQ(A^t_{\{i\}})}{dt} \approx &\sum_{j=1}^{N}\sum_{X_{\mathcal{N}_j\setminus\{i\}}} G_{ij}Q(A^t_{\{j\}})Q(B^t|A^t)^{k_j-n_X}Q(A^t|A^t)^{n_X}\bar{\chi}(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}}X^t_{\mathcal{N}_j\setminus\{i\}})\\ &-\sum_{j=1}^{N}\sum_{X_{\mathcal{N}_j\setminus\{i\}}} G_{ij}Q(B^t_{\{j\}})Q(A^t|B^t)^{n_X+1}Q(B^t|B^t)^{k_j-n_X-1}\bar{\chi}(\Omega^t_{j\to i}|A^t_{\{i\}}B^t_{\{j\}}X^t_{\mathcal{N}_j\setminus\{i\}}),\end{aligned}$$

where k_j is the degree of node j and n_X is the number of type A individuals in state $X_{\mathcal{N}_j\setminus\{i\}}$. Since the transition rate only depends on the number of type A and type B individuals in the neighbourhood of node j and not on their positions, the summand on the right hand side is equal for all states $X_{\mathcal{N}_j\setminus\{i\}}$ which have the same configuration of A and B individuals. The frequency of a certain neighbourhood state across all possible configurations is given by the binomial coefficient, so that

$$\begin{aligned}\frac{dQ(A^t_{\{i\}})}{dt} \approx &\sum_{j=1}^{N}\sum_{n=0}^{k_j-1} G_{ij}\binom{k_j-1}{n}Q(A^t_{\{j\}})Q(B^t|A^t)^{k_j-n}Q(A^t|A^t)^{n}\bar{\chi}(\Omega^t_{j\to i}|n)\\ &-\sum_{j=1}^{N}\sum_{n=0}^{k_j-1} G_{ij}\binom{k_j-1}{n}Q(B^t_{\{j\}})Q(A^t|B^t)^{n+1}Q(B^t|B^t)^{k_j-n-1}\bar{\chi}(\Omega^t_{j\to i}|n),\end{aligned}$$

where $\bar{\chi}(\Omega^t_{A\to B}|n)$ is the rate at which we select one of the type A individuals to reproduce

and replace a type B, given that there are n type A individuals and $k_j - n$ type B individuals in the neighbourhood of the selected node.

Since we have assumed that any pair is equally likely, this assumption only holds when every node in the graph forms k connections, which are chosen at random. Therefore we require that node i is equally likely to be connected to any other node and all nodes are topologically equivalent, so that the probability that a given node of type B is connected to x type A neighbours is given by a binomial distribution with $n = k$ and $p = Q(A^t|B^t)$. Therefore the probability of an individual being type A changes with rate

$$\begin{aligned}\frac{dQ(A^t)}{dt} \approx\ & kQ(A^t|B^t)Q(B^t)\sum_{n=0}^{k-1}\binom{k-1}{n}Q(B^t|A^t)^{k-n}Q(A^t|A^t)^n\bar{\chi}(\Omega^t_{A\to B}|n) \\ & - kQ(B_t|A_t)Q(A^t)\sum_{n=0}^{k-1}\binom{k-1}{n}Q(A^t|B^t)^{n+1}Q(B^t|B^t)^{k-n-1}\bar{\chi}(\Omega^t_{B\to A}|n+1).\end{aligned}$$

This assumption would best represent a network where all nodes have the same degree and edges are chosen at random at every time step. In practice, this assumption is reasonable for networks with little variation in their degree distribution and without significant structural rigidity. Therefore, whilst this assumption may accurately approximate a k-regular random graph it may not perform well on a square-lattice. This may also perform reasonably well on Erdős-Réyni random graphs, since these have reasonably low degree variation, but will not perform well on a scale-free random graph, which has very high degree variation.

We can also apply these assumptions to the pair-level equations to obtain a closed system of equations which are efficient to solve numerically. The resulting model is equivalent to the model in [117], which was justified by using the assumption that the population occupies a regular graph, such that all individuals have degree k, and that all nodes are topologically equivalent, such that every pair of individuals is equally likely to be connected. We have shown that by applying these assumptions to the exact node-level equations (Equation (4.4)) we can derive these models.

Similarly we can obtain a pair approximation model for the dynamics where we select on death by conditioning against the state of the neighbours of node i. Applying analogous assumptions to the previous example then leads to the model in [59]. These models have been shown to yield interesting qualitative results about the relative strengths of different strategies in evolutionary games on graphs. However, the homogeneity assumptions made

result in losing important aspects of the structure, such as how individual nodes in the system can behave differently. In the next sections we will attempt to develop approximation methods which can capture this node-specific information.

As we alluded to earlier, a natural method would be to use Equation (4.7) as a basis for this. However, difficulties in implementing this method on general graphs as well as the number of equations that result leads us to a different direction for the present work.

4.2.2 An unconditioned fitness approximation model

Here we develop a method which removes the need to include the probability of whole neighbourhoods by removing the conditioning in the replacement rate. This causes the replacement rate to only depend on the marginal probabilities of the state of each node rather than the full system state. This assumption also motivated a model in [175] in which the authors construct a population-level approximation describing how the expected number of individuals of each type change with time. Under this assumption, Equation (4.4) becomes

$$\begin{aligned}\frac{dP(A^t_{\{i\}})}{dt} \approx &\sum_{j=1}^{N}\sum_{X_{V\setminus\{i,j\}}} G_{ij}P(B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j\}})\chi(\Omega^t_{j\to i}) \\ &-\sum_{j=1}^{N}\sum_{X_{V\setminus\{i,j\}}} G_{ij}P(A^t_{\{i\}}B^t_{\{j\}}X^t_{V\setminus\{i,j\}})\chi(\Omega^t_{j\to i}).\end{aligned}$$

Since $\chi(\Omega^t_{j\to i})$ is now the same for all system states,

$$\frac{dP(A^t_{\{i\}})}{dt} \approx \sum_{j=1}^{N} G_{ij}P(B^t_{\{i\}}A^t_{\{j\}})\chi(\Omega^t_{j\to i}) - \sum_{j=1}^{N} G_{ij}P(A^t_{\{i\}}B^t_{\{j\}})\chi(\Omega^t_{j\to i}).$$

Adding and subtracting $\sum_{j=1}^{N} G_{ij}P(A^t_{\{i\}}A^t_{\{j\}})\chi(\Omega^t_{j\to i})$ we obtain

$$\frac{dP(A^t_{\{i\}})}{dt} \approx \sum_{j=1}^{N}\Big[G_{ij}\bar{P}(B^t_{\{i\}}A^t_{\{j\}})\chi(\Omega^t_{j\to i}) + G_{ij}P(A^t_{\{i\}}A^t_{\{j\}})\chi(\Omega^t_{j\to i})\Big]$$

$$
\begin{aligned}
&- \sum_{j=1}^{N} \left[G_{ij} P(A^t_{\{i\}} B^t_{\{j\}}) \chi(\Omega^t_{j\to i}) + G_{ij} \bar{P}(A^t_{\{i\}} A^t_{\{j\}}) \chi(\Omega^t_{j\to i}) \right] \\
&\approx \sum_{j=1}^{N} G_{ij} P(A^t_{\{j\}}) \chi(\Omega^t_{j\to i}) - \sum_{j=1}^{N} G_{ij} P(A^t_{\{i\}}) \chi(\Omega^t_{j\to i}),
\end{aligned}
$$

which is a closed set of N equations with at most N summands on the right hand side. Therefore by defining $\bar{P}$ as an approximation to the probability distribution P we obtain the closed system

$$
\frac{d\bar{P}(A^t_{\{i\}})}{dt} = \sum_{j=1}^{N} G_{ij} \bar{P}(A^t_{\{j\}}) \chi(\Omega^t_{j\to i}) - \sum_{j=1}^{N} G_{ij} \bar{P}(A^t_{\{i\}}) \chi(\Omega^t_{j\to i}), \tag{4.12}
$$

which is easy to solve numerically for an arbitrary graph.

Example 4.2.1 (Neutral drift)**.** *In the special case of neutral drift, i.e. when all individuals have identical fitness, the unconditioned fitness model gives the exact fixation probability. With the dynamics of the invasion process under neutral drift we obtain* $\chi(\Omega^t_{j\to i}) = \frac{1}{Nk_j}$*, and therefore Equation (4.12) can be written as*

$$
\frac{d\bar{P}(A^t_{\{i\}})}{dt} = \sum_{j=1}^{N} G_{ij} \bar{P}(A^t_{\{j\}}) \frac{1}{Nk_j} - \sum_{j=1}^{N} G_{ij} \bar{P}(A^t_{\{i\}}) \frac{1}{Nk_j},
$$

which is equivalent to the exact node equation (4.4) for the invasion process under neutral drift [164]. The unconditioned fitness model is also exact for all update mechanisms under neutral drift, but we do not write the equations explicitly here.

As the population size N increases, the solution to Equation (4.12) moves further away from the exact fixation probability obtained either by solving the master equation (4.3) or from the output of stochastic simulations. To obtain a reasonable approximation to the fixation probability from a given initial condition we construct a scaling factor for the constant fitness case by comparing the ratio between the solution of Equation (4.12) on a complete graph to the exact fixation probability on a complete graph. We choose the complete graph because the exact fixation probability can be calculated analytically in this case. Whilst we consider the constant fitness case, it may also be possible to find a suitable scaling factor in the frequency dependent fitness case, however using a complete graph may no longer be appropriate because the relative strength of different strategies in

some games is strongly affected by the average degree of the graph [129].

Example 4.2.2 (Invasion process)**.** *For constant fitness under the dynamics of the invasion process, the exact fixation probability for m initial mutant A individuals on a complete graph is equivalent to the Moran probability [102]:*

$$\rho^A(m) = \frac{1 - \frac{1}{r^m}}{1 - \frac{1}{r^N}}.$$

Since the fixation probability is known, we now need to solve Equation (4.12) on the complete graph to derive the ratio between the two. In the constant fitness case this can be done analytically, with the scaling factor for m initial mutants given by

$$\frac{\rho^A(m)}{\lim\limits_{t\to\infty} A_c(t)} = \frac{\frac{1-\frac{1}{r^m}}{1-\frac{1}{r^N}}}{\frac{1}{r-1}\left(-1 + \sqrt{1 + \frac{m(r^2-1)}{N}}\right)}, \tag{4.13}$$

where $A_c(t) = \frac{1}{N}\sum\limits_{j=1}^{N} \bar{P}(A^t_{\{j\}})$*. The derivation of this can be found in Appendix 4.B.*

We can now define two methods for predicting the fixation probability under any Markovian update mechanism.

- **Method 1** (Unconditioned fitness model) Solve Equation (4.12) to provide an approximation to the dynamics of the evolutionary process.
- **Method 2** (Scaled unconditioned fitness model) Solve Equation (4.12) and then use a scaling factor, the ratio of the exact fixation probability and the solution to Equation (4.12) for the complete graph, to provide an approximation to the fixation probability from a given initial condition.

In Section 4.3 we investigate the numerical performance of these two methods. Note that for the purpose of this work we have found the scaling factor for Method 2 under the invasion process (Equation (4.13)). However, the method can be applied to other update mechanisms, such as death-birth with selection on birth, by finding an appropriate scaling factor, which can be done by solving Equation (4.12) (either analytically or numerically) and comparing to the exact fixation probability on the complete graph. For example, see [70] for the exact fixation probability on a complete graph under the DB-B dynamics.

4.2.3 A contact conditioning approximation model

In Section 4.2.2 we restricted the conditioning so that we only require the marginal probabilities of the individual nodes. However, this removes a significant amount of information from the dynamics. In the evolutionary process, when considering a replacement event the two nodes of most interest are the node selected for birth and the node selected for death. Therefore, here we follow a similar method but retain conditioning on the states of these two key nodes. Since we restrict the conditioning to only the states of the relevant contact, when looking at the term $\chi(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j\}})$ in Equation (4.4) we condition only on the states of the nodes i and j and obtain

$$\chi(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j\}}) \approx \chi(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}}).$$

Under the above condition, Equation (4.4) becomes

$$\begin{aligned}\frac{dP(A^t_{\{i\}})}{dt} \approx &\sum_{j=1}^{N}\sum_{X_{V\setminus\{i,j\}}} G_{ij}P(B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j\}})\chi(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}}) \\ &-\sum_{j=1}^{N}\sum_{X_{V\setminus\{i,j\}}} G_{ij}P(A^t_{\{i\}}B^t_{\{j\}}X^t_{V\setminus\{i,j\}})\chi(\Omega^t_{j\to i}|A^t_{\{i\}}B^t_{\{j\}}). \qquad (4.14)\end{aligned}$$

To see the effect of this assumption on the rates, consider $\chi(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}})$. Here we condition only against node i being in state B and node j being in state A rather than against the entire system state. Consequently in the fitness equation (4.6) we have $P(B^t_{\{i\}}) = 1$ and $P(A^t_{\{j\}}) = 1$ giving

$$f^t_j|B^t_{\{i\}}A^t_{\{j\}} = f_{back} + w\frac{bT_{ij} + a\sum_{l\neq i}G_{jl}P(A^t_{\{l\}}) + b\sum_{l\neq i}G_{jl}P(B^t_{\{l\}})}{\sum_{l=1}^{N}G_{jl}}.$$

In Equation (4.14), the chance of selecting node j is now independent of the state $X^t_{V\setminus\{i,j\}}$ of the remaining nodes which enables the equation to be reduced to

$$\frac{dP(A^t_{\{i\}})}{dt} \approx \sum_{j=1}^{N} G_{ij} P(B^t_{\{i\}} A^t_{\{j\}}) \chi(\Omega^t_{j\to i} | B^t_{\{i\}} A^t_{\{j\}}) - \sum_{j=1}^{N} G_{ij} P(A^t_{\{i\}} B^t_{\{j\}}) \chi(\Omega^t_{j\to i} | A^t_{\{i\}} B^t_{\{j\}}). \tag{4.15}$$

This gives an approximate equation for individuals in terms of pairs. We then need to build equations to describe pair-level probabilities. Similar methodologies have been followed to describe epidemics propagated on networks [166, 168].

Applying the same conditioning to the exact pair-level equation (4.9) we obtain

$$\frac{dP(B^t_{\{i\}} A^t_{\{j\}})}{dt} \approx \tag{4.16}$$

$$\sum_{k=1}^{N} G_{jk} P(B^t_{\{i\}} B^t_{\{j\}} A^t_{\{k\}}) \chi(\Omega^t_{k\to j} | B^t_{\{j\}} A^t_{\{k\}}) - \sum_{k=1}^{N} G_{jk} P(B^t_{\{i\}} A^t_{\{j\}} B^t_{\{k\}}) \chi(\Omega^t_{k\to j} | A^t_{\{j\}} B^t_{\{k\}})$$

$$+ \sum_{k=1}^{N} G_{ik} P(B^t_{\{k\}} A^t_{\{i\}} A^t_{\{j\}}) \chi(\Omega^t_{k\to i} | B^t_{\{k\}} A^t_{\{i\}}) - \sum_{k=1}^{N} G_{ik} P(A^t_{\{k\}} B^t_{\{i\}} A^t_{\{j\}}) \chi(\Omega^t_{k\to i} | A^t_{\{k\}} B^t_{\{i\}}). \tag{4.17}$$

Similar formulae can be constructed for all possible pairs, writing pairs in terms of triples. In a similar way, triples can be written in terms of quads and so on, up to the full system size N which is then closed. Therefore, when using this method we obtain a hierarchy similar to the BBGKY (Bogoliubov–Born–Green–Kirkwood–Yvon) hierarchy [18, 88] in statistical physics. However, here the hierarchy only represents an approximation to the original dynamics. Solving this system exactly is no simpler than evaluating Equation (4.3) since evaluating the hierarchy in full is comparable in numerical complexity, so we wish to find approximation methods to reduce this.

With this hierarchy, we can apply techniques developed in statistical physics to approximate higher-order terms as functions of lower-order terms (as described in Section 1.4.2 for SIS epidemic models). In particular we can close the system of equations (4.15) and (4.17) at the level of pairs by approximating all triples in Equation (4.17) in terms of pair-level and individual-level probabilities. Similar techniques have been applied for many stochastic processes including in epidemiology [83, 91, 166, 168] and evolutionary dynamics [67, 129, 175] leading to models which can be numerically evaluated.

To close the system, we require a functional form that can approximate triple probabilities in terms of individual and pair probabilities. One method is to approximate a triple $P(A^t_{\{i\}}B^t_{\{j\}}C^t_{\{k\}})$ as the product of all possible pairs among these nodes divided by the product of all individuals, i.e.

$$P(A^t_{\{i\}}B^t_{\{j\}}C^t_{\{k\}}) \approx \frac{P(A^t_{\{i\}}B^t_{\{j\}})P(B^t_{\{j\}}C^t_{\{k\}})P(A^t_{\{i\}}C^t_{\{k\}})}{P(A^t_{\{i\}})P(B^t_{\{j\}})P(C^t_{\{k\}})}. \tag{4.18}$$

This closure is commonly attributed to Kirkwood [172] because it is derived from the Kirkwood superposition which approximates triples in terms of pairs in thermodynamics [87, 89]. This is often applied to nodes i, j, k that form a 3-cycle in the graph, which we call a 'closed triple', although it can be applied to any triple of nodes. It has been shown that this closure maximises the entropy of these thermodynamic systems [172], and it also ensures that symmetry is preserved across the triple. This closure has commonly been adapted to probabilistic systems, such as the BBGKY hierarchy [18, 88] and epidemic modelling [82, 166, 169]. However, the Kirkwood closure for probabilities does not define a probability distribution since we can obtain $P(B^t_{\{i\}}A^t_{\{j\}}) + P(B^t_{\{i\}}B^t_{\{j\}}) \neq P(B^t_{\{i\}})$, which has been observed numerically [157]. In spite of this it has been shown to yield accurate approximations in these probabilistic systems [157, 166, 172].

Another closure can be obtained by using conditional probabilities and assuming statistical independence, as described in Section 1.4.2. We have

$$P(A^t_{\{i\}}B^t_{\{j\}}C^t_{\{k\}}) = P(A^t_{\{i\}}|B^t_{\{j\}}C^t_{\{k\}})P(B^t_{\{j\}}C^t_{\{k\}}),$$

which, when we assume statistical independence of nodes i and k given j, simplifies to

$$P(A^t_{\{i\}}B^t_{\{j\}}C^t_{\{k\}}) \approx P(A^t_{\{i\}}|B^t_{\{j\}})P(B^t_{\{j\}}C^t_{\{k\}}) = \frac{P(A^t_{\{i\}}B^t_{\{j\}})P(B^t_{\{j\}}C^t_{\{k\}})}{P(B^t_{\{j\}})}. \tag{4.19}$$

Typically this closure is applied to nodes on a graph where nodes i and j are connected and nodes j and k are connected but where there is no connection between nodes i and k, which we call an 'open triple'. However, it could be applied to any triplet of nodes. This closure method is thought to be most accurate on trees [91, 157, 168], and has been shown to be exact for such graphs under the SIR epidemic model [92, 168, 169].

We can adopt either closure to remove triples and close the system. For example, if we

are using the Kirkwood closure to approximate all triples in Equation (4.17) we obtain the system of equations

$$\frac{d\bar{P}(A^t_{\{i\}})}{dt} = \sum_{j=1}^{N} G_{ij}\bar{P}(B^t_{\{i\}}A^t_{\{j\}})\chi(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}}) - \sum_{j=1}^{N} G_{ij}\bar{P}(A^t_{\{i\}}B^t_{\{j\}})\chi(\Omega^t_{j\to i}|A^t_{\{i\}}B^t_{\{j\}}).$$

$$\begin{aligned}\frac{d\bar{P}(B^t_{\{i\}}A^t_{\{j\}})}{dt} = &\sum_{k=1}^{N} G_{jk}\frac{\bar{P}(B^t_{\{i\}}B^t_{\{j\}})\bar{P}(B^t_{\{j\}}A^t_{\{k\}})\bar{P}(B^t_{\{i\}}A^t_{\{k\}})}{\bar{P}(B^t_{\{i\}})\bar{P}(B^t_{\{j\}})\bar{P}(A^t_{\{k\}})}\chi(\Omega^t_{k\to j}|B^t_{\{j\}}A^t_{\{k\}})\\
&- \sum_{k=1}^{N} G_{jk}\frac{\bar{P}(B^t_{\{i\}}A^t_{\{j\}})\bar{P}(A^t_{\{j\}}B^t_{\{k\}})\bar{P}(B^t_{\{i\}}B^t_{\{k\}})}{\bar{P}(B^t_{\{i\}})\bar{P}(A^t_{\{j\}})\bar{P}(B^t_{\{k\}})}\chi(\Omega^t_{k\to j}|A^t_{\{j\}}B^t_{\{k\}})\\
&+ \sum_{k=1}^{N} G_{ik}\frac{\bar{P}(B^t_{\{k\}}A^t_{\{i\}})\bar{P}(A^t_{\{i\}}A^t_{\{j\}})\bar{P}(B^t_{\{k\}}A^t_{\{j\}})}{\bar{P}(B^t_{\{k\}})\bar{P}(A^t_{\{i\}})\bar{P}(A^t_{\{j\}})}\chi(\Omega^t_{k\to i}|B^t_{\{k\}}A^t_{\{i\}})\\
&- \sum_{k=1}^{N} G_{ik}\frac{\bar{P}(A^t_{\{k\}}B^t_{\{i\}})\bar{P}(B^t_{\{i\}}A^t_{\{j\}})\bar{P}(A^t_{\{k\}}A^t_{\{j\}})}{\bar{P}(A^t_{\{k\}})\bar{P}(B^t_{\{i\}})\bar{P}(A^t_{\{j\}})}\chi(\Omega^t_{k\to i}|A^t_{\{k\}}B^t_{\{i\}}),\end{aligned}$$

where $\bar{P}$ represents the approximation to the probability distribution P. However, note that using this closure for all triples will eventually require equations for every pair of nodes in the system, whether they are connected or not.

It is also useful to use a combination of the two methods whereby the Kirkwood closure (4.18) is used for closed triples, and (4.19) is used for open triples [82, 166]. In this work we shall use this combined approach to obtain a closed system. However, we find that unlike in epidemiology, this standard approach does not produce good results. We therefore also try using just the Kirkwood closure because this permits explicit correlations between nodes which are not linked, although as indicated above, this substantially increases computational complexity because the system of equations will scale with N^2 rather than the number of connected individuals in the graph.

With the contact conditioning model we define two different methods to approximate the evolutionary dynamics.

- **Method 3** (Open and closed triples) Solve Equation (4.15) together with equations for pairs by using two different closures for different types of triples. First consider a triple $P(A^t_{\{i\}}B^t_{\{j\}}Z^t_{\{k\}})$, $Z \in \{A, B\}$, where there is no link between nodes i and k.

