Stochastic Growth Models

by

Eric Foxall
B.A.Sc., University of British Columbia, 2010
M.Sc., University of Victoria, 2011

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Mathematics and Statistics

Stochastic Growth Models

by

Eric Foxall
B.A.Sc., University of British Columbia, 2010
M.Sc., University of Victoria, 2011

Supervisory Committee

Dr. R. Edwards, Co-Supervisor
(Department of Mathematics and Statistics)

Dr. P. van den Driessche, Co-Supervisor
(Department of Mathematics and Statistics)

Dr. A. Quas, Committee Member
(Department of Mathematics and Statistics)

Dr. P. Greenwood, Committee Member
(Department of Mathematics and Statistics, University of British Columbia)

Dr. S. Krone, External Examiner
(Department of Mathematics, University of Idaho)

Supervisory Committee

Dr. R. Edwards, Co-Supervisor
(Department of Mathematics and Statistics)

Dr. P. van den Driessche, Co-Supervisor
(Department of Mathematics and Statistics)

Dr. A. Quas, Committee Member
(Department of Mathematics and Statistics)

Dr. P. Greenwood, Committee Member
(Department of Mathematics and Statistics, University of British Columbia)

Dr. S. Krone, External Examiner
(Department of Mathematics, University of Idaho)

ABSTRACT

This thesis is concerned with certain properties of stochastic growth models. A stochastic growth model is a model of infection spread, through a population of individuals, that incorporates an element of randomness. The models we consider are variations on the contact process, the simplest stochastic growth model with a recurrent infection.

Three main examples are considered. The first example is a version of the contact process on the complete graph that incorporates dynamic monogamous partnerships. To our knowledge, this is the first rigorous study of a stochastic spatial model of infection spread that incorporates some form of social dynamics. The second example is a

non-monotonic variation on the contact process, taking place on the one-dimensional lattice, in which there is a random incubation time for the infection. Some techniques exist for studying non-monotonic particle systems, specifically models of competing populations [38] [12]. However, ours is the first rigorous study of a non-monotonic stochastic spatial model of infection spread. The third example is an additive two-stage contact process, together with a general duality theory for multi-type additive growth models. The two-stage contact process is first introduced in [29], and several open questions are posed, most of which we have answered. There are many examples of additive growth models in the literature [26] [16] [29] [49], and most include a proof of existence of a dual process, although up to this point no general duality theory existed.

In each case there are three main goals. The first is to identify a phase transition with a sharp threshold or "critical value" of the transmission rate, or a critical surface if there are multiple parameters. The second is to characterize either the invariant measures if the population is infinite, or to characterize the metastable behaviour and the time to extinction of the disease, if the population is finite. The final goal is to determine the asymptotic behaviour of the model, in terms of the invariant measures or the metastable states.

In every model considered, we identify the phase transition. In the first and third examples we show the threshold is sharp, and in the first example we calculate the critical value as a rational function of the parameters. In the second example we cannot establish sharpness due to the lack of monotonicity. However, we show there is a phase transition within a range of transmission rates that is uniformly bounded away from zero and infinity, with respect to the incubation time.

For the partnership model, we show that below the critical value, the disease dies out within $C \log N$ time for some $C > 0$, where N is the population size. Moreover we show that above the critical value, there is a unique metastable proportion of infectious individuals that persists for at least $e^{\gamma N}$ time for some $\gamma > 0$.

For the incubation time model, we use a block construction, with a carefully chosen good event to circumvent the lack of monotonicity, in order to show the existence of a phase transition. This technique also guarantees the existence of a non-trivial

invariant measure. Due to the lack of additivity, the identification of all the invariant measures is not feasible. However, we are able to show the following is true. By rescaling time so that the average incubation period is constant, we obtain a limiting process as the incubation time tends to infinity, with a sharp phase transition and a well-defined critical value. We can then show that as the incubation time approaches infinity (or zero), the location of the phase transition in the original model converges to the critical value of the limiting process (respectively, the contact process).

For the two-stage contact process, we can show that there are at most two extremal invariant measures: the trivial one, and a non-trivial upper invariant measure that appears above the critical value. This is achieved using known techniques for the contact process. We can show complete convergence, from any initial configuration, to a combination of these measures that is given by the survival probability. This, and some additional results, are in response to the questions posed by Krone in his original paper [29] on the model.

We then generalize these ideas to develop a theory of additive growth models. In particular, we show that any additive growth model, having any number of types and interactions, will always have a dual process that is also an additive growth model. Under the additional technical condition that the model preserves positive correlations, we can then harness existing techniques to conclude existence of at most two extremal invariant measures, as well as complete convergence.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

My parents, for their constant help and support.

My advisors Rod and Pauline, for their mentoring, support, encouragement, and patience.

My "informal advisor" Anthony, for interesting discussions, research guidance and for always challenging me and stimulating my interest in mathematics.

Grad students at UVic, for a fun and friendly environment in which to work and play.

NSERC, UVic and PIMS IGTC, for funding me with a Scholarship.

Chapter 1

Introduction

This thesis concerns *stochastic growth models*, which are Markov processes $(\xi_t)_{t\geq 0}$ with state space F^V, where F is a finite set, usually $\{0,1\}$ or $\{0,1,2\}$, and $G=(V,E)$ is a connected undirected graph with a finite or countably infinite number of vertices. The vertices V, which we usually refer to as sites, represent individuals in the population, and the edges E represent connections between individuals. We usually use ξ, η, ζ to represent a point in the state space, and we call it a *configuration*.

For us, a growth model usually means spread of an infection, though it may also be thought of as growth and dispersal of a population of organisms, via the correspondence of healthy with vacant and infected with occupied. In our case, the organism considered should probably be a plant, since individual organisms won't be moving around. In any case, the model should have the following properties:

- the 0 state will always mean healthy/vacant,
- every other state in F is an *active* state, i.e., such that a site in an active state can cause other sites to become active,
- the *all-zero* configuration with $\xi(x)=0$ for all $x\in V$ is an absorbing state, and
- from any configuration with $\xi(x)\neq 0$ for at most finitely many x, with positive probability the all-zero configuration is reached at some point in time

In every model we consider, the following *irreducibility* assumption also holds: if $V_0\subset V$ is a finite set and $\phi,\psi : V_0\to F$ are functions such that $\phi(x)\neq 0$ for some $x\in V_0$ then if ξ_0 satisfies $\xi_0(x)=\phi(x)$ for all $x\in V_0$, with positive probability there

is $t > 0$ so that $\xi_t(x) = \psi(x)$ for all $x \in V_0$.

The simplest model with these properties is called the *contact process* on a connected undirected graph $G = (V, E)$ with $F = \{0, 1\}$, and is defined by the two transitions:

- if x is infected then x recovers, i.e., goes from 1 to 0 at rate 1, and
- if x is healthy then x becomes infected, i.e., goes from 0 to 1, at rate λ times the number of infected neighbours of x

where a *neighbour* of x is a site y such that $xy \in E$. The meaning of the transition rates is that if $\xi_t(x) = 1$, for example, the first time s such that $\xi_{t+s}(x) = 0$ is distributed like an exponential random variable with rate 1, and if $\xi_t(x) = 0$, x has k infected neighbours, and the state of x and its neighbours does not otherwise change, the first time s such that $\xi_{t+s}(x) = 1$ is distributed like an exponential random variable with rate $k\lambda$.

The contact process is well-studied; see for example [32] for an introduction or [34], [14] for more recent work. The models we consider are all variations on the theme of the contact process, and we consider three main examples, as follows; since the following description is intended as an advertisement of the main results, we use a bit of terminology that is not introduced till later in this introduction.

- In Chapter 2, we consider a variation of the contact process that incorporates dynamic monogamous partnerships.
- In Chapter 3, we consider a variation of the contact process that is not monotonic.
- In Chapters 4 and 5, we consider a specific example, and then a general theory, of additive multi-type growth models.

There are some examples in the literature of contact processes evolving in a random environment [6] [44] [46]. However, ours is the first rigorous study of a random environment consisting of dynamic social interactions, an area of great interest in epidemiology. Moreover, since our model is constructed on a complete graph, it exhibits mean-field behaviour in the large population limit, and this means that we can obtain

exact results. In other words, we can calculate the critical value, and describe in detail the behaviour of the model in the subcritical, critical and supercritical regimes. This improves on analogous results such as [43] in which a version of the contact process on the complete graph is considered, and the subcritical and supercritical, but not the critical behaviour, are described. For our model, the critical case is non-trivial and requires a detailed analysis.

Monotonicity is a property that greatly simplifies the analysis of growth models. To our knowledge, there is no literature to date on non-monotonic variations of the contact process. However, some tools are available for studying non-monotonic particle systems, as for example the block construction technique that is often used to prove coexistence in models of interacting populations [11, Chapter 4] [12]. In our model the lack of monotonicity is due to a random incubation time between exposure and onset of infectiousness. We make use of block construction techniques and some fairly delicate limiting arguments in order to analyze the phase transition in the model, for arbitrary values of the incubation time and also in the limit as it tends to zero or infinity.

There are many examples of additive growth models, including the contact process itself as well as [16] [29] [49]. In many specific cases the existence of a dual process has been established, although up to this point no general theory existed. Our results show that under a broad definition that includes models with arbitrarily many types with possibly complex interrelations between the types, additivity implies the existence of a dual process. The dual types, and the transitions, are constructed directly from the model using an algebraic construction, which is then shown to be compatible with the graphical representation. This allows classical problems, such as the equivalence of two definitions of the critical value, and characterization of the stationary distributions, to be solved for such models, using known techniques for the contact process.

1.1 Main Goals

When studying growth models, there are three main goals:

- show existence of a phase transition,
- determine and characterize any invariant distributions or metastable states, and
- determine the asymptotic or long-time behaviour.

We discuss these goals in the order just stated, beginning with the first.

1.1.1 Existence of a phase transition

In every model we consider there is at least one transmission parameter λ. Starting the process with exactly one infected site, an important first question is whether for large values of λ the infection tends to survive and spread while for small values of λ it tends to die out. Borrowing a term from statistical physics, if this is the case we say the model exhibits a *phase transition* as the value of λ is varied. For a growth model, identifying a phase transition is the first goal.

For a connected graph with an infinite number of sites, a reasonable condition for survival is that there is a site $x \in V$ so that $\mathbb{P}(|\xi_t| > 0,\ \forall t > 0 \mid \xi_0 = \mathbf{1}(x))$, where the indicator function $\mathbf{1}(x)$ is defined by $\mathbf{1}(x)(y) = 1$ if $y = x$ and $= 0$ if $y \neq x$. In other words, starting from a single infected site there is a positive probability that the disease persists indefinitely. This is called *single-site survival*. Letting $\sigma(x)$ denote the above probability, it is easy to show using our irreducibility assumption and the strong Markov property that for a given set of parameter values either $\sigma(x) = 0$ for all $x \in V$ or else $\sigma(x) > 0$ for all $x \in V$, so the single-site survival condition does not depend on the site considered.

For a finite graph, i.e., with $|V| = N < \infty$, since the state space F^V is finite and the all-zero configuration can be reached from any configuration, eventual recovery is certain. In this case we say the infection tends to spread if, for example, from an initial configuration with a single infected site, there is a positive probability that the infection persists for a long time before dying out. To make this precise we assume the model is defined for any size N of the population, then we can say the infection persists for a long time if with probability $\geq p > 0$ not depending on N, starting from $|\xi_0| = 1$ the infection survives for an amount of time at least $e^{\gamma N}$ for some $\gamma > 0$. In other words, once the infection has become endemic, it can only die out by means of

a sudden extinction event.

The reason why we say "in other words" in this case can be thought of from the perspective of a random walk in the number of infected sites $I(t)$. If $I(t)$ tends to drift towards a value $I^* = i^*N$, then if it falls below $(i^* - \epsilon)N$ it will tend to return above $(i^* - \epsilon/2)N$ before falling again. Therefore, to overcome this drift requires to fall by an amount of order N more or less all at once. In the presence of a fixed amount of upward drift, a sudden decrease of order N has probability of order $e^{-\gamma N}$, so it takes order of $e^{\gamma N}$ attempts before this decrease occurs. This phenomenon is known as *metastability*, so called because the proportion i^*, while not asymptotically stable, is relatively stable for a long period of time.

Ideally, if we fix all other parameters, there is a critical value λ_c of the transmission parameter such that the infection tends to die out when $\lambda < \lambda_c$ and tends to spread when $\lambda > \lambda_c$. In this case, we say the phase transition is *sharp*, and occurs at λ_c. As discussed in Section 1.2.1, we can show this is the case when the model has a nice monotonicity property, as in Chapter 4. In other cases, as in Chapter 3, the best we can do is to define upper and lower values $\lambda^+ \geq \lambda^-$ such that the infection survives when $\lambda > \lambda^+$ and dies out when $\lambda < \lambda^-$.

1.1.2 Characterization of invariant distributions

A second important goal is the characterization of invariant distributions for the process, which are the stochastic analogues of the equilibrium points of a system of differential equations. Let $\mathbb{P}_\mu(\xi_t \in \cdot)$ denote the distribution of the process at time t, started from the measure μ. An invariant distribution is a probability measure μ with $\mathbb{P}_\mu(\xi_t \in A) = \mu(A)$ for all sets of configurations A belonging to the usual σ algebra on F^V.

Except in the case of spontaneous infection, the measure δ_0 that concentrates on the disease-free state is an invariant distribution. A non-trivial invariant distribution μ should satisfy $\mu(\{\xi : |\xi| > 0\}) > 0$, i.e., there is a positive probability with respect to μ that some site is infected. As shown in Theorem 1.2.1, if the model satisfies a certain monotonicity property we can deduce the existence of a unique "largest"

invariant distribution ν, in the sense that for any finite subset $V_0 \subset V$ and any other invariant distribution μ,

$$\nu(\{\xi : \xi(x) = 1, \forall x \in V_0\}) \geq \mu(\{\xi : \xi(x) = 1, \forall x \in V_0\})$$

The distribution ν, when it exists, is called the *upper invariant measure*. Showing that $\nu \neq \delta_0$ is a second way of characterizing survival of the infection.

As noted above, on a finite graph with no spontaneous infection, with probability 1, $|\xi_t| = 0$ for t large enough, which implies that δ_0 is the only invariant distribution. However, there may be a metastable state, which in our case is represented by a fixed proportion of infected individuals around which the process hovers for a long time, before eventually dying out.

In both cases, we can think of a non-trivial invariant distribution or a metastable state as an *endemic* state of the population, where usually the infection persists in the population with some positive proportion of individuals infected at any moment in time. In the simplest case, for a given model there is at most one endemic state, which if it is attracting tells us more or less what happens when the infection survives, and this leads us to our next goal.

1.1.3 Asymptotic behaviour

The third and final goal, after identifying a phase transition and any invariant distributions or metastable states, is to characterize the asymptotic behaviour of the model. In particular, when there is an endemic state, is it attracting? For the contact process (see for example [21] for early work in this direction, or [3] for more recent work) we can show that when single-site survival occurs, *complete convergence* holds in the following sense: from any initial distribution μ_0, letting μ_t denote the distribution at time t, we have

$$\mu_t \Rightarrow \alpha\delta_0 + (1 - \alpha)\nu$$

as $t \to \infty$, where ν is the upper invariant measure, δ_0 concentrates on the configuration with all 0s, $\alpha = \mathbb{P}(\xi_t \text{ dies out})$, and $\Rightarrow$ denotes weak convergence of measures, which for us means that for any $V_0 \subset V$ and $\phi : V_0 \to F$ the probabilities

$\mathbb{P}(\xi_t(x) = \phi(x)\ \forall x \in V_0)$ converge as $t \to \infty$. Part of our work in Chapters 4 and 5 is to show complete convergence for some well-behaved generalizations of the contact process.

For models on a finite graph, we would like to know how quickly the infection dies out when it does, or how quickly it spreads, and how long it persists. If $|V| = N$ and the model is *subcritical*, i.e., the infection tends to die out, it is reasonable to expect the infection to die out within an amount of time that is of order $\log N$. To see this, imagine the case of no transmission, so there are N particles in state 1, each waiting an independent unit exponential amount of time before going to state 0 and remaining there. The probability that all particles are in state 0 is, by independence, $(1 - e^{-t})^N$. Setting $t = \log N + c$ gives $(1 - e^{-c}/N)^N$ which approaches $e^{-e^{-c}}$ as $N \to \infty$; as $c \to -\infty$ this approaches $e^{-e^{\infty}} = e^{-\infty} = 0$ and as $c \to \infty$ it approaches $e^{-e^{-\infty}} = e^0 = 1$, so the time to extinction (all particles in state 0) is equal to $\log N + O(1)$, i.e., is equal to $\log N$ plus fluctuations of constant order.

As discussed earlier, if the model is *supercritical* and initially, there are enough infected individuals, the infection survives for an amount of time that is exponential in N. We show these asymptotics hold for a certain model in Chapter 2.

1.2 Techniques

A number of techniques show up again and again when studying growth models. The techniques we use fall into a few main categories:

- techniques utilizing the graphical representation, including stochastic domination and duality
- comparison of the growth model to an oriented percolation process
- comparison of the growth model to a branching process
- comparison of the growth model to mean-field equations

The first technique is of general usefulness in constructing the process and in comparing the evolution for different choices of initial data, and in particular for proving the existence of a critical value λ_c when the process is well enough behaved. We use the

second technique primarily when studying models on the lattices $\mathbb{Z}^d$, and in particular on $\mathbb{Z}$. The third technique allows us to get rough bounds on λ_c, and is of genuine usefulness in mean-field type models, i.e., models lacking a genuine spatial structure, for determining the short-time behaviour of the infection, when the number of infectious individuals is small relative to the population size. The fourth technique is useful when the so-called mean-field equations, which are a set of ODEs approximating the evolution of certain observables of the process, constitute a good approximation of the stochastic evolution of those observables. We discuss each technique, using the contact process as an example.

1.2.1 The Graphical Construction

The first technique, called the *graphical representation* or graphical construction, is perhaps the most fundamental. This is a way of constructing a growth model (or a more general particle system) from a collection of Poisson processes in spacetime, i.e., on the set $S = G \times [0, \infty)$ in such a way that we can easily view individual realizations of the process. Thus, rather than viewing the process as a collection of random variables $t \mapsto \xi_t$ we can view it as a random function $\omega \mapsto (\xi_t^{(w)})_{t \geq 0}$ where ω is an element of the probability space, which for us corresponds to a specific choice of points in the relevant Poisson processes.

The graphical construction is originally due to Harris [26] and is indispensable in the study of growth models and interacting particle systems in general. The results discussed in this section can be found in the reference book [32].

For concreteness, we show how to construct the contact process using the graphical representation, and then derive some of the basic properties of the contact process from relatively simple graphical arguments. For the sake of contrast, however, we first note how the contact process can be constructed for certain initial data as a continuous-time Markov chain. Given a graph $G = (V, E)$, recall the state space for the contact process on G is $\{0, 1\}^V$, which is equivalent to the set of subsets of V via the correspondence $A = \{x : \xi(x) = 1\}$, so we can think of the process as $(A_t)_{t \geq 0}$ where A_t is the set of infectious sites at time t. In this setting transitions are

- for each $x \in A$, $A \to A \setminus \{x\}$ at rate 1, and
- for each $y \notin A$ such that $xy \in E$ for some $x \in A$, $A \to A \cup \{y\}$ at rate $\lambda n(y)$,

where $n(y)$ is the cardinality of the set $\{x \in A : xy \in E\}$. If $|A_0| < \infty$ then $(A_t)_{t\geq 0}$ can be constructed as in [41] on the countable state space of finite subsets of V, since the transition rate out of any state is finite. Under mild conditions on G, for example if the degree of each vertex is at most M for some fixed M, explosion (i.e. having $|A_t| \to \infty$ in finite time) does not occur, and the process is defined for all $t > 0$.

The above definition shows how we can picture the process as a randomly evolving subset of V, which is nice, but the graphical construction gives us a better sort of "phase portrait". At each site x, the state at x

- goes from $1 \to 0$ at rate 1, and
- goes from $0 \to 1$ at rate $\lambda n(x)$

To construct the process we start with an independent collection of Poisson processes $\{U_x : x \in V\}$ each with intensity 1, and another independent collection $\{U_{xy} : xy \in E\}$ each with intensity λ. We refer to the points in these Poisson processes as the *substructure*, since they underlie the infection process $(\xi_t)_{t\geq 0}$ and will determine its transitions.

If the number of sites is finite, the total rate $|V| + \lambda|E|$ of Poisson processes is finite, so the set of times t such that a point (x, t) (or (xy, t)) is a point in some U_x (or U_{xy}) is well ordered. For later reference we call these the event times. Denoting these times $0 = t_0 < t_1 < t_2 < ...$, in this case the process can be constructed one event at a time: given ξ_0, then inductively, if $\xi_{t_i}(x) = 1$ and $(x, t_{i+1}) \in U_x$ then $\xi_{t_{i+1}}(x) = 0$, and if $\xi_{t_i}(x) = 0$ and $(xy, t_{i+1}) \in U_{xy}$ for some y such that $xy \in E$ and $\xi_{t_i}(y) = 1$ then $\xi_{t_{i+1}}(x) = 1$, and the configuration is otherwise unchanged. Fill in the state at other times by setting, for $t_i < t < t_{i+1}$, $\xi_t = \xi_{t_i}$.

If the number of sites is infinite, it becomes necessary to determine, for each point (x, t) in spacetime, the set $S_{x,t}$ of points (y, s) that can affect the state at (x, t), by tracing backwards in time from (x, t). Provided $S_{x,t}$ is bounded with probability 1, which is true for example if G is of bounded degree, then after restricting to $S_{x,t}$ the event times are well-ordered, and so given ξ_0 we can compute the state at (x, t) recursively, as above.

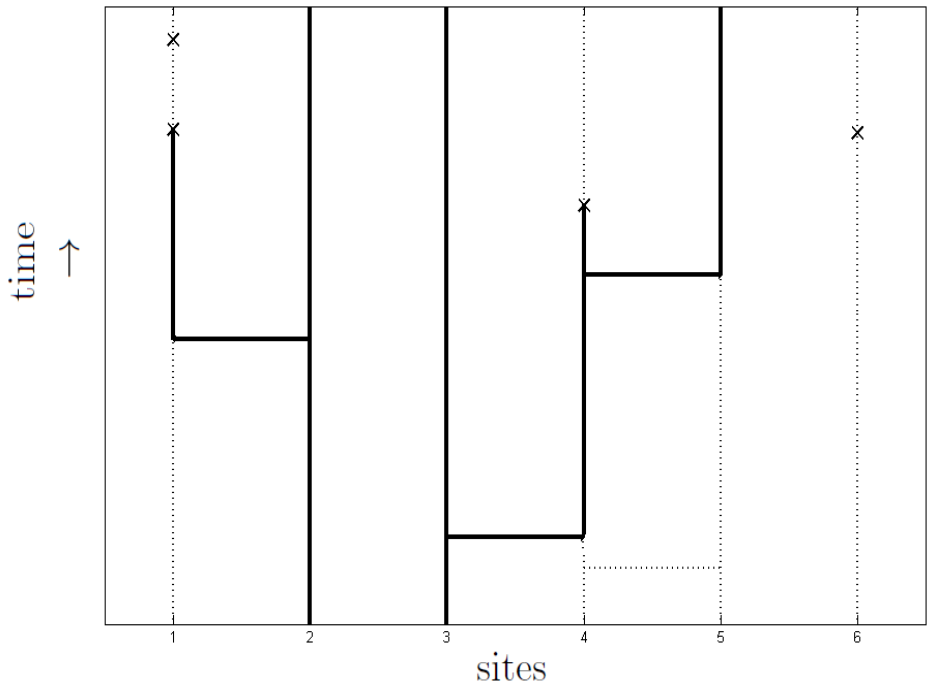

Figure 1.1: An illustration of the graphical construction via active paths. There are six sites, and time evolves in the upward direction; here, only sites 2 and 3 are initially infectious. Crosses denote events in U_x and horizontal segments denote events in U_{xy}, bold if used and dotted if unused. Points (x,t) such that $\xi_t(x) = 1$ are in bold.

For the contact process and some other processes there is a more direct way to determine $\xi_t(x)$, as follows. Say that a list $(v_1, h_1, ..., v_n)$ of segments in S with $v_i = \{x_i\} \times [t_{i-1}, t_i]$ and $h_i = \{x_i, x_{i+1}\} \times \{t_i\}$ is an *active path* if $\xi_{t_0}(x_1) = 1$, $t_i \in U_{x_i x_{i+1}}$ for $i = 1, ..., n-1$ and $(x_i, t) \notin U_{x_i}$ for $t \in [t_{i-1}, t_i]$ and $i = 1, .., n$. Then let $\xi_t(x) = 1$ if and only if there exists an active path from $(y, 0)$ up to (x, t), for some y such that $\xi_0(y) = 1$. A picture is given in Figure 1.1.

Fixing a realization ω of $\{U_x\}_{x \in V}$ and $\{U_{xy}\}_{xy \in E}$, which in the picture corresponds to a set of crosses $\times$ and a set of horizontal lines which we'll denote by $\leftrightarrow$, from this construction we see that if $\xi_0 \leq \xi_0'$ in the sense that $\xi_0(x) = 1 \Rightarrow \xi_0'(x) = 1$ for all x, then determining ξ_t from ξ_0 and ξ_t' from ξ_0', for $t > 0$ we find that $\xi_t \leq \xi_t'$, since adding more infectious sites at time 0 can only increase the number of active paths that are used by the infection. The property that $\xi_0 \leq \xi_0' \Rightarrow \xi_t \leq \xi_t'$ for $t > 0$ pointwise on realizations of the substructure is called *monotonicity*, and a process is *monotone*

if monotonicity holds. Furthermore, it is an example of *stochastic domination*: the processes $(\xi_t)_{t\geq 0}$ and $(\xi'_t)_{t\geq 0}$ are *coupled*, i.e., constructed on a common probability space, in such a way that pointwise, one process dominates the other.

Since the relation $\xi_0 \leq \xi'_0 \Rightarrow \xi_t \leq \xi'_t$ holds for *any* ω in the probability space, it follows immediately, for example, that letting $|\xi_t| = \#\{x : \xi_t(x) = 1\}$ where $\#$ is the cardinality, $\mathbb{P}(|\xi_t| > 0 \ \forall t > 0 \mid \xi_0 = \xi)$ is non-decreasing in the initial configuration ξ. Thus we see that a probabilistic fact about the process can be easily deduced from the geometrical properties of individual realizations of the process in spacetime. Another useful consequence of monotonicity is the following result; for the sake of brevity we refer to Theorem 2.3 in Chapter III of [32] for a few details.

Theorem 1.2.1. *Define* $\overline{\xi}$ *by* $\overline{\xi}(x) = 1$ *for all* x. *If* $(\xi_t)_{t\geq 0}$ *satisfies* $\xi_0 = \overline{\xi}$ *then the distribution of* ξ_t *converges weakly to an invariant distribution* ν, *called the* upper invariant measure.

The upper invariant measure ν *has the property that for each finite* $V_0 \subset V$ *and any invariant distribution* μ,

$$\nu(\{\xi : \xi(x) = 1, \ \forall x \in V_0\}) \geq \mu(\{\xi : \xi(x) = 1, \ \forall x \in V_0\}) \tag{1.2.1}$$

and is the unique invariant distribution with this property.

Proof. For all $\xi \in \{0,1\}^V$ we have $\xi \leq \overline{\xi}$, so on any realization, for $t > 0$, $\xi_t \leq \xi_0$. Placing $\overline{\xi}$ at time t in the graphical representation and evolving up to time $t+s$ while using monotonicity shows that ξ_s dominates ξ_{t+s}, so ξ_t is stochastically decreasing with t, in the sense that for each finite set of sites $V_0 \subset V$,

$$\mathbb{P}(\xi_t(x) = 1 \text{ for all } x \in V_0)$$

is non-increasing with t, and thus converges to a limit. Using convergence of these events and an inclusion-exclusion argument, we can deduce convergence of all finite-dimensional distributions (FDDs), i.e., events of the form $\mathbb{P}(\xi_t(x) = \phi(x) \text{ for all } x \in V_0)$, where $\phi : V_0 \to \{0,1\}$ is any function. We show the inclusion-exclusion argument by an example: if V_0 consists of the four sites $\{x_1, .., x_4\}$, and $(\phi(x_i))_{i=1}^4 = (1,0,1,0)$,

then

$$\begin{aligned}\mathbb{P}(\xi_t(x_i) = \phi(x_i),\ i = 1,2,3,4) &= \mathbb{P}(\xi_t(x_1) = \xi_t(x_3) = 1)\\ &- \mathbb{P}(\xi_t(x_1) = \xi_t(x_2) = \xi_t(x_3) = 1)\\ &- \mathbb{P}(\xi_t(x_1) = \xi_t(x_3) = \xi_t(x_4) = 1)\\ &+ \mathbb{P}(\xi_t(x_i) = 1,\ i = 1,2,3,4)\end{aligned}$$

Since a measure on $\{0,1\}^V$ is determined by its FDDs, we conclude existence of a limiting measure ν to which the distribution of ξ_t converges. From a certain continuity property of the infinitesimal generator of the process that is discussed in [32], we can deduce that ν is invariant.

If (1.2.1) holds and μ is such that it holds with equality for all V_0, from the above inclusion-exclusion argument we conclude that μ and ν have the same FDDs and thus coincide, which proves uniqueness. To see that (1.2.1) holds, start from distribution μ in the graphical construction and note that $\overline{\xi}$ dominates μ. By monotonicity, the distribution of ξ_t with $\xi_0 = \overline{\xi}$ dominates the distribution of the process at time t started from distribution μ, for all $t > 0$. Then, note that μ is invariant, and take the limit as $t \to \infty$ to obtain (1.2.1). □

For the contact process, we can prove a somewhat stronger property than monotonicity. Defining $\xi \vee \xi'$ by $(\xi \vee \xi')(x) = 1$ iff at least one of $\xi(x)$ and $\xi'(x)$ is equal to 1, then letting $\xi_0'' = \xi_0 \vee \xi_0'$ we find that $\xi_t'' = \xi_t \vee \xi_t'$ for $t > 0$, the reason being that the set $\{y \in V : \xi_0''(y) = 1\}$ is the union of the corresponding sets for ξ_0 and ξ_0', and $\xi_t''(x) = 1$ if and only if there is an active path from $(y,0)$ to (x,t) for some y in this set. This property that $\xi_0'' = \xi_0 \vee \xi_0' \Rightarrow \xi_t'' = \xi_t \vee \xi_t'$ for $t > 0$ is called *additivity*, and was first studied in detail by Harris in [26].

We note one more property that follows from the use of active paths. Starting at a point (x,t) and tracing active paths backwards in time from (x,t) in the same way as forwards, i.e., vertically until we encounter a $\times$ label and horizontally along $\leftrightarrow$ labels, there is an active path backwards in time from (x,t) down to (y,s) for $s < t$ if and only if there is an active path forwards in time from (y,s) up to (x,t). Moreover, both the U_x and the U_{xy} are time inversion invariant, that is, their distribution does not change if the direction of time is reversed. Thus to decide whether (x,t) is infected or

more generally, whether a set $A = \{(x_1, t), ..., (x_k, t)\}$ has at least one infected point, it suffices to run a contact process starting from the infected sites A from time t down to time 0, and check if any active paths end up at points $(y, 0)$ that are initially infected. This gives the following useful *duality relation*, where $\mathbf{1}(A)$ is the indicator function of a set $A \subset V$ with $\mathbf{1}(A)(x) = 1$ if and only if $x \in A$, and $= 0$ otherwise:

$$\mathbb{P}(\xi_t(x) = 1 \text{ for some } x \in A \mid \xi_0 = \mathbf{1}(B)) = \mathbb{P}(\xi_t(x) = 1 \text{ for some } x \in B \mid \xi_0 = \mathbf{1}(A))$$

The following is a useful corollary of this fact.

Theorem 1.2.2. *For any $x \in V$, define the survival probability*

$$\sigma(x) = \mathbb{P}(|\xi_t| > 0 \ \forall t > 0 \mid \xi_0 = \mathbf{1}(x))$$

Then $\sigma(x) = \nu(\{\xi : \xi(x) = 1\})$, where ν is the upper invariant measure introduced in Theorem 1.2.1.

Proof. Let $A = V$ and $B = \{x\}$, then let $t \to \infty$ in the duality relation while noting that $\{|\xi_t| > 0 \ \forall t > 0\} = \bigcap_{t>0}\{|\xi_t| > 0\}$. □

For $x, y \in V$ define

$$\sigma(x, y) = \mathbb{P}(\xi_t(y) = 1 \text{ for some } t > 0 \mid \xi_0 = \mathbf{1}(x))$$

Since G is by assumption connected, if $\lambda > 0$ then $\sigma(x, y) > 0$, and using the strong Markov property and monotonicity, $\sigma(x) \geq \sigma(x, y)\sigma(y)$ so it follows that either $\sigma(x) = 0$ for all $x \in V$ or else $\sigma(x) > 0$ for all $x \in V$. This shows the following two notions of survival are equivalent for the contact process:

- survival of the infection starting from a single infectious site, and
- existence of a non-trivial endemic state.

Lastly we show that if the contact process has a phase transition, i.e., if $\sigma(x) > 0$ for some value of λ, then the phase transition is sharp. A priori we do not know whether there is a phase transition at all; we show this in the case of an infinite graph, i.e., when $|V| = \infty$, after we have proved the corresponding result for oriented percolation in the next subsection.

Theorem 1.2.3. *For the contact process there is a well-defined* critical value λ_c *such that* $\sigma(x) = 0$ *for* $\lambda < \lambda_c$ *and all* $x \in V$*, and* $\sigma(x) > 0$ *for* $\lambda > \lambda_c$ *and all* $x \in V$*. A priori* λ_c *may be equal to* 0 *or* ∞*.*

Proof. It suffices to show that $\sigma(x)$ is non-decreasing in λ. This is achieved by making a simultaneous construction of the process for distinct values $\lambda < \lambda'$ of the transmission parameter. First construct the process for the given value λ. Then, add independent Poisson processes $\{U'_{xy} : xy \in E\}$ with rate $\lambda' - \lambda$, to be used only by the second process. Since it has access to the same and possibly more transmission opportunities, the second process dominates the first. By a familiar property, for each $xy \in E$ the union $U_{xy} \cup U'_{xy}$ is a Poisson process with intensity $\lambda + (\lambda' - \lambda) = \lambda'$, so the second process has transmission rate λ'. □

As shown so far for the contact process, the ability to turn geometry into statements about probabilities makes the graphical representation a powerful tool in the study of growth models, and we will continue to see this in later sections and chapters.

1.2.2 Comparison to Oriented Percolation

There is a useful discrete time process that in some sense qualifies as a growth model, and has the advantage that we can estimate directly the probability of survival. It is called oriented percolation, and is defined in two dimensions as follows. We consider the case of site percolation; there is a related definition for so called bond percolation. For a comprehensive article on site percolation see [8].

Let L denote the set of sites $\{(m, n) \in \mathbb{Z}^2 : n \geq 0, m + n \text{ is even }\}$. Construct a probability space by assigning to each $(m, n) \in L$ a uniform $[0, 1]$-valued random variable $u_{m,n}$. Then, given a parameter $p \in [0, 1]$, say that site (m, n) is *open* if $u_{m,n} \leq p$ and *closed* otherwise. An *open path* from (m_0, n) to $(m_j, n + j)$ is a set of sites $(m_0, n), (m_1, n + 1), ..., (m_j, n + j)$ with $(m_i, n + i)$ open, and $m_{i+1} = m_i \pm 1$, for $i = 1, ..., j$, and say $(m, n) \to (k, \ell)$ if there is an open path from (m, n) to (k, ℓ). The *open cluster* $C(m, n)$ of a site (m, n) is the set of sites (k, ℓ) such that $(m, n) \to (k, \ell)$, and $|C(m, n)|$ is the cardinality of $C(m, n)$.

Say that *percolation* occurs from (m, n) if $|C(m, n)| = \infty$, and define

$$\theta(p) = \mathbb{P}_p(\text{percolation occurs from } (0, 0))$$

where $\mathbb{P}_p$ denotes the law of oriented percolation with parameter p. Clearly $\theta(0) = 0$. Given the values of $\{u_{m,n} : (m, n) \in L\}$, the set of open sites is non-decreasing with respect to p. It follows that $\theta(p)$ is non-decreasing with respect to p. Letting $p_c = \sup\{p \in [0, 1] : \theta(p) = 0\}$, $\theta(p) = 0$ if $p < p_c$ and is > 0 if $p > p_c$. Without too much effort we can show the value of p_c is non-trivial. The lower bound is easy, and the proof of the upper bound follows Section 10 of [8].

Theorem 1.2.4. *For p_c as defined above, $1/2 \leq p_c \leq 80/81$.*

Proof. The number of different (upward directed) paths from $(0, 0)$ to $(\cdot, n)$, whether open or not, is 2^n, while the probability that any given path is open is equal to p^n. By a simple union bound, the probability that there exists an open path from $(0, 0)$ to (m, n) for some m is at most $2^n p^n = (2p)^n$, which goes to 0 as $n \to \infty$, if $p < 1/2$.

To see that $p_c < 1$, suppose that $|C(0, 0)| < \infty$, and denote by $\mathcal{C}$ the thickened set obtained from $C(0, 0)$ by filling in a diamond shape with side length $\sqrt{2}$, centered on each $(m, n) \in C(0, 0)$, and denote by $\ell(C(0, 0))$ the outer boundary of $\mathcal{C}$, which if we give it a clockwise orientation is a path from $(-1, 0)$ over and around $\mathcal{C}$ to $(1, 0)$, that keeps $\mathcal{C}$ to its immediate right. Since ℓ consists of directed edges between points in the set $\tilde{L} = \{(m, n) \in \mathbb{Z}^2 : n \geq 0, m + n \text{ is odd }\}$, we can count the number of edges in ℓ, denoted $n(\ell)$, as well as the number of edges of each of the four orientations $\nearrow, \searrow, \swarrow, \nwarrow$ that we denote $\nearrow(\ell)$, etc. Based on its start and endpoints, $\nearrow(\ell) = \swarrow(\ell) + 1$ and $\searrow(\ell) = \nwarrow(\ell) + 1$.

Now let ℓ be any directed path of the type above, i.e., not intersecting itself, and going from $(-1, 0)$ to $(1, 0)$ with edges between vertices in $\tilde{L}$. Say that ℓ is $\nearrow$-*open* if, for each $\nearrow$ edge intersecting the point $(m-1/2, n+1/2)$, the site $(m-1, n+1) \in L$ is closed, and similarly for $\searrow$-open. Note that distinct $\nearrow$ edges correspond to distinct sites, and that distinct $\searrow$ edges correspond to distinct sites. If $\ell = \ell(C(0, 0))$ then in particular, ℓ is $\nearrow$-open and $\searrow$-open, so

$$\mathbb{P}(\ell(C(0, 0)) = \ell) \leq \mathbb{P}(\ell \text{ is } \nearrow\text{-open and } \searrow - \text{ open })$$

Since any ℓ has at most 3 options for the direction of each edge, for any n the set $\{\ell : n(\ell) = n\}$ has cardinality at most 3^n. Moreover, since $n(\ell) = \nearrow(\ell) + \searrow(\ell) + \swarrow(\ell) + \nwarrow(\ell) = 2\nearrow(\ell) + 2\searrow(\ell) - 2$, it follows that $\max(\nearrow(\ell), \searrow(\ell)) \geq (n+2)/4$ so the probability that a given ℓ is both $\nearrow$- and $\searrow$-open is at most $(1-p)^{(n+2)/4}$. Summing on n and noting that $|C(0,0)| < \infty$ is equivalent to $\ell(C(0,0)) = \ell$ for some ℓ,

$$\mathbb{P}(|C(0,0)| < \infty) \leq \sum_{n \geq 1} 3^n (1-p)^{(n+2)/4} \leq \sum_{n \geq 1} (3(1-p)^{1/4})^n$$

which converges for $(1-p)^{1/4} < 1/3$ or $p > 80/81$. To make it < 1, make the same calculation for the cluster $C(A)$ of all sites $(j,k) \in L$ such that $(m,n) \to (j,k)$ for some $(m,n) \in A$, where $A = \{(m,0) \in L : m \in [-N,N]\}$ and N is even, and ℓ now needs to travel from $(-N-1,0)$ to $(N+1,0)$. This forces $n(\ell(C(A))) \geq 2N+2$, so the above series starts at $n = 2N+2$, and adds up to < 1 provided N is large enough. It then suffices to note that for each N, with positive probability, $(0,0) \to (m,N)$ for all $m \in [-N,N]$ such that $(m,N) \in L$, so $\mathbb{P}(|C(0,0)| = \infty) \geq \mathbb{P}((0,0) \to (m,N)$ for all $m \in [-N,N])\mathbb{P}(|C(A)| = \infty)$ which is > 0 provided $\mathbb{P}(|C(A)| = \infty) > 0$. □

Now, consider the contact process on the graph with $V = \mathbb{Z}$ and $E = \{xy : |x - y| = 1\}$, which we call the contact process on $\mathbb{Z}$ with *nearest-neighbour* interactions. Since the spacetime set for this process is two-dimensional and infection paths are directed upwards in time, the model resembles a continuous-time version of oriented percolation. By making an explicit comparison we can establish the existence of a phase transition, in other words, show that $\sigma(x) > 0$ if λ is large enough. This procedure is known as a *block construction* and can be found, for example, in [10]. Since the nearest-neighbour model on $\mathbb{Z}$ is translation-invariant, we denote by $\sigma(\lambda)$ the common value of $\sigma(x)$, for a given value of λ.

Theorem 1.2.5. *$\lambda_c < \infty$ for the contact process on $\mathbb{Z}$ with nearest-neighbour interactions.*

Proof. The proof given here uses a simplified version of the construction from [10]. Given $\epsilon > 0$, we want to discretize the spacetime set $S = \mathbb{Z} \times \mathbb{R}_+$ into the checkerboard pattern of congruent rectangles $\{R_{m,n} : (m,n) \in L\}$ given by

$$R_{m,n} = [0,J] \times [0,T] + (mJ, nT)$$

for some J, T so that $R_{m,n}$ just touches $R_{m\pm1,n+1}$ and does not intersect $R_{m\pm2,n}$, and to define an event $A_{m,n}$ on each rectangle, with the property that:

- $\{A_{m,n} : (m,n) \in L\}$ are independent,
- for each $(m,n) \in L$, $\mathbb{P}(A_{m,n}) = p \geq 1 - \epsilon$ provided λ is large enough, and
- if $\xi_0(0) = 1$ and $(0,0) \to (m,n)$ then either $\xi_{nT}(mJ) = 1$ or $\xi_{nT}((m+1)J) = 1$, or both,

where the meaning of $(0,0) \to (m,n)$ is the same as before, if we think of each site $(m,n) \in L$ as open if $A_{m,n}$ occurs. Defining $C(0,0)$ as before, if $|C(0,0)| = \infty$ and $\xi_0(0) = 1$ then $|\xi_t| > 0 \ \forall t > 0$, so $\sigma(\lambda) \geq \theta(p)$ for such values of λ. Since the set of open sites has the same distribution as in oriented site percolation with parameter p, taking λ large enough that $\epsilon < 1/81$ and using Theorem 1.2.4, $\sigma(\lambda) > 0$ as desired.

The way we have written it, the desired event $A_{m,n}$ is clearly that there are active paths for the contact process lying *inside* the rectangle $R_{m,n}$ from (mJ, nT) to both $(mJ, (n+1)T)$ and $((m+1)J, (n+1)T)$, and from $((m+1)J, nT)$ to both $(mJ, (n+1)T)$ and $((m+1)J, (n+1)T)$. The condition of lying inside the rectangle ensures the independence of the $A_{m,n}$, using the corresponding spatial independence property of Poisson processes.

We may as well assume $J = 1$, in which case, fixing $T > 0$, $A_{m,n}$ is equivalent to the following event: for the contact process on the two-site graph with $V = \{0,1\}$ and $E = \{01\}$, if either 0 or 1 is infectious at time 0 then both 0 and 1 are infectious at time T. To have this it is sufficient that there is a $\leftrightarrow$ label on 01 and no $\times$ labels on either 0 or 1, on the time interval $[0,T]$. Given $\epsilon > 0$, by taking $T > 0$ small enough we can make the probability of a $\times$ label at most $\epsilon/2$, and then by taking λ large enough we can make the probability of a $\leftrightarrow$ arrow at least $1 - \epsilon/2$, which makes $\mathbb{P}(A_{m,n}) \geq 1 - \epsilon$. Since ϵ is arbitrary, the proof is complete. □

Although we don't discuss it here, a similar argument shows that $\theta(p) > 0$ for $p < 1$ close enough to 1, for a similar model called a *k-dependent* percolation model; see [15]. For a k-dependent model, individual sites are open with probability p, and the events $\{(m_i, n_i) \text{ open}\}_{i \in I}$, where I is a finite index set, are independent if $\|(m_i, n_i) - (m_j, n_j)\|_\infty > k$ for all $i, j \in I$ where $\|(m,n)\|_\infty = \max(|m|, |n|)$. The

idea of the proof is that when measuring the probability that ℓ is open, to choose the edges in ℓ more sparsely to ensure they are independent. The motivation for studying k-dependent models is that they arise naturally in certain constructions. In [15] a 1-dependent percolation model arises in a similar construction to the one we made above. Contrary to the situation above, we only have approximate control over the behaviour of active paths, and $R_{m,n}$ now has a non-trivial overlap with $R_{m\pm1,n+1}$ which is needed to ensure that active paths in adjacent rectangles will cross so that they may be concatenated.

1.2.3 Comparison to a Branching Process

As a counterpart to oriented site percolation, which has a definite spatial structure, we now discuss a branching process, which in a sense has minimal spatial structure. In a branching process, particles die and reproduce independently of one another with certain prescribed rates. The reference books [27] and [1] give a good introduction to the theory of branching processes. Here we pick and choose some basic facts about branching processes that will help us to study the contact process, including proofs wherever possible.

In our case, it is useful to study the simplest continuous-time branching process in which each particle dies at rate 1 and produces offspring at rate λ. Letting Z_t denote the number of particles at time t, by the independence assumption we find

- Z_t decreases by 1 at rate Z_t, and
- Z_t increases by 1 at rate λZ_t

The first simple lemma justifies our study of Z_t. Here we use the "random set" notation for the contact process. The intuition is that the contact process can be viewed as a sort of spatial branching process in which births onto already occupied sites are suppressed.

Lemma 1.2.6. *Let $(A_t)_{t\geq0}$ denote the contact process with parameter λ on a graph $G = (V, E)$ in which $\deg x \leq M$ for each $x \in V$, and let $I_t = |A_t|$. Then, I_t is dominated by the branching process Z_t with $Z_0 = I_0$ in which particles die at rate 1 and produce offspring at rate λM.*

Proof. For $A \subset V$ let $E(A) = \{xy \in E : x \in A, y \in A^c\}$, then

- I_t decreases by 1 at rate I_t, and
- I_t increases by 1 at rate $\lambda|E(A)|$

and by assumption, $|E(A)| \leq M|A| = MI_t$. If $Z_t = I_t = x$ then Z_t and I_t both decrease by 1 at rate x, while Z_t increases by 1 at rate $\lambda Mx \geq \lambda|E(A)|$. Given the graphical construction of A_t, we can construct Z_t using the Poisson points that determine I_t, plus some additional independent Poisson processes with rate 1 and rate λ, in such a way that $Z_t \geq I_t$ for $t \geq 0$. Since the idea of the coupling is clear, its details are left to the interested reader. □

Let $m(t) = \mathbb{E}Z_t$. By first conditioning on the value of Z_t and then integrating, we see that $m'(t) = (\lambda - 1)m(t)$, and since $m(0) = 1$ we have $m(t) = e^{(\lambda-1)t}$ for $t > 0$. Since Z_t is integer-valued, $\mathbb{P}(Z_t > 0) \leq \mathbb{E}Z_t = m(t)$, which tends to 0 exponentially fast if $\lambda < 1$. This has the following immediate consequence for the contact process that first appeared in [24], and is also mentioned in [32].

Corollary 1.2.7. *For the contact process on a graph $G = (V, E)$ with $\deg x \leq M$ for each $x \in V$, $\lambda_c \geq 1/M$.*

To get a comparison in the other direction we need to consider the contact process on a graph in which infectious sites behave more or less independently. Of course this is not possible at full occupancy, but if the interaction neighbourhood is large and infectious sites are sparsely scattered through the population this may be a reasonable approximation. For the sake of example, we will do this for the contact process on the complete graph K_N with $|V| = N$ and $E = \{xy : x, y \in V,\ x \neq y\}$. For this model it is useful to replace λ with the rescaled value λ/N, so that a single infectious site in an otherwise healthy population spreads the infection to its neighbours at the total rate $\lambda(N-1)/N \approx \lambda$.

First we determine exactly for what values of λ it is possible for Z_t to survive. Intuition suggests that if $\lambda > 1$ then survival is possible, since then the birth rate exceeds the death rate, and this can be confirmed as follows. Let $\rho(t) = \mathbb{P}(Z_t = 0 \mid Z_0 = 1)$ denote the probability of extinction at or before time t, then since $\rho(t)$ is non-decreasing and $\rho(t) \leq 1$ for $t \geq 0$, the value $\rho^* = \lim_{t\to\infty} \rho(t)$, the probability of eventual extinction, is well-defined.

