

Dualities and genealogies in stochastic population models



Dissertation zur Erlangung
des mathematisch-naturwissenschaftlichen Doktorgrades
“Doctor rerum naturalium”
der Georg-August-Universität Göttingen

im Promotionsprogramm
“PhD School of Mathematical Sciences (SMS)”
der Georg-August University School of Science (GAUSS)

vorgelegt von
Tibor Mach
aus Pilsen

Göttingen, 2017

Betreuungsausschuss:

Prof. Dr. Anja Sturm,
Institut für Mathematische Stochastik, Universität Göttingen

Prof. Dr. Dominic Schuhmacher,
Institut für Mathematische Stochastik, Universität Göttingen

Mitglieder der Prüfungskommission:

Referentin: Prof. Dr. Anja Sturm,
Institut für Mathematische Stochastik, Universität Göttingen

Korreferent: Dr. Jan Swart
Institute of Information Theory and Automation of the Czech Academy of Sciences

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Dominic Schuhmacher
Institut für Mathematische Stochastik, Universität Göttingen

Prof. Dr. Max Wardetzky
Institut für Numerische und Angewandte Mathematik, Universität Göttingen

Jun.-Prof. Dr. Andrea Krajina
Institut für Mathematische Stochastik, Universität Göttingen

Jun.-Prof. Dr. Daniel Rudolf
Institut für Mathematische Stochastik, Universität Göttingen

Tag der mündlichen Prüfung: 20. Dezember 2017

Acknowledgements

First and foremost, I would like to thank my advisor, Anja Sturm, who has guided and supported me during the entirety of my studies in Göttingen, and who has always managed to find the time for a discussion - even shortly after giving birth when she was looking after a newborn baby. That is in no way a matter of course. Herzlichen Dank!

Another huge thanks belongs to Jan Swart in collaboration with whom a large part of this thesis came into existence, who readily agreed to be a referee for this thesis even though that includes having to travel roughly five hundred kilometres just before Christmas, and from whom I learned quite a bit about my own mother tongue. Hartelijk dank!

I also want to thank all the people, present and former, from the Institute for Mathematical Stochastics for a great working atmosphere, for helping me feel at home in a town where I initially knew nobody and for a steady supply of nutritious birthday cakes. Thank you all!

My four years in Göttingen would have been very different without the friends I made here, be it the people I met in the salsa courses, in our little cajón group or when hiking. For those of you who stay in Göttingen, I hope to keep seeing you, albeit not as often, since fortunately you will not be that far. For the rest I hope we will stay in contact and still see each other sometimes, despite the distance. Thank you! Muchas gracias! Muito obrigado!

Finally, I would have probably never started the PhD had it not been for the support of my mother and father, of my grandmother, my aunt and my uncle, as well as my friends from Pilsen. You have always been there for me and that has made everything including this thesis much easier. Díky moc!

Contents

Summary	5
I Particle systems	6
I.1 Overview and previous results	7
I.1.1 History and motivation	7
I.1.2 Basic construction and models	9
I.1.2.1 Voter model	11
I.1.2.2 Contact process	12
I.1.2.3 Cooperative branching process	13
I.1.3 Phase transition	15
I.1.4 Duality	18
I.2 Complete graph and the Mean-field Model	20
I.2.1 Introduction and main results	20
I.2.1.1 The process on the complete graph	20
I.2.1.2 The mean-field limit	22
I.2.1.3 Survival versus extinction	24
I.2.1.4 A measure-valued dual	25
I.2.1.5 Convergence of the measure-valued function	27
I.2.1.6 Coupled processes on the complete graph	30
I.2.1.7 Interpretation of the moments	32
I.2.2 Recursive tree processes	33
I.2.2.1 A recursive tree process in discrete time	33
I.2.2.2 A recursive tree process in continuous time	36
I.2.2.3 A recursive tree process of random probability distributions	37
I.2.3 Distribution determining functions	38
I.2.4 The two-sex model	39
I.2.5 Summary and outlooks	43
I.2.6 Proofs	45
I.2.6.1 Proofs of propositions I.2.1.1 and I.2.1.17	45
I.2.6.2 Proof of propositions I.2.1.2 and I.2.1.18	46
I.2.6.3 Proof of Proposition I.2.1.3	47
I.2.6.4 Proofs of Proposition I.2.1.4 and Theorem I.2.1.19	50
I.2.6.5 Proofs of Propositions I.2.1.6 and I.2.1.8, Lemma I.2.1.7 and Theorem I.2.1.9	52
I.2.6.6 Proof of Proposition I.2.1.10	58
I.2.6.7 Proof of Theorem I.2.1.13	58
I.2.6.8 Proof of Proposition I.2.1.20	59
I.2.6.9 Proof of Lemma I.2.3.1	60
I.2.6.10 Proof of Lemma I.2.3.2	60
I.2.6.11 Proof of Lemma I.2.3.3	60

I.3 Cooperative branching on other lattices	61
I.3.1 Cooperative branching on regular trees	61
I.3.2 The model on the integer lattice	61
I.3.2.1 Survival and nontriviality of the upper invariant law	61
I.3.2.2 Random walk dynamics and deaths	62
I.3.3 Proofs	63
I.3.3.1 Proof of Lemma I.3.1.1	63
I.3.3.2 Proof of Proposition I.3.2.1	65
I.3.3.3 Proof of Proposition I.3.2.2	72
I.3.3.4 Proof of Proposition I.3.2.3	73
I.3.3.5 Proofs of Propositions I.3.2.4 and I.3.2.5	74
II Genealogy of neutral loci in a random environment	76
II.1 Overview and previous results	77
II.2 Multiple neutral loci	81
II.2.1 Model	81
II.2.2 Finite population size process	83
II.2.3 Convergence to the diffusion limit	87
II.2.4 Applications	89
II.2.5 Proof of Theorem II.2.3.1	99
II.2.5.1 Existence of the limiting process	99
II.2.5.2 Convergence	104
Appendix A	106
A.1 A few definitions from graph theory	106
A.2 Pathwise duality for monotone interacting particle systems	106
A.3 Appendix to Section II.2.5	108
Bibliography	112

Summary

In this thesis we study population processes in two different settings. In Part I, which arose in collaboration with Dr. Jan Swart, we consider a so-called cooperative branching process. We construct this process as an interacting particle system which consists of a population of individuals living on a discrete space Λ who reproduce cooperatively - that is to say that in order to produce a new individual, it is necessary that two "parents" meet. The individuals also die independently of each other and in some special cases we consider a version where they can also move in the space and coalesce. The term cooperative branching was coined by Sturm and Swart in [SS15] where they study a variant of this model without deaths and with random walk dynamics on \mathbb{Z} . In this thesis, we study the process in some of its variants on a number of different graphs, namely the complete graph with N vertices, a regular tree of a degree d and the d -dimensional lattice \mathbb{Z}^d .

The main part of Part I is Chapter I.2 where we study the model on a complete graph and its mean-field limit. Here we show connections between a so called mean-field dual process which arises naturally in the study of the mean-field limit of the cooperative branching process and recursive tree processes studied by Aldous and Bandyopadhyay in [AB05]. In particular we show that, roughly speaking, the mean-field dual process corresponds to a continuous-time analogue of these recursive tree processes. In Chapter I.3 we provide some results about the critical parameters for survival and existence of a nontrivial invariant law of the cooperative branching process on various graphs. A more detailed introduction to Part I is provided in Chapter I.1.

In Part II we consider a population of individuals which evolves according to a so called Moran model and in which every individual consists of a chromosome with a finite number of genes such that one gene has an effect on the fitness of the individual and other so called neutral genes do not. We assume that the population is further affected by mutation and recombination which, roughly speaking, is a phenomenon which causes two chromosomes to split and form new chromosomes out of their parts during reproduction. We then study the genealogy of a sample of these neutral genes in a setting where the population has evolved for a long time and has reached stationarity. This is a generalization of a model introduced by Barton, Etheridge and Sturm in [BES04] in which only a single neutral gene is considered. The biological concepts mentioned here and the model we consider are introduced in more detail in Chapter II.1.

Part I

Particle systems

Chapter I.1

Overview and previous results

I.1.1 History and motivation

An interacting particle system, as we will understand the term here, is a countable collection of Markov chains, on the same, and typically countable or finite, state space. In addition to that, the jump rates of each process are affected by the states of (some) other processes. It is the last property which justifies the term *interacting* and which makes these models interesting both mathematically and in terms of applications.

Typically, it is required that there be a geometric structure associated with the index set (which is usually called a *lattice*) of the individual Markov chains, so that we can assume the interactions to be local, i.e. only the current states of the processes which are in some sense close to one of the chains affect its current jump rates. One simple and common example of such a structure is the d -dimensional grid \mathbb{Z}^d equipped with the l_1 metric. However, in large part of this thesis we will be considering the lattice to have a structure of a complete graph and we will be studying behaviour of the limit of a particular interacting particle system on the complete graph as the size of the graph grows beyond all bounds.

The field of interacting particle systems first appeared as a distinct discipline in the late 1960's with the work of F. Spitzer and R.L. Dobrushin (for example [Spi69] and [Dob71]), originally motivated by problems of statistical mechanics such as the Ising model which has been studied since 1925 already. Later, the scope of the field expanded, giving rise to models which describe various phenomena ranging from physics and biology to social networks and computer science. Liggett's 1985 book [Lig85] covers the state of the field at the time, providing a general construction of particle systems and showing several properties of four now-classical models, the stochastic Ising models, the Exclusion process, the Contact process and the Voter model. The latter three are then treated more exhaustively in Liggett's 1999 book [Lig99]. These four models already cover a large area in terms of applications. For example, the contact process on \mathbb{Z}^2 (i.e. with the index set \mathbb{Z}^2) has a natural interpretation as modelling a spread of a population in a two-dimensional space. At a certain rate, an individual gives birth to an offspring at one of its neighbouring unoccupied sites. Individuals then die at a constant rate. Similarly, the Ising model on \mathbb{Z}^d can be interpreted as modelling the state of a collection of atoms in a material, where each atom has either a positive or a negative spin and the spin of each atom is affected by the spin of its neighbours.

While real-world populations rarely live on two-dimensional grids and interact strictly with their neighbours based on a few simple rules, the behaviour of more realistic models often turns out not to be substantially different. On the flip side, while interacting particle systems are

indeed usually rather easy to formulate and construct, ascertaining their exact properties and behaviour often proves remarkably difficult. While a lot of work has been done, particularly on the classical processes mentioned above, many questions about their behaviour remain open even after over 40 years of study, whereas answering others has required development of highly non-trivial techniques.

In this thesis, we will concentrate on a particular model, which shares some similarities with the contact process, at least in its biological interpretation. Just like in the contact process we will assume to have a population of individuals on a countable space which reproduces and dies at certain constant rates. However, we will assume that two parents are required to successfully reproduce. This is why we call this model the *cooperative* branching process. Apart from being more realistic, at least in case of a biological interpretation, this model is also more challenging mathematically as we often need to use more general techniques to study its behaviour.

A variation of the process was studied before in [SS15]. In their paper, Sturm and Swart consider the process on \mathbb{Z} with no deaths, but with additional coalescence dynamics. In addition to the cooperative branching, the individuals act as random walkers who coalesce into a single individual whenever they meet. Sturm and Swart provided an estimate for a rate of branching such that the process eventually dies out (i.e. eventually all sites become empty) whenever branching occurs at a smaller rate and conversely the process survives (i.e. some sites are occupied at all times) with a positive probability if the branching occurs at a faster rate. They also showed a similar result about the probability that a given site is occupied at an arbitrarily late time. They then provided estimates for the behaviour of these probabilities in time. In section I.3.2.2 we will show how their results change if we also allow deaths to occur in addition to the coalescing random walk dynamics. We will see, as was conjectured in [SS15], that when deaths are included the probability of survival and of a site being occupied at a given time behave similarly as the corresponding probabilities for the contact process.

With the exception of Section I.3.2.2, we will however mostly be concerned with a model without the coalescing dynamics. We will study the process on various lattices, namely on \mathbb{Z}^d (in section I.3.2.1), on a d -dimensional regular tree (in section I.3.1) and on a complete graph (in the majority of Chapter I.2). In the model on a complete graph our main object of interest will in fact be what we call the mean-field dual process, which we obtain as a limiting process of the so called dual process on the complete graph. A dual process is a process whose behaviour relates to the behaviour of the cooperative branching process in a prescribed way given by a so called duality function. We introduce the concept of duality in Section I.1.4. Our main motivation to studying the dual process is its connection to the so called recursive tree processes which were studied by Aldous and Bandyopadhyay in [AB05]. We discuss these connections in Section I.2.2.

Part I of this thesis is organized as follows. In the remainder of Chapter I.1 we will provide an overview of the construction of interacting particle systems in general, define a few selected basic models, state some of their properties and discuss duality in the context of interacting particle systems. We will also define the cooperative branching process in its most general form.

In Chapter I.2 we study the cooperative branching process on the complete graph as well as its mean-field dual process. We define the cooperative branching process and its dual process on the complete graph and show that the ratio of occupied sites of the cooperative branching process converges to a solution of the ordinary differential equation (I.2.1.17). Then we show

that we can find a corresponding duality between the solutions of this equation and a particular limit of the dual process which we call the mean-field dual process. We proceed to show some further properties of the mean-field dual, particularly another form of duality between the mean-field dual process and solutions of a measure-valued equation given in (I.2.1.33). We show that the solutions of this equation have a natural interpretation related to the distribution of finite collections of coupled cooperative branching processes on a complete graph and that this can then be used to show some properties about a particular recursive tree process. These results and some open questions related to the mean-field dual process and recursive tree processes are summarized in more detail in Section I.2.5.

Chapter 3 provides various results about the cooperative branching process on regular trees and the d -dimensional integer lattice \mathbb{Z}^d . We summarize these results in section I.1.3.

I.1.2 Basic construction and models

Let Λ be a countable set and S a finite set which we will call the *local state space*. By S^Λ we denote the Cartesian product of a number of copies of S corresponding to the number of elements of Λ . An element $x \in S^\Lambda$ is then of the form $x = (x(i))_{i \in \Lambda}$, where $x(i) \in S$ for all $i \in \Lambda$. We will call the elements of S^Λ *configurations* and we will refer to the elements of Λ as *sites*. For any site $i \in \Lambda$ and any configuration $x \in S^\Lambda$, we will call $x(i)$ the (local) state of (the configuration) x at (the site) i . Now, we define an *interacting particle system* as an S^Λ -valued continuous time Markov process

$$X = (X_t)_{t \geq 0} = (X_t(i), i \in \Lambda)_{t \geq 0}. \quad (\text{I.1.2.1})$$

For any $t \geq 0$ and any site $i \in \Lambda$ we call $X_t(i)$ the state of (the process) X at time t and site i . In many cases, we can characterize the generator G of X by a so called random mapping representation (for a general overview of Markov processes and their generators see for example [EK86]). That is, denoting \mathcal{G} as a set of functions $m : S^\Lambda \rightarrow S^\Lambda$, we would like to write G in the form

$$Gf(x) = \sum_{m \in \mathcal{G}} r_m (f(m(x)) - f(x)), \quad x \in S^\Lambda, \quad (\text{I.1.2.2})$$

where r_m , $m \in \mathcal{G}$ are non-negative constants, which we will call rates. Hence, the process jumps from state $x \in S^\Lambda$ to the state $m(x) \in S^\Lambda$ at an exponential rate given by r_m . We will now state sufficient conditions on the maps $m \in \mathcal{G}$ and their rates r_m under which we can indeed represent G in this way.

We will use the same notation as used in [SS16]. We now provide a general construction of interacting particle systems via stochastic flows corresponding to local maps. For a more detailed treatment and proofs see for example [Swa17], Chapters 2 and 4. For any $m : S^\Lambda \rightarrow S^\Lambda$ we define

$$\mathcal{D}(m) := \{i \in \Lambda; \exists x \in S^\Lambda \text{ such that } m(x)(i) \neq x(i)\},$$

i.e. $\mathcal{D}(m)$ is the set of sites $i \in \Lambda$ such that m can change the state at the site i of at least one configuration $x \in S^\Lambda$. We will say that a site $j \in \Lambda$ is *m -relevant* for $i \in \Lambda$ if changing the state $x(j)$ for some $x \in S^\Lambda$ changes the value of $m(x)$ at site i , that is if

$$\exists x, y \in S^\Lambda \text{ such that } m(x)(i) \neq m(y)(i) \text{ and } x(k) = y(k) \forall k \neq j,$$

and for any $i \in \Lambda$ we will denote by $\mathcal{R}_i(m)$ the set of all sites that are *m -relevant* for i . We will say that a map $m : S^\Lambda \rightarrow S^\Lambda$ is *local* if

- $\mathcal{D}(m)$ is finite
- $\mathcal{R}_i(m)$ is finite for all $i \in \Lambda$
- For any $x, y \in S^\Lambda$ and any $i \in \Lambda$, if $x(j) = y(j)$ for all $j \in \mathcal{R}_i(m)$, then $m(x)(i) = m(y)(i)$.

We will restrict our attention (and construction) to interacting particle systems defined via local maps, hence we assume that \mathcal{G} is a set of local maps on S^Λ and in addition to that we also assume that

$$\sup_{i \in \Lambda} \sum_{m \in \mathcal{G}, \mathcal{D}(m) \ni i} r_m (|\mathcal{R}_i(m)| + 1) < \infty. \quad (\text{I.1.2.3})$$

Under these conditions it can be shown (see for example [Swa17]) that (I.1.2.2) indeed defines a generator of a Markov process and that this process can be constructed via a suitable *stochastic flow* with independent increments (whose definition depends on \mathcal{G}). Let T be a metrizable space. We will call a collection $(\mathbf{X}_{s,t})_{s \leq t, s, t \in \mathbb{R}_+}$ of random maps $\mathbf{X}_{s,t} : T \rightarrow T$ a stochastic flow if

- (i) For each $x \in T$, the value $\mathbf{X}_{s,t}(x)$ is càdlàg as a function of both s and t .
- (ii) $\mathbf{X}_{s,t} = \mathbf{X}_{s-,t} = \mathbf{X}_{s,t-} = \mathbf{X}_{s-,t-}$ almost surely for deterministic $s \leq t$ (and all of these functions are well-defined)
- (iii) $\mathbf{X}_{s,s}$ is the identity map and $\mathbf{X}_{t,u} \circ \mathbf{X}_{s,t} = \mathbf{X}_{s,u}$ for all deterministic $s \leq t \leq u$.

A stochastic flow has independent increments if

$$\mathbf{X}_{t_0, t_1}, \dots, \mathbf{X}_{t_{n-1}, t_n} \quad (\text{I.1.2.4})$$

are independent for any $t_0 < \dots < t_n$. If $(\mathbf{X}_{s,t})_{s \leq t}$ is a stochastic flow with independent increments, then for any $s \in \mathbb{R}$

$$X_t := \mathbf{X}_{s, s+t}(X_0), \quad t \geq 0 \quad (\text{I.1.2.5})$$

is a Markov Process. In particular, if \mathcal{G} is a set of local maps and $r_m, m \in \mathcal{G}$ are such that (I.1.2.3) holds, then we obtain the stochastic flow corresponding to (I.1.2.2) as follows. For each $m \in \mathcal{G}$ we define a Poisson process $N^m = (N_t^m)_{t \geq 0}$ with rate r_m and we assume that the processes $N^m, m \in \mathcal{G}$ are all independent. For each $m \in \mathcal{G}$ let $\tilde{N}^m := \{t > 0; N_{t-}^m \neq N_t^m\}$ be the set of jump times of the process N^m and put $\omega := \{(m, t); t \in \tilde{N}^m \text{ for some } m \in \mathcal{G}\}$. Define

$$\omega_{s,u} := \{(m, t) \in \omega : t \in (s, u]\}.$$

If Λ is finite, then $\omega_{s,u}$ is a finite set and its elements can be ordered so that

$$\omega_{s,u} = \{(m_1, t_1), \dots, (m_n, t_n)\},$$

where $t_1 < \dots < t_n$. Then we can define the stochastic flow from s to u as

$$\mathbf{X}_{s,u} := m_n \circ \dots \circ m_1.$$

In the case when Λ is only countable, the construction becomes more complicated, because $\omega_{s,u}$ does not generally have to be finite. However, using the assumption that \mathcal{G} are local maps and that the summability condition (I.1.2.3) holds, we can show that for every site $i \in \Lambda$ and every

$s \leq u$ there exists a set $\omega_s^{i,u} \subseteq \omega_{s,u}$ which is finite and which contains all events $(m, t) \in \mathcal{G} \times \mathbb{R}_+$ which are relevant for the state of $\mathbf{X}_{s,u}(x)(i)$. We can then define the stochastic flow by

$$\mathbf{X}_{s,u}(x)(i) := m_n \circ \dots \circ m_1(x)(i).$$

In the view of (I.1.2.5), we see that we obtain X_t by successively applying the relevant maps $m \in \mathcal{G}$ to the initial state X_0 in the right order. The collection of the Poisson processes $N^m, m \in \mathcal{G}$ form what we call a *graphical representation* of the process X . For a fixed realization of ω we can visualise the path of X over a finite number of sites by denoting the points $t \in N_m$ for each relevant m with a particular symbol representing the map m . For a concrete example of a graphical representation in case of the contact process on \mathbb{Z} see figure I.1.1.

We will now define two simple classical models we have already mentioned earlier, namely the *voter model* and the *contact process*.

I.1.2.1 Voter model

For each $i, j \in \Lambda$ and $x \in S^\Lambda$ we define the voter map as

$$\text{vot}_{ij}(x)(k) := \begin{cases} x(i) & \text{if } k = j, \\ x(k) & \text{otherwise,} \end{cases} \quad (\text{I.1.2.6})$$

In words, this means that when the voter map vot_{ij} is applied, the local state at site k changes to the local state at site i . Generally, for a fixed $k \in \mathbb{N}$, the Markov process corresponding to the voter model has a generator G_{vot} of the form

$$G_{\text{vot}}f(x) = \sum_{i \neq j} w_{i,j} (f(\text{vot}_{ij}(x)) - f(x)), \quad x \in S^\Lambda, \quad (\text{I.1.2.7})$$

where w_{ij} are non-negative constants for all $i, j \in \Lambda$.