We call this an open triple, and can approximate it as

$$P(A^t_{\{i\}}B^t_{\{j\}}Z^t_{\{k\}}) \approx \frac{P(A^t_{\{i\}}B^t_{\{j\}})P(B^t_{\{j\}}Z^t_{\{k\}})}{P(B^t_{\{j\}})}.$$

If there exists a link between nodes i and k we call this a closed triple, and approximate this using the Kirkwood closure,

$$P(A^t_{\{i\}}B^t_{\{j\}}Z^t_{\{k\}}) \approx \frac{P(A^t_{\{i\}}B^t_{\{j\}})P(B^t_{\{j\}}Z^t_{\{k\}})P(A^t_{\{i\}}Z^t_{\{k\}})}{P(A^t_{\{i\}})P(B^t_{\{j\}})P(Z^t_{\{k\}})}.$$

Using this method it is only necessary to use pairs which have a link between them in the graph, and so it scales with Nd, where d is the average degree of the graph.

- **Method 4** (Kirkwood closure only) Solve Equation (4.15) together with equations for pairs by using the Kirkwood closure for all triples. That is, we approximate any triple $P(A^t_{\{i\}}B^t_{\{j\}}Z^t_{\{k\}})$, $Z \in \{A, B\}$ as

$$P(A^t_{\{i\}}B^t_{\{j\}}Z^t_{\{k\}}) \approx \frac{P(A^t_{\{i\}}B^t_{\{j\}})P(B^t_{\{j\}}Z^t_{\{k\}})P(A^t_{\{i\}}Z^t_{\{k\}})}{P(A^t_{\{i\}})P(B^t_{\{j\}})P(Z^t_{\{k\}})}.$$

This method requires the use of every pair of nodes in the system, not just those which are directly connected, and so scales with N^2.

4.3 Results

4.3.1 A comparison of the different methods: fixation probabilities for constant fitness

Here we investigate the fixation probability of a single initial A individual placed in a given node on the graph under the dynamics of the invasion process. Figure 4.1 compares Method 1 (unconditioned fitness model) under the invasion process against stochastic simulation on a four-node star graph. On such small graphs, Method 1 appears to provide a reasonable approximation to the expected dynamics and to the fixation probability. However, for such small populations exact solutions are easy to obtain, and hence we want to test larger population sizes. When the population size is increased, this method fails to accurately predict the fixation probability, appearing to tend towards zero with increasing

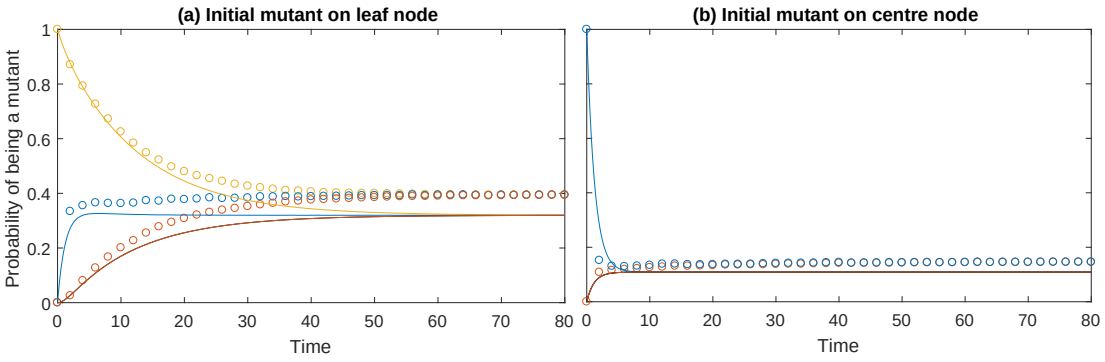

Figure 4.1: Comparison of the marginal probabilities for each node on the graph being a mutant A plotted against time as given by Method 1 (solid lines) versus stochastic simulation of the discrete-time system (circles), when applied to the invasion process on a 4-node star graph. We consider (a) dynamics initiated with a single A individual on a leaf node and (b) dynamics initiated with a single A individual on the central node. Each line represents the marginal probability of a certain node in the graph being occupied by an A individual, the corresponding colours between solid lines and circles represent the same node on the graph. The stochastic process is simulated 10,000 times from the same initial condition until fixation of either the mutant or resident strategy. The probabilities represent, for a given node at a given time, the proportion of simulations for which that node is a mutant. Method 1 is numerically integrated to approximate the probability of each node being a mutant at a given time. This is the constant fitness case where A individuals have fitness 1.2 and B individuals have fitness 1.

population size (for example, see Table 4.1, where it can be seen that increasing the size from 20 to 35 to 50 moves the solution closer to zero on random graphs). To account for this, we use Method 2 (scaled unconditioned fitness model).

Method 2 represents a scaling of the approximation from Method 1 where the scaling is derived analytically from the fixation probability for a complete graph. Consequently, it makes sense to only consider the approximation of the fixation probability rather than the whole time series. Predictions of the fixation probability of a single A individual when placed on various graphs using the different approximation methods are shown in Tables 4.1 and 4.2. We first observe that the accuracy of the method does not significantly differ for different population sizes, so this overcomes the issue with Method 1. For both the Erdős-Réyni [43] and scale-free random graphs, we start the process in three different initial conditions; a high-degree initial node, a low-degree initial node and an average degree initial node. This is because under the dynamics of the invasion process, a low degree node is known to act as an amplifier of selection and a high degree node is known to act as a suppressor [6, 164], and so we potentially expect different performance of the methods when initiated from nodes of different degree. In the random k-regular random

graph, since all nodes have equal degree, we only consider results for one initial node. In addition to the random graphs (Table 4.2), we also investigate a star graph, a square lattice and Zachary's karate club [201]. On these graphs we initiate the dynamics from a high degree and low degree node. We observe that Method 2 performs best on the k-regular random graph and that generally it performs very well on any graph that does not strongly amplify or suppress the average fixation probability compared to the Moran probability, such as the Erdős-Réyni random graph and the square lattice. However on graphs which amplify (or suppress) average fixation probability, such as the scale-free random graph, the approximation becomes less accurate. On the star graph, which significantly amplifies the fixation probability, the approximation is very far from the true value. This is unsurprising because Method 2 is constructed to give the exact fixation probability on complete graphs. For Zachary's karate club, Method 2 provides a reasonable approximation, but does not capture the strong amplifying effect of the low degree node.

In order to improve upon the accuracy of Method 2 we developed the contact conditioning model to retain more information from the system. The contact conditioning model yields a hierarchy which offers no useful reduction in computational complexity, compared to the master equation (4.4). Therefore we developed Method 3 (open and closed triples approximation), analogous to closures used in epidemiology. However, through numerical evaluation we found that this only yields good approximations for simple graphs, such as line graphs and complete graphs for which we have exact analytic results in any case. On other graphs, the fixation probability approximation is equal to 1 (Tables 4.1 and 4.2) for an advantageous mutant of type A, and so this method is not particularly informative.

While the specific reason for this convergence to 1 (or 0 if the mutant is disadvantageous) is unclear, it seems likely that it is associated with graph-wide correlations caused by having two absorbing states. To address this we developed Method 4 (Kirkwood closure only). Through testing multiple graphs we observe (Tables 4.1 and 4.2) that the best results are obtained on Erdős-Réyni and random k-regular graphs, with some accuracy lost on scale-free random graphs. We observe that on the 20 node star graph, inaccuracies result in a significantly amplified approximation when initiated on the low degree leaf nodes, and for the 35 and 50 node star graphs the approximations initiated on the leaf node are close to 1. This is potentially due to the time to convergence on large stars being very long, which allows these inaccuracies to compound so that the system converges to this uninformative solution. This failure does not occur on these stars if we reduce the fitness advantage, suggesting that as the size of the star becomes very large the method

Table 4.1: The fixation probability starting from a single mutant A individual placed on a specific node on single realisations of random graphs. To evaluate the fixation probability using the approximate methods, we solved them until a steady-state was reached and calculated the average probability of a node being a mutant (the methods do not always give exactly the same value for each node). We compare this to the fixation probability as calculated by the proportion of 10,000 stochastic simulations in which the type A individuals fixated. Constant fitness is assumed, where A individuals have fitness 1.2 and B individuals have fitness 1. All graphs were generated to have an average degree of 5.

Graph	Fixation probability				
	Method 1	Method 2	Method 3	Method 4	Simulation
20 node Erdős-Réyni - initial degree 10	0.0193	0.0604	1.0000	0.0654	0.0784
20 node Erdős-Réyni - initial degree 2	0.1055	0.3301	1.0000	0.2874	0.3098
20 node Erdős-Réyni - initial degree 5	0.0424	0.1326	1.0000	0.1343	0.1575
20 node scale-free - initial degree 10	0.0190	0.0594	1.0000	0.0681	0.0783
20 node scale-free - initial degree 2	0.0945	0.2956	1.0000	0.3004	0.3153
20 node scale-free - initial degree 5	0.0475	0.1486	1.0000	0.1490	0.1606
20 node random k-regular	0.0547	0.1711	1.0000	0.1516	0.1722
35 node Erdős-Réyni - initial degree 10	0.0126	0.0671	1.0000	0.0782	0.0940
35 node Erdős-Réyni - initial degree 2	0.0628	0.3346	1.0000	0.3255	0.3191
35 node Erdős-Réyni - initial degree 5	0.0315	0.1679	1.0000	0.1572	0.1730
35 node scale-free - initial degree 10	0.0089	0.0474	1.0000	0.0844	0.0724
35 node scale-free - initial degree 2	0.0444	0.2366	1.0000	0.4743	0.2929
35 node scale-free - initial degree 5	0.0223	0.1188	1.0000	0.1950	0.1546
35 node random k-regular	0.0313	0.1668	1.0000	0.1631	0.1750
50 node Erdős-Réyni - initial degree 10	0.0083	0.0630	1.0000	0.0787	0.0820
50 node Erdős-Réyni - initial degree 2	0.0332	0.2521	1.0000	0.4175	0.3060
50 node Erdős-Réyni - initial degree 5	0.0272	0.2065	1.0000	0.2275	0.2120
50 node scale-free - initial degree 10	0.0056	0.0425	1.0000	0.0872	0.0660
50 node scale-free - initial degree 2	0.0307	0.2331	1.0000	0.3912	0.2840
50 node scale-free - initial degree 5	0.0154	0.1169	1.0000	0.1868	0.1530
50 node random k-regular	0.0219	0.1667	1.0000	0.1533	0.1640

will only work under weak selection. On random graphs, which do not significantly amplify fixation, this issue is also observed, but only when the fitness advantage of one type is sufficiently high. This issue starts when the fitness advantage is at about 50%, below which the solution converges to intermediate values on all random graphs tested. In addition to testing the star graph as an example of an extreme structure, we also tested a square lattice of various sizes, on which we find that Method 4 significantly underestimates the fixation probability. The square lattice is considered as an extreme scenario for this method because it contains many short cycles of order four, for which the correlations are not explicitly captured by the Kirkwood closure, which describes triples. Presenting the star graph and square lattice therefore illustrate the cases where this method is expected to perform least well. Testing Zachary's karate club [201] illustrates how this method might

Table 4.2: The fixation probability starting from a single mutant A individual placed on a specific node on the example graphs. To evaluate the fixation probability using the approximate methods, we solved them until a steady-state was reached and calculated the average probability of a node being a mutant (the methods do not always give exactly the same value for each node). We compare this to the fixation probability as calculated by the proportion of 10,000 stochastic simulations in which the type A individuals fixated. Constant fitness is assumed, where A individuals have fitness 1.2 and B individuals have fitness 1.

Graph	Fixation probability				
	Method 1	Method 2	Method 3	Method 4	Simulation
20 node star - initial degree 1	0.0574	0.1796	1.0000	0.3801	0.2895
20 node star - initial degree 19	0.0030	0.0094	1.0000	0.0217	0.0184
25 node square lattice - initial degree 2	0.0662	0.2546	1.0000	0.1532	0.2388
25 node square lattice - initial degree 4	0.0332	0.1277	1.0000	0.0780	0.1444
34 node Zachary's karate club - initial degree 2	0.0482	0.2498	1.0000	0.4285	0.3160
34 node Zachary's karate club - initial degree 16	0.0061	0.0314	1.0000	0.0461	0.0450
36 node star - initial degree 1	0.0322	0.1717	1.0000	1.0000	0.2971
36 node star - initial degree 35	0.0009	0.0051	1.0000	0.0209	0.0090
36 node square lattice - initial degree 2	0.0483	0.2646	1.0000	0.1363	0.2462
36 node square lattice - initial degree 4	0.0242	0.1326	1.0000	0.0689	0.1385
49 node star - initial degree 1	0.0224	0.1697	1.0000	1.0000	0.3070
49 node star - initial degree 48	0.0005	0.0035	1.0000	0.0260	0.0059
49 node square lattice - initial degree 2	0.0367	0.2734	1.0000	0.1241	0.2494
49 node square lattice - initial degree 4	0.0184	0.1369	1.0000	0.0609	0.1477

perform on a real world graph. On this graph we find that Method 4 provides a reasonable approximation to the fixation probabilities (Table 4.2).

We also observed, as shown in Tables 4.1 and 4.2, that Method 4 performs most accurately when initiated on a node with average to high degree. In addition to approximating the fixation probability, Method 4 can be used to approximate the dynamics across the whole time series, and in particular provides a very accurate approximation to the initial dynamics for all graphs tested (see Figure 4.2 for results on two 20 node graphs as an illustration). This accuracy holds even for the large star graphs when initiated on the leaf node, for which the final approximation was close to 1.

4.3.2 The Hawk-Dove game with the contact conditioning model

So far, we have considered the constant fitness case. Here we briefly consider the effectiveness of Method 4 when applied to the Hawk-Dove game under the dynamics of the invasion process. Method 2 relies on finding a suitable scaling factor, whilst Methods 1 and 3 were both observed in Section 4.3.1 to yield non-informative results on the type of graphs we test here and so we do not investigate these methods in this context.

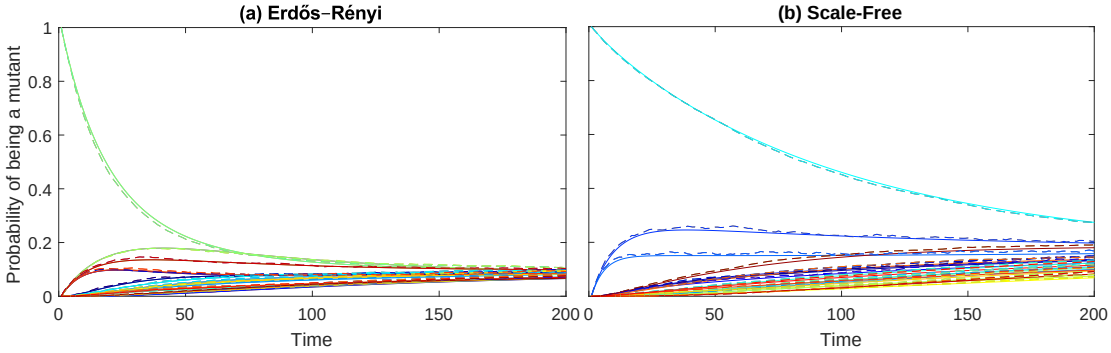

Figure 4.2: Comparison of the early dynamics of the marginal probabilities for each node on the graph being a mutant A plotted against time as given by Method 4 (solid lines) versus stochastic simulation (dashed lines), when applied to the invasion process on (a) an Erdős-Réyni random graph with 20 nodes and average degree of 4 and (b) a scale-free graph with 20 nodes and average degree 4, both initiated with a single A individual in a chosen node. Each line represents the marginal probability of a certain node in the graph being occupied by an A individual, the corresponding colours between the solid lines and dashed lines representing the same node on the graphs. The discrete-time stochastic process was simulated 10,000 times from the same initial condition, from which we obtained the probability for each node being a mutant at a given time as the proportion of simulations for which that node is a mutant. Method 4 was numerically integrated to approximate the probability of each node being a mutant at a given time. We use a dashed line with interpolation between integer time points for the discrete-time system to enable easier comparison of the dynamics. The game considered is the constant fitness case where the A individuals have fitness 1.2 and the B individuals have fitness 1.

The Hawk-Dove game [109, 110] represents a simple model of how animals compete over food, territory and other resources. Animals interact over a resource with either an aggressive or non-aggressive strategy, which we call the Hawk and Dove strategies, respectively. We let the resource yield a payoff V which both players try to obtain. When two Hawks interact, they fight over the resource with one taking the payoff V, and the other accruing a cost C from the fight, and therefore the average payoff received by a Hawk interacting with a Hawk is $(V - C)/2$. When a Hawk meets a Dove, the Dove retreats without a fight receiving a payoff 0, allowing the Hawk to take the whole resource, receiving payoff V. If two Doves meet, they either share the resource, or each takes the whole reward without a fight with probability $1/2$, so that the average payoff received by a Dove from

this interaction is $V/2$. Therefore, in this game the payoff matrix is given by

$$\begin{array}{c} \\ H \\ D \end{array} \begin{array}{c} \begin{array}{cc} H & D \end{array} \\ \begin{pmatrix} (V-C)/2 & V \\ 0 & V/2 \end{pmatrix} \end{array}.$$

Figure 4.3 illustrates results from this game on a scale-free graph, an Erdős-Réyni random graph, a random k-regular graph and a square lattice. Here we add a background fitness to the individuals, so that the fitness of a type X individual is given by $f_{\text{back}} + f_X$. This is to ensure that fitness is always positive. We consider two cases; firstly where the fight cost is low using parameters $f_{\text{back}} = 2$, $w = 1$, $V = 1$ and $C = 1.5$, and secondly where the fight cost is high using parameters $f_{\text{back}} = 2$, $w = 1$, $V = 1$ and $C = 4$. In each case we compare the results of Method 4 to stochastic simulation, initiated with a population consisting of half Hawks and half Doves to minimise the chance of early extinction events. We observe that when the cost is low the approximation is reasonable, with all 3 random graphs providing a good approximation, and some accuracy lost on the square lattice. However, as we increase the cost, C, we observe that the approximation does not perform well. This is because the contact conditioning assumption seems to amplify the strength of the Hawk strategy, with the rate at which an individual becomes a Hawk under this assumption being greater than it will be in the exact case.

4.4 Discussion

Evolutionary graph theory [102] was introduced as a way of adding spatial structure to the stochastic evolutionary dynamics considered by Moran [116]. Analytic results on these stochastic dynamics focused on idealised cases of simple graphs [6, 19]. In order to study arbitrary graphs, methods usually follow certain restrictions, such as focusing on the evolutionary process under weak selection or infinitely large populations [2, 129, 203]. Alternatively, individual-based stochastic simulations give very accurate results but are limited by computational time [12, 105].

The focus of this chapter has been the attempt to develop a general method that can approximate the stochastic dynamics on a wide range of graphs by adapting methods from statistical physics and epidemiology. In doing this, we have provided a derivation of existing (homogenised) pair-approximation models from the master equation [59, 67, 117, 144, 175]

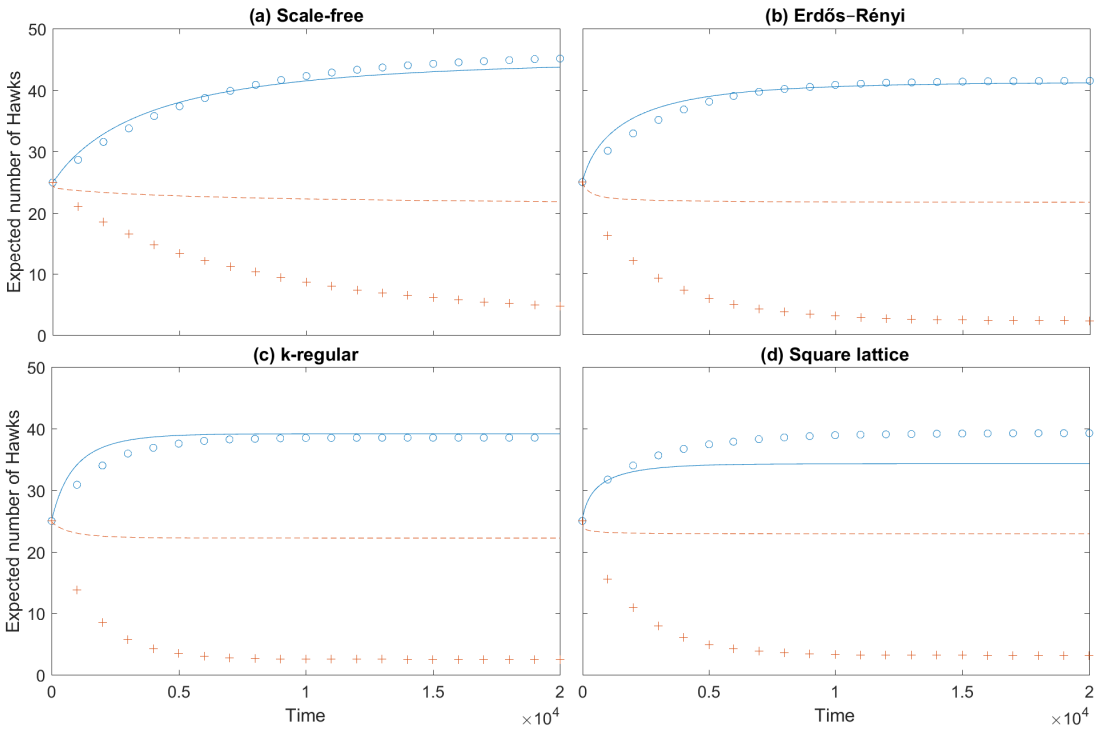

Figure 4.3: Comparison of the expected number of individuals playing the Hawk strategy in a Hawk-Dove game plotted against time as given by Method 4 versus stochastic simulation, when played on (a) a scale-free graph (b) an Erdős-Réyni graph (c) a random k-regular graph and (d) a 7 by 7 square lattice. Except for the square lattice, each graph has 50 nodes and an average degree of approximately 4. The solid lines represent the solution of Method 4 and the circles represent stochastic simulations of the discrete-time system, evaluated every 1000 time steps, in the case where $C = 1.5$. The dashed lines represent the solution of Method 4 and the crosses represent stochastic simulations of the discrete-time system, evaluated every 1000 time steps, in the case where $C = 4$. To generate the stochastic simulation results the discrete-time stochastic process was simulated 10,000 times from the same well mixed initial condition until fixation was reached. By taking the average number of Hawks at each time step we determined the expected number of Hawks at a given time. Method 4 is numerically integrated to give the probability of each node being a Hawk at a given time, from which we obtained the expected number of Hawks by summing over all nodes.

(Section 4.2.1). Additionally, we also derived an individual-level model which has the neutral drift model [164] as a special case (Section 4.2.2).