In a short time h, $Z_h = 0, 1$ or 2 with probability $h + o(h)$, $1 - (1 + \lambda)h + o(h)$ and $\lambda h + o(h)$ respectively, and each particle from time h up to time t evolves independently, so we calculate

$$\begin{aligned}\rho(t) &= h + (1 - (1 + \lambda)h)\rho(t - h) + \lambda h\rho(t - h)^2 + o(h) \\ &= \rho(t - h) + [1 - (1 + \lambda)\rho(t - h) + \lambda\rho(t - h)^2]h + o(h)\end{aligned}$$

and so $\rho'(t) = Q(\rho(t))$ where $Q(x) = 1 - (1 + \lambda)x + \lambda x^2$ is convex and quadratic. Since $\rho(0) = 0$, $Q(0) = 1$ and $Q(1) = Q(1/\lambda) = 0$, $\rho(t)$ increases towards the lesser root of $Q(x)$, so $\rho^* = \min(1, 1/\lambda)$. We have shown the following result.

Theorem 1.2.8. *For the branching process Z_t as defined above let $\theta(\lambda) = \mathbb{P}(Z_t > 0\ \forall t > 0)$ denote the survival probability, as a function of λ. Then*

$$\theta(\lambda) = \begin{cases} 0 & \text{for } \lambda \le 1 \\ 1 - \frac{1}{\lambda} & \text{for } \lambda > 1 \end{cases}$$

In other words, survival is possible exactly when the birth rate exceeds the death rate. The next question is, if Z_t survives, how quickly does it grow? Intuition suggests it grows approximately like $m(t)$. The following result is proved in Chapter II of [27] in discrete time, i.e., for the sequence $(Z_n)_{n=1,2,\ldots}$, under the assumption $\operatorname{Var} Z_1 < \infty$.

Theorem 1.2.9. *For Z_t and $m(t)$ as defined above, if $\lambda > 1$ then as $t \to \infty$, $\tilde{Z}_t = Z_t/m(t)$ converges almost surely to a random variable W satisfying $\mathbb{P}(W > 0) = \mathbb{P}(Z_t > 0\ \forall t > 0) > 0$.*

The proof uses the fact that $\tilde{Z}_t$ is a *martingale*, which means that $\mathbb{E}(\tilde{Z}_{t+s} \mid \tilde{Z}_t) = \tilde{Z}_t$ for $t, s > 0$. Under fairly mild assumptions, martingales converge; see for example Chapter 4 in [13] for some martingale convergence theorems. A useful corollary is the following.

Corollary 1.2.10. *For $\lambda > 1$, let Z_t and W be as in Theorem 1.2.9. For fixed $C > 1/(1 - \lambda)$ and $\epsilon > 0$, $\liminf_{N\to\infty} \mathbb{P}(Z_{C \log N} \ge \epsilon N \mid Z_0 = 1) \ge \mathbb{P}(W > 0) > 0$.*

Proof. In the notation of Theorem 1.2.9, $Z_t \ge \epsilon N$ if and only if $\tilde{Z}_t \ge \epsilon N e^{-(\lambda-1)t}$. Setting $t = C \log N$, $\epsilon N e^{-(\lambda-1)t} = \epsilon N^{1-(\lambda-1)C}$, which $\to 0$ as $N \to \infty$ if $C > 1/(\lambda-1)$. Thus, for each $\delta > 0$, if N is large enough then $\mathbb{P}(Z_{C \log N} \ge \epsilon N) \ge \mathbb{P}(\tilde{Z}_{C \log N} \ge \delta)$, which implies that $\liminf_{N\to\infty} \mathbb{P}(Z_{C \log N} \ge \epsilon N) \ge \mathbb{P}(W \ge \delta)$. By continuity of measure, $\liminf_{N\to\infty} \mathbb{P}(Z_{C \log N} \ge \epsilon N) \ge \mathbb{P}(W > 0)$, as desired. □

Using Theorem 1.2.9 and the fact that $\mathbb{P}(Z_t > 0) \leq m(t)$ we can show the following result. This is analogous to a result from Chapter 2, proved here for the contact process.

Theorem 1.2.11. *Let $(A_t)_{t\geq 0}$ denote the contact process on K_N with transmission rate λ/N per edge, and let $I_t = |A_t|$. Then,*

- *if $\lambda < 1$, there is $C > 0$ not depending on N so that from any initial configuration, $I_{C\log N} = 0$ with probability tending to 1 as $N \to \infty$, and*
- *if $\lambda > 1$, starting from a single infectious site, there are $\epsilon, p, C, N_0 > 0$ not depending on N so that $\mathbb{P}(I_t \geq \epsilon N$ for some $t \leq C\log N) \geq p > 0$ for all $N \geq N_0$.*

Proof. In the case $\lambda < 1$, by monotonicity it is enough to consider the initial configuration with $I_0 = N$. By Lemma 1.2.6, I_t is dominated by the branching process Z_t with $Z_0 = N$ that decreases by 1 at rate Z_t and increases by 1 at rate λ. In this case $m(t) = Ne^{(\lambda-1)t}$ since $m(0) = N$, so $\mathbb{P}(Z_{C\log N} > 0) \leq m(C\log N) = Ne^{(\lambda-1)C\log N} = N^{1+(\lambda-1)C}$ which tends to 0 as $N \to \infty$ provided $C > 1/(1-\lambda)$.

For $\lambda > 1$, we need to lowerbound I_t by a branching process. We note that

- I_t decreases by 1 at rate I_t and
- I_t increases by 1 at rate $I_t(N - I_t)\lambda/N$

and if $\lambda > 1$ then for $\epsilon > 0$, for $x \leq \epsilon N$, $x(N-x)\lambda/N \geq \lambda(1-\epsilon)x$. Letting Z_t denote the branching process with $Z_0 = 1$ that decreases by 1 at rate Z_t and increases by 1 at rate $\lambda(1-\epsilon)Z_t$, since both are Markov chains on $\{0, 1, ...\}$ and their rates satisfy the desired inequality, a simple coupling argument shows that I_t dominates Z_t so long as $Z_t \leq \epsilon N$. If ϵ is such that $\lambda(1-\epsilon) > 1$, then Corollary 1.2.10 applies to show that for $C > 1/(\lambda - 1)$, some $p > 0$ and N large enough, $\mathbb{P}(Z_{C\log N} \geq \epsilon N) \geq p > 0$, and the result follows by comparison. □

1.2.4 Comparison to Mean-Field Equations

Recall the contact process on the complete graph K_N with $V = \{1, ..., N\}$ and $E = \{xy : x, y \in V,\ x \neq y\}$ introduced in the previous section. Starting from a single infectious site and picturing the process as a random subset $A_t \subset V$ of infectious sites, we find for the cardinality $I_t = |A_t|$ that

- I_t decreases by 1 at rate I_t and
- I_t increases by 1 at rate $I_t(N - I_t)\lambda/N$

In particular, I_t is a Markov chain on $\{0, 1, ...\}$. Considering the rescaled value $i_t = I_t/N$ we find that

- i_t decreases by $1/N$ at rate Ni_t and
- i_t increases by $1/N$ at rate $i_t(1 - i_t)\lambda N$

In a small time increment $h > 0$, $\mathbb{E}(i_{t+h} - i_t \mid i_t) = [-i_t + i_t(1 - i_t)\lambda]h + o(h)$, while $\mathbb{E}((i_{t+h} - i_t)^2) = O(h^2)$, which implies $\mathrm{Var}(i_{t+h} - i_t) = O(h^2)$, so we should expect sample paths of i_t to approach solutions to the differential equation

$$i' = -i + \lambda i(1 - i)$$

which we justifiably call the *mean-field equation* or MFE for the contact process on K_N. Although we do not prove it here, using the techniques of Chapter 2 we can show the following approximation theorem.

Theorem 1.2.12. *Let i_0 be fixed, with i_t the above process, and let $i(t)$ be a solution of the mean-field equation with $i(0) = i_0$. For each $\epsilon, T > 0$, there is $\gamma > 0$ so that $\mathbb{P}(|i_t - i(t)| \le \epsilon$ for $0 \le t \le T) \ge 1 - e^{-\gamma N}$.*

Even without this result, we can get some useful information for i_t from the MFE:

- if $\lambda < 1$ the MFE has the unique equilibrium $i = 0$ which is stable on $[0, 1]$, and in Theorem 1.2.11 we showed that $i_t \to 0$ within time $C \log N$ with probability tending to 1 as $N \to \infty$, so i_t and $i(t)$ are at least qualitatively similar in that case.
- If $\lambda > 1$ the MFE has the unique non-trivial equilibrium $i^* = 1 - 1/\lambda$. In this case we can show that starting from $i_0 \ge i^*$ the infection persists for a long time, in the sense that for fixed small enough $\epsilon > 0$, there is $\gamma > 0$ so that

$$\mathbb{P}(i_t \ge i^* - \epsilon \text{ for } 0 \le t \le e^{\gamma N} \mid i^* \ge i_0) \ge 1 - e^{-\gamma N}$$

We give just a sketch of the proof of this last fact; for a rigorous treatment of similar results see Chapters 2. Fix $\epsilon > 0$ such that $i^* - \epsilon > 0$ and $0 < \alpha < 1$, then while

$i_t \in [i^* - \epsilon, i^* - \alpha\epsilon]$, i_t dominates a random walk x_t that decreases by $1/N$ at rate $N(i^* - \alpha\epsilon)$ and increases by $1/N$ at rate $N\lambda(i^* - \epsilon)(1 - (i^* - \epsilon))$. If ϵ is small enough and α close enough to 1 then x_t has positive drift, since $\lambda > 1$ and

$$\begin{aligned}\lambda(i^* - \epsilon)(1 - (i^* - \epsilon)) - (i^* - \alpha\epsilon) &= [\lambda i^*(1 - i^*) - i^*] + (2\lambda i^* - \lambda + \alpha)\epsilon - \lambda\epsilon^2 \\ &= (2\lambda(1 - 1/\lambda) - \lambda + \alpha)\epsilon - \lambda\epsilon^2 \\ &= (\lambda + \alpha - 2)\epsilon - \lambda\epsilon^2\end{aligned}$$

using the definition of i^*. If i_t enters the interval $[i^* - \epsilon, i^* - \alpha\epsilon]$, one can show the probability it exits at the lower end is at most $e^{-\gamma N}$ for some $\gamma > 0$; intuitively this is because it has to make order of N steps down, each of which has probability at most $p < 1/2$ for some p not depending on N. A rigorous proof uses domination of i_t and the formula for the hitting probability.

To push this to the desired result we need to control the amount of time taken at each excursion. One way to do this is to break $[i^* - \epsilon, i^* - \alpha\epsilon]$ into two subintervals U_1, U_2 of equal width. Letting $c_1 < c_2 < c_3$ be the endpoints of U_1 and U_2, if i_t reaches c_2 then it reaches c_1 before c_3 with probability at most $e^{-\gamma N}$, and the time required to reach one of c_1 or c_3 from c_2 is of at least constant order. Iterating this observation and taking a union bound, after $e^{\gamma N/2}$ visits to c_2, which takes order of $e^{\gamma N/2}$ amount of time, i_t reaches c_1 with probability at most $e^{\gamma N/2}e^{-\gamma N} = e^{-\gamma N/2}$. It then suffices to show the lower bound on the time required at each attempt holds with probability $\geq 1 - e^{-\gamma N}$ for some $\gamma > 0$.

1.3 Statement of Results

Here we introduce the models considered in each of the main chapters, and give a brief overview of the main results. Three of the chapters appear or will appear as published papers: Chapter 2 as [20], Chapter 4 as [19] and Chapter 5 as [17]. We also note the paper [18] that concerns a stochastic growth model but is not included here.

1.3.1 Social contact processes and the partner model

In Chapter 2 we consider a stochastic model of infection spread on a graph (V, E) incorporating some form of social dynamics. In other words, we have a process $E_t \subseteq E$

that describes the set of *active* edges as a function of time, and the model behaves like the contact process except that transmission can only occur along active edges.

In our case, the edge process E_t is such that edges become active at some rate r_+, and inactive at some other rate r_-, independently of other edges and not depending on the state (healthy or infectious) of each site. To make things interesting we add the restriction of *monogamy*, which means that $xy \in E \setminus E_t$ can only become active if $xz, yz \notin E_t$ for every $z \in V$. Altogether, this gives the following transitions that determine (V_t, E_t), the set of infectious sites and active edges as a function of time:

- if $x \notin V_t$, $y \in V_t$ and $xy \in E_t$ then $x \in V_t$ at rate λ,
- if $x \in V_t$ then $x \notin V_t$ at rate 1,
- if $xy \notin E_t$ and $xz, yz \notin E_t$ for all $z \in V$ then $xy \in E_t$ at rate r_+/N,
- if $xy \in E_t$ then $xy \notin E_t$ at rate r_-

In this model we think of connected pairs as partners, so we call it the *partner model*.

For simplicity, we study the model on the sequence of complete graphs K_N on N vertices, where N will tend to ∞; this is a reasonable model for, say, the spread of a sexually transmitted infection through a homogeneous population of monogamous homosexual individuals in a big city. We rescale the partner formation rate per edge to r_+/N to ensure that a given individual in a pool of entirely singles finds a partner at total rate approximately r_+. For future reference, we use interchangeably both the words healthy and susceptible, and the words unpartnered and single, to describe respectively an individual that is not infectious, and an individual that does not have a partner. Even in this simple model, as described below, there is a phase transition between extinction and spread of the infection.

For the partner model we are mostly concerned not with the exact values of V_t and E_t but with the total number of susceptible and infectious singles S_t and I_t and the total number of partnered pairs SS_t, SI_t, II_t of the three possible types; as shown in Section 2.5, for each N, $(S_t, I_t, SS_t, SI_t, II_t)$ is a continuous time Markov chain. In general it will be more convenient to work with the rescaled quantities $s_t = S_t/N$, $i_t = I_t/N$, $ss_t = SS_t/N$, $si_t = SI_t/N$, and $ii_t = II_t/N$.

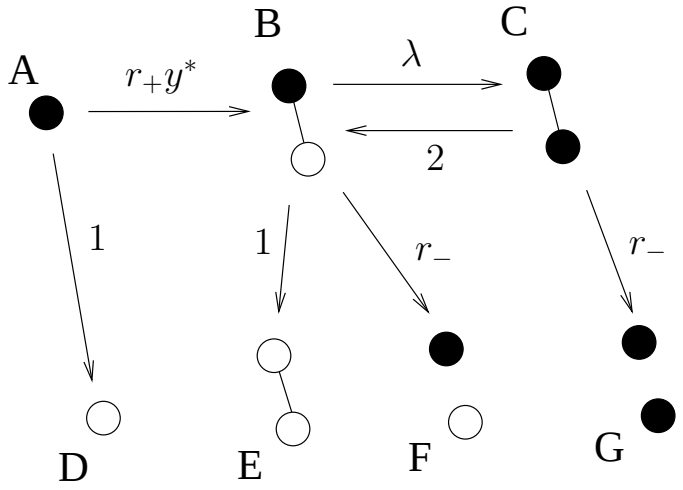

Figure 1.2: Markov Chain used to compute R_0, with transition rates indicated; infectious sites in black

Starting from any configuration, after a short time the proportion of singles $y_t := s_t + i_t$ approaches and remains close to a certain fixed value $y^* \in (0, 1)$ that we can compute and that does not depend on N; a proof of this is given in Section 2.6, and a heuristic argument in Section 2.3. Setting $\alpha = r_+/r_-$, we find

$$y^* = 1/(2\alpha)[-1 + \sqrt{1 + 4\alpha}] \tag{1.3.1}$$

To decide whether the infection can spread we start with $V_0 = \{x\}$ for some $x \in V$ with x single and $y_t \approx y^*$, and keep track of x until the first moment when x either

- recovers without finding a partner, or
- if it finds a partner before recovering, breaks up from that partnership.

This leads to the Markov chain shown in Figure 1.2. Define the *basic reproduction number*

$$R_0 = \mathbb{P}(A \to F) + 2\mathbb{P}(A \to G) \tag{1.3.2}$$

which is the expected number of infectious singles upon absorption of the above Markov chain, starting from state A. As intuition suggests, and Theorem 2.2.2 confirms, the infection can spread if $R_0 > 1$, but cannot spread if $R_0 \leq 1$.

If the dynamics is in equilibrium i.e., $(s_t, i_t, ss_t, si_t, ii_t)$ hovers around a fixed value

$(s^*, i^*, ss^*, si^*, ii^*)$, then in particular the proportion of infectious singles is roughly constant. Three events affect infectious singles:

- $I \to S$, which occurs at rate $I_t = i_t N$,
- $I + I \to II$ which occurs at rate $(r_+/N)\binom{I_t}{2} \approx r_+(i_t^2/2)N$, and
- $S + I \to SI$ which occurs at rate $(r_+/N)I_t S_t = r_+ i_t s_t N$.

If a partnership is formed, then as above, we can compute the expected number of infectious singles upon breakup. Fixing $i_t = i$ for some $i \in [0, y^*]$ and $s_t + i_t = y^*$ in the above rates, define the normalizing constant $z = 1 + r_+ i/2 + r_+(y^* - i) = 1 + r_+(y^* - i/2)$ and the probabilities $p_S = 1/z$, $p_{II} = r_+ i/2z$ and $p_{SI} = r_+(y^* - i)/z$ and referring again to Figure 1.2 let

$$\Delta(i) = p_S \Delta_S + p_{II}\Delta_{II} + p_{SI}\Delta_{SI} \tag{1.3.3}$$

where $\Delta_S = -1$, $\Delta_{II} = -2 + \mathbb{P}(C \to F) + 2\mathbb{P}(C \to G)$ and $\Delta_{SI} = -1 + \mathbb{P}(B \to F) + 2\mathbb{P}(B \to G)$. The function $\Delta(i)$ tracks the expected change in the number of infectious singles, per event affecting one or more infectious singles. Thus, for an equilibrium solution we should have $\Delta(i^*) = 0$. As shown in Lemma 2.4.3, to have a solution $i^* > 0$, we need $R_0 > 1$.

As we will see in Lemma 2.4.1, for fixed r_+, r_-, R_0 is continuous and increasing in λ. Defining

$$\lambda_c = \sup\{\lambda \geq 0 : R_0 \leq 1\} \tag{1.3.4}$$

with $\sup \mathbb{R}_+ := \infty$, it follows that if $\lambda_c = \infty$ then $R_0 < 1$ for all λ, and if $\lambda_c < \infty$ then $R_0 < 1$ if $\lambda < \lambda_c$, $R_0 = 1$ if $\lambda = \lambda_c$ and $R_0 > 1$ if $\lambda > \lambda_c$. The following result gives a formula for λ_c and describes the behaviour of i^* near λ_c. In models exhibiting a phase transition one often seeks a *critical exponent* γ such that for an observable $F(\lambda)$ it holds that $F(\lambda) \sim C(\lambda - \lambda_c)^\gamma$; as we see here, in this model the critical exponent for i^* is equal to 1. The following two theorems are the main results of Chapter 2.

Theorem 1.3.1. *Let $y^*, R_0, \Delta(i)$ and λ_c be as in* (1.3.1), (1.3.2), (1.3.3) *and* (1.3.4) *and let r_+, r_- be fixed. Then, $\lambda_c < \infty \Leftrightarrow r_+ y^* > 1 \Leftrightarrow r_+ > 1 + 1/r_-$ and in this case*

$$\lambda_c = \frac{2}{r_-}\frac{2}{r_+ y^* - 1} + \frac{2}{r_-} + \frac{4}{r_+ y^* - 1} + 1 + \frac{r_-}{r_+ y^* - 1}$$

If $R_0 > 1$ there is a unique solution $i^* \in (0, y^*)$ to the equation $\Delta(i^*) = 0$ and $i^* \sim C(\lambda - \lambda_c)$ as $\lambda \downarrow \lambda_c$, for some constant $C > 0$.

As the following result implies, $R_0 > 1$ is both a necessary and a sufficient condition for spread and long-time survival of the infection, and for the existence of a unique and globally attracting endemic equilibrium.

Theorem 1.3.2. *Fix λ, r_+, r_- and let y^*, R_0 and $\Delta(i)$ be as defined in* (1.3.1), (1.3.2) *and* (1.3.3).

- *If $R_0 \leq 1$, for each $\epsilon > 0$ there are constants $C, T, \gamma > 0$ so that, from any initial configuration, $|V_T| \leq \epsilon N$ with probability $\geq 1 - Ce^{-\gamma N}$.*
- *If $R_0 < 1$, there are constants $C, T, \gamma > 0$ so that, from any initial configuration, all sites are healthy by time $T + C \log N$ with probability tending to 1 as $N \to \infty$.*
- *If $R_0 > 1$, there is a unique vector $(s^*, i^*, ss^*, si^*, ii^*)$, satisfying $i^* > 0$, $s^* + i^* = y^*$ and $\Delta(i^*) = 0$, such that*
  - *for each $\epsilon > 0$, there are constants $C, T, \gamma > 0$ so that, from any initial configuration with $|V_0| \geq \epsilon N$, with probability $\geq 1 - Ce^{-\gamma N}$, $|(s_t, i_t, ss_t, si_t, ii_t) - (s^*, i^*, ss^*, si^*, ii^*)| \leq \epsilon$ for $T \leq t \leq e^{\gamma N}$, and*
  - *there are constants $\delta, p, C, T > 0$ so that, from any initial configuration with $|V_0| > 0$, with probability $\geq p$, $|V_{T + C \log N}| \geq \delta N$.*

To obtain the value of the endemic equilibrium and the behaviour when $|V_0| \geq \epsilon N$, which we call the *macroscopic* regime, we use the *mean-field equations* which are a set of differential equations that give a good approximation to the evolution of $(s_t, i_t, ss_t, si_t, ii_t)$ when N is large. To describe the behaviour when $1 \leq |V_0| \leq \epsilon N$ for small $\epsilon > 0$, which we call the *microscopic* regime, we use comparison to a branching process; if $R_0 < 1$ we bound above and if $R_0 > 1$ we bound below.

1.3.2 The SEIS process

The SEIS model, or susceptible-exposed-infectious-susceptible model, is a model of the spread of an infection that in addition to the usual susceptible (healthy) and infectious classes includes an exposed class that is infected but not yet infectious. The classical model, called the compartment model, is deterministic and consists of a set

of three differential equations describing the evolution of the number of susceptible, exposed and infectious individuals, which for simplicity are taken to be real-valued (see [5], Chapter 2). The model has either a globally stable disease-free state or an unstable disease-free state together with a globally stable endemic state, according as the basic reproduction number for the infection is ≤ 1 or > 1; see [28] for a proof using Lyapunov functions.

Now, the classical SEIS model is deterministic and assumes that the population is well-mixed. Here we consider the SEIS model as a stochastic growth model with state space $\{0, 1, 2\}^V$, where 0 is susceptible, 1 is exposed and 2 is infectious. To distinguish it from the compartment model, we use "SEIS process" to refer to the SEIS model as a stochastic growth model. Given the infection parameter $\lambda > 0$ and incubation time $\tau \geq 0$ the model has the following transitions: at each site x,

- $2 \to 0$ at rate 1 (recovery)
- $1 \to 2$ at rate $1/\tau$ or instantaneously if $\tau = 0$ (onset)
- $0 \to 1$ at rate $\lambda n_2(x)$ (transmission)

where $n_2(x)$ is cardinality of the set $\{xy \in E : y \text{ is in state } 2\ \}$. The case $\tau = 0$ is the contact process with transmission parameter λ. The SEIS process can be constructed graphically as described in Section 1.2.1, using $\times, \star$ and $\leftrightarrow$ labels for recovery, onset and transmission; this is done in detail in Chapter 3. For the SEIS process we cannot directly use the notion of active paths to define the state at a spacetime point (x, t).

Given that $\tau = 0$ gives the contact process, it is natural to ask whether we obtain a limiting process as $\tau \to \infty$. The answer is yes, if we rescale time so that onset occurs at rate 1. We first describe the limit process, then state the sense in which the SEIS process converges to it.

The limit process has the state space $\{0, 1\}^V$ where 1 can be thought of as occupied and 0 as vacant. The process is defined using the dispersal distributions $p(x, \cdot)$, given as follows. Starting the contact process, on the given graph, with x being the only initially infectious site, $p(x, A)$ is the probability that x transmits the infection directly to each site in A, and not to any other sites, before recovering. Each occupied site x becomes vacant at rate 1 (i.e., at the ringing times of an independent Poisson

process with intensity 1), and upon vacating, with probability $p(x, A)$ occupies any vacant sites in A.

We note there is an obvious graphical representation of the limit process: at each site place a Poisson point process with intensity 1 and label $\star$, and at each occurrence of $\star$ at site x sample the dispersal distribution $p(x, \cdot)$, placing a $\rightarrow$ label from x to y for each y to which x disperses, and let the samples be independent. A similar notion of active paths can be defined as for the contact process, and using this it is easy to see the limit process is monotone, and is also monotone in λ. For the latter property, for $\lambda < \lambda'$ make a joint construction by coupling dispersal distributions in the obvious way. Thus the limit process has a critical value that we denote λ_c^∞ such that single-site survival occurs for $\lambda > \lambda_c^\infty$, and does not occur if $\lambda < \lambda_c^\infty$. The following result describes convergence of the SEIS process to the limit process.

Theorem 1.3.3. *For fixed λ, let ξ_t denote the SEIS process on a countable graph with bounded degree, under the rescaling $t \mapsto t/\tau$, and let ζ_t denote the limit process. Let $S = \{t : \xi_t(x) = 2 \text{ for some } x\}$ denote the set of times when the rescaled SEIS process has an infectious site. Fix $T > 0$ and an initial state with no infectious sites and finitely many exposed sites, then for each τ there is a coupling of ξ_t and ζ_t so that with probability tending to 1 as $\tau \to \infty$,*

- *$\zeta_t = \xi_t$ for $t \in [0, T] \setminus S$ and*
- *$\ell(S \cap [0, T]) \to 0$ where ℓ is Lebesgue measure on the line.*

The main idea of the proof is that with probability tending to 1 as $\tau \to \infty$ in the SEIS process, between any two onset events a recovery event occurs, and when this happens the SEIS process behaves like the limit process. The assumption of finitely many initially active sites is necessary.

Unlike the contact process or the limit process, with respect to the obvious graphical representation, for $\tau > 0$ the SEIS process is not monotone in the partial order induced by the order $0 < 1 < 2$ on types (or, it can be checked, for any other order, though $0 < 2 < 1$ is the only other real possibility), since if we take configurations $\eta \leq \eta'$ with $\eta(x) = 1$ and $\eta'(x) = 2$ the 2 can flip to a 0 before the 1 becomes a 2, since type 1 ignores recovery labels. Intuitively, this makes sense because although type

Table 1.1: Lower bounds on $\lambda_c^-(\tau)$

τ	10^4	10^3	100	10	1	0.58	1/10	1/100	10^{-3}	10^{-4}
$\lambda_c^-(\tau) >$	1.57	1.57	1.56	1.45	1.15	1.13	1.24	1.32	1.34	1.34

2 can spread the infection while type 1 cannot, type 1 is not vulnerable to recovery events while type 2 is.

So, lacking monotonicity, we define the following two critical values for the SEIS process; note $\mathbb{P}_{\lambda,\tau}$ denotes the law of the process with parameters λ, τ.

$$\begin{aligned} \lambda_c^-(\tau) &= \sup\{\lambda' : \mathbb{P}_{\lambda,\tau}(\xi_t \text{ dies out } ||\xi_0| < \infty) = 1 \text{ if } \lambda < \lambda'\} \\ \lambda_c^+(\tau) &= \inf\{\lambda' : \mathbb{P}_{\lambda,\tau}(\xi_t \text{ survives } ||\xi_0| > 0) > 0 \text{ if } \lambda > \lambda'\} \end{aligned}$$

Clearly, $\lambda_c^-(\tau) \leq \lambda_c^+(\tau)$ for each τ. The next result gives quantitative estimates on critical values, both for the SEIS process and for the limit process, on $\mathbb{Z}$, i.e., on the graph $G = (V, E)$ with $V = \mathbb{Z}$ and $E = \{xy : |x - y| = 1\}$.

Theorem 1.3.4. *For the SEIS process on* $\mathbb{Z}$, $\lambda_c^+(\tau) < 6.875$ *when* $\tau \leq 1/10$, *and* $\lambda_c^-(\tau)$ *has the lower bounds given in Table 1. For the limit process on* $\mathbb{Z}$, $1.944 < \lambda_c^\infty < 8.563$.

Lower bounds on λ^- are obtained using the method of [51] applied to a monotone process that upperbounds the SEIS process, and the upper bound on λ^+ for small τ is obtained with the method of [10] applied to a monotone process that lowerbounds the SEIS process. In both cases the estimates are achieved with the assistance of a computer and are rigorous up to the rounding error on computations. Unfortunately, in this case each lower bound on λ^- is computed for a single value of τ; it is possible to make guesses by interpolating, but these are not a priori rigorous. Note also that the lower bounds suggest, but again do not prove, that the critical value of the upperbound process has a unique minimum near $\tau = 0.58$ and is otherwise increasing/decreasing. Numerical simulations of the SEIS process on $\mathbb{Z}$ suggest that $\lambda_c^-(\tau) = \lambda_c^+(\tau)$ and that this value increases monotonically from about 1.6 at $\tau = 0$ to about 2.4 as $\tau \to \infty$.

For the limit process, the lower bound is obtained using the method of [51] and the upper bound, using the method of [10]. To distinguish it from other critical values we let λ_c^0 denote the critical value of the contact process. For the contact process on $V = \mathbb{Z}$ with $E = \{xy : |x - y| = 1\}$, $1.539 \leq \lambda_c^0 \leq 1.942$ (lower bound from [51] and upper bound from [33]), and from the upper bound together with our estimate we note that the strict inequality $\lambda_c^\infty > \lambda_c^0$ holds.

Using different methods, we obtain some "qualitative" estimates on critical values.

Theorem 1.3.5. *For the SEIS process on $\mathbb{Z}$,*

- *there exists $\lambda_0 < \infty$ such that $\lambda_c^+(\tau) < \lambda_0$ for all τ,*
- *$\lambda^+(\tau), \lambda^-(\tau) \to \lambda_c^0$ as $\tau \to 0$ and*
- *$\lambda^+(\tau), \lambda^-(\tau) \to \lambda_c^\infty$ as $\tau \to \infty$.*

We show only that $\lambda_0 < \infty$ exists, as it seems difficult to get any sort of realistic estimate. The proof uses a comparison to oriented percolation with a bit of work to get around the lack of monotonicity. Convergence of λ^+, λ^- as $\tau \to 0$ is proved in both cases by passing to a sequence of finite systems and using continuity with respect to parameters. Convergence of λ^+, λ^- as $\tau \to \infty$ is proved in the same way, with a couple of technical points that first need to be proved for the limit process.

1.3.3 New results for the two-stage contact process

Next we consider a natural generalization of the contact process. This model was first introduced in [29], and we can think of it as a model of population growth in which each individual goes through a juvenile stage before maturing and being able to produce offspring. Given an undirected graph $G = (V, E)$ the state space is $\{0, 1, 2\}^V$ and the transitions are

$$\begin{aligned} 0 &\to 1 \text{ at rate } \lambda n_2 \\ 1 &\to 2 \text{ at rate } \gamma \\ 1 &\to 0 \text{ at rate } 1 + \delta \\ 2 &\to 0 \text{ at rate } 1 \end{aligned}$$

where $n_2(x)$ is the cardinality of the set $\{y \in V : xy \in E,\ \xi(y) = 2\}$. State 0 is unoccupied, state 1 represents a juvenile individual and state 2 is a mature individual.

In [29], some basic properties of the model are established. The model is shown to be additive and monotone with respect to the natural partial order on configurations, and is also shown to be monotone increasing in λ and γ and monotone decreasing in δ. Note that for monotonicity with the natural order $0 < 1 < 2$ on types it is important that the rate $1 \to 0$ is at least as large as the rate $2 \to 0$, i.e., that $\delta \geq 0$, otherwise the 1s have some advantage over the 2s, as was the case for the SEIS model. A duality relation is proved, and some bounds are given on the set of parameter values such that the population survives with positive probability, starting from a single occupied site. In Chapter 4 we simplify the proof of the duality relation given in [29] and answer most of the open questions posed in Section 4 of that paper.

The first result is a lower bound on the set of values γ so that the process dies out, i.e., reaches the all 0 configuration, with probability 1. The bound depends only on the maximum degree $M = \max_x \deg x$ of the graph.

Theorem 1.3.6. *If $\gamma < 1/(2M-1)$ then starting from any finite number of occupied sites, the process dies out for any value of λ and δ.*

This answers question 6 in [29], where the author supplies a bound for $\mathbb{Z}^1$ in the case of nearest neighbour interactions, and asks whether a bound exists for other interactions, or for $\mathbb{Z}^d$ with $d > 1$.

The next result shows that the two notions of survival for the two-stage contact process coincide, answering question 1 in [29] affirmatively. For terminology see Sections 4.3.1 and 4.3.5.

Theorem 1.3.7. *For the two-stage contact process on $\mathbb{Z}^d$, single-site survival occurs if and only if the upper invariant measure is non-trivial.*

The proof uses the construction of [3] to show that for both the process and its dual, single-site survival implies the upper invariant measure is non-trivial.

An important question for growth models is that of *complete convergence*, which we show is true for the two-stage contact process, answering question 3 in [29]. Here

λ_c is the critical value for single-site survival as defined in Section 4.3.1 and μ_t denotes the distribution of the process at time t.

Theorem 1.3.8. *If $\lambda > \lambda_c$ then complete convergence holds, i.e.,*

$$\xi_t \Rightarrow \alpha\delta_0 + (1-\alpha)\nu$$

where ν is the upper invariant measure, δ_0 concentrates on the configuration with all 0s and $\alpha = \mathbb{P}_{\mu_0}(\xi_t$ dies out $)$ where $\mathbb{P}_{\mu_0}$ is the law of the process with initial distribution μ_0.

1.3.4 Duality and complete convergence for multi-type additive growth models

Here we consider a fairly general class of growth models on a state space F^V, where F is a finite set. We are interested in additive models, which are defined as follows. Say that an operation $\vee$ on F is a *join* operation if for all $a, b \in F$, $a \vee b \in F$, $a \vee b \geq a, b$, and $c \geq a, b \Rightarrow c \geq a \vee b$. A model is *additive* with respect to $\vee$ if there is a graphical construction of the model in which the same property holds as described in Section 1.2.1 for the contact process, i.e., letting ξ_0, ξ_0', ξ_0'' be initial configurations and letting ξ_t, ξ_t' and ξ_t'' denote the respective processes with those initial configurations, if $\xi_0'' = \xi_0 \vee \xi_0'$ then $\xi_t'' = \xi_t \vee \xi_t'$ for $t > 0$.

Subject to a mild technical condition that ensures the graphical construction leads to a well-defined and unique process, we can show the following for additive growth models.

Theorem 1.3.9. *An additive growth model has a* dual *process that is an additive growth model.*

As it turns out, an additive growth model has a natural extension to what we call a *multi-colour* additive growth model, in which there is a *primitive* set of types $F_p \subset F$ generating F in a unique way, i.e., any $b \in F$ is equal to $a_1 \vee ... \vee a_k$ for some $a_1, ..., a_k \in F_p$ in exactly one way. The upshot of a multi-colour growth model is that the converse of the above result holds.

Theorem 1.3.10. *If a multi-colour growth model has a dual, then the model is additive.*

This is shown by first showing the dual is additive, then showing that *its* dual is additive, and identifies with the original model.

Lastly we show that subject to some additional hypotheses, an additive growth model on the lattice $\mathbb{Z}^d$ exhibits complete convergence in the same sense as for the two-stage contact process described above. To check that the upper invariant measure ν is well-defined it suffices that the model be monotone, that its generator be "nice enough" (the precise condition used in [32] is the Feller condition, which is true of all the models we consider) and that there be a largest configuration to start from. Monotonicity is implied by additivity if we use the partial order defined by $\xi \leq \xi'$ if and only if $\xi' = \xi \vee \xi'$. A largest configuration can be obtained by setting the state at each site to be equal to $\bigvee_{a \in F} a$. We first state the result, and then explain the terminology and the notions required to understand it.

Theorem 1.3.11. *An additive growth model on $\mathbb{Z}^d$ that is translation-invariant, symmetric, has finite range, is irreducible, and has only productive and destructive interactions exhibits complete convergence, that is, letting μ_t denote the distribution at time t, for any μ_0 it holds that*

$$\mu_t \Rightarrow \alpha\nu + (1-\alpha)\delta_0$$

where the upper invariant measure ν is the weak limit starting from the largest configuration and $\alpha = \mathbb{P}(|\xi_t| > 0 \ \forall t > 0)$. Moreover, single-site survival is equivalent to the condition $\nu \neq \delta_0$.

Translation-invariant, symmetric and finite range means that transitions look the same from each site, look the same after permuting coordinates, and can only involve sites separated by distance at most M for some $M > 0$. Irreducibility means that for any finite $V_0 \subset V$ and any $\phi, \psi : V_0 \to F$, it is possible to go from $\xi(x) = \phi(x)$ for all $x \in V_0$ to $\xi(x) = \psi(x)$ for all $x \in V_0$, via some sequence of transitions.

The condition of productive and destructive interactions means that at a given transition, either every configuration will increase or stay the same, or else every configuration will decrease or stay the same. This condition is related to *positive correlations*, defined as follows. Define an *increasing function* $f : F^V \to \mathbb{R}$ to be one with $\xi \leq \xi' \Rightarrow f(\xi) \leq f(\xi')$, and say that a distribution μ on F^V has *positive correlations* if $\mathbb{E}_\mu fg \geq \mathbb{E}_\mu f \mathbb{E}_\mu g$ for all increasing functions f, g. Then letting $\mu_t, t \geq 0$ denote the

distribution of ξ_t, say that ξ_t *preserves positive correlations* if μ_t has positive correlations for $t > 0$ provided μ_0 has positive correlations.

To show complete convergence, we need both the model and its dual to preserve positive correlations, and it is the condition of productive and destructive interactions that ensures this is the case. Note that the two-stage contact process described above can easily be shown to preserve positive correlations, using the following useful condition due to Harris.

Theorem 1.3.12 ([25])**.** *Let ξ_t be a growth model for which monotonicity holds as described in Section 1.2.1. Then, ξ_t preserves positive correlations if and only if transitions are only between comparable states, i.e., letting $\xi_{t^-}(x)$ denote $\lim_{s\to t^-} \xi_s(x)$, for each finite $V_0 \subset V$, either $\xi_t(x) \leq \xi_{t^-}(x)$ for all $x \in V_0$ or $\xi_t(x) \geq \xi_{t^-}(x)$, for all $x \in V_0$.*

Chapter 2

Social Contact Processes and the Partner Model

This chapter consists of the journal article [20] of the same title, accepted for publication to Annals of Applied Probability.

Abstract

We consider a stochastic model of infection spread on the complete graph on N vertices incorporating dynamic partnerships, which we assume to be monogamous. This can be seen as a variation on the contact process in which some form of edge dynamics determines the set of contacts at each moment in time. We identify a basic reproduction number R_0 with the property that if $R_0 < 1$ the infection dies out by time $O(\log N)$, while if $R_0 > 1$ the infection survives for an amount of time $e^{\gamma N}$ for some $\gamma > 0$ and hovers around a uniquely determined metastable proportion of infectious individuals. The proof in both cases relies on comparison to a set of mean-field equations when the infection is widespread, and to a branching process when the infection is sparse.

2.1 Introduction

The contact process is a well studied model of the spread of an infection, in which an undirected graph $G = (V, E)$ determines a collection of sites V and edges E which we can think of as individuals and as links between individuals along which the infection

can be transmitted. Each site is either healthy or infectious; infectious sites recover at a certain fixed rate which is usually normalized to 1, and transmit the infection to each of their neighbours at rate λ.

The contact process has been studied in a variety of different settings, including lattices [7],[3],[32],[34] (to cite just a few), infinite trees [42], power law graphs [45] [37], and complete graphs [43]. In each case there is a critical value λ_c below which the infection quickly vanishes from the graph, and above which the infection has a positive probability of surviving either for all time (if the graph is infinite), or for an amount of time that grows quickly (either exponentially or at least faster than polynomially) with the size of the graph; in the power law case $\lambda_c = 0$ so long-time survival is possible whenever $\lambda > 0$.

In a social context, G might describe a contact network in which an edge connects sites x and y if and only if the corresponding individuals have sufficiently frequent interactions that infection can be spread from one to the other. In the contact process, the contact network is fixed, that is, a given pair of individuals is either connected or not connected for all time. However, we can easily imagine a scenario in which connections form and break up dynamically, which we can model by having edges open and close according to certain rules; here we use the convention of percolation theory, in which "open" means there is a connection across the edge; note this is the opposite of the convention for electric circuits. In this case the edges E represent *possible* connections and we have a process $E_t \subseteq E$ that describes the set of open edges as a function of time. This type of process we will call a *social contact process*, since it involves some form of social dynamics.

In the simplest case, edges open and close independently at some fixed rates r_+ and r_-. In this case the distribution of open edges at a given time converges to the product measure on $\{0,1\}^E$ with density $r_+/(r_- + r_+)$. Estimates on the survival region can then be obtained using the results of [6] and following the pattern of [44]. On the other hand, edge dynamics could depend on the state of the infection; for example, site x might be less likely to connect with site y, if y is infected. If we then relax the tendency to avoid infected sites, then for a given value of λ, we might ask at what point does the infection start to spread, if it does.

Here we consider edges opening and closing independently as described above but with the added restriction of *monogamy*, that is, if two sites are connected (i.e., linked by an edge) then so long as they remain connected, they cannot connect to other sites. In this model we think of connected pairs as partners, so we call it the *partner model*. For simplicity, we study the model on the sequence of complete graphs K_N on N vertices, where N will tend to ∞; this is a reasonable model for, say, the spread of a sexually transmitted infection through a homogeneous population of monogamous homosexual individuals in a big city. We rescale the partner formation rate per edge to r_+/N to ensure that a given individual in a pool of entirely singles finds a partner at total rate approximately r_+. For future reference, we use interchangeably both the words healthy and susceptible, and the words unpartnered and single, to describe respectively an individual that is not infectious, or an individual that does not have a partner. Even in this simple model, as described below, there is a phase transition between extinction and spread of the infection.

2.2 Statement of Main Results

In order to analyze the partner model we should first ensure that it is well-defined, so following [26] we give a graphical construction which makes it easy to visualize its evolution in time and space. We write the model as (V_t, E_t) where $V_t \subseteq V$ is the set of infectious sites at time t and $E_t \subseteq E$ is the set of open edges at time t. In general we assume $\min(r_+, r_-, \lambda) > 0$ since if any of the parameters is equal to zero the dynamics are trivial.

The complete graph $K_N = (V, E)$ has sites $V = \{1, ..., N\}$ and edges $E = \{\{x, y\} : x, y \in \{\{1, ..., N\}, x \neq y\}$. On the spacetime set $K_N \times [0, \infty)$, place independent Poisson point processes (p.p.p.s) along the fibers $\{\cdot\} \times [0, \infty)$ as follows:

- for recovery, at each site with intensity 1 and label $\times$,
- for transmission, along each edge $xy \in E$ with intensity λ and label $\leftrightarrow$,
- for partnership formation, along each edge with intensity r_+/N and label $\uparrow$,
- for partnership breakup, along each edge with intensity r_- and label $\downarrow$.

These define the probability space Ω, whose realizations $\omega \in \Omega$ consist of collections of labelled points on $K_N \times [0, \infty)$. Since the graph is finite, the total intensity of

p.p.p.s is finite, thus with probability 1 events are well-ordered in time. Fixing an admissible initial configuration (V_0, E_0) i.e., such that no two edges xy and yz are both open, we determine (V_t, E_t) as follows. For a well-ordered realization with event times $t_1 < t_2 < t_3 < ...$, suppose (V_{t_i}, E_{t_i}) is known. If the event at time t_{i+1} is

- an $\times$ at site x and $x \in V_{t_i}$ then $V_{t_{i+1}} = V_{t_i} \setminus \{x\}$,
- a $\leftrightarrow$ along edge xy, $xy \in E_{t_i}$, $x \in V_{t_i}$ and $y \notin V_{t_i}$ then $V_{t_{i+1}} = V_{t_i} \cup \{y\}$,
- a $\uparrow$ along edge xy and $xz, zy \notin E_{t_i}$ for all z then $E_{t_{i+1}} = E_{t_i} \cup \{xy\}$,
- a $\downarrow$ along edge xy and $xy \in E_{t_i}$ then $E_{t_{i+1}} = E_{t_i} \setminus \{xy\}$.

Otherwise the configuration is unchanged. This gives (V_t, E_t) at times $t_0 := 0, t_1, t_2, ...$; for $t \in (t_i, t_{i+1})$ set $V_t = V_{t_i}$ and $E_t = E_{t_i}$.

For the partner model we are mostly concerned not with the exact values of V_t and E_t but with the total number of susceptible and infectious singles S_t and I_t and the total number of partnered pairs SS_t, SI_t, II_t of the three possible types; as shown in Section 2.5, for each N, $(S_t, I_t, SS_t, SI_t, II_t)$ is a continuous time Markov chain. In general it will be more convenient to work with the rescaled quantities $s_t = S_t/N$, $i_t = I_t/N$, $ss_t = SS_t/N$, $si_t = SI_t/N$, and $ii_t = II_t/N$.

Starting from any configuration, as shown in Section 2.6, after a short time the proportion of singles $y_t := s_t + i_t$ approaches and remains close to a certain fixed value $y^* \in (0, 1)$. The computation of y^* is given in Section 2.3: setting $\alpha = r_+/r_-$, we find that

$$y^* = 1/(2\alpha)[-1 + \sqrt{1 + 4\alpha}] \tag{2.2.1}$$

To determine the conditions under which the infection can spread we use a heuristic argument. Once we know the correct values, we can then worry about proving they are correct. Suppose we start with $V_0 = \{x\}$ for some $x \in V$ with x single and $y_0 \approx y^*$, and keep track of x until the first moment when x either

- recovers without finding a partner, or
- if it finds a partner before recovering, breaks up from that partnership.

This leads to the continuous time Markov chain shown in Figure 2.1. Each of $A, B, ..., G$ represents a state for the chain, and arrows show possible transitions,

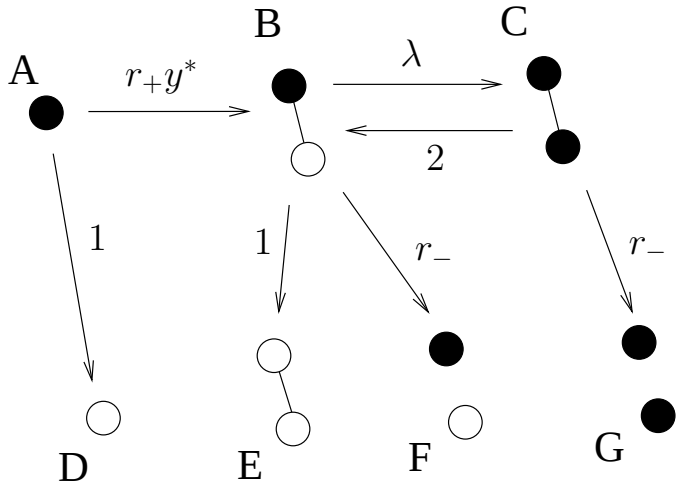

Figure 2.1: Markov Chain used to compute R_0, with transition rates indicated; infectious sites are shaded

with the arrow labelled by the transition rate. Shaded circles represent infectious individuals and unshaded circles, healthy individuals. A pair of circles connected by a line represents a partnered pair. Starting from A, a single infectious site either recovers (goes to D) at rate 1, or finds a healthy partner at rate r_+y^*. Infection takes place at rate λ. If only one individual in a partnership is infectious (state B), then it recovers at rate 1 (state E), and we don't need to worry about them any more, since neither is infectious. If both are infectious (state C), recovery of one or the other occurs at rate 2. While in a partnership, breakup occurs at rate r_-.

Define the *basic reproduction number*

$$R_0 = \mathbb{P}(A \to F) + 2\mathbb{P}(A \to G) \tag{2.2.2}$$

which is the expected number of infectious singles upon absorption of the above Markov chain, starting from state A. As intuition suggests, and Theorem 2.2.2 confirms, the infection can spread if $R_0 > 1$, and cannot spread if $R_0 \leq 1$.

If the dynamics is in equilibrium i.e., $(s_t, i_t, ss_t, si_t, ii_t)$ hovers around a fixed value $(s^*, i^*, ss^*, si^*, ii^*)$, then in particular the proportion of infectious singles is roughly constant. To compute this proportion we again use a heuristic argument. Three events affect infectious singles:

- $I \to S$, which occurs at rate $I_t = i_t N$,
- $I + I \to II$ which occurs at rate $(r_+/N)\binom{I_t}{2} \approx r_+(i_t^2/2)N$, and
- $S + I \to SI$ which occurs at rate $(r_+/N)I_t S_t = r_+ i_t s_t N$.