While Λ can be any countable set, it is often considered to have a structure of an undirected graph (Λ, E) , where Λ is the vertex set and $E = E_\Lambda$ the corresponding set of edges, that is the set of unordered pairs of sites in Λ . In particular let (Λ, E) be a countable, connected, vertex transitive, locally finite graph (see the Appendix, Section A.1 for the definitions of these properties) with vertex set Λ and set of (undirected) edges E . For $i, j \in V$ we denote the (undirected) edge between i and j by $\langle i, j \rangle \in E$. Since Λ is vertex transitive, each vertex has the same degree, which we denote by D . With an appropriate choice of the rates w_{ij} we obtain the nearest neighbour voter model on a (Λ, E) with the generator

$$G_{\text{votn}}f(x) = \frac{1}{D} \sum_{\substack{i, j \in \Lambda \\ \langle i, j \rangle \in E}} (f(\text{vot}_{ij}(x)) - f(x)), \quad x \in S^\Lambda. \quad (\text{I.1.2.8})$$

The maps $\text{vot}_{ij}, i, j \in \Lambda$ are local and by the choice of the edge set E we see that for any $i \in \Lambda$ the set $\{m \in \mathcal{G}; \mathcal{D}(m) \ni i\}$ is finite for the voter model. Hence the condition (I.1.2.3) is met and the voter model is a well-defined Markov process. Note that rescaling by $\frac{1}{D}$ ensures that for each $i \in \Lambda$ the total rate at which the maps $\text{vot}_{ij}, \langle i, j \rangle \in E$ are applied is equal to 1. For example if Λ is the d -dimensional integer lattice \mathbb{Z}^d with nearest neighbour edges, we get $D = 1/2d$.

One possible interpretation of the voter model (which gives it its name) is that of a network of individuals indexed by Λ who each have one of the opinions from the set S . At rate w_{ij}

the individual i talks to individual j and convinces j of his opinion. Another interpretation is that of a population of different species such that the local state of the process at site i denotes the species occupying i . Then at rate w_{ij} the individual occupying site j dies and is replaced by an offspring of the individual who lives at site i . In Part II of this thesis we will be studying a process which in its simplest form corresponds to the voter model with two species on a complete graph.

I.1.2.2 Contact process

The contact process is a process with a local state space $S = \{0, 1\}$ which models the spread of a population in (a countable) space. In a given configuration $x \in S^\Lambda$ we interpret the sites $i \in \Lambda$ for which $x(i) = 0$ as *unoccupied* or *empty* and those where $x(i) = 1$ as *occupied*. Every occupied site becomes unoccupied at a certain non-negative death rate. We interpret this event as the individual occupying site i dying. Individuals produce offspring at a different (also non-negative) branching rate, so that a site i becomes occupied if an individuals occupying one of its neighbouring sites branches/gives offspring to site i . For $a, b \in \{0, 1\}$ we put $a \vee b := \max(a, b)$ and $a \wedge b := \min(a, b)$. Formally, we define the contact process via the *branching* maps \mathbf{bra}_{ij} , $i, j \in \Lambda$ and *death* maps \mathbf{death}_i , $i, j \in \Lambda$, given for $x \in S^\Lambda$ by

$$\begin{aligned} \mathbf{bra}_{ij}(x)(k) &:= \begin{cases} x(i) \vee x(j) & \text{if } k = j, \\ x(k) & \text{otherwise,} \end{cases} \\ \mathbf{death}_i(x)(l) &:= \begin{cases} 0 & \text{if } l = i, \\ x(l) & \text{otherwise.} \end{cases} \end{aligned} \quad (\text{I.1.2.9})$$

The generator of the process on Λ is given by

$$G_{\text{cont}}f(x) := \sum_{i \neq j} b(i, j) (f(\mathbf{bra}_{ij}(x)) - f(x)) + \sum_i d(i) (f(\mathbf{death}_i(x)) - f(x)), \quad x \in S^\Lambda, \quad (\text{I.1.2.10})$$

where $b(i, j)$ and $d(i)$ are non-negative rates. Assuming that (Λ, E) is a countable, connected vertex transitive locally finite graph of degree D , we obtain the generator of the nearest neighbour contact process on Λ as

$$G_{\text{contn}}f(x) := \alpha \frac{1}{D} \sum_{\substack{i, j \in \Lambda \\ \langle i, j \rangle \in E}} (f(\mathbf{bra}_{ij}(x)) - f(x)) + \delta \sum_{i \in \Lambda} (f(\mathbf{death}_i(x)) - f(x)), \quad x \in S^\Lambda, \quad (\text{I.1.2.11})$$

where $\alpha \geq 0$ is the branching rate and $\delta \geq 0$ the death rate. The maps \mathbf{bra}_{ij} and \mathbf{death}_i are local for any $i, j \in \Lambda$ and by the choice of the edge set E we have that for the nearest neighbour contact process the set $\{m \in \mathcal{G}; \mathcal{D}(m) \ni i\}$ is finite. Hence, the condition (I.1.2.3) is met and the nearest neighbour contact process is a well-defined Markov process. Note that the rescaling $\frac{1}{D}$ of the branching rate ensures that for each site $i \in \Lambda$ the maps \mathbf{bra}_{ij} , $\langle i, j \rangle \in E$ are applied at a total rate α .

The contact process is in a sense a simplified version of the cooperative branching process which we will define in the next subsection.

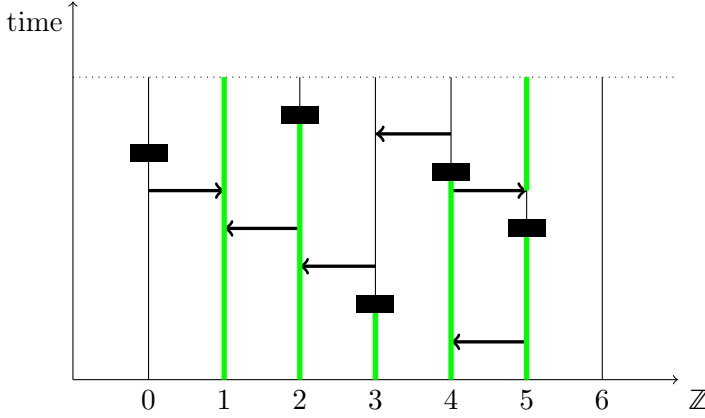


Figure I.1.1: A graphical representation of a contact process on \mathbb{Z} started with sites $i = 1, \dots, 5$ occupied. The arrows and dots denote a particular realization of the Poisson processes associated with the maps in \mathcal{G} . A left arrow from site i to site $j = i - 1$ represents a point of the Poisson process associated with the map \mathbf{bra}_{ij} and analogously for the right arrows. A black rectangle at the site i denotes a point of the Poisson process associated with the death map \mathbf{death}_i . The green lines denote the sites which are occupied at a given time.

I.1.2.3 Cooperative branching process

Let Λ be a countable set and once again let $S := \{0, 1\}$, so that S^Λ denotes the set of all particle configurations $x = (x(i))_{i \in \Lambda}$ with $x(i) \in \{0, 1\}$ for all $i \in \Lambda$. As before we will refer to the sites $i \in \Lambda$ for which $x(i) = 1$, as occupied in the configuration x and as empty otherwise. For each $i, j, k \in \Lambda$ with $i \neq j \neq k \neq i$, we define cooperative branching maps $\mathbf{coop}_{ijk} : S^\Lambda \rightarrow S^\Lambda$ by

$$\mathbf{coop}_{ijk}(x)(l) := \begin{cases} (x(i) \wedge x(j)) \vee x(k) & \text{if } l = k, \\ x(l) & \text{otherwise,} \end{cases} \quad (\text{I.1.2.12})$$

In words, the map \mathbf{coop}_{ijk} has the effect that if in the configuration x , both i and j are occupied by a particle, and k is empty, then the two particles together produce a new particle at k . Apart from the cooperative branching maps, we also consider the death maps \mathbf{death}_i , $i \in \Lambda$ as defined in (I.1.2.9) and the coalescing random walk maps \mathbf{rw}_{ij} , $i, j \in \Lambda$ defined as

$$\mathbf{rw}_{ij}(x)(k) := \begin{cases} 0 & \text{if } k = i, \\ x(i) \vee x(j) & \text{if } k = j, \\ x(k) & \text{otherwise.} \end{cases} \quad (\text{I.1.2.13})$$

In words, if in a configuration x the site i is occupied the map \mathbf{rw}_{ij} causes the particle occupying that site to "move" to site j , so that i becomes unoccupied and j becomes occupied. If j were already occupied in x before applying the map \mathbf{rw}_{ij} the two particles coalesce into one. We will be interested in the interacting particle system $(X_t)_{t \geq 0}$ with state space S^Λ and generator of the form

$$\begin{aligned} Gf(x) := & \sum_{ijk} c(i, j, k) \{f(\mathbf{coop}_{ijk}(x)) - f(x)\} \\ & + \sum_{ij} r(i, j) \{f(\mathbf{rw}_{ij}(x)) - f(x)\} \\ & + \sum_i d(i) \{f(\mathbf{death}_i(x)) - f(x)\}, \end{aligned} \quad (\text{I.1.2.14})$$

where $c(i, j, k)$, $r(i, j)$ and $d(i)$ are nonnegative rates and the first sum runs over all triples of sites i, j, k such that $i \neq j \neq k \neq i$, whereas the second runs over all pairs of sites i, j such that $i \neq j$. Under the summability conditions

$$\sup_i \sum_{j,k} c(i, j, k) < \infty, \quad \sup_i \sum_j r(i, j) < \infty \quad \text{and} \quad \sup_i d(i) < \infty, \quad (\text{I.1.2.15})$$

the cooperative branching, death and coalescing random walk maps all fulfil the condition (I.1.2.3). In this thesis we will concentrate on a particular choice of the rates $c(i, j, k)$, $r(i, j)$ and $d(i)$ which give us the following model as special case of (I.1.2.14) in the same way we obtained models (I.1.2.8) and (I.1.2.11) from the general models of the voter model and contact process, respectively.

Once again, we assume that (Λ, E) is a countable, connected, vertex transitive, locally finite graph of degree D . We now further assume that Λ has at least three elements which by connectedness implies that $D \geq 2$. Let $\alpha \geq 0, \beta \geq 0, \gamma \geq 0, \delta \geq 0$ all be non-negative constants and put

$$\begin{aligned} Gf(x) := & \alpha \frac{1}{D(D-1)} \sum_{\substack{i,j,k \in \Lambda \\ \langle i,j \rangle, \langle j,k \rangle \in E}} \{f(\text{coop}_{ikj}(x)) - f(x)\} \\ & + \beta \frac{1}{2(D-1)} \sum_{\substack{i,j,k \in \Lambda \\ \langle i,j \rangle, \langle j,k \rangle \in E}} \{f(\text{coop}_{ijk}(x)) - f(x)\} \\ & + \gamma \frac{1}{D} \sum_{\substack{i,j \in \Lambda \\ \langle i,j \rangle \in E}} \{f(\text{rw}_{ij}(x)) - f(x)\} \\ & + \delta \sum_{i \in \Lambda} \{f(\text{death}_i(x)) - f(x)\}. \end{aligned} \quad (\text{I.1.2.16})$$

In words, this can be described as follows. For each site $j \in \Lambda$, with rate α , two neighbouring vertices i, k are selected uniformly without replacement from all neighbouring sites of j , and the map coop_{ikj} is applied. For each (undirected) edge $\langle i, j \rangle \in E$, with rate β , one of the vertices neighbouring the edge $\langle i, j \rangle$ is selected at random with probability $\frac{1}{2(D-1)}$ if it is a neighbour of i but not of j (or vice versa) and with probability $\frac{1}{(D-1)}$ if it neighbours both i and j . Then, denoting this vertex by k , the map coop_{ijk} is applied. Finally, each particle jumps with rate γ to a random neighbouring site, coalescing with any particle that may already have been present there, and each particle dies at rate δ . All maps used in the construction of the cooperative branching process are local and by the choice of the edge set E we see that for any $i \in \Lambda$ the set $\{m \in \mathcal{G}; \mathcal{D}(m) \ni i\}$ is finite. Hence the condition (I.1.2.3) is met and the cooperative branching process is a well-defined Markov process. Note that the choice of rescaling of the rates α ensures that for each $j \in \Lambda$ the total rate at which the maps coop_{ikj} , $\langle i, j \rangle, \langle j, k \rangle \in E$ are applied is α . For β and γ the rescaling gives us analogous statements.

In figure I.1.2 we provide an example of the graphical representation of the cooperative branching process with generator (I.1.2.16) on \mathbb{Z} with $\alpha > 0, \delta > 0$ and $\beta = \gamma = 0$.

In this thesis we will be studying the model (I.1.2.16) on three particular graphs - the complete graph with $N \in \mathbb{N}$ vertices in Chapter I.2, the regular tree \mathbb{T}_d , $d \geq 2$ in Section I.3.1 and finally in sections I.3.2.1 and I.3.2.2 we study the model on an integer lattice \mathbb{Z}^d , $d \geq 1$ with nearest-neighbour edges where $D = 2d$.

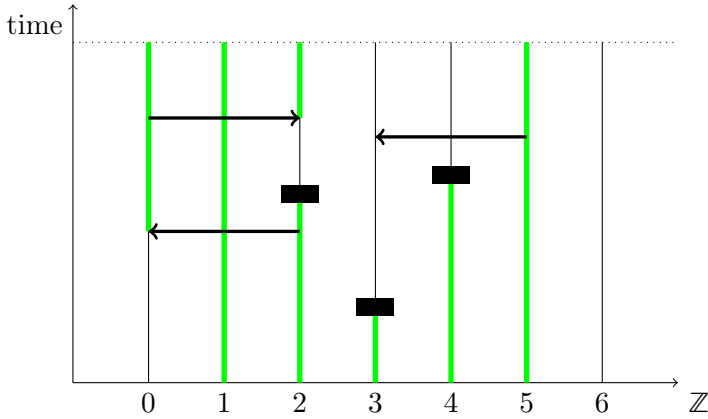


Figure I.1.2: A left arrow from site i over site $j = i - 1$ to site $k = i - 2$ represents a point of the Poisson process associated with the map bra_{ijk} and analogously for the right arrows. Once again, black rectangles represent deaths.

I.1.3 Phase transition

Put $\mathcal{S} = S^\Lambda$, where S is finite and Λ countable let μ and ν be two probability measures on the partially ordered space \mathcal{S} such that a μ -distributed random variable X and a ν -distributed random variable Y can be coupled so that $X \leq Y$. Then we say that μ and ν are stochastically ordered and write $\mu \leq \nu$. It holds that $\mu \leq \nu$ if and only if $\int f(x)d\mu(x) \leq \int f(x)d\nu(x)$ for any monotone bounded measurable function f on \mathcal{S} . For the proof of the equivalence, see for example Theorem 5.1 in [Swa17] as well as Theorem II.2.4 in [Lig85] which [Swa17] refers to for a part of the proof.

Let \mathcal{T} be a σ -algebra on \mathcal{S} . Then we say that a map $K : \mathcal{S} \times \mathcal{T} \rightarrow [0, 1]$ is a probability kernel on \mathcal{S} if the map $x \rightarrow K(x, A)$ is measurable for every measurable set A and the map $A \rightarrow K(x, A)$ is a probability measure on \mathcal{S} for every $x \in \mathcal{S}$. We will say that a probability kernel K on \mathcal{S} is *monotone* if $K(x, \cdot) \leq K(y, \cdot)$ for all $x \leq y \in \mathcal{S}$. A particle interacting system is *monotone* if its transition probability kernels are monotone. The contact process, the voter model as well as the cooperative branching process are all monotone since they can all be represented using monotone maps only (see (A.2.5) for the definition of a monotone map).

We will now assume that $\mathcal{S} = \{0, 1\}^\Lambda$, where Λ is a countable lattice and concentrate on the case when the particle interacting system in question is either the contact process with generator (I.1.2.11) or the cooperative branching process with generator (I.1.2.16). We will refer to the process as $X = (X_t)_{t \geq 0}$. Denote by $1 := 1_\Lambda := \{1\}^\Lambda \in \mathcal{S}$ the configuration in which all sites in Λ are occupied and similarly by $0 := 0_\Lambda := \{0\}^\Lambda \in \mathcal{S}$ the configuration in which every site is empty. Let $(P_t, t \geq 0)$ be the transition kernel of X . We will say that an invariant law ν of X is the *upper invariant law* if $1P_t := \mathbb{P}^1[X_t \in \cdot]$ converges weakly to ν . In other words ν is the upper invariant law of the process X if the distribution of X started in the initial state where all sites are occupied converges weakly to ν . The existence of the upper invariant law is ensured for any monotone interacting particle system (see for example [Lig85], Theorem III.2.3¹). The name *upper* invariant comes from the fact that if ν is an upper invariant law of X , then we have for any invariant law μ of X that $\mu \leq \nu$, where the inequality is in the sense of stochastic order.

¹The theorem is formulated for monotone spin systems only, but its proof remains the same for all monotone interacting particle systems

We will say that

$$\text{a probability law } \nu \text{ on } \mathcal{S} \text{ is non-trivial if } \nu(\mathbf{0}) = 0, \quad (\text{I.1.3.1})$$

that is the empty configuration has probability zero. Given a monotone interacting particle system we can then ask whether the upper invariant law is non-trivial.

For $x \in \mathcal{S}$ we denote by $|x| := \#\{i; x(i) = 1\}$ the number of occupied sites in x . It is clear that the configuration $\mathbf{0}$ is absorbing for both the contact process and the cooperative branching process, that is the process remains constant if there are no occupied sites left. This leads us to studying its survival. We say that the process X survives if there exists a configuration x_0 , $|x_0| < \infty$ such that if $X_0 = x_0$,

$$\mathbb{P}[X_t \neq \mathbf{0} \forall t \geq 0] > 0. \quad (\text{I.1.3.2})$$

That is, we say that the process survives if, when started from a finite configuration (i.e. one with finitely many occupied sites), there is a positive probability that there exist some occupied sites at all times. Otherwise, we say that the process dies out.

In case of the cooperative branching process with random walk dynamics but without deaths (i.e. in the model (I.1.2.16) with $\gamma > 0$ and $\delta = 0$) the process always survives in the sense (I.1.3.2) since the process with a single occupied particle behaves like a random walk. However, we can still define a similar concept. We say that the cooperative branching process *stays active* if there exists an initial state x_0 , $|x_0| < \infty$ such that

$$\mathbb{P}[\forall t \geq 0 \exists s > t \text{ s.t. } |X_s| \neq |X_t|] > 0. \quad (\text{I.1.3.3})$$

In other words, the process stays active if the number of occupied states keeps changing. If (I.1.3.3) does not hold, we say that the process becomes inactive. Note that in the cooperative branching process with deaths (I.1.3.3) and (I.1.3.2) are equivalent.

The non-triviality of the upper invariant law, the survival of the contact process and the cooperative branching process and staying active in the cooperative branching process without deaths all depend on the choice of the underlying lattice Λ as well as the values of the rates α, β, γ and δ . In the cooperative branching process, for fixed values of the rates α, γ, δ , we let

$$\begin{aligned} \beta_{\text{surv}}(\alpha, \gamma, \delta) &= \beta_{\text{surv}} := \inf\{\beta > 0 : \text{the process survives}\}, \\ \beta_{\text{upp}}(\alpha, \gamma, \delta) &= \beta_{\text{upp}} := \inf\{\beta > 0 : \text{the upper invariant law is nontrivial}\}. \end{aligned} \quad (\text{I.1.3.4})$$

In case that $\delta = 0$, we also define

$$\beta_{\text{act}}(\alpha, \gamma) = \beta_{\text{act}} := \inf\{\beta > 0 : \text{the process stays active}\}. \quad (\text{I.1.3.5})$$

We define the rates $\alpha_{\text{surv}}(\beta, \gamma) = \alpha_{\text{surv}}$, $\alpha_{\text{upp}}(\beta, \gamma) = \alpha_{\text{upp}}$ and (in the case $\delta = 0$) the rate $\alpha_{\text{act}}(\beta, \gamma) = \alpha_{\text{act}} := \inf\{\alpha > 0 : \text{the process stays active}\}$ analogously. In the same way we also define the rates $\alpha_{\text{surv}}(\delta) = \alpha_{\text{surv}}$ and $\alpha_{\text{supp}}(\delta) = \alpha_{\text{supp}}$ for the contact process.

Both the cooperative branching process and the contact process are monotone which allows us to use monotone coupling of the respective processes with different rates β to show that they survive for all $\beta > \beta_{\text{surv}}$ and die out for $\beta < \beta_{\text{surv}}$ and their respective upper invariant laws are nontrivial for all $\beta > \beta_{\text{upp}}$ and trivial for $\beta < \beta_{\text{upp}}$ (for an example of such coupling, see the proof of Proposition 5.11. in [Swa17]). By monotone coupling we can also show analogous statements for α_{surv} and α_{upp} . It is also possible to show this for β_{act} and α_{act} (a detailed

proof in the case when Λ is a one-dimensional integer lattice can be found in [SS15], Theorem 3). This is an example of what is usually called phase transition, i.e. a difference of the long term behaviour if a model under and below a certain critical level of a parameter of the model. In the view of that, we call $\alpha_{\text{surv}}, \alpha_{\text{upp}}, \alpha_{\text{act}}, \beta_{\text{surv}}, \beta_{\text{upp}}, \beta_{\text{act}}$ the critical parameters so that for example α_{surv} or β_{surv} are the critical parameters for survival and α_{upp} and β_{upp} the critical parameters for the nontriviality of the upper invariant law. Unless it is clear from the context, we will specify in each case which of the parameters $\alpha, \beta, \gamma, \delta$ we are assuming to be fixed.

In particular, the contact process on \mathbb{Z}^d with death rate $\delta = 1$ survives and its upper invariant law is non-trivial when $\alpha > \alpha(d)$, where $\alpha(d)$ is a constant dependent on d and the process dies out (and there exists no non-trivial invariant law) when $\alpha \leq \alpha(d)$. This can be shown by a comparison to the oriented percolation, see for example [Swa17], Chapter 7. Bezuidenhout and Grimmet ([BG90]) have further shown that in the critical case $\alpha = \alpha_{\text{surv}}$ the contact process dies out. Bezuidenhout and Gray ([BG94]) then expanded on those results to give analogous statements about more general models. In Section I.3.2.1 we will show that their methods can be used in a straightforward manner to show the following.

Proposition I.1.3.1 *Let X be a cooperative branching process with the generator (I.1.2.16) defined on \mathbb{Z}^d equipped with nearest neighbour edges and assume that $\gamma = 0$ (no random walks) and $\delta > 0$. Then*

$$\beta_{\text{upp}} \leq \beta_{\text{surv}},$$

that is the critical parameter for the non-triviality of the upper invariant law is smaller or equal to the critical parameter for survival.

It can be shown for the contact process (see for example [Lig99], Part I.1.) that the corresponding critical parameters α_{upp} and α_{surv} are in fact equal which together with Proposition I.1.3.1 suggests that this might also be the case of the cooperative branching process on \mathbb{Z}^d with no random walks. For the process on a regular tree \mathbb{T}_d we obtain the following estimate of the critical parameter for survival in Section I.3.1.