We start with a representation of the stochastic evolutionary process using a master equation [68], from which we develop exact equations describing individual node probabilities. We then apply ideas for approximating the master equation based around developing

hierarchies of moment equations. Such methods were originally developed in physics [18, 88] and later used in epidemiology and ecology [67, 82, 142, 168, 169]. The key idea behind these techniques is to write deterministic differential equations to describe how the probabilities of the states of individuals and pairs change over time.

We find that a major difference between evolutionary graph theory and other areas in which these methods have been applied is that here, event probabilities depend on the states of all individuals in the population. As a result, we do not obtain a precise BBGKY-like hierarchy, which relies on neighbouring particle-particle interactions. Another difference is that in evolutionary dynamics, we have two absorbing states, which potentially leads to system-wide correlations that cannot be captured on a local level. It is worth noting that some alternative nearest-neighbour interaction evolutionary models, which may yield such a hierarchy directly, have also been considered [178]; however, in this work we have restricted our attention to the classic evolutionary graph theory dynamics.

In spite of these differences, some progress could be made towards approximating evolutionary dynamics. The first step was to write down equations for the rate of change of the state probabilities for individual nodes (Theorem 4.1.1). This led to equations which required conditioning against the probability of the state of the entire system, and therefore required the development of methods to simplify this. Motivated by an objective of deriving homogenised pair-approximation models used in the literature, our first approach was to modify the replacement rate by removing the normalisation by the total fitness (Section 4.2.1). This has the effect of altering the speed at which events occur but does not alter the final fixation probability. The resulting system of equations describes individual and pair probabilities in terms of the probability of their entire neighbourhoods. This could provide a basis to accurately approximate the fixation probability by finding appropriate moment closures to express the neighbourhoods as functions of individual and pair probabilities. However, this is difficult to implement and the number of equations increases exponentially with the maximum degree of the graph, making it infeasible in general without further approximation. By making further assumptions about the graph such that all individuals and pairs of a given type are identical and interchangeable, we were able to derive the homogenised pair approximation models [59, 117], which have been shown to give interesting results for various evolutionary games.

To obtain an approximation which is numerically feasible in general, we first ignored any conditioning, similar to a model in [175] which uses this assumption to construct a population level approximation. The resulting model (Equation (4.12)) was found to work

well for small graphs and contains the exact neutral drift model [164] as a special case. However, as population size increases, the predictions for the fixation probability of a single mutant individual were observed to tend to zero. By solving this system for the fixation probability on a complete graph, we obtained a scaling factor which enabled this model to give a reasonable prediction of fixation probability from a given initial condition with a single mutant individual on any graph. Due to the construction of this method, it will perform best on graphs which yield average fixation probability close to the Moran probability.

To generate a more accurate model and one which does not require an artificial scaling factor, we investigated models with some level of conditioning (Section 4.2.3). Conditioning against a single node results in the same level of complexity as conditioning against pairs of nodes and so we elected to produce results for the latter. In this case, we conditioned against the pair of nodes directly involved in the replacement event. However, in order to use this model on large graphs, we require the use of moment closure approximations. We found that the standard method used in other areas with different closures for open and closed triples [82, 166] was not effective here because while it provides very good results on simple structures, on most graphs it predicts fixation probabilities of either zero or one. It seems likely that this is caused by neglecting important graph-wide correlations across open triples associated with the two absorbing states of the system.

By using the Kirkwood closure method for all triples, including open ones, we obtained a method which provides informative predictions on the majority of graphs tested. We investigated square lattices and star-type graphs, as these are two extreme population structures which we use as worst case scenarios. The lattice is extreme as moment closure methods do not perform well on such graphs. The star is extreme because this type of graph significantly amplifies the fixation probability, which seems to amplify the accumulated error in the approximation methods. For all three types of random graph considered, and Zachary's karate club, this method provides a reasonable approximation to the fixation probability. When the degree of the initial mutant node is not low the approximation can be very accurate. However, if we initiate on a low degree node, the method performs less well, potentially due to such nodes amplifying the fixation probability in the invasion process, again leading to inaccuracies in the solution being amplified. Despite potential inaccuracies in the fixation probability approximation, we observe that this method is particularly accurate for the early-time behaviour of these systems for any graph, and therefore can give interesting insights into this behaviour. The method is computationally

feasible for reasonably large N, however, the computational complexity scales with N^2 rather than with N which is more typical for epidemic models. Nevertheless, this still represents a significant reduction over the master equation which scales with 2^N.

The novelty of this work is the adaption of well-established techniques from other fields to the study of evolutionary dynamics at the level of individual nodes. The contribution is two-fold. Firstly we have obtained insight into existing models by deriving them from the master equation. Secondly, the advantage of looking at node-level quantities rather than a homogenised model is that we gain the ability to compare dynamics from different initial conditions on the same graph, which is not present in many other approximation methods. Furthermore, the initial dynamics of Method 4 are very accurate (Figure 4.2), allowing us to see how the probability of each node being a mutant flows through the population. Although we chose to work in continuous time here and examples study the invasion process, similar methods could be followed directly in discrete-time and the methods are applicable to any Markovian update rule.

Appendix 4.A Proof of Theorem 4.1.1

Proof. By total probability rules we have that

$$\frac{dP(A^t_{\{i\}})}{dt} = \frac{d\left[\sum_{X_{V\setminus\{i\}}} P(A^t_{\{i\}}X^t_{V\setminus\{i\}})\right]}{dt} = \sum_{X_{V\setminus\{i\}}} \frac{dP(A^t_{\{i\}}X^t_{V\setminus\{i\}})}{dt}, \tag{4.20}$$

where $X_{V\setminus\{i\}}$ is the state of the nodes in the system not including i.

Consider a set state $X_{V\setminus\{i\}}$ of the remaining nodes. The rate of change in the full system state probability $P(A^t_{\{i\}}X^t_{V\setminus\{i\}})$ is given by

$$\begin{aligned}\frac{dP(A^t_{\{i\}}X^t_{V\setminus\{i\}})}{dt} = &\sum_{Y_{V\setminus\{i\}}} P(A^t_{\{i\}}Y^t_{V\setminus\{i\}})\chi(A^t_{\{i\}}Y^t_{V\setminus\{i\}} \to A^t_{\{i\}}X^t_{V\setminus\{i\}}) \\ &+ P(B^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(B^t_{\{i\}}X^t_{V\setminus\{i\}} \to A^t_{\{i\}}X^t_{V\setminus\{i\}}) \\ &- \sum_{Y_{V\setminus\{i\}}} P(A^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(A^t_{\{i\}}X^t_{V\setminus\{i\}} \to A^t_{\{i\}}Y^t_{V\setminus\{i\}})\end{aligned}$$

$$- P(A^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(A^t_{\{i\}}X^t_{V\setminus\{i\}} \to B^t_{\{i\}}X^t_{V\setminus\{i\}}), \tag{4.21}$$

where $\chi(A^t_{\{i\}}X^t_{V\setminus\{i\}} \to B^t_{\{i\}}X^t_{V\setminus\{i\}})$ is the rate at which the system moves from state $A^t_{\{i\}}X^t_{V\setminus\{i\}}$ to state $B^t_{\{i\}}X^t_{V\setminus\{i\}}$.

Consider the terms which involve changing the state of the individual in node i in Equation (4.21). By expanding the rate into the sum of separate event rates we obtain

$$P(B^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(B^t_{\{i\}}X^t_{V\setminus\{i\}} \to A^t_{\{i\}}X^t_{V\setminus\{i\}}) = \\ P(B^t_{\{i\}}X^t_{V\setminus\{i\}})\sum_{j=1}^{N} G_{ij}\chi(\Omega^t_{j\to i}|B^t_{\{i\}}X^t_{V\setminus\{i\}})\mathbb{1}_{(A^t_{\{j\}}\in X^t_{V\setminus\{i\}})},$$

and

$$P(A^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(A^t_{\{i\}}X^t_{V\setminus\{i\}} \to B^t_{\{i\}}X^t_{V\setminus\{i\}}) = \\ P(A^t_{\{i\}}X^t_{V\setminus\{i\}})\sum_{j=1}^{N} G_{ij}\chi(\Omega^t_{j\to i}|A^t_{\{i\}}X^t_{V\setminus\{i\}})\mathbb{1}_{(B^t_{\{j\}}\in X^t_{V\setminus\{i\}})},$$

where $\mathbb{1}_{(B^t_{\{j\}}\in X^t_{V\setminus\{i\}})}$ is an indicator function on the event $B^t_{\{j\}}$ being part of the event $X^t_{V\setminus\{i\}}$. That is, the state of node j in the state X is type B. The $\chi(\Omega^t_{j\to i}|A^t_{\{i\}}X^t_{V\setminus\{i\}})$ term is the rate at which the individual in node j replaces the individual in node i, given that the system is in state $A^t_{\{i\}}X^t_{V\setminus\{i\}}$, as defined in Definition 4.1.1. Rearranging these and substituting into Equation (4.21) gives

$$\begin{aligned}\frac{dP(A^t_{\{i\}}X^t_{V\setminus\{i\}})}{dt} =& \sum_{j=1}^{N} G_{ij}P(B^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(\Omega^t_{j\to i}|B^t_{\{i\}}X^t_{V\setminus\{i\}})\mathbb{1}_{(A^t_{\{j\}}\in X^t_{V\setminus\{i\}})} \\ &- \sum_{j=1}^{N} G_{ij}P(A^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(\Omega^t_{j\to i}|A^t_{\{i\}}X^t_{V\setminus\{i\}})\mathbb{1}_{(B^t_{\{j\}}\in X^t_{V\setminus\{i\}})} \\ &+ \sum_{Y_{V\setminus\{i\}}} P(A^t_{\{i\}}Y^t_{V\setminus\{i\}})\chi(A^t_{\{i\}}Y^t_{V\setminus\{i\}} \to A^t_{\{i\}}X^t_{V\setminus\{i\}}) \\ &- \sum_{Y_{V\setminus\{i\}}} P(A^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(A^t_{\{i\}}X^t_{V\setminus\{i\}} \to A^t_{\{i\}}Y^t_{V\setminus\{i\}}).\end{aligned}$$

By substituting this into Equation (4.20) we obtain

$$\begin{aligned}\frac{dP(A^t_{\{i\}})}{dt} = &\sum_{X_{V\setminus\{i\}}}\sum_{j=1}^{N} G_{ij}P(B^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(\Omega^t_{j\to i}|B^t_{\{i\}}X^t_{V\setminus\{i\}})\mathbb{1}_{(A^t_{\{j\}}\in X^t_{V\setminus\{i\}})}\\ &- \sum_{X_{V\setminus\{i\}}}\sum_{j=1}^{N} G_{ij}P(A^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(\Omega^t_{j\to i}|A^t_{\{i\}}X^t_{V\setminus\{i\}})\mathbb{1}_{(B^t_{\{j\}}\in X^t_{V\setminus\{i\}})}\\ &+ \sum_{X_{V\setminus\{i\}}}\sum_{Y_{V\setminus\{i\}}} P(A^t_{\{i\}}Y^t_{V\setminus\{i\}})\chi(A^t_{\{i\}}Y^t_{V\setminus\{i\}}\to A^t_{\{i\}}X^t_{V\setminus\{i\}})\\ &- \sum_{X_{V\setminus\{i\}}}\sum_{Y_{V\setminus\{i\}}} P(A^t_{\{i\}}X^t_{V\setminus\{i\}})\chi(A^t_{\{i\}}X^t_{V\setminus\{i\}}\to A^t_{\{i\}}Y^t_{V\setminus\{i\}}).\end{aligned}$$

Clearly the last two sums cancel, so we can simplify this to

$$\begin{aligned}\frac{dP(A^t_{\{i\}})}{dt} =&\sum_{j=1}^{N}\sum_{X_{V\setminus\{i,j\}}} G_{ij}P(B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j\}})\chi(\Omega^t_{j\to i}|B^t_{\{i\}}A^t_{\{j\}}X^t_{V\setminus\{i,j\}})\\ &-\sum_{j=1}^{N}\sum_{X_{V\setminus\{i,j\}}} G_{ij}P(A^t_{\{i\}}B^t_{\{j\}}X^t_{V\setminus\{i,j\}})\chi(\Omega^t_{j\to i}|A^t_{\{i\}}B^t_{\{j\}}X^t_{V\setminus\{i,j\}}),\end{aligned}$$

as required. □

Appendix 4.B Derivation of the scaling factor (Equation 4.13)

Consider a system with rate of change given by

$$\frac{d\bar{P}(A^t_{\{i\}})}{dt} = \sum_{j=1}^{N} G_{ij}\bar{P}(A^t_{\{j\}})\chi(\Omega^t_{j\to i}) - \sum_{j=1}^{N} G_{ij}\bar{P}(A^t_{\{i\}})\chi(\Omega^t_{j\to i}).$$

Since we are interested in the complete graph, we have that $G_{ij} = 1$ for $j \neq i$, and $G_{i,i} = 0$. Let A_c denote the average probability that a node is of type A on the complete graph at time t. That is

$$A_c(t) = \frac{1}{N}\sum_{j=1}^{N}\bar{P}(A^t_{\{j\}}) = \frac{S}{N}.$$

Since we are considering constant fitness we have

$$\chi(\Omega^t_{j\to i}) = \frac{\bar{P}(A^t_{\{j\}})(r-1)+1}{\sum_{k=1}^{N}\bar{P}(A^t_{\{k\}})(r-1)+1} = \frac{\bar{P}(A^t_{\{j\}})(r-1)+1}{N+(r-1)S},$$

which gives us

$$\frac{dS}{dt} = \sum_{i=1}^{N}\frac{d\bar{P}(A^t_{\{i\}})}{dt} = \frac{\sum_{i,j=1}^{N}(\bar{P}(A^t_{\{j\}})-\bar{P}(A^t_{\{i\}}))(\bar{P}(A^t_{\{j\}})(r-1)+1)}{N+(r-1)S}.$$

Writing $G = \sum_{i,j=1}^{N}(\bar{P}(A^t_{\{j\}})-\bar{P}(A^t_{\{i\}}))\bar{P}(A^t_{\{j\}})$, and $H = \sum_{i,j=1}^{N}(\bar{P}(A^t_{\{j\}})-\bar{P}(A^t_{\{i\}}))$ we have

$$\frac{dS}{dt} = \frac{(r-1)G+H}{N+(r-1)S}.$$

Clearly $H = 0$, so we obtain

$$\frac{dS}{dt} = \frac{(r-1)G}{N+(r-1)S}.$$

Note that $\sum_{i,j=1}^{N}(\bar{P}(A^t_{\{j\}})-\bar{P}(A^t_{\{i\}}))^2 = \sum_{i,j=1}^{N}\bar{P}(A^t_{\{j\}})^2+\bar{P}(A^t_{\{i\}})^2-2\bar{P}(A^t_{\{j\}})\bar{P}(A^t_{\{i\}}) = 2G$, so that

$$\frac{dG}{dt} = \frac{1}{2}\frac{d}{dt}\Big(\sum_{i,j=1}^{N}(\bar{P}(A^t_{\{j\}})-\bar{P}(A^t_{\{i\}}))^2\Big) = \sum_{i,j=1}^{N}(\bar{P}(A^t_{\{j\}})-\bar{P}(A^t_{\{i\}}))\frac{d(\bar{P}(A^t_{\{j\}})-\bar{P}(A^t_{\{i\}}))}{dt}.$$

Considering the last term on the right hand side we have

$$\begin{aligned}\frac{d}{dt}(\bar{P}(A^t_{\{i\}})-\bar{P}(A^t_{\{j\}})) = \frac{1}{N+(r-1)S}\sum_{k=1}^{N}\big(&\bar{P}(A^t_{\{k\}})(\bar{P}(A_k)^t-\bar{P}(A^t_{\{i\}}))+\bar{P}(A^t_{\{k\}})(\bar{P}(A^t_{\{j\}})\\ &-\bar{P}(A^t_{\{k\}}))\big)(r-1)+(\bar{P}(A^t_{\{k\}})-\bar{P}(A^t_{\{i\}}))\\ &+(\bar{P}(A^t_{\{j\}})-\bar{P}(A^t_{\{k\}}))\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sum_{k=1} \bar{P}(A^t_{\{k\}})(\bar{P}(A^t_{\{j\}}) - \bar{P}(A^t_{\{i\}}))(r-1) + (\bar{P}(A^t_{\{j\}}) - \bar{P}(A^t_{\{i\}}))}{N + (r-1)S} \\
&= \frac{(\bar{P}(A^t_{\{j\}}) - \bar{P}(A^t_{\{i\}}))\big((r-1)S + N\big)}{N + (r-1)S} \\
&= -(\bar{P}(A^t_{\{i\}}) - \bar{P}(A^t_{\{j\}})).
\end{aligned}
$$

Thus,

$$
\frac{dG}{dt} = \sum_{i,j=1}^{N} (\bar{P}(A^t_{\{j\}}) - \bar{P}(A^t_{\{i\}}))^2 = -2G \implies G = Ae^{-2t} = (N-m)me^{-2t},
$$

since $G(0) = (N-m)m$. Therefore we have

$$
\frac{dS}{dt} = \frac{(r-1)(N-m)me^{-2t}}{N + (r-1)S}
$$

$$
\Rightarrow NS + \frac{r-1}{2}S^2 = -\frac{1}{2}(r-1)(N-m)me^{-2t} + C.
$$

At $t = 0$ we have $S = \sum \bar{P}(A^t_{\{j\}}) = m$, which gives

$$
C = Nm + \big(\frac{r-1}{2}\big)Nm = Nm\big(\frac{r+1}{2}\big),
$$

and so we can solve to obtain

$$
S = \frac{\big(-N \pm \sqrt{N^2 + 4\frac{r-1}{2}\big(Nm\frac{r+1}{2} - (N-m)m\frac{r-1}{2}e^{-2t}\big)}\big)}{r-1}.
$$

Only the positive root makes sense, so we obtain

$$
A_c = \frac{1}{r-1}\big(-1 + \sqrt{1 + \frac{m(r^2-1)}{N} - (r-1)^2\frac{(N-m)m}{N^2}e^{-2t}}\big).
$$

Thus, we have $\lim_{t\to\infty} A_c(t) = \frac{1}{r-1}\big(-1 + \sqrt{1 + \frac{m(r^2-1)}{N}}\big)$.

Chapter 5

Approximating the quasi-stationary distribution in network-based SIS

Understanding the stochastic dynamics of the Markovian network-based SIS model (Section 1.4.1) can be challenging since a large number of simulations are required, which uses a lot of computational time and yields relatively little in theoretical insight. Deterministic SIS models, such as mean-field models [96, 184, 185, 187] and pair-approximation models [50, 61, 83, 82, 91, 107, 167], on the other hand, can yield theoretical insights. One example is the epidemic threshold, below which the pathogen goes extinct and above which the pathogen reaches a stable endemic equilibrium solution [136]. These deterministic models can be interpreted as modified versions of stochastic SIS dynamics, which are obtained by making statistical independence assumptions. However, the presence of this stable endemic equilibrium means it is not clear how to relate these results back to the underlying stochastic process, since the only stable solution to the stochastic model is the disease-free state. Although this is the only stable solution, the time it takes to be reached can be extremely long. Therefore, the dynamics of the pathogen before absorption are important to understand, which is given by the quasi-stationary distribution (Section 1.1.9).

Quasi-stationary behaviour is commonly used to describe the long-term behaviour of finite Markov chains with absorbing states. Examples include: modelling the spread of a computer virus across a network with cure and reinfection [85, 119, 139, 191], chemical

reactions in which materials or catalysts can be exhausted [31, 32, 133, 137, 146], and wildlife management models [73, 93, 111, 135, 148, 147, 161]. Within Markovian SIS dynamics, various statistics have been derived using the concept of the QSD [5, 9, 10, 62]. This includes use by Wilkinson and Sharkey [193] to derive a measure of the invasion probability, by Ferreira and colleagues [44] to approximate the epidemic threshold, and by Nåsell [122] to account for the influence of epidemic and demographic forces on the time to extinction.

These statistics provide theoretical insights into the epidemic, however the calculation of the QSD can require a large number of stochastic simulations, and therefore it is necessary to derive approximation methods. Thus far, approximations have focused on well-mixed populations. Kriscio and Lefevre [95] used a conditional birth-and-death process to approximate the QSD of Markovian SIS epidemic dynamics, which has since been extended by Nåsell [120, 121]. Allen and Burgin [4] used a system of ordinary differential equations to approximate the conditional probability distribution, which yields the QSD as a steady-state.

In network-structured populations, the endemic equilibrium of the deterministic SIS models approximates the quasi-stationary distribution when sufficiently above the epidemic threshold [91, 184]. The aim of this chapter is to extend these methods (Section 1.4.2) to approximate the expected number of infected individuals in the QSD for all parameter values. Whilst we focus on SIS dynamics, the techniques we use are applicable to other Markov processes. For example, in Markovian network-based SIR the methods can be applied similarly by building on the pair-based approximations in [168]. The complication in extending this work to SIR is in defining the absorbing states, since there will be multiple disease-free states.

This chapter is structured as follows. In Section 5.1, we recap the Markovian network-based SIS modelling framework and the master equation. In Section 1.4.2 we described moment-closure approximations to the SIS model. These are hard to relate to the stochastic process, so we aim to develop a novel modelling framework that is directly related to this process. We do this via the quasi-stationary distribution, which we define in Section 5.2. Sections 5.3 and 5.4 focus on deriving individual-based and pair-based approximation methods that capture the QSD. We then analyse the performance of the proposed methods on different contact networks in Section 5.5.

5.1 Markovian SIS dynamics on a contact network

In this section, we briefly recap the SIS dynamics, as defined in Section 1.4.1. We consider a population of N individuals. The model is described by a continuous-time Markov chain on this population, parameterised by an $N \times N$ irreducible square matrix T with non-negative entries and an $N \times 1$ vector γ with positive entries. The matrix T is the transmission matrix of the epidemic, and represents the rate at which individuals infect each other. The vector γ represents the rates at which each individual in the population recovers.

Let $\sigma_\alpha \in \{S, I\}^N$ denote a state of the population. We assume throughout that state σ_1 corresponds to the all susceptible state. Let $\Sigma_i(t)$ denote the status of individual i at time t, and for a given state σ_α, let $\sigma_{\alpha i}$ denote the status of individual i in that state. The time evolution of the Markov chain is captured by the master equation,

$$\frac{d\mathbf{p}(t)}{dt} = Q\mathbf{p}(t), \tag{5.1}$$

where $P_\alpha(t) = P(\Sigma(t) = \sigma_\alpha)$ is the probability that the system is in state σ_α at time $t \geq 0$, and Q is a matrix of transition rates (obtained from Table 1.1). In particular, $P_1(t)$ denotes the probability that all individuals are susceptible at time t. Although this can be solved to determine the future behaviour, in many cases this is infeasible since the matrix Q grows exponentially with N.