If a partnership is formed, then using these rates and Figure 2.1, we can compute the expected number of infectious singles upon breakup. Fixing $i_t = i$ for some $i \in [0, y^*]$ and $s_t + i_t = y^*$, define the normalizing constant $z = 1 + r_+ i/2 + r_+(y^* - i) = 1 + r_+(y^* - i/2)$ and the probabilities $p_S = 1/z$, $p_{II} = r_+ i/(2z)$ and $p_{SI} = r_+(y^* - i)/z$ and let

$$\Delta(i) = p_S \Delta_S + p_{II}\Delta_{II} + p_{SI}\Delta_{SI} \tag{2.2.3}$$

where $\Delta_S = -1$, $\Delta_{II} = -2 + \mathbb{P}(C \to F) + 2\mathbb{P}(C \to G)$ and $\Delta_{SI} = -1 + \mathbb{P}(B \to F) + 2\mathbb{P}(B \to G)$. The function $\Delta(i)$ tracks the expected change in the number of infectious singles, per event affecting one or more infectious singles. Thus, for an equilibrium solution we should have $\Delta(i^*) = 0$. As shown in Lemma 2.4.3, to have a solution with $i^* > 0$, we need $R_0 > 1$.

As shown in Lemma 2.4.1, for fixed r_+, r_-, R_0 is continuous and increasing in λ. Defining

$$\lambda_c = \sup\{\lambda \geq 0 : R_0 \leq 1\} \tag{2.2.4}$$

with $\sup \mathbb{R}_+ := \infty$, it follows that if $\lambda_c = \infty$ then $R_0 < 1$ for all λ, and if $\lambda_c < \infty$ then $R_0 < 1$ if $\lambda < \lambda_c$, $R_0 = 1$ if $\lambda = \lambda_c$ and $R_0 > 1$ if $\lambda > \lambda_c$. In models exhibiting a phase transition one often seeks a *critical exponent* γ such that for an observable $F(\lambda)$ it holds that $F(\lambda) \sim C(\lambda - \lambda_c)^\gamma$. As we see in the statement of the upcoming Theorem 2.2.1, here the critical exponent for i^* is equal to 1.

The following two theorems are the main results of this paper. The first result tells us where and when we should expect a phase transition to occur. In particular, it gives a formula for λ_c and describes the behaviour of i^* near λ_c.

Theorem 2.2.1. *Let $y^*, R_0, \Delta(i)$ and λ_c be as in* (2.2.1), (2.2.2), (2.2.3) *and* (2.2.4) *and let r_+, r_- be fixed. Then, $\lambda_c < \infty \Leftrightarrow r_+ y^* > 1 \Leftrightarrow r_+ > 1 + 1/r_-$ and in this case*

$$\lambda_c = \frac{2}{r_-}\frac{2}{(r_+ y^* - 1)} + \frac{2}{r_-} + \frac{4}{r_+ y^* - 1} + 1 + \frac{r_-}{r_+ y^* - 1}$$

If $R_0 = R_0(\lambda) > 1$ there is a unique solution $i^*(\lambda) \in (0, y^*)$ to the equation $\Delta(i^*) = 0$ and $i^*(\lambda) \sim C(\lambda - \lambda_c)$ as $\lambda \downarrow \lambda_c$, for some constant $C > 0$.

The second result shows that our heuristics are correct. More precisely, $R_0 > 1$ is a necessary and sufficient condition for spread and long-time survival of the infection. Moreover, when $R_0 > 1$ there is a unique and globally stable endemic equilibrium with $i^* > 0$ given by $\Delta(i^*) = 0$.

Theorem 2.2.2. *Fix λ, r_+, r_- and let y^*, R_0 and $\Delta(i)$ be as defined in* (2.2.1), (2.2.2) *and* (2.2.3).

- *If $R_0 \leq 1$, for each $\epsilon > 0$ there are constants $C, T, \gamma > 0$ so that, from any initial configuration, with probability $\geq 1 - Ce^{-\gamma N}$, $|V_T| \leq \epsilon N$.*
- *If $R_0 < 1$ there are constants $C, T, \gamma > 0$ so that, from any initial configuration, with probability tending to 1 as $N \to \infty$ all sites are healthy by time $T + C \log N$.*
- *If $R_0 > 1$, there is a unique vector $(s^*, i^*, ss^*, si^*, ii^*)$, satisfying $i^* > 0$, $s^* + i^* = y^*$ and $\Delta(i^*) = 0$, such that*
  - *for each $\epsilon > 0$, there are constants $C, T, \gamma > 0$ so that, from any initial configuration with $|V_0| \geq \epsilon N$, with probability $\geq 1 - Ce^{-\gamma N}$, $|(s_t, i_t, ss_t, si_t, ii_t) - (s^*, i^*, ss^*, si^*, ii^*)| \leq \epsilon$ for $T \leq t \leq e^{\gamma N}$, and*
  - *there are constants $\delta, p, C, T > 0$ so that, from any initial configuration with $|V_0| > 0$, with probability $\geq p$, $|V_{T + C \log N}| \geq \delta N$.*

To obtain the value of the endemic equilibrium and the behaviour when $|V_0| \geq \epsilon N$, which we call the *macroscopic* regime, we use the *mean-field equations* (MFE) introduced in Section 2.5, which are a set of differential equations that give a good approximation to the evolution of $(s_t, i_t, ss_t, si_t, ii_t)$ when N is large. To describe the behaviour when $1 \leq |V_0| \leq \epsilon N$ for small $\epsilon > 0$, which we call the *microscopic* regime, we use comparison to a branching process; if $R_0 < 1$ we bound above and if $R_0 > 1$ we bound below.

The paper is laid out as follows. Sections 2.3 and 2.4 contain the heuristic calculations that allow us to determine $y^*, R_0, \lambda_c, \Delta(i)$ and prove Theorem 2.2.1. In Section 2.3 we give an informal description of the edge dynamics and compute y^*. In Section 2.4 we analyze $R_0, \lambda_c, \Delta(i)$ and prove Theorem 2.2.1, in two parts: Proposition

2.4.2 and Proposition 2.4.4. In Section 2.5 we introduce the mean-field equations and characterize their dynamics. In Sections 2.6, 2.7 and 2.8 we consider the stochastic process and prove Theorem 2.2.2. In Section 2.6 we develop the tools needed to relate the stochastic model to the mean-field equations. In Section 2.7 we prove the macroscopic part of Theorem 2.2.2, and in Section 2.8 we prove the microscopic part.

2.3 Proportion of Singles

Starting from the total number of singles $Y_t = S_t + I_t$ the transitions are

- $Y \to Y - 2$ at rate $(r_+/N)Y(Y-1)/2$
- $Y \to Y + 2$ at rate $(N - Y)r_-/2$

which for $y_t := Y_t/N$ gives

- $y \to y - 2/N$ at rate $[r_+ y(y - 1/N)/2]N = (r_+ y^2/2)N - r_+ y/2$
- $y \to y + 2/N$ at rate $[(1-y)r_-/2]N$

Combining these transitions gives

$$\frac{d}{dt}\mathbb{E}(y_t \mid y_t = y) = -r_+ y^2 + r_-(1-y) + \frac{r_+ y}{N}$$

In Lemma 2.6.6 we make a rigorous statement about the behaviour of y_t. For now, though, some heuristics are helpful. Letting $y = Y/N$ and Δy denote the increment in y over a time step of size $1/N$ we find $\mathbb{E}\Delta y = O(1/N)$ while $\mathbb{E}(\Delta y)^2 = O(1/N^2)$ which means $\mathrm{Var}(\Delta y) = O(1/N^2)$. This suggests that as $N \to \infty$ we should expect the sample paths of y to approach solutions to the differential equation

$$y' = -r_+ y^2 + r_-(1-y) \tag{2.3.1}$$

Notice the right-hand side is positive at $y = 0$, negative at $y = 1$ and strictly decreases with y, so there is a unique and globally stable equilibrium for $y \in [0,1]$, that lies in $(0,1)$. Setting $y' = 0$ and letting $\alpha = r_+/r_-$ gives the equation $\alpha y^2 + y - 1 = 0$ which has the unique solution $y^* = 1/(2\alpha)[-1 + \sqrt{1+4\alpha}]$ in $[0,1]$. Notice that $y^* \sim 1 - \alpha$ as $\alpha \to 0^+$ and $y^* \sim 1/\sqrt{\alpha}$ as $\alpha \to \infty$.

2.4 Survival Analysis

In this section we analyze R_0, λ_c and $\Delta(i)$ which are defined in Section 2.2. We begin with R_0 defined in (2.2.2). Define the recruitment probability $p_r = r_+y^*/(1+r_+y^*) = \mathbb{P}(A \to E \cup F \cup G)$ which is the probability of finding a partner before recovering and depends only on r_+, r_-. Define $a = 1 + \lambda + r_-$, $b = 2 + r_-$ which are the rates at which the Markov chain of Figure 2.1 jumps away from states B and C, respectively. Also, let

$$\sigma = \sum_{k=0}^{\infty} \left(\frac{\lambda}{a}\frac{2}{b}\right)^k = \frac{ab}{ab - 2\lambda}$$

It is easy to check that $ab > 2\lambda$. Notice that any path from A to $E \cup F \cup G$ must go to B and then goes around the B,C loop some number of times before being absorbed at E, F or G, and σ accounts for this looping. Summing probabilities over all possible paths we find

$$\mathbb{P}(A \to F) = p_r\sigma\frac{r_-}{a} \quad \text{and} \quad \mathbb{P}(A \to G) = p_r\sigma\frac{\lambda}{a}\frac{r_-}{b}$$

so we obtain the explicit expression

$$R_0 = p_r\sigma r_-(1 + 2\lambda/b)/a$$

which after re-substituting and a bit of algebra gives

$$R_0 = p_r r_- \frac{b + 2\lambda}{ab - 2\lambda} = p_r r_- \frac{2 + r_- + 2\lambda}{2 + 3r_- + \lambda r_- + r_-^2} \tag{2.4.1}$$

Lemma 2.4.1. *Fixing r_+ and r_-, R_0 is continuous and increasing with respect to λ.*

Proof. Continuity is obvious from the formula above. We write $R_0(\lambda)$ and compute the derivative $R_0'(\lambda)$, noting that p_r is fixed. Letting $c_1 = 2 + r_-$, $c_2 = 2$, $c_3 = 2 + 3r_- + r_-^2$ and $c_4 = r_-$, $R_0(\lambda) = p_r r_-(c_1 + c_2\lambda)/(c_3 + c_4\lambda)$ so $R_0'(\lambda) = p_r r_-(c_2c_3 - c_1c_4)/(c_3 + c_4\lambda)^2$ and $c_2c_3 - c_1c_4 = 4 + 4r_- + r_-^2 > 0$ so $R_0'(\lambda) > 0$. □

From this it follows that for fixed r_+, r_-, if $R_0(\lambda) = 1$ has a solution then it is unique and is equal to λ_c. So, setting $R_0 = 1$ gives

$$p_r r_-(2 + r_- + 2\lambda_c) = 2 + 3r_- + \lambda_c r_- + r_-^2 \tag{2.4.2}$$

To get a handle on this equation we first examine the limit of large r_+ i.e., quick formation of partnerships. As noted in Section 2.3, $y^* \sim 1/\sqrt{\alpha} = \sqrt{r_-}/\sqrt{r_+}$ as $\alpha = r_+/r_- \to \infty$, so for fixed r_-, $r_+ y^* \sim \sqrt{r_- r_+} \to \infty$, and so $p_r \to 1$, as $r_+ \to \infty$. Setting $p_r = 1$ in the equation above, after cancelling like terms and dividing both sides by r_- gives

$$\lambda_c = 1 + 2/r_-$$

for fixed r_-, when $r_+ = \infty$. For the contact process on a large complete graph $\lambda_c = 1$, so here the only difference is the term $2/r_-$ which makes it harder for the infection to spread when partnerships last a long time.

Accounting for p_r, we still get a fairly nice expression. From (2.4.2), putting all terms involving λ_c on the left and all other terms on the right gives

$$\lambda_c r_-(2p_r - 1) = 2 + (3 - 2p_r)r_- + r_-^2(1 - p_r)$$

Letting $\beta = 2p_r - 1$ then substituting for β and dividing by r_- gives

$$\lambda_c \beta = 2/r_- + (2 - \beta) + (1/2)r_-(1 - \beta) \tag{2.4.3}$$

This equation suggests that we view $\lambda\beta$ as a sort of force of infection, which makes sense as λ is the transmission rate and $\beta = 2p_r - 1$ measures the chance of finding a partner before recovering. Although β depends on r_-, $-1 \le \beta \le 1$ regardless, so we see from (2.4.3) that for fixed λ and r_+, if r_- is either too small or too large, the infection cannot spread. The reason for this can be understood as follows: if r_- is too small, partners tend both to recover before breaking up and transmitting the infection to anyone else, whereas if r_- is too large, partnerships do not last long enough for transmission to occur.

Using (2.4.3) we can now prove the first assertion of Theorem 2.2.1.

Proposition 2.4.2. *For fixed r_+, r_- and λ_c given by (2.2.4), $\lambda_c < \infty$ if and only if $r_+ y^* > 1$, if and only if $r_+ > 1 + 1/r_-$ and in this case*

$$\lambda_c = \frac{2}{r_-}\frac{2}{(r_+ y^* - 1)} + \frac{2}{r_-} + \frac{4}{r_+ y^* - 1} + 1 + \frac{r_-}{r_+ y^* - 1}$$

Proof. It is easy to check, using the formula $y^* = (r_-/(2r_+))(-1 + (1 + 4r_+/r_-)^{1/2})$,

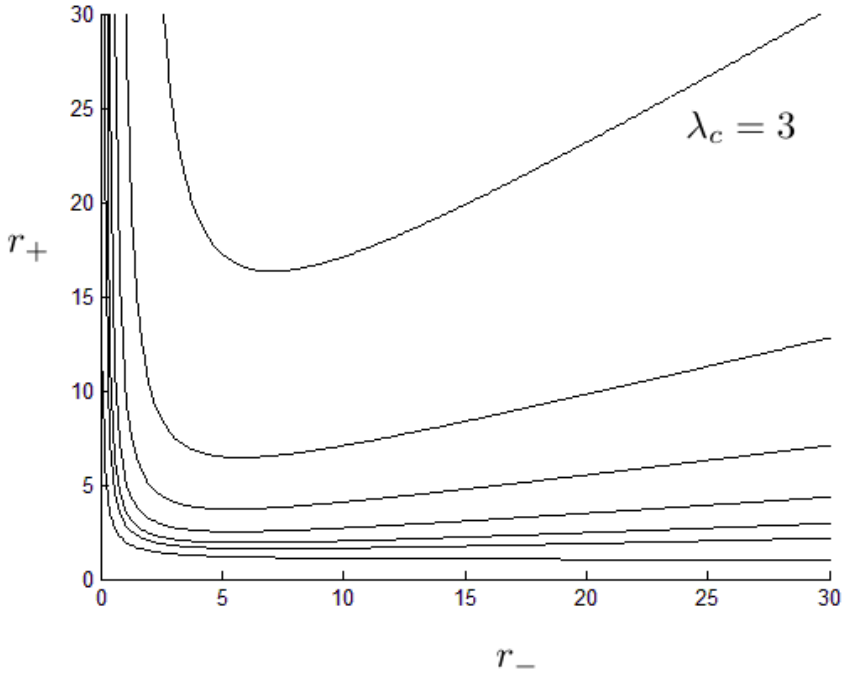

Figure 2.2: Level curves of λ_c depicted in the r_+, r_- plane. Starting from the top curve and going down, $\lambda_c = 3, 5, 8, 13, 21, 34, \infty$.

that $r_+ y^* > 1$ if and only if $r_+ > 1 + 1/r_-$. Since $\beta \in [-1, 1]$, the right-hand side of (2.4.3) is positive, so to have a solution it is necessary that $\beta > 0$; dividing by β on both sides shows that it is also sufficient. Then, observe that $\beta > 0$ if and only if $r_+ y^* > 1$. To get the formula for λ_c, divide by β in (2.4.3) and observe that $\beta^{-1} - 1 = 2/(r_+ y^* - 1)$. □

Figure 2.2 shows level curves of λ_c in the r_+, r_- plane. Using the formula for λ_c we can see how it scales in various limits of r_+, r_- and α. First we see what happens when we speed up and slow down the partnership dynamics. Let $\alpha = r_+/r_-$ be fixed (and by extension, y^*) and let r_-^* denote the unique value of r_- such that $r_+ y^* = 1$. We find that

- $\lambda_c \downarrow 1 + 1/(\alpha y^*)$ as $r_+ \uparrow \infty$ (fast partner dynamics)
- $\lambda_c(r_+ y^* - 1) \downarrow 4/r_-^* + 4 + r_-^*$ as $r_+ y^* \downarrow 1$ (slow partner dynamics)

In particular, in the limit of fast partner dynamics λ_c approaches its value for the contact process on a complete graph, plus a correction for the proportion of available

singles. In the slow limit i.e., as the recruitment probability approaches $1/2$, λ_c diverges like $1/(r_+y^*-1)$, with a proportionality that itself diverges as r_-^* approaches either 0 or ∞. Now we fix $r_+ > 1$ and vary r_-. Note that $y^* \downarrow 0$ as $r_- \downarrow 0$.

- as $r_- \uparrow \infty$, $y^* \uparrow 1$, $\alpha \downarrow 0$ and $\lambda_c/r_- \downarrow 1/(r_+ - 1)$, and
- as $r_+y^* \downarrow 1$, $\lambda_c(r_+y^* - 1) \downarrow 4/r_-^* + 4 + r_-^*$.

Here, in both limits λ_c diverges, in the first case like r_- and in the second case like $1/(r_+y^* - 1)$. Finally we fix r_- and vary r_+, and we find that

- as $r_+ \uparrow \infty$, $y^* \sim 1/\sqrt{\alpha} = \sqrt{r^-/r^+}$ and $\lambda_c \to 1 + 2/r_-$, and
- as $r_+y^* \downarrow 1$, $\lambda_c(r_+y^* - 1) \downarrow 4/r_- + 4 + r_-$

The first limit agrees with the previous large r_+ approximation, and the second limit shows that when r_+y^* is close to 1, $\lambda(r_+y^* - 1)/2 \approx \lambda(r_+y^* - 1)/(r_+y^* + 1) = \lambda\beta$ behaves like the force of infection and we require again that r_- be neither too small nor too large in order for the infection to be able to spread.

We now examine $\Delta(i)$, defined in (2.2.3).

Lemma 2.4.3. $\Delta(0) = R_0 - 1$, *and*

- *if* $R_0 < 1$ *the equation* $\Delta(i) = 0$ *has no solution* $i \in [0, y^*]$,
- *if* $R_0 = 1$, *the equation* $\Delta(i) = 0$ *has the unique solution* $i = 0$ *and*
- *if* $R_0 > 1$ *the equation* $\Delta(i^*) = 0$ *has a unique solution* $i^* \in (0, y^*)$.

Proof. Letting $z = 1 + r_+(y^* - i/2)$ we recall the definition:

$$\Delta(i) = p_S\Delta_S + p_{II}\Delta_{II} + p_{SI}\Delta_{SI} \tag{2.4.4}$$

with $p_S = 1/z$, $p_{SI} = r_+(y^* - i)/z$, $p_{II} = r_+i/(2z)$, $\Delta_S = -1$, $\Delta_{II} = -2 + \mathbb{P}(C \to F) + 2\mathbb{P}(C \to G)$ and $\Delta_{SI} = -1 + \mathbb{P}(B \to F) + 2\mathbb{P}(B \to G)$, where probabilities are with respect to the Markov chain in Figure 2.1.

First we show $\Delta(0) = R_0 - 1$. If $i = 0$ then $p_S = 1/(1 + r_+y^*) = \mathbb{P}(A \to D)$, $p_{II} = 0$ and $p_{SI} = r_+y^*/(1 + r_+y^*) = \mathbb{P}(A \to B)$ so

$$\begin{aligned}\Delta(0) &= -\mathbb{P}(A \to D) + \mathbb{P}(A \to B)(-1 + \mathbb{P}(B \to F) + 2\mathbb{P}(B \to G)) \\ &= -\mathbb{P}(A \to D \cup B) + \mathbb{P}(A \to F) + 2\mathbb{P}(A \to G) \\ &= -1 + R_0\end{aligned}$$

It is easy to check that $\Delta_{II} \leq 0$, so if $\Delta_{SI} \leq 0$ then $\Delta(i) < 0$ for $i \in [0, y^*]$, since $p_S > 0$ and $\Delta_S < 0$, and the other terms are ≤ 0. Since $\partial_i z = -r_+/2$, we find

$$\partial_i p_S = r_+/2z^2 > 0 \text{ and } \partial_i p_{II} = r_+/(2z) + r_+^2 i/(4z^2) > 0$$

and since $p_{SI} = 1 - (p_S + p_{II})$, $\partial_i p_{SI} = -\partial_i p_S - \partial_i p_{II} < 0$. If $\Delta_{SI} > 0$ it follows that $\partial_i \Delta(i) < 0$ so if $R_0 < 1$ but $\Delta_{SI} > 0$ then $\Delta(i) \leq \Delta(0) < 0$ for $i \in [0, y^*]$. If $R_0 \geq 1$ then since $0 \leq \Delta(0) = p_S\Delta_S + p_{SI}\Delta_{SI}$ and $\Delta_S < 0$, it follows that $\Delta_{SI} > 0$ and so $\partial_i \Delta(i) < 0$. If $R_0 = 1$, then since $\Delta(0) = 0$ it follows that $i = 0$ is the only solution in $[0, y^*]$ to the equation $\Delta(i) = 0$. If $i = y^*$ then $p_{SI} = 0$ so $\Delta(y^*) < 0$, and clearly $\Delta(i)$ is continuous on $[0, y^*]$. Therefore, if $R_0 > 1$ then since $\Delta(0) > 0$, by the intermediate value theorem the equation $\Delta(i^*)$ has a solution $i^* \in (0, y^*)$, and since $\partial_i \Delta(i) < 0$ the solution is unique. □

Write $\Delta(i)$ as $\Delta(\lambda, i)$ to emphasize the λ dependence. By Lemma 2.4.3 and since $R_0 = 1 \Leftrightarrow \lambda = \lambda_c$ and $R_0 > 1 \Leftrightarrow \lambda > \lambda_c$, for fixed r_+, r_- such that $r_+y^* > 1$, we have a function $i^*(\lambda)$ defined for $\lambda \geq \lambda_c$ satisfying $\Delta(\lambda, i^*(\lambda)) = 0$ such that $i^*(\lambda_c) = 0$ and $i^*(\lambda) > 0$ for $\lambda > \lambda_c$. Next we see how i^* behaves for $\lambda > \lambda_c$ near λ_c. As usual, C^1 means continuously differentiable.

Proposition 2.4.4. *For fixed r_+, r_- such that $r_+y^* > 1$, $i^* \sim C(\lambda - \lambda_c)$ as $\lambda \downarrow \lambda_c$ for some constant $C > 0$.*

Proof. Clearly p_S, p_{SI} and p_{II} depend only on i and are C^1 in a neighbourhood of 0. Also, Δ_S is fixed and Δ_{SI} and Δ_{II} depend only on λ and are rational functions of λ whose range lies in a bounded interval, thus are C^1 in a neighbourhood of λ_c. Glancing at (2.4.4), this means that $\Delta(\lambda, i)$ is C^1 in a neighbourhood of $(\lambda_c, 0)$. If $\lambda \geq \lambda_c$ then $R_0 \geq 1$, so as shown in the proof of Lemma 2.4.3, $\partial_i \Delta(\lambda, i) < 0$ and in particular, $\partial_i \Delta(\lambda_c, 0) \neq 0$. Applying the implicit function theorem, there is a unique C^1 function $i^*(\lambda)$ defined in a neighbourhood of λ_c (and thus coinciding with the

previous definition of $i^*(\lambda)$ when $\lambda \geq \lambda_c$) satisfying $\Delta(\lambda, i^*(\lambda)) = 0$ and, noting that $i^*(\lambda_c) = 0$,

$$i^*(\lambda) \sim -(\lambda - \lambda_c)\frac{\partial_\lambda \Delta(\lambda_c, 0)}{\partial_i \Delta(\lambda_c, 0)}$$

as $\lambda \downarrow \lambda_c$. A straightforward Markov chain coupling argument shows that $\partial_\lambda \Delta_{SI}, \partial_\lambda \Delta_{II} > 0$, which implies $\partial_\lambda \Delta(\lambda, i) > 0$. Since $\partial_i \Delta(\lambda, i) < 0$, the result follows. □

2.5 Mean-Field Equations

A set of differential equations defined below are indispensable to our analysis of the partner model as they enable a (better and better as N increases) approximate description of the model, when N is large. First we write down the transitions for the variables introduced in Section 2.2 that track the total number of singles and pairs of various types; there are ten such transitions. Since the transition rates of the set of variables $(S_t, I_t, SS_t, SI_t, II_t)$ depends only on its present value, it is a continuous time Markov chain.

- $I \to I - 1$ and $S \to S + 1$ at rate I,
- $S \to S - 2$ and $SS \to SS + 1$ at rate $(r_+/N)S(S-1)/2$,
- $S \to S - 1$, $I \to I - 1$ and $SI \to SI + 1$ at rate $(r_+/N) \cdot S \cdot I$,
- $I \to I - 2$ and $II \to II + 1$ at rate $(r_+/N)I(I-1)/2$,
- $SI \to SI - 1$ and $SS \to SS + 1$ at rate SI,
- $II \to II - 1$ and $SI \to SI + 1$ at rate $2II$,
- $SI \to SI - 1$ and $II \to II + 1$ at rate λSI,
- $SS \to SS - 1$ and $S \to S + 2$ at rate r_SS,
- $SI \to SI - 1$, $S \to S + 1$ and $I \to I + 1$ at rate r_SI, and
- $II \to II - 1$ and $I \to I + 2$ at rate r_II.

Focusing now on the rescaled quantities $(s_t, i_t, ss_t, si_t, ii_t) = (S_t, I_t, SS_t, SI_t, II_t)/N$ and noting the relation $s_t + i_t + 2(ss_t + si_t + ii_t) = 1$, we shall ignore ss_t since it plays no role in the calculations that follow. Also, it will be convenient to use $y_t := s_t + i_t$ instead of s_t. Doing so, the above transitions become

- $i \to i - 1/N$ at rate iN,
- $y \to y - 2/N$ at rate $[r_+(y-i)(y-i-1/N)/2]N = [r_+(y-i)^2/2]N - r_+(y-i)/2$,
- $y \to y - 2/N$, $i \to i - 1/N$ and $si \to si + 1/N$ at rate $r_+(y-i)iN$,
- $y \to y - 2/N$, $i \to i - 2/N$ and $ii \to ii + 1/N$ at rate $[r_+ i(i - 1/N)/2]N = (r_+ i^2/2)N - r_+ i/2$,
- $si \to si - 1/N$ at rate siN,
- $ii \to ii - 1/N$ and $si \to si + 1/N$ at rate $2iiN$,
- $si \to si - 1/N$ and $ii \to ii + 1/N$ at rate λsiN,
- $y \to y + 2/N$ at rate $[r_-((1-y)/2 - (si + ii))]N$,
- $si \to si - 1/N$, $y \to y + 2/N$ and $i \to i + 1/N$ at rate $r_- siN$, and
- $ii \to ii - 1/N$, $y \to y + 2/N$ and $i \to i + 2/N$ at rate $r_- iiN$.

As we did for y_t in Section 2.3, we derive some differential equations that approximate the evolution of (y_t, i_t, si_t, ii_t); since we already have an equation for y_t we focus on i_t, si_t, ii_t. We have

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}(i_t \mid i_t = i) &= -(1 + r_+(y-i) + 2r_+(i - 1/N)/2)i + r_-(si + 2ii) \\
\frac{d}{dt}\mathbb{E}(si_t \mid si_t = si) &= r_+(y-i)i + 2ii - (1 + \lambda + r_-)si \\
\frac{d}{dt}\mathbb{E}(ii_t \mid ii_t = ii) &= r_+ i(i - 1/N)/2 + \lambda si - (2 + r_-)ii
\end{aligned}$$

and as before, in a time step of size $1/N$ the increment in each variable has expected value $O(1/N)$ while its square has expected value $O(1/N^2)$. Adding in the y' equation (2.3.1), this suggests again that in the limit as $N \to \infty$ we should expect the sample paths of (y_t, i_t, si_t, ii_t) to approach solutions to the *mean-field equations* (MFE)

$$\begin{aligned}
y' &= -r_+ y^2 + r_-(1-y) \qquad (2.5.1) \\
i' &= -(1 + r_+ y)i + r_-(si + 2ii) \\
si' &= r_+(y-i)i - (1 + \lambda + r_-)si + 2ii \\
ii' &= r_+ i^2/2 + \lambda si - (2 + r_-)ii
\end{aligned}$$

It is sometimes convenient to replace si with $ip := si + ii$, where the ip stands for "infected partnership". Since $si = ip - ii$, both forms lead to the same solutions. After the change of variables we have

$$\begin{aligned} y' &= -r_+y^2 + r_-(1-y) \\ i' &= -(1+r_+y)i + r_-(ip+ii) \\ ip' &= r_+(y-i/2)i - (1+r_-)ip + ii \\ ii' &= r_+i^2/2 + \lambda ip - (2+r_-+\lambda)ii \end{aligned} \tag{2.5.2}$$

We will often use the shorthand $u' = F(u)$ for the MFE (2.5.1) or (2.5.2), where $u \in \mathbb{R}^4$. In both cases the MFE have the form $y' = f(y), u' = G(y,u)$, where $u \in \mathbb{R}^3$ i.e., the y dynamics does not depend on the other 3 variables, but it does influence them; systems of this form are often referred to as *skew product*. The next three results have natural analogues for the stochastic model, and in fact the analogue of Lemma 2.5.2 shows up in Section 2.7 as Lemma 2.7.1. First we show the domain of interest is an invariant set.

Lemma 2.5.1. *The following set is invariant for the MFE:*

$$\Lambda := \{(y,i,ip,ii) \in \mathbb{R}^4_+ : i \le y \le 1, ii \le ip \le (1-y)/2\}$$

Proof. We examine the boundary and use the form (2.5.2) of the MFE. If $y = 0$ then $y' > 0$ and if $y = 1$ then $y' < 0$, so $[0,1]$ is invariant for y. Let $u = (i, ip, ii)$. If $u = (0,0,0)$ then $u' = (0,0,0)$, so $(0,0,0)$ is invariant for u. If $u \neq (0,0,0)$ and $u_j = 0$ for coordinate j then $u'_j > 0$ (note for ip' that since $i \le y$, if $i > 0$ then $y - i/2 > 0$). If $i = y \neq 0$ then since $ip + ii \le (1-y)$, $i' \le -y - r_+y^2 + r_-(1-y) = -y + y' < y'$. If $i = y = 0$ then $i' \le -y + y' = y'$ and since $y' > 0$, $i'' \le -y' + y'' < y''$. For the remainder we may assume $i < y$. If $ii = ip \neq 0$ then $ii' = r_+i^2/2 - (2+r_-)ip \le r_+(y-i/2)i - (2+r_-)ip < ip'$ while if $ii = ip = 0$ then we may assume $i > 0$ in which case $ii' = r_+i^2/2 < r_+(y-i/2)i = ip'$.

□

Written in the form (2.5.2) the MFE have a useful monotonicity property which is described in the following lemma.

Lemma 2.5.2. *Let $(y(t), u(t))$ and $(y(t), v(t))$ be solutions to the MFE written in (y,i,ip,ii) coordinates, and say that $u \le v \Leftrightarrow u_j \le v_j \ \forall j \in \{1,2,3\}$. If $u(0) \le v(0)$*

then $u(t) \leq v(t)$ *for* $t > 0$.

Proof. Since trajectories are continuous it suffices to check that if $u \leq v$, $u \neq v$ and $u_j = v_j$ then $u_j' < v_j'$. Referring to (2.5.2), i' increases with ip and ii, ip' increases with i and ii (note $\partial_i(y - i/2)i = y - i$ and $i \leq y$) and ii' increases with i and ip. □

For what follows we set $y = y^*$ in which case the MFE are three-dimensional. Since Λ is invariant,

$$\Lambda^* := \{(y, u) \in \Lambda : y = y^*\}$$

is also invariant. Since $\Lambda^* \cong \{(i, ip, ii) \in \mathbb{R}^3_+ : i \leq y^*, ii \leq ip \leq (1 - y^*)/2\}$ is three-dimensional, elements of Λ^* are usually written as a three-vector in either (i, si, ii) or (i, ip, ii) coordinates.

Lemma 2.5.3. *For the MFE* (2.5.2) *with* $y = y^*$*, written as* $u' = G(y^*, u)$*, say that* $u = (i, ip, ii)$ *is* increasing *if every coordinate of* $G(y^*, u)$ *is* > 0*. For the MFE with* $y = y^*$ *and any solution* $u(t)$*,*

- *if* $(0, 0, 0)$ *is the only equilibrium then* $u(t) \to (0, 0, 0)$ *as* $t \to \infty$*, and*
- *if there is a unique equilibrium* $u^* \neq (0, 0, 0)$ *and a sequence of non-zero increasing states tending to* $(0, 0, 0)$*, then for* $u(0) \neq (0, 0, 0)$*,* $u(t) \to u^*$ *as* $t \to \infty$*.*

Proof. Defining $\overline{u} := (y^*, (1 - y^*)/2, (1 - y^*)/2)$, $\overline{u} \geq v$ for all $v \in \Lambda^*$, so letting $\overline{u}(t)$ be the solution to the MFE with $\overline{u}(0) = \overline{u}$, for $s \geq 0$, $\overline{u}(0) \geq \overline{u}(s)$. Since $y = y^*$, by monotonicity (Lemma 2.5.2) $\overline{u}(t) \geq \overline{u}(t + s)$ for $t > 0$, so $\overline{u}(t)$ is non-increasing in t. Since Λ^* is compact, $\lim_{t\to\infty} \overline{u}(t)$ exists and by continuity of the MFE is an equilibrium. If $(0, 0, 0)$ is the only equilibrium, then since $\overline{u}(t) \geq (0, 0, 0)$, $\overline{u}(t) \to (0, 0, 0)$ as $t \to \infty$, so for any solution $v(t)$, since $\overline{u}(0) \geq v(0)$, $\overline{u}(t) \geq v(t)$ for $t > 0$, and since $v(t) \geq (0, 0, 0)$, $v(t) \to (0, 0, 0)$.

If $u(0)$ is increasing then $u(0) \neq (0, 0, 0)$ and by continuity of the MFE there is $\epsilon > 0$ so that $u(s) \geq u(0)$ for $0 \leq s \leq \epsilon$. By monotonicity $u(t + s) \geq u(t)$ for $0 \leq s \leq \epsilon$ and if $(k - 1)\epsilon \leq s \leq k\epsilon$, by iterating at most k times $u(t + s) \geq u(t)$, so $u(t)$ is increasing for all time. As in the previous case, $\lim_{t\to\infty} u(t)$ exists and is an equilibrium which in this case is not $(0, 0, 0)$. If there is a unique equilibrium $u^* \neq (0, 0, 0)$, and if for any non-zero solution $v(t)$ there is $T > 0$ so that $v(T) \geq u$ for some increasing u, then setting $u(T) = u$, since $\overline{u}(t) \geq v(t) \geq u(t)$ for $t \geq T$ and $\lim_{t\to\infty} \overline{u}(t) = \lim_{t\to\infty} u(t) = u^*$ it follows that $\lim_{t\to\infty} v_t = u^*$. If $v(0) \neq (0, 0, 0)$ then

for $t > 0$, $v_j(t) > 0$ in each coordinate j; this follows from the fact that for $j = 1, 2, 3$, $v_j' \geq -Cv_j$ for some C, and if $v_j = 0$ but $v_k > 0$ for some $k \neq j$ then $v_j' > 0$. Thus, fixing $T > 0$, if $v(0) \neq (0, 0, 0)$ then since $\epsilon := \min_j v_j(T) > 0$, if there is a sequence of increasing states tending to $(0, 0, 0)$ there is an increasing state u with $\max_j u_j \leq \epsilon$ and thus $v(T) \geq u$, as desired. □

As the next result shows, on Λ^* the MFE have a simple dynamics with a bifurcation at $R_0 = 1$. Since we refer back to quantities from Section 2.4, in this proof we mostly use (i, si, ii) coordinates.

Theorem 2.5.4. *For the MFE,*

- *if $R_0 \leq 1$ there is the unique equilibrium $(0, 0, 0)$ which is attracting on Λ^* and*
- *if $R_0 > 1$ there is a unique positive equilibrium (i^*, s^*, ii^*) satisfying $\Delta(i^*) = 0$ which is attracting on $\Lambda^* \setminus \{(0, 0, 0)\}$.*

Proof. By Lemma 2.5.3 it is enough to show that if $R_0 \leq 1$ then $(0, 0, 0)$ is the only equilibrium, and that if $R_0 > 1$ there is a unique equilibrium $(i^*, si^*, ii^*) \neq (0, 0, 0)$ satisfying $\Delta(i^*) = 0$, and a sequence of increasing states converging to $(0, 0, 0)$. Treating si, ii as a separate system with input function i, we have the non-homogenous linear system

$$\begin{pmatrix} si' \\ ii' \end{pmatrix} = \begin{pmatrix} -a & 2 \\ \lambda & -b \end{pmatrix} \begin{pmatrix} si \\ ii \end{pmatrix} + r_+ i \begin{pmatrix} (y^* - i) \\ i/2 \end{pmatrix}$$

or, in matrix form, $v' = Kv + Li$, with $v = (si, ii)^\top$, $K = \left(\begin{smallmatrix} -a & 2 \\ \lambda & -b \end{smallmatrix}\right)$ and $L = r_+((y^* - i), i/2)^\top$, whose solution is given by

$$v(t) = \Phi(t)v(0) + \int_0^t \Phi(t - s)L(s)i(s)ds \tag{2.5.3}$$

where $\Phi(t) = \exp(Kt)$ is the solution of the associated homogenous system - note that $\Phi(t)$ is the restriction of the transition semigroup for the continuous-time Markov chain from Figure 2.1 to the states B and C. Substituting the solution for the si, ii system into the equation for i, we have

$$i'(t) = -(1 + r_+ y^*)i(t) + r_-(1, 2)\left[\Phi(t)v(0) + \int_0^t \Phi(t - s)L(s)i(s)ds\right] \tag{2.5.4}$$

where $(1, 2)$ is a row vector that multiplies the column vector in the square brackets. This equation depends only on i, the initial values $v(0) = (si(0), ii(0))^\top$ and the so-

lution matrix $\Phi(t)$.

Linearizing (2.5.4) around $(i, si, ii) = (0, 0, 0)$ and using the ansatz $i(t) = \exp(\mu t)$ we obtain

$$\mu e^{\mu t} = -(1 + r_+ y^*)e^{\mu t} + r_-(1, 2)\left[\Phi(t)v_0 + \int_0^t \Phi(t - s)e^{\mu s} ds\right] L_0$$

where $L_0 = r_+(y^*, 0)^\top$, and using $\Phi(t) = \exp(Kt)$ the integral in the square brackets is

$$e^{Kt}\int_0^t e^{(\mu I - K)s} ds = e^{Kt}(\mu I - K)^{-1}(e^{(\mu I - K)t} - I) = (\mu I - K)^{-1}(e^{\mu t} - e^{Kt})$$

where I is the identity matrix. Letting $t \to \infty$ and noting $\Phi(t) = e^{Kt} \to 0$ since K is a stable matrix, we obtain the eigenvalue equation

$$\mu = -(1 + r_+ y^*) + r_-(1, 2)(\mu I - K)^{-1} L_0$$

which, expanding, is

$$\mu = -(1 + r_+ y^*) + r_- \frac{\mu + b + 2\lambda}{(\mu + b)(\mu + a) - 2\lambda} r_+ y^* \tag{2.5.5}$$

and setting $\mu = 0$ gives the equation

$$1 = \frac{r_+ y^*}{1 + r_+ y^*} \frac{r_-}{ab - 2\lambda}(b + 2\lambda)$$

which, comparing to (2.4.1), is exactly $R_0 = 1$. Recalling that $ab - 2\lambda > 0$,

$$\begin{aligned} \frac{d}{d\mu}\left(\frac{\mu + b + 2\lambda}{(\mu + b)(\mu + a) - 2\lambda}\right) &= \frac{(\mu + b)(\mu + a) - 2\lambda - (\mu + b + 2\lambda)(2\mu + b + a)}{[(\mu + b)(\mu + a) - 2\lambda]^2} \\ &= \frac{-2\lambda - [(\mu + b)^2 + 2\lambda(2\mu + b + a)]}{[(\mu + b)(\mu + a) - 2\lambda]^2} \end{aligned}$$

is negative when $\mu \geq 0$. Setting $\mu = 0$ in (2.5.5), the right-hand side is positive if $R_0 > 1$, so since both sides are continuous in μ, the left-hand side is equal to 0 at $\mu = 0$ and increases unboundedly as μ increases and the right-hand side decreases with μ it follows that (2.5.5) has a positive solution $\mu > 0$ when $R_0 > 1$.

To obtain the increasing states mentioned in Lemma 2.5.3 we show that for $R_0 > 1$ the unstable eigenvector of the linearized system near $(0,0,0)$ is strictly positive when viewed in (i, ip, ii) coordinates; we can then take for the initial states small multiples of the eigenvector. To show the eigenvector is strictly positive, linearize (2.5.3) around $(i, si, ii) = (0,0,0)$ with input $i(t) = \exp(\mu t)$, substitute the solution form $v(t) = v\exp(\mu t)$ and let $t \to \infty$ to obtain $v = (\mu I - K)^{-1}L_0$ which has positive entries, which implies that in (ip, ii) coordinates it also has positive entries.

It remains to look for non-zero equilibria. Focusing again on (2.5.4), as our steady state assumption we suppose the system was started in the distant past and has remained in equilibrium up to the present time. Since $\Phi(t) \to 0$ as $t \to \infty$ we ignore $\Phi(t)v(0)$, and letting $\Phi_\infty = \int_0^\infty \Phi(s)ds = -K^{-1}$, $\int_0^t \Phi(t-s)L(s)i(s)ds$ becomes $\int_{-\infty}^t \Phi(t-s)L^\dagger i^\dagger ds = \Phi_\infty L^\dagger i^\dagger$ where $L^\dagger = r_+((y^* - i^\dagger), i^\dagger/2)^\top$ and $i^\dagger$ are the equilibrium values, and we obtain

$$(1 + r_+ y^*) = r_-(1,2)\Phi_\infty L^\dagger$$

Notice that $r_-(1,2)\Phi_\infty$ returns the expected number of infectious singles that result from an SI or an II partnership upon breakup, so we have $r_-(1,2)\Phi_\infty = (1+\Delta_{SI}, 2+\Delta_{II})$ and

$$\begin{aligned}(1 + r_+ y^*) &= r_+[(y^* - i^\dagger)(1 + \Delta_{SI}) + (i^\dagger/2)(2 + \Delta_{II})] \\ &= r_+ y^* + r_+[(y^* - i^\dagger)\Delta_{SI} + (i^\dagger/2)\Delta_{II}]\end{aligned}$$

and subtracting $r_+ y^*$, $1 = r_+(y^* - i^\dagger)\Delta_{SI} + r_+(i^\dagger/2)\Delta_{II}$ which comparing with (2.4.4) is exactly the equation $\Delta(i^\dagger) = 0$, as desired. By Lemma 2.4.3, we have the unique solution $i^\dagger = i^*$ if $R_0 > 1$, and there is no positive solution when $R_0 \leq 1$. Using the steady state assumption and (2.5.3) gives $(si^\dagger, ii^\dagger) = \Phi_\infty L^\dagger i^\dagger$, that is, $si^\dagger, ii^\dagger$ are uniquely determined by $i^\dagger$. This proves uniqueness of the non-zero equilibrium when $R_0 > 1$ and uniqueness of $(0,0,0)$ as an equilibrium when $R_0 \leq 1$. □

Remark 2.5.5. *Setting $y = y^*$ in* (2.5.1) *and writing the remaining equations in*

matrix form, we have $u' = Au$ *with* $u = (i, si, ii)^\top$ *and*

$$A = \begin{pmatrix} -(1 + r_+ y^*) & r_- & 2r_- \\ r_+(y^* - i) & -a & 2 \\ r_+ i & \lambda & -b \end{pmatrix}$$

that depends on u*. Using the technique of [50], if we evaluate* A *at* $i = 0$ *and write it as* $F - V$ *with*

$$F = \begin{pmatrix} 0 & 0 & 0 \\ r_+ y^* & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; V = \begin{pmatrix} (1 + r_+ y^*) & -r_- & -2r_- \\ 0 & a & -2 \\ 0 & -\lambda & b \end{pmatrix}$$

and define $R_0 = \rho(FV^{-1})$ *where* ρ *is the spectral radius, then it can be verified that this definition of* R_0 *coincides with the one given in* (2.2.2)*. Then, according to Theorem 2 of [50],* $R_0 < 1$ *implies* $(0, 0, 0)$ *is locally asymptotically stable, while* $R_0 > 1$ *implies it is unstable.*

2.6 Approximation by the Mean-Field Equations

In this section we show how to approximate the sample paths of (y_t, i_t, si_t, ii_t) with solutions to the MFE (2.5.1), and use this to get some control on y_t. Unless otherwise noted, for a vector, $|\cdot|$ denotes the ℓ^∞ norm, that is, $|u| = \max_i |u_i|$. We begin with a useful definition.

Definition 2.6.1. *An event* A *depending on a parameter* n *is said to hold* with high probability *or whp in* n *if there exists* $\gamma > 0$ *and* n_0 *so that* $\mathbb{P}(A) \geq 1 - e^{-\gamma n}$ *when* $n \geq n_0$*.*

When possible, probability estimates are given more or less explicitly, but we will occasionally use this definition to reduce clutter, especially in Section 2.7. We begin with a well-known large deviations result for Poisson random variables; since it is not hard to prove, we supply the proof. For a reference to large deviations theory see Section 1.9 in [13].

Lemma 2.6.2. *Let X be Poisson distributed with mean μ, then*

$$\begin{aligned}\mathbb{P}(X > (1+\delta)\mu) &\leq e^{-\delta^2\mu/4} \text{ for } 0 < \delta \leq 1/2,\\ \mathbb{P}(X < (1-\delta)\mu) &\leq e^{-\delta^2\mu/2} \text{ for } \delta > 0\end{aligned}$$

Proof. We deal separately with $X > (1+\delta)\mu$ and $X < (1-\delta)\mu$. For $t > 0$ and using Markov's inequality we have

$$\mathbb{P}(X > (1+\delta)\mu) = \mathbb{P}(e^{tX} > e^{(1+\delta)t\mu}) \leq \mathbb{E}e^{tX}e^{-(1+\delta)t\mu}$$

Notice that

$$\mathbb{E}e^{tX} = \sum_{k\geq 0} e^{tk}e^{-\mu}\frac{\mu^k}{k!} = e^{-\mu}\sum_{k\geq 0}\frac{(e^t\mu)^k}{k!} = e^{-\mu}e^{e^t\mu} = \exp((e^t-1)\mu)$$

so $\mathbb{E}e^{tX}e^{-(1+\delta)t\mu} = \exp(\mu(e^t - 1 - (1+\delta)t))$. Minimizing $e^t - 1 - (1+\delta)t$ gives $t = \log(1+\delta)$ and thus $(1+\delta) - 1 - (1+\delta)\log(1+\delta) = \delta - (1+\delta)\log(1+\delta)$. Since $\log(1+\delta) \geq \delta - \delta^2/2$ this is at most $\delta - (1+\delta)(\delta - \delta^2/2) = -\delta^2/2 + \delta^3/2$ which is $\leq -\delta^2/4$ for $0 < \delta \leq 1/2$.

For the other direction we take a similar approach. For $t > 0$ and using Markov's inequality we have

$$\mathbb{P}(X < (1-\delta)\mu) = \mathbb{P}(e^{-tX} > e^{-(1-\delta)t\mu}) \leq \mathbb{E}e^{-tX}e^{(1-\delta)t\mu}$$

and using $\mathbb{E}e^{-tX} = \exp((e^{-t}-1)\mu)$ the right-hand side above is $\exp(\mu(e^{-t} - 1 + (1-\delta)t))$. Minimizing $e^{-t} - 1 + (1-\delta)t$ gives $-t = \log(1-\delta)$ and thus $(1-\delta) - 1 - (1-\delta)\log(1-\delta) = -\delta - (1-\delta)\log(1-\delta)$. Since $\log(1-\delta) \geq -\delta - \delta^2/2$ this is at most $-\delta + (1-\delta)(\delta + \delta^2/2) = -\delta^2/2 - \delta^3/2 \leq -\delta^2/2$. □

For the next three results we use the notation $u_t = (y_t, i, si_t, ii_t)$. First we give an a priori bound on the change in u_t over a short period of time.