Proposition I.1.3.2 *Let X be the process with generator (I.1.2.16) defined on a regular tree \mathbb{T}_d with $\alpha = \gamma = 0$ and $\delta = 1$. Then*

$$\beta_{\text{surv}} \geq \frac{d}{d-1}. \tag{I.1.3.6}$$

Sturm and Swart [SS15] studied the cooperative branching process with the generator (I.1.2.16) on \mathbb{Z} in the special case when $\alpha = \delta = 0$ and $\gamma = 1$. In this case, rather than survival as in (I.1.3.2) we are interested in whether the process stays active in the sense (I.1.3.3). Sturm and Swart showed that $\beta_{\text{act}} \in [1, \infty)$ and $\beta_{\text{upp}} \in [1, \infty)$ and that for $\beta < \frac{1}{2}$ when started from an initial state $X_0 = x_{\text{pair}} := \mathbf{1}_{\{0,1\}}$ with only two neighbouring occupied particles, the probability $\mathbb{P}^{x_{\text{pair}}} [|X_t| \geq 2]$ of staying active² until time t has a polynomial decay (to zero). Under the same conditions they also showed that the same holds for the probability $\mathbb{P}^{\mathbf{1}_{\mathbb{Z}}}[X_t(0) = 1]$ of a particular site being occupied at time t with the initial state $X_0 = \mathbf{1}_{\mathbb{Z}}$ with all sites $i \in \mathbb{Z}$ occupied. They also conjectured that this decay becomes exponential if deaths are allowed to occur in addition to random walk dynamics, i.e. if $\delta > 0$. In section I.3.2.2 we confirm their conjecture, namely we show the following

²By the recurrence of of a one-dimensional random walk this probability is indeed equivalent to the left-hand side of (I.1.3.3)

Proposition I.1.3.3 *Let X be the cooperative branching process with generator (I.1.2.16) on \mathbb{Z} (with nearest neighbour edges) and such that $\alpha = 0$, $\gamma = 1$ and $\delta > 0$ is fixed. Then for all $\beta \leq 1$ it holds that*

$$\mathbb{P}^{\text{xpair}}[|X_t| \geq 1] \leq 2e^{-\delta t} \quad \text{and} \quad \mathbb{P}^{\mathbf{1z}}[X_t(0) = 1] \leq e^{-\delta t}. \quad (\text{I.1.3.7})$$

We also show corresponding lower bounds for these probabilities in Propositions I.3.2.2 and I.3.2.3. Some of our results as well as other results cited in this section are obtained using duality, which is a concept that we introduce in the next section.

I.1.4 Duality

We say that Markov processes $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ with state spaces S and T respectively are dual to each other if there exists a measurable function $\psi : S \times T \rightarrow \mathbb{R}$ such that

$$\mathbb{E}[\psi(X_t, Y_0)] = \mathbb{E}[\psi(X_0, Y_t)], \quad t \geq 0 \quad (\text{I.1.4.1})$$

whenever the initial state X_0 of X is independent of Y and likewise the initial state Y_0 of Y is independent of X . We will say that X is subdual to Y if the left-hand side of (I.1.4.1) is lower or equal than the right-hand side. This notion of duality with respect to a duality function is classical and the above definition can be found for example in [CR84].

We can also define a duality between two maps. Let S and T be sets and $m : S \rightarrow S$ and $\hat{m} : T \rightarrow T$ functions. We then say that m and \hat{m} are dual with respect to (the duality function) ψ if

$$\psi(m(x), y) = \psi(x, \hat{m}(y)), \quad x \in S, y \in T. \quad (\text{I.1.4.2})$$

For processes which can be constructed from their corresponding stochastic flows as in (I.1.2.5), we can also define a stronger version of duality as follows. Let $(\mathbf{X}_{s,t})_{s \leq t}$ and $(\mathbf{Y}_{s,t})_{s \leq t}$ be stochastic flows with independent increments such that $\mathbf{X}_{s,t} : T_1 \rightarrow T_1$ and $\mathbf{Y}_{s,t} : T_2 \rightarrow T_2$ for all $s, t \in \mathbb{R}$ where T_1 and T_2 are some metrizable spaces. We say that $(\mathbf{X}_{s,t})_{s \leq t}$ and $(\mathbf{Y}_{s,t})_{s \leq t}$ are dual to each other with respect to the duality function $\psi : T_1 \times T_2 \rightarrow \mathbb{R}$ if

- (i) $(\mathbf{X}_{t_0, t_1}, \mathbf{Y}_{-t_1, -t_0}), \dots, (\mathbf{X}_{t_{n-1}, t_n}, \mathbf{Y}_{-t_n, -t_{n-1}})$ are independent for any $t_0 < \dots < t_n$.
- (ii) For each $x \in T_1$, $y \in T_2$ and $s \leq u$, the function $[s, u] \ni t \rightarrow \psi(\mathbf{X}_{s,t}, \mathbf{Y}_{-u,-t}(y))$ is almost surely constant.

If the processes X and Y are constructed via stochastic flows $(\mathbf{X}_{s,t})_{s \leq t}$ and $(\mathbf{Y}_{s,t})_{s \leq t}$, respectively, which are dual to each other, then we say that X and Y are pathwise dual. As noted above, this is a stronger notion of duality than (I.1.4.1), since in particular we obtain from the duality of their stochastic flows that for any $x \in T_1$ and $y \in T_2$ it holds that

$$\psi(x, \mathbf{Y}_{-u,-s}(y)) = \psi(\mathbf{X}_{s,u}(x), y) \quad \text{almost surely,}$$

which implies that the processes X and Y are also dual in the sense (I.1.4.1). The term pathwise duality with respect to a function has been first introduced by Jansen and Kurt in [JK14]. The same article also includes a survey of various other notions of duality.

We will now construct a process which is pathwise dual to the cooperative branching process X with generator (I.1.2.14) without random walks. This includes all variations of the cooperative branching process that we study in this thesis except for Section I.3.2.2 where we

consider the random walk dynamics as well. Assume that X does not exhibit any random walk dynamics, that is $r(i, j) = 0$ for all $i, j \in \Lambda$. Let $\mathcal{S}_{\text{fin}} := \{y \in \mathcal{S}(\Lambda) : \sum_i y(i) < \infty\}$ denote the space of all finite particle configurations on Λ and denote by

$$\mathcal{H}(\Lambda) := \mathcal{P}_{\text{fin}}(\mathcal{S}_{\text{fin}}(\Lambda)) := \{E \subset \mathcal{S}_{\text{fin}} : |E| < \infty\}$$

the set of all finite subsets of $\mathcal{S}_{\text{fin}}(\Lambda)$. Let $\mathbf{bran}_{kij} : \mathcal{S}(\Lambda) \rightarrow \mathcal{S}(\Lambda)$ be the map defined as

$$\mathbf{bran}_{kij}(e)(l) := \begin{cases} e(k) \vee e(l) & \text{if } l = i \text{ or } l = j, \\ 0 & \text{if } l = k, \\ e(l) & \text{otherwise,} \end{cases} \quad (\text{I.1.4.3})$$

which can be described in words by saying that if in the configuration e there is a particle at site k , then this particle disappears but two new particles appear at the sites i and j , provided that these are empty. Next, for any $E \in \mathcal{H}(\Lambda)$, let $\mathbf{bran}_{kij}(E)$ denote the image of E under the map \mathbf{bran}_{kij} and define maps acting on the space $\mathcal{H}(\Lambda)$ by

$$\mathbf{coop}_{ijk}^\bullet(E) := E \cup \mathbf{bran}_{kij}(E). \quad \text{and} \quad \mathbf{death}_i^\bullet(E) := \{e \in E : e(i) = 0\}. \quad (\text{I.1.4.4})$$

Let $(Y_t)_{t \geq 0}$ be the continuous-time Markov process with state space $\mathcal{H}(\Lambda)$ and generator

$$G_\bullet f(Y) := \sum_{ijk} c(i, j, k) \{f(\mathbf{coop}_{ijk}^\bullet(Y)) - f(Y)\} + \sum_i d(i) \{f(\mathbf{death}_i^\bullet(Y)) - f(Y)\}. \quad (\text{I.1.4.5})$$

Note that $\mathcal{H}(\Lambda)$ is countable, and hence $(Y_t)_{t \geq 0}$ is a continuous-time Markov chain. It follows from the Proposition 28 of [SS16] that under the summability conditions (I.1.2.15), the process $(Y_t)_{t \geq 0}$ is non-explosive.

The following result is an immediate consequence of Theorem A.2.2, where we choose the dual (A.2.1) as $S' := S$ but equipped with the reversed order and the bijection $x \rightarrow x'$ chosen as the identity function.

Proposition I.1.4.1 (Cooperative branching duality, no random walks) *Let $\chi : \mathcal{S}(\Lambda) \times \mathcal{H}(\Lambda) \rightarrow \{0, 1\}$ be defined by*

$$\chi(b, E) := 1_{\{\exists e \in E \text{ s.t. } e \leq b\}} \quad (b \in \mathcal{S}(\Lambda), E \in \mathcal{H}(\Lambda)). \quad (\text{I.1.4.6})$$

Let $X = (X_t)_{t \geq 0}$ be an interacting particle system with generator as in (I.1.2.14) with $r(i, j) = 0$ for all i, j and let $Y = (Y_t)_{t \geq 0}$ be the continuous-time Markov process with generator as in (I.1.4.5). Then X and Y can be coupled in such a way that they are pathwise dual.

The process $(Y_t)_{t \geq 0}$ in Proposition I.1.4.1 is very similar to the process $(Y_t^\bullet)_{t \geq 0}$ from Theorem A.2.2, but not quite the same. The difference is that according to our definition $\mathbf{coop}_{ijk}^\bullet(\{111\}) = \{111, 110\}$, whereas (A.2.10) gives $\mathbf{coop}_{ijk}^\bullet(\{111\}) = \{110\}$. Here 110 is a short notation for the particle configuration for which $(x(i), x(j), x(k)) = (1, 1, 0)$. Since $\chi(x, \{111, 110\}) = \chi(x, \{110\})$ for any $x \in \mathcal{S}(\Lambda)$, this has no effect on the duality.

Chapter I.2

Complete graph and the Mean-field Model

I.2.1 Introduction and main results

I.2.1.1 The process on the complete graph

In this subsection, we concentrate on the case where $\Lambda = \Lambda_N := \{1, \dots, N\}$ is a finite set with $N \in \mathbb{N}$, $N \geq 3$ elements and the rates $c(i, j, k)$ and $d(i)$ from (I.1.2.14) are given by

$$c(i, j, k) := \begin{cases} \frac{\alpha}{(N-1)(N-2)} & \text{if } i \neq j \neq k \neq i, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad d(i) := \delta, \quad (\text{I.2.1.1})$$

where $\alpha, \delta \geq 0$ are fixed constants. For this particular choice of rates, the dynamics of $(X_t)_{t \geq 0}$ can be described as follows. For each site $k \in \Lambda_N$, with rate α , two sites i, j are drawn at random such that i, j, k are all different, and the cooperative branching map coop_{ijk} is applied. Moreover, for each $i \in \Lambda$, with rate δ , the map death_i is applied. Since all sites play an equal role, we can think of each site as being a neighbour of each other site, i.e., we picture Λ_N as a complete graph with N vertices.

From now on, unless stated otherwise, we will assume that the death rate δ is equal to 1. Note that by time rescaling any positive $\delta > 0$ can be reduced to this case.

Let $(X_t^N)_{t \geq 0}$ denote the cooperative branching process on Λ_N with generator as in (I.1.2.14) and rates as in (I.2.1.1). Since all sites play an equal role, the fraction of sites that is occupied by a particle

$$\bar{X}_t^N := \frac{1}{N} \sum_{i \in \Lambda_N} X_t^N(i), \quad t \geq 0 \quad (\text{I.2.1.2})$$

is a Markov process with state space $\{0, \frac{1}{N}, \dots, 1\}$ that jumps from

$$\begin{aligned} \bar{x} &\mapsto \bar{x} + \frac{1}{N} & \text{at rate} & \quad \alpha N(1 - \bar{x})\bar{x}(\bar{x} - \frac{1}{N})\frac{N-1}{N-1}\frac{N-2}{N-2}, \\ \bar{x} &\mapsto \bar{x} - \frac{1}{N} & \text{at rate} & \quad N\bar{x}. \end{aligned} \quad (\text{I.2.1.3})$$

Here $N(1 - \bar{x})(N - 1)\bar{x}(N - 2)(\bar{x} - \frac{1}{N})$ is the number of ordered triples of sites such that the first site is empty and the other two are occupied.

The process \bar{X}_t^N contains all information about X_t^N modulo permutations of the sites. Similarly, for the dual process $(Y_t^N)_{t \geq 0}$ of $(X_t^N)_{t \geq 0}$, we will also use the fact that all sites in

Λ_N play an equal role to define a Markov process $(\bar{Y}_t^N)_{t \geq 0}$ such that \bar{Y}_t^N keeps track of the information in Y_t^N modulo permutations of the sites. We will describe the process Y_t^N using hypergraphs.

By definition, a *hypergraph* is a pair (V, E) where V is a set whose elements are called *vertices* and E is a subset of the set $\mathcal{P}(V)$ of all subsets of V . Elements $e \in E$ are called *hyperedges*. A hypergraph is *finite* if V and hence also E are finite sets. An *isolated vertex* is an element $v \in V$ such that $v \notin e$ for all $e \in E$. A hypergraph without isolated vertices is uniquely characterized by its edge set E . Two hypergraphs $H = (V, E)$ and $H' = (V', E')$ are *isomorphic* if there exists a bijection $\pi : V \rightarrow V'$ such that $E' = \{\pi(e) : e \in E\}$, where $\pi(e) := \{\pi(i) : i \in e\}$ denotes the image of the set of $e \subset V$ under the bijection π .

For any set Λ , identifying subsets with their indicator functions, there is a natural isomorphism $\mathcal{P}_{\text{fin}}(\Lambda) \cong \mathcal{S}_{\text{fin}}$ and hence we can identify $\mathcal{H}(\Lambda)$ with the set

$$\mathcal{P}_{\text{fin}}(\mathcal{P}_{\text{fin}}(\Lambda)) \cong \mathcal{H}(\Lambda) = \mathcal{P}_{\text{fin}}(\mathcal{S}_{\text{fin}}(\Lambda)). \quad (\text{I.2.1.4})$$

Each $E \in \mathcal{H}(\Lambda)$ defines a finite hypergraph without isolated vertices, with the edge set E and the vertex set

$$V_E := \bigcup E := \{i \in \mathbb{N} : \exists e \in E \text{ s.t. } i \in e\}. \quad (\text{I.2.1.5})$$

Clearly each finite hypergraph without isolated vertices is isomorphic to an element of $\mathcal{H}(\mathbb{N})$. We define an equivalence relation \sim on $\mathcal{H}(\mathbb{N})$ by

$$E \sim E' \quad \text{iff} \quad E \text{ is isomorphic to } E'. \quad (\text{I.2.1.6})$$

Now, let $\bar{E} := \{E' \in \mathcal{H} : E' \sim E\}$ denote the equivalence class containing E and let

$$\bar{\mathcal{H}} := \{\bar{E} : E \in \mathcal{H}(\mathbb{N})\} \quad (\text{I.2.1.7})$$

denote the set of all equivalence classes. We call $\bar{\mathcal{H}}$ the space of all hypergraphs without isolated vertices *modulo isomorphisms*. Set

$$\bar{\mathcal{H}}_N := \bar{\mathcal{H}}(\Lambda_N) = \{\bar{E} \in \bar{\mathcal{H}} : E \in \mathcal{H}(\Lambda_N)\} \quad \text{with} \quad \Lambda_N = \{1, \dots, N\}. \quad (\text{I.2.1.8})$$

Then we can describe $\bar{\mathcal{H}}_N$ as the space of all hypergraphs that have at most N vertices and no isolated vertices, modulo isomorphisms.

Let $(Y_t^N)_{t \geq 0}$ denote the dual process of $(X_t^N)_{t \geq 0}$, which has the generator in (I.1.4.5) with rates of the form (I.2.1.1). Then Y_t^N takes values in the space $\mathcal{H}(\Lambda_N)$. From now on, we will often view Y_t^N as a hypergraph without isolated vertices, i.e., as a set of subsets of Λ_N . In our present language of hypergraphs,

$$\text{coop}_{ijk}^\bullet(E) = E \cup \text{bran}_{kij}(E) \quad \text{and} \quad \text{death}_k^\bullet(E) = \{e \in E : k \notin e\}, \quad (\text{I.2.1.9})$$

where $\text{bran}_{kij}(E)$ denotes the image of $E \in \mathcal{H}(\Lambda_N)$ under the map bran_{kij} , and

$$\text{bran}_{kij}(e) = \begin{cases} (e \setminus \{k\}) \cup \{i, j\} & \text{if } k \in e, \\ e & \text{otherwise.} \end{cases} \quad (\text{I.2.1.10})$$

Hence, $(Y_t^N)_{t \geq 0}$ is a continuous-time Markov chain that jumps from a state $E \in \mathcal{H}(\Lambda_N)$ to another state $F \in \mathcal{H}(\Lambda_N)$ at rate

$$r_N(E, F) = \frac{\alpha}{(N-1)(N-2)} \sum_{ijk} 1_{\{\text{coop}_{ijk}^\bullet(E) = F\}} + \sum_k 1_{\{\text{death}_k^\bullet(E) = F\}}, \quad (\text{I.2.1.11})$$

where the first sum runs over all ordered triples $(i, j, k) \in \{1, \dots, N\}^3$ such that $i \neq j \neq k \neq i$, and the second sum runs over all $k \in \{1, \dots, N\}$.

For any $E \in \mathcal{H}(\Lambda_N)$, we let $\bar{E} \in \bar{\mathcal{H}}_N$ denote the equivalence class that contains E . Then $(\bar{Y}_t^N)_{t \geq 0}$ is a process taking values in the space $\bar{\mathcal{H}}_N$ defined in (I.2.1.8). It follows from (I.2.1.11) that the process $(\bar{Y}_t^N)_{t \geq 0}$ jumps from a state $\bar{E} \in \bar{\mathcal{H}}_N$ to another state $\bar{F} \in \bar{\mathcal{H}}_N$ with rate

$$\bar{r}_N(\bar{E}, \bar{F}) := \frac{\alpha}{(N-1)(N-2)} \sum_{ijk} 1_{\{\overline{\text{coop}}_{ijk}^\bullet(E) = \bar{F}\}} + \sum_k 1_{\{\overline{\text{death}}_k^\bullet(E) = \bar{F}\}}, \quad (\text{I.2.1.12})$$

where we have used implicitly that since all sites play an equal role, these rates depend only on the equivalence classes $\bar{E}, \bar{F} \in \bar{\mathcal{H}}_N$ and not on the choice of the representatives $E, F \in \mathcal{H}(\Lambda_N)$.

The duality of Proposition I.1.4.1 gives rise to a duality between the processes $(\bar{X}_t^N)_{t \geq 0}$ and $(\bar{Y}_t^N)_{t \geq 0}$ which we describe now. For any countable set Λ , $b \in \mathcal{S}(\Lambda)$, and $E \in \mathcal{H}(\Lambda)$, we define $\text{Thin}_b(E) \in \mathcal{H}(\Lambda)$ by

$$\text{Thin}_b(E) := \{e \in E : b(i) = 1 \forall i \in e\}. \quad (\text{I.2.1.13})$$

We can think of sites $i \in \{1, \dots, N\}$ such that $b(i) = 1$ (resp. $b(i) = 0$) as being open (resp. closed). Then the effect of the map Thin_b is to throw away all hyperedges that contain a closed vertex. In terms of thinning, the duality function χ from (I.1.4.6) can be written as

$$\chi(b, E) = 1_{\{\text{Thin}_b(E) \neq \emptyset\}} \quad (b \in \mathcal{S}(\Lambda), E \in \mathcal{H}(\Lambda)), \quad (\text{I.2.1.14})$$

where \emptyset denotes the empty hypergraph whose edge and vertex sets are empty.

For $p \in \{0, \frac{1}{N}, \dots, 1\}$, let B_p^N be a random variable that is uniformly distributed on the set $\{b \in \{0, 1\}^N, \text{ s.t. } \bar{b} = p\}$, where $\bar{b} := N^{-1} \sum_{i=1}^N b(i)$. We also assume that B_p^N is independent of X^N for each $N \in \mathbb{N}$ and $p \in \{0, \frac{1}{N}, \dots, 1\}$. We define a function $\bar{\phi}^N : \{0, \frac{1}{N}, \dots, 1\} \times \bar{\mathcal{H}}_N \rightarrow [0, 1]$ by

$$\bar{\phi}^N(p, \bar{E}) := \mathbb{P}[\text{Thin}_{B_p^N}(E) \neq \emptyset], \quad (\text{I.2.1.15})$$

where we have implicitly used that the right-hand side depends only on the equivalence class $\bar{E} \in \bar{\mathcal{H}}$ and not on the choice of the representative $E \in \mathcal{H}$.

Proposition I.1.4.1 implies that the processes $(\bar{X}_t^N)_{t \geq 0}$ and $(\bar{Y}_t^N)_{t \geq 0}$ are dual in the following sense.

Proposition I.2.1.1 (Duality on the complete graph) *Let $(\bar{X}_t^N)_{t \geq 0}$ and $(\bar{Y}_t^N)_{t \geq 0}$ be Markov processes with jump rates as in (I.2.1.3) and (I.2.1.12). Then, assuming that \bar{X}_0^N is independent of \bar{Y}_t^N and \bar{X}_t^N is independent of \bar{Y}_0^N , we have that*

$$\mathbb{E}[\bar{\phi}^N(\bar{X}_t^N, \bar{Y}_0^N)] = \mathbb{E}[\bar{\phi}^N(\bar{X}_0^N, \bar{Y}_t^N)]. \quad (\text{I.2.1.16})$$

The proof of Proposition I.2.1.1 can be found in Subsection I.2.6.1.

I.2.1.2 The mean-field limit

We are interested in the process $(\bar{X}_t^N)_{t \geq 0}$ from (I.2.1.3) and its dual $(\bar{Y}_t^N)_{t \geq 0}$ with jump rates as in (I.2.1.12) in the mean-field limit $N \rightarrow \infty$. As $N \rightarrow \infty$, the process $(\bar{X}_t^N)_{t \geq 0}$ converges to a solution of an ordinary differential equation.

Proposition I.2.1.2 (Mean-field forward process) *Let $(\bar{X}_t^N)_{t \geq 0}$ be Markov processes as in (I.2.1.3) started in deterministic initial states \bar{X}_0^N that converge in probability as $N \rightarrow \infty$ to some deterministic $u_0 \in [0, 1]$. Let $(u_t)_{t \geq 0}$ denote the solution of the ODE*

$$\frac{\partial}{\partial t} u_t = \alpha(1 - u_t)u_t^2 - \delta u_t, \quad t \geq 0 \quad (\text{I.2.1.17})$$

with initial state u_0 . Then, for each $T < \infty$ and $\varepsilon > 0$,

$$\mathbb{P}[|\bar{X}_t^N - u_t| \leq \varepsilon \ \forall t \in [0, T]] \xrightarrow{N \rightarrow \infty} 1. \quad (\text{I.2.1.18})$$

The proof of Proposition I.2.1.2 can be found in Subsection I.2.6.2.