5.2 The quasi-stationary distribution

Let us construct a vector $\rho(t)$, such that $\rho_\alpha(t)$ is the conditional probability that the system is in state σ_α at time t given that at least one individual is infected; i.e. $\rho_\alpha(t) = P(\Sigma(t) = \sigma_\alpha | \Sigma(t) \neq \sigma_1)$, where σ_1 is the disease-free state. We have

$$\begin{aligned} \rho_\alpha(t) = \frac{P(\Sigma(t) = \sigma_\alpha \cap \Sigma(t) \neq \sigma_1)}{P(\Sigma(t) \neq \sigma_1)} &= \frac{P(\Sigma(t) = \sigma_\alpha \cap \Sigma(t) \neq \sigma_1)}{1 - P(\Sigma(t) = \sigma_1)} \\ &= \frac{P(\Sigma(t) = \sigma_\alpha)}{1 - P(\Sigma(t) = \sigma_1)} = \frac{P_\alpha(t)}{(1 - P_1(t))}, \end{aligned} \tag{5.2}$$

for $\alpha \neq 1$. For $\alpha = 1$, $\rho_1(t) = 0$ for all t. Here we have assumed that $P_1(t) \neq 1$. Using (5.2) and the master equation (5.1), the time derivative of $\rho_\alpha(t)$ is given by

$$\frac{d\rho_\sigma}{dt} = \begin{cases} 0 & \text{if } \alpha = 1 \\ \frac{d\rho_\alpha}{dt} = \frac{d}{dt}\left(\frac{P_\alpha(t)}{(1-P_1(t))}\right) = \frac{(QP)_\alpha}{1-P_1} + \frac{P_\alpha (QP)_1}{(1-P_1)^2} & \text{if } \alpha = 2, 3, \ldots, 2^N. \end{cases} \tag{5.3}$$

We suppress the explicit time dependence of P and ρ in favour of compactness.

The state space for the Markov chain is finite and consists exhaustively of one absorbing state and a communicating class of transient states. The non-absorbing states form a communicating class of transient states because the contact network is strongly connected and the vector γ of recovery rates is positive. Thus, there exists a unique quasi-stationary distribution (QSD), independent of initial conditions, which is equivalent to the limiting conditional distribution [33]. This QSD is similar to a true stationary distribution in that it is invariant under the conditional distribution. Therefore, the QSD, q, defined as

$$q_\alpha = \begin{cases} 0 & \text{if } \alpha = 1 \\ \frac{P^*_\alpha}{1-P^*_1}, & \text{if } \alpha = 2, 3, \ldots, 2^N, \end{cases} \tag{5.4}$$

for some P^*, is a globally stable equilibrium of system (5.3). Therefore, to find q we need to find some P^* satisfying

$$0 = \frac{(QP^*)_\alpha}{1 - P^*_1} + \frac{P^*_\alpha (QP^*)_1}{(1 - P^*_1)^2} \qquad \alpha = 2, 3, \ldots, 2^N. \tag{5.5}$$

However, finding P^* directly is in many cases infeasible since the size of Q grows exponentially with the population size.

5.3 Individual-based approximations

Instead of solving Equation (5.5) exactly, approximations can be made by making assumptions regarding correlations between individual nodes, similar to the standard approximation models described in Section 1.4.2.

5.3.1 Node-level equations

The probability of a node i being infected under the conditional distribution ρ is given by $\rho(\Sigma_i(t) = I) = \sum_{\alpha:\sigma_{\alpha i}=I} \rho_\alpha(t)$. Taking the derivative and using (5.2),

$$\frac{d}{dt}\left(\rho(\Sigma_i(t) = I)\right) = \sum_{\alpha:\sigma_{\alpha i}=I} \frac{d\rho_\alpha}{dt} = \frac{\sum_{\alpha:\sigma_{\alpha i}=I}(QP)_\alpha}{1-P_1} + \frac{(QP)_1}{(1-P_1)^2}\sum_{\alpha:\sigma_{\alpha i}=I} P_\alpha. \tag{5.6}$$

The numerator on the first term corresponds to the rate of change in the probability that node i is infected, which is given by $\langle \dot{I}_i \rangle$ in Equation (1.19). The summation in the second term corresponds to the probability that node i is infected, $\langle I_i \rangle$. Therefore, we can write

$$\frac{d}{dt}\left(\rho(\Sigma_i(t) = I)\right) = \frac{\langle \dot{I}_i \rangle}{1-P_1} + \frac{(QP)_1}{(1-P_1)^2}\langle I_i \rangle.$$

$(QP)_1$ is the rate at which the system enters the absorbing state. The system can only reach the absorbing state from a state with a single infected individual, in node j for example, which transitions to the all susceptible state at rate γ_j. Therefore $(QP)_1 = \sum_j \gamma_j \langle I_j S \rangle$, where we use $\langle I_j S \rangle$ to denote the probability that node j is infected and all other nodes are susceptible. Using this along with Equation (1.19), we obtain

$$\frac{d}{dt}\left(\rho(\Sigma_i(t) = I)\right) = \frac{\sum_j T_{ij}\langle S_i I_j \rangle - \gamma_i \langle I_i \rangle}{1-P_1} + \frac{\langle I_i \rangle}{(1-P_1)^2}\sum_j \gamma_j \langle I_j S \rangle. \tag{5.7}$$

Approximating (5.7) by assuming that the states of individuals are independent gives

$$\frac{d}{dt}\left(\rho(\Sigma_i(t) = I)\right) \approx \frac{\sum_j T_{ij}\langle S_i \rangle\langle I_j \rangle - \gamma_i \langle I_i \rangle}{1-\prod_k \langle S_k \rangle} + \frac{\langle I_i \rangle}{(1-\prod_k \langle S_k \rangle)^2}\sum_j \gamma_j \langle I_j \rangle \prod_{k\neq j} \langle S_k \rangle. \tag{5.8}$$

The probability of node i being infected in the quasi-stationary distribution is given by $\langle I_i^* \rangle/(1-P_1^*)$, where P^* is a solution to Equation (5.5). Therefore, to find the approximation to this quantity under this independence assumption, we first need to find a steady state of Equation (5.8), which is given by vectors $\langle X \rangle^*$ and $\langle Y \rangle^*$ satisfying,

$$0 \;=\; \frac{\sum_j T_{ij}\langle X_i \rangle^*\langle Y_j \rangle^* - \gamma_i \langle Y_i \rangle^*}{1-\prod_k \langle X_k \rangle^*} + \frac{\langle Y_i \rangle^*}{(1-\prod_k \langle X_k \rangle^*)^2}\sum_j \gamma_j \langle Y_j \rangle^* \prod_{k\neq j} \langle X_k \rangle^*, \tag{5.9}$$

for all i. From $\langle X\rangle^*$ and $\langle Y\rangle^*$, the probability that i is infected in the QSD is approximated by computing $\langle Y_i\rangle^*/(1-\prod_k\langle X_k\rangle^*)$. However, we are only interested in solutions of (5.9) that are feasible; i.e. $\langle Y_i\rangle^* \in [0,1], \langle X_i\rangle^* = 1-\langle Y_i\rangle^*$ for all i. To obtain such a solution, define

$$\begin{aligned}\langle \dot{Y}_i\rangle &= \sum_j T_{ij}\langle X_i\rangle\langle Y_j\rangle - \gamma_i\langle Y_i\rangle + \frac{\langle Y_i\rangle \sum_j \gamma_j\langle Y_j\rangle \prod_{k\neq j}\langle X_k\rangle}{1-\prod_k\langle X_k\rangle} \\ \langle X_i\rangle &= 1-\langle Y_i\rangle. \end{aligned} \tag{5.10}$$

Equation (5.10) is positively invariant in $[0,1]^N$ (see Appendix 5.B.1), so therefore provided $\langle X_i\rangle \in [0,1]$ and $\langle Y_i\rangle \in [0,1]$ at $t=0$, any solution will be a feasible solution to Equation (5.9). The equilibrium can be found by solving Equation (5.10) until a steady state is reached. Note that $\langle Y_i\rangle = 0$ for all i is an equilibrium of this system; however, since we are studying the QSD, only non-zero solutions are interesting.

Theorem 1. *For a complete network, when the transmission rate is non-zero there exists a steady-state solution to Equation (5.10) in the interval* $(0,1)$.

Proof. See Appendix 5.B.2.

We have not proven uniqueness, but this holds numerically for all parameters tested. This existence and uniqueness also holds numerically for complex networks. Since there is always a solution in $(0,1)$, the solution only approaches 0 in the limit $T_{ij}\to 0$ for all (i,j) (or $\gamma\to\infty$).

Theorem 2. *As $T_{ij}\to 0$ for all (i,j) (or $\gamma\to\infty$), the QSD approximation, $\sum_i\langle Y_i\rangle/(1-\prod_k\langle X_k\rangle)$, approaches 1.*

Proof. See Appendix 5.B.3.

This is consistent with the true QSD, for which the expected number of infected individuals is bounded below by 1, and shows that this model has the required properties. It has been observed that the steady-state of the standard individual-based model (Equation 1.19) captures the dynamics of the meta-stable state when sufficiently above the epidemic threshold [184]. This meta-stable state corresponds with the QSD, and hence in this region the standard model captures the QSD. This can be seen by comparing the standard model to the QSD model. In Equation (5.10), since the probability of each node being

susceptible decreases as the transmission rate increases, if the population is sufficiently large the product over susceptible nodes approaches zero. In this case, Equation (5.10) tends towards the standard individual-based model.

5.3.2 Degree heterogeneous population-level equations

The node-level equations give detailed insight into the dynamics of individual nodes in the QSD, however the number of equations scales with N. To build approximations with a reduced number of equations, population-level models can be constructed for undirected networks. The rate of change in the expected number of infected individuals with a given degree, under the conditional distribution, is found by taking the sum over the probability that each node with this degree is infected

$$\sum_{i:k_i=k}\sum_{\alpha:\sigma_{\alpha i}=I}\frac{d\rho_\alpha}{dt} = \sum_{i:k_i=k}\left(\frac{\sum_j T_{ij}\langle S_iI_j\rangle-\gamma_i\langle I_i\rangle}{1-P_1}+\frac{\langle I_i\rangle}{(1-P_1)^2}\sum_j\gamma_j\langle I_jS\rangle\right).$$

The numerator in the first term on the right-hand side is the rate of change in the probability that an individual is infected. Taking the sum over all nodes with the same degree, this gives the rate of change in the expected number of infected individuals with that degree, which is given by Equation (1.18). Taking the sum of $\langle I_i\rangle$ over all nodes with the same degree gives the expected number of infected nodes with that degree. Therefore, assuming $T_{ij}=\tau$ whenever $T_{ij}>0$, we obtain

$$\sum_{i:k_i=k}\sum_{\alpha:\sigma_{\alpha i}=I}\frac{d\rho_\alpha}{dt} = \frac{\tau\sum_{l\in\mathcal{M}}[S_kI_l]-\gamma[I_k]}{1-P_1}+\frac{[I_k]}{(1-P_1)^2}\sum_j\gamma\langle I_jS\rangle, \tag{5.11}$$

where $[A_k]$ is the expected number of individuals with degree k in state A, $[A_kB_l]$ is the expected number of pairs between individuals of degree k and degree l, in states A and B respectively, and k_i is the degree of node i. Above, and throughout, all expected numbers are with respect to the standard probability measure P. Assuming that the states of individuals are independent, (5.11) becomes

$$\sum_{i:k_i=k}\sum_{\alpha:\sigma_{\alpha i}=I}\frac{d\rho_\alpha}{dt} \approx \frac{\tau\sum_{l\in\mathcal{M}}|C_{k,l}|\frac{[S_k]}{|C_k|}\frac{[I_l]}{|C_l|}-\gamma[I_k]}{1-\prod_j\langle S_j\rangle}+\frac{[I_k]}{(1-\prod_j\langle S_j\rangle)^2}\sum_j\gamma\langle I_j\rangle\prod_{k\neq j}\langle S_k\rangle$$

where $|C_k|$ is the number of degree k nodes in the network and $|C_{k,l}|$ is the number of pairs between degree k and degree l nodes. This equation is not closed, since the final term and the denominators depend on node-level quantities. However, from (1.22) the node-level quantities can be approximated by assuming $\langle S_j \rangle = [S_k]/|C_k|$, where k is the degree of node j. Therefore

$$\prod_i \langle S_i \rangle \approx \prod_{l \in \mathcal{M}} \left(\frac{[S_l]}{|C_l|} \right)^{|C_l|},$$

and

$$\gamma \langle I_j \rangle \prod_{i \neq j} \langle S_i \rangle \approx \frac{[I_k]}{|C_k|} \left(\frac{[S_k]}{|C_k|} \right)^{|C_k|-1} \prod_{l \in \mathcal{M}: l \neq k} \left(\frac{[S_l]}{|C_l|} \right)^{|C_l|}, \tag{5.12}$$

where k is the degree of node j. Multiplying Equation (5.12) by the number of degree k nodes, $|C_k|$, we obtain the probability of a single degree k node being infected, which we denote $\tilde{P}(I_k = 1)$. Therefore, we obtain

$$\sum_{i:k_i=k} \sum_{\alpha:\sigma_{\alpha i}=I} \frac{d\rho_\alpha}{dt} \approx \frac{\tau \sum_{l \in \mathcal{M}} |C_{k,l}| \frac{[S_k]}{|C_k|} \frac{[I_l]}{|C_l|} - \gamma [I_k]}{(1 - \prod_l (\frac{[S_l]}{|C_l|})^{|C_l|})} + \frac{[I_k]}{(1 - \prod_l (\frac{[S_l]}{|C_l|})^{|C_l|})^2} \sum_{l \in \mathcal{M}} \gamma \tilde{P}(I_l = 1).$$

To find a steady-state, we need to find vectors $\langle X \rangle^*$ and $\langle Y \rangle^*$ satisfying

$$0 = \frac{\tau \sum_{l \in \mathcal{M}} |C_{k,l}| \frac{[X_k]^*}{|C_k|} \frac{[Y_l]^*}{|C_l|} - \gamma [Y_k]^*}{(1 - \prod_l (\frac{[X_l]^*}{|C_l|})^{|C_l|})} + \frac{[Y_k]^*}{(1 - \prod_l (\frac{[X_l]^*}{|C_l|})^{|C_l|})^2} \sum_{l \in \mathcal{M}} \gamma \tilde{P}(Y_l = 1)^*$$

from which we can approximate the expected number of infected degree k individuals in the QSD by computing $[Y_k]^*/(1 - \prod_l (\frac{[X_l]^*}{|C_l|})^{|C_l|})$. We require $[Y_k]^* \in [0, |C_k|]$, $[X_k]^* = |C_k| - [Y_k]^*$ for all i. Such a solution can be found by defining

$$\begin{aligned} \dot{[Y_k]} &= \tau \sum_{l \in \mathcal{M}} |C_{k,l}| \frac{[X_k]}{|C_k|} \frac{[Y_l]}{|C_l|} - \gamma [Y_k] + \frac{[Y_k] \sum_{l \in \mathcal{M}} \gamma \tilde{P}(Y_l = 1)}{(1 - \prod_l (\frac{[X_l]}{|C_l|})^{|C_l|})} \\ [X_k] &= |C_k| - [Y_k] \\ \tilde{P}(Y_k = 1) &= |C_k| \frac{[Y_k]}{|C_k|} \left(\frac{[X_k]}{|C_k|} \right)^{|C_k|-1} \prod_{l \in \mathcal{M}: l \neq k} \left(\frac{[X_l]}{|C_l|} \right)^{|C_l|}, \end{aligned} \tag{5.13}$$

and specifying that $[Y_k(0)] \in [0, |C_k|]$ for all k and calculating the steady-state. Any solution will be a valid solution, since Equation (5.13) is bounded such that $[Y_k]^* \in [0, |C_k|]$

for all k (this can be shown using a method similar to Appendix 5.B.1).

5.4 Pair-based approximations

Assuming statistical independence of individuals yields a system with low computational complexity, however the assumption is not necessarily realistic. This is because the pathogen spreads through contact between neighbouring individuals, and hence it is likely that the state of an individual is dependent on its neighbours. To account for this pairwise correlation, we can instead assume statistical independence of pairs. Doing this closes the system by describing triple and higher order terms as functions of individual and pair terms.

5.4.1 Node-level equations

Following the logic of the individual-based derivation, we arrive at the following pair-based model (see Appendix 5.C),

$$\begin{aligned}
\dot{\langle Y_i\rangle} &= \sum_j T_{ij}\langle X_iY_j\rangle - \gamma_i\langle Y_i\rangle + \frac{\langle Y_i\rangle\sum_j \gamma_j\langle\widetilde{Y_jX}\rangle}{1-\langle\sigma_1\rangle},\\
\dot{\langle X_iY_j\rangle} &= \sum_{k\in\mathcal{N}_j\setminus i} T_{jk}\frac{\langle X_iX_j\rangle\langle X_jY_k\rangle}{\langle X_j\rangle} - \sum_{k\in\mathcal{N}_i\setminus j} T_{ik}\frac{\langle Y_kX_i\rangle\langle X_iY_j\rangle}{\langle X_i\rangle}\\
&\quad -(T_{ij}+\gamma_j)\langle X_iY_j\rangle + \gamma_i\langle Y_iY_j\rangle + \frac{\langle X_iY_j\rangle\sum_j \gamma_j\langle\widetilde{Y_jX}\rangle}{1-\langle\sigma_1\rangle},\\
\langle X_i\rangle &= 1-\langle Y_i\rangle,\\
\langle X_iX_j\rangle &= \langle X_i\rangle - \langle X_iY_j\rangle,\\
\langle Y_iY_j\rangle &= \langle Y_i\rangle - \langle Y_iX_j\rangle,
\end{aligned}$$

where

$$\langle\widetilde{Y_jX}\rangle = \frac{\prod_{x\in\mathcal{N}_j}\langle Y_jX_x\rangle\prod_{y\neq j}\prod_{x\in\mathcal{N}_y:x<y,x\neq j}\langle X_yX_x\rangle}{\prod_{x\neq j}\langle X_x\rangle^{k_x-1}\langle Y_j\rangle^{k_j-1}}$$

and

$$\langle \sigma_1 \rangle = \prod_y \prod_{x \in \mathcal{N}_y : x<y} \frac{\langle X_y X_x \rangle}{\langle X_y \rangle^{n_y - 1}}.$$

To approximate the probability that node i is infected in the QSD, compute $\lim_{t\to\infty} \langle Y_i(t) \rangle / (1 - \langle \sigma_1(t) \rangle)$.

5.4.2 Population-level equations

The node-level pair-based model can quickly rise in computational cost if the degree of the network is large. We therefore derive the following population-level pair-based model for undirected networks, by following a similar approach to the individual-based derivation (see Appendix 5.D)

$$\begin{aligned}
[\dot{Y_k}] &= -\gamma [Y_k] + \tau \sum_{l \in \mathcal{M}} [X_k Y_l] + \frac{[Y_k] \sum_{l \in \mathcal{M}} \gamma \tilde{P}(Y_l = 1)}{1 - \langle \sigma_1 \rangle}, \\
[\dot{X_k Y_l}] &= \tau \Big(\sum_{m \in \mathcal{M}} \frac{l-1}{l} \frac{[X_k X_l][X_l Y_m]}{[X_l]} - \sum_{m \in \mathcal{M}} \frac{k-1}{k} \frac{[Y_m X_k][X_k Y_l]}{[X_k]}, \\
&\quad -[X_k Y_l] \Big) + \gamma([Y_k Y_l] - [X_k Y_l]) + \frac{[X_k Y_l] \sum_{l \in \mathcal{M}} \gamma \tilde{P}(Y_l = 1)}{1 - \langle \sigma_1 \rangle} \\
[\dot{Y_k Y_l}] &= \tau \Big(\sum_{m \in \mathcal{M}} \frac{k-1}{k} \frac{[Y_m X_k][X_k Y_l]}{[X_k]} + \sum_{m \in \mathcal{M}} \frac{l-1}{l} \frac{[Y_k X_l][X_l Y_m]}{[X_l]} \Big), \\
&\quad + \tau([X_k Y_l] + [Y_k X_l] - 2\gamma [Y_k Y_l] + \frac{[Y_k Y_l] \sum_{l \in \mathcal{M}} \gamma \tilde{P}(Y_l = 1)}{1 - \langle \sigma_1 \rangle} \\
[X_k] &= |C_k| - [Y_k], \\
[X_k X_l] &= |C_{k,l}| - [Y_k Y_l] - [X_k Y_l] - [X_l Y_k],
\end{aligned}$$

where

$$\tilde{P}(Y_k = 1) = |C_k| \frac{\prod_{k \neq k_j} \prod_{l \leq k : l \neq k_j} \left(\frac{[X_k X_l]}{|C_{k,l}|} \right)^{|C_{k,l}|} \left(\frac{[Y_{k_j} X_k]}{|C_{k_j,k}|} \right)^{\frac{|C_{k_j,k}|}{|C_{k_j}|}} \left(\frac{[X_{k_j} X_k]}{|C_{k_j,k}|} \right)^{|C_{k_j,k}| - \frac{|C_{k_j,k}|}{|C_{k_j}|}}}{\prod_{k \neq k_j} \left(\frac{[X_k]}{|C_k|} \right)^{|C_k|(k-1)} \left(\frac{[Y_{k_j}]}{|C_{k_j}|} \right)^{(k_j - 1)} \left(\frac{[X_{k_j}]}{|C_{k_j}|} \right)^{(|C_{k_j}| - 1)(k_j - 1)}}$$

and

$$\langle \sigma_1 \rangle \approx \prod_k \prod_{l \leq k} \left(\frac{\left(\frac{[X_k X_l]}{|C_{k,l}|} \right)^{|C_{k,l}|}}{\left(\frac{[X_k]}{|C_k|} \right)^{|C_k|(k-1)}} \right).$$

In these expressions, we use the convention that $0^0 = 1$. The expected number of infected degree k nodes in the QSD can be approximated by computing $\lim_{t\to\infty}[Y_k(t)]/(1-\langle\sigma_1(t)\rangle)$.