Lemma 2.6.3. *Let $u_t = (y_t, i_t, si_t, ii_t)$. There are constants $C, \gamma > 0$ so that for all $h > 0$ and fixed t,*

$$\mathbb{P}(\sup_{t\leq s\leq t+h} |u_s - u_t| \leq Ch) \geq 1 - e^{-\gamma Nh}$$

Proof. Looking to the transitions listed in Section 2.5, jumps in u_t are of size $\leq 2/N$ and occur at total rate $\leq MN$ for some $M > 0$ that depends only on parameters. Thus in a time step $h > 0$ the number of events affecting u_t is stochastically bounded above by a Poisson random variable X with mean MNh, so if $X \leq x$ then $|u_s - u_t| \leq 2x/N$ for all $s \in [t, t+h]$. By Lemma 2.6.2, $\mathbb{P}(X > (1+\delta)MNh) \leq e^{-\delta^2 MNh/4}$ for $0 < \delta \leq 1/2$. Taking $\delta = 1/4$ and $C = 2(1+\delta)M$, $\gamma = \delta^2 M/4$ completes the proof. □

Let $u' = F(u)$ denote the MFE (2.5.1). As N becomes large, for small $h > 0$ we expect that with probability tending to 1, $u_{t+h} = u_t + hF(u_t) + o(h)$. Using Lemma 2.6.3 and re-using the estimate from Lemma 2.6.2 we obtain a quantitative bound on the remainder.

Lemma 2.6.4. *Let $u_t = (y_t, i_t, si_t, ii_t)$. For each $\epsilon > 0$ there are constants $C, \gamma > 0$ so that for small enough $h > 0$,*

$$\mathbb{P}(|u_{t+h} - u_t - hF(u)| \leq \epsilon h) \geq 1 - Ce^{-\gamma Nh}$$

Proof. Let $Q_j(u)$, $j = 1, ..., 10$, denote the transition rates of the ten transitions introduced in Section 2.5, as a function of u, and let $X_j(t,h)$ denote the number of type j transitions occurring in the time interval $[t, t+h]$. For each j, $Q_j(u) = Nq_j(u) + R_j(u)$ where $q_j(u)$ is a quadratic function of u and $R_j(u)$ is a remainder that satisfies $|R_j(u)| \leq M$ for some $M > 0$ and all $u \in [0,1]^4$. It is easily verified that if $u_t = u$ and $X_j(t,h) = Nq_j(u)h$ for each j then $u_{t+h} = u + hF(u)$. Since each transition changes u by at most $2/N$, it is therefore enough to show that there are constants $C, \gamma > 0$ so that for each j, small enough $h > 0$, and all u,

$$\mathbb{P}(|X_j(t,h) - Nq_j(u)h| \leq \epsilon Nh/20 \mid u_t = u) \geq 1 - Ce^{-\gamma Nh}$$

Since the domain of $q_j(u)$ is a subset of $[0,1]^4$ and thus bounded it follows that q_j is bounded and Lipschitz continuous i.e., for some $L > 0$ and all v, u in the domain of q_j, $q_j(u) \leq L$ and $|q_j(v) - q_j(u)| \leq L|v-u|$, and in particular, $|Q_j(v) - Q_j(u)| \leq NL|v-u| + 2M$; for what follows, take $L \geq \epsilon$. Let $A(t,h)$ be the event

$$\{\sup_{t \leq s \leq t+h} |u_s - u_t| \leq C_1 h\},$$

from Lemma 2.6.3, then on the event $\{u_t = u\} \cap A(t,h)$,

$$\sup_{t\le s\le t+h} |Q_j(u_s) - Nq_j(u)| \le \sup_{t\le s\le t+h} |Q_j(u_s) - Q_j(u)| + |Q_j(u) - Nq_j(u)| \le N(LC_1 h + 3M/N)$$

For ease of notation let $q = q_j(u)$ and let $r = LC_1 h + 3M/N$, and note that $r \to 0$ as $\max(h, 1/N) \to 0$. Then, on $\{u_t = u\} \cap A(t,h)$, $X_j(t,h)$ is stochastically bounded above and below respectively by Poisson random variables with means $Nh(q+r)$ and $Nh(q-r)$, so from Lemma 2.6.2 it follows that for $0 < \delta \le 1/2$,

$$\mathbb{P}(\{|X_j(t,h) - Nhq| \le Nh(q\delta + r(1+\delta))\} \cap \{u_t = u\} \cap A(t,h)) \ge 1 - 2e^{-Nh(q-r)\delta^2/4} \tag{2.6.1}$$

Recalling that $q \le L$, let $h, \delta, 1/N > 0$ be chosen small enough that $L\delta + r(1+\delta) \le \epsilon/20$, then $Nh(q\delta + r(1+\delta)) \le \epsilon Nh/20$. To bound the probability uniformly in q we split into two cases according as $q \ge q\delta + r(1+\delta)$ or not i.e., as $q \ge r(1+\delta)/(1-\delta)$ or not. If $q \ge r(1+\delta)/(1-\delta)$ then letting $\gamma_1 = r[(1+\delta)/(1-\delta) - 1]\delta^2/4$ which is > 0 it follows that $Nh(q-r)\delta^2/4 \ge \gamma_1 Nh$. If $q < q\delta + r(1+\delta)$ the lower bound on $X_j(t,h) - Nhq$ is trivial and so in that case

$$\mathbb{P}(\{|X_j(t,h) - Nhq| \le Nh(q\delta + r(1+\delta))\} \cap \{u_t = u\} \cap A(t,h)) \ge 1 - e^{-Nh(q+r)\delta^2/4}$$

Letting $\gamma_2 = r\delta^2/4$ which is > 0 it follows that $Nh(q+r)\delta^2/4 \ge \gamma_2 Nh$. Letting γ_3 be such that $\mathbb{P}(A(t,h)) \ge 1 - e^{-\gamma_3 Nh}$ and letting $\gamma = \min(\gamma_1, \gamma_2, \gamma_3)$ and $C = 3$ completes the proof. □

Using the above estimate we obtain finite-time control on the evolution of u_t, as N becomes large.

Proposition 2.6.5. *Let $u_t = (y_t, i_t, si_t, ii_t)$. For each $\epsilon, T > 0$ there are constants $\delta, C, \gamma > 0$ so that from any initial condition u_0 and any solution $u(t)$ to the MFE (2.5.1) satisfying $|u_0 - u(0)| \le \delta$,*

$$\mathbb{P}(\sup_{0\le t\le T} |u_t - u(t)| \le \epsilon) \ge 1 - Ce^{-\gamma N}$$

Proof. The proof is analogous to the proof in numerical analysis that the Euler method is $O(h)$ accurate. Fix $h = T/M$ for integer M and define events $A_1, ..., A_m$

as follows: $A_1 = B_1 \cap D_1$ and given A_{j-1}, $A_j = A_{j-1} \cap B_j \cap D_j$ where

$$B_j = \{ \sup_{h(j-1) \le t \le hj} |u_t - u_{hj}| \le C_1 h \}$$

is the event from Lemma 2.6.3 and

$$D_j = \{ |u_{hj} - u_{h(j-1)} - hF(u_{h(j-1)})| \le \mu h \}$$

is the event from Lemma 2.6.4, for $\mu > 0$ to be chosen. If $\mu, h > 0$ are fixed and h is small enough then there are constants $C, \gamma > 0$ so that $\mathbb{P}(B_j \cap D_j) \ge 1 - (C/M)e^{-\gamma N}$, and since $A_M = \bigcap_{j=1}^{M} (B_j \cap D_j)$, $\mathbb{P}(A_M) \ge 1 - Ce^{-\gamma N}$. For $j = 1, ..., M$ let

$$E_j = \sup_{\omega \in A_j} |u_{hj}(\omega) - u(hj)|$$

where ω denotes an element of the probability space for the partner model. Letting $u' = F(u)$ denote (2.5.1) we have

$$u(hj) - u(h(j-1)) = \int_{h(j-1)}^{hj} F(u(s))ds$$

Since $F(u)$ is quadratic in u and its domain is bounded, it is bounded and Lipschitz continuous i.e., for some $L > 0$ and all u, v in the domain, $|F(u)| \le L$ and $|F(v) - F(u)| \le L|v - u|$. From the first inequality it follows that $|u(s) - u(h(j-1))| \le L(s - h(j-1))$ for $s \ge h(j-1)$ and from this and the second inequality it follows that

$$\begin{aligned}
|u(hj) - u(h(j-1)) - hF(u(h(j-1)))| &= \left| \int_{h(j-1)}^{hj} (F(u(s)) - F(u(h(j-1))))ds \right| \\
&\le \int_{h(j-1)}^{hj} |F(u(s)) - F(u(h(j-1)))|ds \\
&\le \int_{h(j-1)}^{hj} L|u(s) - u(h(j-1))|ds \\
&\le \int_{h(j-1)}^{hj} L^2(s - hj)ds = L^2 \int_0^h sds = L^2h^2/2
\end{aligned}$$

Also,

$$\begin{aligned}|u_{hj}-u(hj)| &= |u_{hj}-u_{h(j-1)}-hF(u_{h(j-1)})+u_{h(j-1)}-u(h(j-1))\\ &\quad +hF(u_{h(j-1)})-hF(u(h(j-1)))+u(h(j-1))+hF(u(h(j-1)))-u(hj)|\\ &\leq |u_{hj}-u_{h(j-1)}-hF(u_{h(j-1)})|+|u_{h(j-1)}-u(h(j-1))|\\ &\quad +|hF(u_{h(j-1)})-hF(u(h(j-1)))|+|u(hj)-u(h(j-1))-hF(u(h(j-1)))|\end{aligned}$$

so using the definition of A_j, letting $E_0 := |u_0 - u(0)| \leq \delta$ and using once more Lipschitz continuity of F it follows that for $j = 1, ..., M$,

$$E_j \leq \mu h + E_{j-1} + hLE_{j-1} + L^2h^2/2 = (1+hL)E_{j-1} + h(\mu + hL^2/2)$$

Setting $q = (1+hL)$ and $r = \mu + hL^2/2$ and iterating the inequality $E_j \leq qE_{j-1} + hr$ we find $E_M \leq q^M E_0 + [(q^M-1)/(q-1)]hr \leq q^M[E_0 + hr/(q-1)] = (1+hL)^M[E_0 + hr/(hL)] = (1+LT/M)^M[E_0 + r/L] \leq e^{LT}[E_0 + r/L] \leq e^{LT}[\delta + r/L]$ and the same inequality holds for all $E_j, j = 1, ..., M$. Since on A_j, $|u_s - u_{hj}| \leq C_1 h$ for $h(j-1) \leq s \leq hj$, on A_M we find for $j = 1, ..., M$ and $h(j-1) \leq s \leq hj$ that

$$\begin{aligned}|u_s - u(s)| &\leq |u_s - u_{hj}| + |u_{hj} - u(hj)| + |u(hj) - u(s)|\\ &\leq C_1 h + E_j + Lh \leq h(C_1 + L) + e^{LT}[\delta + r/L]\end{aligned}$$

and taking $h, \mu, \delta > 0$ small enough, this is $\leq \epsilon$. □

Our first application of Proposition 2.6.5 is to control y_t.

Lemma 2.6.6. *For each $\epsilon > 0$ there are constants $C, T, \gamma > 0$ so that from any value $y_0 \in [0,1]$,*

$$\mathbb{P}(\sup_{T \leq t \leq e^{\gamma N}} |y_t - y^*| \leq \epsilon) \geq 1 - Ce^{-\gamma N}$$

Moreover, if $|y_0 - y^*| \leq 2\epsilon/3$ we may take $T = 0$.

Proof. Let $y' = f(y)$ denote the y' equation in (2.5.1) and let $\phi(t,y)$, $\phi : [0,1] \times \mathbb{R}_+ \to [0,1]$ denote the flow for this equation i.e., the unique function satisfying $\partial_t \phi(t,y) = f(\phi(t,y))$ and $\phi(0,y) = y$ for each (t,y) in its domain. Since $\phi(t,0) \leq \phi(t,y) \leq \phi(t,1)$ and $\lim_{t\to\infty} \phi(t,y) = y^*$ for each $y \in [0,1]$, for each $\epsilon > 0$ there is $T > 0$ so that $|\phi(T,y) - y^*| \leq \epsilon/3$ for all $y \in [0,1]$. Letting $y(t) = \phi(t, y_0)$ and using Proposition 2.6.5, there are constants $C_1, \gamma_1 > 0$ depending on ϵ but not on y_0 so that with

probability $\geq 1 - C_1 e^{-\gamma_1 N}$, $|y_T - y^*| \leq |y_T - y(T)| + |y(T) - y^*| \leq \epsilon/3 + \epsilon/3 = 2\epsilon/3$. Then, for $t \geq 0$ and $y \in [y^* - (2\epsilon/3), y^* + (2\epsilon/3)]$,

$$y^* - (2\epsilon/3) \leq \phi(t, y^* - (2\epsilon/3)) \leq \phi(t, y) \leq \phi(t, y^* + (2\epsilon/3)) \leq y^* + (2\epsilon/3)$$

and since all solutions approach y^* there is $h > 0$ so that $\phi(h, y^* - 2\epsilon/3) \geq y^* - \epsilon/3$ and $\phi(h, y^* + 2\epsilon/3) \leq y^* + \epsilon/3$. Thus for the given value of h and any solution $y(t)$ of $y' = f(y)$, if $|y(T) - y^*| \leq 2\epsilon/3$ then $|y(t) - y^*| \leq 2\epsilon/3$ for $t \geq T$ and $|y(T+h) - y^*| \leq \epsilon/3$. Given y_T such that $|y_T - y^*| \leq 2\epsilon/3$ and setting $y(T) = y_T$, by Proposition 2.6.5 there are constants $C_2, \gamma_2 > 0$ so that $\sup_{T \leq t \leq T+h} |y_t - y(t)| \leq \epsilon/3$ with probability $\geq 1 - C_2 e^{-2\gamma_2 N}$, in which case

$$\sup_{T \leq t \leq T+h} |y_t - y^*| \leq \sup_{T \leq t \leq T+h} |y_t - y(t)| + \sup_{T \leq t \leq T+h} |y(t) - y^*| \leq \epsilon/3 + 2\epsilon/3 = \epsilon$$

and $|y_{T+h} - y^*| \leq |y_{T+h} - y(T+h)| + |y(T+h) - y^*| \leq \epsilon/3 + \epsilon/3 = 2\epsilon/3$ with the same probability. Iterating this for $e^{\gamma_2 N}/h$ time steps we find that

$$\sup_{T \leq t \leq e^{\gamma_2 N}} |y_t - y^*| \leq \max_{i \in \{1, \ldots, e^{\gamma_2 N}/h\}} \sup_{T + (i-1)h \leq t \leq T + ih} |y_t - y^*| \leq \epsilon$$

with probability $\geq 1 - (C_2/h) e^{\gamma_2 N} e^{-2\gamma_2 N} = 1 - (C_2/h) e^{-\gamma_2 N}$, then choose $C = C_1 + C_2/h$ and $\gamma = \min(\gamma_1, \gamma_2)$. Note that if $|y_0 - y^*| \leq 2\epsilon/3$, the iteration step is immediately applicable, in which case we may take $T = 0$. □

2.7 Macroscopic Behaviour

In this section we prove the macroscopic side of Theorem 2.2.2 i.e., when $|V_0| \geq \epsilon N$. We begin with the analogue of Lemma 2.5.2 for the partner model, which we refer to later on as monotonicity. As for the MFE, define $ip_t := si_t + ii_t$.

Lemma 2.7.1. *Let $\leq$ denote the partial order on $\mathbb{R}^3$ given by $u \leq v \Leftrightarrow u_j \leq v_j$, $\forall j \in \{1, 2, 3\}$, and let $u_t = (i_t, ip_t, ii_t)$. If $(V_t^{(1)}, E_t^{(1)})$ and $(V_t^{(2)}, E_t^{(2)})$ are two copies of the partner model with $E_0^{(1)} = E_0^{(2)}$ and $V_0^{(1)} \subseteq V_0^{(2)}$ then with respect to the coupling given by the graphical construction, $E_t^{(1)} = E_t^{(2)}$ and $V_t^{(1)} \subseteq V_t^{(2)}$ for $t > 0$ and correspondingly $y_t^{(1)} = y_t^{(2)}$ and $u_t^{(1)} \leq u_t^{(2)}$.*

Proof. If $E_0^{(1)} = E_0^{(2)}$ then $E_t^{(1)} = E_t^{(2)} =: E_t$ for $t > 0$. Given $\{E_t : t \geq 0\}$, the only transitions affecting $V_t^{(1)}$ and $V_t^{(2)}$ are recovery of infectious sites and transmission

from infectious to healthy sites along open edges, both of which preserve the order $V_t^{(1)} \subseteq V_t^{(2)}$. The equality $y_t^{(1)} = y_t^{(2)}$ follows directly from $|E_t^{(1)}| = |E_t^{(2)}|$ and the inequality $u_t^{(1)} \leq u_t^{(2)}$ follows directly from $V_t^{(1)} \subseteq V_t^{(2)}$. □

Using Proposition 2.6.5 and Lemma 2.7.1 we can prove the macroscopic part of Theorem 2.2.2 when $R_0 \leq 1$. In this section u_t will generally refer to (i_t, si_t, ii_t) or (i_t, ip_t, ii_t), with y_t written separately.

Proposition 2.7.2. *If $R_0 \leq 1$, for each $\epsilon > 0$ there are constants $C, T, \gamma > 0$ so that, from any initial configuration, with probability $\geq 1 - Ce^{-\gamma N}$, $|V_T| \leq \epsilon N$.*

Proof. By Lemma 2.7.1 it is enough to show the result holds when $V_0 = V$ i.e., everyone is initially infectious; in this case $y_0 = 1 - 2E_0/N$, $i_0 = y_0$ and $ip_0 = ii_0 = (1 - y_0)/2$. Let $u_t = (i_t, ip_t, ii_t)$ and let $(y(t), u(t))$ be the solution to the MFE with $y(0) = y_0$ and $u(0) = u_0$. By Lemma 2.6.6 and Proposition 2.6.5, for each $\delta > 0$ there are constants $C_1, T_1, \gamma_1 > 0$ so that with probability $\geq 1 - C_1 e^{-\gamma_1 N}$, $|y_{T_1} - y^*| \leq \delta$ and $|u_{T_1} - u(T_1)| \leq \delta$, so with the same probability $|(y_{T_1}, u_{T_1}) - (y^*, u(T_1))| \leq \delta$.

Recall the set Λ^* and let $(y^*, \overline{u}(t))$ be the solution to the MFE with $\overline{u}(0) = (y^*, (1 - y^*)/2, (1 - y^*)/2)$. As shown in the proof of Lemma 2.5.3, $\overline{u}(t)$ decreases to an equilibrium. Since $R_0 \leq 1$, $(0, 0, 0)$ is the only equilibrium, so $\overline{u}(t) \to (0, 0, 0)$ as $t \to \infty$. Moreover, $\overline{u}(0) \geq v$ for each $v \in \Lambda^*$ so for any solution $(y^*, u(t))$, $\overline{u}(0) \geq u(0)$. By Lemma 2.5.2, $\overline{u}(t) \geq u(t)$ for $t \geq 0$, so there is T_2 not depending on $u(0)$ so that $|u(T_2)| \leq \epsilon/2$. Using Proposition 2.6.5, there are constants $C_2, \gamma_2, \delta > 0$ not depending on $u(0)$ so that with probability $\geq 1 - C_2 e^{-\gamma_2 N}$, if $|(y_0, u_0) - (y^*, u(0))| \leq \delta$ then $|u_{T_2}| \leq |u(T_2)| + |u_{T_2} - u(T_2)| \leq \epsilon/2 + \epsilon/2 = \epsilon$. Letting $T = T_1 + T_2$, $C = C_1 + C_2$ and $\gamma = \min(\gamma_1, \gamma_2)$ and combining the two steps completes the proof. □

Using similar ideas we can prove the macroscopic part of Theorem 2.2.2 when $R_0 > 1$. Before showing approach to equilibrium, we first have to show long time survival of the infection, and to do that, we need the following result concerning the MFEs.

Lemma 2.7.3. *Suppose $R_0 > 1$ and let $v \in \mathbb{R}^3$ with $|v| = 1$ be an unstable eigenvector of the MFEs on Λ^* as given in the proof of Theorem 2.5.4, written in (i, ip, ii) coordinates. For $0 < \delta' \leq \delta$ let $(y(t), u(t))$ be a solution to the MFE with $|y(0) - y^*| \leq \delta$ and $u(0) := (i(0), ip(0), ii(0)) = \delta' v$. If $\delta > 0$ is small enough then there is $T > 0$ so that $\min_j u_j(T) \geq 2\delta'$ for all $0 < \delta' \leq \delta$.*

Proof. First write the MFE (2.5.2), without the y equation, in matrix form as follows:

$$\begin{pmatrix} i' \\ ip' \\ ii' \end{pmatrix} = \begin{pmatrix} -(1+r_+y) & r_- & r_- \\ r_+(y^*-i/2) & -(1+r_-) & 1 \\ r_+i/2 & \lambda & -(2+r_-+\lambda) \end{pmatrix} \begin{pmatrix} i \\ ip \\ ii \end{pmatrix} \tag{2.7.1}$$

The y dynamics proceeds as in (2.5.1), and note $|y(t)-y^*| \le |y(0)-y^*|$ for $t>0$. Write (2.7.1) as $u' = A(i,y)u$ with $u = (i, si, ii)^\top$ to emphasize the dependence on i, y. As noted in the proof of Theorem 2.5.4, if $R_0 > 1$ then $A := A(0, y^*)$ has a positive eigenvalue $\mu > 0$ with positive eigenvector v such that $|v| = 1$, so the system $v' = Av$ has solutions $v(t) = cve^{\mu t}$ for any $c > 0$. Let $|\cdot|$ denote the operator norm and let

$$L = \sup_{(i,y)\in[0,1]^2} |A(i,y)|$$

then any solution $u(t)$ to (2.7.1) has $|u(t)| \le |u(0)|e^{Lt}$ for $t > 0$. Fix $T > 0$, then for each $\epsilon > 0$, by continuity there is $\delta > 0$ so that if $\max(y - y^*, i) \le e^{LT}\delta$ then $|A(i,y) - A| \le \epsilon$. Let $|y(0) - y^*| \le \delta$ and for $0 < \delta' \le \delta$ let $u(t)$ be the solution to (2.7.1) with $u(0) = \delta' v$, then for $0 \le t \le T$,

$$\begin{aligned} |(u-v)'| &= |A(i,y)u - Av| \le |(A(i,y)-A)u| + |(A(u-v)| \\ &\le |A(i,y)-A||u| + |A||u-v| \\ &\le \epsilon|u| + L|u-v| \\ &\le \epsilon e^{Lt}\delta' + L|u-v| \end{aligned}$$

Letting $v(0) = u(0)$, defining $E(t) := |u(t) - v(t)|$, noting that $E(0) = 0$ and integrating,

$$E(T) \le e^{LT}\epsilon\delta' T$$

Since $v(T) = \delta v e^{\mu T}$,

$$\begin{aligned} \min_j u_j(T) &\ge \min_j v_j(T) - \max_j |v_j(T) - u_j(T)| \\ &\ge \delta' e^{\mu T} \min_j v_j - \epsilon e^{LT}\delta' T \\ &= \delta' e^{\mu T}(\min_j v_j - \epsilon e^{(L-\mu)T} T) \end{aligned}$$

Choose $T > 0$ so that $e^{\mu T} \min_j v_j/2 \ge 2$, then choose $\epsilon > 0$ so that $\epsilon e^{(L-\mu)T} T \le$

$\min_j v_j/2$, then it follows that $\min_j u_j(T) \geq 2\delta'$.

□

Now we can show long-time survival of the infection when $R_0 > 1$ and $|V_0| \geq \epsilon N$.

Lemma 2.7.4. *Suppose $R_0 > 1$. For each $\epsilon > 0$, there are constants $\delta, C, \gamma > 0$ so that if $|V_0| \geq \epsilon N$ then*

$$\mathbb{P}(\inf_{0 \leq t \leq e^{\gamma N}} |V_t| \geq \delta N) \geq 1 - Ce^{-\gamma N}$$

Proof. Recall that an event holds with high probability or whp in N if for N large enough it occurs with probability $\geq 1 - Ce^{-\gamma N}$ for some $C, \gamma > 0$. If $|V_0| \geq \epsilon N$ then $\max(i_0, ip_0, ii_0) \geq \epsilon/3$, so in view of Lemma 2.7.1 it is enough to prove the result starting from $u_0 := (i_0, ip_0, ii_0) \in \mathcal{E} := \{(\epsilon/3, 0, 0), (0, \epsilon/3, 0), (0, 0, \epsilon/3)\}$. For $\delta_1 > 0$, by Proposition 2.6.6 there are $T, \gamma_1 > 0$ so that whp $|y_t - y^*| \leq \delta_1$ for $T \leq t \leq e^{\gamma_1 N}$. If $u(0) \neq (0, 0, 0)$ then for $t > 0$, $\min_j u_j(t) > 0$; this is shown for $u(0) \in \Lambda^*$ in the proof of Lemma 2.5.3, but the same proof applies if $y \neq y^*$. Also, since $(0, 0, 0)$ is an equilibrium solution, by uniqueness of solutions $u(t) \neq (0, 0, 0)$ for $0 \leq t \leq T$, so by continuity of solutions $\inf\{|u(t)| : 0 \leq t \leq T\} > 0$. Therefore, there exists $0 < \delta_2 \leq \delta_1$ so that $\min_j u_j(T) \geq \delta_2$ and $\inf\{\max_j u_j(t) : 0 \leq t \leq T\} \geq \delta_2$ for all $u(0) \in \mathcal{E}$. For $u_0 = u(0) \in \mathcal{E}$ with $y_0 = y(0) \in [0, 1]$, by Proposition 2.6.5, whp $|u_t - u(t)| \leq \delta_2/2$ for $0 \leq t \leq T$ in which case $\min(i_T, ip_T, ii_T) \geq \delta_2/2$ and $\inf\{\max(i_t, ip_t, ii_t) : 0 \leq t \leq T\} \geq \delta_2/2$, which means that for the eigenvector v with $|v| = 1$ mentioned in the proof of Lemma 2.7.3, $(i_T, si_T, ii_T) \geq (\delta_2/2)v$, and also $|V_t| \geq (\delta_2/2)N$ for $0 \leq t \leq T$.

Taking $y(t) = y_t$ and $u(T) = (\delta_2/2)v$, if $|y_t - y^*| \leq \delta_1$ then by Lemma 2.7.3 there is $h > 0$ so that $\min_j u_j(T + h) \geq \delta_2$, and as before there is $\delta_3 > 0$ so that $\inf\{\max_j u_j(t) : T \leq t \leq T + h\} \geq \delta_3$. By Lemma 2.7.1 and the last paragraph, it is enough to consider the case $u_T = u(T) = (\delta_2/2)v$. Letting $\delta = \min(\delta_2/2, \delta_3/2)$ and using Proposition 2.6.5, with probability $\geq 1 - Ce^{\gamma_2 N}$, $|u_t - u(t)| \leq \delta$ for $T \leq t \leq T + h$, in which case $u_{T+h} \geq (\delta_2/2)v$ and $|V_t| \geq N \min(i_t, ip_t, ii_t) \geq (\delta_3/2)N$ for $T \leq t \leq T + h$. Letting $\gamma = \min(\gamma_1/2, \gamma_2/2)$ and iterating for $e^{\gamma N}/h$ time steps as in the proof of Lemma 2.6.6, whp $|V_t| \geq N \min(i_t, ip_t, ii_t) \geq (\delta_3/2)N$ for $T \leq t \leq e^{\gamma N}$. Combining with the previous estimate, whp $|V_t| \geq \delta N$ for $0 \leq t \leq e^{\gamma N}$ as we wanted to show. □

We now wrap up the macroscopic side of Theorem 2.2.2.

Proposition 2.7.5. *Suppose $R_0 > 1$ and let (y^*, i^*, ip^*, ii^*) with $i^* > 0$ be the non-trivial equilibrium solution to the MFE (2.5.2). Let $u_t = (i_t, ip_t, ii_t)$ and let $u^* = (i^*, ip^*, ii^*)$. For each $\epsilon > 0$, there are constants $C, T, \gamma > 0$ so that if $|V_0| \geq \epsilon N$ then*

$$\mathbb{P}(\sup_{T \leq t \leq e^{\gamma N}} |(y_t, u_t) - (y^*, u^*)| \leq \epsilon) \geq 1 - Ce^{-\gamma N}$$

Proof. We begin with the lower bound. As shown in the proof of Lemma 2.7.4 there are $T_1, h_1, \delta_1, \gamma_1 > 0$ so that whp $\min(i_t, ip_t, ii_t) \geq \delta_1$, and thus $u_t \geq \delta_1 v$, for $t = T_1 + kh_1$, $k = 1, ..., (e^{\gamma_1 N} - T_1)/h_1$, where v with $|v| = 1$ is the eigenvector from Lemma 2.7.3. Let $y(0) = y^*$ and $u(0) := (i(0), ip(0), ii(0)) = \delta_1 v$. If $\delta_1 > 0$ is small enough then $u_j'(0) > 0$ in each coordinate and since $u^* \neq (0, 0, 0)$ is unique, as shown in the proof of Lemma 2.5.3 $u(t)$ is increasing with respect to (i, ip, ii) coordinates and $\lim_{t\to\infty} u(t) = u^*$, and in particular $u(t) \leq u^*$ for $t \geq 0$. We will need the stronger fact $u_j(t) < u_j^*$ for $j = 1, 2, 3$. Looking to the equations for i', ip', ii' in (2.5.2), the derivative of each variable increases with the other two variables, and of course is equal to 0 at u^*. If we had $i(t) = i^*$, then since $ip(t) \leq ip^*$ and $ii(t) \leq ii^*$ we would have $i' < 0$ which contradicts the fact that $u(t)$ is increasing, and the same applies to $ip(t)$ and $ii(t)$.

Using the above facts, there is T_2 so that $u(T_2) \geq u^* - \epsilon/2$, and since $0 < \min_j(u_j^* - u_j(T_2)) =: \epsilon' \leq \epsilon$, there is h_2 so that $u(T_2 + h_2) \geq u^* - \epsilon'/2$. By Proposition 2.6.5, there is $\delta_2 > 0$ so that if $u_0 = u(0)$ and $|y_0 - y^*| \leq \delta_2$ then whp $|u_t - u(t)| \leq \epsilon'/2$ for $T_2 \leq t \leq T_2 + h_2$ in which case $u_t \geq u^* - \epsilon$ for $T_2 \leq t \leq T_2 + h_2$ and $u_{T_2+h_2} \geq u^* - \epsilon'$ which means that $u_{T_2+h_2} \geq u(T_2)$. By Lemma 2.6.6 there are T_3, γ_2 so that whp $|y_t - y^*| \leq \min(\delta_2, \epsilon)$ for $T_3 \leq t \leq e^{\gamma_2 N}$. Let k be such that $T_1 + kh_1 \geq T_3$ and let $T_4 = T_1 + kh_1$, then setting $u(T_4) = \delta_1 v$, whp $u_{T_4} \geq u(T_4)$ so it is enough to consider the case where $u_{T_4} = u(T_4)$. Letting $T = T_4 + T_2$, then for some $\gamma_3 > 0$, with probability $\geq 1 - Ce^{-\gamma_3 N}$, $u_t \geq u^* - \epsilon$ for $T \leq t \leq T + h_2$ and $u_{T+h_2} \geq u(T)$. Letting $\gamma = \min(\gamma_2/2, \gamma_3/2)$ and iterating for $(e^{\gamma N} - T)/h_2$ time steps (subtracting T to make sure y_t stays in bounds) as in the proof of Lemma 2.6.6 it follows that $u_t \geq u^* - \epsilon$ for $T \leq t \leq e^{\gamma N}$.

To prove the upper bound it is enough to consider any value of y_0 and let $u_0 = (y_0, (1/2)(1 - y_0), (1/2)(1 - y_0))$. Setting $y(0) = y^*$ and $u(0) = (y^*, (1/2)(1 -$

$y^*), (1/2)(1-y^*))$, then as shown in the proof of Lemma 2.5.3, $u(t)$ decreases to u^*. Moreover, $u_j(t) - u_j^* > 0$ for the same reason as above, so there is T_1 so that $u(T_1) \leq u^* + \epsilon/2$, and since $0 < \min_j(u_j(T_1) - u_j^*) =: \epsilon' \leq \epsilon$, there is h so that $u(T_1 + h) \geq u^* - \epsilon'/2$. By Proposition 2.6.5, there is $\delta > 0$ so that if $\max(|u_0 - u(0)|, |y_0 - y(0)|) \leq \delta$ then whp $|u_t - u(t)| \leq \epsilon'/2$ for $T_1 \leq t \leq T_1 + h$ in which case $u_t \leq u^* + \epsilon$ for $T_1 \leq t \leq T_1 + h$ and $u_{T_1+h} \leq u^* + \epsilon'$ which means that $u_{T_1+h} \leq u(T_1)$. By Lemma 2.6.6 there are T_2, γ_1 so that whp $|y_t - y^*| \leq \delta$ for $T_2 \leq t \leq e^{\gamma_1 N}$. Letting $T = T_1 + T_2$ and setting $u(T_2) = (y^*, (1/2)(1-y^*), (1/2)(1-y^*)$ and $u_{T_2} = (y_{T_2}, (1/2)(1-y_{T_2}), (1/2)(1-y_{T_2}))$, then for some $\gamma_2 > 0$, with probability $\geq 1 - Ce^{-\gamma_2 N}$, $u_t \leq u^* + \epsilon$ for $T \leq t \leq T + h$ and $u_{T+h} \leq u(T)$. Letting $\gamma = \min(\gamma_1/2, \gamma_2/2)$ and iterating for $(e^{\gamma N} - T)/h$ time steps it follows as before that $u_t \leq u^* + \epsilon$ for $T \leq t \leq e^{\gamma N}$. □

In the next section we use a comparison to prove that if $R_0 < 1$ the infection disappears quickly from the population. To make this work we will need a complementary result to Lemma 2.7.4.

Lemma 2.7.6. *If $R_0 \leq 1$ then for each $\epsilon > 0$ there are $C, T, \gamma > 0$ so that*

$$\mathbb{P}(\sup_{T \leq t \leq e^{\gamma N}} |V_t| \leq \epsilon N) \geq 1 - Ce^{-\gamma N}$$

Proof. The proof is similar to that of Lemma 2.7.4. Letting $\overline{u} = (y^*, (1-y^*)/2, (1-y^*)/2)$ as in Lemma 2.5.3 and letting $(y^*, \overline{u}(t))$ be the solution to the MFE with $\overline{u}(0) = \overline{u}$, since $\overline{u}(t)$ decreases to $(0,0,0)$ and $\overline{u} \geq v$ for all $v \in \Lambda^*$, there is T_1 so that for any solution $(y^*, u(t))$, $|u(T_1)| \leq \epsilon/6$, and since $\epsilon' := \min_j u_j(T_1) > 0$, there is h so that $|u(T_1 + h)| \leq \epsilon'/2$. There is $\delta > 0$ so that if $\max(|u_0 - u(0)|, |y_0 - y^*|) \leq \delta$ then whp $|u_t - u(t)| \leq \min(\epsilon'/2, \epsilon/6)$ for $T_1 \leq t \leq T_1 + h$ in which case $|u_t| \leq \epsilon/3$ for $T_1 \leq t \leq T_1 + h$ and $|u_{T_1+h}| \leq \epsilon'$ which means $u_{T_1+h} \leq u(T_1)$. There are $\gamma_1, T_2 > 0$ so that whp $|y_t - y^*| \leq \delta$ for $T_2 \leq t \leq e^{\gamma_1 N}$. By monotonicity it is enough to consider $u_{T_2} = \overline{u}$. Letting $u(T_2) = u_{T_2}$ and $T = T_1 + T_2$, there are C_1, γ_2 so that with probability $\geq 1 - C_1 e^{\gamma_2 N}$, $|u_t| \leq \epsilon/3$ for $T \leq t \leq T + h$ and $u_T \leq u(T)$. Letting $\gamma = \min(\gamma_1, \gamma_2)$ and iterating for $(e^{\gamma N} - T)/h$ time steps, whp $|u_t| \leq \epsilon/3$ and thus $|V_T| \leq \epsilon N$ for $T \leq t \leq e^{\gamma N}$. □

2.8 Microscopic Behaviour

In this section we compare the partner model in the regime $|V| \leq \epsilon N$ for small $\epsilon > 0$ to a branching process to get decisive information when $R_0 \neq 1$.

2.8.1 Subcritical Case: $R_0 < 1$

First we introduce the comparison process to use when $R_0 < 1$.

Definition 2.8.1. *Define the* upperbound process *(UBP)* $B_t = (\mathcal{I}_t, \mathcal{SI}_t, \mathcal{II}_t)$ *on state space* $\{0, 1, 2, ...\}^3$ *with parameter* $0 \leq \delta \leq y^*$ *by the following transitions:*

- $\mathcal{I} \to \mathcal{I} - 1$ *at rate* $\mathcal{I}$,
- $\mathcal{I} \to \mathcal{I} - 1$ *and* $\mathcal{SI} \to \mathcal{SI} + 1$ *at rate* $r_+(y^* - \delta)\mathcal{I}$,
- $\mathcal{SI} \to \mathcal{SI} + 1$ *at rate* $2r_+\delta\mathcal{I}$,
- $\mathcal{II} \to \mathcal{II} + 1$ *at rate* $r_+\delta\mathcal{I}$,
- $\mathcal{SI} \to \mathcal{SI} - 1$ *at rate* $\mathcal{SI}$,
- $\mathcal{SI} \to \mathcal{SI} - 1$ *and* $\mathcal{I} \to \mathcal{I} + 1$ *at rate* $r_-\mathcal{SI}$,
- $\mathcal{SI} \to \mathcal{SI} - 1$ *and* $\mathcal{II} \to \mathcal{II} + 1$ *at rate* $\lambda\mathcal{SI}$,
- $\mathcal{II} \to \mathcal{II} - 1$ *and* $\mathcal{SI} \to \mathcal{SI} + 1$ *at rate* $2\mathcal{II}$,
- $\mathcal{II} \to \mathcal{II} - 1$ *and* $\mathcal{I} \to \mathcal{I} + 2$ *at rate* $r_-\mathcal{II}$

Note the UBP describes the evolution of the total number of particles of each of the three types $\mathcal{I}, \mathcal{SI}, \mathcal{II}$ in a multi-type continuous-time branching process; for an introduction to branching processes see [27]. We now show that for fixed $R_0 < 1$, if $\delta > 0$ is small enough the UBP quickly dies out.

Lemma 2.8.2. *For fixed* λ, r_+, r_- *let* B_t *denote the UBP with parameter* δ' *and let* R_0 *be as defined in* (2.4.1). *If* $R_0 < 1$ *there are* $C, \delta > 0$ *so that if* $|B_0| \leq N$ *and* $\delta' \leq \delta$ *then*

$$\mathbb{P}(|B_{C \log N}| = 0) \to 1 \text{ as } N \to \infty$$

Proof. For a multi-type continuous time branching process $B_t = (b_1(t), ..., b_n(t))$, with $b_j(t)$ denoting the number of type j particles alive at time t, we can extract some useful information from the *mean matrix* M_t defined by $m_{ij}(t) = \mathbb{E}(b_j(t) \mid b_k(0) = \delta_{ik})$. Since particles evolve independently, $\mathbb{E}(B_t) = B_0 M_t$ and it is not hard to show that M_t satisfies the equation

$$\frac{d}{dt} M_t = A M_t$$

and therefore $M_t = \exp(At)$, where $A = (r_{ij})$ is the matrix whose entries r_{ij} give the rate at which a particle of type i produces particles of type j. If $\mathrm{Re}(\lambda) < 0$ for each eigenvalue λ of A, then letting $\gamma_0 = \min\{|\,\mathrm{Re}(\lambda)| : \lambda \in \sigma(A)\}$ where $\sigma(\cdot)$ denotes the spectrum, from standard matrix theory it follows that for any $\gamma_1 < \gamma_0$, there is $C_1 > 0$ so that $m_{ij} \leq C_1 e^{-\gamma_1 t}$ for each pair ij. Since each $b_i(t)$ is valued on non-negative integers,

$$\begin{aligned} \mathbb{P}(B_t \neq (0, ..., 0)) &\leq \sum_i \mathbb{P}(b_i(t) \neq 0) \leq \sum_i \mathbb{E} b_i(t) \\ &= \sum_{ij} b_i(0) m_{ij}(t) \leq |B(0)| n^2 C_1 e^{-\gamma_1 t} \end{aligned}$$

If $|B(0)| \leq N$ then letting $t = C \log N$ for $C > 1/\gamma_1$ and setting $\gamma = C\gamma_1 - 1$ and $C_2 = n^2 C_1$ we find

$$\mathbb{P}(B_{C \log N} \neq (0, ..., 0)) \leq N C_2 e^{-\gamma_1 C \log N} = N C_2 N^{-\gamma_1 C} = C_2 N^{1 - \gamma_1 C} = C_2 N^{-\gamma}$$

which tends to 0 as $N \to \infty$. In our case,

$$A = A(\delta) = \begin{pmatrix} -(1 + r_+(y^* - \delta)) & r_+(y^* + \delta) & r_+ \delta \\ r_- & -(1 + r_- + \lambda) & \lambda \\ 2r_- & 2 & -(2 + r_-) \end{pmatrix}$$

Letting $\sigma(A)$ denote the spectrum and defining the *spectral abcissa* $\mu(A) := \max\{\mathrm{Re}(\lambda) : \lambda \in \sigma(A)\}$, if $\mu(A(\delta)) < 0$, then the real part of each eigenvalue of A is negative, and the above argument applies. By continuity of eigenvalues in the entries of a matrix, it is enough to show $\mu(A(0)) < 0$, since then there is $\delta > 0$ so that if $\delta' \leq \delta$ then

$\mu(A(\delta')) \leq \mu(A(0))/2 < 0$. Setting $\delta = 0$,

$$A(0) = \begin{pmatrix} -(1+r_+y^*) & r_+y^* & 0 \\ r_- & -(1+r_-+\lambda) & \lambda \\ 2r_- & 2 & -(2+r_-) \end{pmatrix}$$

and looking to Section 2.5 we see that $A(0,0)$ is the (transpose of the) linearized matrix at $(0,0,0)$ for the MFE on Λ^*, which we denote A. As noted in Remark 2.5.5, $(0,0,0)$ is locally asymptotically stable when $R_0 < 1$, and in the proof of Theorem 2 in [50] this is done by showing that $\mu(A) < 0$. □

We now complete the proof of the case $R_0 < 1$ in Theorem 2.2.2.

Proposition 2.8.3. *If $R_0 < 1$ there are constants $C, T, \gamma > 0$ so that, from any initial configuration,*

$$\mathbb{P}(|V_{T+C\log N}| = 0) \to 1 \text{ as } N \to \infty$$

Proof. Let $U_t := (I_t, SI_t, II_t)$ denote variables in the partner model and for $\delta > 0$ such that $y^* - \delta \geq 0$ and $y^* + \delta \leq 1$, let B_t denote the UBP with parameter δ. We first describe a coupling with the property that $U_0 \leq B_0 \Rightarrow U_t \leq B_t$ for $t > 0$, with respect to the usual partial order $U \leq V \Leftrightarrow U_j \leq V_j, j = 1, 2, 3$. For $j = 1, ..., 10$ define a countable number of independent Poisson point processes (ppp's) $\{e_j(n) : n = 1, 2, ...\}$ with respective rates $1, r_+, r_+, 1, r_-, \lambda, 2, r_-, r_+, r_-$, together with independent uniform $[0,1]$ random variables attached to each event in $e_2(n), e_3(n), e_9(n), n = 1, 2, ...$ These correspond to the nine transitions listed in the definition of the UBP, except that the second and third transition in the UBP are lumped into e_2, plus an additional transition for $S + S \to SS$ and one for $SS \to S + S$. Note that the rates of e_2, e_3, e_9 appear too large at the moment and are corrected in the next paragraph.

Construct the UBP one transition at a time as follows, letting $(\mathcal{I}, \mathcal{SI}, \mathcal{II})$ denote the present state. Each event in $e_1(1), ..., e_1(\mathcal{I})$ reduces $\mathcal{I}$ by 1. For an event in e_2, e_3 let p denote the corresponding uniform $[0,1]$ random variable. If an event in $e_2(1), ..., e_2(\mathcal{I})$ occurs and $p \leq (y^* - \delta)$, reduce $\mathcal{I}$ by 1 and increase $\mathcal{SI}$ by 1, while if $y^* - \delta < p \leq y^* + \delta$ simply increase $\mathcal{SI}$ by 1. If an event in $e_3(1), ..., e_3(\mathcal{I})$ occurs and $p \leq \delta$, increase $\mathcal{II}$ by 1. Each event in $e_4(1), ..., e_4(\mathcal{SI})$ reduces $\mathcal{SI}$ by 1, each event in $e_5(1), ..., e_5(\mathcal{SI})$ reduces $\mathcal{SI}$ by 1 and increases $\mathcal{I}$ by 1, each event in $e_6(1), ..., e_6(\mathcal{SI})$

reduces $\mathcal{SI}$ by 1 and increases $\mathcal{II}$ by 1, each event in $e_7(1), ..., e_7(\mathcal{II})$ reduces $\mathcal{II}$ by 1 and increases $\mathcal{SI}$ by 1, and each event in $e_8(1), .., e_8(\mathcal{II})$ reduces $\mathcal{II}$ by 1 and increases $\mathcal{I}$ by 2. It can be checked that the transition rates are correct.

Similarly, construct the Markov chain $(S_t, I_t, SS_t, SI_t, II_t)$ for the partner model as follows, letting (S, I, SS, SI, II) denote the present state. Define $\alpha_t = y_t - y^* - i_t$ and $\beta_t = i_t/2 - 1/(2N)$ and note that α_t and β_t are piecewise constant in time. Each event in $e_1(1), ..., e_1(I)$ reduces I by 1 and increases S by 1. If an event in $e_2(1), ..., e_2(I)$ occurs and $p \leq y^* + \alpha_t$ reduce S and I by 1 and increase SI by 1. If an event in $e_3(1), ..., e_3(I)$ occurs and $p \leq \beta_t$ reduce I by 2 and increase II by 1. Each event in $e_4(1), ..., e_4(SI)$ reduces SI by 1 and increases SS by 1, each event in $e_5(1), ..., e_5(SI)$ reduces SI by 1 and increases S and I by 1, and events in e_6, e_7, e_8 have the same effect as before. If an event in $e_9(1), ..., e_9(S)$ occurs and $p \leq s_t/2 - 1/(2N)$ reduce S by 2 and increase SS by 1, and each event in $e_{10}(1), ..., e_{10}(SS)$ reduces SS by 1 and increases S by 2. Recalling that $U_t := (I_t, SI_t, II_t)$, if $U_0 \leq B_0$ and $\sup_{s \leq t} \max(|\alpha_s|, \beta_s) \leq \delta$ then $U_t \leq B_t$ since (as can be easily checked) the order is preserved at each transition.

By Lemma 2.6.6 there are $T_1, \gamma_1 > 0$ so that whp $|y_t - y^*| \leq \delta/2$ for $T_1 \leq t \leq e^{\gamma N}$ and since $R_0 < 1$, by Lemma 2.7.6 there are T_2, γ_2 so that $|V_t| \leq (\delta/2)N$ and thus $i_t \leq \delta/2$ for $T_2 \leq t \leq e^{\gamma_2 N}$. Letting $T = \max(T_1, T_2)$ and $\gamma = \min(\gamma_1, \gamma_2)$, whp $\max(|\alpha_t|, \beta_t) \leq \delta$ for $T \leq t \leq e^{\gamma N}$. Setting $B_T = U_T$ and using Lemma 2.8.2 completes the proof. □

2.8.2 Supercritical Case: $R_0 > 1$

We introduce the comparison process for $R_0 > 1$, which is similar to the UBP, but different.

Definition 2.8.4. *Define the* lowerbound process *(LBP)* $B_t = (\mathcal{I}_t, \mathcal{SI}_t, \mathcal{II}_t)$ *on state space* $\{0, 1, 2, ...\}^3$ *with parameters* $\delta \geq 0$ *such that* $y^* - \delta \geq 0$ *by the following transitions:*

- $\mathcal{I} \to \mathcal{I} - 1$ *at rate* $(1 + 2r_+\delta)\mathcal{I}$,
- $\mathcal{I} \to \mathcal{I} - 1$ *and* $\mathcal{SI} \to \mathcal{SI} + 1$ *at rate* $r_+(y^* - \delta)\mathcal{I}$,
- $\mathcal{I} \to \mathcal{I} - 2$ *at rate* $r_+\delta\mathcal{I}$,

- $\mathcal{SI} \to \mathcal{SI} - 1$ *at rate* $\mathcal{SI}$,
- $\mathcal{SI} \to \mathcal{SI} - 1$ *and* $\mathcal{I} \to \mathcal{I} + 1$ *at rate* $r_-\mathcal{SI}$,
- $\mathcal{SI} \to \mathcal{SI} - 1$ *and* $\mathcal{II} \to \mathcal{II} + 1$ *at rate* $\lambda\mathcal{SI}$,
- $\mathcal{II} \to \mathcal{II} - 1$ *and* $\mathcal{SI} \to \mathcal{SI} + 1$ *at rate* $2\mathcal{II}$,
- $\mathcal{II} \to \mathcal{II} - 1$ *and* $\mathcal{I} \to \mathcal{I} + 2$ *at rate* $r_-\mathcal{II}$

As before, the LBP describes the evolution of the total number of particles of each of the three types $\mathcal{I}, \mathcal{SI}, \mathcal{II}$ in a multi-type continuous-time branching process. We now show that for fixed $R_0 > 1$, if $\delta > 0$ is small enough then the LBP survives.