We next set out to describe the mean-field limit of the process $(\bar{Y}_t^N)_{t \geq 0}$ from (I.2.1.12), which is our main object of interest. The main idea is easily explained. We first observe that for $E \in \mathcal{H}_{\Lambda_N}$ we have $\text{coop}_{ijk}^\bullet(E) = E$ and $\text{death}_k^\bullet(E) = E$ if k is not a vertex of E . In view of this, formula (I.2.1.11) can be described in words as follows. For each vertex k of E , with rate α , two sites $i, j \in \{1, \dots, N\}$ are drawn at random in such a way that $i \neq j \neq k \neq i$, and the map $\text{coop}_{ijk}^\bullet$ is applied. Moreover, for each vertex k of E , with rate δ the map death_k^\bullet is applied. For large N , the probability that the random sites i, j are already vertices of E is small, so in the limit $N \rightarrow \infty$, the map $\text{coop}_{ijk}^\bullet$ always adds two new vertices to E .

To formulate this more formally, for each $E \in \mathcal{H}(\mathbb{N})$, let $V_E := \bigcup E$ denote its vertex set and choose in some arbitrary way two sites $i_E, j_E \in \mathbb{N}$ such that $i_E, j_E \notin V_E$ and $i_E \neq j_E$. Define

$$\text{coop}_k^\bullet(E) := \text{coop}_{i_E j_E k}^\bullet(E), \quad E \in \mathcal{H}, \quad (\text{I.2.1.19})$$

where $\text{coop}_{ijk}^\bullet$ is the map from (I.2.1.9). Recall from (I.2.1.7) that $\bar{\mathcal{H}}$ denotes the set of all finite hypergraphs without isolated vertices, modulo isomorphisms. Let $(\bar{Y}_t)_{t \geq 0}$ denote the continuous-time Markov chain with state space $\bar{\mathcal{H}}$ that jumps from a state $\bar{Y} \in \bar{\mathcal{H}}$ to another state $\bar{Z} \in \bar{\mathcal{H}}$ with rate

$$r(\bar{E}, \bar{F}) := \alpha \sum_{k \in V_E} 1_{\{\overline{\text{coop}_k^\bullet(E)} = \bar{F}\}} + \sum_{k \in V_E} 1_{\{\overline{\text{death}_k^\bullet(E)} = \bar{F}\}}. \quad (\text{I.2.1.20})$$

where once again we have implicitly used that these rates depend only on the equivalence classes $\bar{E}, \bar{F} \in \bar{\mathcal{H}}$ and not on the choice of the representatives $E, F \in \mathcal{H}$. Informally, the dynamics of $(\bar{Y}_t)_{t \geq 0}$ can be described as follows. Let \bar{E} be the current state of $(\bar{Y}_t)_{t \geq 0}$. Then

- (i) For each vertex k of \bar{E} ¹, at rate α , two new vertices i, j are added to \bar{E} and for every hyperedge e of \bar{E} that contains k , a new hyperedge $e' := (e \setminus \{k\}) \cup \{i, j\}$ is added to \bar{E} .
- (ii) For each vertex k of \bar{E} , at rate 1, all hyperedges that contain k are removed from \bar{E} .

We call $(\bar{Y}_t)_{t \geq 0}$ the *mean-field dual process*.

Proposition I.2.1.3 (Mean-field limit of dual) *The mean-field dual process is nonexplosive. Moreover, if $(\bar{Y}_t^N)_{t \geq 0}$ are continuous-time Markov chains with values in $\bar{\mathcal{H}}_N$ and jump rates as in (I.2.1.12), and*

$$\mathbb{P}[\bar{Y}_0^N \in \cdot] \xrightarrow{N \rightarrow \infty} \mathbb{P}[\bar{Y}_0 \in \cdot] \quad (\text{I.2.1.21})$$

for some $\bar{\mathcal{H}}$ -valued random variable \bar{Y}_0 , then

$$\mathbb{P}[(\bar{Y}_t^N)_{0 \leq t \leq T} \in \cdot] \xrightarrow{N \rightarrow \infty} \mathbb{P}[(\bar{Y}_t)_{0 \leq t \leq T} \in \cdot], \quad 0 \leq T < \infty \quad (\text{I.2.1.22})$$

¹Or more precisely of an arbitrary representative E of the equivalence class \bar{E} .

where $(\bar{Y}_t)_{t \geq 0}$ is the mean-field dual process started in \bar{Y}_0 and \rightarrow denotes convergence of probability measures in total variation norm distance on the space of piecewise constant, right-continuous functions from $[0, \infty)$ into $\bar{\mathcal{H}}$, equipped with the Skorohod topology.

The proof of Proposition I.2.1.3 can be found in Subsection I.2.6.3.

The duality for processes on the complete graph (Proposition I.2.1.1) gives in the mean-field limit $N \rightarrow \infty$ rise to a duality between the mean-field dual process and solutions to the ODE (I.2.1.17). To describe the duality function, for any $p \in [0, 1]$, let B_p be a random variable with values in $\mathcal{S}(\mathbb{N})$ such that $(B_p(i))_{i \in \mathbb{N}}$ are i.i.d. Bernoulli random variables with mean p . We define $\bar{\phi} : [0, 1] \times \bar{\mathcal{H}} \rightarrow [0, 1]$ by

$$\bar{\phi}(p, \bar{E}) := \mathbb{P}[\text{Thin}_{B_p}(E) \neq \emptyset], \quad E \in \mathcal{H}, \quad (\text{I.2.1.23})$$

where we have implicitly used that the right-hand side depends only on the equivalence class $\bar{E} \in \bar{\mathcal{H}}$ and not on the choice of the representative $E \in \mathcal{H}$.

Proposition I.2.1.4 (Mean-field duality) *Let $(\bar{Y}_t)_{t \geq 0}$ be the mean-field dual process described above and let $(u_t)_{t \geq 0}$ be a $[0, 1]$ -valued solution to the ODE (I.2.1.17). Then*

$$\mathbb{E}[\bar{\phi}(u_0, \bar{Y}_t)] = \mathbb{E}[\bar{\phi}(u_t, \bar{Y}_0)], \quad t \geq 0. \quad (\text{I.2.1.24})$$

The proof of Proposition I.2.1.4 can be found in Subsection I.2.6.4.

I.2.1.3 Survival versus extinction

We will be interested in the mean-field dual process $(\bar{Y}_t)_{t \geq 0}$, which is the continuous-time Markov chain with state space $\bar{\mathcal{H}}$ and jump rates as in (I.2.1.20). Our first result says that the mean-field dual process survives with positive probability if and only if $\alpha \geq 4$. Below, \emptyset again denotes the hypergraph whose edge and vertex sets are both empty and $\bar{\emptyset} \in \bar{\mathcal{H}}$ is the same thing modulo permutation of sites.

Proposition I.2.1.5 (Survival versus extinction) *Let $(\bar{Y}_t)_{t \geq 0}$ be the mean-field dual process with branching rate α .*

(a) *If $\alpha < 4$, then the process started in any initial law satisfies*

$$\mathbb{P}[\exists t \geq 0 \text{ s.t. } \bar{Y}_t = \bar{\emptyset}] = 1. \quad (\text{I.2.1.25})$$

(b) *If $\alpha \geq 4$, then the process started in any deterministic initial state $\bar{Y}_0 \neq \bar{\emptyset}$ satisfies*

$$\mathbb{P}[\bar{Y}_t \neq \bar{\emptyset} \forall t \geq 0] > 0. \quad (\text{I.2.1.26})$$

Proof Since the proof contains some concepts that will be needed in what follows, we give it here. For $\alpha < 4$, the only fixed point of (I.2.1.17) is $x_0 := 0$. For $\alpha \geq 4$, there are additional fixed points at

$$x_1 := \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{\alpha}} \quad \text{and} \quad x_2 := \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{\alpha}}. \quad (\text{I.2.1.27})$$

For $\alpha > 4$, the fixed points x_0 and x_2 are stable while x_1 is unstable and separates the domains of attraction of x_0 and x_2 . For $\alpha = 4$, the fixed points x_1 and x_2 coincide in one point that is stable from the right but unstable from the left.

Let $(u_t)_{t \geq 0}$ be the solution to the ODE (I.2.1.17) with initial state $u_0 = 1$. Then the mean-field duality (Proposition I.2.1.4) implies that

$$\mathbb{P}[\bar{Y}_t \neq \bar{\emptyset}] = \mathbb{E}[\bar{\phi}(1, \bar{Y}_t)] = \mathbb{E}[\bar{\phi}(u_t, \bar{Y}_0)], \quad t \geq 0. \quad (\text{I.2.1.28})$$

If $\alpha < 4$, then $\lim_{t \rightarrow \infty} u_t = 0$ which implies part (a) of the proposition as $(Y_t)_{t \geq 0}$ takes values on a countable state space and so $\lim_{t \rightarrow \infty} Y_t = \emptyset$ if and only if $\exists t \geq 0$ such that $Y_t = \emptyset$. On the other hand, if $\alpha \geq 4$, then $\lim_{t \rightarrow \infty} u_t = x_2 > 0$, which, together with the continuity of \mathbb{P} , shows that for the process started in a deterministic initial state

$$\mathbb{P}[\bar{Y}_t \neq \bar{\emptyset} \ \forall t \geq 0] = \lim_{t \rightarrow \infty} \mathbb{P}[\bar{Y}_t \neq \bar{\emptyset}] = \bar{\phi}(x_2, \bar{Y}_0), \quad (\text{I.2.1.29})$$

which is positive if $\bar{Y}_0 \neq \bar{\emptyset}$. ■

I.2.1.4 A measure-valued dual

The duality of Proposition I.2.1.4 allows us to calculate expectations of the form $\mathbb{E}[\bar{\phi}(p, \bar{Y}_t)]$, where $p \in [0, 1]$ is a constant. In the present section, we generalize this duality by using a different duality function that allows us to calculate the expectation of more general functions of \bar{Y}_t .

Let $E \in \mathcal{H}$ be a deterministic hypergraph and let μ be a probability measure on $[0, 1]$. Let $\omega = (\omega_i)_{i \in \mathbb{N}}$ be an i.i.d. collection of $[0, 1]$ -valued random variables with a common law μ , and conditional on ω let $B_\omega = (B_\omega(i))_{i \in \mathbb{N}}$ be a collection of independent Bernoulli random variables with $\mathbb{P}[B_\omega(i) = 1 | \omega] = \omega_i$ ($i \in \mathbb{N}$). Let $\mathcal{M}_1[0, 1]$ denote the space of all probability measures on $[0, 1]$. We define a function $\rho : \mathcal{M}_1[0, 1] \times \bar{\mathcal{H}} \rightarrow \mathcal{M}_1[0, 1]$ by

$$\rho(\mu, \bar{E}) := \mathbb{P}[\mathbb{P}[\text{Thin}_{B_\omega}(E) \neq \emptyset | \omega] \in \cdot], \quad (\text{I.2.1.30})$$

i.e., $\rho(\mu, \bar{E})$ is the law of the $[0, 1]$ -valued random variable $\mathbb{P}[\text{Thin}_{B_\omega}(E) = \emptyset | \omega]$. Note that (I.2.1.30) is a good definition since the right-hand side does not depend on the choice of the representative E in the equivalence class \bar{E} . An equivalent way to characterize $\rho = \rho(\mu, \bar{E})$ is by the formula

$$\int f \, d\rho := \mathbb{E}[f(\mathbb{P}[\text{Thin}_{B_\omega}(E) \neq \emptyset | \omega])], \quad f \in B_b[0, 1], \quad (\text{I.2.1.31})$$

where $B_b[0, 1]$ denotes the space of bounded measurable functions $f : [0, 1] \rightarrow \mathbb{R}$. We will prove that the mean-field dual process has a deterministic dual with respect to the duality function ρ .

To define this dual process consider the nonlinear map ψ acting on probability measures on $[0, 1]$ defined by

$$\psi(\mu) := \mathbb{P}[\omega_1 + (1 - \omega_1)\omega_2\omega_3 \in \cdot] \quad \text{if } \omega_1, \omega_2, \omega_3 \text{ are i.i.d. with common law } \mu. \quad (\text{I.2.1.32})$$

The deterministic dual that we are looking for will be given by the solutions to the differential equation

$$\frac{\partial}{\partial t} \mu_t = (\delta_0 - \mu_t) + \alpha(\psi(\mu_t) - \mu_t). \quad (\text{I.2.1.33})$$

We interpret (I.2.1.33) as follows. For any measure μ and real measurable function f , write $\langle \mu, f \rangle := \int f \, d\mu$. Let $\mathcal{C}[0, 1]$ denote the space of continuous functions $f : [0, 1] \rightarrow \mathbb{R}$. We say that a function $[0, \infty) \ni t \mapsto \mu_t \in \mathcal{M}_1[0, 1]$ solves (I.2.1.33) if for any $f \in \mathcal{C}[0, 1]$, the function $[0, \infty) \ni t \mapsto \langle \mu_t, f \rangle \in \mathbb{R}$ is continuously differentiable and solves

$$\frac{\partial}{\partial t} \langle \mu_t, f \rangle = (\langle \delta_0, f \rangle - \langle \mu_t, f \rangle) + \alpha(\langle \psi(\mu_t), f \rangle - \langle \mu_t, f \rangle). \quad (\text{I.2.1.34})$$

We need the following results.

Proposition I.2.1.6 (Existence) For each probability measure μ_0 on $[0, 1]$ there exists a $\mathcal{M}_1([0, 1])$ -valued solution $(\mu_t)_{t \geq 0}$ of the equation (I.2.1.33) with initial state μ_0 .

The proof of Proposition I.2.1.6 can be found in Subsection I.2.6.5.

The next Lemma tells us that the k -th moment of μ_t only depends on the moments lower than k .

Lemma I.2.1.7 (Moment equations) For $k \in \mathbb{N}$ let $p_t(k)$ be the k -th moment of a particular solution $(\mu_t)_{t \geq 0}$ of the equation (I.2.1.33) such that $\mu_t \in \mathcal{M}_1([0, 1])$ for all $t \geq 0$. Let X be a μ_t -distributed random variable. Then $p_t(k)$ solves the equation

$$\begin{aligned} \frac{\partial}{\partial t} p_t(k) &= -p_t(k) + \alpha \left(\sum_{l=1}^k \binom{k}{l} \mathbb{E}[X^{k-l}(1-X)^l] p_t^2(l) \right) \\ &= -p_t(k) + \alpha \left(\sum_{l=1}^k \binom{k}{l} \sum_{j=0}^l \binom{l}{j} (-1)^j p_t(k-l+j) p_t^2(l) \right), \end{aligned} \quad (\text{I.2.1.35})$$

where we put $p_t(0) := 1$.

The proof of Lemma I.2.1.7 can be found in Subsection I.2.6.5.

Proposition I.2.1.8 (Uniqueness) For each probability measure μ_0 on $[0, 1]$, there exists a unique $\mathcal{M}_1([0, 1])$ -valued solution $(\mu_t)_{t \geq 0}$ of the equation (I.2.1.33) with initial state μ_0 .

The proof of Proposition I.2.1.8 can be found in Subsection I.2.6.5.

We are now ready to formulate the measure duality, which is our first main result. Below, we define the expectation of a random measure in the usual way, i.e., if ν is a $\mathcal{M}_1[0, 1]$ -valued random variable, then $\mathbb{E}[\nu]$ is the probability measure on $[0, 1]$ defined by $\langle \mathbb{E}[\nu], f \rangle := \mathbb{E}[\langle \nu, f \rangle]$ ($f \in B_b[0, 1]$).

Theorem I.2.1.9 (Measure-valued dual) Let $(\bar{Y}_t)_{t \geq 0}$ be the mean-field dual and let $(\mu_t)_{t \geq 0}$ solve (I.2.1.33). Then

$$\mathbb{E}[\rho(\mu_0, \bar{Y}_t)] = \mathbb{E}[\rho(\mu_t, \bar{Y}_0)] \quad (t \geq 0). \quad (\text{I.2.1.36})$$

The proof of Theorem I.2.1.9 can be found in Subsection I.2.6.5.

Let $\mu \in \mathcal{M}_1[0, 1]$ be a probability measure on $[0, 1]$, $E \in \mathcal{H}$ be a hypergraph and for $b = \{b_i\}_{i \in \cup E} \in \{0, 1\}^{\cup E}$ define

$$\text{Thin}_b(E) := \{e \in E; b_i = 1 \forall i \in e\}.$$

Denote by $T_E := \{b \in \{0, 1\}^{\cup E} \text{ s.t. } \text{Thin}_b(E) \neq \emptyset\}$ the set of all configurations b such that thinning E with b does result in an empty set. Let l denote the linear function $l(x) = x$ and put $p := \int_0^1 x d\mu(x)$. Then using the fact that $B_\omega(i), i \in \cup E$ are independent and identically distributed Bernoulli random variables with success probability p we see from (I.2.1.30) that

$$\begin{aligned} \langle \rho(\mu, E), l \rangle &= \mathbb{E}[\mathbb{P}[\text{Thin}_{B_\omega}(E) \neq \emptyset | \omega]] = \sum_{b \in T_E} \mathbb{E}[\mathbb{P}[B_\omega = b | \omega]] \\ &= \sum_{b \in T_E} \mathbb{P}[B_\omega = b] = \mathbb{P}[\text{Thin}_{B_p}(E) \neq \emptyset] = \bar{\phi}(\langle \mu, l \rangle, \bar{E}). \end{aligned} \quad (\text{I.2.1.37})$$

Hence, we see from (I.2.1.37) that

$$\langle \rho(\mu, \bar{E}), l \rangle = \bar{\phi}(\langle \mu, l \rangle, \bar{E}), \quad \mu \in \mathcal{M}_1[0, 1], \quad E \in \mathcal{H}. \quad (\text{I.2.1.38})$$

If $(\mu_t)_{t \geq 0}$ solves the differential equation (I.2.1.33), then it is not hard to see that $u_t := \langle \mu_t, l \rangle$ ($t \geq 0$) solves (I.2.1.17), so using (I.2.1.38) we obtain from (I.2.1.36) that

$$\mathbb{E}[\bar{\phi}(u_0, \bar{Y}_t)] = \mathbb{E}[\langle \rho(\mu_0, \bar{Y}_t), l \rangle] = \mathbb{E}[\langle \rho(\mu_t, \bar{Y}_0), l \rangle] = \mathbb{E}[\bar{\phi}(u_t, \bar{Y}_0)], \quad t \geq 0, \quad (\text{I.2.1.39})$$

which shows that the duality of Proposition I.2.1.4 is a special case of the more general Theorem I.2.1.9.

Let $\{\{1\}\} \in \mathcal{H}$ denote the hypergraph that contains only a single hyperedge $\{1\}$ which in turn contains only the vertex 1, and let $\bar{1} := \overline{\{\{1\}\}} \in \bar{\mathcal{H}}$ denote the same thing modulo isomorphisms, i.e., $\bar{1}$ is the hypergraph consisting of a single hyperedge containing a single vertex. We let $\mathbb{P}^{\bar{1}}$ denote the law of the mean-field dual process started in $\bar{Y}_0 = \bar{1}$. Proposition I.2.1.4 allows us to calculate, for any $p \in [0, 1]$, the expectation $\mathbb{E}^{\bar{1}}[\bar{\phi}(p, \bar{Y}_t)]$. The following lemma shows that we can in fact obtain the whole distribution of $\bar{\phi}(p, \bar{Y}_t)$.

Proposition I.2.1.10 (Law of thinning probability) *Let $p \in [0, 1]$ and let $(\mu_t)_{t \geq 0}$ be the solution of (I.2.1.33) with initial state $\mu_0 = \delta_p$. Then*

$$\mathbb{P}^{\bar{1}}[\bar{\phi}(p, \bar{Y}_t) \in \cdot] = \mu_t \quad (t \geq 0). \quad (\text{I.2.1.40})$$

The proof of Proposition I.2.1.10 can be found in subsection I.2.6.6.

I.2.1.5 Convergence of the measure-valued function

Recall from the proof of Proposition I.2.1.5 that for $\alpha \geq 4$, the ODE (I.2.1.17) has three fixed points $x_0 < x_1 \leq x_2$. It follows from Proposition I.2.1.4 and the Markov property of $(\bar{Y}_t)_{t \geq 0}$ that $\bar{\phi}(x_m, \cdot)$, $m = 0, 1, 2$ are harmonic functions of the Markov process $(\bar{Y}_t)_{t \geq 0}$. Therefore, the processes $(H_t^m)_{t \geq 0}$ defined by

$$H_t^m := \bar{\phi}(x_m, \bar{Y}_t), \quad m = 0, 1, 2, \quad t \geq 0 \quad (\text{I.2.1.41})$$

are bounded martingales. Hence, the almost sure limits

$$H_\infty^m := \lim_{t \rightarrow \infty} H_t^m \quad \text{a.s.}, \quad m = 0, 1, 2 \quad (\text{I.2.1.42})$$

exist. By Proposition I.2.1.10,

$$\mathbb{P}^{\bar{1}}[H_t^m \in \cdot] = \mu_t^m \quad (\text{I.2.1.43})$$

where $(\mu_t^m)_{t \geq 0}$ is given by the unique solution of (I.2.1.33) with initial state $\mu_0^m := \delta_{x_m}$. The process H_t^0 is identically zero, but the other two processes are more interesting.

Proposition I.2.1.11 *Assume that $\alpha \geq 4$ and let $\mathcal{S} := \{\bar{Y}_t \neq \bar{\emptyset} \forall t \geq 0\}$ denote the event that $(\bar{Y}_t)_{t \geq 0}$ survives. Then one has*

$$H_\infty^2 = 1_{\mathcal{S}} \quad \text{a.s.} \quad (\text{I.2.1.44})$$

Furthermore, it holds that $\mathbb{P}[H_\infty^2 = 1] = x_2$.