Special Case: k-regular networks

For a regular network, where all individuals have identical degree, the population-level model becomes simplified since $\mathcal{M} = \{k\}$, where k is the degree of the network. This model becomes

$$\begin{aligned}
[\dot{Y_k}] &= -\gamma[Y_k] + \tau[X_k Y_k] + \frac{[Y_k]\gamma\tilde{P}(I_k \to \sigma_0)}{1 - \langle\sigma_1\rangle}, \\
[\dot{X_k Y_k}] &= \tau\Big(\frac{k-1}{k}\frac{[X_k X_k][X_k Y_k]}{[X_k]} - \frac{k-1}{k}\frac{[Y_k X_k][X_k Y_k]}{[X_k]}, \\
&\quad -[X_k Y_k]\Big) + \gamma([Y_k Y_k] - [X_k Y_k]) + \frac{[X_k Y_k]\gamma\tilde{P}(I_k \to \sigma_0)}{1 - \langle\sigma_1\rangle}, \\
[\dot{Y_k Y_k}] &= \tau\Big(\frac{k-1}{k}\frac{[Y_k X_k][X_k Y_k]}{[X_k]} + \frac{k-1}{k}\frac{[Y_k X_k][X_k Y_k]}{[X_k]}\Big), \\
&\quad +\tau([X_k Y_k] + [Y_k X_k] - 2\gamma[Y_k Y_k] + \frac{[Y_k Y_k]\gamma P(I_k \to \sigma_0)}{1 - \langle\sigma_1\rangle}, \\
\langle\sigma_1\rangle &= \frac{\left(\frac{[X_k X_k]}{|C_{k,k}|}\right)^{|C_{k,k}|}}{\left(\frac{[X_k]}{|C_k|}\right)^{|C_k|(k-1)}}, \\
\tilde{P}(I_k = 1) &= |C_k| \frac{\left(\frac{[Y_k X_k]}{|C_{k,k}|}\right)^{\frac{|C_{k,k}|}{|C_k|}} \left(\frac{[X_k X_k]}{|C_{k,k}|}\right)^{|C_{k,k}| - \frac{|C_{k,k}|}{|C_k|}}}{\left(\frac{[Y_k]}{|C_k|}\right)^{(k-1)} \left(\frac{[X_k]}{|C_k|}\right)^{(|C_k|-1)(k-1)}},
\end{aligned}$$

which is efficient to solve numerically. To approximate the expected number of infected degree k individuals in the QSD, calculate $\lim_{t\to\infty}[Y_k(t)]/(1 - \langle\sigma_1(t)\rangle)$.

5.5 Numerical results

In this section, we investigate how the methods developed in this chapter perform when used to approximate the expected number of infected individuals in the QSD for various networks and parameter values.

We assume that: the transition rate for any pair of connected individuals is equal (taking $T_{ij} = \tau$ whenever $T_{ij} > 0$ and zero otherwise), the network is undirected, and infected individuals recover at the same rate; i.e. $\gamma_i = \gamma$ for all $i \in \mathcal{V}$. In the case of a complete network, the epidemic threshold of the population-level individual-based model occurs when $\tau \times k/\gamma = 1$, where k is the degree of the network (i.e. $k = N - 1$). We therefore plot the expected number of infected against $\tau \times \bar{d}/\gamma$, where $\bar{d}$ is the average degree of the network, to ensure that all networks are tested over a similar range of transmission strengths. We assume $\gamma = 1$ throughout, so that the ratio can be changed by changing τ.

The standard individual-based and pair-based models (Section 1.4.2) have a non-zero equilibrium solution in the region of parameter space where $\tau \times \bar{d}/\gamma$ is large [91, 184]. When the transmission rate is sufficiently large, the models proposed converge to the standard model, and therefore this endemic equilibrium approximates the expected number of infected individuals in the QSD. To demonstrate this, the dynamics for these standard models are compared to the QSD approximation methods. We are particularly interested in how our methods perform for low values of $\tau \times \bar{d}/\gamma$, at which the standard models will not capture the QSD.

To calculate the true QSD, we simulate the stochastic model and average over all simulations that have not gone extinct. Simulations are run until $t = 300$ in all results shown, since by this point all cases reached a steady-state. We compare this solution to the steady-state of the QSD approximation methods and the standard models.

5.5.1 Impact of network structure

To test the methods, consider three networks: the complete network, the square lattice, and Zachary's karate club [201]. The complete network represents a well-mixed population, in which all individuals are connected to each other. The square lattice represents an extreme structure whereby all individuals occupy the nodes of a square grid, and is a commonly used network when adding structure to population dynamics. Although there is a lot of symmetry across the network, the rigorous structure with multiple loops can prove challenging for moment closure approximation methods. Zachary's karate club is an

example of a real world network, formed from interactions between members of a karate club.

Figure 5.1(a) compares the node-level individual-based model with stochastic simulations. Below the epidemic threshold (where the standard model switches from zero to an endemic equilibrium), the QSD method captures the behaviour reasonably accurately. As $\tau \times \bar{d}/\gamma$ increases, the approximation diverges, with differing levels of performance on each of the networks tested. This individual-based method performs best on the complete network, on which it provides a good approximation to the expected number of infected for all parameter values. Some level of accuracy is also observed on Zachary's Karate club. However, on the square lattice this method does not perform well when above the epidemic threshold, significantly overestimating the expected number of infected individuals in the QSD. This is because the structure of the lattice results in significant local correlations on the network, and therefore assuming statistical independence of individual nodes is unrealistic.

Using the population-level individual-based model, little accuracy is lost (Figure 5.1(b)). The same pattern of performance occurs across the three networks, and by overlaying the results, the population-level model is almost indistinguishable from the node-level model on the resolution of the graph, showing that the QSD is mainly determined by the degree distribution.

Since the assumption of individual-level statistical independence is unrealistic, we developed a node-level pair-based model for the QSD. Figure 5.2(a) shows the accuracy of this approximation, which is significantly improved over the individual-based models on all networks. On the complete network and Zachary's karate club, this approximation is very accurate, and on the lattice it loses some accuracy but still significantly outperforms the individual-based approximation. The loss of accuracy on the lattice is expected, since pair-approximation methods are generally considered to perform weakly on such structures.

Although the pair-based model is computationally feasible, for large networks it can be slow. Therefore, we derived a population-level pair-based model. Again, little accuracy is lost for all networks (Figure 5.2(b)), with the result being indistinguishable from the node-level model.

For each of the methods proposed, a stationary solution is reached for all parameter values on all networks. These solutions appear to be unique and lower bounded by 1. Therefore, the proposed methods satisfy the basic properties of the QSD. Sufficiently above the epidemic threshold, our models and the standard models coincide (Figures 5.1 and 5.2),

showing that the standard models approximate the expected number of infected individuals in the QSD in this region. However, as the transmission rate decreases, the steady-states of the standard models deviate from this, eventually tending to the disease-free equilibrium. Therefore, the standard models are not a reliable measure of the QSD since they do not capture this for all parameter values, and the endemic equilibrium in the intermediate range (between the disease-free equilibrium and coinciding with the QSD model) is hard to relate to any properties of the underlying stochastic process. The models we propose are more robust for providing insight into the stochastic epidemic model.

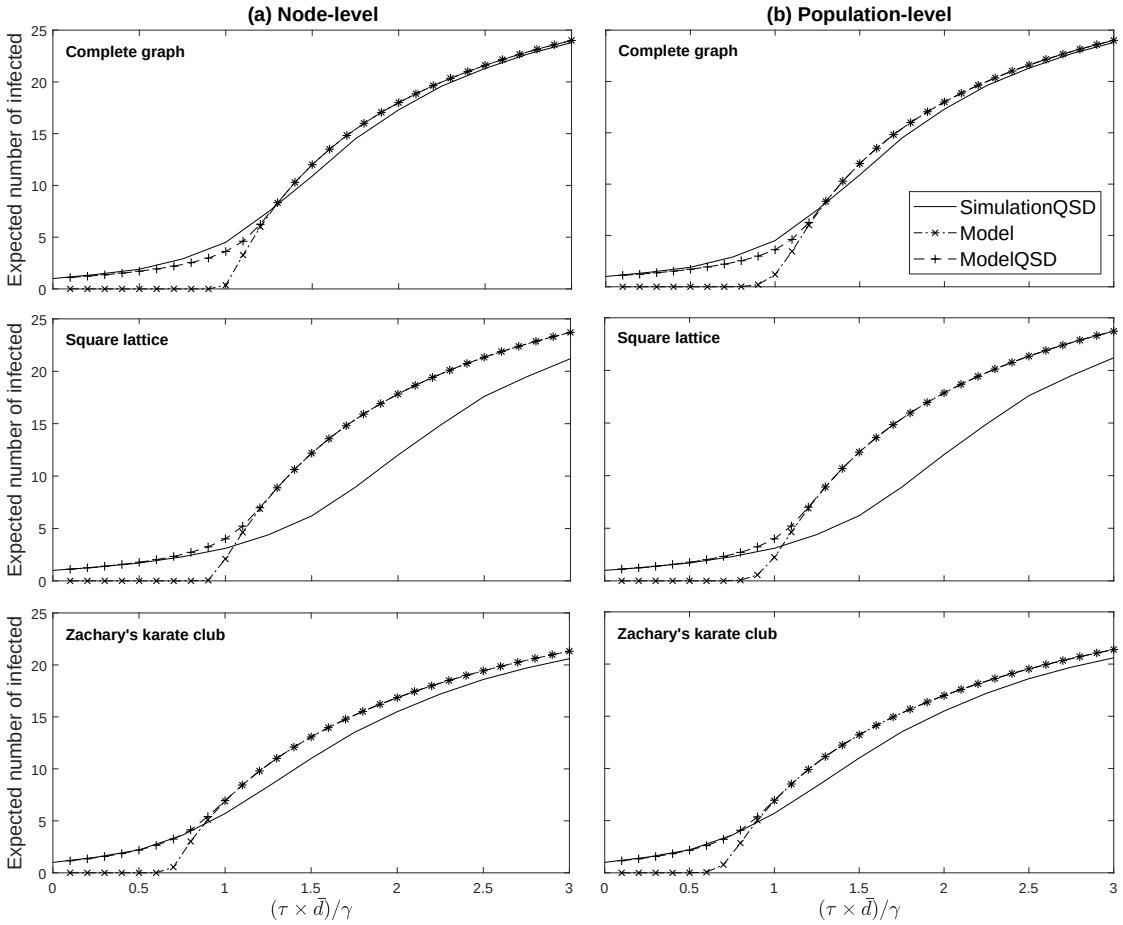

Figure 5.1: The expected number of infected individuals in the QSD as calculated by the individual-based model versus stochastic simulation on a 36 node complete network, 36 node square lattice and the 34 node karate club network, for a range of parameters. The right shows the population-level methods and the left shows the node-level methods. The solid lines represent the average of 10,000 stochastic simulations conditioned against extinction, the "+" marks the QSD method and the "×" marks the standard model.

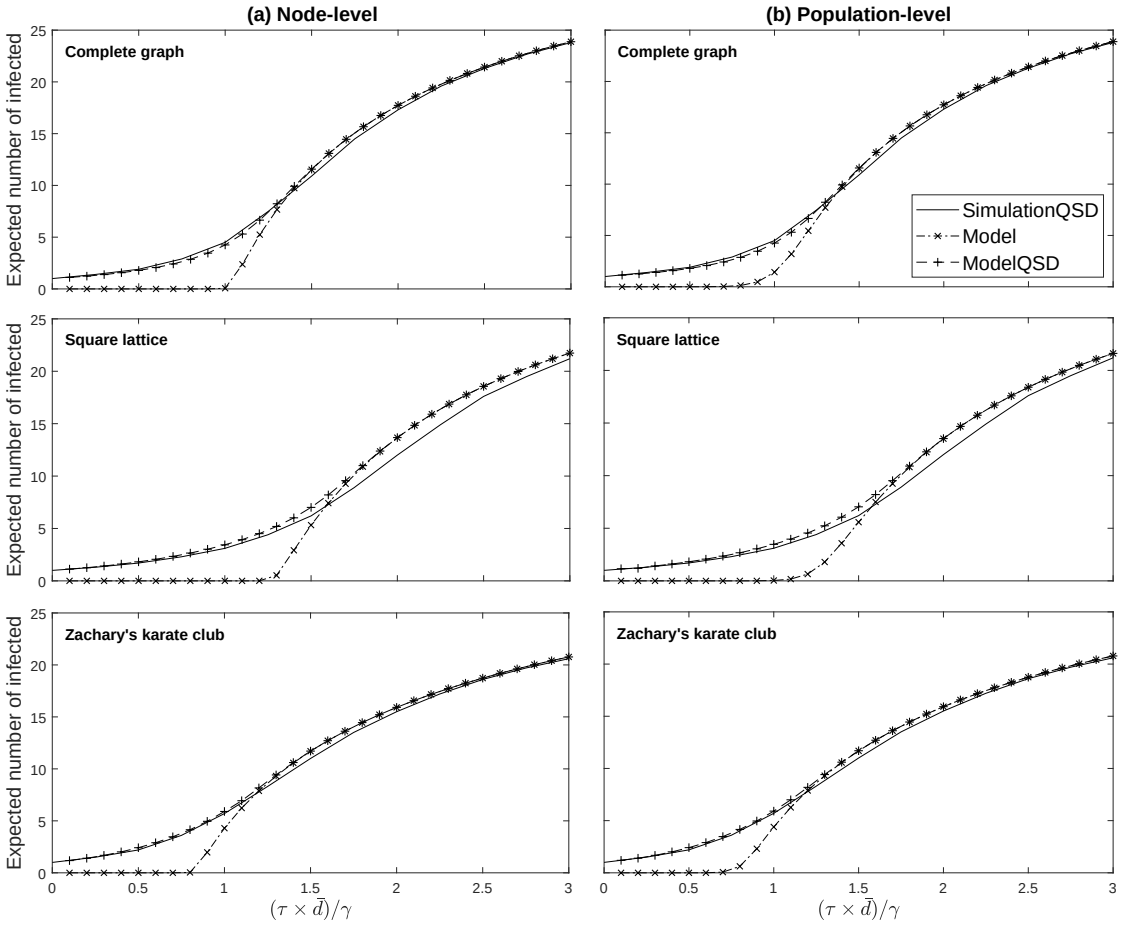

Figure 5.2: The expected number of infected individuals in the QSD as calculated by the pair-based model versus stochastic simulation on a 36 node complete network, 36 node square lattice and the 34 node karate club network, for a range of parameters. The right shows the population-level methods and the left shows the node-level methods. The solid lines represent the average of 10,000 stochastic simulations conditioned against extinction, the "+" marks the QSD method and the "×" marks the standard model.

5.5.2 Impact of network size

We now investigate how increasing the size of the population affects the accuracy of the results, testing a 100 node lattice and 225 node lattice. Here the square lattice is chosen because this presented itself as the worst case, with other networks expected to perform better.

Since the population-level models perform similarly to the node-level models at captur-

ing the expected number of infected, with significantly reduced computational cost, in this section we only use these models to approximate the dynamics. As the size of the lattice increases, the accuracy of the approximation does not significantly change (Figure 5.3), and therefore this method can accurately scale up to large networks.

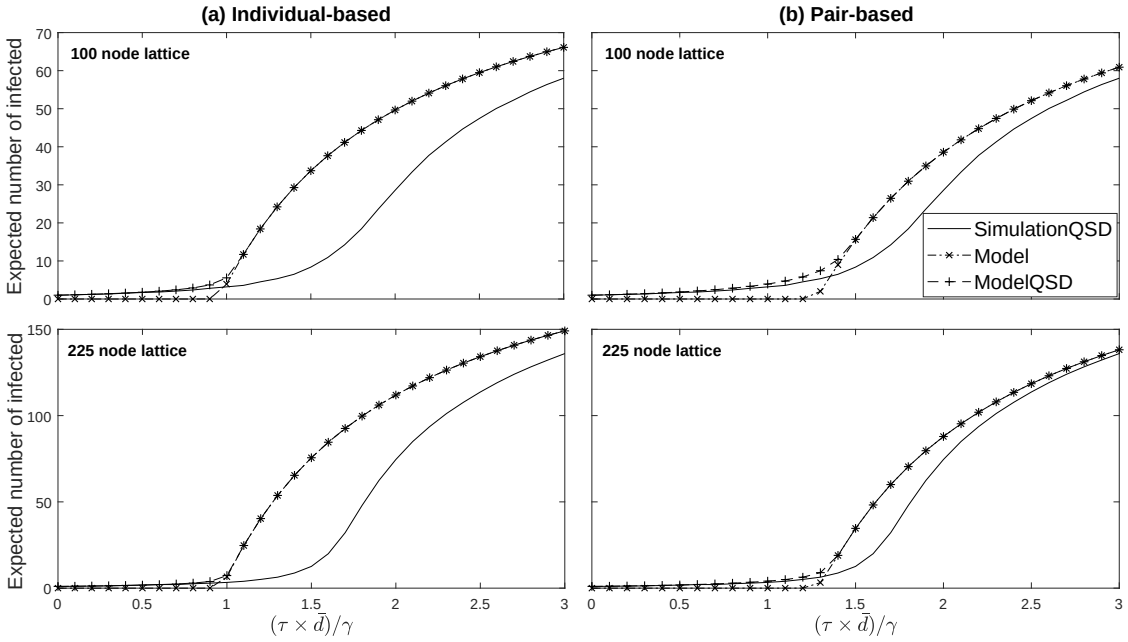

Figure 5.3: The expected number of infected individuals in the QSD as calculated by the population-level models versus stochastic simulation on a 100 node square lattice and 225 node square lattice for a range of parameter values. The plot shows the individual-based methods and the right shows the pair-based methods. The solid lines represent the average of 10,000 stochastic simulations conditioned against extinction, the "+" marks the QSD method and the "×" marks the standard model.

5.5.3 Marginal probabilities

Similar accuracy is observed between the node-level and population-level models, and so we can argue that the population-level models are superior due to the reduced computational cost. However, other statistics of the QSD can depend on node-level quantities, which are only captured by the node-level equations. One such property is the invasion probability.

The invasion probability from a given initial state in Markovian network-based SIS dynamics can be defined as quasi-invasion [193], which is the probability of that state being infected in the QSD. Such a statistic shows the strength of each node at spreading the

Table 5.1: The marginal probability of individual nodes being infected in the QSD. The table shows a selection of 5 nodes on Zachary's karate club, for a low transmission rate and high transmission rate.

High		Low	
Simulation	QSD pair-based model	Simulation	QSD pair-based model
0.5750	0.6142	0.1511	0.1640
0.2315	0.2513	0.0396	0.0421
0.4364	0.4599	0.1039	0.1099
0.3643	0.3874	0.0735	0.0771
0.3799	0.4002	0.0813	0.0868

infection through the network, and is a measure of the likelihood that a certain individual will cause the pathogen to spread. Table 5.1 shows that the node-level pair-based model accurately captures the marginal probabilities of each node being infected in the QSD, and therefore this method accurately approximates the invasion probability from a given initially infected single node. Population-level methods are not applicable here, illustrating when the node-level models are advantageous.

5.6 Discussion

Defining metrics to describe the spread of epidemics can allow different systems, whether this be different contact networks or different pathogens, to easily be compared. In the standard deterministic SIS model [96, 184, 185, 187], one such metric is the epidemic threshold, below which the pathogen will go extinct and above which the pathogen will reach a steady endemic equilibrium solution [96, 184]. More complicated deterministic models have been developed, such as pair-approximations models [50, 61, 83, 82, 91, 107, 167], in which this threshold behaviour can also be observed [82, 107]. However, no endemic solution exists in the stochastic model, making it hard to relate the insights gained from these models to the underlying stochastic process.

Many measures have been discussed and defined that provide insights into the stochastic dynamics [5, 9, 10, 44, 62, 122], such as the invasion probability [193]. Of particular interest when trying to find such information is the quasi-stationary distribution, since the disease-free state gives little insight into the impact the pathogen has had on the population. The quasi-stationary distribution is the steady-state of the system given that the absorbing

state has not been reached; i.e. the expected long term behaviour of the system when the pathogen does not die out. For network-based Markovian SIS dynamics, a unique QSD will always exist [181]. The QSD is therefore rich in information regarding the impact the pathogen may have on the population. This is especially relevant in the case when the transmission rate is high, because here the time to pathogen extinction is likely to be long, and apparently stable behaviour is observed before extinction.

The QSD can be calculated exactly from the master equation describing this Markov process; however, this system scales exponentially and quickly becomes unfeasible. The focus of this chapter was to derive computationally feasible methods to approximate the expected number of infected individuals in the QSD, providing a deterministic framework that can be linked directly to the underlying stochastic dynamics.

Where the population is homogeneous, the issue of reconciling the deterministic to stochastic models has been investigated by looking at the QSD [4]. Extending the SIS model to structured populations, the most similar work is that of [184], which investigates the endemic equilibrium solution to the individual-based approximation to the SIS model. The endemic steady-state solution has also been investigated in pair-based SIS approximations [91]. These provided a basis for the methods developed here, but we amended the systems so that the equilibrium solutions directly approximate the expected number of infected individuals in the QSD. The new methods are important, since although we have shown that the existing methods approximate this when the transmission rate is high, as this rate decreases an endemic equilibrium solution is attained that is not close to the QSD. It is in this intermediate state and the lower state, where the stable solution is the disease-free equilibrium, that our methods provide a strong advantage over the existing ones. In both of these cases, our techniques captured the dynamics much more accurately, as was shown in all figures.

We introduced various methods in this chapter for approximating the expected number of infected individuals in the QSD. Our first approach was an individual-based approximation, based on the assumption that the states of neighbouring nodes are independent. Although the assumption of individual-level independence is not particularly realistic, on the complete network this method gave a good approximation to the expected number of infected individuals in the QSD. However, for more structured networks the approximation became less accurate. From this node-level individual-based model, we developed a degree heterogeneous population-level model, again using the assumption of independence between individuals. Little accuracy was lost when computing the expected number of

infected individuals compared to the node-level model, and the computational cost was reduced.

Assuming independence of individuals does not accurately capture correlations between neighbours, due to infection spreading through contact. Therefore, we developed a node-level method based on assuming independence at the level of pairs. This model accurately captures the dynamics of the expected number of infected individuals in the QSD on all networks tested.

We then derived a degree heterogeneous population-level pair-approximation, in order to reduce the computational cost. Again, little accuracy is lost in the population-level model. This suggests that the expected number of infected individuals in the QSD is mainly determined by the degree distribution of the network. One advantage however of the node-level models is the insight these can give into the dynamics of individual nodes in the population. The population-level models lose this information since they group all individuals with the same degree into one class, whereas the node-level model accurately captures the marginal probability of each node being infected in the QSD.

These methods facilitate efficient and reliable analysis of how different network structures can alter the statistics of the QSD. We have shown how different networks can change the expected number of infected individuals in the QSD. The proposed methods also grant insight into other statistics. One example is the invasion probability [193], which can be calculated directly from the node-level models proposed. Being able to quickly calculate statistics such as these improves understanding of the differences in infection risks that various population structures can yield, which can have important treatment/prevention implications. The potential future applications of this work include extending the methods to other models for population dynamics.

Appendix 5.A Proof of existence of an endemic equilibrium for the standard pair-based model

Proof. In [96], a theorem is proven regarding the existence of stable endemic solutions for ordinary differential equation epidemic models. Here we demonstrate that the standard pair-based SIS model [107] (Equations 1.19 and 1.20 with the closure from Equation 1.21) satisfies the requirements for this proof, and therefore has a stable endemic equilibrium.

Consider an ODE of the form

$$\frac{dy}{dt} = Ay + N(y). \tag{5.14}$$

If the following statements hold, then there exists a threshold above which an endemic equilibrium exists.