Lemma 2.8.5. *Let B_t denote the LBP with parameter δ'. If λ, r_+, r_- are such that $R_0 > 1$ then there are $C, \delta > 0$ so that if $\delta' \leq \delta$ then $\liminf_{N\to\infty} \mathbb{P}(B_{C\log N} \neq (0,0,0)) > 0$ and*

$$\mathbb{P}(|B_{C\log N}| \geq \delta N \mid B_{C\log N} \neq (0,0,0)) \to 1 \text{ as } N \to \infty$$

Proof. As in the proof of Lemma 2.8.2 define the mean matrix $M(t) = \exp(At)$ and the spectral abcissa $\mu(A)$. If $\delta' = 0$ for both the UBP and the LBP they coincide, in which case A is the transpose of the linearized matrix at $(0,0,0)$ of the MFE on Λ^*. As shown in the proof of Theorem 2.5.4, if $R_0 > 1$ then $\mu(A) > 0$. By continuity of eigenvalues in the entries of a matrix, there is $\delta > 0$ so that if $\delta' \leq \delta$ then $\mu(A(\delta')) \geq \mu(A)/2 > 0$. As shown in V.7 of [27], if $M(t)$ is such that for some $t_0 > 0$ and each entry $m_{ij}(t)$ of $M(t)$ one has $m_{ij}(t_0) > 0$ (which is the case here), then $\mu(A) =: \lambda_1$ is an eigenvalue of A, and if $\lambda_1 > 0$ the process is said to be *supercritical*. In this case, $B_t e^{-\lambda_1 t} \to Wv$ where v is a left eigenvector of A with eigenvalue λ_1 and W is a real-valued random variable. Setting $t = C\log N$ with $C > 1/\lambda_1$ and letting $\gamma = C\lambda_1 > 1$, $B_{C\log N} N^{-\gamma} \to Wv$, so for each $\epsilon > 0$,

$$\liminf_{N\to\infty} \mathbb{P}(|B_{C\log N}| \geq \delta N) \geq \lim_{N\to\infty} \mathbb{P}(|B_{C\log N}| \geq \epsilon N^\gamma) = \mathbb{P}(W|v| \geq \epsilon)$$

and letting $\epsilon \to 0^+$ and using continuity of measure,

$$\liminf_{N\to\infty} \mathbb{P}(|B_{C\log N}| \geq \delta N) \geq \mathbb{P}(W > 0)$$

Under a mild regularity assumption on the offspring distribution that holds trivially in this case, $\mathbb{P}(W > 0) = \lim_{t\to\infty} \mathbb{P}(B_t \neq (0,0,0)) > 0$. Since $|B_t| \geq \delta N$ implies

$B_t \neq (0,0,0)$, this means $\limsup_{N\to\infty} \mathbb{P}(|B_{C\log N}| \geq \delta N) \leq \lim_{t\to\infty} \mathbb{P}(B_t \neq (0,0,0)) = \mathbb{P}(W > 0)$, so $\lim_{N\to\infty} \mathbb{P}(|B_{C\log N}| \geq \delta N)$ exists and is equal to $\mathbb{P}(W > 0)$. The result then follows by observing that for $t, x > 0$, $\mathbb{P}(|B_t| \geq x \mid B_t \neq (0,0,0)) = \mathbb{P}(|B_t| \geq x)/\mathbb{P}(B_t \neq (0,0,0))$. □

We now complete the proof of Theorem 2.2.2.

Proposition 2.8.6. *If $R_0 > 1$, there are constants $\delta, p, C, T > 0$ so that if $|V_0| > 0$ then $\mathbb{P}(|V_{T+C\log N}| \geq \delta N) \geq p$.*

Proof. We use the same approach as in the proof of Proposition 2.8.3. Let $U_t := (I_t, SI_t, II_t)$ denote variables in the partner model and for $\delta_1 > 0$ such that $\delta_1 \leq 1$, $y^* - \delta_1 \geq 0$ and $y^* + \delta_1 \leq 1$, let B_t denote the LBP with parameter δ_1. Let $e_1, ..., e_{10}$ be as in the proof of Proposition 2.8.3.

Construct the LBP one transition at a time as follows, letting $(\mathcal{I}, \mathcal{SI}, \mathcal{II})$ denote the present state. Each event in $e_1(1), ..., e_1(\mathcal{I})$ reduces $\mathcal{I}$ by 1. For an event in e_2, e_3 let p denote the corresponding uniform $[0,1]$ random variable. If an event in $e_2(1), ..., e_2(\mathcal{I})$ occurs and $p \leq (y^* - \delta_1)$, reduce $\mathcal{I}$ by 1 and increase $\mathcal{SI}$ by 1, while if $y^* - \delta_1 < p \leq y^* + \delta_1$ simply reduce $\mathcal{I}$ by 1. If an event in $e_3(1), ..., e_3(\mathcal{I})$ occurs and $p \leq \delta_1$, reduce $\mathcal{I}$ by 2. Events in e_4, e_5, e_6, e_7, e_8 have the same effect as in the dynamics of the UBP. The Markov chain $(S_t, I_t, SS_t, SI_t, II_t)$ for the partner model is constructed in the same way as in the proof of Proposition 2.8.3, with α_t, β_t defined in the same way, and it is easy to check in this case that if $U_0 \geq B_0$ and $\sup_{s\leq t} \max(|\alpha_s|, \beta_s) \leq \delta_1$ then $U_t \geq B_t$.

Define the stopping time $\tau = \inf\{t : |U_t| \geq \delta_1 N/2\}$ and note that $|V_\tau| \geq (\delta_1/2)N$. By Lemma 2.7.4 and using the strong Markov property, there are $\delta, \gamma > 0$ so that whp $|V_t| \geq \delta N$ for $\tau \leq t \leq \tau + e^{\gamma N}$. There are $T, \gamma > 0$ so that whp $|y_t - y^*| \leq \delta_1/2$ for $T \leq t \leq e^{\gamma N}$. If $\tau \leq T$, then since T is fixed, we are done. If $t < \tau$ then $i_t \leq \delta_1/2$, so letting $B_T = U_T$, if $T \leq t < \tau$ then $\max(|\alpha_t|, \beta_t) \leq \delta_1$, so $U_t \geq B_t$ for $T \leq t < \tau$. The result follows from this inequality and from Lemma 2.8.5. □

Acknowledgements

The authors wish to thank Chris Hoffman for the suggestion to study the model on the complete graph, as well as the referee for a thorough reading of the article and for

helpful comments. The authors' research is partly supported by NSERC Discovery Grants (P.vdD. and R.E.) and by an NSERC PGSD2 Graduate Scholarship (E.F.).

Chapter 3

The SEIS Process

Abstract

The susceptible-exposed-infectious-susceptible (SEIS) model is well-known in mathematical epidemiology as a model of infection in which there is an average incubation time between the moment of infection and the onset of infectiousness. The compartment model is well studied, but the corresponding particle system has so far received no attention. We define the model and estimate critical values, also proving convergence of critical values in the limit of small and large incubation time, and identify a limiting process to which the SEIS model converges in the limit of large incubation time.

3.1 Introduction

The SEIS model is a model of the spread of an infection that in addition to the usual susceptible and infectious classes includes an exposed class that is infected but not yet infectious; it can be used to model infections such as gonorrhea in which there is a short latent stage before the onset of infectiousness, and in which recovery from the infection confers no immunity. The classical model, usually called a compartment model, is deterministic and consists of a set of three differential equations describing the evolution of the number of susceptible, exposed and infectious individuals (the three compartments), which for simplicity are taken to be real-valued; for a formal definition see [5], Chapter 2. The model has either a globally stable disease-free state or an unstable disease-free state together with a globally stable endemic state, ac-

cording as the basic reproduction number for the infection is ≤ 1 or > 1; see [28] for a proof using Lyapunov functions.

Now, the classical SEIS model is deterministic and assumes that the population is well-mixed. However, there is a natural way to define an SEIS model that incorporates both spatial and random effects, using an interacting particle system (see [32] for an introduction to interacting particle systems). For the simpler SIS model with no exposed class, this system is called the contact process, and has been well studied over the last forty years, in a variety of spatial settings including the d-dimensional integer lattices, trees, random graphs, and even more general sequences of finite graphs; [32] and [34] give an overview of results up to 1985 and 1999 respectively, and [14] includes a survey of results on random graphs up to about 2009; a recent result on fairly general sequences of finite graphs can be found in [36].

For the contact process on the d-dimensional lattice $\mathbb{Z}^d$ with a single initially infectious site there is a critical value λ_c of the infection parameter λ such that for $\lambda \leq \lambda_c$ the process dies out with probability 1, and for $\lambda > \lambda_c$ the process survives with a positive probability, spreading linearly in time and converging to a non-trivial invariant measure when it survives; for a proof of convergence see [34], Part I, and for a proof of linear spread in $d = 1$ and linear spread with convergence to a limiting shape in $d \geq 2$, see [7] and [9].

According to numerical simulations in $d = 1, 2$, the SEIS model behaves in the same way as the contact process, in this case with a critical value that varies slightly with the incubation time, and spreading linearly in time when it survives. However, for the SEIS model it is not clear how to prove this, because of the absence of a property called monotonicity that enables much of the analysis of the contact process. Nevertheless, we can show that the infection survives when the infection parameter is large enough, uniformly in the incubation time, and we can obtain reasonable bounds when the incubation time is either very large or very small. In addition, in the limit of large incubation time the model, when properly rescaled in time, approaches a limit process, and we describe the limit process and the convergence to the limit process. We begin by describing the process and summarizing the main results.

3.2 Main Results

To distinguish it from the compartment model, we use "SEIS process" to refer to the SEIS model as an interacting particle system. Given a finite or countably infinite connected undirected graph $G = (V, E)$ with bounded degree i.e., for some $d < \infty$, $|\{y \in V : xy \in E\}| \leq d$ for each $x \in V$, the SEIS process with infection parameter $\lambda > 0$ and incubation time $\tau \geq 0$ is defined as follows. Letting 0 denote susceptible, 1 denote exposed and 2 denote infectious, each site $x \in V$ is in one of the states $0, 1$ or 2 (that we later refer to as *types*), with transitions

- $2 \to 0$ at rate 1 (recovery)
- $1 \to 2$ at rate $1/\tau$ or instantaneously if $\tau = 0$ (onset)
- $0 \to 1$ at rate $\lambda n_2(x)$ (transmission)

where $n_2(x)$ is cardinality of the set $\{xy \in E : y \text{ is in state } 2\}$. The case $\tau = 0$ is the contact process with transmission parameter λ. The meaning of "rate" is that in the absence of other transitions, each transition occurs after an amount of time which is exponentially distributed with parameter given by the rate.

A standard reference on particle systems, and methods for constructing them, can be found in [32]. Since it will help us later on, we follow [26] and use a graphical representation to construct the process. We begin with the *spacetime set* $\mathcal{S} = G \times [0, \infty)$, which we picture as a copy of G extruded upward along fibers in the increasing time direction; this is particularly easy to imagine if G is a planar graph. When required we use the topology on $\mathcal{S}$ with base $\{a\} \times (t, t') : a \in V \cup E, t < t'\}$. Place independent 1-dimensional Poisson point processes (p.p.p.'s) along fibers $\{\cdot\} \times [0, \infty)$ as follows:

- at each site $x \in V$, recovery with intensity 1 and label $\times$,
- at each site $x \in V$, onset with intensity $1/\tau$ and label $\star$ if $\tau > 0$, or omitting if $\tau = 0$, and
- along each edge $xy \in E$, transmission with intensity λ and label $\leftrightarrow$.

This furnishes the probability space Ω, which we can think of as a random labelling of $\mathcal{S}$, and which we refer to as the *substructure*; the notation $\mathbb{P}$ is used to denote

the law of the substructure, and when necessary, we write for example $\mathbb{P}_\lambda$ or $\mathbb{P}_\tau$ to emphasize the dependence on parameters. Define the following notation: for a Borel measurable set $R \subset \mathcal{S}$ let $\mathcal{F}(R)$ denote the σ-algebra generated by the restriction of the substructure to R, and for $t > 0$ let $\mathcal{F}(t) = \mathcal{F}(G \times [0, t])$. If we view the substructure as a function of time U_t then U_t is adapted to the filtration $\mathcal{F}(t)$ and it follows from the strong Markov property applied to U_t that for any stopping time s the law of $\mathbb{E}(U_{t+s}|\mathcal{F}(s))$ is the same as the law of U_t. Also, it follows from standard properties of p.p.p's that if $\{R_i : i = 1, 2, ...\}$ are pairwise disjoint then $\{\mathcal{F}(R_i) : i = 1, 2, ...\}$ are independent; the same is true if the sets are disjoint up to measure zero in the measure on $\mathcal{S}$ given by the product of counting measure on edges and vertices of G with Lebesgue measure on $[0, \infty)$. Both of these facts will be useful throughout the paper.

Given an initial configuration $\eta_0 \in \{0, 1, 2\}^V$, to determine $\eta_t(x)$ for each realization $\omega \in \Omega$ consider the set $T_t(x)$ of points $(u, s) \in G \times [0, t]$ that can reach (x, t) by moving either upwards in time along vertices or horizontally along transmission labels $\leftrightarrow$. In order to compute $\eta_t(x)$ from the transition labels it suffices to compute $\eta_s(y)$ for $(y, s) \in T_t(x)$. By a simple comparison, $|\{u \in G : (u, t - s) \in T_t(x)\}|$ is bounded above by a branching process with no deaths in which pairs of offspring are produced at rate λd, so with probability 1, $T_t(x)$ is a bounded set, and it follows easily (for example, by considering the events $T_t(x) \subset G_n \times [0, t]$ for a sequence of graphs G_n with $\cup G_n = G$) that the number of labels in $T_t(x)$ is almost surely finite. Denote the timing of labels by $t_1 < t_2 < ... < t_m$, then given $\eta_{t_i}(x)$ for x such that $(x, t_i) \in T_t(x)$, if the label at time t_{i+1} is

- $\times$ at x and $\eta_{t_i}(x) = 2$ then $\eta_{t_{i+1}}(x) = 0$,
- $\star$ at x and $\eta_{t_i}(x) = 1$ then $\eta_{t_{i+1}}(x) = 2$,
- $\leftrightarrow$ along xy and $\eta_{t_i}(x) = 0$, $\eta_{t_i}(y) = 2$ then $\eta_{t_{i+1}}(x) = 1$ if $\tau > 0$ and $\eta_{t_{i+1}}(x) = 2$ if $\tau = 0$.

otherwise nothing happens. Then, let $\eta_t(x) = \eta_{t_m}(x)$. The reader may easily verify that this approach defines $\eta_t(x)$ for all $x \in V$, $t \geq 0$ in a consistent manner. For what follows, say that x is *active* at time t if $\eta_t(x) \neq 0$. Letting $\mathcal{C}_0$ denote the set of configurations having only finitely many active sites, if $\eta_0 \in \mathcal{C}_0$ then bounding the number of active sites by a branching process in which each particle produces offspring at rate λ

it follows that $\eta_t \in \mathcal{C}_0$ for $t > 0$, and η_t behaves like a continuous time Markov chain on $\mathcal{C}_0$ in the sense of [41], with transition rates as specified in the description of the model.

The above graphical representation supplies a natural coupling of the process for all choices of η_0, namely the one in which, for each η_0, η_t is determined from η_0 for $t > 0$ via the substructure. With respect to this coupling, the reader may verify that the contact process, which is the case $\tau = 0$, is *monotone* in the partial order $\eta \leq \eta' \Leftrightarrow \forall x, \eta(x) \leq \eta'(x)$ in the sense that $\eta_0 \leq \eta'_0$ implies $\eta_t \leq \eta'_t$ for $t > 0$. In fact, the process is also monotone with respect to λ: if $\lambda < \lambda'$ then we can couple processes η_t with parameter λ and η'_t with parameter λ' so that $\eta_0 \leq \eta'_0$ implies $\eta_t \leq \eta'_t$ for $t > 0$. To do so, for transmission events place independent p.p.p.'s along each edge

- with intensity λ and label $\leftrightarrow$
- with intensity $\lambda' - \lambda$ and label $\leftrightarrow'$

with recovery events as before, and for transmission events, for η_t use only the labels $\leftrightarrow$, while for η'_t use both the labels $\leftrightarrow$ and $\leftrightarrow'$. Using this fact, and the fact that any configuration with a positive and finite number of active sites can reach any other such configuration, it follows directly that there is a critical value of the transmission parameter that we denote λ_c^0 (which may a priori be equal to 0 or ∞) such that the infection survives with positive probability when $\lambda > \lambda_c^0$ and $|\eta_0| \geq 1$ ($|\eta|$ denotes the number of active sites in η) and dies out with probability 1 when $\lambda < \lambda_c^0$ and $|\eta_0| < \infty$, where *survival* means $|\eta_t| > 0\ \forall t$ and *dying out* means $|\eta_t| = 0$ for t large enough. Whenever we refer to critical values in what follows, they will have this property; the only exceptions are the upper and lower critical values for the SEIS process defined below, for which the above property is split between the two.

Given that $\tau = 0$ gives the contact process it is natural to ask whether we obtain something as $\tau \to \infty$. The answer is yes, if we rescale time so that onset occurs at rate 1. We first describe the limit process, then state the sense in which the SEIS process converges to it. The limit process has the state space $\{0, 1\}^V$ where 1 can be thought of as occupied and 0 as vacant. It is defined using the *dispersal distributions* $p(x, \cdot)$ given by letting $p(x, A)$ be equal to the probability that for the contact process with the single infectious site x, transmission from x to every site in A occurs, followed by recovery at x, without transmission to any sites in A^c, and ignoring subsequent transmissions from other newly infected sites. Each occupied site x becomes vacant

at rate 1, at which point, with probability $p(x, A)$ all the vacant sites in A become occupied.

There is an obvious graphical representation of the limit process: at each site place a p.p.p. with intensity 1 and label $\star$, and at each occurrence of $\star$ at site x sample the dispersal distribution $p(x, \cdot)$, placing a $\rightarrow$ label from x to y for each y to which x disperses, and let the samples be independent. The rest of the construction follows the same pattern so we omit the details. It is easy to see the limit process is monotone, and is also monotone in λ; to see the latter property, for $\lambda < \lambda'$ make a joint construction by coupling dispersal distributions in the obvious way. Thus the limit process has a critical value that we denote λ_c^∞ such that the same dichotomy holds as for the contact process above. The following result describes convergence of the SEIS process to the limit process.

Theorem 3.2.1. *For fixed λ, let η_t denote the SEIS process on a countable graph with bounded degree, under the rescaling $t \mapsto t/\tau$, and let ζ_t denote the limit process. Let $S = \{t : \eta_t(x) = 2 \text{ for some } x\}$ denote the set of times when the rescaled SEIS process has an infectious site. Fix $T > 0$ and an initial state with no infectious sites and finitely many exposed sites, then for each τ there is a coupling of η_t and ζ_t so that with probability tending to 1 as $\tau \to \infty$,*

- *$\zeta_t = \eta_t$ for $t \in [0, T] \setminus S$ and*
- *$\ell(S \cap [0, T]) \to 0$ where ℓ is Lebesgue measure on the line.*

The main idea of the proof is that with probability tending to 1 as $\tau \to \infty$ in the SEIS process, between any two onset events a recovery event occurs, and when this happens the SEIS process behaves like the limit process. The assumption of finitely many initially active sites is necessary.

Unfortunately, unlike the contact process or the limit process, with respect to the graphical representation given above, for $\tau > 0$ the SEIS process is not monotone in the partial order induced by the order $0 < 1 < 2$ on types (or, it can be checked, for any other order, though $0 < 2 < 1$ is the only other real possibility), since if we take configurations $\eta \leq \eta'$ with $\eta(x) = 1$ and $\eta'(x) = 2$ the 2 can flip to a 0 before the 1 becomes a 2, since type 1 ignores the $\times$ labels. Intuitively, this makes sense

Table 3.1: Lower bounds on $\lambda_c^-(\tau)$

τ	10^4	10^3	100	10	1	0.58	1/10	1/100	10^{-3}	10^{-4}
$\lambda_c^-(\tau) >$	1.57	1.57	1.56	1.45	1.15	1.13	1.24	1.32	1.34	1.34

because although type 2 can spread the infection while type 1 cannot, type 1 is not vulnerable to recovery events while type 2 is. Of course, it is possible to search for other graphical representations, or even more general types of coupling, to try to show monotonicity. After a long search, we have found no such coupling, but the reader is encouraged to try!

So, lacking monotonicity, we define the following two critical values for the SEIS process; note $\mathbb{P}_{\lambda,\tau}$ denotes the law of the process with parameters λ, τ.

$$\begin{aligned} \lambda_c^-(\tau) &= \sup\{\lambda' : \mathbb{P}_{\lambda,\tau}(\eta_t \text{ dies out } ||\eta_0| < \infty) = 1 \text{ if } \lambda < \lambda'\} \\ \lambda_c^+(\tau) &= \inf\{\lambda' : \mathbb{P}_{\lambda,\tau}(\eta_t \text{ survives } ||\eta_0| > 0) > 0 \text{ if } \lambda > \lambda'\} \end{aligned}$$

Clearly, $\lambda_c^-(\tau) \leq \lambda_c^+(\tau)$ for each τ. The next result gives quantitative estimates on critical values, both for the SEIS process and for the limit process, on $\mathbb{Z}$, i.e., on the graph $G = (V, E)$ with $V = \mathbb{Z}$ and $E = \{xy : |x - y| = 1\}$.

Theorem 3.2.2. *For the SEIS process on* $\mathbb{Z}$, $\lambda_c^+(\tau) < 6.875$ *when* $\tau \leq 1/10$, *and* $\lambda_c^-(\tau)$ *has the lower bounds given in Table 1. For the limit process on* $\mathbb{Z}$, $1.944 < \lambda_c^\infty < 8.563$.

Lower bounds on λ^- are obtained using the method of [51] applied to a monotone process that upperbounds the SEIS process, and the upper bound on λ^+ for small τ is obtained with the method of [10] applied to a monotone process that lowerbounds the SEIS process. In both cases the estimates are achieved with the assistance of a computer and are rigorous up to the rounding error on computations. Unfortunately, in this case each lower bound on λ^- is computed for a single value of τ; it is possible to make guesses by interpolating, but these are not a priori rigorous. Note also that the lower bounds suggest, but again do not prove, that the critical value of the upperbound process has a unique minimum near $\tau = 0.58$ and is otherwise increasing/decreasing. Numerical simulations of the SEIS process on $\mathbb{Z}$ suggest that

$\lambda_c^-(\tau) = \lambda_c^+(\tau)$ and that this value increases monotonically from about 1.6 at $\tau = 0$ to about 2.4 as $\tau \to \infty$.

For the limit process, the lower bound is obtained using the method of [51] and the upper bound, using the method of [10]. Note that for the contact process, $1.539 \leq \lambda_c \leq 1.942$ (lower bound from [51] and upper bound from [33]), and from the upper bound together with our estimate we note that the strict inequality $\lambda_c^\infty > \lambda_c^0$ holds.

Using different methods, we obtain some "qualitative" estimates on critical values.

Theorem 3.2.3. *For the SEIS process on $\mathbb{Z}$,*

- *there exists $\lambda_0 < \infty$ such that $\lambda_c^+(\tau) < \lambda_0$ for all τ,*
- *$\lambda^+(\tau), \lambda^-(\tau) \to \lambda_c^0$ as $\tau \to 0$ and*
- *$\lambda^+(\tau), \lambda^-(\tau) \to \lambda_c^\infty$ as $\tau \to \infty$.*

Here we show only that $\lambda_0 < \infty$ exists, as it appears difficult to get any sort of realistic estimate. The proof uses the block construction idea of [10] with a bit of extra work to get around the lack of monotonicity. Convergence of λ^+, λ^- as $\tau \to 0$ is proved with the help of the results of [15] and [51], in both cases by passing to a sequence of finite systems and using continuity with respect to parameters. Convergence of λ^+, λ^- as $\tau \to \infty$ is proved in the same way, with a couple of technical points that first need to be proved for the limit process.

The paper is laid out as follows. In Section 3.3 we prove Theorem 3.2.1. In Section 3.4 we prove Theorem 3.2.2. In Section 3.5 we prove Theorem 3.2.3.

3.3 Theorem 3.2.1: Convergence to the Limit Process

Here we prove Theorem 3.2.1. We begin with a useful lemma. Using the graphical representation given in Section 3.1, construct the SEIS process η_t rescaled by $t \mapsto t/\tau$, so that onset occurs at rate 1, recovery at rate τ and transmission at rate $\lambda\tau$. Recall that $\mathcal{S} = G \times [0, \infty)$ denotes the spacetime set.

Lemma 3.3.1. *Let $G = (V, E)$ be a finite graph. In the rescaled SEIS process, for each $T > 0$ with probability tending to 1 as $\tau \to \infty$, for each onset label $\star$ at a point $(x, t) \in \mathcal{S}$ there is a $t' > t$ and a recovery label $\times$ at (x, t') such that there are no onset labels in $V \times (t, t']$.*

Proof. Let $\{(x_i, t_i) : i = 1, 2, ...\}$ be the set of points $(x, t) \in \mathcal{S}$ such that there is a $\star$ label at (x, t), with $t_1 < t_2 < ...$; since the total intensity of $\star$ labels is finite, with probability 1 the times can be ordered in this way. Say a *discrepancy* occurs at time t_{i+1} if in the interval $\{x_i\} \times (t_i, t_{i+1})$ there are no $\times$ labels, then the desired event holds if the first discrepancy occurs after time T. The intensity of $\times$ labels at each site is τ and the intensity of $\star$ labels is $|V|$, so for each N, with probability $[\tau/(|V| + \tau)]^N$ which $\uparrow 1$ as $\tau \to \infty$, there are no discrepancies up to time t_{N+1}. Since $\mathbb{P}(t_{N+1} > T) \uparrow 1$ as $N \to \infty$, the result follows. □

Proof of Theorem 3.2.1. We prove the result when G is a finite graph. The result for infinite graphs is implied by the following fact that can be seen from the construction of the process. Fix $T > 0$ and let η_0 be an initial condition for the SEIS process with no infectious site and finitely many exposed sites. Define the graph distance $\rho(x, y)$ to be the least number of edges in any path between x and y with $\rho(x, y) = \infty$ if there is no path from x to y. For $k \geq 0$ let G_k be the graph induced by the set of vertices $y \in V$ such that $\rho(x, y) < k$ for some x such that $\eta_0(x) = 1$ and let $\mathcal{S}_k = G_k \times [0, \infty)$, then define η_t^k for $0 < t \leq T$ using the restriction of the substructure to $\mathcal{S}_k$. Then, $\eta_t^k = \eta_t$ for $0 < t \leq T$ with probability tending to 1 as $k \to \infty$.

First use the graphical representation to build two independent substructures $U_t^{(1)}$ and $U_t^{(2)}$ with respective filtrations $\mathcal{F}^{(1)}(t)$ and $\mathcal{F}^{(2)}(t)$, one as for the rescaled SEIS process and another as for the limit process. Construct the rescaled SEIS process η_t from η_0 using $U_t^{(1)}$ and let $\{(x_i, t_i); i = 1, ..., m\}$ with $t_1 < t_2 < ... < t_m$ denote the points (x, t) at which $\eta_t(x)$ goes from 1 to 2. For $i = 1, ..., m$ let $s_i = \min\{t > t_i : \eta_t(x) = 0\}$ be the first recovery time of x_i after t_i. Say that a *discrepancy* occurs if $t_{i+1} < s_i$ for some $i \in \{1, ..., m - 1\}$. So long as no discrepancy has occurred we will use $U_t^{(1)}$ to construct the limit process ζ_t; $U_t^{(2)}$ will help to construct ζ_t in the event of a discrepancy.

For $i = 1, ..., m - 1$ define the stopping times $r_i = \max s_i, t_{i+1}$, and let $\zeta_t = \eta_t$ for $t \in [0, t_1]$. For $t \in [0, t_1]$ let $\zeta_t = \eta_t$. Then, working inductively, suppose

η_t is determined for $t \in [0, r_{j-1}]$, is measurable with respect to $\mathcal{F}^{(1)}(r_{j-1})$, and no discrepancy has occurred up to time r_{j-1} i.e., $r_i = t_{i+1}$ for $i = 1, ..., j-1$. To determine ζ_t for $t \in [r_{j-1}, t_{j+1}]$, use the $\leftrightarrow$ labels in $\{x_j\cdot\} \times (t_j, s_j)$ and the $\times$ label at (x_j, s_j) to obtain the propagation distribution at (x_j, t_j) and use the $\star$ label at (x_{j+1}, t_{j+1}) to obtain the next onset transition; note these transitions depend only on $\zeta_{r_{j-1}}$ and $\mathbb{E}(U_{t+r_{j-1}}|\mathcal{F}(r_{j-1}))$ so have the correct distribution, and are measurable with respect to $\mathcal{F}(r_j)$. If $r_j = t_{j+1}$ then ζ_t is determined for $t \in [0, r_j]$ and no discrepancy has occurred up to time r_j. If $r_j = s_j$ a discrepancy has occurred; in this case, use the second substructure to determine ζ_t for $t \in [t_{j+1}, T]$, noting that the second substructure is independent of the first. Proceeding in this way determines ζ_t for $t \in [0, T]$, and by Lemma 3.3.1, with probability tending to 1 as $\tau \to \infty$, there are no discrepancies in the time interval $[0, T]$ and so $\eta_t = \zeta_t$, $0 \le t \le T$. It is left to the reader to show that for $S = \{t : \eta_t(x) = 2 \text{ for some } x\}$, $\ell(S \cap [0, T]) \to 0$ with probability tending to 1 as $\tau \to \infty$; to do so it suffices to prove a slight refinement of Lemma 3.3.1. □

3.4 Theorem 3.2.2: Quantitative Estimates

In this section we prove Theorem 3.2.2, in three parts: estimate of λ^+, estimates of λ^-, estimate of λ_c^∞. First we define the lowerbound and upperbound processes, which we denote $\underline{\eta_t}$ and $\overline{\eta_t}$. The idea is to modify some transitions in the SEIS process so that we end up with a monotone process that either lowerbounds or upperbounds the original process. The definition is only relevant for $\tau > 0$, since if $\tau = 0$, in both cases it coincides with the contact process.

3.4.1 Lowerbound Process

Starting from the graphical representation for the SEIS process, to obtain $\underline{\eta_t}$, construct the process as if it was the SEIS process, except that whenever an exposed site sees a recovery label $\times$, it becomes healthy. As it turns out, this gives a particular case of what is called the two-stage contact process [29], [19], which is known to be monotone, as well as monotone increasing in λ and monotone decreasing in τ, with respect to the partial order on configurations induced by the order $0 < 1 < 2$ on types; given what was shown for the contact process in Section 3.1, this is not hard to

check. Intuitively, the reason why the stated monotonicity holds is because now the exposed type is in every sense weaker, in its ability to spread the infection, than the infectious type. Monotonicity in λ is intuitively clear, and monotonicity in τ can be explained by saying that the longer an exposed site has to wait to become infectious, the less it will spread the infection. This gives the existence of a critical value $\underline{\lambda_c}(\tau)$ that is non-decreasing in τ. It is also not hard to check that $\underline{\eta_0}$ is a genuine lower bound, that is, if $\underline{\eta_0} \leq \eta_0$ then $\underline{\eta_t} \leq \eta_t$ for $t > 0$, with respect to the partial order just described. This implies in particular that $\underline{\lambda_c}(\tau) \geq \lambda^+(\tau)$ for each τ. As shown in [19], for a graph of bounded degree, $\underline{\lambda_c}(\tau) \to \infty$ at a finite value of τ, so this upper bound is only useful for small values of τ.

3.4.2 Upperbound Proces

To get $\overline{\eta_t}$ we first picture η_t as follows. Recall that $\mathcal{S} = G \times [0, \infty)$ is the spacetime set, which we picture as a copy of G extruded upward in the increasing time direction. Given η_0, and determining the process on each realization for all time, if $\eta_t(x) = 1$ for $t \in [t_1, t_2)$, draw a thick dashed line on the line segment $\{x\} \times [t_1, t_2)$ in $\mathcal{S}$, and if $\eta_t(x) = 2$ draw a thick solid line; if $\eta_t(x) = 0$ leave it blank. If the infection is transmitted along an edge $e \in E$ then draw a thick solid line with an arrow pointing in the direction it was transmitted.

Now, modify the graphical representation so that transmission labels are directed. That is, for $xy \in E$, with intensity λ place transmission labels $\rightarrow$ from x to y and with intensity λ, place independent transmission labels $\leftarrow$ from y to x. Clearly, this does not change the law of η_t. Then, to define $\overline{\eta_t}$ we simply allow both a dashed line *and* a solid line to exist at the same site, at the same time; that is, if both x and y are infectious and there is a transmission label from x to y, then y becomes "both" infectious and exposed; with respect to the above visualization, along y there is both a dashed line and a solid line, each behaving as it would in the SEIS process. This is why we make labels directed; if both x and y have a solid line and a $\leftrightarrow$ appears along edge xy there is no way to tell which of the two sites x, y will receive a dashed line.

So, if y has both a dashed and solid line and the next event is

- an onset label $\star$, the dashed line coalesces with the solid line, and only a solid

line remains, and if it is

- a recovery label $\times$, the solid line is knocked out and only the dashed line persists

Then, as in the SEIS process, only a solid line is able to use the transmission labels. Also, at most one line of each type is allowed at a single site, so if x already has a dashed line and there is a transmission event to x, it still has only one dashed line. In order to refer to it, we denote by type 3 the presence of both a dashed and solid line. It is not hard to check that $\overline{\eta_t}$ is monotone, and is monotone increasing in λ, with respect to the partial order on configurations induced by the order $0 < 1, 2 < 3$ on types; it is not, however, monotone in τ, effectively because in this process, as in the SEIS process, types 1 and 2 are not comparable. Thus for each τ, $\overline{\eta_t}$ has a critical value $\overline{\lambda}_c(\tau)$ whose variation in τ is not known a priori. Clearly, $\overline{\eta_t}$ is an upper bound for η_t in the sense that $\overline{\eta_0} \geq \eta_0$ implies $\overline{\eta_t} \geq \eta_t$ for $t > 0$, with respect to the partial order just described, which implies that $\overline{\lambda}_c(\tau) \leq \lambda^-(\tau)$, for each τ.

3.4.3 Some Definitions

We introduce a couple of definitions that will be useful in this section and the next section. Notice that, given a finite state space S and transition rates q_{ij} for $i, j \in S$, we can construct a continuous time Markov chain on S as follows. Given that $X_t = i$, then corresponding to the set of k such that $q_{ik} > 0$ we have an independent collection of exponential random variables s_k with rate q_{ik}, and setting $s = \min s_k$ and $j = \operatorname{argmin} s_k$, we let $X_{t+s} = j$. Note this is the same Markov chain as the one in which each i to j transition is generated by a Poisson point process with intensity q_{ij}. It can be checked by a calculation that $p_{ij} = \mathbb{P}(X_{t+s} = j | X_t = i) = q_{ij}/q_i$ if $q_i \neq 0$, where $q_i = \sum_{k \neq i} q_{ik}$; if $q_i = 0$ the chain remains at i once it has arrived there.

Definition 3.4.1. *The discrete time Markov chain on S determined by the transition probabilities p_{ij} defined above is called the* embedded jump chain.

Note that the embedded jump chain encodes the changes of state in the continuous time chain. In fact, the continuous time chain can be reconstructed from the embedded jump chain by waiting an independent exponential time of rate q_i at each state i before making a transition; see [41] for details.

Our next task is to define a notion of path for the infection in spacetime.

Definition 3.4.2. *A* path *in spacetime is a list of alternating vertical and horizontal line segments*

$$(v_1, h_1, ..., v_{m-1}, h_{m-1}, v_m)$$

in $\mathcal{S}$ *with each* $v_i = \{x_i\} \times (t_{i-1}, t_i)$, $t_{i-1} < t_i$ *and each* $h_i = \{x_i x_{i+1}\} \times \{t_i\}$. *The* base *of a path is the point* (x_1, t_0), *and the* end *is the point* (x_m, t_m).

The notion of active path is here defined only for the contact, lowerbound, and limit processes. Although it is easy to define for the upperbound process we will have no need for it, and for the SEIS process the definition will be a bit different.

Definition 3.4.3. *For the contact process, a path is* active *if for* $i = 1, ..., m-1$ *there is a* $\leftrightarrow$ *label at* h_i *and for* $i = 1, .., m$ *there are no* $\times$ *labels on* v_i. *For the lowerbound process, a path is active if in addition, for* $i = 1, ..., m-1$ *there is a* $\star$ *label on* v_i. *For the limit process, a path is active if instead of a* $\leftrightarrow$ *label at* h_i, x_{i+1} *belongs to the offspring of* (x_i, t_i).

Say a point (x, t) is *active* for a process η_t if $\eta_t(x) \neq 0$. It is easy to check for the contact process, the lowerbound process and the limit process that if (x, t) is active and there is an active path with base (x, t) and end (y, s) then (y, s) is active; one way to express this is that active paths take active points to active points. It is also true that (y, s) is active if and only if there is an active point $(x, 0)$ and an active path with base $(x, 0)$ and end (y, s). This implies a useful property called *additivity* which is discussed in the proof of Lemma 3.5.2 in Section 3.5.4.

3.4.4 Estimate of λ^+

We will use the method described in [10], applied to the lowerbound process on $\mathbb{Z}$, to obtain upper bounds on $\underline{\lambda_c}(\tau)$. Our first task is to describe the method and to justify its usage in this setting. Recall that $\mathcal{S} = G \times [0, \infty)$ is the spacetime set, for $R \subset \mathcal{S}$, $\mathcal{F}(R)$ is the σ-algebra generated by the restriction of the substructure of R. In what follows, we say an event holds "on R" if it is $\mathcal{F}(R)$-measurable.

Letting $L := \{(m, n) \in \mathbb{Z}^2 : n \geq 0, m + n \text{ is even}\}$, define *oriented site percolation* on L to be the process in which sites in L are independently *open* with probability p and *closed* with probability $1 - p$, and there is a *path* from (k, l) to (m, n)

or $(k,l) \to (m,n)$ if there is a list $(k,l) = (k_1,l_1),(k_2,l_2),...,(k_j,l_j) = (m,n)$ such that $l_{i+1} = l_i + 1$ and $k_{i+1} = k_i \pm 1$ for $i = 1,...,j-1$, and (k_i,l_i) is open for $i = 1,...,j-1$; note the last site is not required to be open. The *cluster* of (m,n) is the set $C(m,n) = \{(k,l) \in L : (m,n) \to (k,l)\}$. We say that *percolation occurs* from (m,n) if $|C(m,n)| = \infty$. If $p < 1$ is close enough to 1, then percolation occurs from $(0,0)$ with positive probability; as shown in [10], $p \geq 0.819$ suffices.

The idea of the method of [10] is to use the above result for oriented percolation to prove survival of the infection in the process of interest. In our case the discussion applies to the lowerbound process, the contact process, the limit process, and any process whose active paths take active points to active points as described above. Embed the spacetime set $(\mathbb{Z}, \{xy : |x-y| = 1\}) \times [0,\infty)$ into the half-space $\{(x,y) \in \mathbb{R}^2 : y \geq 0\}$ then draw the rectangles $\{R_{m,n} : (m,n) \in L\}$ defined by

$$R_{m,n} = R_{0,0} + (mK, nT)$$

for some K,T to be determined, where $R_{0,0} = [0,J] \times [0,T]$ for some integer $K \leq J < 2K$ and for a set S and a point r, $S + r := \{s + r : s \in S\}$; the range of J ensures that $R_{m,n}$ does not intersect $R_{m-2,n}$ or $R_{m+2,n}$, but does intersect $R_{m-1,n+1}$ and $R_{m+1,n+1}$ along its top edge; let j denote the number of sites along which $R_{m,n}$ intersects $R_{m-1,n+1}$, which is also the number of sites along which $R_{m,n}$ intersects $R_{m+1,n+1}$.

Fix a parameter $i \in \{1,...,j\}$. Say that a configuration η is good for $R_{m,n}$ if among either the leftmost or rightmost j sites in $R_{m,n}$ there are at least i distinct sites $x_1,...,x_i$ such that $\eta(x_k) \neq 0$ for $k = 1,...,i$. For $(m,n) \in L$, and for η that is good for $R_{m,n}$, define $A_{m,n}(\eta)$ as follows: $A_{m,n}(\eta)$ occurs if for at least i of the j leftmost sites $y_1,...,y_i$ and at least i of the j rightmost sites $y_{i+1},...,y_{2i}$ in $R_{m,n}$ at time $(n+1)T$, for $p = 1,...,2i$ there is an $q \in \{1,...,i\}$ such that there is an active path lying in $R_{m,n}$ with base (x_q, nT) and end $(y_p, (n+1)T)$. See Figure 3.1 for a picture.

If η_{nT} is good for $R_{m,n}$ and $A_{m,n}(\eta_{nT})$ occurs, then $\eta_{(n+1)T}$ is good for both $R_{m-1,n+1}$ and $R_{m+1,n+1}$. Moreover, given η_{nT} that is good for $R_{m_1,n}, R_{m_2,n}, ...$, the events $A_{m_1,n}(\eta_{nT}), A_{m_2,n}(\eta_{nT}), ...$ are independent. It is then straightforward to show that if for each $(m,n) \in L$, $\mathbb{P}(A_{m,n}) \geq p$ for every configuration that is good for $R_{m,n}$

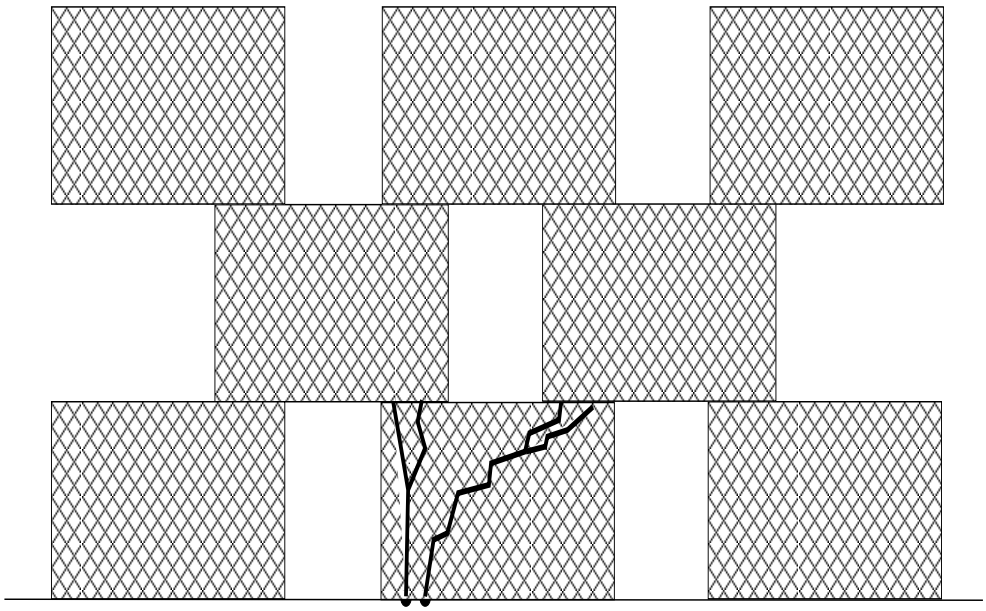

Figure 3.1: Depiction of rectangles $R_{m,n}$ for $-2 \leq m \leq 2$ and $0 \leq n \leq 2$, as well as the event $A_{0,0}(\eta_0)$ in $R_{0,0}$, with $i = 2$.

then if η_0 is good for $R_{0,0}$, the set $\{(m,n) \in L :$ there is an active point for η_t in $R_{m,n}\}$ stochastically dominates the cluster $C(0,0)$ of oriented site percolation with parameter p. Moreover, taking η at time 0 that is good for $R_{0,0}$ together with the set of labellings of $R_{0,0}$ belonging to the event $A_{0,0}$ given η and translating both by (mK, nT) gives η' that is good for $R_{m,n}$ together with the set of labellings belonging to $A_{m,n}$ given η'. Thus, by translation invariance of the law of the substructure, to show the lowerbound process survives with positive probability starting from η_0 that is good for $R_{0,0}$, it suffices to show that $\mathbb{P}(A_{0,0}) \geq 0.819$ for each η_0 that is good for $R_{0,0}$.

Using blocks with $J = 7$ and $K = 6$ and setting T so that there are an average of 650 labels in R (650 just chosen to match the choice in [10]), following [10] we estimate numerically the transition matrix P for the embedded jump chain corresponding to the lowerbound process restricted to the sites in R, counting redundant transitions (i.e. points in the Poisson process that have no effect) so that the rate of transitions is fixed and is equal to the total intensity of p.p.p.'s, that we denote γ, and then by computing the first couple of thousand terms in the sum

$$\sum_{i=1}^{\infty} e^{-\gamma T} \frac{\gamma^i}{i!} P^i$$

by monotonicity we obtain a lower bound on the entries of the transition matrix at time T in the continuous time chain. If $\lambda = 6.875$ and $\tau = 1/10$, then letting $i = 1$, with respect to the estimated transition matrix we find that for each configuration favourable for $R_{0,0}$, $A_{0,0}$ has probability at least 0.819, which implies $\underline{\lambda_c}(1/10) \leq 6.875$ and by monotonicity in τ, that $\lambda_c(\tau) < 6.875$ when $\tau \leq 1/10$, so that for the SEIS process $\lambda_c^+(\tau) < 6.875$ when $\tau \leq 1/10$.

3.4.5 Estimates of λ^-

For the upperbound process we use the method of [51]. Starting with the upperbound process on $\mathbb{Z}$ with $\tau > 0$ and initial configuration $\eta_0(x) = 3$ for $x \geq 0$ and $\eta_0(x) = 0$ for $x < 0$, for integer $m \geq 0$ we define the modified process $\eta_t^m(x)$ by evolving like the upperbound process but with the added constraint $\eta_t^m(x) = 3$ for $x > l_t^m + m$, where $l_t^m = \inf\{x : \eta_t^m(x) \neq 0\}$. The vector $v(t) = (\eta_t^m(l_t^m), ..., \eta_t^m(l_t^m + m)$ evolves like a finite state continuous time homogenous Markov chain on the state space $S = \{\eta \in \{0,1,2,3\}^{m+1} : \eta(0) \neq 0\}$, so we let X_n denote the embedded jump chain, which

evolves on the same space. For this chain each state communicates with the state defined by $X(x) = 3, x = 0, ..., m$, so the chain is irreducible on S with a unique invariant measure that we denote μ. Letting $k(i,j)$ denote the number of distinct transitions in η_t^m that take $v(t)$ from i to j, and letting $\Delta(i,j,r)$ be the increment in l_t^m and $p(i,j,r)$ be the transition probability for the r^{th} transition, $1 \leq r \leq k(i,j)$, we define λ_m as $\sup\{\lambda : \mathbb{E}_\mu \Delta > 0\}$, where

$$\mathbb{E}_\mu \Delta = \sum_{i,j \in S} \mu(i) \sum_{r=1}^{k(i,j)} p(i,j,r)\Delta(i,j,r) \tag{3.4.1}$$

is the average increment in l_t^m at each transition. Since $\eta_t^m \geq \eta_t$ and $\mathbb{E}_\mu \Delta > 0$ implies $l_t^m \to \infty$, if $\lambda < \lambda_m$ then $l_t \to \infty$. A simple coupling argument as in [7] then shows that the upperbound process on $\mathbb{Z}$ started from a finite number of active sites dies out with probability 1 when $l_t \to \infty$, which implies $\lambda_m \leq \overline{\lambda}_c$. To estimate λ_m we construct the embedded jump chain, compute μ, and iterate to find λ such that $\mathbb{E}_\mu \Delta \approx 0$. With $m = 3$ we obtain the table of values given in Theorem 3.2.2.

3.4.6 Estimate of λ_c^∞

Defining an active path for the limit process as in Definition 3.4.3, we have that active paths take active points to active points, so we can use the method of [10] as described. Doing so with $L = 10, j = 4$ and $i = 2$ and T chosen to give an average of 650 labels gives the upper bound on λ_c^∞.

To get a lower bound we use again the method of [51]. From the limit process ζ_t, for $m \geq 0$ the modified process ζ_t^m is defined by evolving like the limit process but with the added constraint $\zeta_t^m(x) = 1$ for $x > l_t^m + m$, with l_t^m as defined above. The definition of λ_m, μ and $\mathbb{E}_\mu \Delta$ are as above, with μ now supported on the state space $\{\eta \in \{0,1\}^{m+1} : \eta(0) \neq 0\}$. We give the coupling argument that shows $l_t \to \infty$ implies the limit process started from a finite number of active sites dies out with probability 1: letting $\mathbf{1}$ denote the indicator function define ζ_t^-, ζ_t^0 and ζ_t^+ by $\zeta_0^- = \mathbf{1}(x \leq 0)$, $\zeta_0^0 = \mathbf{1}(x = 0)$ and $\zeta_0^+ = \mathbf{1}(x \geq 0)$, then by monotonicity $\zeta_t^0 \leq \min \zeta_t^-, \zeta_t^+$. If $l_t^+ = \inf\{x : \zeta_t^+(x) \neq 0\} \to \infty$ then by symmetry $r_t^- = \sup\{x : \zeta_t^-(x) \neq 0\} \to -\infty$, moreover $r_t^- < l_t^+$ implies $\zeta_t^0 \equiv 0$, so there is almost surely finite time s so that $\zeta_t^0 \equiv 0$ for $t > s$. An analogous argument works when ζ_t^0 is replaced with ζ_t^N defined by $\zeta_0^N = \mathbf{1}(x \in [-N, N])$. Computing λ_m for $m = 8$ in the same way as above gives the

lower bound on λ_c^∞.