Proof Put $u_0 = 1$. Using the duality (I.2.1.24) we see for any $\bar{Y}_0 \in \bar{\mathcal{H}}$ that

$$\mathbb{P}^{\bar{Y}_0}[\bar{Y}_t \neq \bar{\emptyset}] = \mathbb{E}^{\bar{Y}_0}[\bar{\phi}(1, \bar{Y}_t)] = \bar{\phi}(u_t, \bar{Y}_0) \xrightarrow{t \rightarrow \infty} \bar{\phi}(x_2, \bar{Y}_0) = \mathbb{E}^{\bar{Y}_0}[H_\infty^2], \quad (\text{I.2.1.45})$$

where in the last equality we used the martingale property of H^2 . From (I.2.1.45) we see that $\mathbb{P}^{\bar{Y}_0}[\mathcal{S}] = \mathbb{E}^{\bar{Y}_0}[H_\infty^2]$. From the definition of H_∞^2 , we see immediately that $H_\infty^2 = 0$ almost surely on the event \mathcal{S}^c . Hence, we get that

$$\mathbb{E}^{\bar{Y}_0}[H_\infty^2] = \mathbb{E}^{\bar{Y}_0}[H_\infty^2 | \mathcal{S}] \mathbb{P}^{\bar{Y}_0}[\mathcal{S}] + \mathbb{E}^{\bar{Y}_0}[H_\infty^2 | \mathcal{S}^c] \mathbb{P}^{\bar{Y}_0}[\mathcal{S}^c] = \mathbb{E}^{\bar{Y}_0}[H_\infty^2 | \mathcal{S}] \mathbb{E}^{\bar{Y}_0}[H_\infty^2] \quad (\text{I.2.1.46})$$

Equation (I.2.1.46) shows that $H_\infty^2 = 1$ almost surely on the event that the process \bar{Y} survives. It remains to note that H_∞^2 is a Bernoulli random variable by (I.2.1.44) and so $\mathbb{P}[H_\infty^2 = 1] = x_2$ by (I.2.1.42) and the fact that $(H_t^2)_{t \geq 0}$ is a martingale such that $\mathbb{E}[H_t^2] = x_2$ for all $t \geq 0$. ■

Lemma I.2.1.12 *Assume that $\mathbb{P}[\bar{Y}_0 \neq \emptyset] > 0$. Then for all $\alpha > 4$ it holds that $\mathbb{P}[H_\infty^1 = 1 | \mathcal{S}] < 1$ and $\mathbb{P}[H_\infty^1 \in (0, 1)] > 0$.*

Proof Since we will need some concepts from the proof in the next paragraph, we provide it here. Function $\bar{\phi}$ is strictly monotone in the first coordinate and so $\mathbb{E}[H_0^1] < \mathbb{E}[H_0^2]$ as long as $\mathbb{P}[\bar{Y}_0 \neq \emptyset] > 0$. Since $(H_t^1)_{t \geq 0}$ and $(H_t^2)_{t \geq 0}$ are bounded martingales and H_∞^1 and H_∞^2 their respective limits, it also holds that $\mathbb{E}^{\bar{Y}_0}[H_\infty^1] < \mathbb{E}^{\bar{Y}_0}[H_\infty^2]$ and so by Proposition I.2.1.11 we have $\mathbb{E}^{\bar{Y}_0}[H_\infty^1] < \mathbb{P}^{\bar{Y}_0}[\mathcal{S}]$. Using the same calculation as in the Proof of Proposition I.2.1.11 (concretely as in (I.2.1.46)), we see that $\mathbb{P}[H_\infty^1 = 1 | \mathcal{S}] < 1$.

Unlike H_∞^2 , H_∞^1 is not a Bernoulli random variable. That can be seen as follows. We know by Lemma I.2.1.7 that the second moment $p_t(2)$ of μ_t solves the equation

$$\frac{\partial}{\partial t} p_t(2) = \alpha p_t^3(2) + \alpha p_t^2(2)(1 - 2p_t(1)) - p_t(2)(1 + 2\alpha p_t^2(1)) + 2\alpha p_t^3(1). \quad (\text{I.2.1.47})$$

If we set $\mu_0 = \delta_{x_1}$, with x_1 as in (I.2.1.27), then the first moment $p_t(1)$ stays constant in t and equal to x_1 , so (I.2.1.47) turns into

$$\frac{\partial}{\partial t} p_t(2) = \alpha p_t^3(2) + \alpha p_t^2(2)(1 - 2x_1) - p_t(2)(1 + 2\alpha x_1^2) + 2\alpha x_1^3. \quad (\text{I.2.1.48})$$

Using the fact that x_1 is a fixed point of the equation (I.2.1.17), we can easily check that it is also a fixed point of (I.2.1.48) (We will denote it by $m_3 := x_1$) and using that we can calculate the other two fixed points m_1 and m_2 of (I.2.1.48) as $m_1 = \frac{x_1 - 1 - \sqrt{13x_1^2 - 6x_1 + 1 + 4/\alpha}}{2}$ and

$$m_2 = \frac{x_1 - 1 + \sqrt{13x_1^2 - 6x_1 + 1 + 4/\alpha}}{2}. \quad (\text{I.2.1.49})$$

Let $F : [0, 1] \rightarrow \mathbb{R}$ by defined as $F(p) := \alpha p^3 + \alpha p^2(1 - 2x_1) - p(1 + 2\alpha x_1^2) + 2\alpha x_1^3$. Then we can write

$$\frac{\partial}{\partial p} F(p)|_{p=p_t(2)} = 3\alpha p_t^2(2) + 2\alpha p_t(2)(1 - 2x_1) - 1 - 2\alpha x_1^2. \quad (\text{I.2.1.50})$$

When $p_t(2) = x_1$, the right-hand side of (I.2.1.50) turns into

$$2\alpha x_1 - 3\alpha x_1^2 - 1. \quad (\text{I.2.1.51})$$

Since $x_1 = \frac{1}{2} - \sqrt{\frac{1}{2} - \frac{1}{\alpha}} \neq 0$ for $\alpha > 4$ and x_1 is a root of the equation (I.2.1.17), we see that $\alpha x_1 - \alpha x_1^2 - 1 = 0$. Plugging that into (I.2.1.51), we see that (I.2.1.51) is positive whenever $x_1 > 2x_1^2$. Since $x_1^2 = x_1 - 1/\alpha$, we get from (I.2.1.27) that this holds whenever $\alpha > 4$. The fixed points x_1 and m_2 are given by (I.2.1.27) and (I.2.1.49), respectively, so by a simple

calculation we get that also $x_1 > m_2$ for all $\alpha > 4$. It holds therefore for all $\alpha > 4$ that $x_1 > m_2 > m_1$. Hence, as $\frac{\partial}{\partial p}F(p)$ is positive for $p = x_1$,

$$x_1 \text{ and } m_1 \text{ are unstable fixed points of the equation (I.2.1.48).} \quad (\text{I.2.1.52})$$

Since there are no other fixed points and since by (I.2.1.42) we know that the limit H_∞^1 exists, it follows that m_2 is the second moment of H_∞^1 provided that H_0^1 is not Bernoulli distributed (i.e. provided that $p_0(2) \neq x_1$). However from the definition (I.2.1.41) of H_t^1 it is clear that H_0^1 cannot be Bernoulli distributed, since $\phi(x, \bar{Y}_0) < 1$ almost surely for any $0 \leq x < 1$. Hence we see that $\mathbb{P}[H_\infty^1 \in (0, 1)] > 0$ for $\alpha > 4$, since a Bernoulli distribution on $\{0, 1\}$ with a mean x_1 also has all higher moments equal to x_1 . ■

More generally, we want to study fixed points of (I.2.1.33) and their domains of attraction.

Theorem I.2.1.13 *Let $\alpha \geq 4$, the equation (I.2.1.33) has the fixed points*

$$\nu_0 := \delta_0, \quad \nu_1 := (1 - x_1)\delta_0 + x_1\delta_1, \quad \text{and} \quad \nu_2 := (1 - x_2)\delta_0 + x_2\delta_1, \quad (\text{I.2.1.53})$$

where x_1 and x_2 are the fixed points (I.2.1.27) of (I.2.1.17). There exists at least one more fixed point ν_3 of (I.2.1.33) which corresponds to the law of $H_\infty^1 = \lim_{t \rightarrow \infty} \bar{\phi}(x_1, \bar{Y}_t)$ with $\bar{Y}_0 := \bar{1}$. The fixed point ν_3 has the form

$$\nu_3 = (1 - x_2)\delta_0 + x_2\hat{\nu} \quad (\text{I.2.1.54})$$

for some probability law $\hat{\nu}$ on $[0, 1]$. For $k \in \mathbb{N}$ and $\nu \in \mathcal{M}_1[0, 1]$ denote by $p^\nu(k) := \int_0^1 x^k d\nu$ the k -th moment of ν . Then we have

$$p^{\nu_3}(1) = x_1, \quad p^{\nu_3}(2) = m_2, \quad (\text{I.2.1.55})$$

where m_2 is defined in (I.2.1.49) and the domains of attraction of ν_0 , ν_1 and ν_2 are as follows:

$$\begin{aligned} \{\nu \in \mathcal{M}_1[0, 1]; p^\nu(1) < x_1\} & \quad \text{for } \nu_0, \\ \{\nu \in \mathcal{M}_1[0, 1]; p^\nu(1) = x_1, p^\nu(2) = x_1\} & \quad \text{for } \nu_1, \\ \{\nu \in \mathcal{M}_1[0, 1]; p^\nu(1) > x_1\} & \quad \text{for } \nu_2. \end{aligned} \quad (\text{I.2.1.56})$$

The proof of Theorem I.2.1.13 can be found in section I.2.6.7.

Remark I.2.1.14 For $\alpha < 4$ it is easy to see that ν_0 is the only fixed point of (I.2.1.33). Similarly, for $\alpha = 4$, the fixed points ν_1 and ν_2 are equal (since $x_1 = x_2$) and the only fixed points of (I.2.1.33) are ν_0 with the domain of attraction $\{\nu \in \mathcal{M}_1[0, 1]; p^\nu(1) < x_1\}$ and ν_2 with the domain of attraction $\{\nu \in \mathcal{M}_1[0, 1]; p^\nu(1) \geq x_1\}$. The last statement follows from the fact that in this case, the only two fixed points of the first two moment equations (I.2.1.35) are (x_0, x_0) and (x_1, x_1) and $p_t(1) \rightarrow x_0$ as $t \rightarrow \infty$ whenever $p_0(1) < x_1$.

Remark I.2.1.15 Under the conditions of Proposition I.2.1.11, if $\alpha > 4$, simulations suggest that $0 < H_\infty^1 < 1$ almost surely on the event \mathcal{S} , which is to say that $\hat{\nu}$ is a distribution concentrated on $(0, 1)$.

In the following two subsections we study the connections between the mean-field dual and the function μ_t further and provide an interpretation for the moments of μ_t which will follow from Theorem I.2.1.19 and Proposition I.2.1.20.

I.2.1.6 Coupled processes on the complete graph

Put $\Lambda_N = \{1, \dots, N\}$ and let Π_N denote the group of all permutations of Λ_N . For each $\pi \in \Pi_N$, we define $T_\pi : \mathcal{S}(\Lambda_N) \rightarrow \mathcal{S}(\Lambda_N)$ by

$$(T_\pi x)(i) := x(\pi(i)) \quad (i \in \Lambda_N). \quad (\text{I.2.1.57})$$

For any set $E \subset \mathcal{S}(\Lambda_N)$, we also write $T_\pi(E) := \{T_\pi(e) : e \in E\}$ for the image of E under T_π . Note that the equivalence relation (I.2.1.6) can be written for $E, E' \in \mathcal{H}(\Lambda_N)$ in terms of T_π as

$$E \sim E' \quad \text{if and only if} \quad E = T_\pi(E) \quad \text{for some } \pi \in \Pi_N.$$

For $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{S}(\Lambda_N)^n$, we define

$$\bar{\mathbf{x}}^\sigma := \frac{1}{N} \sum_{i \in \Lambda_N} 1_{\{\mathbf{x}(i) = \sigma\}} \quad (\sigma \in \{0, 1\}^n), \quad (\text{I.2.1.58})$$

where $\mathbf{x}(i) = (x_1(i), \dots, x_n(i))$ and $\sigma = (\sigma_1, \dots, \sigma_n) \in \{0, 1\}^n$. In words, for each $\sigma \in \{0, 1\}^n$, $\bar{\mathbf{x}}^\sigma$ is the ratio of sites which are occupied in configurations x_i such that $\sigma_i = 1$ and which are empty in configurations x_j such that $\sigma_j = 0$. Note that $\bar{\mathbf{x}} := (\bar{\mathbf{x}}^\sigma, \sigma \in \{0, 1\}^n)$ contains all information about (x_1, \dots, x_n) modulo a joint permutation of the sites, i.e., $\bar{\mathbf{x}} = \bar{\mathbf{y}}$ if and only if there exists a $\pi \in \Pi_N$ such that $(x_1, \dots, x_n) = (T_\pi y_1, \dots, T_\pi y_n)$. We let \mathcal{M}_N^n denote the space of all functions $\bar{\mathbf{x}} : \{0, 1\}^n \rightarrow \{0, 1/N, \dots, 1\}$ such that $\sum_{\sigma \in \{0, 1\}^n} \bar{\mathbf{x}}^\sigma = 1$.

Let $(\mathbf{X}_t)_{t \geq 0}$ with

$$\mathbf{X}_t = (X_t^1, \dots, X_t^n) \quad (\text{I.2.1.59})$$

be the Markov process with state space $\mathcal{S}(\Lambda_N)^n$ such that for each $k = 1, \dots, n$ each component X^k is a cooperative branching process on Λ_N with generator as in (I.1.2.14) and rates as in (I.2.1.1), and these components are coupled using the same graphical representation in terms of the maps coop_{ijk} and death_i . Then $(\bar{\mathbf{X}}_t)_{t \geq 0} = (\bar{\mathbf{X}}_t^N)_{t \geq 0}$ is a Markov process with state space \mathcal{M}_N^n .

In the limit $N \rightarrow \infty$ for fixed n , the process $(\bar{\mathbf{X}}_t)_{t \geq 0}$ converges to a solution of an ODE that is dual to the mean-field dual process with values in $\bar{\mathcal{H}}$. In the rest of this section we will show this convergence and duality. First, we provide the duality between $(\bar{\mathbf{X}}_t)_{t \geq 0}$ and $(\bar{Y}_t^N)_{t \geq 0}$. The construction is very similar to that used in Proposition I.2.1.1. For a $\mathbf{p} \in \mathcal{M}_N^n$ let $\mathbf{B}_\mathbf{p}^N := (B_\mathbf{p}^{N,k}, k \in \{1, \dots, n\})$ be a random vector uniformly distributed on the set $\{\mathbf{b} \in \mathcal{S}(\Lambda_N)^n; \bar{\mathbf{b}}(\sigma) = \mathbf{p}(\sigma)\}$. Define $\phi_N^n : \mathcal{M}_N^n \times \bar{\mathcal{H}}_N \rightarrow [0, 1]$

$$\phi_N^n(\mathbf{p}, \bar{E}) := \mathbb{P}[\text{Thin}_{B_\mathbf{p}^{N,k}}(E) \neq \emptyset, k \in \{1, \dots, n\}]. \quad (\text{I.2.1.60})$$

For $n = 1$, the above definitions of $\mathbf{B}_\mathbf{p}^N$ and ϕ_N^n correspond to B_p^N and $\bar{\phi}^N$ which we defined in (I.2.1.15). We will base our duality on the following fact.

Proposition I.2.1.16 (Duality for coupled processes) *Let χ be as in (I.1.4.6). For $i = 1, \dots, n$, let $X^i = (X_t^i)_{t \geq 0}$ be interacting particle systems with generator as in (I.1.2.14), which are constructed using the same graphical representation in terms of the maps coop_{ijk} and death_i . Let $(Y_t)_{t \geq 0}$ be the continuous-time Markov process with generator as in (I.1.4.5). Then, assuming that (X_0^1, \dots, X_0^n) is independent of Y_t and (X_t^1, \dots, X_t^n) is independent of Y_0 , we have that*

$$\mathbb{E}\left[\prod_{i=1}^n \chi(X_t^i, Y_0)\right] = \mathbb{E}\left[\prod_{i=1}^n \chi(X_0^i, Y_t)\right] \quad (t \geq 0). \quad (\text{I.2.1.61})$$

Proof Proposition I.2.1.16 is an immediate consequence of Proposition I.1.4.1, i.e. of the fact that we have a pathwise duality. \blacksquare

We use Proposition (I.2.1.16) to obtain a duality between the dual process and the coupled processes on the complete graph, which is the same as the duality (I.2.1.16) in the case when $n = 1$.

Proposition I.2.1.17 *Let $\bar{\mathbf{X}}_0$ be independent of $(\bar{Y}_t^N)_{t \geq 0}$ and \bar{Y}_0^N independent of $(\bar{\mathbf{X}}_t)_{t \geq 0}$. Then*

$$\mathbb{E}[\phi_N^n(\bar{\mathbf{X}}_t, \bar{Y}_0^N)] = \mathbb{E}[\phi_N^n(\bar{\mathbf{X}}_0, \bar{Y}_t^N)]. \quad (\text{I.2.1.62})$$

The proof of Proposition I.2.1.17 can be found in Section I.2.6.1.

We will now describe the dynamics of the process $(\bar{\mathbf{X}}_t)_{t \geq 0}$. Let $\bar{\mathbf{x}}$ be the current state of the process and put $\mathbf{0} := \{0, \dots, 0\} \in \{0, 1\}^n$. First, we note that each site $i \in \{1, \dots, N\}$ is counted exactly once in the sense that it adds $1/N$ to the count $\bar{\mathbf{x}}^\sigma$ for one and only one $\sigma \in \{0, 1\}^n$. This means that any death event at site i only decreases $\bar{\mathbf{x}}^{\sigma'}$ by $1/N$ for the particular σ' such that $\mathbf{x}^k(i) = 1$ if and only if $\sigma'(k) = 1$. It also increases $\bar{\mathbf{x}}^{\mathbf{0}}$ by one and leaves $\bar{\mathbf{x}}^\sigma$ unchanged for all other σ . Cooperative branching to a site i changes $x_j(i)$ from 0 to 1 for $j \in J$ where $J \subseteq \{1, \dots, n\}$. In terms of $\bar{\mathbf{x}}$ this means a change $\bar{\mathbf{x}}^\sigma \rightarrow \bar{\mathbf{x}}^\sigma + 1/N$ and $\bar{\mathbf{x}}^{\sigma^-} \rightarrow \bar{\mathbf{x}}^{\sigma^-} - 1/N$ for some $\sigma \neq \sigma^-$, $\sigma^- \leq \sigma \in \{0, 1\}^n$ while $\bar{\mathbf{x}}(\sigma)$ remains unchanged for all other σ . The quantity $\bar{\mathbf{x}}^{\sigma^-}$ should be thought of as the ratio of sites which can be the target ("birth site") of the branching event so that in that branching event the site becomes occupied in exactly the coupled processes with indices i such that $\sigma_i^- = 0$ but $\sigma_i = 1$. To that end, we put for $\sigma^- \neq \sigma$, $\sigma^- \leq \sigma$

$$J(\sigma, \sigma^-) := \{\sigma' \in \{0, 1\}^n; \sigma'_j = 1 \text{ whenever } \sigma_j - \sigma_j^- = 1\} \quad (\text{I.2.1.63})$$

and

$$K(\sigma, \sigma^-) := \{\sigma' \in J(\sigma, \sigma^-); \sigma'_j = 0 \text{ whenever } \sigma_j = 0\}. \quad (\text{I.2.1.64})$$

In words, $J(\sigma, \sigma^-)$ is the set of all σ' whose marginals are 1 whenever the corresponding marginals of σ are 1 but those of σ^- are 0. This should be thought of as follows. In an event which increases $\bar{\mathbf{x}}^\sigma$ and decreases $\bar{\mathbf{x}}^{\sigma^-}$, each $\sigma \in J(\sigma, \sigma^-)$ gives us a possible set of indices of $\mathbf{x} = (x_1, \dots, x_n)$ such that one of the parent sites is occupied in configuration x_j if and only if $\sigma_j = 1$. This is necessary in order for it to be possible to "occupy" the site to which the branching occurs at configurations x_j such that $\sigma_j = 1$ but $\sigma_j^- = 0$. In the branching event where $\bar{\mathbf{x}}^\sigma$ increases and $\bar{\mathbf{x}}^{\sigma^-}$ decreases, it is also necessary that the "birth site" does not become occupied in any configuration x_j such that $\sigma_j = 0$. That means that at least one of the parent sites of the event has to be unoccupied at x_j whenever $\sigma_j = 0$ which is where the definition of $K(\sigma, \sigma^-)$ comes from.

Denote by $O^{\bar{\mathbf{x}}}(\sigma, \sigma^-) := \sum_{\sigma' \in J(\sigma, \sigma^-)} \bar{\mathbf{x}}(\sigma')$ the frequency of sites occupied by the configurations which need to branch to one of the sites corresponding to σ^- in order to increase $\bar{\mathbf{x}}(\sigma)$ by $1/N$ and by $\hat{O}^{\bar{\mathbf{x}}}(\sigma, \sigma^-) := \sum_{\sigma' \in K(\sigma, \sigma^-)} \bar{\mathbf{x}}(\sigma')$ we denote the frequency of such sites where the configurations which cannot branch in this event are excluded. Then the dynamics of $(\bar{\mathbf{X}}_t)_{t \geq 0}$ are as follows.

$$\begin{aligned} \bar{x}^\sigma &\mapsto \bar{x}^\sigma + \frac{1}{N}, \quad \bar{x}^{\sigma^-} \mapsto \bar{x}^{\sigma^-} - \frac{1}{N} & \text{at rate} & \quad \alpha N \bar{\mathbf{x}}^{\sigma^-} (O^{\bar{\mathbf{x}}}(\sigma, \sigma^-) - \frac{1}{N}) \hat{O}^{\bar{\mathbf{x}}}(\sigma, \sigma^-), \\ \bar{x}^\sigma &\mapsto \bar{x}^\sigma - \frac{1}{N}, \quad \bar{x}^{\mathbf{0}} \mapsto \bar{x}^{\mathbf{0}} + \frac{1}{N} & \text{at rate} & \quad N \bar{x}^\sigma, \end{aligned} \quad (\text{I.2.1.65})$$

where again $\sigma^- \neq \sigma$, $\sigma^- \leq \sigma$. As $N \rightarrow \infty$, the process $\bar{\mathbf{X}}^N$ then converges a solution of a 2^n -dimensional ODE.

Proposition I.2.1.18 (*Coupled mean-field forward process*) Let $(\bar{\mathbf{X}}_t^N)_{t \geq 0}$ be Markov processes as in (I.2.1.65) started in deterministic initial states $\bar{\mathbf{X}}_0^N$ that converge in probability as $N \rightarrow \infty$ to some deterministic $\mathbf{p}_0 \in \mathcal{M}_\infty^n := \{r : \{0, 1\}^n \rightarrow [0, 1] \text{ s.t. } \sum_{\sigma \in \{0, 1\}^n} r(\sigma) = 1\}$. Let $(\mathbf{p}_t)_{t \geq 0}$ denote the solution of the ODE

$$\begin{aligned} \frac{\partial}{\partial t} \mathbf{p}_t(\sigma) &= \alpha \sum_{\sigma^- \leq \sigma, \sigma^- \neq \sigma} \mathbf{p}_t(\sigma^-) O^{\mathbf{p}_t}(\sigma, \sigma^-) \hat{O}^{\mathbf{p}_t}(\sigma, \sigma^-) \\ &\quad - \alpha \sum_{\sigma \leq \sigma^+, \sigma \neq \sigma^+} \mathbf{p}_t(\sigma) O^{\mathbf{p}_t}(\sigma^+, \sigma) \hat{O}^{\mathbf{p}_t}(\sigma^+, \sigma) - \mathbf{p}_t(\sigma), \quad \sigma \in \{0, 1\}^n \setminus \mathbf{0}, \\ \frac{\partial}{\partial t} \mathbf{p}_t(\mathbf{0}) &= (1 - \mathbf{p}_t(\mathbf{0})) - \alpha \sum_{\sigma \neq \mathbf{0}} \mathbf{p}_t(\sigma) O^{\mathbf{p}_t}(\sigma, \mathbf{0}) \hat{O}^{\mathbf{p}_t}(\sigma, \mathbf{0}) \end{aligned} \quad (\text{I.2.1.66})$$

with initial state \mathbf{p}_0 , where $O^{\mathbf{p}}(\sigma, \sigma^-) := \sum_{\sigma' \in J(\sigma, \sigma^-)} \mathbf{p}(\sigma')$ and $\hat{O}^{\mathbf{p}}(\sigma, \sigma^-) := \sum_{\sigma' \in K(\sigma, \sigma^-)} \mathbf{p}(\sigma')$. Then, for each $T < \infty$ and $\varepsilon > 0$,

$$\mathbb{P}[\|\bar{\mathbf{X}}_t^N - \mathbf{p}_t\| \leq \varepsilon \forall t \in [0, T]] \xrightarrow{N \rightarrow \infty} 1. \quad (\text{I.2.1.67})$$

The proof of Proposition I.2.1.18 can be found in section I.2.6.2.