1. A compact convex set C on the domain of N is positively invariant, with $y = 0 \in C$
2. $\lim_{y \to 0} ||N(y)||/||y|| = 0$
3. There exists $r > 0$ and a real eigenvector w or A^T such that $(w \cdot y) \geq r||y|| \ \ \forall y \in C$
4. $(w \cdot N(y)) \leq 0 \ \ \forall y \in C$
5. $y = 0$ is the largest positively invariant set contained in $H = \{y \in C | (w \cdot N(y)) = 0\}$

The first step is to write the pair-based model in the form (5.14). The pair-based model is given by

$$\begin{aligned}
\langle \dot{I}_i \rangle &= \tau \sum_{j}^{N} A_{ij} \langle S_i I_j \rangle - \gamma \langle I_i \rangle \\
\langle \dot{S_i I_j} \rangle &= \tau \sum_{k \neq i}^{N} A_{jk} \frac{\langle S_i S_j \rangle \langle S_j I_k \rangle}{\langle S_j \rangle} - \tau \sum_{k \neq j}^{N} A_{ik} \frac{\langle I_k S_i \rangle \langle S_i I_j \rangle}{\langle S_i \rangle} - \tau \langle S_i I_j \rangle - \gamma \langle S_i I_j \rangle + \gamma \langle I_i I_j \rangle,
\end{aligned}$$

where $\langle S_i \rangle = 1 - \langle I_i \rangle$, $\langle I_i I_j \rangle = \langle I_j \rangle - \langle S_i I_j \rangle$ and $\langle S_i S_j \rangle = \langle S_i \rangle - \langle S_i I_j \rangle$.

This can be rewritten as

$$\begin{aligned}
\langle \dot{I}_i \rangle &= \tau \sum_{j}^{N} A_{ij} \langle S_i I_j \rangle - \gamma \langle I_i \rangle \\
\langle \dot{S_i I_j} \rangle &= -(\tau + 2\gamma) \langle S_i I_j \rangle + \gamma \langle I_j \rangle + \tau \sum_{k \neq i}^{N} A_{jk} \langle S_j I_k \rangle \\
&\quad - \tau \sum_{k \neq i}^{N} A_{jk} \frac{\langle I_i S_j \rangle \langle S_j I_k \rangle}{\langle S_j \rangle} - \tau \sum_{k \neq j}^{N} A_{ik} \frac{\langle I_k S_i \rangle \langle S_i I_j \rangle}{\langle S_i \rangle}.
\end{aligned}$$

Defining $y_i = \langle I_i \rangle$ for $1 \leq i \leq N$ and $y_i = \langle S_1 I_{i-N} \rangle$ for $N + 1 \leq i \leq 2N$, $y_i = \langle S_2 I_{i-2N} \rangle$ for $2N + 1 \leq i \leq 3N$, and so on, we can write the pair-based model in the form of

Equation (5.14). Compiling the linear terms into the matrix A, we see that A is only negative on the diagonal. The remaining non-linear terms define the function $N(y)$, which only assigns negative values to each input. Now it is required to check if the properties hold.

Property (1.) holds because the system is invariant on the set $C = \{0 \leq \langle I_i \rangle \leq 1; 0 \leq \langle S_i I_j \rangle \leq 1\}$. Property (2.) holds because as $y \to 0$ the denominator of all terms, $1 - \langle I_i \rangle$, goes to one, and the numerator is of the form $y_i y_j$, which goes to zero faster than y_i and y_j. Property (3.) holds because A is irreducible since all the equations are coupled. Since A is only negative on the diagonal, by the Perron-Frobenius theorem, A^T must have an eigenvector w such that $w_i > 0$ for all i. Property (4.) holds because the function $N(y)$ is negative, so $(w \cdot N(y)) \leq 0$, since $w_i > 0$ for all i. We now need to test property (5.).
Property (5.) If $y \in H$ then $(w \cdot N(y)) = 0$. This implies that

$$w_i \tau \sum_{k \neq i} \frac{A_{jk} \langle I_i S_j \rangle \langle S_j I_k \rangle}{1 - \langle I_j \rangle} = 0$$

and

$$w_i \tau \sum_{k \neq j} \frac{A_{ik} \langle I_k S_i \rangle \langle S_i I_j \rangle}{1 - \langle I_i \rangle} = 0,$$

for all pairs (i, j). If we assume that $y \in H$ and $y \neq 0$, then $y_h \neq 0$ for some h. If we assume that $y_h = \langle S_i I_j \rangle \neq 0$, then we must have $\langle S_i I_k \rangle = 0$, for all $k \in \mathcal{N}_i$. Also, we require $\langle S_j I_k \rangle = 0$ for some k or $\langle I_i S_j \rangle = 0$. Therefore, $\langle S_i I_j \rangle = 0$ for some (i, j). We now need to investigate whether such a state can be invariant.

Define $S = \{i : y_i = 0\}$ and $S' = \{i : y_i \neq 0\}$, both of which are non-empty by the above argument. Since A is irreducible, there must exist a pair $k \in S$ and $h \in S'$ such that dy_k/dt depends on y_h.

First assume that $y_h = \langle S_i I_j \rangle$ and $y_k = \langle I_i \rangle$. We have

$$\frac{dy_k}{dt} = \tau \sum_{j \neq i} A_{ij} \langle S_i I_j \rangle$$

If this state is invariant, then $dy/dt = 0$, which implies that $dy_k/dt = 0$ for all k. This can only be the case if $\langle S_k I_j \rangle = 0$ for all j. However, we have assumed that $\langle S_i I_j \rangle \neq 0$, so this is not the case and $dy_k/dt \neq 0$.

Now assume $y_k = \langle S_j I_i \rangle$, which gives

$$\frac{dy_k}{dt} = \gamma\langle I_i\rangle + \tau\sum_{m\neq j}^{N} A_{im}\langle S_i I_m\rangle - \tau\sum_{m\neq j}^{N} A_{im}\frac{\langle I_j S_i\rangle\langle S_i I_m\rangle}{\langle S_i\rangle}.$$

Since $\langle I_j S_i\rangle/\langle S_i\rangle \leq 1$, the sum of the last two terms cannot be negative. Therefore we have $\langle I_i\rangle = 0$. However, as has been shown by assuming $\langle I_i\rangle = 0$, this case is not possible. Therefore, $dy_k/dt \neq 0$. Therefore, if $\langle S_i I_j\rangle \neq 0$ for some pair (i, j) and $y \in H$, then this state cannot be invariant.

Now assume that $y_h = \langle I_i\rangle \in S'$ for some i, and consider $y_k = \langle S_j I_i\rangle \in S$. Since $\langle S_x I_y\rangle = 0$ for all (x, y), we have

$$\frac{dy_k}{dt} = \gamma\langle I_i\rangle.$$

Since $\langle I_i\rangle \in S'$, $dy_k/dt \neq 0$. Therefore, there are no invariant sets in H such that $y \neq 0$, and $y = 0$ is the largest positively invariant set in H. □

Appendix 5.B Node-level individual-based QSD model

5.B.1 Proof that the individual-based node-level QSD model is invariant on $[0, 1]^N$.

Proof. To prove that the model is invariant we use the method from [96]. Along the boundaries to the set we are interested in, we either have $\langle Y_i\rangle = 0$ and $\langle X_i\rangle = 1$ or $\langle Y_i\rangle = 1$ and $\langle X_i\rangle = 0$. To show the system is invariant, we need to show that along these boundaries the trajectories do not point away from this set.

First consider $\langle Y_i\rangle = 0$. At this boundary, we have

$$\langle\dot{Y}_i\rangle = \sum_j T_{ij}\langle Y_j\rangle.$$

If $\langle Y_i\rangle \in [0, 1]$, this cannot be negative, and therefore at $\langle Y_i\rangle = 0$ the trajectory in the i direction cannot leave the set $[0, 1]^N$. Now consider $\langle Y_i\rangle = 1$. We have

$$\langle\dot{Y}_i\rangle = -\gamma_i + \gamma_i\prod_{k\neq i}\langle X_k\rangle.$$

The product in this equation is in $[0,1]$ if $\langle X_k \rangle \in [0,1]$ for all k. Therefore, this equation can never be positive, so along this boundary the trajectory cannot leave the set $[0,1]$. Therefore, this model is invariant on $[0,1]^N$. □

5.B.2 Proof of a non-zero solution for the complete network

Proof. On the complete network, if we start with a fully infected population the probability that each node is infected at a given time will be equal for all nodes, and therefore we can denote $\langle I_i \rangle = a$ for all $i \in \mathcal{V}$. Therefore, we can write the rate of change in the node probabilities as

$$\dot{a} = -\gamma a + \tau(N-1)(1-a) + a\frac{\gamma N a(1-a)^{N-1}}{1-(1-a)^N}.$$

In the steady-state $\dot{a} = 0$. Clearly $a = 0$ is a solution to this equation. If we rule out $a = 0$ then we obtain

$$(1-a)(N-1)\left(\frac{\tau}{\gamma} + \frac{N}{N-1}\frac{a(1-a)^{N-2}}{1-(1-a)^N}\right) = 1.$$

To find solutions we are interested in solutions to $f(a) = 0$ for $a \in [0,1]$, where

$$f(a) = (1-a)(N-1)\left(\frac{\tau}{\gamma} + \frac{N}{N-1}\frac{a(1-a)^{N-2}}{1-(1-a)^N}\right) - 1.$$

To see if a solution exists within this interval we check the signs at the end points.

At $a = 1$

$$f(1) = -1 < 0$$

the function is negative.

At $a = 0$

$$f(0) = (N-1)\frac{\tau}{\gamma} - 1 + \lim_{a\to 0} N\frac{a(1-a)^{N-1}}{1-(1-a)^N},$$

$$\lim_{a\to 0} N\frac{a(1-a)^{N-1}}{1-(1-a)^N} = \lim_{a\to 0} N\frac{(1-a)^{N-1} + (N-1)(1-a)^{N-2}}{N(1-a)^{N-1}} = 1$$

$$\implies f(0) = (N-1)\frac{\tau}{\gamma} > 0 \text{ if } \frac{\tau}{\gamma} > 0.$$

Therefore as long as the transmission rate τ is greater than zero there exists a solution to $f(a) = 0$ in the open interval $(0,1)$, since $f(a)$ is non-singular on $(0,1)$. □

5.B.3 Proof that the node-level individual-based model is lower bounded by 1

Proof. Consider the node-level individual-based model; i.e.

$$\langle \dot{Y}_i \rangle = -\gamma_i \langle Y_i \rangle + \sum_j T_{ij} \langle X_i \rangle \langle Y_j \rangle + \frac{\langle Y_i \rangle}{1 - \prod_k \langle X_k \rangle} \sum_j \gamma_j \langle Y_j \rangle \prod_{k \neq j} \langle X_k \rangle. \tag{5.15}$$

To approximate the QSD we calculate $\langle Y_i^* \rangle / (1 - \prod_k \langle X_k^* \rangle)$, where $\langle Y^* \rangle$ and $\langle X^* \rangle$ are steady-state solutions to (5.15). To understand the behaviour of this quantity, take its derivative with respect to time,

$$\begin{aligned}
&\frac{d}{dt}\left(\frac{\langle Y_i \rangle}{1 - \prod_j \langle X_j \rangle} \right) = \\
&- \gamma_i \langle Y_i^c \rangle + \sum_j T_{ij} \langle X_i \rangle \langle Y_j^c \rangle + \langle Y_i^c \rangle \sum_j \gamma_j \langle Y_j^c \rangle \prod_{k \neq j} \langle X_k \rangle \\
&+ \langle Y_i^c \rangle \sum_j \left(\prod_{k \neq j} \langle X_k \rangle \right) \left[\gamma_j \langle Y_j^c \rangle - \sum_k T_{jk} \langle X_j \rangle \langle Y_k^c \rangle - \langle Y_j^c \rangle \sum_k \gamma_k \langle Y_k^c \rangle \prod_{m \neq k} \langle X_m \rangle \right],
\end{aligned} \tag{5.16}$$

where $\langle Y_i^c \rangle = \langle Y_i \rangle / (1 - \prod_k \langle X_k \rangle)$. Since the marginal probability that node i is infected in the QSD is given by $\langle Y_i \rangle / (1 - \prod_j \langle X_j \rangle)$ in the equilibrium, the expected number of infected individuals in the QSD is approximated by summing this over i in the equilibrium. If we assume Equation (5.16) is in equilibrium for all i, which is the case if $\langle X \rangle$ and $\langle Y \rangle$ are in equilibrium, summing over i gives

$$\begin{aligned}
&\sum_i \gamma_i \langle Y_i^c \rangle + \sum_i \langle Y_i^c \rangle \sum_j \left(\prod_{k \neq j} \langle X_k \rangle \right) \left[\sum_k T_{jk} \langle X_j \rangle \langle Y_k^c \rangle + \langle Y_j^c \rangle \sum_k \gamma_k \langle Y_k^c \rangle \prod_{m \neq k} \langle X_m \rangle \right] \\
&\quad = \sum_i \sum_j T_{ij} \langle X_i \rangle \langle Y_j^c \rangle + \sum_i \langle Y_i^c \rangle \sum_j \gamma_j \langle Y_j^c \rangle \prod_{k \neq j} \langle X_k \rangle + \langle Y_i^c \rangle \sum_j \left(\prod_{k \neq j} \langle X_k \rangle \right) \gamma_j \langle Y_j^c \rangle.
\end{aligned} \tag{5.17}$$

For the complete network, the steady-state $\langle Y_i \rangle$ only goes to zero as the transmission rate goes to zero, which is also true for all networks tested. As $\langle Y_i \rangle$ goes to zero, $\langle X_i \rangle$ goes to

one. Therefore, to test the lower bound on the expected number of infected predicted by this model, set $\langle Y_i \rangle = 0$ and $\langle X_i \rangle = 1$. This gives

$$\sum_i \gamma_i \langle Y_i^c \rangle + \sum_i \langle Y_i^c \rangle \sum_j \left[\sum_k T_{jk} \langle Y_k^c \rangle + \langle Y_j^c \rangle \sum_k \gamma_k \langle Y_k^c \rangle \right] = \sum_i \sum_j T_{ij} \langle Y_j^c \rangle + \sum_i \langle Y_i^c \rangle \sum_j \gamma_j \langle Y_j^c \rangle + \langle Y_i^c \rangle \sum_j \gamma_j \langle Y_j^c \rangle. \tag{5.18}$$

Using the notation $A = \sum_i \gamma_i \langle Y_i^c \rangle$, $B = \sum_j \sum_k T_{jk} \langle Y_k^c \rangle$, and $x = \sum_i \langle Y_i^c \rangle$, we obtain

$$A + Bx + Ax^2 = B + 2Ax,$$

which, for $A \neq 0$, has solutions

$$x = \frac{2A - B \pm B}{2A}.$$

The two solutions are $x = 1$ and $x = (A - B)/A$. If we assume that $\langle Y_i \rangle$ is only equal to zero fo all i if T_{ij} goes to zero for all pairs (i, j), as we have proven for the complete network, then $B = 0$, so we end up with a single solution $x = 1$. If $A = 0$, then $x = 1$ or $B = 0$. If $A = 0$ and $B = 0$, then any value of x is permitted. However, numerically we observe that $A \neq 0$, so we do not consider this case.

Therefore, as the solution to the node-level individual-based model approaches zero for all individuals, the approximation to the expected number of infected individuals in the QSD is lower bounded by 1, which is consistent with the true QSD. □

Appendix 5.C Node-level pair-based model derivation

If we do not assume independence at the level of individuals, we need to find equations describing pair probabilities in the conditional distribution. We have

$$\begin{aligned}
\frac{d}{dt}\left(\rho(\Sigma_i(t)=I)\right) = \sum_{\alpha:\sigma_{\alpha i}=I}\frac{d\rho_\alpha}{dt} &= \frac{\sum_j T_{ij}\langle S_iI_j\rangle - \gamma_i\langle I_i\rangle}{1-P_1} + \frac{\langle I_i\rangle}{(1-P_1)^2}\sum_j \gamma_j\langle I_jS\rangle,\\
\frac{d}{dt}\left(\rho(\Sigma_i(t)=S,\Sigma_j=I)\right) = \sum_{\substack{\alpha:\sigma_{\alpha i}=S,\\ \sigma_{\alpha j}=I}}\frac{d\rho_\alpha}{dt} &= \frac{\sum_{k\in\mathcal{N}_j\setminus i} T_{jk}\langle S_iS_jI_k\rangle}{1-P_1} - \frac{\sum_{k\in\mathcal{N}_i\setminus j} T_{ik}\langle I_kS_iI_j\rangle}{1-P_1}\\
&\quad -\frac{(T_{ij}+\gamma_j)\langle S_iI_j\rangle}{1-P_1} + \frac{\gamma_i\langle I_iI_j\rangle}{1-P_1} + \frac{\langle S_iI_j\rangle}{(1-P_1)^2}\sum_j\gamma_j\langle I_jS\rangle,
\end{aligned} \tag{5.19}$$

where $\langle A_i\rangle$ is shorthand for the marginal probability $P(\Sigma_i(t)=A)$ with $A\in\{S,I\}$, $\langle A_iB_j\rangle$ is shorthand for $P(\Sigma_i(t)=A,\Sigma_j(t)=B)$ with $A,B\in\{S,I\}$, $\langle A_iB_jC_k\rangle$ is shorthand for $P(\Sigma_i(t)=A,\Sigma_j(t)=B,\Sigma_k(t)=C)$ with $A,B,C\in\{S,I\}$, and $\langle I_jS\rangle$ is shorthand for $P(\Sigma_j=I,\Sigma_k=S \text{ for all } k\neq j)$. We can simplify this system by assuming statistical independence at the level of pairs.

As we have described in Section 1.4.2, we approximate the triples in terms of pairs and individuals by assuming

$$\langle A_iB_jC_k\rangle \approx \frac{\langle A_iB_j\rangle\langle B_jC_k\rangle}{\langle B_j\rangle}.$$

Under this assumption, Equation (5.19) becomes

$$\begin{aligned}
\frac{d}{dt}\left(\rho(\Sigma_i(t)=I)\right) &= \frac{\sum_j T_{ij}\langle S_iI_j\rangle - \gamma_i\langle I_i\rangle}{1-P_1} + \frac{\langle I_i\rangle}{(1-P_1)^2}\sum_j \gamma_j\langle I_jS\rangle,\\
\frac{d}{dt}\left(\rho(\Sigma_i(t)=S,\Sigma_j=I)\right) &= \frac{\sum_{k\in\mathcal{N}_j\setminus i} T_{jk}\frac{\langle S_iS_j\rangle\langle S_jI_k\rangle}{\langle S_j\rangle}}{1-P_1} - \frac{\sum_{k\in\mathcal{N}_i\setminus j} T_{ik}\frac{\langle I_kS_i\rangle\langle S_iI_j\rangle}{\langle S_i\rangle}}{1-P_1}\\
&\quad -\frac{(T_{ij}+\gamma_j)\langle S_iI_j\rangle}{1-P_1} + \frac{\gamma_i\langle I_iI_j\rangle}{1-P_1} + \frac{\langle S_iI_j\rangle}{(1-P_1)^2}\sum_j\gamma_j\langle I_jS\rangle.
\end{aligned}$$

Note that $\langle S_i\rangle = 1-\langle I_i\rangle$, $\langle I_iI_j\rangle = \langle I_j\rangle - \langle S_iI_j\rangle$ and $\langle S_iS_j\rangle = \langle S_i\rangle - \langle S_iI_j\rangle$. Both $\langle I_jS\rangle$ and the ground state probability, P_1, are of full system size, and therefore, following [50],

a natural pair approximation for these is

$$\langle I_j S\rangle \approx \langle \widetilde{I_j S}\rangle = \frac{\prod_{x\in\mathcal{N}_j}\langle I_j S_x\rangle \prod_{y\neq j}\prod_{x\in\mathcal{N}_y:x<y,x\neq j}\langle S_y S_x\rangle}{\prod_{x\neq j}\langle S_x\rangle^{k_x-1}\langle Y_j\rangle^{k_j-1}}$$

and

$$P_1 \approx \langle\sigma_1\rangle = \prod_y \prod_{x\in\mathcal{N}_y:x<y}\frac{\langle S_y S_x\rangle}{\langle S_y\rangle^{n_y-1}},$$

where $\mathcal{N}_j$ is the set of individuals that can infect or be infected by j; i.e. $i\in\mathcal{N}_j$ if $T_{ij}>0$ or $T_{ji}>0$.

In the QSD, both the pair-level and individual-level conditional probabilities are in a steady-state, so both equations in Equation (5.19) are equal to zero. Therefore, to find the approximation to the QSD under the pair-level independence assumption, we need to find vectors $\langle X^*\rangle$, $\langle Y^*\rangle$, and matrices $\langle XX^*\rangle$,$\langle XY^*\rangle$, and $\langle YY^*\rangle$ satisfying,

$$\begin{aligned}
0 &= \frac{\sum_j T_{ij}\langle X_i Y_j\rangle^* - \gamma_i\langle Y_i\rangle^*}{1-\langle\sigma_1\rangle} + \frac{\langle Y_i\rangle^*}{(1-\langle\sigma_1\rangle)^2}\sum_j \gamma_j\langle\widetilde{Y_j X}\rangle^*,\\
0 &= \frac{\sum_{k\in\mathcal{N}_j\setminus i} T_{jk}\frac{\langle X_i X_j\rangle^*\langle X_j Y_k\rangle^*}{\langle X_j\rangle^*}}{1-\langle\sigma_1\rangle} - \frac{\sum_{k\in\mathcal{N}_i\setminus j} T_{ik}\frac{\langle Y_k X_i\rangle^*\langle X_i Y_j\rangle^*}{\langle X_i\rangle^*}}{1-\langle\sigma_1\rangle}\\
&\quad -\frac{(T_{ij}+\gamma_j)\langle X_i Y_j\rangle^*}{1-\langle\sigma_1\rangle} + \frac{\gamma_i\langle Y_i Y_j\rangle^*}{1-\langle\sigma_1\rangle} + \frac{\langle X_i Y_j\rangle^*}{(1-\langle\sigma_1\rangle)^2}\sum_j\gamma_j\langle\widetilde{Y_j X}\rangle^*,
\end{aligned}$$

which, once solved, can be used to find the probability that i is infected in the QSD by computing $\langle Y_i\rangle^*/(1-\langle\sigma_1\rangle^*)$. However, we require solutions $\langle Y_i\rangle^*$ and $\langle X_i Y_j\rangle^*\in[0,1]$ which satisfy $\langle X_i\rangle^* = 1-\langle Y_i\rangle^*$ for all i and $\langle X_i X_j\rangle = \langle X_i\rangle - \langle X_i Y_j\rangle$, and $\langle Y_i Y_j\rangle = \langle Y_j\rangle - \langle X_i Y_j\rangle$ for all i,j in order to be valid solutions to our original problem.