3.5 Theorem 3.2.3: Qualitative Estimates

In this section we prove Theorem 3.2.3, in five parts: existence of λ_0, upper bound on λ^+ as $\tau \to 0$, lower bound on λ^- as $\tau \to 0$, upper bound on λ^+ as $\tau \to \infty$, lower bound on λ^- as $\tau \to \infty$.

3.5.1 Existence of λ_0

To show the existence of λ_0 we use a comparison to oriented site percolation, in the spirit of [10]. As in the previous section, let $L := \{(m,n) \in \mathbb{Z}^2 : n \geq 0, m+n \text{ is even}\}$ and for T to be determined, define the set of rectangles $\{R_{m,n} : (m,n) \in L\}$ by $R_{0,0} = R = [0,3] \times [0,T]$ and

$$R_{m,n} = R + (2m, nT)$$

where for a set S and a point r, $S+r := \{s+r : s \in S\}$. For the SEIS process on $\mathbb{Z}$, the graphical representation embeds in a natural way into the set $\{(x,y) \in \mathbb{R}^2 : y \geq 0\}$, as do the rectangles $R_{m,n}$. In light of the upper bound on $\lambda^+(\tau)$ given for $\tau \leq 1/10$ in Theorem 3.2.2, it is enough to show that survival occurs with positive probability when $\lambda > \lambda_0$ and $\tau \geq \tau_0 > 0$ for some $\tau_0 \leq 1/10$; we phrase it in this way because we will be able to take $\tau_0 > 0$ as small as we choose.

For each m, let x_m denote the site such that $R_{m,n} = [x_m, x_m + 3] \times [nT, (n+1)T]$, when $m+n$ is even. To show survival we define, for each $p < 1$, a value λ_0 such that for each λ and τ there is a time T and a collection of events $A_{m,n}$ each one depending only on the corresponding $\mathcal{F}(R_{m,n})$, the restriction of the substructure to $R_{m,n}$, with the property that if $\eta_{nT}(y) \neq 0$ for some $y \in [x_m, x_m + 3]$ and $A_{m,n}$ occurs, then $\eta_{(n+1)T}(y) \neq 0$ for some $y \in [x_m, x_m + 1]$ *and* some $y \in [x_m + 2, x_m + 3]$, and such that if $\lambda > \lambda_0$, then $\mathbb{P}(A_{m,n}) \geq p$. Then, provided $\eta_0(0) \neq 0$, if $\lambda > \lambda_0$ the set of (m,n) such that $\eta_t(x) \neq 0$ for some $(x,t) \in R_{m,n}$ dominates the cluster $C(0,0)$ of oriented percolation with parameter p, so if p is chosen close enough to 1 we conclude that survival in the SEIS process occurs with positive probability uniformly in τ, for any value $\lambda > \lambda_0$.

For simplicity we define $A_{m,n}$ to have the property that if $\eta_{nT}(y) \neq 0$ for some $y \in [x_m, x_m+1]$ then $\eta_{(n+1)T}(y) \neq 0$ for some $y \in [x_m, x_m+1]$ and some $y \in [x_m+2, x_m+3]$. By reflection symmetry of the substructure, if $\mathbb{P}(A_{m,n}) \geq 1-\epsilon/2$ then the probability that the same conclusion holds with the weaker hypothesis $\eta_{nT}(y) \neq 0$ for some $y \in [x_m, x_m+3]$ is at least $1-\epsilon$. By translation-invariance it is enough to define $A_{0,0}$, for which $x_m = 0$. Letting $B(t) = (\eta_t(0), \eta_t(1))$ and $C(t) = (\eta_t(2), \eta_t(3))$, say that either is *active* at time t if at least one of its two coordinates is not zero. Then, it is enough that $A_{0,0}$ have the following two properties:

1. if $B(s)$ is active for some $s \in [0, T)$ then $B(t)$ is active for $s < t \leq T$, and the same is true for $C(t)$, and

2. if $B(0)$ is active (which it is by the assumption $\eta_0(0) \neq 0$) then for some $s \in [0, T)$, $C(s)$ is active

Before choosing T it is convenient to rescale the model in time so that $\star$ labels occur at rate 1, $\times$ labels at rate τ and $\leftrightarrow$ labels at rate $\lambda\tau$; this is ok since T is allowed to depend on τ, and is convenient since after rescaling, it won't have to.

To get property 1 for both $B(t)$ and $C(t)$ (and a bit more besides) it is enough that the following event depending on a fixed parameter $h > 0$, that we denote E_1, hold: in the time interval $[0, h]$ there is a $\leftrightarrow$ label on each of the edges $01, 12, 23$ before there is a $\times$ label on any of the sites $\{0, 1, 2, 3\}$ and everywhere on the set $[0, 3] \times [0, T]$, after any $\star$ label on $\{0, 1, 2, 3\}$ and within at most h amount of time, there is a $\leftrightarrow$ on each of the edges $01, 12, 23$ before there is a $\times$ label on any of the sites $\{0, 1, 2, 3\}$. Since after rescaling, $\star$ labels occur at rate 1, $\times$ labels at rate τ and $\leftrightarrow$ labels at rate $\lambda\tau$, it is a simple exercise to show that given $h, \tau_0 > 0$, for each $\epsilon, T > 0$ we can choose λ_0 so that $\mathbb{P}(E_1) \geq 1 - \epsilon$ if $\lambda > \lambda_0$ and $\tau > \tau_0$.

To get property 2, it is enough to show that $t \leq T$ where t is the first time such that $\eta_t(1) = 2$ and a $\leftrightarrow$ on 12 occurs at time t. If E_1 holds, it is enough to have one of the following:

- $\eta_0(1) = 2$ and $T \geq h$,

- $\eta_0(1) = 1$ and there is a $\star$ label at $(1, t)$ for some $t \leq T - h$,

- $\eta_0(0) = 2$ and there is a $\star$ label at $(1, t)$ for some $h < t \leq T - h$, and
- $\eta_0(0) = 1$ and there is a $\star$ label at $(0, t_1)$ for some $t_1 \leq T - 2h$ and a $\star$ label at $(1, t_2)$ for some $t_2 \in (t_1 + h, T - h]$

Since we can choose $T \geq h$ and since one of the four conditions on $B(0)$ holds by assumption, we denote by E_2 the intersection of the label events just described, so that E_2 depends only on $\mathcal{F}(R_{0,0})$ as desired. Given $\epsilon > 0$, it is not hard to check that for T large enough, $\mathbb{P}(E_2) \geq 1 - \epsilon/2$. Then, we can choose λ_0 so that if $\tau > \tau_0$ and $\lambda > \lambda_0$, $\mathbb{P}(E_1) \geq 1 - \epsilon/2$, so that $\mathbb{P}(E_1 \cap E_2) \geq 1 - \epsilon$, and this completes the proof.

3.5.2 Upper bound on λ^+ as $\tau \to 0$

We now show that $\lambda^+(\tau), \lambda^-(\tau) \to \lambda_c^0$, the critical value of the contact process, as $\tau \to 0$. We first use the lowerbound process to show that $\limsup_{\tau \to 0} \lambda^+(\tau) \leq \lambda_c^0$. In the proof we mention a 1-dependent oriented site percolation process with parameter p; this is a model in which each site is open with probability p, and sites $(m_1, n_1), ..., (m_k, n_k)$ are independent provided $|m_i - m_j| + |n_i - n_j| > 2$ for $i \neq j$. The definition of paths, clusters and percolation is the same as before. As shown in [15], for a 1-dependent oriented site percolation process, if $p = 1 - \epsilon$ for $\epsilon > 0$ small enough, then percolation from $(0, 0)$ occurs with positive probability.

In [15] it is shown for the contact process that if $\lambda > \lambda_c^0$ then for each $\epsilon > 0$, for a suitable choice of rectangles $R_{m,n} \in \mathcal{S}$ and $\mathcal{F}(R_{m,n})$-measurable events $A_{m,n}$ with indicator $I_{m,n}$, the set $\{(m, n) \in L : I_{m,n} = 1\}$ dominates a 1-dependent oriented site percolation process with parameter $p = 1 - \epsilon$. The 1-dependence arises from the fact that $R_{m,n}$ overlaps with each of $R_{m\pm1,n\pm1}$ on a set of positive measure in $\mathcal{S}$; see Fig. 1 in [15] for a picture. The events $A_{m,n}$ are such that for appropriate choice of initial configuration with finitely many active sites, the set of $(m, n) \in L$ such that there is an active point in $R_{m,n}$ stochastically dominates $C(0, 0)$, so to show survival it is *sufficient* to have $\mathbb{P}(A_{m,n}) \geq 1 - \epsilon$ for $\epsilon > 0$ small enough.

Now, since each rectangle $R_{m,n}$ is bounded, if $\mathbb{P}_\lambda(A_{m,n}) > p$ then for small enough $\delta > 0$, $\mathbb{P}_{\lambda-\delta}(A_{m,n}) > p$. To see this, proceed as in Section 3.1: first generate transmission labels $\leftrightarrow$ and $\leftrightarrow'$ using independent p.p.p's with intensity $\lambda - \delta$ and δ respectively, then let the process with transmission parameter λ use both types of labels, and let

the process with transmission parameter $\lambda - \delta$ use only the labels $\leftrightarrow$. Since R is a bounded region in spacetime, with probability tending to 1 as $\delta \to 0$ there are no $\leftrightarrow'$ labels in R, and on this event the p.p.p.'s for the two processes agree on R. Viewing the contact process as the lowerbound process with $\tau = 0$, a similar argument shows that if $\mathbb{P}_{\lambda,0}(A_{m,n}) > p$ then for small enough δ and τ, $\mathbb{P}_{\lambda-\delta,\tau}(A_{m,n}) > p$; for a detailed argument see [19], but the main idea is that for a bounded region in spacetime, with probability tending to 1 as $\tau \to 0$, between any two consecutive labels of type $\leftrightarrow$ or $\times$ in time there is a $\star$ label at every site. Therefore if $\lambda > \lambda_c^0$ then taking $p = 1 - \epsilon$ where $\epsilon > 0$ is such that percolation occurs with positive probability in a 1-dependent oriented site percolation model with parameter p, for small enough $\delta, \tau > 0$, started from some initial configuration with finitely many active sites the lowerbound process with parameters $\lambda - \delta, \tau$ survives with positive probability. In other words, $\lambda > \lambda_c^0$ implies $\lambda > \lambda_c(\tau)$ for small enough $\tau > 0$, and noting that $\lambda^+(\tau) \leq \lambda_c(\tau)$ the desired result follows.

3.5.3 Lower bound on λ^- as $\tau \to 0$

Next we show that $\liminf_{\tau\to 0} \lambda^-(\tau) \geq \lambda_c^0$, using the upperbound process and a comparison. In Section 3.4 we described the method of [51] that gives, for each $\tau > 0$, a sequence of lower bounds $\lambda_0(\tau) \leq \lambda_1(\tau) \leq \lambda_2(\tau) \leq ... \leq \overline{\lambda}_c(\tau)$, each is which is determined by a process η_t^m that approximates the upperbound process; we did not show that the λ_m are increasing but this follows from the observation $\eta_t^m \geq \eta_t^{m+k}$ for $k \geq 0$. For $\tau = 0$ which is the contact process, starting from the process with initial configuration $\eta_0(x) = 2$ for $x \geq 0$ and $\eta_0(x) = 0$ for $x < 0$, then for integer $m \geq 0$ define η_t^m by evolving like the contact process but with the added constraint $\eta_t^m(x) = 2$ for $x > l_t^m + m$, with again $l_t^m = \inf\{x : \eta_t^m(x) \neq 0\}$. The definition of λ_m, μ and $\mathbb{E}_\mu \Delta$ are as before. Since $\eta_t^m \geq \eta_t^{m+k}$ we have again $\lambda_0(0) \leq \lambda_1(0) \leq ...\lambda_c^0$. It is shown in [51] that $\lambda_m(0) \uparrow \lambda_c^0$; the proof relies on the fact, proved as Theorem 4 in [22], that $\lambda_c^0 = \sup\{\lambda : \alpha > 0\}$, where for $l_t = \inf\{x : \eta_t(x) \neq 0\}$, $\alpha = \lim_{t\to\infty} l_t/t$ was shown to exist in [7].

For $\tau > 0$, μ is supported on $S = \{\eta \in \{0,1,2,3\}^{m+1} : \eta(0) \neq 0\}$ and for $\tau = 0$, μ is supported on $S_0 = \{\eta \in \{0,2\}^{m+1} : \eta(0) \neq 0\} \subset S$. Define $S_1 = \{\eta \in S : \eta(x) \in \{1,3\}$ for some x and $\eta(y) \in \{0,2\}, y \neq x\}$, the set of states with exactly one exposed site. Modify the embedded jump chain X_n when $\tau = 0$ to include inter-

mediate transitions to configurations with an exposed i.e. type 1 or type 3 site; μ is then supported on $S_0 \cup S_1$ and from any state in S_1 there is a unique transition with probability 1 to the corresponding state in S_0 in which onset of the exposed site has occurred. Since this modification preserves the sign of $\mathbb{E}_\mu\Delta$ the value of λ_m is unchanged. Writing $\mathbb{E}_\mu\Delta(m,\lambda,\tau)$ to emphasize the dependence and noting that $\lambda_m(\tau) = \sup\{\lambda : \mathbb{E}_\mu\Delta(m,\lambda,\tau) > 0\}$, to prove the result it suffices to show that for each m and λ, $\mathbb{E}_\mu\Delta(m,\lambda,\tau)$ is continuous in τ at $\tau = 0$, since if $\lambda < \lambda_c^0$ then $\lambda < \lambda_m(0)$ for some m and $\mathbb{E}_\mu\Delta(m,\lambda,0) > 0$ and then by continuity $\mathbb{E}_\mu\Delta(m,\lambda,\tau) > 0$ and thus $\lambda < \lambda_m(\tau)$ for small enough $\tau > 0$ and the result follows from the inequalities $\lambda_m(\tau) \le \overline{\lambda}_c(\tau) \le \lambda^-(\tau)$.

Recall equation (3.4.1):

$$\mathbb{E}_\mu\Delta = \sum_{i,j\in S} \mu(i) \sum_{r=1}^{k(i,j)} p(i,j,r)\Delta(i,j,r)$$

Thus to show $\mathbb{E}_\mu\Delta(m,\lambda,\tau) \to \mathbb{E}\Delta(m,\lambda,0)$ as $\tau \to 0$ it suffices to show $\mu(i)(m,\lambda,\tau) \to \mu(i)(m,\lambda,0)$ for $i \in S$ and $p(i,j,r)(m,\lambda,\tau) \to p(i,j,r)(m,\lambda,0)$ as $\tau \to 0$ for $i,j \in S$ and $r = 1,...,k(i,j)$. In fact, we can make a further reduction.

Lemma 3.5.1. *Let $p(i,j)(s)$, $0 \le s \le 1$, be a family of transition probabilities on a finite state space S such that for each s, $p(i,j)(s)$ has a unique invariant measure $\mu(i)(s)$. If $p(i,j)(s) \to p(i,j)(0)$ as $s \to 0$ for each $i,j \in S$ then $\mu(i)(s) \to \mu(i)(0)$ as $s \to 0$ for $i \in S$.*

Proof. Let $n = |S|$ and define the simplex $\Lambda = \{x \in \mathbb{R}^n : x_i \ge 0, i = 1,...,n,\ \sum_i x_i = 1\}$ that corresponds to probability measures on S. Suppose by way of contradiction that there is a sequence (s_k) tending to 0 and an $\epsilon > 0$ such that $\max_i |\mu(i)(s_k) - \mu(i)(0)| > \epsilon$ for each k. By compactness of Λ there is a subsequence (s_{k_m}) tending to 0 and an element $\mu^* \in \Lambda$ with $\mu^* \ne \mu(0)$ such that $\mu(i)(s_{k_m}) \to \mu^*(i)$ for each i. However, since $p(i,j)(s_{k_m}) \to p(i,j)(0)$ for each (i,j) and each $\mu(s_{k_m})$ is invariant for $p(i,j)(s_{k_m})$, μ^* is invariant for $p(i,j)(0)$, and by assumption of uniqueness, $\mu^* = \mu(0)$, a contradiction. □

By the lemma above, since for X_n, $p(i,j) = \sum_{r=1}^{k(i,j)} p(i,j,r)$, and μ is determined from $p(i,j)$ it suffices to show $p(i,j,r)(m,\lambda,\tau) \to p(i,j,r)(m,\lambda,0)$ as $\tau \to 0$. For

each $i, j \in S$ and $r \in \{1, ..., k(i,j)\}$ the transition rate q_{ijr} for η_t^m is of the form $1/\tau$, λ, 2λ or 1. Using the formula $p_{ijr} = q_{ijr} / \sum_{(k,s)\neq(i,r)} q_{iks}$ and writing $p_{ijr}(m, \lambda, \tau)$ to emphasize the dependence, fix m and λ. For $i \in S_0$, $p_{ijr}(m, \lambda, \tau)$ does not depend on τ and agrees with $p(i,j,r)(m, \lambda, 0)$ provided we include intermediate transitions in X_n as discussed above, and for $i \in S_1$ there is a unique $j \in S_0$ and $r \in \{1, ..., k(i,j)\}$ such that $p_{ijr}(m, \lambda, \tau) \uparrow 1$ as $\tau \downarrow 0$; j and r are determined by forcing onset to occur, which again agrees with X_n. Defining $S_2 = S \setminus (S_0 \cup S_1)$, for $i \in S_2$ each $p(i,j,r)$ converges to some number $a(i,j,r)$ such that $\sum_{j\in S, r=1,...,k(i,j)} a(i,j,r) = 1$ and such that with respect to $a(i,j,r)$, $S_0 \cup S_1$ is accessible from S_2; since for $\tau = 0$, $S_0 \cup S_1$ is invariant, we can define $p(i,j,r)(m, \lambda, 0) = a(i,j,r)$ without affecting X_n, so convergence of the $p(i,j,r)$ is proved, & we are done.

3.5.4 Upper bound on λ^+ as $\tau \to \infty$

We use the same approach as in the case $\tau \to 0$, and begin by describing the construction of [15] in somewhat better detail.

As mentioned in Section 3.5.2, the idea of [15] is that given $p < 1$, we can choose rectangles $R_{m,n}$ and initial data for the process so that the set of $(m,n) \in L$ such that there is an active point in $R_{m,n}$ dominates the cluster $C(0,0)$ in a 1-dependent oriented percolation model with parameter p. The specific event on $R_{m,n}$ that allows this is the existence of active paths going from the bottom centre of $R_{m,n}$ to the top left and top right, with the property that if, say, $R_{m,n}$ and $R_{m-1,n+1}$ both have these paths, then said paths can be concatenated to form a longer active path through both rectangles.

Our strategy is first to show that the construction of [15] applies to the limit process, so that if $\lambda > \lambda_c^\infty$ then for any $p < 1$ we can choose $R_{m,n}$ for the limit process that have the desired active paths, in a 1-dependent way, with probability $\geq p$ for each (m,n). Then, by defining a condition for initial data at the base of $R_{m,n}$, and an event on each rectangle $R_{m,n}$, that allow us to deal with the possibility of discrepancies as encountered in the proof of Theorem 3.2.1 in Section 3.3, we can show that with nearly the same probability, for τ large enough the SEIS process has the same active paths, and these paths can be concatenated. First we address applicability of the

construction of [15] to the limit process.

Lemma 3.5.2. *The comparison to oriented percolation given in [15] is valid for the limit process on $\mathbb{Z}$, i.e., if $\lambda > \lambda_c^\infty$ and $\epsilon > 0$ then suitably rescaled, the limit process dominates oriented site percolation with parameter $p > 1 - \epsilon$.*

Proof. First note that our definition of the critical value is the same as theirs, namely of survival with positive probability starting from a single infectious site. As mentioned at the end of Section 2 of [15], their construction is valid for a broader class of models they call "nearest neighbour additive growth models" that includes the limit process. The key properties required are additivity, described in Section 4 of [7], and the coupling property described in Lemmas 3.1 and 3.4 of [7], both of which are easily verified to hold for the limit process. Additivity means that for configurations ζ and ζ', defining $\zeta \vee \zeta'$ for each x by $(\zeta \vee \zeta)(x) = \max(\zeta(x), \zeta'(x))$, then with respect to the coupling given by the graphical construction, if $\zeta_0'' = \zeta_0 \vee \zeta_0'$ then $\zeta_t'' = \zeta_t \vee \zeta_t'$ for $t > 0$. The desired coupling property is that, if we let $\zeta_0^1(x) \equiv 1$, $\zeta_0^+(x) = \mathbf{1}(x \geq 0)$ and $\zeta_0^0(x) = \mathbf{1}(x = 0)$ then for $l_t^+ = \inf\{x : \zeta_t^+(x) \neq 0$, $l_t^0 = \inf\{x : \zeta_t^0(x) \neq 0\}$ or $= \infty$ if $\zeta_t^0(x) = 0$ for all x, and $r_t^0 = \inf\{x : \zeta_t^0(x) \neq 0\}$ or $= -\infty$ if that set is empty, it holds that $\zeta_t^0 = \zeta_t^+ \cap [l_t^0, r_t^0] = \zeta_t^1 \cap [l_t, r_t]$ i.e., on the interval $[l_t, r_t]$, ζ_t^0 agrees with ζ_t^+ and with ζ_t^1, and if $r_t^0 > -\infty$ then $l_t^0 = l_t^+$. The coupling property can be checked by examining transitions occurring near the endpoints l_t^0, r_t^0. □

Next we introduce two slightly different definitions of active path for the SEIS process that will be helpful. An active path literally implies that all points along that path are active, while a potentially active path will, under some additional conditions on the state of the process near the base of the path and on the surrounding substructure, also be active.

Definition 3.5.3. *For the SEIS process η_t, given η_0 and $\tau > 0$, a path as defined in Definition 3.4.2 is* active *if for $i = 1, ..., m-1$, $\eta_{t_i}(x_i) = 2$, $\eta_{t_i}(x_{i+1}) = 1$, there is a $\leftrightarrow$ label at h_i, and for $i = 1, ..., m$, $\eta_t(x_i) \neq 0$ for $t \in (t_{i-1}, t_i)$.*

Definition 3.5.4. *For the SEIS process and a path $\gamma = (v_1, h_1, ..., v_{m-1}, h_{m-1}, v_m)$ as defined in Definition 3.4.2 say that γ is* potentially active *if*

1. *for $i = 1, ..., m-1$, there is a $\leftrightarrow$ label at h_i,*
2. *for $i = 1, ..., m-1$ there is a $\star$ label at a point $(x_i, t) \in v_i = \{x_i\} \times (t_{i-1}, t_i)$ called the* activating *label, which is such that there are no $\star$ labels in $\{x_i\} \times (t_{i-1}, t)$ and no $\times$ labels in $\{x_i\} \times (t, t_i)$, and*

3. *for any $\star$ label at a point $(x_m, t) \in v_m = \{x_m\} \times (t_{m-1}, t_m)$ there are no $\times$ labels in $\{x_m\} \times (t, t_m)$.*

Next we generalize the condition on the initial configuration given in Theorem 3.2.1, and the event described in Lemma 3.3.1, to a larger class of spacetime sets. Together these are probably the simplest conditions under which the SEIS process is well behaved.

Definition 3.5.5. *Let $R \subset \mathcal{S}$ be a set in spacetime which is the closure of an open set, and define the* base *of R as*

$$\text{base}(R) = \{(x,t) \in R : (x, t-\epsilon) \notin R \text{ for all small enough } \epsilon > 0\}$$

For the rescaled SEIS process, say that R is onset-ordered *if the event described in Lemma 3.3.1 holds on R i.e. if there is a $\star$ label at $(x,t) \in R$ and $\{x\} \times (t, t') \subset R$ for $t < t'$ then for any $\star$ label at a point (y, s), $s \in (t, t')$ there is a $\times$ label at $\{x\} \times s'$ for some $s' \in (t, s)$. Say that η_t is* good *for R if $\eta_t(x) \neq 2$ for $(x,t) \in \text{base}(R)$.*

The following result allows us to promote a potentially active path to an active path, when the above-stated conditions hold.

Lemma 3.5.6. *Let R be as in Definition 3.5.5. If R is onset-ordered and η_t is good for R then*

1. *for each $t \geq 0$, $\eta_t(x) = 2$ for at most one x in the set $\{x \in V : (x,t) \in R\}$, and*
2. *a potentially active path γ with base (y, s) satisfying $\eta_s(y) = 1$ is active in the sense of Definition 3.5.3.*

Proof. If $\eta_t(x) = 2$ for $(x,t) \in R$ then by tracing back in time along the fiber $\{x\} \times \mathbb{R}^+$, there is a point (x, s), $s \leq t$ such that if $s < t$ then $\{x\} \times (s, t) \in R$ and there are no $\times$ labels in $\{x\} \times (s, t)$, and such that either

1. $(x, s) \in \text{int}(R)$, the interior of R, $\eta_{s-\epsilon}(x) = 1$ for all small enough $\epsilon > 0$ and there is a $\star$ label at (x, s) or
2. $(x, s) \in \text{base}(R)$ and $\eta_s(x) = 2$

If η_t is good for R then the second case does not occur. Suppose $\eta_t(x) = \eta_t(y) = 2$ for some $x \neq y$, and let s_x, s_y be the times such that (x, s_x) and (y, s_y) have the property

stated above; we may assume that $s_x \leq s_y$, and since labels almost surely do not occur simultaneously, that $s_x < s_y$. But then there is an interval $\{x\} \times (s_x, t) \in R$ and a $\star$ label at a point $(y, s_y) \in R$, $s_x < s_y < t$, such that there are no $\times$ labels in the interval $\{x\} \times (s_x, s_y)$, so R is not onset-ordered. The second statement follows from the first statement, the definition of potentially active path, and the transition rules. □

The next result allows us, under fairly mild conditions, to concatenate a collection of potentially active paths into a longer potentially active path. The words before, after, starts, etc. are with respect to the order of events in time.

Lemma 3.5.7. *Let R be as in Definition 3.5.5 and suppose $\gamma_1, ..., \gamma_k$ are potentially active paths on R with respective bases $(x_1, t_1), ..., (x_k, t_k)$ satisfying*

1. *$t_i < t_{i+1}$, $i = 1, ..., k-1$ i.e., γ_i starts before γ_{i+1}, and*
2. *(x_{i+1}, t_{i+1}) does not intersect γ_i, $i = 1, ..., k-1$ i.e., γ_{i+1} does not start somewhere on γ_i,*

and with intersection points $(y_i, s_i) \in \gamma_i \cap \gamma_{i+1}$, $i = 1, ..., k-1$, $y_i \in V$, satisfying

1. *$s_i < s_{i+1}$, $i = 1, ..., k-2$,*
2. *$y_i \neq y_{i+1}$, $i = 1, ..., k-2$, and*
3. *y_{k-1} is not on the last vertical segment of γ_k.*

If R is onset-ordered then the path γ obtained by concatenating $\gamma_1, ..., \gamma_k$ through the points (y_i, s_i), $i = 1, .., k-1$ is a potentially active path.

Proof. For $i \in 1, .., k-1$ let u_i, v_i be the vertical segments in γ_i, γ_{i+1} respectively that satisfy $(y_i, s_i) \in u_i \cap v_i$ and let w_i be the vertical segment in γ containing (y_i, s_i). For $i = 1, ..., k-1$, w_i is not the last vertical segment in γ, $w_i \subset u_i \cup v_i$ and more precisely, $w_i = \{y_i\} \times [\min u_i, \max v_i]$ where $\min u_i$ is the lowest point (in time) on u_i and $\max v_i$ is the highest point on v_i; for $i = 1, ..., k-2$ this follows from the assumption $y_i \neq y_{i+1}$ and for $i = k-1$ it follows from the assumption y_{k-1} is not on the last verical segment of γ_k. Except for the segments w_i, $i = 1, ..., k-1$, the segments of γ correspond to segments in the paths $\gamma_1, ..., \gamma_k$, so it suffices to check for $i = 1, .., k-1$ that on w_i there is an activating label i.e. a $\star$ label with no $\times$ labels on w_i after it and no $\star$ labels on w_i before it; we consider separately the cases

$\min v_i < \min u_i$, $\min u_i = \min v_i$ and $\min u_i < \min v_i$.

If $\min v_i < \min u_i$ then since the base of γ_i comes before the base of γ_{i+1} there is a vertical segment u_{i-1} in γ_i that precedes u_i. Since neither u_{i-1} nor v_i are the last vertical segments on their respective paths, there is a $\star$ label on u_{i-1} with no $\times$ labels on u_{i-1} after it, and a $\star$ label ℓ on v_i with the analogous property. Since R is onset-ordered, it follows that there are no $\star$ labels between the time ℓ occurs and $\max v_i$, so ℓ cannot occur before $\min u_i$. Thus ℓ occurs after $\min u_i$ and lies on w_i. It is easy to check that ℓ is an activating label for w_i.

If $\min u_i = \min v_i$ then $w_i = v_i$ and w_i inherits the activating label from v_i. If $\min u_i < \min v_i$, then since γ_{i+1} does not start on γ_i, there is a vertical segment v_{i-1} preceding v_i. Any $\star$ label on u_i has no $\times$ labels on u_i after it. From the same argument as in the previous paragraph with v_{i-1} playing the role of u_{i-1} and u_i playing the role of v_i, it follows that there are no $\star$ labels on $u_i \setminus v_i$, and so w_i inherits the activating label from v_i. □

After assembling the above ideas, we can prove in a straightforward way the domination of a 1-dependent oriented percolation process. Once we have shown this is true, the rest of the proof of the upper bound on λ^+ proceeds in the same way as when $\tau \to 0$, which we leave to the reader.

Lemma 3.5.8. *For fixed $\lambda > \lambda_c^\infty$ and $\epsilon > 0$, for τ large enough the SEIS process with parameters λ, τ dominates a 1-dependent oriented percolation process with parameter $p \geq 1 - \epsilon$.*

Proof. First, construct the limit process ζ_t and apply the construction of [15] to obtain rectangles $R_{m,n} = (mK, nT) + [-J, J] \times [0, 1.2T]$ such that the corresponding events $A_{m,n}$ have $\mathbb{P}(A_{m,n}) \geq 1 - \epsilon/3$. As specified more precisely in [15], each event $A_{m,n}$ corresponds to the existence of some active paths in $R_{m,n}$ going from the base of $R_{m,n}$ to some locations in the top of $R_{m,n}$, such that if $A_{m_i,n_i} = 1$ for each (m_i, n_i) in a path $(0,0) = (m_1, n_1), ..., (m_k, n_k)$, then the resulting paths intersect to produce an active path from some point $(y, 0)$, $y \in [-J, J]$ to a point (x, t) in R_{m_k,n_k}.

Independently, construct the rescaled SEIS process η_t and superimpose onto it the rectangles $R_{m,n}$ from the last paragraph. For $\delta > 0$, define the region $P_{m,n} = (mK, NT) + [-J, J] \times [-\delta, 0]$ that lies just below $R_{m,n}$ in spacetime, and say that

$P_{m,n}$ is good for $R_{m,n}$ if, depending only on the labelling on $P_{m,n}$, for all possible configurations $\eta_{nT-\delta}$, η_{nT} is good for $R_{m,n}$. It is easy to check that for τ large enough and $\delta > 0$ small enough, with probability $\geq 1 - \epsilon/3$, $P_{m,n}$ is good for $R_{m,n}$, due to the resulting plenitude of $\times$ labels and lack of $\star$ labels on $P_{m,n}$. Suppose η_t is good for $R_{m,n}$, then construct a new copy $\zeta_t^{(m,n)}$ of the limit process only on $R_{m,n}$ using the method of Section 3.3 with initial configuration $\{\eta_{nT}(x) : (x, nT) \in R_{m,n}\}$, and note that when $R_{m,n}$ is onset-ordered, to each active path in $\zeta_t^{(m,n)}$ corresponds a potentially active path for η_t on $R_{m,n}$ in the obvious way. Let $Q_{m,n} = R_{m,n} \cup P_{m,n}$ and define new events $B_{m,n}$ on $Q_{m,n}$ by

$$B_{m,n} = \{P_{m,n} \text{ is good for } R_{m,n},\ R_{m,n} \text{ is onset-ordered and } A_{m,n} \text{ holds for } \zeta_t^{(m,n)}\}$$

If τ is large enough then $\mathbb{P}(R_{m,n} \text{ is onset-ordered }) \geq 1 - \epsilon/3$, and if $\delta > 0$ is small enough then the oriented percolation model defined by the events $A'_{m,n}$ is 1-dependent, so for τ large enough and $\delta > 0$ small enough, $A'_{m,n}$ is 1-dependent and $P(A'_{m,n}) \geq 1 - \epsilon$.

If η_t is good for a finite collection of sets then it is good for the union of those sets. Moreover, if $(m_1, n_1), ..., (m_k, n_k)$ is a lattice path, $R_{m_1,n_1}, ..., R_{m_k,n_k}$ are onset-ordered and $A_{m_1,n_1}, ..., A_{m_k,n_k}$ holds for $\zeta_t^{(m_1,n_1)}, ..., \zeta_t^{(m_k,n_k)}$, then it is easy to see from the geometry of the construction in [15] that the potentially active paths for η_t in the rectangles $R_{m_1,n_1}, ..., R_{m_k,n_k}$ corresponding to the relevant active paths for $\zeta_t^{(m_1,n_1)}, ..., \zeta_t^{(m_k,n_k)}$ satisfy the conditions of Lemma 3.5.7. Thus if $B_{m_1,n_1} \cap ... \cap B_{m_k,n_k}$ holds for a path $(m_1, n_1), ..., (m_k, n_k)$ then in each of the rectangles $R_{m_1,n_1}, ..., R_{m_k,n_k}$ there is a potentially active path for η_t, such that the concatenation of those paths is a potentially active path through the union $R_{m_1,n_1} \cup .... \cup R_{m_k,n_k}$ which, since η_t is good for the union, is an active path for η_t, and the desired stochastic domination follows. □

3.5.5 Lower bound on λ^- as $\tau \to \infty$

An argument given in [22] shows that the coupling property described in the proof of Lemma 3.5.2 implies that $\lambda_c^\infty = \sup\{\lambda : \alpha > 0\}$, where $\alpha = \lim_{t\to\infty} l_t^+/t$; note that our definition of critical value agrees with the one in [22], namely as survival with positive probability, started from a single infectious site. With this fact, the proof given in [51] that $\lambda_m \uparrow \lambda_c$ applies without modification to the limit process. We then

proceed in the same way as when $\tau \to 0$. In this case we shall have S as before, $S_0 = \{\eta \in \{0,1,2,3\} : \eta(x) \in \{0,1\}, x = 1, ..., m+1\}$ and $S_1 = \{\eta \in \{0,1,2,3\} : \eta(x) = 2 \text{ for some } x \text{ and } \eta(y) \notin \{2,3\}, y \neq x\}$, and modify X_n when $\tau = \infty$ to include intermediate transitions through the infectious state and propagation before recovery. Since the proof is analogous, the details are omitted.

Acknowledgements

The author's research is partly supported by an NSERC PGSD2 Scholarship.

Chapter 4

New Results for the Two-Stage Contact Process

This chapter consists of the journal article [19] of the same title, appearing in Journal of Applied Probability 52.1 (March 2015).

Abstract

Here we continue the work started by Steve Krone on the two-stage contact process. We give a simplified proof of the duality relation, and answer most of the open questions posed in that paper. We also fill in the details of an incomplete proof.

4.1 Introduction

The contact process is a stochastic spatial model of population growth that was introduced in the 1970's [24] and has since been widely studied. In its simplest form, each site is in one of two states, occupied or vacant, and a great deal is known about this model (see for example the books [32] and [34] of Liggett). More recently, variants of the model have been studied in which there is more than one type of individual [38], [30] or more than one stage of development [29]. In [49], a multitype framework is used to model infection spread between households located at the vertices of a homogenous tree. In [31], a two-stage model of infection spread is studied on scale-free networks.

Here we consider the two-stage contact process introduced in [29]. It is a natural generalization of the contact process in which there is an intermediate juvenile type that must mature before it can produce offspring. More precisely, it is a growth model on $\mathbb{Z}^d$ defined by the rates

$$\begin{aligned} 0 &\to 1 \text{ at rate } \lambda n_2 \\ 1 &\to 2 \text{ at rate } \gamma \\ 1 &\to 0 \text{ at rate } 1+\delta \\ 2 &\to 0 \text{ at rate } 1 \end{aligned}$$

where $n_2(x)$ is the cardinality of the set $\{y \in \mathbb{Z}^d : 0 < \|y-x\|_\infty < r,\ y \text{ is in state } 2\}$ for some fixed $r \geq 1$. The state space for the process is $\{0,1,2\}^{Z^d}$, so that each site is either unoccupied, recently occupied, or occupied by a mature organism that can give birth at other sites. Aside from the choice of neighbourhood, there are three parameters λ, γ and δ, respectively the transmission rate, the maturation rate, and the juvenile death rate.

A number of basic properties of the process are proved in [29], including additivity and monotonicity with respect to parameters (increasing in λ and γ and decreasing in δ), as well as a duality relation, and some bounds on the survival region (the set of parameters for which an initially finite population has a chance of surviving for all time).

Here we consider the process in the more general setting of a countable graph (V, E) with finite maximum degree, proving some results in this setting and more precise results on $\mathbb{Z}^d$. We simplify the proof of the duality relation given in [29] and answer most of the open questions posed in Section 4 of that paper. As we shall see, for the two-stage contact process there is a critical value of the maturation rate below which survival does not occur (Theorem 4.1.1). Also, it shares many of the properties of the contact process; in particular, there is complete convergence (Theorem 4.1.3). The following is a summary of the main results.

Our first result is an upper bound on the set of values γ so that the process dies out, i.e., reaches the all 0 state with probability 1. The bound depends only on the maximum degree $M = \max_x \deg x$ of the graph.

Theorem 4.1.1. *If $\gamma < 1/(2M-1)$ then starting from any finite number of occupied sites, the process dies out, no matter the value of λ and δ.*

This answers question 6 in [29], where the author supplies a bound for $\mathbb{Z}^1$ in the case of nearest neighbour interactions, and asks whether a bound exists for other interactions, or for $\mathbb{Z}^d$ with $d > 1$.

Our next result shows that two notions of survival for the two-stage contact process coincide, answering question 1 in [29] affirmatively. For terminology see Sections 4.3.1 and 4.3.5.

Theorem 4.1.2. *For the two-stage contact process on $\mathbb{Z}^d$, single-site survival occurs if and only if the upper invariant measure is non-trivial.*

The proof uses the construction of [3] to show that for both the process and its dual, single-site survival implies the upper invariant measure is non-trivial.

An important question for growth models is that of *complete convergence*, which we show is true for the two-stage contact process, answering question 3 in [29]. Here λ_c is the critical value for single-site survival as defined in Section 4.3.1 and μ_t denotes the distribution of the process at time t. The $\Rightarrow$ denotes weak convergence.

Theorem 4.1.3. *If $\lambda > \lambda_c$ then complete convergence holds, i.e.,*

$$\xi_t \Rightarrow \alpha\delta_0 + (1-\alpha)\nu$$

where ν is the upper invariant measure, δ_0 concentrates on the configuration with all 0's and $\alpha = \mathbb{P}_{\mu_0}(\xi_t$ dies out $)$ where $\mathbb{P}_{\mu_0}$ is the law of the process with initial distribution μ_0.

We now summarize the organization of the paper. In Section 4.2 we construct the process and prove the duality relation. In Section 4.3.1 we recall the critical values defined in [29]. In Section 4.3.2 we fill in some missing details in the proof of Proposition 3.6 in [29]. In Section 4.3.3 we prove Theorem 4.1.1. In Section 4.3.3 we give a sufficient condition for the edge speed of the process in one dimension to characterize survival, providing a partial answer to question 2 in [29]. In Section 4.3.5 we prove Theorem 4.1.2, and in Section 4.3.6 we prove Theorem 4.1.3. We discuss the survival region in Section 4.3.7, and using the construction from the proof of Theorem 4.1.2

we find that the process dies out on the boundary of the survival region, providing a partial answer to question 5. We argue that question 4 appears not to have an affirmative answer, and we give some informal arguments as to why this should be so.

4.2 Construction and duality

We recall briefly the construction of the process. Here the process ξ_t lives on the state space $\{0,1,2\}^V$ where V is the vertex set of an undirected graph (V,E), with $V = \mathbb{Z}^d$ and $E = \{xy : 0 < \|x-y\|_\infty < r\}$ for some $r \geq 1$ being common choices. The state space is equipped with the partial order $\xi \leq \xi' \Leftrightarrow \xi(x) \leq \xi'(x)$ for each $x \in V$, where $0 < 1 < 2$ is the order on the state at each site. The process is *attractive* if there exists a coupling so that $\xi_0 \leq \xi_0' \Rightarrow \xi_t \leq \xi_t'$ for $t > 0$. It is *additive* if $\xi_0 = \xi_0' \vee \xi_0'' \Rightarrow \xi_t = \xi_t' \vee \xi_t''$, where $(\xi \vee \xi')(x) = \max(\xi(x), \xi'(x))$ for each x. It is *monotone increasing* (*decreasing*) with respect to a parameter λ if $\xi_0 \leq \xi_0'$ and $\lambda \leq \lambda'$ $(\lambda \geq \lambda') \Rightarrow \xi_t \leq \xi_t'$. We shall often use the word *active* to refer to a site or a point in spacetime where the state is not 0.

We can construct the process on any undirected graph (V,E) by taking $n_2(x)$ to be the cardinality of the set $\{y \in V : xy \in E,\ y \text{ is in state } 2\}$. Assign independent Poisson processes to each of the events:

- death of 1's and 2's at each site, at rate 1
- death of 1's at each site at the additional rate δ
- transmission across each edge at rate λ
- maturation at each site, at rate γ

Place the events on the spacetime graph $V \times \mathbb{R}^+$ and fix a configuration at time 0. The configuration at later times can then be determined from the events on the graph. To ensure it is well-defined it suffices to work backwards from a point (x,t) on the spacetime graph and ensure that with probability 1, only finitely many events occur that can influence the state of (x,t). For this to be true it suffices that the graph has finite maximum degree, i.e., for some M we have $\deg x \leq M < \infty$ for each $x \in V$;

the desired property then follows by comparison with a branching process in which births occur at rate λM.

Additivity of the process is immediate from this construction and from the fact that each transition is additive. Monotonicity with respect to parameters can be established in the usual way; for example, to compare processes with identical values of γ and δ and transmission rates $\lambda < \lambda'$ on the same graph, simply add a point process at rate $\lambda' - \lambda$ for the extra transmission events in the second process, and note that this tends to give larger configurations as the process evolves.

For each δ, there is a dual process ζ_t given by the rates

$$\begin{aligned} 0 &\to 1 \text{ at rate } \lambda n_2 \\ 1 &\to 2 \text{ at rate } \gamma \\ 2 &\to 1 \text{ at rate } \delta \\ 1,2 &\to 0 \text{ at rate } 1 \end{aligned}$$

and which Krone calls the "on-off" process because of the $2 \to 1$ transition. Note the dual is similar to the original process, in that type 0 represents a vacant state, and type 2's give birth to type 1's. Define the compatibility relation $\xi \sim \zeta \Leftrightarrow \xi(x) \sim \zeta(x)$ for some x, where $1, 2 \sim 2$ and $2 \sim 1$. Notice that type 2 in the dual process corresponds to type 1 or type 2 in the original process and that dual type 1 corresponds to original type 2.

The interpretation of compatibility is that the configuration ξ is strong enough to be compatible with ζ at some site, and the stronger the dual type, the easier it is to match up. We give a simple proof of the following fact, the proof of which occupies several pages in [29].

Proposition 4.2.1. *The dual process has the property that*

$$\xi_t \sim \zeta_0 \Leftrightarrow \zeta_t \sim \xi_0$$

with the dual running down the (same) spacetime graph from time t to time 0, so that ζ_s is on the time line $t - s$.

Proof. We start from the above condition to construct the dual, showing that it has the stated transitions and rates. The proof is given for the case $|V| < \infty$, that is, when the set of sites is finite, since only finitely many events occur in a finite time and we can proceed by induction on the events. To extend this to the case $|V| = \infty$ fix a finite subset V_0 and let $V_k = \{y \in V : \rho(y, V_0) \leq k\}$, where ρ is the graph distance. Denoting by ${}_{V_k}\xi_t$ the process constructed using the events on the subset $V_k \times \mathbb{R}^+$ of the spacetime graph, there is an almost surely finite value of k_0 so that ${}_{V_k}\xi_s(x) = \xi_s(x)$ for $x \in V_0$ and $0 \leq s \leq t$ when $k \geq k_0$, and this suffices to make the extension.

Say that a set of (forward) configurations Λ is *dualizable* if there is a dual configuration ζ so that

$$\Lambda = \{\xi : \xi \sim \zeta\}$$

Note that ζ is unique, if it exists. For fixed ζ_0 and $0 \leq s \leq t$ let $\Lambda_s = \{\xi_{t-s} : \xi_t \sim \zeta_0\}$. Clearly, Λ_0 is dualizable with dual configuration ζ_0. If Λ_s is dualizable, denote by ζ_s its dual configuration. Suppose there is an event at time s, and use the notation ξ_{t-s^+} and Λ_{s^+} to denote the state just prior to its occurrence (which corresponds to just after its occurrence in backward time, which is why we denote it s^+ rather than s^-). Note that

$$\Lambda_{s^+} = \{\xi_{t-s^+} : \xi_{t-s} \in \Lambda_s\}$$

Suppose that Λ_s is dualizable with dual configuration ζ_s, then $\Lambda_{s^+} = \{\xi_{t-s^+} : \xi_{t-s} \sim \zeta_s\}$. We show that Λ_{s^+} is dualizable by producing its dual configuration ζ_{s^+}. A type 2 death at x (i.e., a rate 1 death event) kills both active types, so $\zeta_{s^+}(x) = 0$ whatever the value of $\zeta_s(x)$; this causes the dual $1, 2 \to 0$ transition at rate 1. A type 1 death at x (i.e., a rate δ death event) kills only type 1. If $\zeta_s(x) = 2$ i.e., a 1 or a 2 is sufficient for compatibility after the event, then a 2 is required for compatibility before, so $\zeta_{s^+}(x) = 1$; this is the dual $2 \to 1$ transition at rate δ. A (forward) transmission event from $y \to x$ leads to a 1 at x after the event, if y is in state 2 just before the event, so $\zeta_{s^+}(y) = 1$ if $\zeta_s(x) = 2$; this is the dual transmission event. A maturation event at x causes a $1 \to 2$ transition, so that $\zeta_{s^+}(x) = 2$ if $\zeta_s(x) = 1$; this is the dual $1 \to 2$ transition at rate γ. For values of $\zeta_s(x)$ not mentioned, or for sites that aren't involved in the transition, it is easily verified that $\zeta_{s^+}(x) = \zeta_s(x)$. This finishes the induction step and establishes the dual transitions, completing the proof. □

Before moving on, we note that the dual process is also additive and monotone increasing in λ and γ, and monotone decreasing in δ, a fact which is noted in [29] and which we use later.

4.3 Main Results

4.3.1 Critical values for survival

Denoting by ξ_t^o the process starting a single mature site (the "o" stands for "origin", which if the process lives on the lattice, we can without loss of generality set to be the initially occupied site), we say ξ_t^o survives if

$$\mathbb{P}(\forall t > 0, \exists x : \xi_t^o(x) \neq 0) > 0$$

and dies out otherwise. Defining the critical value

$$\lambda_c(\gamma, \delta) = \inf\{\lambda > 0 : \xi_t^o \text{ survives } \}$$

it follows by monotonicity that λ_c is an increasing function of δ and a decreasing function of γ and that ξ_t^o survives if $\lambda > \lambda_c$. For each δ, by taking γ and λ large enough and comparing to a (suitably scaled in time) 1-dependent bond percolation diagram it is possible to show that ξ_t survives, which implies that $\lambda_c(\delta, \gamma) < \infty$ if γ is large enough. The first proof of this type is given by Harris for the contact process in [24]; its application to the two-stage process is noted in [29].

For fixed δ the parameter space for the process is the quadrant $\{(\lambda, \gamma) : \lambda \geq 0, \gamma \geq 0\}$, and by identifying the survival region $\mathcal{S} = \{(\lambda, \gamma) : \xi_t^o \text{ survives } \}$ we obtain a phase diagram for survival. We can define the critical lines

$$\lambda_*(\delta) = \inf\{\lambda : \xi_t^o \text{ survives for some } (\lambda, \gamma)\}$$
$$\gamma_*(\delta) = \inf\{\gamma : \xi_t^o \text{ survives for some } (\lambda, \gamma)\}$$

that bound the survival region below, and on the left. From monotonicity it follows that $\lambda_*(\delta) \geq \lambda_c(\infty, 0)$, the critical value for the contact process, and also that $\gamma_*(\delta) \geq \gamma_*(0)$, the left-hand critical line when $\delta = 0$. We shall have more to say about the survival region in Section 4.3.7. First, we complete a proof given in [29]

that characterizes λ_* for any value of δ.