Note that for $n = 1$ we obtain the equation (I.2.1.17). We now define independent identically distributed random variables $B_{\mathbf{p}}(i) = \{B_{\mathbf{p}}^k(i), k \in \{1, \dots, n\}\}$, $i \in \mathbb{N}$ on $\{0, 1\}^n$ with distribution given by $\mathbf{p} = (p(\sigma), \sigma \in \{0, 1\}^n) \in \mathcal{M}_\infty^n$. Let $\bar{\phi}^n : \mathcal{M}_\infty^n \times \bar{\mathcal{H}} \rightarrow [0, 1]$ be defined as

$$\bar{\phi}^n(\mathbf{p}, \bar{E}) := \mathbb{P}[\text{Thin}_{B_{\mathbf{p}}^k(\cdot)}(E) \neq \emptyset, k \in \{1, \dots, n\}] \quad (E \in \mathcal{H}), \quad (\text{I.2.1.68})$$

where we have implicitly used that the right-hand side depends only on the equivalence class $\bar{E} \in \bar{\mathcal{H}}$ and not on the choice of the representative $E \in \mathcal{H}$.

Theorem I.2.1.19 (Moment duality) Let $n \in \mathbb{N}$ and let $(\bar{Y}_t)_{t \geq 0}$ be the mean-field dual process. Furthermore, let $(\mathbf{p}_t)_{t \geq 0}$ be a \mathcal{M}_∞^n -valued solution to the ODE (I.2.1.66). Then

$$\mathbb{E}[\bar{\phi}^n(\mathbf{p}_0, \bar{Y}_t)] = \mathbb{E}[\bar{\phi}^n(\mathbf{p}_t, \bar{Y}_0)] \quad (t \geq 0). \quad (\text{I.2.1.69})$$

The proof of Theorem I.2.1.19 can be found in section I.2.6.4.

I.2.1.7 Interpretation of the moments

In the present section, we will assume that the random variables X_0^1, \dots, X_0^n which form \mathbf{X}_0 are independent and identically distributed (for any $n, N \in \mathbb{N}$) which will significantly simplify the equations (I.2.1.66). For any $n, k \in \mathbb{N}$ such that $k \leq n$ let $\{0, 1\}_k^n := \{\sigma \in \{0, 1\}^n; \sum_{i=1}^n \sigma(i) = k\}$ be the subset of $\{0, 1\}^n$ where each element is a permutation of $\{0\}^{n-k} \times \{1\}^k$. Then for $(p_t)_{t \geq 0}$ as in (I.2.1.66) we have that

$$\mathbf{p}_t(\sigma) = \mathbf{p}_t(\sigma'), \quad t \geq 0, k \leq n, \sigma, \sigma' \in \{0, 1\}_k^n.$$

In other words $\mathbf{p}_t(\sigma)$ depends only on the number of indices $i = 1, \dots, n$ for which $\sigma(i) = 1$ and we can therefore define

$$\begin{aligned} \bar{p}_t(l, n) &:= \mathbf{p}_t(\sigma), \quad \sigma \in \{0, 1\}_{n-l}^n \\ \bar{p}_t(n) &:= \mathbf{p}_t(\mathbf{0}, n). \end{aligned} \quad (\text{I.2.1.70})$$

Here, $\bar{p}_t(l, n)$ is the limit as $N \rightarrow \infty$ of the ratio of sites at which $n - l$ coupled processes are occupied.

Proposition I.2.1.20 For each $n \in \mathbb{N}$ and $N \in \mathbb{N}$ let $(\mathbf{X}_t^N)_{t \geq 0} = (X_t^1, \dots, X_t^n)$ be a Markov process defined as in (I.2.1.59) and assume that X_0^1, \dots, X_0^n are independent and identically distributed. Assume that for any $n \in \mathbb{N}$, $\overline{\mathbf{X}}_0^N$ converges to $\mathbf{p}_0 \in \mathcal{M}_\infty^n$ in probability as $N \rightarrow \infty$. Let μ_0 be a probability measure on $[0, 1]$ whose k -th moment is equal to $p_0(k)$. Then $\bar{p}_t(n)$ is the solution of (I.2.1.35) and

$$\bar{p}_t(l, n) = \mathbb{E}[X^{n-l}(1-X)^l], \quad (\text{I.2.1.71})$$

for a μ_t -distributed random variable X , where $(\mu_t)_{t \geq 0}$ is a solution to (I.2.1.33).

The proof of Proposition I.2.1.20 can be found in the subsection I.2.6.8.

Proposition I.2.1.20 tells us that the k -th moment of μ_t is the limit as $N \rightarrow \infty$ of the probability that k cooperative branching processes on a complete graph with N vertices, which are started in independent and identically distributed initial states and coupled using the same graphical representation all simultaneously occupy a given site. For example, for $n = 2$, we have by Proposition (I.2.1.20) that

$$\begin{aligned} \mathbf{p}_t(\{1, 1\}) &= \bar{p}_t(2) = \int x^2 \mu_t(dx), \\ \mathbf{p}_t(\{0, 1\}) = \mathbf{p}_t(\{1, 0\}) &= \bar{p}_t(1, 2) = \int x(1-x) \mu_t(dx), \\ \mathbf{p}_t(\{0, 0\}) &= \bar{p}_t(2, 2) = \int (1-x)^2 \mu_t(dx). \end{aligned} \quad (\text{I.2.1.72})$$

These quantities have the following interpretation. Consider two mean-field models X^1, X^2 on $\Lambda_N = \{1, \dots, N\}$ that are coupled using the same graphical representation and let X_0^1 and X_0^2 be independent and identically distributed (the last assumption is necessary for the symmetry between $\mathbf{p}_t(\{0, 1\})$ and $\mathbf{p}_t(\{1, 0\})$). By Proposition I.2.1.16, such coupled processes are naturally dual to Y . Now $\mathbf{p}_t(0, 1)$, for example, is the limit as $N \rightarrow \infty$ of the probability that at time t , a given site $i \in \Lambda_N$ has $X_t^1(i) = 0$ and $X_t^2(i) = 1$.

I.2.2 Recursive tree processes

In this section, we will see the connections between the measure-valued function $(\mu_t)_{t \geq 0}$ (as well as the mean-field dual process $(\bar{Y}_t)_{t \geq 0}$) and recursive tree processes studied by Aldous and Bandyopadhyay in [AB05]. We start with defining and constructing a recursive tree process which corresponds to the limit $\lim_{t \rightarrow \infty} \mu_t$ of the measure-valued function $(\mu_t)_{t \geq 0}$.

I.2.2.1 A recursive tree process in discrete time

We define $g : \{0, 1\} \times \{0, 1\}^3 \rightarrow \{0, 1\}$ for any $\theta, x, y, z \in \{0, 1\}$ as

$$g_\theta(x, y, z) := \begin{cases} 0 & \text{if } \theta = 0, \\ (x \wedge y) \vee z & \text{if } \theta = 1. \end{cases} \quad (\text{I.2.2.1})$$

Comparing (I.2.2.1) to (I.1.2.12) and (I.1.2.9), we see that $g_1(x, y, z) = \text{coop}_{ijk}(b)(k)$ for $b \in \mathcal{S}(\Lambda)$ such that $b(i) = x$, $b(j) = y$ and $b(k) = z$ and that $g_0(x, y, z) = \text{death}_i(b)(i)$ (for any $b \in \mathcal{S}(\Lambda)$).

Denote by $\mathcal{M}_1(\{0, 1\})$ the space of all probability measures on $\{0, 1\}$, i.e. the space of all Bernoulli distributions. For any $\mu \in \mathcal{M}_1(\{0, 1\})$ let X_μ^1, X_μ^2 and X_μ^3 be i.i.d. random variables

with law μ . For a fixed $\alpha \geq 0$ let Θ be a Bernoulli distributed random variable with parameter $\alpha/(\alpha+1)$ and independent of X_μ^1, X_μ^2 and X_μ^3 and let $T_\alpha : \mathcal{M}_1(\{0, 1\}) \rightarrow \mathcal{M}_1(\{0, 1\})$ be defined as

$$T_\alpha(\mu) := \mathbb{P}[g_\Theta(X_\mu^1, X_\mu^2, X_\mu^3) \in \cdot]. \quad (\text{I.2.2.2})$$

Using the terminology of [AB05], we will call the equation

$$T_\alpha(\mu) = \mu \quad (\text{I.2.2.3})$$

the *recursive distributional equation* or RDE. We notice that the solutions of (I.2.2.3) correspond to the solutions of the mean-field equation (I.2.1.17). Indeed, if μ is a solution of (I.2.2.3), then

$$\mathbb{E}[g_\Theta(X_\mu^1, X_\mu^2, X_\mu^3)] = \frac{\alpha}{\alpha+1}(p + (1-p)p^2), \quad (\text{I.2.2.4})$$

where $p = \mathbb{P}[X_1^\mu = 1]$. Since (I.2.2.4) has the same dynamics as (I.2.1.17), both equations have the same fixed points and since the solution μ of (I.2.2.3) is a Bernoulli distribution with parameter p , we see that the parameters of these solutions are equal to the fixed points of (I.2.1.17). In other words, if $\alpha < 4$ then $\mu = \text{Ber}(0)$ is the only solution of (I.2.2.3) and if $\alpha \geq 4$, then μ is a solution of (I.2.2.3) if and only if $\mu = \text{Ber}(x_i)$, $i \in \{0, 1, 2\}$, with $x_0 = 0$ and x_i , $i \in \{1, 2\}$ as in (I.2.1.27), where by $\text{Ber}(p)$ we denote the Bernoulli distribution with parameter p .

We will now follow [AB05] and construct a so called *recursive tree process* or RTP corresponding to a solution of (I.2.2.3). Let \mathbb{T} be the space of all finite words $\mathbf{i} = i_1 i_2 \cdots i_n$, $n \geq 0$ made of letters from the alphabet $\{1, 2, 3\}$. By \emptyset we denote the empty word. We interpret \mathbb{T} as a ternary tree with root \emptyset and for $n \in \mathbb{N}$ and $\mathbf{i} = i_1 i_2 \cdots i_n \in \mathbb{T}$ we interpret the word $\mathbf{i} i_{n+1} = i_1 i_2 \cdots i_n i_{n+1}$ as the i_{n+1} -th child of the parent \mathbf{i} . We denote by $|\mathbf{i}| = |i_1 i_2 \cdots i_n| := n$ the length of the word \mathbf{i} . We let $\mathbb{T}_d := \{\mathbf{i} \in \mathbb{T} : |\mathbf{i}| \leq d\}$ be the subspace of all words of length at most d , which we can interpret as the set of individuals from the first d generations and $\mathbb{T}_d^* := \{\mathbf{i} \in \mathbb{T} : |\mathbf{i}| = d\}$ the set of all words of length exactly d , i.e. the collection of individuals from the d -th generation. Let μ be a probability measure on $\{0, 1\}$. For each $d \geq 0$ we can construct collections of random variables $(\Theta_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_{d-1}}$ and $(\Theta_{\mathbf{i}}, X^{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_d}$ such that

- (i) $(\Theta_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_{d-1}}$ are i.i.d. with law $\text{Ber}(\alpha/(\alpha+1))$.
- (ii) $(X^{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_d^*}$ are i.i.d. with law μ and independent of $(\Theta_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_{d-1}}$
- (iii) $X^{\mathbf{i}} = g_{\Theta_{\mathbf{i}}}(X^{\mathbf{i}1}, X^{\mathbf{i}2}, X^{\mathbf{i}3})$ for all $\mathbf{i} \in \mathbb{T}_{d-1}$.

It follows for each $0 \leq d' \leq d$ that

$$(X^{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_{d'}^*} \text{ are i.i.d. with law } T_\alpha^{d-d'}(\mu). \quad (\text{I.2.2.5})$$

If μ is furthermore a solution of (I.2.2.3) then by Kolmogorov's extension theorem there exists a collection of random variables $(\Theta_{\mathbf{i}}, X^{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_d}$ which is unique in distribution and such that

- (i) $(\Theta_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ are i.i.d. with law $\text{Ber}(\alpha/(\alpha+1))$.
- (ii) $(X^{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_d^*}$ are i.i.d. with law μ for all $d \geq 0$
- (iii) $X^{\mathbf{i}} = g_{\Theta_{\mathbf{i}}}(X^{\mathbf{i}1}, X^{\mathbf{i}2}, X^{\mathbf{i}3})$ for all $\mathbf{i} \in \mathbb{T}$.

Using the same terminology as [AB05], we will call $(\Theta_{\mathbf{i}}, X^{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_d}$ the recursive tree process (corresponding to μ). Put

$$\mathcal{G}_\emptyset := \sigma\{\Theta_{\mathbf{i}} : \mathbf{i} \in \mathbb{T}\}. \quad (\text{I.2.2.6})$$

By Definition 7 of [AB05] a recursive tree process is called *endogenous* if X_\emptyset is \mathcal{G}_\emptyset -measurable. In other words, a recursive tree process is endogenous if the knowledge of $(\Theta_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ is enough to determine the value X_\emptyset at the root of the tree.

Let $n \in \mathbb{N}$ and let $\nu \in \mathcal{M}_1(\{0, 1\}^n)$ be a probability distribution on $\{0, 1\}^n$. Denote by $(X^{i,j}, j \in \{1, \dots, N\})_{i \in \{1, 2, 3\}}$ a collection of i.i.d. random variables with law ν which are independent of Θ . Then we define $T_\alpha^{(n)} : \mathcal{M}_1(\{0, 1\}^n) \rightarrow \mathcal{M}_1(\{0, 1\}^n)$ as

$$T_\alpha^{(n)}(\nu) := \mathbb{P}[(g_\Theta(X^{1,j}, X^{2,j}, X^{3,j}), j \in \{1, \dots, n\}) \in \cdot]. \quad (\text{I.2.2.7})$$

For any $\mu \in \mathcal{M}_1(\{0, 1\})$, define $\mu^\nearrow \in \mathcal{M}_1(\{0, 1\}^2)$ as $\mu^\nearrow := \mathbb{P}[(X, X) \in \cdot]$, where X is a μ -distributed random variable. We say that the recursive tree process corresponding to $\mu \in \mathcal{M}_1(\{0, 1\})$ has the *bivariate uniqueness* property if μ^\nearrow is the only fixed point of $T^{(2)}$ in the space $\mathcal{P}_\mu^{(2)}$ of probability laws on $\{0, 1\}^2$ whose first and second marginals are both equal to μ . Aldous and Bandyopadhyay show that bivariate uniqueness and endogeneity are equivalent. Specifically, the following holds.

Theorem I.2.2.1 *The following conditions are equivalent:*

- (i) *The recursive tree process corresponding to μ is endogenous.*
- (ii) *The recursive tree process corresponding to μ has the bivariate uniqueness property.*
- (iii) *$(T^{(2)})^n(\mu \otimes \mu) \Rightarrow \mu^\nearrow$ as $n \rightarrow \infty$, where \Rightarrow denotes weak convergence.*

Proof This is a special case of Theorem 11 of [AB05]. ■

Assume that $\alpha > 4$ and let $\mu = \text{Ber}(x_i)$ for some $i \in \{0, 1, 2\}$, where $x_0 = 0$ and x_1 and x_2 are as in (I.2.1.27). Let $(X_t^1)_{t \geq 0}$ and $(X_t^2)_{t \geq 0}$ be two coupled cooperative branching processes on a complete graph with N vertices which are started in i.i.d. initial states such that for each $j = 1, 2$, X_0^j has a distribution given by a product measure such that $X_0^j(i) \sim \mu$, $i \in \{1, \dots, N\}$ as in Subsection I.2.1.6. Then from Subsection I.2.1.6 we see that $(T^{(2)})^n(\mu \otimes \mu)$ is the discrete-time analogue of the limit as $N \rightarrow \infty$ of the distribution of local state $(X_t^1(i), X_t^2(i))$ at an arbitrary fixed site $i \in \{1, \dots, N\}$. We see by Proposition I.2.1.20 that this distribution is characterized by the first two moments (see (I.2.1.35)) of $(\mu_t)_{t \geq 0}$. In particular, the limit $\mu_{T^\infty}^{(2)} := \lim_{n \rightarrow \infty} (T^{(2)})^n(\mu \otimes \mu)$ exists and is characterized by $\mu_{T^\infty}^{(2)}(\{(0, 1)\} \cup \{(1, 1)\}) = p_\infty(1)$ and $\mu_{T^\infty}^{(2)}(\{(1, 1)\}) = p_\infty(2)$, where $p_\infty(k) := \lim_{t \rightarrow \infty} p_t(k)$, $k = 1, 2$ with $p_t(k)$ as in (I.2.1.35) and such that $p_0(1) = x_i = \mu(\{1\})$ and $p_0(2) = x_i^2 = \mu(\{1\})^2$. Here $p_\infty(1)$ corresponds to the limit (taken first as $N \rightarrow \infty$ and then $t \rightarrow \infty$) of the probability that $X_t^2(i) = 1$ and $p_\infty(2)$ is the limit of the probability that $X_t^1(i) = X_t^2(i) = 1$. We notice that $\mu_{T^\infty}^{(2)}$ is equal to μ^\nearrow if and only if $p_\infty(1) = p_\infty(2)$. From Theorem I.2.1.13 we see that this is the case only for $\mu = \text{Ber}(x_0) = \delta_0$ and $\mu = \text{Ber}(x_2)$, since ν_0 and ν_2 as in (I.2.1.53) are the only fixed points whose first and second moments are equal. When $\mu = \text{Ber}(x_1)$, we see from (I.2.1.56) that $(p_\infty(1), p_\infty(2)) = (x_1, m_2)$. In other words, the recursive tree processes corresponding to $\mu = \text{Ber}(x_0)$ and $\mu = \text{Ber}(x_2)$ are endogenous, whereas the RTP corresponding to $\mu = \text{Ber}(x_1)$ is not. In the next subsection, we conjecture a generalization of the concept of recursive tree processes to continuous time.

I.2.2.2 A recursive tree process in continuous time

We now let $(\sigma_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ be a collection of i.i.d. exponentially distributed random variables with $\mathbb{P}[\sigma_{\mathbf{i}} > t] = e^{-(\alpha+1)t}$ for all $t \geq 0$ and we assume that $(\sigma_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ is independent of the random variables $(\Theta_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$. We interpret $\sigma_{\mathbf{i}}$ as the amount of time the individual \mathbf{i} lives. We write $\mathbf{i} \prec \mathbf{j}$ if $\mathbf{j} = \mathbf{i}\mathbf{k}$ for some $\mathbf{k} \neq \emptyset$, i.e. if \mathbf{i} is an ancestor of \mathbf{j} . We let $\tau_{\mathbf{i}}^- := \sum_{\mathbf{j} \prec \mathbf{i}} \sigma_{\mathbf{j}}$ be the time the individual \mathbf{i} is born and $\tau_{\mathbf{i}}^+ := \tau_{\mathbf{i}}^- + \sigma_{\mathbf{i}}$ the time the individual \mathbf{i} dies. Analogously to the discrete-time case of the previous subsection, for each $t \geq 0$ we denote by $\mathbb{T}_t^* := \{\mathbf{i} \in \mathbb{T} : t \in [\tau_{\mathbf{i}}^-, \tau_{\mathbf{i}}^+)\}$ the (random) collection of individuals who are alive at time t and by $\mathbb{T}_t := \bigcup_{0 \leq s \leq t} \mathbb{T}_s^*$ we denote the collection of all individuals alive at time t and all their ancestors.

For a fixed $\mu \in \mathcal{P}(\{0, 1\})$ and $t > 0$, conditionally on $(\Theta_{\mathbf{i}}, \sigma_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ we let $(X^{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_t^*}$ be i.i.d. random variables with law μ . We define $X^{\mathbf{i}}$, $\mathbf{i} \in \mathbb{T}_t$ inductively so that

$$X^{\mathbf{i}} = g_{\Theta_{\mathbf{i}}}(X^{\mathbf{i}1}, X^{\mathbf{i}2}, X^{\mathbf{i}3}), \quad \mathbf{i} \in \mathbb{T}_t \setminus \mathbb{T}_t^*. \quad (\text{I.2.2.8})$$

In the light of (I.2.2.5), we conjecture that the continuous-time analogue holds, namely that

$$(X^{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}_s^*} \text{ are i.i.d. with law } \mu_{t-s}, \quad (\text{I.2.2.9})$$

where $(\mu_s)_{0 \leq s \leq t}$ is a solution to the differential equation

$$\frac{\partial}{\partial s} \mu_s = (\alpha + 1)(T_{\alpha}(\mu_s) - \mu_s), \quad 0 \leq s \leq t. \quad (\text{I.2.2.10})$$

With g defined as in (I.2.2.1), the equation (I.2.2.10) has the form

$$\frac{\partial}{\partial s} \mu_s = (\delta_0 - \mu_s) + \alpha(\psi'(\mu_t) - \mu_t), \quad (\text{I.2.2.11})$$

where we define $\psi' : \mathcal{M}_1(\{0, 1\}) \rightarrow \mathcal{M}_1(\{0, 1\})$ as

$$\psi'(\mu) := \mathbb{P}[(X^1 \wedge X^2) \vee X^3 \in \cdot] \quad \text{if } X^1, X^2, X^3 \text{ are i.i.d. with common law } \mu.$$

Here we interpret the measure-valued equation (I.2.2.11) in the same sense we interpreted (I.2.1.33). Specifically, we say that a function $[0, t] \ni s \rightarrow \mu_s \in \mathcal{M}_1(\{0, 1\})$ solves (I.2.2.11), if for any function $f : \{0, 1\} \rightarrow \mathbb{R}$ the function $[0, t] \ni s \mapsto \langle \mu_s, f \rangle = \int f d\mu_s \in \mathbb{R}$ is continuously differentiable and solves

$$\frac{\partial}{\partial s} \langle \mu_s, f \rangle = (\langle \delta_0, f \rangle - \langle \mu_s, f \rangle) + \alpha(\langle \psi'(\mu_t), f \rangle - \langle \mu_s, f \rangle).$$

Notice that a $\mu \in \mathcal{M}_1(\{0, 1\})$ is a fixed point of (I.2.2.11) if and only if it is solution of the RDE (I.2.2.3). This means in particular that if $(\Theta_{\mathbf{i}}, X^{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ is a recursive tree process corresponding to a solution of the RDE (I.2.2.3), then for each $t \geq 0$, conditionally on $(\Theta_{\mathbf{i}}, \sigma_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$, the random variables $X^{\mathbf{i}}$, $\mathbf{i} \in \mathbb{T}_t^*$ are i.i.d. with law μ .