By calculating the equilibrium of the system,

$$\begin{aligned}
\langle\dot{Y_i}\rangle &= \sum_j T_{ij}\langle X_i Y_j\rangle - \gamma_i\langle Y_i\rangle + \frac{\langle Y_i\rangle\sum_j\gamma_j\langle\widetilde{Y_j X}\rangle}{1-\langle\sigma_1\rangle},\\
\langle\dot{X_i Y_j}\rangle &= \sum_{k\in\mathcal{N}_j\setminus i} T_{jk}\frac{\langle X_i X_j\rangle\langle X_j Y_k\rangle}{\langle X_j\rangle} - \sum_{k\in\mathcal{N}_i\setminus j} T_{ik}\frac{\langle Y_k X_i\rangle\langle X_i Y_j\rangle}{\langle X_i\rangle}
\end{aligned}$$

$$
\begin{aligned}
& -(T_{ij}+\gamma_j)\langle X_iY_j\rangle + \gamma_i\langle Y_iY_j\rangle + \frac{\langle X_iY_j\rangle \sum_j \gamma_j \langle \widetilde{Y_jX}\rangle}{1-\langle\sigma_1\rangle}, \\
\langle X_i\rangle &= 1-\langle Y_i\rangle, \\
\langle X_iX_j\rangle &= \langle X_i\rangle - \langle X_iY_j\rangle, \\
\langle Y_iY_j\rangle &= \langle Y_i\rangle - \langle Y_iX_j\rangle,
\end{aligned}
$$

where

$$
\langle \widetilde{Y_jX}\rangle = \frac{\prod_{x\in\mathcal{N}_j}\langle Y_jX_x\rangle \prod_{y\neq j}\prod_{x\in\mathcal{N}_y:x<y,x\neq j}\langle X_yX_x\rangle}{\prod_{x\neq j}\langle X_x\rangle^{k_x-1}\langle Y_j\rangle^{k_j-1}}
$$

and

$$
\langle\sigma_1\rangle = \prod_y \prod_{x\in\mathcal{N}_y:x<y} \frac{\langle X_yX_x\rangle}{\langle X_y\rangle^{n_y-1}}.
$$

we can approximate the probability that i is infected in the QSD by computing $\lim_{t\to\infty}\langle Y_i(t)\rangle^*/(1-\langle\sigma_0(t)\rangle^*)$.

Appendix 5.D Population-level pair-based mode derivation

To obtain a population-level pair-based model, we sum over nodes with the same degree (and pairs of nodes with the same pair of degrees); i.e.

$$
\begin{aligned}
\sum_{i:k_i=k}\sum_{\alpha:\sigma_{\alpha i}=I}\frac{d\rho_\alpha}{dt} &= \frac{\tau\sum_{l\in\mathcal{M}}[S_kI_l]-\gamma[I_k]}{1-P_1} + \frac{[I_k]}{(1-P_1)^2}\sum_j\gamma\langle I_jS\rangle, \\
\sum_{\substack{i,j:k_i=k,\\ k_j=l}}\sum_{\substack{\alpha:\sigma_{\alpha i}=S,\\ \sigma_{\alpha j}=I}}\frac{d\rho_\alpha}{dt} &= \frac{\tau\sum_{m\in\mathcal{M}}[S_kS_lI_m]-\tau\sum_{m\in\mathcal{M}}[I_mS_kI_l]-\tau[S_kI_l]+\gamma[I_kI_l]-\gamma[S_kI_l]}{1-P_1} \\
&\quad + \frac{[S_kI_l]}{(1-P_1)^2}\sum_j\gamma\langle I_jS\rangle,
\end{aligned}
$$

where $[A_kB_lC_h]$ is the expected number of triples between degree k, degree l and degree h individuals in states A, B and C respectively.

As described in Section 1.4.2, we can express the triple terms as

$$[A_l S_k I_m] \approx \frac{k-1}{k} \frac{[A_l S_k][S_k I_m]}{[S_k]}. \tag{5.20}$$

We can set equations (5.20) to zero and use the approximation (5.20) to find equations describing the QSD.

A solution to the resulting system can be found by finding an equilibrium of

$$\begin{aligned}
[\dot{Y_k}] &= -\gamma[Y_k] + \tau \sum_{l \in \mathcal{M}} [X_k Y_l] + \frac{[Y_k] \sum_{l \in \mathcal{M}} \gamma \tilde{P}(Y_l = 1)}{1 - \langle \sigma_1 \rangle} \\
[\dot{X_k Y_l}] &= \tau\Big(\sum_{m \in \mathcal{M}} \frac{l-1}{l} \frac{[X_k X_l][X_l Y_m]}{[X_l]} - \sum_{m \in \mathcal{M}} \frac{k-1}{k} \frac{[Y_m X_k][X_k Y_l]}{[X_k]} \\
&\quad -[X_k Y_l]\Big) + \gamma([Y_k Y_l] - [X_k Y_l]) + \frac{[X_k Y_l] \sum_{l \in \mathcal{M}} \gamma \tilde{P}(Y_l = 1)}{1 - \langle \sigma_1 \rangle} \\
[\dot{Y_k Y_l}] &= \tau\Big(\sum_{m \in \mathcal{M}} \frac{k-1}{k} \frac{[Y_m X_k][X_k Y_l]}{[X_k]} + \sum_{m \in \mathcal{M}} \frac{l-1}{l} \frac{[Y_k X_l][X_l Y_m]}{[X_l]} \Big) \\
&\quad +\tau([X_k Y_l] + [Y_k X_l] - 2\gamma[Y_k Y_l] + \frac{[Y_k Y_l] \sum_{l \in \mathcal{M}} \gamma \tilde{P}(Y_l = 1)}{1 - \langle \sigma_1 \rangle} \\
[X_k] &= |C_k| - [Y_k] \\
[X_k X_l] &= |C_{k,l}| - [Y_k Y_l] - [X_k Y_l] - [X_l Y_k],
\end{aligned} \tag{5.21}$$

where $\tilde{P}(Y_l = 1) = |C_l| \langle Y_i X \rangle$ for some i with $k_i = l$. Here

$$\langle Y_i X \rangle = \frac{\prod_{x,y \neq i} G_{xy} \langle X_x X_y \rangle \prod_x G_{ix} \langle Y_i X_x \rangle}{\prod_{x \neq i} \langle X_x \rangle^{k_x - 1} \langle Y_i \rangle^{k_i - 1}},$$

which requires node-level terms. We can approximate this by population-level quantities using

$$\langle S_i \rangle \approx \frac{[S_{k_i}]}{|C_{k_i}|}, \tag{5.22}$$

and

$$\langle S_i S_j \rangle \approx \frac{[S_{k_i} S_{k_j}]}{|C_{k_i,k_j}|}, \tag{5.23}$$

based on the discussion in Section 1.4.2. This gives

$$\langle Y_j X \rangle \approx \frac{\prod\limits_{k \neq k_j} \prod\limits_{l \leq k : l \neq k_j} \left(\frac{[X_k X_l]}{|C_{k,l}|} \right)^{|C_{k,l}|} \left(\frac{[Y_{k_j} X_k]}{|C_{k_j,k}|} \right)^{\frac{|C_{k_j,k}|}{|C_{k_j}|}} \left(\frac{[X_{k_j} X_k]}{|C_{k_j,k}|} \right)^{|C_{k_j,k}| - \frac{|C_{k_j,k}|}{|C_{k_j}|}}}{\prod\limits_{k \neq k_j} \left(\frac{[X_k]}{|C_k|} \right)^{|C_k|(k-1)} \left(\frac{[Y_{k_j}]}{|C_{k_j}|} \right)^{(k_j - 1)} \left(\frac{[X_{k_j}]}{|C_{k_j}|} \right)^{(|C_{k_j}|-1)(k_j-1)}}. \tag{5.24}$$

To approximate the ground state recall that in the previous section we have shown that a natural approximation to the ground state probability under the assumption of pair-level independence is

$$\langle \sigma_1 \rangle \approx \prod_i \prod_{j<i} \frac{G_{ij} \langle X_i X_j \rangle}{\langle X_i \rangle^{n_i - 1}}.$$

Using Equations (5.22) and (5.23) we can approximate this in terms of population level quantities, which yields

$$\langle \sigma_1 \rangle \approx \prod_k \prod_{l \leq k} \left(\frac{\left(\frac{[X_k X_l]}{|C_{k,l}|} \right)^{|C_{k,l}|}}{\left(\frac{[X_k]}{|C_k|} \right)^{|C_k|(k-1)}} \right). \tag{5.25}$$

By substituting equations (5.25) and (5.24) into Equation (5.21) we obtain a closed system of equations.

Chapter 6

Conclusion

Evolution is the process through which species change and adapt [34, 51, 98, 104, 110, 176, 182]. Traditional evolutionary models focussed on well-mixed populations, where all individuals interact with each other. However, real populations are often structured, since individuals only interact with a subset of the population. This structure can be represented by a network (or graph), which illustrates how the individuals in the population interact and compete with each other. Such population structure has been found to have significant outcomes on the evolutionary process [102]. Modelling evolution on a network in such a way, which is known as evolutionary graph theory, has been widely studied theoretically [6, 70, 106, 140]. For example, results have shown that population structure changes the time to fixation [68, 69] and the probability of mutants fixating [60], and affects the dynamics of evolutionary games [129]. However, despite the widespread theoretical interest, the biological applications of this framework have been limited, potentially due to unrealistic assumptions underpinning the model, such as birth and death being coupled.

In this thesis, we have been motivated by adding further biological realism to this process. In Chapter 2, we built upon the framework of Champagnat [22] to develop a biologically motivated framework for evolution on networks. Through this, we have revealed the underlying assumptions behind existing structured models, such as evolutionary graph theory. By applying different limits to the dynamics, we have shown in what circumstances these assumptions are realistic. Through constructing this model, we also derive epidemic models, such as SIS. This provides further understanding into pathogen evolution, which is important for tackling antimicrobial resistance, HIV and other healthcare problems characterised by evolving pathogens.

Aside from the improved biological realism, the proposed framework also enables the analysis of clonal interference in network-structured populations. Clonal interference is where new mutations hinder the chance of existing mutants fixating in a population [53]. In the evolutionary graph theory framework, mutants have an infinite time to fixate since the mutation rate is assumed to be negligible. However, network structure can significantly amplify the time to fixation relative to a well-mixed population [177]. Therefore, this assumption of no mutations before fixation is potentially unrealistic. By adding mutation, we have shown that if the mutation rate is sufficiently high, structures that would amplify selection under evolutionary graph theory can hinder selection due to this clonal interference.

Using this novel framework, we have investigated certain properties of interest, mainly focusing on the insights this can give into evolutionary graph theory. However, the framework provides a whole range of potential research areas that we have yet to explore. A future direction which will be particularly interesting is investigating how this framework can be applied to real populations, and what insights it can give into how these evolve under different selective pressures.

Another aspect present in biological populations that is often omitted from evolutionary models is the presence of environmental variation. This variation results in the conditions experienced by individuals changing across time. In well-mixed populations, this has led to the investigation of evolutionary bet-hedging [14, 100, 160, 162, 173, 174, 186], where evolution selects for species that have reduced mean fitness but are sufficiently less sensitive to these variations. In Chapter 3, we investigated evolutionary bet-hedging in network-structured populations.

There are two types of variation considered; between-generation and within-generation. Between-generational variation acts on all individuals of a given type identically at a given time, and has been found to be important to the evolutionary process. Within-generational variation acts on individuals of a given species independently at any given time. In well-mixed populations, such variation has been found to not play a part on the evolutionary process. Using the framework of evolutionary graph theory, we have shown that provided the population has sufficient structure governing how individuals compete, within-generational variation can play a large part on evolution. This is supported by work using metapopulations [97, 170, 171, 199], and provides theoretical insight into many empirical observations that have identified species that evolved to counter such variation. A limitation of this current work is that we have only investigated network-structure using the

evolutionary graph theory framework. Potential future extensions may involve investigating how bet-hedging strategies would evolve under the novel dynamics we proposed in Chapter 2.

The improved realism obtained through adding ecological dynamics and environmental variation comes at the cost of increased computational effort. The analyses performed therefore relies mostly on stochastic simulations, though we have obtained some analytical results where possible. These analytical results provide insight into the scenarios considered, but do not allow systematic analysis of different network structures, since they can only be applied to a restrictive subset of networks. Such analysis is also challenging through stochastic simulations, since these provide little theoretical insight into the underlying dynamics and can be very time consuming for large population sizes. Therefore, approximation methods can be powerful, since they can significantly reduce the computational time and provide theoretical insight into the dynamics.

A common approach for approximating network-structured dynamics is moment-closure approximation, first used in statistical physics and more recently applied to epidemic modelling. In Chapter 4, we applied such moment-closure methods to the standard evolutionary graph theory dynamics. The application of these methods immediately ran into challenges, because despite the similarity to network-based epidemic models, evolutionary graph theory has extra correlations and dependencies that complicated the methods. By amending the methods to account for these difficulties, we constructed a novel approximation framework that can study evolutionary dynamics for arbitrary networks and update rules. We demonstrated the performance of these models on different population structures to illustrate when they perform well and when they do not.

Although these moment-closure methods are commonly applied to epidemic models, these methods exhibit endemic equilibrium solutions which are not present in the underlying stochastic models. This can make it challenging to link theoretical results from these models back to the stochastic process. To address this, in Chapter 5 we developed novel moment-closure methods to approximate the quasi-stationary distribution of the stochastic SIS epidemic model. The quasi-stationary distribution describes the behaviour of the epidemic if the pathogen does not go extinct. This is biologically relevant, since in many epidemics, particularly those with evolution and competition between pathogens, we do not observe extinction on a relevant timescale. The novel moment-closure methods we developed capture the QSD accurately on all networks considered.

Currently, this work has focused on the QSD for SIS dynamics. A future direction will

be to extend these methods to capture other epidemic dynamics, such as SIR and SEIR, as well as the transient behaviour of the conditional distribution. Using the work we have conducted on moment-closure in evolutionary graph theory, along with other similar works such as the pair-approximation models and neighbourhood approximation models in [59], it may be possible to extend these QSD methods to evolutionary dynamics. This would facilitate the analysis of evolutionary systems that have not gone to fixation, which could be important as often population structure can significantly amplify the time to fixation in evolutionary games, so we might expect to observe prolonged coexistence.

We have considered these moment-closure methods in the standard evolutionary graph theory dynamics and SIS epidemic dynamics as an illustration of their potential. We find that these methods significantly reduce the computational time and investigate in what scenarios they are most reliable. This provides a framework upon which approximations to the more realistic evolutionary dynamics developed in Chapters 2 and 3 can be developed. Future work therefore entails extending these methods to capture these more complicated dynamics.

Bibliography

[1] R. Albert and A. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.

[2] B. Allen, G. Lippner, Y. Chen, B. Fotouhi, N. Momeni, S. Yau, and M. A. Nowak. Evolutionary dynamics on any population structure. *Nature*, 544(7649):227–230, 2017.

[3] B. Allen and C. E. Tarnita. Measures of success in a class of evolutionary models with fixed population size and structure. *Journal of Mathematical Biology*, 68(1-2):109–143, 2014.

[4] L. J. S. Allen and A. M. Burgin. Comparison of deterministic and stochastic SIS and SIR models in discrete time. *Mathematical Biosciences*, 163(1):1 – 33, 2000.

[5] H. Andersson and T. Britton. Stochastic epidemics in dynamic populations: quasi-stationarity and extinction. *Journal of Mathematical Biology*, 41(6):559–580, 2000.

[6] T. Antal, S. Redner, and V. Sood. Evolutionary dynamics on degree-heterogeneous graphs. *Physical Review Letters*, 96(18):188104, 2006.

[7] T. Antal and I. Scheuring. Fixation of strategies for an evolutionary game in finite populations. *Bulletin of Mathematical Biology*, 68(8):1923–1944, 2006.

[8] K. Argasinski and M. Broom. Ecological theatre and the evolutionary game: how environmental and demographic factors determine payoffs in evolutionary games. *Journal of Mathematical Biology*, 67(4):935–962, 2013.

[9] J. R. Artalejo, A. Economou, and M. J. Lopez-Herrero. The maximum number of infected individuals in SIS epidemic models: Computational techniques and

quasi-stationary distributions. *Journal of Computational and Applied Mathematics*, 233(10):2563–2574, 2010.

[10] J. R. Artalejo, A. Economou, and M. J. Lopez-Herrero. Stochastic epidemic models with random environment: quasi-stationarity, extinction and final size. *Journal of Mathematical Biology*, 67(4):799–831, 2013.

[11] F. Baquero and J. Blázquez. Evolution of antibiotic resistance. *Trends in Ecology & Evolution*, 12(12):482–487, December 1997.

[12] V. C. Barbosa, R. Donangelo, and S. R. Souza. Early appraisal of the fixation probability in directed networks. *Physical Review E*, 82(4), 2010.

[13] M. S. Bartlett. Stochastic processes or the statistics of change. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2(1):44–64, 1953.

[14] H. J. E. Beaumont, J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey. Experimental evolution of bet hedging. *Nature*, 462(7269):90–93, 2009.

[15] A. A. Berryman. On principles, laws and theory in population ecology. *Oikos*, 103(3):695–701, 2003.

[16] A. Beutel, B. A. Prakash, R. Rosenfeld, and C. Faloutsos. Interacting viruses in networks: can both survive? In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434, 2012.

[17] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4):175 – 308, 2006.

[18] M. Born and H. S. Green. A general kinetic theory of liquids. i. the molecular distribution functions. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 188(1012):10–18, 1946.

[19] M. Broom, C. Hadjichrysanthou, and J. Rychtář. Evolutionary games on graphs and the speed of the evolutionary process. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2117):1327–1346, 2010.

[20] M. Broom and J. Rychtář. An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 464(2098):2609–2627, 2008.

[21] M. Broom and J. Rychtár. *Game-theoretical models in biology.* Chapman and Hall/CRC, 2013.

[22] N. Champagnat, R. Ferrière, and S. Méléard. Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. *Theoretical Population Biology*, 69(3):297–321, 2006.

[23] N. Champagnat and S. Méléard. Invasion and adaptive evolution for individual-based spatially structured populations. *Journal of Mathematical Biology*, 55(2):147–188, 2007.

[24] D. Cohen. Optimizing reproduction in a randomly varying environment. *Journal of Theoretical Biology*, 12(1):119–129, 1966.

[25] G. W. A. Constable and H. Kokko. The rate of facultative sex governs the number of expected mating types in isogamous species. *Nature Ecology & Evolution*, 2(7):1168, 2018.

[26] S. P. Courtney. Why insects move between host patches: some comments on 'risk-spreading'. *Oikos*, pages 112–114, 1986.

[27] J. Cremer, A. Melbinger, and E. Frey. Evolutionary and population dynamics: A coupled approach. *Physical Review E*, 84(5), 2011.

[28] R. Cressman and Y. Tao. The replicator equation and other game dynamics. *Proceedings of the National Academy of Sciences*, 111(Supplement 3):10810–10817, 2014.

[29] J. F. Crow and M. Kimura. *An introduction to population genetic theory.* Harper & Row, New York, 1970.

[30] P. Czuppon and G. W. A. Constable. Invasion and extinction dynamics of mating types under facultative sexual reproduction. *Genetics*, 213(2):567–580, 2019.

[31] S. Dambrine and M. Moreau. Note on the stochastic theory of a self-catalytic chemical reaction. i. *Physica A: Statistical Mechanics and its Applications*, 106(3):559–573, 1981.

[32] S. Dambrine and M. Moreau. Note on the stochastic theory of a self-catalytic chemical reaction. ii. *Physica A: Statistical Mechanics and its Applications*, 106(3):574–588, 1981.

[33] J. N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing continuous-time finite markov chains. *Journal of Applied Probability*, 4(1):192–196, 1967.

[34] C. Darwin. *On the origin of species, 1859*. Routledge, 2004.

[35] R. Dickman and M. M. De Oliveira. Quasi-stationary simulation of the contact process. *Physica A: Statistical Mechanics and its Applications*, 357(1):134–141, 2005.

[36] U. Dieckmann and R. Law. The dynamical theory of coevolution: A derivation from stochastic ecological processes. *Journal of Mathematical Biology*, 34(5-6):579–612, 1996.

[37] O. Diekmann. A beginners guide to adaptive dynamics. *Summer School on Mathematical Biology*, pages 63–100, 2002.

[38] R. Durrett. Coexistence in stochastic spatial models. *The Annals of Applied Probability*, 19(2):477–496, 2009.

[39] R. Durrett and S. Levin. Spatial models for species-area curves. *Journal of Theoretical Biology*, 179(2):119–127, 1996.

[40] K. T. D. Eames and M. J. Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences*, 99(20):13330–13335, 2002.

[41] D. J. D. Earn, J. Dushoff, and S. A. Levin. Ecology and evolution of the flu. *Trends in Ecology & Evolution*, 17(7):334–340, 2002.

[42] R. Erban, J. Chapman, and P. Maini. A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv preprint arXiv:0704.1908*, 2007.

[43] P. Erdős and A. Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5(1):17–60, 1960.

[44] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Physical Review E*, 86(4):041125, 2012.

[45] R. A. Fisher. Xxi.—on the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, 42:321–341, 1923.

[46] R. A. Fisher. *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford, 1930.

[47] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *The Annals of Applied Probability*, 14(4):1880–1919, 2004.

[48] C. W. Fox and C. M. Rauter. Bet-hedging and the evolution of multiple mating. *Evolutionary Ecology Research*, 5(2):273–286, 2003.

[49] S. A. Frank and M. Slatkin. Evolution in a variable environment. *The American Naturalist*, 136(2):244–260, 1990.

[50] M. Frasca and K. J. Sharkey. Discrete-time moment closure models for epidemic spreading in populations of interacting individuals. *Journal of Theoretical Biology*, 399:13–21, 2016.

[51] C. Fraser, K. Lythgoe, G. E. Leventhal, G. Shirreff, T. D. Hollingsworth, S. Alizon, and S. Bonhoeffer. Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. *Science*, 343(6177):1243727, 2014.

[52] M. Frean, P. B. Rainey, and A. Traulsen. The effect of population structure on the rate of evolution. *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1762):20130211, 2013.

[53] P. J. Gerrish and R. E. Lenski. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102:127, 1998.

[54] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[55] J. H. Gillespie. Natural selection for within generation variance in offspring number. *Genetics*, 76(3):601–606, 1974.

[56] C. A. Gravenmier, M. Siddique, and R. A. Gatenby. Adaptation to stochastic temporal variations in intratumoral blood flow: the warburg effect as a bet hedging strategy. *Bulletin of Mathematical Biology*, 80(5):954–970, 2018.