4.3.2 Correction to Proposition 3.6

In Krone, Proposition 3.6 it is claimed that $\lambda_*(\delta) = \lambda_c(\infty, 0)$ for the process on $\mathbb{Z}^d$, for any dimension d. However, the proof given covers only the case $d = 1$. This is because the paper to which it refers gives a finite spacetime condition for survival only when $\lambda > \lambda_c^{(1)}$, the critical value for the contact process in one dimension. Here we use the more general construction of [3], plus a perturbation argument, to show that $\lambda_*(\delta) \leq \lambda_c(\infty, 0)$ in any dimension, which combined with the previous inequality implies the desired result.

In [3] it is shown for the contact process that if $\lambda > \lambda_c$ and $\epsilon > 0$, we can place a latticework structure over an effectively two-dimensional region in $\mathbb{Z}^d \times \mathbb{R}^+$ and make a $1:1$ correspondence between certain spacetime boxes contained in this structure and the set $\{(x, y) \in \mathbb{Z}^2 : y \geq 0, x + y \text{ is even}\}$ with the property that when the process starts with a large disc of active sites in the box corresponding to (x, y), then with probability $> 1 - \epsilon$ it can produce a large disc of active sites in the boxes corresponding to both $(x - 1, y + 1)$ and $(x + 1, y + 1)$. In their paper, they then show that if one decreases λ slightly, this property still holds, and using results for oriented percolation in two dimensions, conclude that the process still survives.

In our case it suffices to show that the property still holds when γ is decreased slightly from ∞, i.e., when γ is large enough. From this we may then conclude that if $\lambda > \lambda_c(\infty, 0)$ then $\lambda > \lambda_c(\gamma)$ for some γ, which implies that $\lambda > \lambda_*$, or $\lambda_c(\infty, 0) \geq \lambda_*$, and combining the inequalities, $\lambda_c(\infty, 0) = \lambda_*$.

It is sufficient to show that on a bounded spacetime region, when γ is large enough and the two processes are started from the same configuration (with mature sites in the place of active sites in the two-stage process), with high probability,

- between any two transmission events incident at a given site, there is a maturation event, and
- if at a fixed time the contact process has a certain set of active sites, then in the two-stage process those sites are all mature sites

The first condition ensures that no connections are cut due to a juvenile site being unable to give birth at a neighbouring site. The second condition ensures that if the contact process has produced a large disc of active sites, then the two-stage process has produced a large disc of mature sites.

To satisfy both conditions, it suffices to ensure that maturation events occur arbitrarily often, since on a finite spacetime region $B \subset \mathbb{Z}^d \times \mathbb{R}^+$, for each $\epsilon > 0$ there is a $\delta > 0$ so that with probability $> 1 - \epsilon$, the waiting time between transmission events is $\geq \delta$ everywhere on B. However, for each $\delta > 0$ and $\epsilon > 0$ there is a γ_0 so that if $\gamma > \gamma_0$, with probability $> 1 - \epsilon$ the waiting time between maturation events is $< \delta$ everywhere on B, thus for $\gamma > \gamma_0$ the conditions hold.

The two assertions of the last paragraph (those regarding waiting times) require proof, and it suffices to consider a spacetime region which is a single interval of length L. To prove the first assertion, notice that with high probability a finite number N of events occur in the interval, and with probability $e^{-\delta\lambda N}$ which $\to 1$ as $\delta \to 0$, each event takes time $\geq \delta$ to occur. To prove the second assertion, break up the interval into pieces of length δ, so that the number of events on each piece is distributed like a Poisson random variable with mean δ. The probability that on each interval at least one event has occurred is $(1 - e^{-\delta\gamma})^{L/\delta}$ which $\to 1$ as $\gamma \to \infty$, for fixed δ.

4.3.3 Critical maturation rate (q.6)

In [29], a lower bound on $\gamma_*(\delta)$ is given for the process on $\mathbb{Z}$ with nearest-neighbour interactions which is about $1/4$ when $\delta = 0$ and increases towards 1 as $\delta \to \infty$. Here we answer question 6 in that paper, which asks for lower bounds on $\gamma_*(\delta)$ in other settings; we obtain here a simple lower bound on $\gamma_*(0)$ (and by monotonicity, on $\gamma_*(\delta)$) that works for any graph of finite maximum degree, and depends only on the maximum degree. A graph has finite maximum degree if there is a number M so that $\deg x \leq M < \infty$ for each $x \in V$.

Proposition 4.3.1. *If $\delta = 0$ and $\gamma < 1/(2M - 1)$ the process dies out for any value of λ.*

Proof. It suffices to show this for $\lambda = \infty$, i.e., when the $0 \to 1$ transition at x is instantaneous if $n_2(x) > 0$. The result is obtained by estimating the average number

of offspring of a site x in state 1. The transition $1 \to 2$ occurs with probability $\gamma/(1+\gamma)$, since $1 \to 0$ at rate 1 and $1 \to 2$ at rate γ. If the $1 \to 2$ transition occurs at x, then each unoccupied neighbour of x becomes occupied. In order for x to send a second offspring to a neighbour y, the existing offspring at y has to die off. Denoting by N_t a Poisson process with rate 1 (representing the number of deaths at y, starting from the moment the $1 \to 2$ transition occurs at x) and by X_t an independent exponential random variable (representing death of the mature organism at x), the number of additional offspring produced at y is equal to N_{X_t}. Intuitively, we might expect $\mathbb{E}N_{X_t} = \mathbb{E}N_{\mathbb{E}X_t} = 1$, and computing, we confirm that

$$\begin{aligned}
\mathbb{E}N_{X_t} &= \int_0^\infty \sum_k k x^k \frac{e^{-x}}{k!} e^{-x} dx \\
&= \int_0^\infty \sum_k k \frac{x^k}{k!} e^{-2x} dx \\
&= \sum_k k \int_0^\infty \frac{x^k}{k!} e^{-2x} dx \\
&= \sum_k k 2^{-(k+1)} \\
&= 1
\end{aligned}$$

Thus the expected number of offspring at each initially unoccupied neighbouring site is $1 + \mathbb{E}N_{X_t} = 1 + 1 = 2$, so the expected number of offspring of a site in state 2 is at most $2M$. Since the probability of making the $1 \to 2$ transition before dying is $\gamma/(1+\gamma) = 1/(1+1/\gamma)$ the expected number of offspring of a site in state 1 is at most $2M/(1+1/\gamma)$. Setting this < 1 and comparing to a branching process gives the result. □

From Proposition 4.3.1 we conclude that $\gamma_*(0) \geq 1/(2M-1)$, so that $\gamma_*(\delta) \geq \gamma_*(0) \geq 1/(2M-1)$ for any δ, proving Theorem 4.1.1. For the nearest-neighbour process on $\mathbb{Z}^d$ we have $M = 2d$, giving $\gamma_*(\delta) \geq 1/(4d-1)$, which is $1/3$ for $d = 1$, $1/7$ for $d = 2$, etc.

4.3.4 Single-site survival and edge speed (q.2)

Let ξ_t^- denote the process starting from type 2's on $\mathbb{Z}^-$, and let $r_t = \sup\{x : \xi_t^-(x) \neq 0\}$ denote the right edge of ξ_t^-. A result of Durrett shows that r_t/t converges to

a constant α as $t \to \infty$. It is asked in [29] (question 2 in Section 4) whether $\lambda_c = \inf\{\lambda : \alpha(\gamma, \delta) > 0\}$. Here we do not prove this, but we give a sufficient condition for it to be true. To get a sense of what it means, note that this property is equivalent to the property that ξ_t is supercritical (i.e., $\lambda > \lambda_c$) if and only if the right edge of the process started from a half-line of mature sites has a positive spreading speed. For the equivalence of these statements, note that α is upper semi-continuous in λ, since it is the infimum of a family of continuous functions as described in [7].

One side of the result is easy; letting ξ_t^+ denote the process starting from type 2's on $\mathbb{Z}^+$ and ℓ_t its left edge, by attractiveness $\xi_t^o \leq \min \xi_t^+, \xi_t^-$, so $\xi_t^o(x) = 0$ for $x > r_t$ and for $x < l_t$. If $\alpha < 0$ then by symmetry $\ell_t/t \to -\alpha > 0$. Since $r_t \to -\infty$ and $l_t \to \infty$, eventually $r_t < l_t$ and $\xi_t^o(x) = 0$ for all x, i.e., ξ_t^o dies out.

For the converse, for $x \in \mathbb{Z}$ denote by C_x the "active cluster" of x, i.e., the set of spacetime points (y, t) such that if site x is initially in state 2, then site y is active at time t, and denote by $|C_x|$ its width, that is, $|C_x| = \sup\{|y - x| : (y, t) \in C_x \text{ for some } y, t\}$. On the event that the process does not die out, the number of active sites tends to ∞, so if ξ_t^o survives, $\mathbb{P}(|C_x| = \infty) > 0$ and so $\mathbb{E}|C_x| = \infty$ for each x, thus if $\mathbb{E}|C_x| < \infty$ then $\lambda \leq \lambda_c$. By analogy with percolation theory [23] we might guess that the converse holds, i.e., that if $\lambda < \lambda_c$ then $\mathbb{E}|C_x| < \infty$; this is proved, for example, for the contact process in [4]. We do not pursue this here, but instead show that if $\mathbb{E}|C_x| < \infty$ then $\alpha \leq 0$. Thus a sufficient condition for edge speed to characterize single-site survival is for the subcritical process to have a finite expected size.

Proposition 4.3.2. *If $\mathbb{E}|C_x| < \infty$ then $\alpha \leq 0$.*

Proof. If $\mathbb{E}|C_x| < \infty$ but $\alpha > 0$ then each C_x is bounded almost surely, but for each $\epsilon > 0$ eventually $r_t/t > \alpha - \epsilon$, which means with probability 1 there is an infinite sequence of sites (x_k) in $\mathbb{Z}^-$ and (y_k) in $\mathbb{Z}^+$ with $x_{k+1} < x_k$ for each k, and an infinite sequence of times (t_k) with $t_k \to \infty$ such that for each k, $(x_k, 0) \to (y_k, t_k)$. This is because the cluster of any finite collection of sites is almost surely bounded, which means that later activity of the process must originate from sites which are progressively further to the left; note that this implies also that the stated paths must be disjoint, although we will not need this here. In any case, the event $|C_x| \geq |x|$

occurs for infinitely many $x \in \mathbb{Z}^-$. However,

$$\sum_{x \in \mathbb{Z}^-} \mathbb{P}(|C_x| \geq |x|) = \mathbb{P}(|C_0 \geq 0|) + \mathbb{E}|C_x| < \infty$$

so applying the Borel-Cantelli lemma, $|C_x| \geq |x|$ occurs infinitely often with probability zero, contradicting our assumption. □

4.3.5 Equality of critical values (q.1)

For any attractive growth model there is at least one other characterization of survival aside from single-site survival, or divergence of the expected cluster size, which is the existence of a non-trivial upper invariant measure ν, obtained as the weak limit of the distribution of the process started from its largest initial configuration (in the case of ξ_t, when started from all sites in state 2). For either the two-stage contact process or the on-off process, this weak limit exists by attractiveness, and from the Feller property is an invariant measure for the system; see [32], Chapter, Theorem 2.3 on page 135 for a proof. The proof is for spin systems but generalizes without modification to any attractive system with a largest configuration.

It is possible that $\nu = \delta_0$, the measure that concentrates on the configuration with all 0's; we say that ν is non-trivial if $\nu \neq \delta_0$, equivalently, if ν assigns positive density at each site, that is, $\nu(\{\xi : \xi(x) \neq 0\}) > 0$ for each x. In [29] (question 1 in Section 4), it is asked whether single-site survival is equivalent to this property. First we show that single-site survival of either the two-stage contact process, or of the on-off process, implies that $\nu \neq \delta_0$, which supplies one direction of the proof. We then use the duality relation to observe that

$$\nu(\{\xi : \xi(o) \neq 0\}) = \mathbb{P}(\zeta_t^o \text{ survives })$$

where ξ_t is the two-stage contact process and ζ_t is the on-off process, and that the same property holds when ξ_t and ζ_t are exchanged in the formula. Thus if the two-stage contact process has a non-trivial stationary distribution, then the on-off process has single-site survival, which means that the on-off process has a non-trivial stationary distribution, which means that the two-stage contact process has single-site survival, which supplies the other direction of the proof, and shows that the two notions of

survival are in fact equivalent, proving Theorem 4.1.2. Therefore, it suffices to show that single-site survival of the two-stage process, or of the on-off process, implies that $\nu \neq \delta_0$.

For the (single-stage) contact process η_t on $\mathbb{Z}^d$, if $\lambda > \lambda_c$ then the method described in [3] allows us to conclude that under a suitable renormalization and started from a finite number of active sites, η_t dominates a supercritical 1-dependent oriented site percolation process in two dimensions, for which it is known that the origin is active for a positive fraction of the time, and from which it follows that $\liminf_t \mathbb{P}(\eta_t^o(0) \neq 0) > 0$, which since $\eta_0^o \leq \eta_0^1$ and by attractiveness implies that $\nu(\{\eta : \eta(0) \neq 0\}) = \lim_{t\to\infty} \mathbb{P}(\eta_t^1(0) \neq 0) > 0$, where η_t^1 is the contact process started from all sites active. The following lemma allows to conclude the same fact for the two-stage contact process, whenever the interaction neighbourhood is symmetric about permutation and sign change of coordinates, and such that with some probability, any site can infect any other site; the first condition we call coordinate symmetry, and the second we call irreducibility. Note the interaction neighbourhood must of course be finite.

Lemma 4.3.3. *The construction in [3] is valid for the two-stage contact process and for the on-off process on $\mathbb{Z}^d$, for any coordinate-symmetric and irreducible interaction neighbourhood.*

Proof. The basic strategy is to repeat the construction, replacing "infected site" with "active site". Most of the steps go through without modification, and the required changes are addressed below.

In the construction in [3], nearest-neighbour interactions are assumed. This condition can be relaxed by redefining the "sides" of the box to be a region whose width is equal to the interaction range of the process. In this way, we can control transmission from the sides of the rectangle to the outside world as is done in the nearest-neighbour case; see [2] for an example where this modification is carried out.

When widening the sides, it is necessary to make sure that a large finite disc can be produced at an extra distance corresponding to the range of the interaction, but this can be prescribed. Irreducibility is required to ensure that, starting from a single infectious site, all sites in a large finite disc can be made infectious with a certain probability.

Coordinate symmetry is implicit in the construction, and it is assumed when proving that the process reaches each orthant of the top and sides of the box with high probability.

When defining the event where a single site produces a large finite disc, it doesn't matter too much which type is produced in the disc; arbitrarily, we pick mature sites. However, we should require that a juvenile site can produce the disc, since whatever a juvenile site can produce, a mature site can produce.

The construction uses the property that the process dies out if its population dips below a certain value infinitely often. This is a property that holds for any growth model in which there is a finite number of active types, and each active type dies before reproducing with a certain probability.

When arguing along the sides of the box, it is also important that active sites die at a certain uniform rate; for example, the argument would not carry over directly if juvenile sites had to mature before being subject to death events. Note that the argument in [3] along the sides of the box is somewhat convoluted; for a clearer proof of this part see [34], Part 1, Proposition 2.8.

For the (usual) contact process, the application of the FKG inequality given in the construction follows from the invariance of positive correlations (see [32], Theorem 2.14 on page 80), for which it is sufficient that the process be attractive and that its transitions occur only between comparable states, a property which is true of the two-stage contact process and of the on-off process.

We have addressed the changes needed to adapt the construction of [3] to the two-stage and on-off processes. The rest of the construction consists of geometrical or "resetting" arguments, and no modification is required. □

It follows from the Lemma and from the discussion preceding it that for the two-stage contact process or the on-off process on $\mathbb{Z}^d$, whenever there is single-site survival ($\lambda > \lambda_c$) the upper invariant measure assigns a positive density at each site, i.e., $\nu(\{\xi : \xi(x) \neq 0\}) > 0$ for each x. The proof is now complete.

4.3.6 Complete convergence (q.3)

Let δ_0 be the measure that concentrates on the configuration with all 0's, and let ν be the upper invariant measure introduced in the previous section. For an attractive growth model, complete convergence means that

$$\xi_t \Rightarrow \alpha\delta_0 + (1-\alpha)\nu$$

as $t \to \infty$, where $\Rightarrow$ denotes weak convergence and $\alpha = \mathbb{P}(\xi_t \text{ dies out })$. In [29] it is asked whether complete convergence holds for the two-stage contact process, when $\lambda > \lambda_c$. We follow [16], Section 5, where the argument is used for the contact process; the idea is originally due to Griffeath [21]. Fix an arbitrary configuration ξ_0, and a dual configuration ζ_0 with finitely many active sites; doing this for all such ζ_0 we will recover the finite-dimensional distributions of the upper invariant measure. We have that

$$\xi_{2t} \sim \zeta_0 \Leftrightarrow \xi_t \sim \zeta_t$$

where ζ_s, $0 \le s \le t$ is constructed on the same spacetime graph as ξ_t and run from time $2t$ down to time t, with initial configuration ζ_0. Use the notation $\xi \neq 0$ to denote "not identically zero". Then observe that

$$\mathbb{P}(\xi_t \sim \zeta_t) = \mathbb{P}(\xi_t \neq 0, \zeta_t \neq 0) - \mathbb{P}(\xi_t \neq 0, \zeta_t \neq 0, \xi_t \nsim \zeta_t)$$

Since they are built over disjoint parts of the graph, ξ_s, $0 \le s \le t$ and ζ_s, $0 \le s \le t$ are independent, so

$$\mathbb{P}(\xi_t \neq 0, \zeta_t \neq 0) = \mathbb{P}(\xi_t \neq 0)\mathbb{P}(\zeta_t \neq 0)$$

for each $t > 0$. Using the duality relation, $\mathbb{P}(\zeta_t \neq 0) = \mathbb{P}(\xi_t \sim \zeta_0)$ with ξ_0 in this case being the configuration with all 2's. Letting $t \to \infty$

$$\mathbb{P}(\xi_t \neq 0)\mathbb{P}(\zeta_t \neq 0) \to (1-\alpha)\nu(\{\xi : \xi \sim \zeta_0\})$$

To have complete convergence, it therefore suffices to show that

$$\mathbb{P}(\xi_t \neq 0, \zeta_t \neq 0, \xi_t \nsim \zeta_t) \to 0$$

as $t \to \infty$. A method for doing this is outlined in [16] for a certain class of growth models. They use a restart argument to show that whenever the process survives, and suitably rescaled, it eventually dominates a two-dimensional oriented percolation process, which is known to have a positive density of sites. Using this fact it is then argued that if run for long enough, the process and its dual intersect with high probability. In Section 5 of [3] it is noted that, using their construction and the ideas from [16], the same can be concluded for the contact process in $\mathbb{Z}^d$. Noting the equality of critical values proved in the previous section, if $\lambda > \lambda_c$ the construction of [3] can be applied to both the two-stage contact process and the on-off process, and the same arguments apply to show that the two processes eventually intersect with high probability, which proves Theorem 4.1.3.

4.3.7 Structure of the survival region (q.5 and q.4)

Continuing the analysis of the survival region $\mathcal{S}$ begun in Section 4.3.1, we show that the process dies out on the boundary ∂S. By monotonicity $\lambda_c(\gamma)$ can have only jump discontinuities, which means that the boundary of the survival region is the set

$$\{(\gamma, \lambda) : \gamma \geq \gamma_*, \lambda_c^-(\gamma) \geq \lambda \geq \lambda_c^+(\gamma)\}$$

where $\lambda_c^-(\gamma)$ and $\lambda_c^+(\gamma)$ are the left-hand and right-hand limits of λ_c at γ; set $\lambda_c^-(\gamma_*) = \infty$.

It follows from Lemma 4.3.3 that survival of the two-stage contact process is given by a finite spacetime condition of the form "a certain event happens with probability $> 1 - \epsilon$", where ϵ is sufficiently small. Moreover, the probability of this event is continuous in λ and γ (also δ, but we will not use this fact here). This is because by a small enough change in parameters, on a finite spacetime region we can ensure that the probability of even one more or one fewer transmission/maturation events can be made arbitrarily small. This implies that $\mathcal{S}$ is an open subset of the plane, in any dimension and for any value of δ. Since $\mathcal{S}$ is open it follows that the process dies

out on its boundary ∂S, which includes the critical values $\lambda_c(\gamma)$.

It seems that question 4, namely, whether there is a formula for λ_c in terms of $\lambda_c(\infty, 0)$, γ and δ, should be false. One good reason to believe this is that for the contact process, a sequence of approximants is known that converges to the critical value, and these are roots of successively more complicated rational functions, as shown in [51]. There is no obvious reason to believe why the introduction of an additional stage to the process should lead to a critical value which is any simpler to determine, even if the critical value of the contact process is used in the expression.

Remaining questions for the survival region include whether $\lambda_c(\gamma)$ is continuous, whether it is strictly decreasing on $\{\gamma > \gamma_*\}$ and whether

$$\lim_{\gamma \to \gamma_*^+} \lambda_c(\gamma) = \infty$$

We believe the answers are respectively yes, yes, and yes, but we are not sure how to prove this.

Chapter 5

Duality and Complete Convergence for Multi-Type Additive Growth Models

This chapter consists of the journal article [17] of the same title, to appear in Advances in Applied Probability 48.1 (March 2016). Moreover, it includes Sections 5.5 and 5.6 that do not appear in the above article, yet provide some additional context and connections with other models.

Abstract

We consider a class of multi-type particle systems having similar structure to the contact process and show that additivity is equivalent to the existence of a dual process, extending a result of Harris. We prove a necessary and sufficient condition for the model to preserve positive correlations. We then show that complete convergence on $\mathbb{Z}^d$ holds for a large subclass of models including the two-stage contact process and a household model, and give examples.

5.1 Introduction

The contact process [32],[34] is a model of infection spread in which each individual is in one of two states, either healthy or infectious. Specifically, given an undirected graph $G = (S, E)$ where S stands for "sites", we define the contact process $(\eta_t)_{t\geq 0}$ on

G with rate $\lambda > 0$ to have the state space $\{0,1\}^S$ and transitions

- $0 \to 1$ at site $x \in S$ at rate $\lambda n(x)$ (transmission) and
- $1 \to 0$ at site $x \in S$ at rate 1 (recovery)

where $n(x)$ is the cardinality of the set $\{y \in S : xy \in E, \eta_t(y) = 1\}$. A graphical method for constructing the process, originally due to [26], is described in Section 5.2. This construction allows us to define multiple copies of the process, with different initial data, on the same probability space. With respect to this construction, the model has a number of convenient properties including

- *additivity*: define $\eta \vee \eta'$ by $(\eta \vee \eta')(x) = \max(\eta(x), \eta'(x))$, for $x \in S$. Suppose η_0 and η_0' are given and let $\eta_0'' = \eta_0 \vee \eta_0'$. Denote by $\eta_t, \eta_t', \eta_t''$ the processes with respective initial data η_0, η_0' and η_0''. Then, additivity means that the equality $\eta_t'' = \eta_t' \vee \eta_t$ holds for all $t > 0$.
- The existence of a *dual* process: the contact process has a dual process $(\zeta_t)_{t \geq 0}$ with state space $\{0,1\}^S$, with the following property. For fixed $t > 0$, $(\zeta_s)_{0 \leq s \leq t}$ can be constructed using the graphical construction for the contact process starting at time t and going down to time 0. Moreover, when defined in this way, $\eta_t(x) = \zeta_0(x) = 1$ for some x if and only if for every $s \in [0,t)$, there is an x for which $\eta_s(x) = \zeta_{t-s}(x) = 1$. Note that for the contact process, the dual behaves in the same way as the contact process i.e. it is *self-dual*.
- Preservation of *positive correlations*: define the partial order $\eta \leq \eta'$ by $\eta(x) \leq \eta'(x)$ for all $x \in S$, and an *increasing function* $f : \{0,1\}^S \to \mathbb{R}$ to be one with $\eta \leq \eta' \Rightarrow f(\eta) \leq f(\eta')$, and say that a distribution μ on $\{0,1\}^S$ has *positive correlations* if $\mathbb{E}_\mu fg \geq \mathbb{E}_\mu f \mathbb{E}_\mu g$ for all increasing functions f, g. Then letting $\mu_t, t \geq 0$ denote the distribution of η_t, if μ_0 has positive correlations then μ_t has positive correlations for $t > 0$.

Note that as stated, the first two properties hold pointwise on realizations of the process, when it is constructed graphically. These properties are all described in [26], and as shown in [34], they suffice to prove *complete convergence* (defined precisely in Section 5.2) for the contact process on the lattices $\mathbb{Z}^d$, which characterizes not only the set of extremal invariant measures as a function of the infection parameter but also the convergence to a combination of these measures, from any initial configuration. Since the proof relies only on the three properties mentioned above and not, for

example, on the number of infectious types, it is natural to suppose the same results hold for a much wider class of models.

To this end, we consider a class of processes that we call *multi-type particle systems*, with a finite set of types $F = \{0, 1, ..., N\}$, so that the state space of configurations is F^S. The local dynamics is given by a family of transition rates

$$\{c_T(\phi, \psi) : \phi, \psi \in F^T, T \subseteq S \text{ finite } \}$$

that give the rate at which $\eta|_T$, the restriction of η to T, flips to ψ when its value is ϕ. These types of systems are well-studied, particularly those in which $F = \{0, 1\}$; the interested reader may consult [32] for sufficient conditions on the rates so that the process is uniquely defined. In Section 5.2.2 we give a simple and more easily proved sufficient condition for this to hold, and a graphical construction that shows existence and uniqueness.

For the next definition, let $\underline{0}$ denote the configuration η with $\eta(x) = 0$ for all x. Our focus is on *growth models*, where F is a partially ordered set (generally not with the usual order on $\{0, 1, ..., N\}$) with a distinguished *passive type* 0 and *active types* $\{1, ..., N\}$ satisfying

- $0 \leq a$ for each $a \in F$,
- $\underline{0}$ is absorbing, and
- $\underline{0}$ is reachable with positive probability from any configuration with only finitely many active sites.

We can generalize the above three properties to growth models as follows:

- to define additivity for a multi-type growth model, we define a *join* operation $\vee$ to be a symmetric binary operation defined for all pairs $a, b \in F$, that satisfies $a \vee b \in F$, $a \vee b \geq a, b$ and $c \geq a, b \Rightarrow c \geq a \vee b$. Note the $\leq$ and $\vee$ extend from F to F^S in the obvious way: $\eta \leq \eta' \Leftrightarrow \eta(x) \leq \eta'(x)\ \forall x$ and $(\eta \vee \eta')(x) = \eta(x) \vee \eta'(x)$ for all x. If it has a join operation, the process is said to be *additive* if with respect to an appropriate graphical construction called an *event coupling*, it satisfies the property described above i.e., $\eta''_t = \eta_t \vee \eta'_t$ for $t > 0$ whenever $\eta''_0 = \eta_0 \vee \eta'_0$. A precise definition is given in Definition 5.2.6.

- Duality is defined in a similar way. In this case we want a process $(\zeta_t)_{t\geq 0}$ with state space $\tilde{F}^S$ where $\tilde{F}$ is a certain collection of subsets of F, and the duality relation is that $\eta_t(x) \in \zeta_0(x)$ for some x if and only if for every $s \in [0,t)$, there is an x for which $\eta_s(x) \in \zeta_{t-s}(x)$, where $(\zeta_s)_{0\leq s\leq t}$ is again constructed graphically from time t down to time 0. A precise definition is given in Definition 5.3.4.

- Preservation of positive correlations is defined in the same way as above, since it depends only on the existence of a partial order.

In general there is more than one way to define a dual process. The definition we will use here gives a dual process valued on F^S for some finite set F and is well suited to models of growth and reproduction of an organism. More general dual processes are possible (see for example [40]) in which the state space consists of collections of finite subsets of S, or the duality relation may differ (see for example [32] for a more general definition of duality). The existence of a dual process is a key component in the proof of complete convergence, as first described in [21].

The term growth model is borrowed from [16], however here we work in continuous rather than discrete time, and with systems having generally more than one type of active particle. Examples of additive growth models are the two-stage contact process [29], [19] the household model studied in [48] and a spatial analogue of any multi-type branching process (see [35] for an introduction to these processes). On the other hand, the multitype contact process from [39] is not an additive growth model, since it is not additive.

The paper is organized as follows. In Section 5.2 we introduce the graphical method and describe the set of couplings that it furnishes, exploring some basic properties of these couplings and giving a precise definition of additivity. In addition, we state the main results of the paper. In Section 5.3 we define the dual process and prove Theorem 5.2.7. In Section 5.4 we define multi-colour systems and prove Theorems 5.2.8 and 5.2.9. In Section 5.7 we discuss positive correlations. In Section 5.8 we prove Theorem 5.2.10, and in Section 5.9 we conclude with some examples.

5.2 Graphical method and main results

We begin by introducing a graphical method with the intention of describing a large but manageable class of couplings for multi-type particle systems. The usual setting for particle systems is an undirected graph $V = (S, E)$ where S is a countable set. Generally the edge set determines which sites can interact, but in the construction below, the edge set is not used explicitly. Graphical methods have been extensively applied to particle systems; see [26] for one of the first papers on the subject.

Definition 5.2.1. *A* local mapping *is a function $e : F^T \to F^T$ for some finite $T \subseteq S$. An* event structure *is a family of triples $(T_i, e_i, r_i)_{i\in I}$ where I is some index set, each $e_i : F^{T_i} \to F^{T_i}$ for the corresponding T_i, and $r_i > 0$ is the associated rate.*

Our goal is to turn an event structure into a device for constructing a process $(\eta_t)_{t\geq 0}$, given an arbitrary $\eta_0 \in F^S$. Since η_t is constructed from a collection of space-time events, we call this the *event construction*. First define a family of independent Poisson point processes $\{u_i\}$ such that each u_i has intensity r_i; we can think of each u_i as an alarm clock that rings at random times.

If $|S| < \infty$ (i.e., the set of sites is finite) and for each $T \subseteq S$ at most finitely many mappings e_i send $F^T \to F^T$, then for each $t > 0$, only finitely many ringing events occur in the interval $(0, t]$. Thus, given the initial configuration η_0 we can determine η_t for $t > 0$ by accounting for each event. If u_i rings at time s, use η_{s^-} to denote the state just prior to the event. Letting $\phi = \eta_{s^-}|_T$ be the restriction of η_{s^-} to T, we let

$$\eta_s|_T = e_i(\phi)$$

and $\eta_s(x) = \eta_{s^-}(x)$ for $x \notin T$. In general, to determine the state at each spacetime point $(x, t) \in \mathcal{S}$, it is sufficient that with probability 1, only finitely many ringing events occur within the set of spacetime points (y, s) that can affect the state at (x, t), as follows.

For $s < t$, say there is an *influence path* from (y, s) to (x, t) if there is a finite sequence of sites $y = x_0, x_1, ..., x_k = x$ and a sequence of times $s = s_0 < s_1 < ... < s_k = t$ so that for $j = 1, ..., k$, $\{x_{j-1}, x_j\} \subset T_i$ for some $i \in I$ such that u_i rings at time s_j. Define the *influence set* $S_{x,t}$ to be the set of points (y, s), $s < t$ such that there is

an influence path from (y, s) to (x, t). Roughly speaking, $S_{x,t}$ is the set of points in $\mathcal{S}$ whose state can affect the state at (x, t). In order to determine the state at (x, t) from the state at time 0, it is sufficient that the set of times $s < t$ such that some u_i rings, for some $(y, s) \in \mathcal{S}_{x,t}$ such that $y \in T_i$, is a finite set. This is because then, we can proceed one event at a time to determine the state at (x, t).

- the set of points $\mathcal{S}_{x,t} \subset \mathcal{S}$ that can affect the state at (x, t) is a bounded set, and that
- the total rate of Poisson processes affecting points in $\mathcal{S}_{x,t}$ is bounded

since then, only finitely many ringing events occur on $\mathcal{S}_{x,t}$ and we can proceed as before to determine the state at (x, t). The following condition is sufficient for this to hold.

Lemma 5.2.2. *Given an event structure $(T_i, e_i, r_i)_{i \in I}$, if there is an M and an L so that*

- *for each $x \in S$, $\sum_{i \in I : x \in T_i} |T_i| \leq M$, and*
- *each $r_i \leq L$*

then for each spacetime point $(x, t) \in \mathcal{S}$, with probability 1 at most finitely many ringing events occur within the influence set $\mathcal{S}_{x,t}$.

Proof. Starting at (x, t) and looking backwards in time, the cardinality of the set $\{y \in S : (y, t - s) \in \mathcal{S}_{x,t}\}$ is dominated by a branching process Z_s with $Z_0 = 1$ in which each individual gives birth to M individuals at rate LM, so the cardinality of the set $U_{x,t} = \{y \in S : (y, s) \in \mathcal{S}_{x,t} \text{ for some } s < t\}$ is dominated by $\sum_{i=0}^{n} Z_{s_i}$ where $0 = s_0 < s_1 < s_2 < ... < s_n < t < s_{n+1}$ are the jump times of Z_s. It is easy to show this last sum is finite with probability 1. To obtain the second conclusion, note that $\mathcal{S}_{x,t} \subset U_{x,t} \times [0, t]$, and that for each $y \in S$, the intensity $\sum_{i \in I : y \in T_i} r_i$ of Poisson processes that may influence y is by assumption at most LM.

□

Remark 5.2.3. *These are fairly mild conditions, since we assume only that the change of the state at each site depend on the state of at most M other sites, and that*

each type of event occurs at rate at most L. Note, however, that there are well-behaved systems not satisfying the first condition; see for example the nearest particle systems of Chapter 7 in [32].

To construct a multi-type particle system from its transition rates, one way to do it is for each T, ϕ, ψ such that $c_T(\phi, \psi) > 0$, define the mapping e_i that sends ϕ to ψ while leaving other local configurations unchanged, and the rate $r_i = c_T(\phi, \psi)$. Letting $c_T = \sum_{\phi,\psi} c_T(\phi, \psi)$ for each $T \subset S$, to satisfy the conditions of Lemma 5.2.2 it suffices, for example, to have an M so that for each $x \in S$, $\sum_{T \subset S: x \in T} c_T \leq M$. We call this the *independent* event construction, since an independent point process is assigned to each transition. There may be other constructions, which we now describe.

Recall that a coupling is a probability space on which two or more processes are defined. In the event construction, copies of the process starting from any set of initial configurations are automatically coupled, since they are determined from the same set of Poisson processes.

For a multi-type particle system, the independent event construction leads to a certain coupling of realizations. Using the same point processes to determine some transitions, we can obtain other couplings.

Definition 5.2.4. *An* event coupling *for a multi-type particle system is an event structure $(T_i, e_i, r_i)_{i \in I}$ such that*

- *corresponding to each transition $c_T(\phi, \psi)$ is a subcollection I_c of indices i such that*
  1. *$T_i \supseteq T$,*
  2. *$e_i(\phi')|_T = \psi$ and $e_i(\phi')|_{T_i - T} = \phi'|_{T_i - T}$ whenever $\phi'|_T = \phi$, and*
  3. *$c_T(\phi, \psi) = \sum_{i \in I_c} r_i$*
- *aside from this, $e_i(\phi) = \phi$, and*
- *if e_i is assigned to the transitions $c_{T_j}(\phi_j, \psi_j)$, $j = 1, 2, ...$ then*

$$\{\eta : \eta|_{T_j} = \phi_j\} \cap \{\eta : \eta|_{T_k} = \phi_k\} = \emptyset, \; j \neq k \tag{5.2.1}$$

In the context of an event coupling the e_i are called *transition mappings*. The first condition is so that the correct transitions occur, and at the correct rate. The second condition is so that nothing else happens. The third condition is to ensure that we respect the independence of distinct transitions $c_T(\phi, \psi)$.

We note that from any valid event coupling we obtain a coupling of any number of copies of the process starting from any collection of initial configurations. In certain cases, some of these couplings may be more useful than others, for example if the resulting evolution of the system respects a certain natural partial order on the states, as we now describe.

Definition 5.2.5. *A mapping $e : F^T \to F^T$ is* attractive *if $\phi \leq \phi' \Rightarrow e(\phi) \leq e(\phi')$. An event coupling is attractive if each of its mappings is attractive. A process is attractive if it has an attractive event coupling.*

Note that attractiveness of a process is equivalent to the existence of an event coupling so that $\eta_0 \leq \eta'_0 \Rightarrow \eta_t \leq \eta'_t$ for all $t > 0$.

The historical reason for the word "attractive" is given in [32], Chapter 2, Section 2; in that context, the "spins" at adjacent sites tend to align, which means that the spin state at one site tends to be attracted to the spin state at adjacent sites.

Definition 5.2.6. *A mapping $e : F^T \to F^T$ is* additive *if $e(\phi \vee \phi') = e(\phi) \vee e(\phi')$. An event coupling is additive if each of its mappings is additive, and a process is additive if it has an additive event coupling.*

Note that additivity of a process is equivalent to the existence of an event coupling so that $\eta_t = \eta'_t \vee \eta''_t$ whenever $\eta_0 = \eta'_0 \vee \eta''_0$.

If a and b are *comparable* i.e., $a \leq b$ or $b \leq a$, then $a \vee b$ is just $\max(a, b)$. From this it follows that if $\eta_0 \leq \eta'_0$, then $\eta_0 \vee \eta'_0 = \eta'_0$ and so $\eta_t \vee \eta'_t = \eta'_t$ and in particular, $\eta_t \leq \eta'_t$. Thus additivity implies attractiveness.

We now summarize the main results, which are proved subject to the conditions of Lemma 5.2.2. Note a growth model is defined in the Introduction, additivity in Definition 5.2.6, and a dual is defined in Definition 5.3.4.

Theorem 5.2.7. *Every additive growth model with an additive event coupling that satisfies the conditions of Lemma 5.2.2 has a dual which is an additive growth model.*

The above result was proved in [26] for $F = \{0, 1\}$, using graphical methods, and also in [16] for their definition of additive growth model, which is a generalization of discrete-time oriented percolation in two dimensions. As mentioned above, the contact process is self-dual, and in Example 5.3.6 we show that for each N there is an N-stage contact process which is also self-dual.

In [26] it is shown that for $F = \{0, 1\}$ the converse is true, that is, the existence of a dual implies additivity. In order to show this we first need to define a slightly nicer type of growth model.

For an additive growth model we define the set of *primitive* types to be those $a \in F$ such that $a \neq b \vee c$ for any $b, c \in F$ with $b \neq a$, $c \neq a$. If we imagine primitive types a, b as individual organisms, we can think of $a \vee b$ as a state in which both an a and a b organism live together at a single site. We can then expand the set of types F so that it distinguishes all combinations of primitive types; a growth model that satisfies this condition is called *multi-colour*.

Our next result shows that any additive growth model can be "lifted" to an additive multi-colour growth model, showing that being multi-colour is a natural condition. In the statement $F_* \supseteq F$, the understanding is that for each $b \in F$, write b in exactly one way as $a_1 \vee ... \vee a_k$ for some primitive types $a_1, ..., a_k$, then the containment holds under the identification $a_1 \vee .... \vee a_k \in F$ with $\{a_1, ..., a_k\} \in F_*$.

Theorem 5.2.8. *Every additive growth model η with types F and an additive event coupling that satisfies the conditions of Lemma 5.2.2 has a* lift *ξ with types $F_* \supseteq F$ which is an additive multi-colour growth model, and is such that*

$$\pi(\xi_0) = \eta_0 \Rightarrow \pi(\xi_t) = \eta_t \text{ for } t > 0$$

where $\pi : F_*^S \to F^S$ is given by $\pi(\xi)(x) = \pi(\xi(x))$ for some $\pi : F_* \to F$ that preserves the join.

Since as constructed, ξ has the same event coupling as η, if η's event coupling satisfies the conditions of Lemma 5.2.2, then so does ξ's. If we focus on multi-colour

growth models, we can show the following. Because of Theorem 5.2.7, we only need to show that existence of a dual implies additivity.

Theorem 5.2.9. *For a multi-colour growth model with an event coupling that satisfies the conditions of Lemma 5.2.2, additivity is equivalent to the existence of a dual process.*

For an additive multi-colour growth model we can give a precise characterization of the property of preserving positive correlations. Since it requires some extra definitions, the statement is deferred to Section 5.7 and is given as Theorem 5.7.2.

By taking the join of all configurations we see that an additive model has a largest configuration. If with respect to the graphical construction, $\eta_0 \geq \eta'_0$ implies $\eta_t \geq \eta'_t$ for $t > 0$ we say the model is *attractive*. It follows from the definition of $\vee$ that $\eta \geq \eta' \Leftrightarrow \eta = \eta \vee \eta'$, so an additive model is attractive. For an attractive model with a largest configuration, the distribution of the process η_t started from that configuration is decreasing in time, since for $s < t$, $\eta_0 \geq \eta_{t-s}$ and evolving by s forward in time, by attractiveness $\eta_s \geq \eta_t$. By first examining cylinder sets of configurations whose probabilities decrease in time, and then using these to determine finite-dimensional distributions, we can show the distribution of the process converges to an *upper invariant measure* ν; details of this are given in [32]. We say the model exhibits *complete convergence* if for any initial configuration η_0 we have

$$\eta_t \Rightarrow (1 - \sigma(\eta_0))\delta_0 + \sigma(\eta_0)\nu$$

where δ_0 is the measure that concentrates on the all zero configuration and $\sigma(\eta_0) = \mathbb{P}(\eta_t \neq \underline{0}\ \forall t > 0\}$ is the probability of survival starting from η_0, and $\Rightarrow$ denotes weak convergence of measures.

The following result is proved in Section 5.8, and uses the construction of [3]. On $\mathbb{Z}^d$" just means that the set of sites is the d-dimensional integer lattice $\mathbb{Z}^d$, where $d \geq 1$ is any integer. Irreducible means that any active type at one site can produce any active type at any other site via some sequence of transitions. Translation invariant and symmetric means that transitions are invariant under translation and under reflection in each coordinate. Finite range means that there is an $M > 0$ so that $c_T(\phi, \psi) = 0$ if $\operatorname{Diam} T > M$, where $\operatorname{Diam} T = \max_{x,y \in T} \mathrm{d}(x, y)$, and $\mathrm{d}(x, y)$ is the graph distance. For an additive model, a productive interaction is a transition

mapping such that $e_i(\phi) \geq \phi$ for all $\phi \in F^{T_i}$, and a destructive interaction is one in which $e_i(\phi) \leq \phi$; an equivalent definition is given in Definition 5.7.1, just prior to the statement of Theorem 5.7.2.

Theorem 5.2.10. *An additive multi-colour growth model on $\mathbb{Z}^d$ that is irreducible, has finite range, is translation invariant, symmetric and has an additive event coupling that has only productive and destructive interactions exhibits complete convergence. Moreover, single-site survival is equivalent to having $\nu \neq \delta_0$.*

It is easy to check that any additive event coupling which is both translation invariant and has finite range satisfies the conditions of Lemma 2.1, which means the result of Theorem 5.2.7 applies to the models considered in the statement of Theorem 5.2.10. Since the proof of this theorem in [3] is lengthy, and since the generalization to multi-colour growth models is straightforward, we only give a sketch of the proof in which we indicate how to make the necessary modifications.

5.3 Growth models and duality

In order to say more about additivity, we now focus on growth models, defined in the introduction, which are natural choices for either population growth or disease spread. If we think of the active types as representing organisms in various stages of development this means that vacancy is the "lowest" type, that no spontaneous birth occurs, and that a finite population has a chance of dying out. For a population model, unoccupied territory is the passive type, and for a disease model in which each site represents a particular individual, the passive type might be a healthy individual.

For a growth model we are interested in questions of *survival*, that is, whether the process started from a finite number of active sites has a chance of avoiding the $\underline{0}$ state for all time.

We will show that each additive growth model has an additive counterpart going backwards in time; the first step is to define its configurations.

Definition 5.3.1. *For a growth model with types F and join $\vee$, a set of active types $E \subseteq F$ is*

- increasing *if $a \in E, a \leq b \Rightarrow b \in E$, and is*

- decomposable *if* $a \vee b \in E$ *implies* $a \in E$ *or* $b \in E$.

The dual types $\tilde{F}$ *are the set of increasing and decomposable sets of active types* E*, together with a passive type* 0*, with the partial order* $0 \leq E$ *for each* $E \in \tilde{F}$ *and* $E \leq E'$ *if* $E \subseteq E'$.

A partially ordered set of dual configurations $\tilde{F}^S$ is defined in the same way as for F^S. From the following proposition we see that the dual types come equipped with a join.

Proposition 5.3.2. *The set of dual types has a join* $\vee$ *given by* $E \vee E' = E \cup E'$.

Proof. If E and E' are increasing and $a \in E \cup E'$ then $b \geq a$ implies $b \in E \cup E'$, so $E \cup E'$ is increasing, and that if E and E' are decomposable and $b \vee c \in E \cup E'$ then $b \in E \cup E'$ or $c \in E \cup E'$, so $E \cup E'$ is decomposable; therefore, $E \cup E' \in \tilde{F}$ whenever $E, E' \in \tilde{F}$. If $E'' \geq E, E'$ then $E'' \supseteq E, E'$ so $E'' \supseteq E \cup E'$, moreover $E \cup E' \geq E, E'$. □

In order to relate configurations in F^S with dual configurations in $\tilde{F}^S$ we introduce the following compatibility relation.

Definition 5.3.3. *A configuration* $\eta \in F^S$ *is* compatible *with a dual configuration* $\zeta \in \tilde{F}^S$*, denoted* $\eta \sim \zeta$*, if for some* x*,* $\eta(x)$ *is an active type and* $\eta(x) \in \zeta(x)$*. For local configurations* $\phi \in F^T, \theta \in \tilde{F}^T$ *this reads*

$$\phi \sim \theta \Leftrightarrow \phi(x) \in \theta(x) \neq 0 \text{ for some } x \in T$$

When the $\neq 0$ is understood it is usually omitted. With the compatibility relation comes the natural identification $\zeta \leftrightarrow \{\eta : \eta \sim \zeta\}$ (or, for local configurations, $\theta \leftrightarrow \{\phi : \phi \sim \theta\}$) of dual configurations with the set of configurations with which they are compatible.

Definition 5.3.4. *A growth model with configurations in* F^S *has a* dual *if there is an event coupling for the model and a multi-type particle system with configurations* $\tilde{F}^S$ *so that in the event construction for the process, the following* duality relation *holds:*

$$\eta_t \sim \zeta_0 \Leftrightarrow \eta_0 \sim \zeta_t$$

where the process η_s, $0 \le s \le t$ is run forward in time from 0 to t, and the dual ζ_s, $0 \le s \le t$, is run backward in time from t to 0 (so that ζ_s is on the timeline $t - s$), using the dual mappings given by

$$\tilde{e}(\theta) = \{\phi \in F^T : e(\phi) \sim \theta\}$$

Note the identification $\theta \leftrightarrow \{\phi : \phi \sim \theta\}$ is assumed on the right-hand side of the equation.

Remark 5.3.5. *Duality for particle systems (the use of relations pairing particle systems with other systems going backwards in time) comes in more than one form; the form described here is known as coalescing duality in Chapter III of [32]. An application of this form of duality can be found in [16], where it is used to help prove complete convergence of the growth models described in that paper. Other forms of duality exist; see for example the paper [47] in which a form of duality is used for nonattractive contact processes.*

Returning to the task at hand (and to the definition of duality just given), we now prove that every additive growth model has an additive dual.

Proof of Theorem 5.2.7. Given the additive coupling of η_s, $0 \le s \le t$, for all possible η_0, and given the dual configuration ζ_0, set $\Lambda_0 = \{\eta : \eta \sim \zeta_0\}$ and

$$\begin{aligned} \Lambda_s &= \{\eta_{t-s} : \eta_t \sim \zeta_0\} \\ \Lambda_{s^+} &= \{\eta_{t-s^+} : \eta_t \sim \zeta_0\} \end{aligned}$$

where η_{t-s^+} is the configuration just preceding an event, if an event occurs at time s. Suppose for the moment that $|S| < \infty$ so that events happen one at a time. In the case $|S| = \infty$, we suppose the hypothesis of Lemma 5.2.2 are satisfied. Then, for any finite $T \subset S$ and $t > 0$, run the process on an increasing sequence of sets T_i with $\cup_i T_i = S$, that is, for each i, construct the process using only the events that occur on T_i. Then from the conclusion of Lemma 5.2.2 there exists an almost surely finite i_0 such that if $i \ge i_0$, then $\eta_s|_T$, $0 \le s \le t$, run on T_i, is equal to $\eta_s|_T$, $0 \le s \le t$, for the process itself. The duality relation then follows from the proof in the finite case.