We consider the differential equation

$$\frac{\partial}{\partial t} \mu_t = (\alpha + 1)(T_{\alpha}^{(2)}(\mu_t) - \mu_t), \quad t \geq 0. \quad (\text{I.2.2.12})$$

and in the light of Theorem I.2.2.1 we conjecture that the continuous analogue of Theorem I.2.2.1 holds:

Conjecture I.2.2.2 *The following is equivalent:*

- (i) *The recursive tree process corresponding to μ is endogenous.*

(ii) μ^\nearrow is the only fixed point of (I.2.2.12) in $\mathcal{P}_\mu^{(2)}$.

(iii) The solution of (I.2.2.12) started in $\mu_0 = \mu \otimes \mu$ converges weakly to μ^\nearrow as $t \rightarrow \infty$.

Let $\alpha > 4$. Note that if the Conjecture I.2.2.2 holds, then by the continuous-time version of the arguments in the last paragraph of the previous section, the recursive tree processes corresponding to $\mu = \text{Ber}(x_0)$ and $\mu = \text{Ber}(x_2)$ are endogenous and RTP corresponding to $\mu = \text{Ber}(x_1)$ is not. Indeed, the solution $(\mu_t)_{t \geq 0}$ of (I.2.2.12) is equal to the limit as $N \rightarrow \infty$ of the distribution of local states at a fixed site of two coupled cooperative branching processes on a complete graph with N vertices started in the same i.i.d. initial states as in the previous subsection.

I.2.2.3 A recursive tree process of random probability distributions

In this section we will show how the fixed points of (I.2.1.33) correspond to recursive tree processes. First, we define $\bar{g}_\theta : \{0, 1\} \times [0, 1]^3 \rightarrow [0, 1]$ for $\theta \in \{0, 1\}$ and $\omega_1, \omega_2, \omega_3 \in [0, 1]$ by

$$\bar{g}_\theta(\omega_1, \omega_2, \omega_3) := \begin{cases} 0 & \text{if } \theta = 0, \\ \omega_1 + (1 - \omega_1)\omega_2\omega_3 & \text{if } \theta = 1. \end{cases} \quad (\text{I.2.2.13})$$

Let $\nu \in \mathcal{M}_1[0, 1]$ and let $\omega_1, \omega_2, \omega_3$ be i.i.d. random variables with law μ and Θ an independent random variable with law $\text{Ber}(\alpha/(\alpha + 1))$. Analogously to (I.2.2.7) we define $\bar{T}_\alpha : \mathcal{M}_1[0, 1] \rightarrow \mathcal{M}_1[0, 1]$ as

$$\bar{T}_\alpha(\nu) := \mathbb{P}[\bar{g}_\Theta(\omega_1, \omega_2, \omega_3) \in \cdot]. \quad (\text{I.2.2.14})$$

Now, we denote by

$$\mathcal{G}_i := \sigma\{\Theta_{ij} : j \in \mathbb{T}\}, \quad i \in \mathbb{T}. \quad (\text{I.2.2.15})$$

the σ -algebra generated by all random variables Θ_k corresponding to the individuals k whose ancestor is the individual i . Note that this is a generalization of \mathcal{G}_\emptyset which we defined in Subsection I.2.2.1. If $(\Theta_i, X^i)_{i \in \mathbb{T}}$ is a recursive tree process corresponding to $\mu = \text{Ber}(x_k)$, $k \in \{0, 1, 2\}$, then we define $((\omega_i)_{i \in \mathbb{T}}) \in [0, 1]^{\mathbb{T}}$ by

$$\omega_i := \mathbb{P}[X^i = 1 | \mathcal{G}_i], \quad i \in \mathbb{T}. \quad (\text{I.2.2.16})$$

If $\mu \in \mathcal{M}_1(\{0, 1\})$ is a solution of (I.2.2.3), then $(\Theta_i, \omega_i)_{i \in \mathbb{T}}$ is itself a recursive tree process corresponding to a $\nu = \nu^\mu \in \mathcal{M}_1[0, 1]$. If the recursive tree process corresponding to μ is endogenous, then X_i is \mathcal{G}_i measurable, hence $\omega_i = X^i$ for all $i \in \mathbb{T}$ and the two recursive tree processes are identical. However, this is no longer the case in the non-endogenous case. Let $\alpha > 4$ and $\mu = \text{Ber}(x_1)$, in which case the RTP $(\Theta_i, X^i)_{i \in \mathbb{T}_d}$ is not endogenous. Consider the differential equation

$$\frac{\partial}{\partial t} \mu_t = (\alpha + 1)(\bar{T}_\alpha(\mu) - \mu_t), \quad t \geq 0. \quad (\text{I.2.2.17})$$

From (I.2.2.14), we immediately see that (I.2.2.17) is the same as (I.2.1.33). Since the space $[0, 1]$ is isomorphic to $\mathcal{P}(\{0, 1\})$ and in the light of Proposition I.2.1.10, it makes sense to view the solutions $(\mu_t)_{t \geq 0}$ of (I.2.2.17) at each $t \geq 0$ as elements of $\mathcal{M}_1(\mathcal{M}_1(\{0, 1\}))$, i.e. as probability distributions taking values in the set of probability distributions on $\{0, 1\}$. Let ω_t be $\mathcal{M}_1(\{0, 1\})$ -valued μ_t -distributed random variable. For each $n \in \mathbb{N}$ we define the n -th moment measure $\mu_t^{(n)} \in \mathcal{M}_1(\{0, 1\}^n)$ of μ_t by the formula

$$\mu_t^{(n)}(\{x_1, \dots, x_n\}) := \mathbb{E}[\omega_t(\{x_1\}) \cdots \omega_t(\{x_n\})], \quad x = (x_1, \dots, x_n) \in \{0, 1\}^n. \quad (\text{I.2.2.18})$$

We see from Proposition I.2.1.20 that the n -th moment measure of a solution of (I.2.2.17) solves the n -variate equation

$$\frac{\partial}{\partial t} \mu_t = (\alpha + 1)(T_\alpha^{(n)}(\mu_t) - \mu_t). \quad (\text{I.2.2.19})$$

In particular, as we have already seen in the previous subsection, the second moment measures solve the bivariate equation (I.2.2.12).

In the next section we will see that the duality between the coupled cooperative branching processes on a complete graph with N vertices and the dual process $(\bar{Y}_t^N)_{t \geq 0}$ with the jump rates I.2.1.12 characterizes the distribution of the minimal elements of $(\bar{Y}_t^N)_{t \geq 0}$.

I.2.3 Distribution determining functions

This section is an excerpt from yet unpublished results of Jan Swart who provided all statements and proofs of statements of Section I.2.3.

Lemma I.2.3.1 (Distribution of the forward process) *Let Λ be a countable set and let X, X' be random variables with values in $\mathcal{S}(\Lambda)$. Assume that*

$$\mathbb{E}[\chi(X, \{y\})] = \mathbb{E}[\chi(X', \{y\})] \quad (y \in \mathcal{S}_{\text{fin}}(\Lambda)). \quad (\text{I.2.3.1})$$

Then X and X' are equal in distribution.

The proof of Lemma I.2.3.1 can be found in subsection I.2.6.9.

Lemma I.2.3.1 shows that the duality of Proposition I.1.4.1 uniquely determines the law of the process X . We next ask whether it also determines the law of Y . Let now Λ be a countable lattice and let $E \in \mathcal{H}(\Lambda)$. Recall that an element $e \in E$ is called *minimal* if $\nexists f \in E$ s.t. $f \leq e$, $f \neq e$. Let $E^\circ := \{e \in E : e \text{ is minimal}\}$. It is easy to see that $\forall e \in E \exists e' \in E^\circ$ s.t. $e' \leq e$, and as a result

$$\chi(x, E) = \chi(x, E^\circ) \quad (x \in \mathcal{S}(\Lambda), E \in \mathcal{H}(\Lambda)). \quad (\text{I.2.3.2})$$

Thus, in a sense, only the set Y_t° of minimal elements of the dual process Y_t really matters and it is clear from I.2.3.2 that the duality of Proposition I.1.4.1 cannot determine the law of $(Y_t)_{t \geq 0}$. If Λ is a finite set then by Lemma 1 of [SS16] the process $(Y_t^\circ)_{t \geq 0}$ is itself a Markov process with state space $\mathcal{H}^\circ(\Lambda) := \{E^\circ : E \in \mathcal{H}(\Lambda)\}$ (and using the Proposition 30 of [SS16] the same can be shown for Λ which is only countable) and by Proposition I.1.4.1 and (I.2.3.2), the process $(Y_t^\circ)_{t \geq 0}$ of minimal elements of the dual process Y_t is also dual to $(X_t)_{t \geq 0}$ with the duality function χ . Therefore, we can ask the same question about Y_t° . However, the next lemma shows that even the marginals of the process Y° are not determined by the duality.

Lemma I.2.3.2 (Insufficiently many functions) *Let Λ be a countable set with at least two elements. Then there exist random variables Y, Y' with values in $\mathcal{H}^\circ(\Lambda)$ such that*

$$\mathbb{E}[\chi(x, Y)] = \mathbb{E}[\chi(x, Y')] \quad (x \in \mathcal{S}(\Lambda)), \quad (\text{I.2.3.3})$$

and such that Y and Y' are not equal in distribution.

The proof of Lemma I.2.3.2 can be found in subsection I.2.6.10.

Recall that $\Lambda_N = \{1, \dots, N\}$ and note that if $E \in \mathcal{H}^\circ(\Lambda_N)$ and $E \sim E'$, then also $E' \in \mathcal{H}^\circ(\Lambda_N)$. Hence, we can define

$$\bar{\mathcal{H}}^\circ(\Lambda_N) := \{\bar{E} : E \in \mathcal{H}^\circ(\Lambda_N)\}. \quad (\text{I.2.3.4})$$

analogously to how we defined $\overline{\mathcal{H}}(\Lambda_N)$ in (I.2.1.8) and we can define the process $(\overline{Y}_t^\circ)_{t \geq 0} = (\overline{Y}_t^{\circ, N})_{t \geq 0}$ as $\overline{\mathcal{H}}^\circ(\Lambda_N)$ -valued process of equivalence classes of Y_t° analogously to how we defined the process $(\overline{Y}_t^N)_{t \geq 0}$ in (I.2.1.12). By (I.2.3.2), we see that the duality (I.2.1.62) holds also for the process $(\overline{Y}_t^\circ)_{t \geq 0}$. The following Lemma tells us that this duality characterizes the distribution of $(\overline{Y}_t^\circ)_{t \geq 0}$.

Lemma I.2.3.3 (Products determine the law) *Let Λ be a finite set and let Y, Y' be random variables with values in $\mathcal{H}^\circ(\Lambda)$. Assume that*

$$\mathbb{E}\left[\prod_{i=1}^n \chi(x_i, Y)\right] = \mathbb{E}\left[\prod_{i=1}^n \chi(x_i, Y')\right] \quad (\text{I.2.3.5})$$

for each $x_1, \dots, x_n \in \mathcal{S}_{\text{fin}}(\Lambda)$. Then Y and Y' are equal in distribution.

The proof of Lemma I.2.3.3 can be found in subsection I.2.6.11

Remark I.2.3.4 Since for any two hyperedges $y \leq y'$ we have that $y \subseteq y'$, it is easy to see that $\overline{\phi}^n(p, \{y, y'\}) \neq \emptyset$ if and only if $\overline{\phi}^n(p, \{y\}) \neq \emptyset$ for any $n \in \mathbb{N}$, and so Theorem I.2.1.19 also holds with $(\overline{Y}_t)_{t \geq 0}$ replaced by $(\overline{Y}_t^\circ)_{t \geq 0}$.

I.2.4 The two-sex model

So far, we have always considered the cooperative branching process to model a population of identical individuals. This is unrealistic for many biological populations where individuals are male and female and one individual of each sex is required to produce an offspring. In this section, we will discuss a modification of the model (I.1.2.16) which includes two sexes, so that an offspring can only be produced when the two parents are of opposite sex and provide some justification for why we study the simpler version of the cooperative branching process in the rest of the thesis instead. We will limit ourselves to a situation without random walk dynamics (i.e. $\gamma = 0$) so that the individuals only die and reproduce via the cooperative branching. We will consider three variants of the two-sex model (which, as we will see, all have very similar dynamics) on a complete graph and in their mean-field limits.

In the first model, two individuals of opposite sex can simultaneously occupy the same site and the only interaction between the sexes occurs during the cooperative branching events. Let $\{\{0, 1\}^2\}^\Lambda$ be the space of all functions $x = (m_x, f_x) : \Lambda \rightarrow \{0, 1\}^2$, where $m_x, f_x : \Lambda \rightarrow \{0, 1\}$. We interpret $m_x(i) = 1$ (resp. $m_x(i) = 0$) for some $i \in \Lambda$ as the site i being occupied (resp. unoccupied) by a male particle and analogously $f_x(i) = 1$ (resp. $f_x(i) = 0$) indicates that the site is occupied (resp. unoccupied) by a female particle. For each $i, j \in \Lambda$ let the function pair_{ij} be defined for $x \in \{\{0, 1\}^2\}^\Lambda$ as

$$\text{pair}_{ij}(x) := (m_x(i) \wedge f_x(j)) \vee (f_x(i) \wedge m_x(j)). \quad (\text{I.2.4.1})$$

Function pair_{ij} tells us whether there is a pair of two individuals of different sex on the sites

i and j . For each $i, j, k \in \Lambda$, we then define maps on $\{\{0, 1\}^2\}^\Lambda$ by

$$\begin{aligned}
\text{coop}^{1,\text{m}}_{ijk}(x)(l) &:= \begin{cases} (\text{pair}_{ij}(x) \vee m_x(k), f_x(l)) & \text{if } l = k, \\ x(l) & \text{otherwise,} \end{cases} \\
\text{coop}^{1,\text{f}}_{ijk}(x)(l) &:= \begin{cases} (m_x(l), \text{pair}_{ij}(x) \vee f_x(k)) & \text{if } l = k, \\ x(l) & \text{otherwise,} \end{cases} \\
\text{death}^{1,\text{m}}_i(x)(l) &:= \begin{cases} (0, f_x(l)) & \text{if } l = i, \\ x(l) & \text{otherwise,} \end{cases} \\
\text{death}^{1,\text{f}}_i(x)(l) &:= \begin{cases} (m_x(l), 0) & \text{if } l = i, \\ x(l) & \text{otherwise.} \end{cases}
\end{aligned} \tag{I.2.4.2}$$

Map $\text{coop}^{1,\text{m}}_{ijk}$ can be described in words as follows. If there is a pair of a male and a female particles at sites i and j and there is no male particle at site k , then the pair at i and j produces a male offspring at k . The map $\text{coop}^{1,\text{f}}_{ijk}$ is the same, except that now the offspring is female. Since in this model, two individuals can occupy the same site as long as their sex is different, we also introduce two separate death maps for each sex.

The generator of the two-sex process which uses the cooperative branching and death dynamics (I.2.4.2) has the following form

$$\begin{aligned}
Gf(x) &:= \alpha \frac{1}{D(D-1)} \sum_{\substack{i,j,k \in \Lambda \\ \langle i,j \rangle, \langle j,k \rangle \in E}} \{f(\text{coop}^{1,\text{m}}_{ikj}(x)) + f(\text{coop}^{1,\text{f}}_{ikj}(x)) - 2f(x)\} \\
&+ \beta \frac{1}{2(D-1)} \sum_{\substack{i,j,k \in \Lambda \\ \langle i,j \rangle, \langle j,k \rangle \in E}} \{f(\text{coop}^{1,\text{m}}_{ijk}(x)) + f(\text{coop}^{1,\text{f}}_{ijk}(x)) - 2f(x)\} \\
&+ \delta \sum_{i \in \Lambda} \{f(\text{death}^{1,\text{m}}_i(x)) + f(\text{death}^{1,\text{f}}_i(x)) - 2f(x)\}.
\end{aligned} \tag{I.2.4.3}$$

Since a male and a female particle can occupy the same site simultaneously, it would perhaps be more natural to define cooperative branching in such a way that pairs of male and female particles which occupy the same site produce offspring rather than those which occupy neighbouring sites. However, defined as in (I.2.4.3) the process resembles more the process with only one sex which we study in the rest of this section and since we only study the mean-field behaviour of the process, the two different ways of producing offspring have no influence on our results. Note that the parameters α and β in the two-sex model correspond to 2α and 2β in (I.1.2.16). This will be the case with the two other variants of the two-sex model as well.

We will now consider the model (I.2.4.3) on the complete graph $\Lambda = K_N$ with N vertices when $\beta = 0$ and $\delta = 1$. In the mean-field limit $N \rightarrow \infty$, we obtain (in the same way as in Proposition I.2.1.2) the following differential equations for the fractions $m(t) := m_X(t) := N^{-1} \sum_{i=1}^N m_X(t, i)$ and $f(t) := f_X(t) := N^{-1} \sum_{i=1}^N f_X(t, i)$ of sites occupied by male and female particles, respectively.

$$\begin{aligned}
\frac{\partial}{\partial t} m(t) &= \alpha m(t) f(t) (1 - m(t)) - m(t) \\
\frac{\partial}{\partial t} f(t) &= \alpha m(t) f(t) (1 - f(t)) - f(t).
\end{aligned} \tag{I.2.4.4}$$

For $\alpha < 4$, the only fixed point of the system (I.2.4.4) is (x_0, x_0) , where $x_0 = 0$. For $\alpha \geq 4$, there are two additional fixed points at (x_1, x_1) and (x_2, x_2) where x_1 and x_2 are as in (I.2.1.27).

When $\alpha > 4$, x_1 and x_2 are distinct and (x_1, x_1) is an unstable fixed point which separates the domains of attraction of the stable fixed points x_0 and (x_2, x_2) .

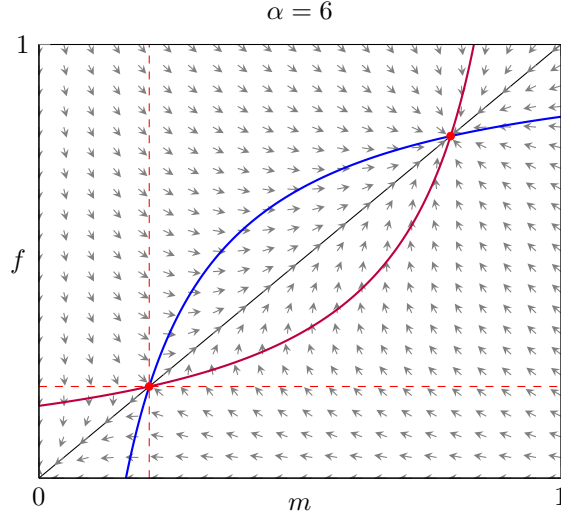


Figure I.2.1: A phase diagram of the mean-field model (I.2.4.4). The blue and the purple curve denote the points where the fraction of male and the fraction of female particles does not change, respectively.

In Figure I.2.1 we see the phase diagram of the mean-field limit. From any starting point, the process approaches the diagonal, i.e. the points in which the sex ratio is $1/2$, in other words where $m(t) = f(t)$. Once it reaches the diagonal, its dynamics is the same as the dynamics of the model (I.2.1.17). The process started in any point in the upper-right square (denoted by the dotted red lines) survives and is attracted to the upper stable point. If started in any of the points of the lower-left square, the process dies out. We conjecture that in each of the two rectangles on the sides, there exists a particular solution which converges to the lower unstable fixed-point (x_1, x_1) and whose path separates domains of attraction of the stable points (x_0, x_0) (extinction) and (x_2, x_2) (survival).

Now, we will consider a model in which there can only be a single individual per site. In this case, the model changes depending on whether we allow the offspring of a branching event to replace the particles present at the “birth site” or not. We will first study the model in which births are permitted only to those sites which are not occupied by a particle of a different sex than the offspring. This corresponds to the following maps.

$$\begin{aligned}
 \text{coop}^{2,m}_{ijk}(x)(l) &:= \begin{cases} ((\text{pair}_{ij}(x) \vee m_x(k)) \wedge (1 - f_x(k)), f_x(l)) & \text{if } l = k, \\ x(l) & \text{otherwise,} \end{cases} \\
 \text{coop}^{2,f}_{ijk}(x)(l) &:= \begin{cases} (m_x(l), (\text{pair}_{ij}(x) \vee f_x(k)) \wedge (1 - m_x(k))) & \text{if } l = k, \\ x(l) & \text{otherwise,} \end{cases} \\
 \text{death}^{2,i}(x)(l) &:= \begin{cases} (0, 0) & \text{if } l = i, \\ x(l) & \text{otherwise.} \end{cases}
 \end{aligned} \tag{I.2.4.5}$$

Since a birth to a site occupied by the same sex has no effect on the process, this effectively

describes a model in which births happen to empty sites only. This gives us the generator

$$\begin{aligned}
Gf(x) := & \alpha \frac{1}{D(D-1)} \sum_{\substack{i,j,k \in V \\ \langle i,j \rangle, \langle j,k \rangle \in E}} \{f(\text{coop}^{2,m}_{ikj}(x)) + f(\text{coop}^{2,f}_{ikj}(x)) - 2f(x)\} \\
& + \beta \frac{1}{2(D-1)} \sum_{\substack{i,j,k \in V \\ \langle i,j \rangle, \langle j,k \rangle \in E}} \{f(\text{coop}^{2,m}_{ijk}(x)) + f(\text{coop}^{2,f}_{ijk}(x)) - 2f(x)\} \\
& + \delta \sum_{i \in V} \{f(\text{death}^2_i(x)) - f(x)\}.
\end{aligned} \tag{I.2.4.6}$$

Again setting $\beta = 0$ and $\delta = 1$ and following the structure of the proof of Proposition (I.2.1.2) we obtain in the mean-field limit the following equations for the fractions of occupied sites.

$$\begin{aligned}
\frac{\partial}{\partial t} m(t) &= \alpha m(t) f(t) (1 - m(t) - f(t)) - m(t) \\
\frac{\partial}{\partial t} f(t) &= \alpha m(t) f(t) (1 - m(t) - f(t)) - f(t).
\end{aligned} \tag{I.2.4.7}$$

The dynamics of the model is similar to that of (I.2.4.4). If $\alpha < 8$, the only fixed point is again $(x_0, x_0) = (0, 0)$. For $\alpha \geq 8$, there exist two additional fixed points (x_1, x_1) and (x_2, x_2) , where $x_1 = \frac{1}{4}(1 - \sqrt{\frac{\alpha-8}{\alpha}})$ and $x_2 = \frac{1}{4}(1 + \sqrt{\frac{\alpha-8}{\alpha}})$. As in (I.2.4.4), (x_1, x_1) is an unstable fixed point, whereas the fixed points (x_0, x_0) and (x_2, x_2) are stable.