[57] B. T. Grenfell, O. G. Pybus, J. R. Gog, J. L. N. Wood, J. M. Daly, J. A. Mumford, and E. C. Holmes. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303(5656):327–332, 2004.

[58] P. Groisman and M. Jonckheere. Simulation of quasi-stationary distributions on countable spaces. *arXiv preprint arXiv:1206.6712*, 2012.

[59] C. Hadjichrysanthou, M. Broom, and I. Z. Kiss. Approximating evolutionary dynamics on networks using a neighbourhood configuration model. *Journal of Theoretical Biology*, 312:13–21, 2012.

[60] C. Hadjichrysanthou, M. Broom, and J. Rychtář. Evolutionary games on star graphs under various updating rules. *Dynamic Games and Applications*, 1(3):386–407, 2011.

[61] C. Hadjichrysanthou and K. J. Sharkey. Epidemic control analysis: designing targeted intervention strategies against epidemics propagated on contact networks. *Journal of Theoretical Biology*, 365:84–95, 2015.

[62] T. J. Hagenaars, C. A. Donnelly, and N. M. Ferguson. Spatial heterogeneity and the persistence of infectious diseases. *Journal of Theoretical Biology*, 229(3):349–359, 2004.

[63] W. P. Hanage, B. G. Spratt, K. M. E. Turner, and C. Fraser. Modelling bacterial speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1475):2039–2044, 2006.

[64] I. Hanski and O. Ovaskainen. Metapopulation theory for fragmented landscapes. *Theoretical Population Biology*, 64(1):119–127, 2003.

[65] I. Hanski, T. Schulz, S. C. Wong, V. Ahola, A. Ruokolainen, and S. P. Ojanen. Ecological and genetic basis of metapopulation persistence of the Glanville fritillary butterfly in fragmented landscapes. *Nature Communications*, 8:14504, 2017.

[66] T. E. Harris. Contact Interactions on a Lattice. *The Annals of Probability*, 2(6):969–988, 1974.

[67] C. Hauert and G. Szabó. Game theory and physics. *American Journal of Physics*, 73(5):405–414, 2005.

[68] L. Hindersin, M. Möller, A. Traulsen, and B. Bauer. Exact numerical calculation of fixation probability and time on graphs. *BioSystems*, 150:87–91, 2016.

[69] L. Hindersin and A. Traulsen. Counterintuitive properties of the fixation time in network-structured populations. *Journal of The Royal Society Interface*, 11(99):20140606, 2014.

[70] L. Hindersin and A. Traulsen. Most undirected random graphs are amplifiers of selection for birth-death dynamics, but suppressors of selection for death-birth dynamics. *PLoS Computational Biology*, 11(11):1–14, 2015.

[71] L. Hindersin, B. Werner, D. Dingli, and A. Traulsen. Should tissue structure suppress or amplify selection to minimize cancer risk? *Biology Direct*, 11(1):41, 2016.

[72] R. A. Holley and T. M. Liggett. Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model. *The Annals of Probability*, 3(4):643–663, 1975.

[73] C. S. Holling. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4(1):1–23, 1973.

[74] K. R. Hopper. *Risk-spreading and bet-hedging in insect population biology*, volume 44 of *Annual Review of Entomology*. 1999.

[75] K. R. Hopper, J. A. Rosenheim, T. Prout, and S. J. Oppenheim. Within-generation bet hedging: A seductive explanation? *Oikos*, 101(1):219–222, 2003.

[76] C. C. Horvitz and D. W. Schemske. Effects of plant size, leaf herbivory, local competition and fruit production on survival, growth and future reproduction of a neotropical herb. *Journal of Ecology*, 90(2):279–290, 2002.

[77] W. Huang, C. Hauert, and A. Traulsen. Stochastic game dynamics under demographic fluctuations. *Proceedings of the National Academy of Sciences*, 112(29):9064–9069, 2015.

[78] Y Jin, W. Wang, and S. Xiao. An SIRS model with a nonlinear incidence rate. *Chaos, Solitons & Fractals*, 34(5):1482–1497, 2007.

[79] K. C. Kao and G. Sherlock. Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae. *Nature Genetics*, 40(12):1499–1504, 2008.

[80] S. Karlin and H. M. Taylor. *A First Course in Stochastic Processes*. Gulf Professional Publishing, 1975.

[81] K. Kaveh, N. L. Komarova, and M. Kohandel. The duality of spatial death–birth and birth–death processes and limitations of the isothermal theorem. *Royal Society Open Science*, 2(4):140465, 2015.

[82] M. J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society B: Biological Sciences*, 266(1421):859–867, 1999.

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

[84] D. G. Kendall. An artificial realization of a simple "birth-and-death" process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(1):116–119, 1950.

[85] J. O. Kephart, S. R. White, and D. M. Chess. Computers and epidemiology. *IEEE Spectrum*, 30(5):20–26, 1993.

[86] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A.*, 115(772):700–721, 1927.

[87] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, 3(5):300–313, 1935.

[88] J. G. Kirkwood. The statistical mechanical theory of transport processes ii. transport in gases. *The Journal of Chemical Physics*, 15(1):72–76, 1947.

[89] J. G. Kirkwood and E. M. Boggs. The radial distribution function in liquids. *The Journal of Chemical Physics*, 10(6):394–402, 1942.

[90] 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.

[91] I. Z. Kiss, J. C. Miller, and P. L. Simon. Mathematics of epidemics on networks. *Cham: Springer*, 2017.

[92] I. Z. Kiss, C. G. Morris, F. Sélley, P. L. Simon, and R. R. Wilkinson. Exact deterministic representation of markovian SIR epidemics on networks with and without loops. *Journal of Mathematical Biology*, 70(3):437–464, 2015.

[93] D. R. Klein. The introduction, increase, and crash of reindeer on St. Matthew island. *The Journal of Wildlife Management*, pages 350–367, 1968.

[94] W. C. Krumbein and M. F. Dacey. Markov chains and embedded markov chains in geology. *Journal of the International Association for Mathematical Geology*, 1(1):79–96, 1969.

[95] R. J. Kryscio and C. Lefevre. *On the extinction of the SIS stochastic logistic epidemic*, pages 213–228. Statistical Methods in Computer Security. 2004.

[96] A. Lajmanovich and J. A. Yorke. A deterministic model for gonorrhea in a nonhomogeneous population. *Mathematical Biosciences*, 28(3-4):221–236, 1976.

[97] L. Lehmann and F. Balloux. Natural selection on fecundity variance in subdivided populations: kin selection meets bet-hedging. *Genetics*, 2007.

[98] S. Lehtinen, F. Blanquart, N. J. Croucher, P. Turner, M. Lipsitch, and C. Fraser. Evolution of antibiotic resistance is linked to any genetic mechanism affecting bacterial duration of carriage. *Proceedings of the National Academy of Sciences*, 114(5):1075–1080, 2017.

[99] R. Levins. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15(3):237–240, 1969.

[100] S. F. Levy, N. Ziv, and M. L. Siegal. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biology*, 10(5), 2012.

[101] M. Y. Li and J. S. Muldowney. Global stability for the SEIR model in epidemiology. *Mathematical biosciences*, 125(2):155–164, 1995.

[102] E. Lieberman, C. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, 2005.

[103] T. M. Liggett. *Interacting particle systems*, volume 276. Springer Science & Business Media, 2012.

[104] K. A. Lythgoe, L. Pellis, and C. Fraser. Is HIV short-sighted? Insights from a multistrain nested model. *Evolution*, 67(10):2769–2782, 2013.

[105] W. Maciejewski, F. Fu, and C. Hauert. Evolutionary game dynamics in populations with heterogenous structures. *PLOS Computational Biology*, 10:1–16, 04 2014.

[106] N. Masuda. Directionality of contact networks suppresses selection pressure in evolutionary dynamics. *Journal of Theoretical Biology*, 258(2):323–334, 2009.

[107] A. S. Mata and S. C. Ferreira. Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks. *Europhysics Letters*, 103(4):48003, 2013.

[108] L. Matthews, D. T. Haydon, D. J. Shaw, M. E. Chase-Topping, M. J. Keeling, and M. E. J. Woolhouse. Neighbourhood control policies and the spread of infectious diseases. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1525):1659–1666, 2003.

[109] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.

[110] J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246(5427):15, 1973.

[111] L. D. Mech. The wolves of isle royale. 1966.

[112] A. Melbinger, J. Cremer, and E. Frey. Evolutionary game theory in growing populations. *Physical Review Letters*, 105(17), 2010.

[113] J. A. J. Metz, S. A. H. Geritz, G. Meszena, F. J. A. Jacobs, and J. S. van Heerwaarden. Adaptive dynamics: A geometrical study of the consequences of nearly faithful reproduction. http://pure.iiasa.ac.at/id/eprint/4497/, 1995.

[114] M. Möller, L. Hindersin, and A. Traulsen. Exploring and mapping the universe of evolutionary graphs identifies structural properties affecting fixation probability and time. *Communications Biology*, 2(1):1–9, 2019.

[115] D. Mollison. Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(3):283–313, 1977.

[116] P. A. P. Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(1):60–71, 1958.

[117] S. Morita. Extended pair approximation of evolutionary game on complex networks. *Progress of Theoretical Physics*, 119(1):29–38, 2008.

[118] H. J. Muller. Some genetic aspects of sex. *The American Naturalist*, 66(703):118–138, 1932.

[119] W. H. Murray. The application of epidemiology to computer viruses. *Computers & Security*, 7(2):139–145, 1988.

[120] I. Nåsell. The quasi-stationary distribution of the closed endemic SIS model. *Advances in Applied Probability*, 28(3):895–932, 1996.

[121] I. Nåsell. On the quasi-stationary distribution of the stochastic logistic epidemic. *Mathematical Biosciences*, 156(1-2):21–40, 1999.

[122] I. Nåsell. On the time to extinction in recurrent epidemics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):309–330, 1999.

[123] M. E. J. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66(3), 2002.

[124] R. Noble, D. Burri, J. N. Kather, and N. Beerenwinkel. Spatial structure governs the mode of tumour evolution. Preprint, Cancer Biology, 2019.

[125] J. R. Norris and J. R. Norris. *Markov chains*. Number 2. Cambridge University Press, 1998.

[126] M. A. Nowak. *Evolutionary dynamics*. Harvard University Press, 2006.

[127] M. A. Nowak, C. E. Tarnita, and T. Antal. Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1537):19–30, 2010.

[128] T. Ohta. Near-neutrality in evolution of genes and gene regulation. *Proceedings of the National Academy of Sciences*, 99(25):16134–16137, 2002.

[129] H. Ohtsuki, C. Hauert, E. Lieberman, and M. A. Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092):502–505, 2006.

[130] B. Øksendal. *Stochastic differential equations*. Springer, 2003.

[131] A. Oliver. High frequency of hypermutable pseudomonas aeruginosa in cystic fibrosis lung infection. *Science*, 288(5469):1251–1253, May 2000.

[132] H. Olofsson, J. Ripa, and N. Jonzén. Bet-hedging as an evolutionary game: The trade-off between egg size and number. *Proceedings of the Royal Society B: Biological Sciences*, 276(1669):2963–2969, 2009.

[133] I. Oppenheim, K. E. Shuler, and G. H. Weiss. Stochastic theory of nonlinear rate processes with multiple stationary states. *Physica A: Statistical Mechanics and its Applications*, 88(2):191–214, 1977.

[134] C. E. Overton, M. Broom, C. Hadjichrysanthou, and K. J. Sharkey. Methods for approximating stochastic evolutionary dynamics on graphs. *Journal of Theoretical Biology*, 468:45–59, 2019.

[135] A. G. Pakes. Limit theorems for the population size of a birth and death process allowing catastrophes. *Journal of Mathematical Biology*, 25(3):307–325, 1987.

[136] R. Parshani, S. Carmi, and S. Havlin. Epidemic threshold for the susceptible-infectious-susceptible model on random networks. *Physical Review Letters*, 104(25):258701, 2010.

[137] R. W. Parsons and P. K. Pollett. Quasistationary distributions for autocatalytic reactions. *Journal of Statistical Physics*, 46(1-2):249–254, 1987.

[138] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925, 2015.

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

[140] K. Pattni, M. Broom, J. Rychtář, and L. J. Silvers. Evolutionary graph theory revisited: when is an evolutionary process equivalent to the moran process? *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 471(2182), 2015.

[141] A. Pavlogiannis, J. Tkadlec, K. Chatterjee, and M. A. Nowak. Construction of arbitrarily strong amplifiers of natural selection using evolutionary graph theory. *Communications Biology*, 1(1):71, 2018.

[142] L. Pellis, T. House, and M. J. Keeling. Exact and approximate moment closures for non-markovian network epidemics. *Journal of Theoretical Biology*, 382:160–177, 2015.

[143] J. Peña and G. Nöldeke. Variability in group size and the evolution of collective action. *Journal of Theoretical Biology*, 389:72–82, 2016.

[144] J. Pena, H. Volken, E. Pestelacci, and M. Tomassini. Conformity hinders the evolution of cooperation on scale-free networks. *Physical Review E*, 80(1):016110, 2009.

[145] T. Philippi. Bet-hedging germination of desert annuals: variation among populations and maternal effects in lepidium lasiocarpum. *The American Naturalist*, 142(3):488–507, 1993.

[146] P. K. Pollett. On the problem of evaluating quasistationary distributions for open reaction schemes. *Journal of Statistical Physics*, 53(5-6):1207–1215, 1988.

[147] P. K. Pollett. The determination of quasistationary distributions directly from the transition rates of an absorbing markov chain. *Mathematical and Computer Modelling*, 22(10-12):279–287, 1995.

[148] P. K. Pollett and S. Kumar. On the long-term behaviour of a population that is subject to large-scale mortality or emigration. In *Proceedings of the 8th National*

Conference of the Australian Society for Operations Research, volume 196, page 207, 1987.

[149] H. R. Pulliam. Sources, sinks, and population regulation. *The American Naturalist*, 132(5):652–661, 1988.

[150] R. R. Regoes, S. Hamblin, and M. M. Tanaka. Viral mutation rates: modelling the roles of within-host viral dynamics and the trade-off between replication fidelity and speed. *Proceedings of the Royal Society B: Biological Sciences*, 280(1750):20122047, 2013.

[151] J. Renton and K. M. Page. Evolution of cooperation in an epithelium. *Journal of the Royal Society Interface*, 16(152):20180918, 2019.

[152] S. H. Rice. A stochastic version of the price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC Evolutionary Biology*, 8(1):262, 2008.

[153] S. H. Rice and A. Papadopoulos. Evolution with stochastic fitness and stochastic migration. *PloS One*, 4(10):e7130, 2009.

[154] B. Roche, J. M. Drake, and P. Rohani. An agent-based model to study the epidemiological and evolutionary dynamics of influenza viruses. *BMC Bioinformatics*, 12(1):87, 2011.

[155] K. Rock, S. Brand, J. Moir, and M. J. Keeling. Dynamics of infectious diseases. *Reports on Progress in Physics*, 77(2):026602, 2014.

[156] D. A. Roff. Defining fitness in evolutionary models. *Journal of Genetics*, 87(4):339–348, 2008.

[157] T. Rogers. Maximum-entropy moment-closure for stochastic systems on networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(05):P05007, 2011.

[158] R. B. Root and P. M. Kareiva. The search for resources by cabbage butterflies (pieris rapae): ecological consequences and adaptive significance of markovian movements in a patchy environment. *Ecology*, 65(1):147–165, 1984.

[159] J. N. Rosenquist. The spread of alcohol consumption behavior in a large social network. *Annals of Internal Medicine*, 152(7):426, 2010.

[160] A. Sarhan and H. Kokko. Multiple mating in the glanville fritillary butterfly: A case of within-generation bet hedging? *Evolution*, 61(3):606–616, 2007.

[161] V. B. Scheffer. The rise and fall of a reindeer herd. *The Scientific Monthly*, 73(6):356–362, 1951.

[162] J. Seger. What is bet-hedging? *Oxford Surveys in Evolutionary Biology*, 4:182–211, 1987.

[163] P. Shakarian, P. Roos, and A. Johnson. A review of evolutionary graph theory with applications to game theory. *Biosystems*, 107(2):66–80, 2012.

[164] P. Shakarian, P. Roos, and G. Moores. A novel analytical method for evolutionary graph theory problems. *BioSystems*, 111(2):136–144, 2013.

[165] M. Shaked and J. G. Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007.

[166] K. J. Sharkey. Deterministic epidemiological models at the individual level. *Journal of Mathematical Biology*, 57(3):311–331, 2008.

[167] K. J. Sharkey. Deterministic epidemic models on contact networks: correlations and unbiological terms. *Theoretical Population Biology*, 79(4):115–129, 2011.

[168] K. J. Sharkey, I. Z. Kiss, R. R. Wilkinson, and P. L. Simon. Exact equations for SIR epidemics on tree graphs. *Bulletin of Mathematical Biology*, 77(4):614–645, 2015.

[169] K. J. Sharkey and R. R. Wilkinson. Complete hierarchies of SIR models on arbitrary networks with exact and approximate moment closure. *Mathematical Biosciences*, 264(1):74–85, 2015.

[170] M. Shpak. Evolution of variance in offspring number: the effects of population size and migration. *Theory in Biosciences*, 124(1):65–85, 2005.

[171] M. Shpak and S. R. Proulx. The role of life cycle and migration in selection for variance in offspring number. *Bulletin of Mathematical Biology*, 69(3):837–860, 2007.

[172] A. Singer. Maximum entropy formulation of the kirkwood superposition approximation. *The Journal of Chemical Physics*, 121(8):3657–3666, 2004.

[173] J. Starrfelt and H. Kokko. Bet-hedging-a triple trade-off between means, variances and correlations. *Biological Reviews*, 87(3):742–755, 2012.

[174] M. P. H. Stumpf, Z. Laidlaw, and V. A. A. Jansen. Herpes viruses hedge their bets. *Proceedings of the National Academy of Sciences*, 99(23):15234–15237, 2002.

[175] G. Szabó and G. Fath. Evolutionary games on graphs. *Physics Reports*, 446(4-6):97–216, 2007.

[176] L. Ternent, R. J. Dyson, A. Krachler, and S. Jabbari. Bacterial fitness shapes the population dynamics of antibiotic-resistant and-susceptible bacteria in a model of combined antibiotic and anti-virulence treatment. *Journal of Theoretical Biology*, 372:1–11, 2015.

[177] J. Tkadlec, A. Pavlogiannis, K. Chatterjee, and M. A. Nowak. Fixation probability and fixation time in structured populations. *arXiv preprint arXiv:1810.02687*, 2018.

[178] A. Traulsen, J. C. Claussen, and C. Hauert. Coevolutionary dynamics: From finite to infinite populations. *Physical Review Letters*, 95:238701, 2005.

[179] A. Traulsen and C. Hauert. *Stochastic Evolutionary Game Dynamics*, volume 2 of *Reviews of Nonlinear Dynamics and Complexity*, pages 25–61. 2010.

[180] J. Tufto. Genetic evolution, plasticity, and bet-hedging as adaptive responses to temporally autocorrelated fluctuating selection: a quantitative genetic model. *Evolution*, 69(8):2034–2049, 2015.

[181] E. A. van Doorn and P. K. Pollett. Quasi-stationary distributions for discrete-state models. *European Journal of Operational Research*, 230(1):1–14, 2013.

[182] C. H. van Dorp, M. van Boven, and R. J. De Boer. Immuno-epidemiological modeling of HIV-1 predicts high heritability of the set-point virus load, while selection for CTL escape dominates virulence evolution. *PLoS Computational Biology*, 10(12):e1003899, 2014.

[183] N. G. Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[184] P. Van Mieghem. The N-intertwined SIS epidemic network model. *Computing*, 93(2-4):147–169, 2011.

[185] P. Van Mieghem, J. Omic, and R. Kooij. Virus spread in networks. *IEEE/ACM Transactions on Networking (TON)*, 17(1):1–14, 2009.

[186] D. L. Venable. Bet hedging in a guild of desert annuals. *Ecology*, 88(5):1086–1090, 2007.

[187] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.*, pages 25–34. IEEE, 2003.

[188] S. A. Ward and A. F. G. Dixon. Spreading the risk, and the evolution of mixed strategies: seasonal variation in aphid reproductive biology. In *Advances in Invertebrate Reproduction 3: Proceedings of 3rd International Symposium, International Society of Invertebrate Reproduction.* Amsterdam: Elsevier Science Publishers, 1984.

[189] P. J. Watson. Multiple paternity as genetic bet-hedging in female sierra dome spiders, linyphia litigiosa (linyphiidae). *Animal Behaviour*, 41(2):343–360, 1991.

[190] J. H. Werren. Sex ratio evolution under local mate competition in a parasitic wasp. *Evolution*, 37(1):116–124, 1983.

[191] J. C. Wierman and D. J. Marchette. Modeling computer virus prevalence with a susceptible-infected-susceptible model with reintroduction. *Computational Statistics & Data Analysis*, 45(1):3–23, 2004.

[192] G. Wild and P. D. Taylor. Fitness and evolutionary stability in game theoretic models of finite populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1555):2345–2349, 2004.

[193] R. R. Wilkinson and K. J. Sharkey. An exact relationship between invasion probability and endemic prevalence for markovian SIS dynamics on networks. *PloS One*, 8(7):e69028, 2013.

[194] R. R. Wilkinson and K. J. Sharkey. Impact of the infectious period on epidemics. *Physical Review E*, 97(5):052403, 2018.

[195] R. Wise. The relentless rise of resistance? *Journal of Antimicrobial Chemotherapy*, 54(2):306–310, 2004.

[196] S. Wright. The distribution of gene frequencies under irreversible mutation. *Proceedings of the National Academy of Sciences of the United States of America*, 24(7):253, 1938.

[197] S. Wright. The genetical structure of populations. *Annals of Eugenics*, 15(1):323–354, 1949.

[198] Y. Yasui. Female multiple mating as a genetic bet-hedging strategy when mate choice criteria are unreliable. *Ecological Research*, 16(4):605–616, 2001.

[199] Y. Yasui and F. Garcia-Gonzalez. Bet-hedging as a mechanism for the evolution of polyandry, revisited. *Evolution*, 70(2):385–397, 2016.

[200] Y. Yasui and J. Yoshimura. Bet-hedging against male-caused reproductive failures may explain ubiquitous cuckoldry in female birds. *Journal of Theoretical Biology*, 437:214–221, 2018.

[201] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977.

[202] J. Zhang, J. J. Cunningham, J. S. Brown, and R. A. Gatenby. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature Communications*, 8(1):1816, 2017.

[203] W. Zhong, J. Liu, and L. Zhang. Evolutionary dynamics of continuous strategy games on graphs and social networks under weak selection. *BioSystems*, 111(2):102–110, 2013.

[204] J. Zukewich, V. Kurella, M. Doebeli, and C. Hauert. Consolidating birth-death and death-birth processes in structured populations. *PLoS One*, 8(1):e54639, 2013.