Say that a collection of configurations $\Lambda \subset F^S$ is *dualizable* if there is a dual

configuration ζ, called the minimal configuration, so that

$$\Lambda = \{\eta : \eta \sim \zeta\}$$

If such a ζ exists it is unique. It is immediate that Λ_0 is dualizable. Suppose that Λ_s is dualizable and that there is an event at time s. We have $\Lambda_{s^+} = \{\eta_{t-s^+} : \eta_{t-s} \in \Lambda_s\}$, so we examine the mapping $e : F^T \to F^T$ that corresponds to the event at time s. Let ζ be the minimal configuration for Λ_s and let θ be its restriction to T. Then $\eta \in \Lambda_{s^+}$ if and only if $e(\phi) \sim \theta$ where ϕ is the restriction of η to T. Equivalently,

$$\eta \in \Lambda_{s^+} \Leftrightarrow e(\phi)(y) \in \theta(y) \text{ for some } y \in T$$

For each $y \in T$ we have by additivity that $e(\phi)(y) = \vee_x e(\phi_x)(y)$ where $\phi_x(x) = \phi(x)$ and $\phi_x(y) = 0$ if $y \neq x$. Since $\theta(x)$ is decomposable it then follows that

$$e(\phi)(y) \in \theta(y) \Leftrightarrow e(\phi_x)(y) \in \theta(y) \text{ for some } x \in T$$

We then set $\delta_x(a)$ to be the local configuration ϕ with $\phi(x) = a$ and $\phi(y) = 0$ for $y \neq x$ and since

$$e(\delta_x(a \vee b)) = e(\delta_x(a)) \vee e(\delta_x(b))$$

it follows that if $e(\delta_x(a \vee b))(y) \in \theta(y)$, then either $e(\delta_x(a))(y) \in \theta(y)$ or $e(\delta_x(b))(y) \in \theta(y)$, since $\theta(x)$ is decomposable. By attractiveness and since $\theta(x)$ is increasing, if $\phi \leq \phi'$ and $e(\phi)(y) \in \theta(y)$ then $e(\phi')(y) \in \theta(y)$. It follows that for each $x, y \in T$ the set

$$\theta_{+,y}(x) = \{a \in F : e(\delta_x(a))(y) \in \theta(y)\}$$

is increasing and decomposable, thus is a dual type. Thus, setting

$$\zeta_{s^+}(x) = \{a \in F : \delta_x(a) \in \theta(y) \text{ for some } y \in T\}$$

when $x \in T$ (and $\zeta_{s^+}(x) = \zeta_s(x)$ when $x \notin T$) we observe that $\zeta_{s^+}(x) = \cup_y \theta_{+,y}(x)$ is a dual type, since dual types are closed under unions, and moreover that

$$\eta \in \Lambda_{s^+} \Leftrightarrow \eta(x) \in \zeta_{s^+}(x) \text{ for some } x$$

or in other words, Λ_{s^+} is dualizable, and ζ_{s^+} is its minimal configuration. This completes the induction step and shows that an additive process has a dual which is

defined in the manner just described. Dual mappings $\tilde{e} : \tilde{F}^T \to \tilde{F}^T$ can be given as

$$\tilde{e}(\theta)(x) = \{a \in F : e(\delta_x(a)) \sim \theta\}$$

with the property that $e(\phi) \sim \theta \Leftrightarrow \phi \sim \tilde{e}(\theta)$. We observe that $\phi \sim \tilde{e}(\theta \vee \theta') \Leftrightarrow e(\phi) \sim \theta \vee \theta' \Leftrightarrow e(\phi)(x) \in \theta(x) \cup \theta(x)$ for some $x \Leftrightarrow e(\phi) \sim \theta$ or $e(\phi) \sim \theta' \Leftrightarrow \phi \sim \tilde{e}(\theta)$ or $\phi \sim \tilde{e}(\theta') \Leftrightarrow \phi \sim \tilde{e}(\theta) \vee \tilde{e}(\theta')$, which shows that the dual is additive. By definition of the dual types and the duality relation, it follows that the dual is a growth model.

□

In order to push the duality relation backwards through an event it was necessary to decouple the interactions in the forward process, first by sites, then by type. It was in order to decouple by type that we required dual types be decomposable. In general, considerable complexity can arise, however in the following simple case we can easily identify the dual.

Example 5.3.6 (An N-stage contact process)**.** *Take $F = \{0, 1, ..., N\}$ with primitive types $1, ..., N$ totally ordered in the usual ordering of integers. Let G be an undirected graph, then define a process on G using the transitions*

- *$i \to i + 1$ at rate γ, for $i = 1, ..., N - 1$,*
- *$i \to 0$ at rate 1, for $i \geq 1$, and*
- *$0 \to 1$ at rate λ times the number of neighbours in state N.*

The set of dual types is 0 and $E_j = \{i \in F : i \geq j\}$, for $j = 1, ..., N$, with $E_j \subset E_k$ for $j > k$. It is not hard to check that the dual transitions are

- *$E_{i+1} \to E_i$ at rate γ, for $i = 1, ..., N - 1$,*
- *$E_i \to 0$ at rate 1, for $i \geq 1$, and*
- *$0 \to E_N$ at rate λ times number of neighbours in state E_1.*

Introducing the correspondence $E_j \leftrightarrow N + 1 - j$, the set of dual types becomes $\{0, 1, ..., N\}$ with the usual order, and the transitions are identical to those in the original process. Thus we see that this process is self-dual.

5.4 Primitivity and colour decomposition

The goal of this section is to prove Theorem 5.2.9. We will make our way towards multi-colour systems in a few steps. First we need a couple of definitions.

Definition 5.4.1. *For a growth model with a join, an active type $a \in F$ is* primitive *if $a = b \vee c \Rightarrow a = b$ or $a = c$, and is* compound *otherwise. The primitive types are denoted F_p.*

It is possible to have $a, b \in F_p$ with $a < b$; this is true for example when F is totally ordered, since in that case all active types are primitive. We say that $a, b \in F$ are *incomparable* or $a <> b$ if $a \not\leq b$ and $b \not\leq a$, which is the opposite of being comparable.

Definition 5.4.2. *A set $C \subset F_p$ satisfying $a = b$ or $a <> b$ for $a, b \in C$ is called a* colour combination. *We denote by $\mathcal{C}$ the set of colour combinations.*

Notice that $\mathcal{C}$ is partially ordered by $C \leq C' \Leftrightarrow$ for each $a \in C$, there exists $b \in C'$ such that $b \geq a$. There is also a join defined by

$$C \vee C' = \{a \in C \cup C' : a \not\leq b \text{ for any } b \in C \cup C'\}$$

Definition 5.4.3. *For the set of types F of a growth model with a join, the* multi-colour expansion *F_* is the set $\mathcal{C}$ of colour combinations of F, together with a passive type 0 satisfying $0 \leq a$ for each $a \in F_*$.*

From the definition of the join on $\mathcal{C}$ it follows that the sets $\{a\}$, for $a \in F_p$, are the primitive types in F_*.

There is a mapping $\pi : F_* \to F$ defined by $\pi(0) = 0$ and

$$\pi : C \mapsto \bigvee_{a \in C} a$$

for $C \in \mathcal{C}$. Notice that for $b \in F$, $\pi^{-1}(b) = \{C \in \mathcal{C} : b = \bigvee_{a \in C} a\}$. If $\pi(C) = a$ we say that C is a *decomposition* of a, and if a has a unique decomposition we use $C(a)$ to denote it. Notice that $C(a \vee b) = C(a) \vee C(b)$, if they exist.

Proposition 5.4.4. *The mapping $\pi : F_* \to F$ has the following properties:*

- $a \leq b \Rightarrow \pi(a) \leq \pi(b)$.
- $\pi(a \vee b) = \pi(a) \vee \pi(b)$.
- π *is surjective.*
- $\pi^{-1}(a) = \{a\}$ *for each* $a \in F_p$.

Proof. The first and second assertions are clear. To see that π is surjective, first note that F can be partitioned into $\{0\}$, primitive types, and compound types. By definition, $\pi(0) = 0$. If $a \in F_p$, the set $\{a\}$ belongs to $\mathcal{C}$ and $\pi(\{a\}) = a$. Finally, if $b \in F$ is compound, then by repeatedly breaking up joins, we see that b can be written as

$$b = \bigvee_{a \in C} a$$

for some $C \in \mathcal{C}$, and $\pi(C) = b$.

For each $a \in F_p$, from primitivity it follows that a has no decomposition containing more than one element, moreover $\pi(\{c\}) \neq a$ if $c \neq a$. Since $\pi(\{a\}) = a$ it follows that $\pi^{-1}(a) = \{a\}$. □

For future use we note that π can be defined on configurations by letting $\pi(\eta)(x) = \pi(\eta(x))$ for each x.

From Proposition 5.4.4 it follows that

- each primitive type has exactly one decomposition,
- each compound type has at least one decomposition, and
- primitive types in F *lift* to primitive types in F_* in the sense that
  1. the sets $\{a\}$ are the primitive types in F_*,
  2. $\pi^{-1}(a) = a$ for each $a \in F_p$ and
  3. the order is preserved under π^{-1}

From the lifting property we see that the sets of primitive types F_{*p} and F_p are isomorphic, π being a natural isomorphism between them.

Remark 5.4.5. *A compound type in F can have more than one decomposition and thus more than one preimage under π, as in the case $F = \{0, 1, 2, 3, 4\}$ with $0 < 1, 2, 3 < 4$, so that $F_p = \{1, 2, 3\}$ and $4 = 1 \vee 2 = 1 \vee 3 = 2 \vee 3 = 1 \vee 2 \vee 3$. Here F_* can be labelled $\{0, 1, 2, 3, 4, 5, 6, 7\}$ with $\pi(a) = a$ and $\pi^{-1}(a) = a$ for $a = 0, 1, 2, 3$, $4 = 1 \vee 2$, $5 = 1 \vee 3$, $6 = 2 \vee 3$ and $7 = 1 \vee 2 \vee 3$, and with $\pi^{-1}(4) = \{4, 5, 6, 7\}$. If we add $1 < 2$ to the order on F then F_p is still $\{1, 2, 3\}$ but $4 = 1 \vee 3 = 2 \vee 3$ only (in the sense of colour combinations). In this case F_* can be labelled $\{0, 1, 2, 3, 4, 5\}$ with $\pi(a) = a$ and $\pi^{-1}(a) = a$ for $a = 0, 1, 2, 3$, $4 = 1 \vee 3$ and $5 = 2 \vee 3$, and with $\pi^{-1}(4) = \{4, 5\}$.*

Definition 5.4.6. *A set of types F with a join is* multi-colour *if every compound type has exactly one decomposition. If F is multi-colour and $b \in F$, $C(b)$ denotes its decomposition.*

For any set of types F with a join, its multi-colour expansion F_* is multi-colour. From Proposition 5.4.4, F being multi-colour is equivalent to $\pi : F_* \to F$ being injective, which is equivalent to F_* and F being isomorphic, π being a natural isomorphism between them. Thus for a multi-colour system there is a $1 : 1$ correspondence between the set of active types and the set of colour combinations.

Before proving Theorem 5.2.8, we consider the following simple example of a multi-type system and its lift.

Example 5.4.7 (A 3-type system)**.** *Take the model with $F = \{0, 1, 2, 3, 4\}$ with incomparable primitive types $F_p = \{1, 2, 3\}$ and $4 = 1 \vee 2 = 1 \vee 3 = 2 \vee 3 = 1 \vee 2 \vee 3$. Each non-zero type dies at rate 1, and produces an individual of type $1, 2$ or 3 at rate $\lambda/3$. Then the lift has types $F_* = \{0, 1, 2, 3, 1 \vee 2, 1 \vee 3, 2 \vee 3, 1 \vee 2 \vee 3\}$, and the transitions are the same: each non-zero type dies at rate 1, and produces an individual of type $1, 2$ or 3 at rate $\lambda/3$. The only difference is that if, for example, a 2 lands on a site occupied by a 3, then in the lift the resulting type is $2 \vee 3$ instead of just 4; in the lift we distinguish the various joins of primitive types.*

We now prove Theorem 5.2.8. For the proof, define the *minimal* elements $\min(E)$ of a set E to be the elements $a \in E$ such that $a \not\succ b$ for any $b \in E$. Each set E has a *layer partition* $E_1, E_2, ..., E_k$ given by $E_1 = \min(E)$ and $E_i = \min(E - E_{i-1})$ for $i > 1$.

Proof of Theorem 5.2.8. We take transition mappings $e : F^T \to F^T$ to new mappings

$$e_* : F_*^T \to F_*^T$$

Having done this, ξ_t is constructed using the transitions e_*. To ensure that $\pi(x_0) = \eta_0 \Rightarrow \pi(\xi_t) = \eta_t$ for $t > 0$ it suffices to show that $\pi \circ e_* = e \circ \pi$. Since F_{*p} and F_p are isomorphic we use a to denote both $a \in F_{*p}$ and $\pi(a) \in F_p$.

Partition F_p into layers $F_{p1}, ..., F_{pk}$. For a set B let $\vee B$ denote $\vee_{b \in B} b$. For $a \in F_1$ and $x \in T$ let

$$e_*(\delta_x(a)) = \vee \pi^{-1}(e(\delta_x(a)))$$

Since for $a, b \in F_*$, $\pi(a) = \pi(b) \Rightarrow \pi(a \vee b) = \pi(a) \vee \pi(b) = \pi(b)$, it follows that $\pi(e_*(\delta_x(a)) = e(\delta_x(a))$. To extend e_* to other layers in an additive way we first make the following observations. If $a < a'$ then $a' = a \vee a'$ and

$$e(\delta_x(a')) = e(\delta_x(a \vee a')) = e(\delta_x(a) \vee \delta_x(a')) = e(\delta_x(a)) \vee e(\delta_x(a'))$$

by additivity of e. If $\phi' = \phi \vee \phi'$ and $\psi \in \pi^{-1}(\phi)$, $\psi' \in \pi^{-1}\phi'$ then

$$\pi(\psi \vee \psi') = \pi(\psi) \vee \pi(\psi') = \phi \vee \phi' = \phi'$$

Thus, for $i > 1$ suppose e_* is defined on $\delta_x(a)$ for each $x \in T$, $a \in F_{pj}$ and $j < i$. Then let

$$e_*(\delta_x(a')) = \vee \pi^{-1}(e(\delta_x(a'))) \vee \bigvee \{e_*(\delta_x(a)) : a < a', a \in F_{pj}, j < i\}$$

It follows from the last observation that $\pi(e_*(\delta_x(a'))) = e(\delta_x(a'))$, moreover $e_*(\delta_x(a) \vee \delta_x(a')) = e_*(\delta_x(a'))$ for any $a < a', a \in F_{pj}$ for some $j < i$. Doing this for each i, we find that $e_*(\delta_x(a))$ is defined for each $x \in F^T$ and each $a \in F_p$, and is additive on that set.

For each compound type $b \in F_*$ and $x \in T$, let

$$e_*(\delta_x(b)) = \bigvee \{e_*(\delta_x(a)) : a \in C(b)\}$$

Finally, for arbitrary $\phi \in F_*^T$ let $e_*(\phi) = \bigvee_x e_*(\phi_x)$ where $\phi_x(x) = \phi(x)$ and $\phi_x(y) = 0$

for $y \neq x$. Both extensions preserves additivity, and so e_* is defined and additive on all of F_*^T.

To check that $\pi(e_*(\phi)) = e(\pi(\phi))$ observe that $\pi(e_*(\delta_x(a))) = e(\pi(\delta_x(a)))$ for $a \in F_p$ and so

$$\begin{aligned}
\pi(e_*(\phi)) &= \pi(\vee e_*(\delta_x(a))) \\
&= \vee\pi(e_*(\delta_x(a))) \\
&= \vee e(\pi(\delta_x(a))) \\
&= e(\pi(\vee\delta_x(a))) \\
&= e(\pi(\phi))
\end{aligned}$$

where the $\vee$ is over $x \in T$ and $a \in C(\phi(x))$. □

Remark 5.4.8. *If ξ_t is the lift of η_t and $\eta_t = \underline{0}$ then $\xi_t = \underline{0}$, since the configuration with all 0's has only itself as a preimage under π. Therefore η_t survives, i.e., $\eta_t \neq \underline{0}$ for all $t > 0$, if and only if ξ_t survives. For this reason, when studying questions of survival (which are usually the interesting questions for growth models), it is sufficient to restrict our attention to multi-colour systems.*

Having lifted to a multi-colour system, we now consider duality in that context, for which the following fact is helpful.

Lemma 5.4.9. *If F is multi-colour and $a \in F$ is primitive then*

$$a \leq b \vee c \Rightarrow a \leq b \text{ or } a \leq c$$

Proof. It suffices to think in terms of colour combinations. If $a \leq b \vee c$ then $a \leq a'$ for some $a' \in C(b \vee c)$. Since $C(b \vee c) = C(b) \vee C(c) \subset C(b) \cup C(c)$, $a' \in C(b)$ or $a' \in C(c)$, so that $a \leq b$ or $a \leq c$. □

If F is not multi-colour this may not hold, as in the case $F = \{0, 1, 2, 3, 4\}$ with $0 < 1, 2, 3 < 4$, in which case $3 \leq 4 = 1 \vee 2$ but $3 \not\leq 1, 2$.

From Lemma 5.4.9 we conclude that for a multi-colour system the set

$$E_a = \{a \in F : b \geq a\}$$

is not only increasing, but also decomposable, when a is primitive. Thus for a multi-colour growth model and each primitive a, E_a is a dual type. We show next that these are exactly the primitive dual types.

We say that $C(c) \succ C(b)$ if $a' \geq a$ for some $a' \in C(c)$, $a \in C(b)$. Note that $\succ$ is reflexive but not necessarily antisymmetric or transitive; for example, let $b_i = a_i \vee a_{i+1}$ for some set of incomparable a_i, then $C(b_i) \succ C(b_j) \Leftrightarrow |i-j| \leq 1$.

Define also $C \sqcup C' = \{a \in C \cup C' : a \not\succ b \text{ for any } b \in C \cup C'\}$, or $= 0$ (the passive type) if $C \cap C' = \emptyset$.

Proposition 5.4.10. *If F is multi-colour then $\tilde{F}$ is multi-colour with primitive types*

$$\tilde{F}_p = \{E_a : a \in F_p\}$$

and F identifies with $\tilde{F}$ on active types via

$$b \leftrightarrow E_b = \{c \in F : C(c) \succ C(b)\}$$

Moreover,

$$E_a \vee E_b = \{c \in F : C(c) \succ C(a) \sqcup C(b)\}$$

In particular, the identification $a \leftrightarrow E_a$ of F_p with $\tilde{F}_p$ is order-reversing.

Proof. We first exhibit the identification of F with $\tilde{F}$, then deduce the rest. For $E \in \tilde{F}$ define the *minimal types*

$$E_m = \min(E) = \{a \in E : a \not\succ b \ \forall b \in E\}$$

Each minimal type $a \in E_m$ is primitive, since if $a = b \vee c$ then since E is decomposable, $b \in E$ or $c \in E$; if $b \in E$ then $a \not\succ b$, but since $a \geq b$ it follows that $a = b$. Moreover, $a <> b$ for $a, b \in E_m$ thus E_m is a colour combination.

Since E is increasing, $E \supseteq \{b : b \geq a \text{ for some } a \in E_m\} = \bigcup_{a \in E_m} E_a$, moreover $b \in E \Rightarrow b \in E_m$ or $b > a$ for some $a \in E_m$ which means $E \subseteq \bigcup_{a \in E_m} E_a$, thus $E = \bigcup_{a \in E_m} E_a$. Therefore $E = \{c \in F : C(c) \succ E_m\}$, or $E = E_b = \{c \in F : C(c) \succ C(b)\}$ where $b \in F$ is the unique active type that satisfies $C(b) = E_m$. To each active type

b corresponds a different set E_b, which yields the identification of F with $\tilde{F}$. The expression for $E_a \vee E_b$ follows easily.

Observe that if $a \in F_p$ and $a \in C(b) \sqcup C(c)$ then either $a \in C(b)$ or $a \in C(c)$, so that if $E_a = E_b \vee E_c$ then $a \in \min E_b$ or $a \in \min E_c$. Supposing the former, $E_a \subset E_b$ which means $E_a = E_b$, i.e., E_a is primitive. On the other hand, if b is compound then $C(b) = \sqcup_{a \in C(b)} C(a)$, so that $E_b = \vee_{a \in C(b)} E_a$ and E_b is not primitive. Thus $\tilde{F}_p = \{E_a : a \in F_p\}$.

Since $E_b \neq E_c$ when $C(b) \neq C(c)$, the dual is multi-colour. If $a, b \in F_p$ are primitive and $a < b$ then $C(a) \sqcup C(b) = C(a)$, so $E_a > E_b$.

□

We now approach the *double dual*. If a growth model has a dual, then from the proof of Theorem 5.2.7 it follows that its dual is additive, and so the dual has a dual, which is the double dual, whose set of types we denote by $\overline{F}$, using λ to denote an element of $\overline{F}$. For a multi-colour system, it follows from an application of Proposition 5.4.10 to $\tilde{F}$ that the double dual is multi-colour and has primitive types

$$\lambda_a = \{E \in \tilde{F} : E \supseteq E_a\}$$

where $a \in F_p$, which are in $1:1$ correspondence with F_p. Since duality is order-reversing on primitive types, double duality is order-preserving on primitive types. By identifying active types with colour combinations, we obtain an order-preserving identification of F with $\overline{F}$ given by

$$b \leftrightarrow \lambda_b = \{E \in \tilde{F} : E \supseteq E_a \text{ for some } a \in C(b)\}$$

For a local configuration $\phi \in F^T$, let $\Xi_\phi \in \overline{F}^T$ denote the corresponding local double dual configuration, under the above identification. The following relationship holds for compatibility.

Lemma 5.4.11. *Let $\phi \in F^T$ and $\theta \in \tilde{F}^T$, then*

$$\phi \sim \theta \Leftrightarrow \theta \sim \Xi_\phi$$

Proof. It is sufficient to have $b \in E \Leftrightarrow E \in \lambda_b$, for each $b \in F$ and $E \in \tilde{F}$, and it

is not hard to check that $b \in E_c \Leftrightarrow C(b) \succ C(c) \Leftrightarrow \exists a \in C(b) : C(a) \succ C(c) \Leftrightarrow C(a) \sqcup C(c) = C(c) \Leftrightarrow E_a \vee E_c = E_c \Leftrightarrow E_c \supseteq E_a \Leftrightarrow E_c \in \lambda_b$. □

We now prove Theorem 5.2.9 by showing that if an additive multi-colour growth model has a dual, then it identifies with its double dual, which since the double dual is additive, implies that the process itself is additive.

Proof of Theorem 5.2.9. It suffices to show that transitions in the double dual commute with the identification of F^S with $\overline{F}^S$, i.e., to show that for double dual transition mappings $\overline{e}$,

$$\overline{e}(\Xi_\phi) = \Xi_{e(\phi)}$$

Transition mappings $\overline{e}$ of the double dual are given by

$$\overline{e}(\Xi) = \{\theta : \tilde{e}(\theta) \sim \Xi\}$$

where the identification $\Xi \leftrightarrow \{\theta : \theta \sim \Xi\}$ is implicit. However, $\tilde{e}(\theta) \sim \Xi_\phi \Leftrightarrow \phi \sim \tilde{e}(\theta) \Leftrightarrow e(\phi) \sim \theta \Leftrightarrow \theta \sim \Xi_{e(\phi)}$, using Lemma 5.4.11 and the duality relation for the process on F^S. Therefore,

$$\overline{e}(\Xi_\phi) = \{\theta : \theta \sim \Xi_{e(\phi)}\} = \Xi_{e(\phi)}$$

as desired. □

5.5 Percolation viewpoint

A percolation model (percolation process) is any stochastic model with some sort of spatial structure and a notion of *path* from one point to another. We distinguish three models in increasing order of resemblance to a particle system.

1. The classical percolation model [23]: take the random subgraph of the graph (V, E) with $V = \mathbb{Z}^2$ and $E = \{xy : \|x-y\|_\infty = 1\}$ in which each edge is included independently with probability $p \in (0, 1)$. Sites $x, y \in V$ are *connected* if they are linked by a path of edges. Percolation occurs if $(0, 0)$ belongs to an infinite cluster of connected sites.

2. Oriented percolation in two dimensions [8]: take the random subgraph of the directed graph (V, E) with $V = \{(x, y) \in \mathbb{Z}^2 : y \geq 0, x + y \text{ is even}\}$ and edges

from $(x, y) \to (x-1, y+1)$ and from $(x, y) \to (x+1, y+1)$ for each $(x, y) \in V$, again with each edge included independently with probability p. Percolation occurs if there is an infinite path of directed edges starting at $(0, 0)$.

3. The contact process (this one *is* a particle system): take the particle system with types $\{0, 1\}$ and transitions

   - $0 \to 1$ at site x at rate $\lambda n_1(x)$ and
   - $1 \to 0$ at each site at rate 1

   where $n_1(x)$ is the cardinality of the set $\{y : xy \in E, y \text{ is in state } 1\}$. The spacetime event map for the process leads to a percolation model on the spacetime set $V \times \mathbb{R}^+$ in which the interval $\{x\} \times [s, t]$ is an upward directed edge if there are no deaths on the interval, and $(x, t) \to (y, t)$ is a directed edge whenever there is a transmission event from x to y at time t. Percolation occurs from $(x, 0)$ if there is an infinite path (i.e., a path of infinite length) of directed edges starting at $(x, 0)$.

For the contact process it is not hard to confirm that percolation from $(x, 0)$ implies survival starting from the single active site x, and vice versa. For a multi-particle system it may seem less clear how to prescribe a percolation process. However, for an additive multi-colour system it can be easily achieved, provided that we colour code the edges, as we show now.

On the spacetime set $V \times \mathbb{R}^+$, the interval $\{x\} \times [s, t)$ is an edge of every colour if there are no events on that interval, since any colour can propagate upwards along it. If an event occurs at time t in the transition mapping $e : F^T \to F^T$, there is an edge from $(x, t^-) \to (y, t)$ (where $x = y$ is allowed) with initial colour a and terminal colour b if $b \in C(e(\delta_x(a))(y))$, that is, if type a at x *produces* type b at y through e. Edges $\{x\} \times [s, t)$ with colour a and $\{y\} \times [t, u)\}$ with colour b can be linked by an edge from $(x, t^-) \to (y, t)$ if it has initial colour a and terminal colour b. Percolation occurs from type a at $(x, 0)$ if there is an infinite path of directed coloured edges starting at $(x, 0)$ with colour a, and it characterizes survival:

Theorem 5.5.1. *For an additive multi-colour particle system, survival occurs starting from the single type a active site x if and only if percolation occurs from type a at $(x, 0)$.*

Proof. If there is a $t > 0$ so that all coloured paths from $(x, 0)$ are contained in $V \times [0, t)$ then it follows by additivity that the process has died out by time t. If there is an infinite path, then it follows again by additivity that the process survives. □

5.6 Population viewpoint

Multi-colour systems can be understood in a natural way as models of interacting populations. Let each colour correspond to a type of organism (plant, animal, fungus etc.). Recalling what it means to be a colour combination, at each site we can have up to one each of any collection of organisms of different types, provided the types are incomparable. Thus the restrictions on the population are that

- there can be at most one organism of a given type, at a given site, and
- organisms of comparable type cannot exist together at the same site.

These are realistic assumptions, if we think of populations having a carrying capacity, and of stronger types excluding weaker types. If we wanted to model a carrying capacity $K > 1$ we could just assign a primitive type to each number $1, 2, 3, ..., K$ of organisms of a certain type and then order them $1 < 2 < ... < K$.

At an event with transition mapping $e : F^T \to F^T$, for $a, b \in F_p$ and $x, y \in T$ we say that a at x *produces* b at y, denoted $e : (a, x) \to (b, y)$, if

$$b \in C(e(\delta_x(a))(y))$$

that is, if b is one of the organisms in $e(\phi)(y)$, when ϕ is the local configuration having only the organism a at site x and no other organisms at sites $y \in T, y \neq x$.

Using this notion of production, we can define a multi-type branching process in which each individual gives birth to individuals of the appropriate types at the appropriate rates. We can then ask how the spatial process differs from the branching process, which is answered by the following result.

Theorem 5.6.1. *For an additive multi-colour system, the only interactions between organisms are due to the effects of crowding and exclusion.*

Proof. For an additive system, to compute the effect of a transition we first focus on each organism (x, a) present before the event and include everything that it produces. Then, if for example (x, a) produces (z, c) but (y, b) produces (z, d) with $d > b$, organism d *excludes* organism b; if (y, b) also produces (z, b) then only one copy of organism b is retained at z. Thus, accounting for each organism's production and then clearing any duplicate or excluded organisms gives the result of a transition. □

With this characterization in hand we now describe which types of interaction fail to be additive:

1. If the system exhibits *inhibition*, that is, the presence of one type of organism decreases the production of another organism, then it cannot be additive, or even attractive; here's why. The introduction of an organism increases the configuration. However, for some organism its introduction decreases the production of another organism, so we have a transition e with $\phi > \phi'$ but $e(\phi) < e(\phi')$.

2. If the system exhibits *cooperation*, that is, multiple organisms can produce together what they could not produce alone, then $e(\vee_{x,a}\delta_x(a)) > \vee_{x,a}e(\delta_x(a))$ for some collection of organisms (x, a). Thus the system may be attractive, but cannot be additive.

In an additive system, organisms neither inhibit nor enhance each other's production. In this sense, additive systems are the least interactive of all interacting particle systems. However, mathematically, they are relatively easy to analyze and have fairly strong properties, and biologically, they lie on the boundary between inhibition and cooperation, and make good models of situations where, to first order of approximation, the only interaction between organisms arises from the fact that they take up space.

5.7 Positive Correlations

We begin by recalling some definitions mentioned in the introduction. On a partially ordered state space, we say a function f is *increasing* if $\eta \leq \eta' \Rightarrow f(\eta) \leq f(\eta')$, and is *decreasing* if $-f$ is increasing. For a particle system, increasing functions include

indicators of the events $\{\eta(x) \geq a\}$ for some a, x. A measure μ on a partially ordered state space has positive correlations if

$$\mathbb{E}_\mu fg \geq \mathbb{E}_\mu f \mathbb{E}_\mu g$$

for all increasing functions f, g, provided the expectations are defined. It can be checked, for example, that deterministic configurations, that is, measures concentrated on a single configuration, have positive correlations. Letting μ_t denote the distribution of the process at time t, we say a particle system preserves positive correlations or preserves PC if for any μ_0,

$$\mu_0 \text{ has positive correlations } \Rightarrow \mu_t \text{ has positive correlations for all } t > 0$$

It is shown in [32], Chapter II, that a particle system preserves PC if and only if all transitions are between comparable states, that is, for each e, ϕ, either $e(\phi) \geq \phi$ or $e(\phi) \leq \phi$.

To describe the following it is helpful to introduce a bit of language. At an event with transition mapping $e : F^T \to F^T$, for $a, b \in F_p$ and $x, y \in T$ we say that a at x *produces* b at y, denoted $e : (a, x) \to (b, y)$, if

$$b \in C(e(\delta_x(a))(y))$$

that is, if b is one of the organisms in $e(\phi)(y)$, when ϕ is the local configuration having only the organism a at site x and no other organisms at sites $y \in T, y \neq x$.

For an additive system, we can use this result to characterize PC in terms of production. Say that an organism (x, a) *waxes* if (x, a) produces at least (x, a) (birth or persistence) and *wanes* if it produces at most (x, b) for some $b \leq a$ and nothing else (persistence, demotion or death).

Definition 5.7.1. *A transition mapping e is said to be* productive *if each (x, a) waxes, and* destructive *if each (x, a) wanes. An organism (x, a)* compensates for the loss of *another organism (y, b) in a transition if (y, b) is demoted or dies, but (x, a) produces (y, b).*

Theorem 5.7.2. *For an additive multi-colour system, if each transition is either*

productive or destructive, then the system preserves PC. A necessary and sufficient condition is that for each transition mapping,

- *each organism either waxes or wanes, and*
- *if (x, a) waxes then it compensates for the loss of every organism that wanes.*

Proof. The sufficient condition follow easily from additivity, using the fact that $\phi \leq \psi, \phi' \leq \psi' \Rightarrow \phi \vee \phi' \leq \psi \vee \psi'$. If a transition mapping has both a waxing organism (x, a) that is not waning ((x, a) produces both (x, a) and something else, possibly (x, b) for $b > a$ or $b <> a$, or possibly another organism elsewhere) and a waning organism (y, b), and (x, a) does not compensate for (y, b)'s loss, then letting $\phi = \delta_x(a) \vee \delta_y(b)$, $e(\phi)(y) < b = \phi(y)$ but $e(\phi)(z) > \phi(z)$ for some z, so $e(\phi) <> \phi$. If the condition given does hold, then for any collection of organisms (x, a) joined into a local configuration ϕ, if at least one of them is waxing then $e(\phi) \geq \phi$ since all losses are compensated for by the waxing organism, and if they are all waning then $e(\phi) \leq \phi$. □

As shown in the following example, it is possible that a model preserves PC, but its dual does not.

Example 5.7.3. *Define the* dandelion process *with types $\{0, 1\}$ as follows. For each $x \in V$, fix a dispersal distribution $p(x, \cdot)$ which is an atomic measure on subsets of V. At each site x, at rate 1, if x is occupied then it dies and disperses, occupying the unoccupied sites in the set A with probability $p(x, A)$. Equivalently, for each site $x \in V$ and each subset $A \subset V$, at rate $p(x, A)$ (possibly $= 0$), if x is occupied then it dies and disperses to the sites in A. This process is additive and multi-colour. The dual to the dandelion process is the* helper process *in which at rate $p(x, A)$, x becomes (or remains) unoccupied unless some site in A is occupied, in which case x remains (or becomes) occupied. Since the dandelion process has death/dispersal transitions, it does not preserve PC. In the helper process, however, all waning organisms have their losses compensated for, so it follows that the helper process preserves PC.*

5.8 Complete Convergence

We focus now on additive multi-colour growth models having only productive and destructive transitions, which by Theorem 5.7.2 preserve positive correlations; for

convenience we call these models *simple*. We first note the following fact.

Lemma 5.8.1. *The dual of a simple model is simple (and thus preserves PC).*

Proof. Reversing a productive transition (i.e., looking at the dual transition) gives a productive transition, and the same is true of a destructive transition. □

As in the Introduction, we say the model has single-site survival or more succinctly, *survives* if, started from a single active site, with positive probability there are active sites for all time.

Proof sketch of Theorem 5.2.10. There are three main steps:

- Apply the construction of [3] to both the process and its dual to show that for any $\epsilon > 0$, starting from a large enough patch of infectious sites,

  $$\text{survival} \;\Rightarrow\; \text{dominates oriented percolation with parameter } 1-\epsilon$$

  where the domination is in spacetime, with respect to a rescaled block construction,

- Conclude that the process survives if and only if its dual survives.

- Use the argument of [16] to show complete convergence.

First we use step 1 to show step 2, i.e. that if a simple model survives, then so does it dual. Applying the result of step 1, since supercritical oriented percolation has a nontrivial stationary distribution, we find survival $\Rightarrow \nu \neq \delta_0$, which using the duality relation shows that process survives $\Rightarrow$ dual survives, and applying the same argument to the dual and noting the dual of the dual is the process itself, the converse also holds.

We now address step 1. The construction of [3] works roughly as follows: starting from a large patch of infectious sites in $\mathbb{Z}^d$, on a large spacetime box $R_{0,0} \subset \mathbb{R}^d \times \mathbb{R}^+$, there will be a large number of infectious sites on the top *and* sides of the box, including a large number on *each* orthant of the top of the box – this step uses positive correlations. From these infectious sites, with good probability in every coordinate direction, at least one of them will produce a large patch of infectious sites. Using, say, the positive and negative directions in the first dimension, with good probability the process spreads in both directions and can be restarted with a large patch

of infectious sites in a *deterministic* location above and to the left and right of the starting location, with spacetime boxes $R_{-1,1}$ and $R_{1,1}$, which gives the comparison to oriented percolation.

A large number of additional geometrical details are addressed in [3]. Note that attractiveness, translation invariance, coordinate symmetry, and positive correlations are needed. We also want the model to be irreducible, so that from a single site we can produce a large finite disc of active sites with some probability. In [3] the interactions are nearest neighbour, but the proof is extended in [2] to include finite range interactions. Finally, to show that survival implies the existence a large number of active sites on the top and sides of the spacetime box $R_{0,0}$, the zero state must be reachable from any configuration with finitely many active sties. Notice that if it were not, survival with positive probability would be trivial. It can then be checked that the proof given in [3] extends to the case of an additive growth model that preserves positive correlations, as we have defined it, and step 1 is complete.

It remains to show step 3. The proof given in [16] relies on a calculation in [21] that gives a relatively simple condition for complete convergence. Fix an arbitrary configuration η_0, and a dual configuration ζ_0 with finitely many active sites; doing this for all such ζ_0 we will recover the finite-dimensional distributions of the upper invariant measure. We have that

$$\eta_{2t} \sim \zeta_0 \Leftrightarrow \eta_t \sim \zeta_t$$

where ζ_s, $0 \leq s \leq t$ is run from time $2t$ down to time t, with initial configuration ζ_0. Use the notation $\eta \neq 0$ to denote "not identically zero". Then observe that

$$\mathbb{P}(\eta_t \sim \zeta_t) = \mathbb{P}(\eta_t \neq 0, \zeta_t \neq 0) - \mathbb{P}(\eta_t \neq 0, \zeta_t \neq 0, \eta_t \nsim \zeta_t)$$

Since they use disjoint parts of the Poisson processes, η_s, $0 \leq s \leq t$ and ζ_s, $0 \leq s \leq t$ are independent, so

$$\mathbb{P}(\eta_t \neq 0, \zeta_t \neq 0) = \mathbb{P}(\eta_t \neq 0)\mathbb{P}(\zeta_t \neq 0)$$

for each $t > 0$. Using the duality relation, $\mathbb{P}(\zeta_t \neq 0) = \mathbb{P}(\overline{\eta}_t \sim \zeta_0)$ with $(\overline{\eta}_t)_{t \geq 0}$ the

process started from the largest possible configuration. Letting $t \to \infty$

$$\mathbb{P}(\eta_t \neq 0)\mathbb{P}(\zeta_t \neq 0) \to (1-\alpha)\nu(\{\eta : \eta \sim \zeta_0\})$$

To have complete convergence, it therefore suffices to show that

$$\mathbb{P}(\eta_t \neq 0, \zeta_t \neq 0, \eta_t \nsim \zeta_t) \to 0$$

that is, that when both processes survive, they intersect with probability tending to 1. This is shown in [16], using only a comparison to oriented percolation, and thus applies when we use the construction of [3]. □

5.9 Examples

We conclude with some examples of additive multi-type growth models.

5.9.1 Two-Stage Contact Process

This is a natural generalization of the contact process introduced in [29] and further studied in [19] in which (viewing the contact process as a model of population growth) there is an intermediate juvenile type that must mature before it can produce offspring. There are three types $0, 1, 2$ and the transitions at a site $x \in V$ are

- $2 \to 0$ at rate 1
- $1 \to 0$ at rate $1 + \delta$
- $1 \to 2$ at rate γ
- $0 \to 1$ at rate $\lambda n_2(x)$

where $n_2(x)$ is the cardinality of the set $\{xy \in E : y \text{ is in state } 2\}$. In order to obtain an additive process we take the event structure with transition mappings

- recovery of both type 1 and 2 at each site at rate 1
- recovery of type 1 at the additional rate δ
- onset i.e., $1 \to 2$ at rate γ

- transmission along each edge at rate λ

In other words, it is enough to make sure that whenever a 2 recovers, a 1 also recovers. It is then easy to check that the resulting process has the correct transition rates and is additive with respect to the join operation on types given by $a \vee b = \max a, b$, so the partial order is $0 < 1 < 2$. Since each transition is either productive or destructive, the model preserves positive correlations. Taking the model on $\mathbb{Z}^d$ with nearest-neighbour interactions, the conditions of Theorem 5.2.10 are satisfied, so complete convergence holds. This was shown in [19], with the proof given along the same lines but specifically for the two-stage contact process.

5.9.2 Bipartite Infection Model

Consider the following model with primitive types $\{0, m, f, \}$ where m stands for infectious male and f stands for infectious female. We can construct a two-sex model of infection spread by defining the following transitions on primitive types:

- $m \to 0$ at rate 1
- $f \to 0$ at rate 1
- $m \to m \vee f$ at rate λ
- $f \to m \vee f$ at rate λ
- $0 \to f$ at rate $\lambda n_m(x)$
- $0 \to m$ at rate $\lambda n_f(x)$

where $n_f(x)$ is the cardinality of the set $\{xy : y \text{ has an } m \text{ type particle}\}$ and $n_m(x)$ is the cardinality of the set $\{xy : y \text{ has an } f \text{ type particle}\}$, and then extending transitions to the compound type $m \vee f$ in the obvious way (for example $m \vee f \to f$ and $m \vee f \to m$ each at rate 1 from either m or f recovering). Thus in this model we can imagine that at each site there is one male and one female each of which is either healthy or infectious, and males can only transmit to females, and females to males, either at the same site, or at neighbouring sites. Taking the event construction in which each of the above transitions, at each site/edge, is assigned its own transition mapping, the model is additive with respect to the join with $0 \vee m = m$, $0 \vee f = f$

and $m \vee f$ assigned its own (compound) type, moreover each transition is either productive or destructive. Taking the model on $\mathbb{Z}^d$ with nearest-neighbour interactions, the conditions of Theorem 5.2.10 are satisfied, so complete convergence again holds.

5.9.3 Household Model

Consider the following model introduced in [48], which is somewhat similar to the N-stage contact process considered in Example 5.3.6. The set of types is $\{0, 1, ..., N\}$, with parameters λ, γ and the transitions are

- $i \to 0$ at rate 1
- $i \to i + 1$ at rate $i\gamma$
- $0 \to 1$ at rate $\lambda n_N(x)$

where $n_N(x)$ is the cardinality of the set $\{xy : y \text{ is in state } N\}$. They also examine the variant where $n_N(x)$ is replaced by $\lambda \sum_{xy \in E} \eta(y)$, the sum of the values at neighbouring sites. In the first case, take the event construction in which at each site there is a transition mapping for recovery $i \to 0$ and one mapping for each of the $i \to i + 1$ transitions, $i = 1, ..., N - 1$, and along each edge, there is a mapping for transmission at rate λ. In the second case take the same event construction except with N transition mappings for transmission $e_1, ..., e_N$ each at rate λ, with exactly the mappings $e_1, ..., e_i$ being used when $\eta(y) = i$. In both cases the resulting process is additive with respect to the join given by $a \vee b = \max a, b$, and each transition is either productive or destructive. Taking the model on $\mathbb{Z}^d$ with nearest-neighbour interactions, the conditions of Theorem 5.2.10 are satisfied, so complete convergence again holds.

Acknowledgements

The author's research is supported by an NSERC PGSD2 scholarship.

Bibliography

[1] K.B. Athreya and P.E. Ney. *Branching Processes*. Springer, 1972.

[2] C. Bezuidenhout and L. Gray. Critical attractive spin systems. *Annals of Probability*, 23(3):1160–1194, 1994.

[3] C. Bezuidenhout and G. Grimmett. The critical contact process dies out. *Annals of Probability*, 18(4):1462–1482, 1990.

[4] C. Bezuidenhout and G. Grimmett. Exponential decay for subcritical contact and percolation processes. *Annals of Applied Probability*, 19(3):984–1009, 1991.

[5] F. Brauer, P. van den Driessche, and J. Wu. *Mathematical Epidemiology*. Springer, 2008.

[6] E.I. Broman. Stochastic domination for a hidden markov chain with applications to the contact process in a randomly evolving environment. *Annals of Probability*, 35:2263–2293, 2007.

[7] R. Durrett. On the growth of one-dimensional contact processes. *Annals of Probability*, 8(5):890–907, 1980.

[8] R. Durrett. Oriented percolation in two dimensions. *Annals of Probability*, 12(4):999–1040, 1984.

[9] R. Durrett. The contact process, 1974-1989. *Mathematics of Random Media*, pages 1–18, 1992.

[10] R. Durrett. Stochastic growth models: bounds on critical values. *Journal of Applied Probability*, 29:11–20, 1992.

[11] R. Durrett. Ten lectures on particle systems. *Lecture Notes in Math.*, 1608:97–201, 1995.

[12] R. Durrett. Coexistence in stochastic spatial models. *Annals of Applied Probability*, 19(2):477–496, 2009.

[13] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2010.

[14] R. Durrett. *Random Graph Dynamics*. Cambridge University Press, 2010.

[15] R. Durrett and D. Griffeath. Supercritical contact processes on Z. *The Annals of Probability*, 11(1):1–15, 1983.

[16] R. Durrett and R.H. Schonmann. Stochastic growth models. *Percolation Theory and Ergodic Theory of Infinite Particle Systems*, 8:85–119, 1987.

[17] E. Foxall. Duality and complete convergence for multi-type additive growth models. *To appear in Adv. Appl. Prob, 48.1 (March 2016), preprint on the ArXiv*.

[18] E. Foxall. Convergence and monotonicity for a model of spontaneous infection and transmission. *Advances in Applied Probability*, 46(2):560–584, 2014.

[19] E. Foxall. New results for the two-stage contact process. *J. Appl. Prob.*, 52(1):258–268, 2015.

[20] E. Foxall, R. Edwards, and P. van den Driessche. Social contact processes and the partner model. *To appear in Annals of Applied Probability, preprint on the ArXiv*.

[21] D. Griffeath. Limit theorems for nonergodic set-valued markov processes. *Annals of Probability*, 6(3):379–387, 1978.

[22] D. Griffeath. The basic contact processes. *Stochastic Processes and their Applications*, 11(2):151–185, 1981.

[23] G. Grimmett. *Percolation*. Springer, second edition, 1999.

[24] T.E. Harris. Contact interactions on a lattice. *Annals of Probability*, 2(6):969–988, 1974.

[25] T.E. Harris. A correlation inequality for markov processes in partially ordered state spaces. *Annals of Probability*, 5(3):451–454, 1977.

[26] T.E. Harris. Additive set valued markov processes and graphical methods. *Annals of Probability*, 6(3):355–378, 1978.

[27] T.E. Harris. *The Theory of Branching Processes*. Courier Dover Publications, 2002.

[28] A. Korobeinikov. Lyapunov functions and global properties for SEIR and SEIS epidemic models. *Mathematical Medicine and Biology*, 21:75–83, 2004.

[29] S.M. Krone. The two-stage contact process. *The Annals of Applied Probability*, 9(2):331–351, 1999.

[30] N. Lanchier. Two-scale multitype contact process: coexistence in spatially explicit metapopulations. *Markov Processes and Related Fields*, 17:151–186, 2011.

[31] Y. Li and D. Han. A two-stage contact process on scale-free networks. *Journal of Statistical Physics*, 153:312–324, 2013.

[32] T.M. Liggett. *Interacting Particle Systems*. Springer, 1985.

[33] T.M. Liggett. Improved upper bounds for the contact process critical value. *Annals of Probability*, 23(2):697–723, 1995.

[34] T.M. Liggett. *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. Springer, 1999.

[35] C.J. Mode. *Multitype Branching Processes: Theory and Applications*, volume 34. American Elsevier Pub. Co., 1971.

[36] T. Mountford, J.C. Mourrat, D. Valesin, and Q. Yao. Exponential extinction time of the contact process on finite graphs. *preprint on the ArXiv*, 2012.

[37] T. Mountford, D. Valesin, and Q. Yao. Metastable densities for the contact process on power law random graphs. *Electronic Journal of Probability*, 18(103):1–36, 2013.

[38] C. Neuhauser. Ergodic theorems for the multitype contact process. *Probability Theory and Related Fields*, 91:467–506, 1992.

[39] C. Neuhauser. Ergodic theorems for the multitype contact process. *Probability Theory and Related Fields*, 91(3-4):467–506, 1992.

[40] C. Neuhauser. A long range sexual reproduction process. *Stoch. Process. Appl.*, 53:193–220, 1994.

[41] J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.

[42] R. Pemantle. The contact process on trees. *Annals of Probability*, pages 2089–2116, 1992.

[43] J. Peterson. The contact process on the complete graph with random vertex-dependent infection rates. *Stochastic Processes and their Applications*, 121(3):609–629, 2011.

[44] D. Remenik. The contact process in a dynamic random environment. *Annals of Applied Probability*, 18(6):2392–2420, 2008.

[45] R. Durrett S. Chatterjee. Contact processes on random graphs with power law degree distributions have critical value 0. *Annals of Probability*, 37(6):2332–2356, 2009.

[46] J.E. Steif and M. Warfheimer. The critical contact process in a randomly evolving environment dies out. *Latin American Journal of Probability and Mathematical Statistics-ALEA*, 4:337–357, 2008.

[47] A. Sudbury. Dual families of interacting particle systems on graphs. *Journal of Theoretical Probability*, 13(3):695–716, 2000.

[48] N. Sugimine, N. Masuda, N. Konno, and K. Aihara. On global and local critical points of extended contact process on homogeneous trees. *Mathematical biosciences*, 213(1):13–17, 2008.

[49] N. Sugimine, N. Masuda, N. Konno, and K. Aihara. On global and local critical points of extended contact processes on homogenous trees. *Mathematical Biosciences*, 213(1):13–17, 2008.

[50] P. van den Driessche and J. Watmough. Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1):29–48, 2002.

[51] H. Ziezold and C. Grillenberger. On the critical infection rate of the one-dimensional basic contact process: numerical results. *Journal of Applied Probability*, 25:1–8, 1988.