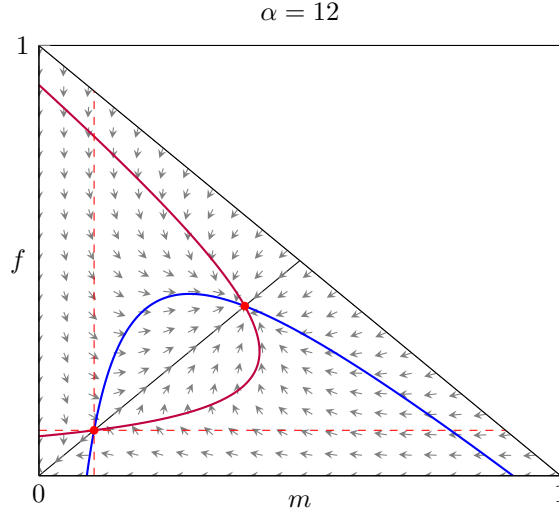


Figure I.2.2: A phase diagram of the mean-field model (I.2.4.7).

Finally, we will consider a variant of the two-sex model in which branching can occur even to the sites which are already occupied at which point the individual occupying that site is replaced by the offspring of the cooperative branching event. This kind of cooperative branching corresponds to the maps

$$\begin{aligned}
\text{coop}^{3,m}_{ijk}(x)(l) &:= \begin{cases} ((\text{pair}_{ij}(x) \vee m_x(k), (1 - \text{pair}_{ij}(x)) \wedge f_x(k)) & \text{if } l = k, \\ x(l) & \text{otherwise,} \end{cases} \\
\text{coop}^{3,f}_{ijk}(x)(l) &:= \begin{cases} ((1 - \text{pair}_{ij}(x)) \wedge m_x(k), (\text{pair}_{ij}(x) \vee f_x(k)) & \text{if } l = k, \\ x(l) & \text{otherwise.} \end{cases}
\end{aligned} \tag{I.2.4.8}$$

Since deaths occur the same way as in the model (I.2.4.6), we obtain the same generator as in (I.2.4.6), only with $\text{coop}^{2,m}_{ijk}$ and $\text{coop}^{2,f}_{ijk}$ replaced by $\text{coop}^{3,m}_{ijk}$ and $\text{coop}^{3,f}_{ijk}$, respectively. Setting $\beta = 0$ and $\gamma = 1$, the mean-field equations for the occupied sites are

$$\begin{aligned}\frac{\partial}{\partial t}m(t) &= \alpha m(t)f(t)(1 - m(t)) - \alpha f(t)m^2(t) - m(t) \\ \frac{\partial}{\partial t}f(t) &= \alpha m(t)f(t)(1 - f(t)) - \alpha f^2(t)m(t) - f(t),\end{aligned}\tag{I.2.4.9}$$

where the additional terms $\alpha f(t)m^2(t)$ and $\alpha f^2(t)m(t)$ correspond to the rate at which individuals are being replaced by the branching event offspring of the opposite sex. In this model, the fixed points are exactly the same (and appear for the same values of α) as in the model (I.2.4.7).

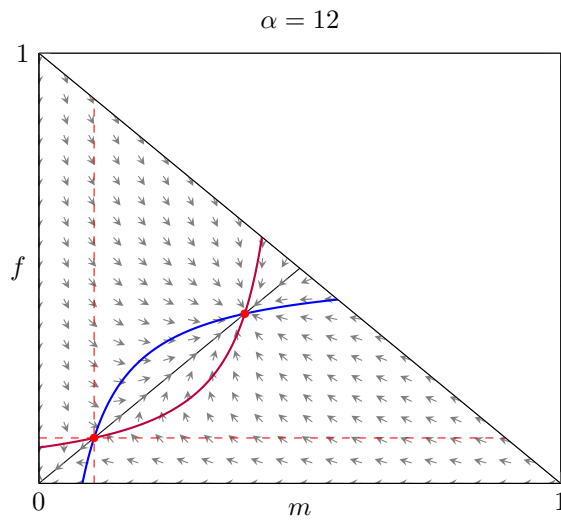


Figure I.2.3: A phase diagram of the mean-field model (I.2.4.9).

We see that all three considered models, at least in the mean-field limits (I.2.4.4), (I.2.4.7) and (I.2.4.9), the ratio of the two sexes approaches 1/2 and has the same dynamics as the model with identical particles once it does so. Hence, while from a biological perspective, the model with two sexes is more realistic for modelling populations of many species, the mean-field behaviour limit behaviour indicates that its asymptotic behaviour is similar to that of the simpler model I.1.2.16. Another reason not to study the two-sex models directly is that unless we permit male and female particles to inhabit the same site simultaneously, the model is no longer monotone and so many of the tools we used to study the model (I.1.2.16) are no longer applicable. Model (I.2.4.2) is monotone as the only interaction between male and female particles in model (I.2.4.2) happens through cooperative branching (unlike the other two two-sex models).

I.2.5 Summary and outlooks

In Chapter I.2 we studied the cooperative branching process on a complete graph with N vertices and particularly the mean-field dual process, which we obtained in I.2.1.3 as the limit of the dual process on the complete graph as $N \rightarrow \infty$. We have seen in Proposition I.2.1.2 that the ratio of occupied sites of the cooperative branching process on the complete graph converges to a solution of the differential equation (I.2.1.17) as $N \rightarrow \infty$ and in Proposition

I.2.1.4 we showed that there is a duality between the solutions of that equation and the mean-field dual process. In Proposition I.2.1.5 we used this duality to show the conditions under which the mean-field dual process survives. In Section I.2.1.4 we showed that the duality of the Proposition (I.2.1.4) is just a special case of a more general duality (I.2.1.36) between the mean-field dual process and a probability measure-valued function $(\mu_t)_{t \geq 0}$ defined in (I.2.1.33). In Section I.2.1.6 we provide an interpretation for the function $(\mu_t)_{t \geq 0}$ by showing that for each $k \in \mathbb{N}$ and $t \geq 0$ the k -th moment of μ_t arises naturally as a limit of the probability that a given site is occupied at time t simultaneously in k cooperative branching processes on the complete graph which are coupled using the same graphical representation and started in i.i.d. initial states. In particular, the first moment of $(\mu_t)_{t \geq 0}$ corresponds to the solution of (I.2.1.17). For the corresponding duality (I.2.1.62) the results of Jan Swart in Section I.2.3 show that this duality given for any number of coupled cooperative branching processes characterizes the distribution of the process \bar{Y}° of the minimal elements of the dual process.

In Section I.2.1.5, the fact that the duality of Proposition I.2.1.4 along with the fixed points (I.2.1.27) of equation (I.2.1.17) give rise to bounded martingales of the form (I.2.1.41), we show that the processes $(H_t^1)_{t \geq 0}$ and $(H_t^2)_{t \geq 0}$ generally have different limits as $t \rightarrow \infty$. We show that $H_\infty^2 = \lim_{t \rightarrow \infty} H_t^2$ is Bernoulli distributed with a success probability equal to the probability of survival of the mean-field dual process but we have not been able to retrieve the exact distribution of $H_\infty^1 = \lim_{t \rightarrow \infty} H_t^1$. These results lead us to study the fixed points of equation (I.2.1.33) and their domains of attraction. We saw that the only fixed points of (I.2.1.33) which have the first moment different from the fixed point x_1 of the equation (I.2.1.17) are ν_0 and ν_2 as given in (I.2.1.53), where ν_2 is also the distribution of H_∞^2 . We then showed that ν_1 (which is also defined in (I.2.1.53)) is an unstable fixed point of (I.2.1.33) and that there exists at least one additional fixed point ν which corresponds to the distribution of H_∞^1 .

In Section I.2.2 we studied the connections between the mean-field dual process and recursive tree processes studied by Aldous and Bandyopadhyay in [AB05]. In particular, we have shown that the fixed points ν_0 and ν_1 of (I.2.1.53) correspond to endogenous recursive tree processes and from the duality (I.2.1.36) between the mean-field dual process and the $\mathcal{M}_1[0, 1]$ -valued solutions of the equation (I.2.1.33) and based on the interpretation of its moments given by Proposition I.2.1.20 we conjecture that we can define a continuous time version of recursive tree processes with an analogous characterization of endogeny. Theorem I.2.1.13 also tells us that there exists a nonendogenous recursive probability distribution-valued tree process corresponding to ν_3 .

Finally, in Section I.2.4 we briefly discussed a variant of the cooperative branching processes with individuals of different sexes and saw that at least in the mean-field model, the behaviour of such processes does not fundamentally differ from the simpler model (I.1.2.14).

There are several questions that remain open. First, the conjectures (I.2.2.9) and I.2.2.2 should be shown rigorously. While Proposition I.2.1.11 gives us a partial insight into the form of the fixed point ν_3 , the exact value of ν_3 remains undetermined. Furthermore, we would like to show whether ν_i , $i = 0, 1, 2, 3$ are the only fixed points of (I.2.1.33) or whether there exist additional fixed points (and hence additional corresponding non-endogenous recursive tree processes). Finally, while the results provided by Jan Swart in Section I.2.3 show that the distribution of the process $(\bar{Y}_t^\circ)_{t \geq 0}$ of the minimal elements of the dual process $(\bar{Y}_t^N)_{t \geq 0}$ on the complete graph is characterized by the duality of the Proposition I.2.1.17, it remains to be shown whether the same can be said of the corresponding process of minimal elements of the mean-field dual process and the duality of Theorem I.2.1.19.

I.2.6 Proofs

We will group related proofs together so they do not necessarily appear in the same order as the statements that they prove. In Section (I.2.6.1) we prove Propositions I.2.1.1 and I.2.1.17. In fact, Proposition I.2.1.17 is just a multidimensional version of I.2.1.1. We stated the Proposition I.2.1.1 separately to avoid unnecessary notation before we need it.

I.2.6.1 Proofs of propositions I.2.1.1 and I.2.1.17

Proof of Proposition I.2.1.1 Recall that we define $T_\pi : \mathcal{S}(\Lambda_N) \rightarrow \mathcal{S}(\Lambda_N)$ as $(T_\pi x)(i) := x(\pi(i))$, $i \in \Lambda_N$ and let π be a permutation on Λ_N . Then by Proposition I.1.4.1 we see that

$$\mathbb{E}[\chi(T_\pi(X_t^N), Y_0^N)] = \mathbb{E}[\chi(T_\pi(X_0^N), Y_t^N)].$$

Hence, for a random Π uniformly distributed on the set S_N of all permutations on Λ_N and independent of both $X^N = (X_t^N)_{t \geq 0}$ and $Y^N = (Y_t^N)_{t \geq 0}$, we get

$$\mathbb{E}[\chi(T_\Pi(X_t^N), Y_0^N)] = \mathbb{E}[\chi(T_\Pi(X_0^N), Y_t^N)].$$

Since Π is independent of X^N and uniformly distributed we note that $B_{\overline{X}_t}^N$ has the same distribution as $T_\Pi(X_t^N)$. Since Π is also independent of Y^N , X_t^N is independent of Y_0^N and Y_t^N is independent of X_0^N , we see that both $T_\Pi(X_t^N)$ and $B_{\overline{X}_t}^N$ are independent of Y_0^N and $T_\Pi(X_0^N)$ and $B_{\overline{X}_0}^N$ are independent of Y_t^N . Therefore, we also get that

$$\mathbb{E}[\chi(B_{\overline{X}_t}^N, Y_0^N)] = \mathbb{E}[\chi(B_{\overline{X}_0}^N, Y_t^N)].$$

Since $\chi(b, E) = \mathbf{1}_{\{\text{Thin}_b(E) \neq \emptyset\}}$, we obtain (I.2.1.16). This completes the proof of Proposition I.2.1.1. \blacksquare

Proof of Proposition I.2.1.17 Let $\pi \in \Pi_N$. Recall the duality function χ from (I.1.4.6) (which can also be written as (I.2.1.14)). By Proposition I.2.1.16 we see that

$$\mathbb{E}[\Pi_{i=1}^n \chi(T_\pi(X_t^i), Y_0)] = \mathbb{E}[\Pi_{i=1}^n \chi(T_\pi(X_0^i), Y_t)]. \quad (\text{I.2.6.1})$$

Hence, for a random π uniformly distributed on the set S_N of all permutations on Λ_N and independent of both $\mathbf{X} = (\mathbf{X}_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$, we get that (I.2.6.1) also holds. Since π is independent of $\overline{\mathbf{X}}$ and uniformly distributed, it follows for each $k \in \{1, \dots, n\}$ that $B_{\overline{\mathbf{X}}_t}^{N,k}$ has the same distribution as $T_\pi(X_t^k)$. Since π is also independent of Y , $\overline{\mathbf{X}}$ is independent of Y_0 and Y_t is independent of $\overline{\mathbf{X}}_0$, we see that for each $k \in \{1, \dots, n\}$ both $T_\pi(X_t^k)$ and $B_{\overline{\mathbf{X}}_t}^{N,k}$ are independent of Y_0 and $T_\pi(X_0^k)$ and $B_{\overline{\mathbf{X}}_0}^{N,k}$ are independent of Y_t . Therefore, we also get that

$$\mathbb{E}[\Pi_{i=1}^n \chi(B_{\overline{\mathbf{X}}_t}^{N,k}, Y_0)] = \mathbb{E}[\Pi_{i=1}^n \chi(B_{\overline{\mathbf{X}}_0}^{N,k}, Y_t)].$$

Since $\Pi_{i=1}^n \chi(b_i, E) = \mathbf{1}_{\{\text{Thin}_{b_i}(E) \neq \emptyset, i \in \{1, \dots, n\}\}}$ for any $b = (b_1, \dots, b_n) \in (\Lambda_N)^n$ and $E \in \mathcal{H}(\Lambda_N)$, we obtain (I.2.1.62). \blacksquare

In Section I.2.6.2 we prove Propositions I.2.1.2 and I.2.1.18. Once again, Proposition I.2.1.18 is a multidimensional version of I.2.1.2.

I.2.6.2 Proof of propositions I.2.1.2 and I.2.1.18

Proof of Proposition I.2.1.2 We apply [Swa17, Thm 3.2] which in turn is a simple consequence of [DN08, Thm 4.1]. Let $S_N := \{0, \frac{1}{N}, \dots, 1\}$ and for $\bar{x}, \bar{y} \in S_N$, let

$$r_N(\bar{x}, \bar{y}) := \begin{cases} \alpha N(1 - \bar{x})\bar{x}(\bar{x} - \frac{1}{N}) & \text{if } \bar{y} = \bar{x} + \frac{1}{N}, \\ \delta N\bar{x} & \text{if } \bar{y} = \bar{x} - \frac{1}{N}, \\ 0 & \text{otherwise} \end{cases} \quad (\text{I.2.6.2})$$

denote the rate at which \bar{X}_t jumps from \bar{x} to \bar{y} . Define functions $\beta_N, \gamma_N : S_N \rightarrow \mathbb{R}$ by

$$\begin{aligned} \beta_N(\bar{x}) &:= \sum_{\bar{y} \in S_N} r_N(\bar{x}, \bar{y})(\bar{y} - \bar{x}), \\ \gamma_N(\bar{x}) &:= \sum_{\bar{y} \in S_N} r_N(\bar{x}, \bar{y})(\bar{y} - \bar{x})^2. \end{aligned} \quad (\text{I.2.6.3})$$

Note that β and γ describe the local drift and quadratic variation of \bar{X}_t . By [Swa17, Thm 3.2], in order to prove Proposition I.2.1.2, it suffices to check that

$$(i) \quad \sup_{\bar{x} \in S_N} |\beta_N(\bar{x}) - b(\bar{x})| \xrightarrow{N \rightarrow \infty} 0 \quad \text{and} \quad (ii) \quad \sup_{\bar{x} \in S_N} \gamma_N(\bar{x}) \xrightarrow{N \rightarrow \infty} 0, \quad (\text{I.2.6.4})$$

where $b : [0, 1] \rightarrow \mathbb{R}$ is the function

$$b(u) := \alpha(1 - u)u^2 - \delta u \quad (u \in [0, 1]). \quad (\text{I.2.6.5})$$

(Note that b is Lipschitz, which is one of the requirements of [Swa17, Thm 3.2].) We observe that

$$\beta_N(\bar{x}) = \alpha(1 - \bar{x})\bar{x}(\bar{x} - \frac{1}{N}) - \delta\bar{x} = \alpha(1 - \bar{x})\bar{x}^2 - \delta\bar{x} + O(\frac{1}{N}) \quad (\text{I.2.6.6})$$

and

$$\gamma_N(\bar{x}) = \frac{1}{N}(\alpha(1 - \bar{x})\bar{x}(\bar{x} - \frac{1}{N}) + \delta\bar{x}) = O(\frac{1}{N}), \quad (\text{I.2.6.7})$$

where $O(\frac{1}{N})$ denotes a term that can be estimated as $|O(\frac{1}{N})| \leq K/N$ where K is a constant that does not depend on \bar{x} . This completes the proof of Proposition I.2.1.2. \blacksquare

Proof of Proposition I.2.1.18 The proof is mostly analogous to that of Proposition I.2.1.2. Theorem 4.1. of [DN08] holds even in higher dimensions and so we can use Theorem 3.2. of [Swa17] with just a few small changes. We define functions $\beta_N : \mathcal{M}_N^n \rightarrow \mathbb{R}^{2^n}$ and $\gamma_N : \mathcal{M}_N^n \rightarrow \mathbb{R}^{2^n}$ analogously to I.2.6.3, the only difference being that these functions are now \mathbb{R}^{2^n} -dimensional. By $|\cdot|$ we now denote the Euclidean metric on \mathbb{R}^{2^n} and we require that

$$(i) \quad \sup_{\bar{\mathbf{x}} \in \mathcal{M}_N^n} |\beta_N(\bar{\mathbf{x}}) - b(\bar{\mathbf{x}})| \xrightarrow{N \rightarrow \infty} 0 \quad \text{and} \quad (ii) \quad \sup_{\bar{\mathbf{x}} \in \mathcal{M}_N^n} |\gamma_N(\bar{\mathbf{x}})| \xrightarrow{N \rightarrow \infty} 0, \quad (\text{I.2.6.8})$$

where $b : \mathcal{M}_\infty^n \rightarrow \mathbb{R}^{2^n}$ is defined as

$$\begin{aligned} b(\mathbf{p})(\sigma) &= \alpha \sum_{\sigma^- \leq \sigma, \sigma^- \neq \sigma} \mathbf{p}(\sigma^-) O^{\mathbf{p}}(\sigma, \sigma^-) \hat{O}^{\mathbf{p}}(\sigma, \sigma^-) \\ &\quad - \alpha \sum_{\sigma \leq \sigma^+, \sigma \neq \sigma^+} \mathbf{p}(\sigma) O^{\mathbf{p}}(\sigma^+, \sigma) \hat{O}^{\mathbf{p}}(\sigma^+, \sigma) \mathbf{p}(\sigma), \quad \sigma \in \{0, 1\}^n \setminus \mathbf{0}, \\ b(\mathbf{p})(\mathbf{0}) &= (1 - \mathbf{p}(\mathbf{0})) - \alpha \sum_{\sigma \neq \mathbf{0}} \mathbf{p}(\mathbf{0}) O^{\mathbf{p}}(\sigma, \mathbf{0}) \hat{O}^{\mathbf{p}}(\sigma, \mathbf{0}). \end{aligned} \quad (\text{I.2.6.9})$$

We see that b is Lipschitz (it is a bounded polynomial) and that $b(\mathbf{p}) + \mathbf{p}$ for all $\mathbf{p} \in \mathcal{M}_\infty^n$ (i.e. the solutions do not leave \mathcal{M}_∞^n) in other words, the equation $\frac{\partial}{\partial t} \mathbf{p}_t = b(\mathbf{p}_t)$ has a unique \mathcal{M}_∞^n -valued solution $(\mathbf{p}_t)_{t \geq 0}$ for each initial state $\mathbf{p}_0 \in \mathcal{M}_\infty^n$.

For $\bar{\mathbf{x}} \in \mathcal{M}_N^n$ and $\sigma \neq \sigma^- \in \{0, 1\}^n$ let $\bar{\mathbf{x}}_{\sigma, \sigma^-} \in \mathcal{M}_N^n$ be defined as $\bar{\mathbf{x}}_{\sigma, \sigma^-}(\sigma) = \bar{\mathbf{x}}(\sigma) + 1/N$, $\bar{\mathbf{x}}_{\sigma, \sigma^-}(\sigma^-) = \bar{\mathbf{x}}(\sigma^-) - 1/N$ and $\bar{\mathbf{x}}_{\sigma, \sigma^-}(\sigma') = \bar{\mathbf{x}}(\sigma')$ for all $\sigma, \sigma^- \neq \sigma' \in \mathcal{M}_N^n$. For $\bar{\mathbf{x}}, \bar{\mathbf{y}} \in \mathcal{M}_N^n$, we let

$$r_N(\bar{\mathbf{x}}, \bar{\mathbf{y}}) := \begin{cases} \alpha N \bar{\mathbf{x}}(\sigma^-) (O^{\bar{\mathbf{x}}}(\sigma, \sigma^-) - \frac{1}{N}) \hat{O}^{\bar{\mathbf{x}}}(\sigma, \sigma^-) & \text{if } \bar{\mathbf{y}} = \bar{\mathbf{x}}_{\sigma, \sigma^-} \text{ and } \sigma \geq \sigma^-, \\ N \bar{\mathbf{x}}(\sigma) & \text{if } \bar{\mathbf{y}} = \bar{\mathbf{x}}_{\sigma, \sigma}, \\ 0 & \text{otherwise} \end{cases} \quad (\text{I.2.6.10})$$

Then, for $\sigma \in \{0, 1\}^n \setminus \mathbf{0}$, $\beta_N(\sigma)$ is given by

$$\begin{aligned} \beta_N(\bar{\mathbf{x}})(\sigma) = & \alpha \sum_{\sigma^- \leq \sigma, \sigma^- \neq \sigma} \bar{\mathbf{x}}(\sigma^-) (O^{\bar{\mathbf{x}}}(\sigma, \sigma^-) - 1/N) \hat{O}^{\bar{\mathbf{x}}}(\sigma, \sigma^-) \\ & - \alpha \sum_{\sigma \leq \sigma^+, \sigma \neq \sigma^+} \bar{\mathbf{x}}(\sigma) (O^{\bar{\mathbf{x}}}(\sigma^+, \sigma) - 1/N) \hat{O}^{\bar{\mathbf{x}}}(\sigma^+, \sigma) - \bar{\mathbf{x}}(\sigma), \end{aligned} \quad (\text{I.2.6.11})$$

and

$$\beta_N(\bar{\mathbf{x}})(\mathbf{0}) = (1 - \bar{\mathbf{x}}(\mathbf{0})) - \alpha \sum_{\sigma \neq \mathbf{0}} \bar{\mathbf{x}}(\mathbf{0}) (O^{\bar{\mathbf{x}}}(\sigma, \mathbf{0}) - 1/N) \hat{O}^{\bar{\mathbf{x}}}(\sigma, \mathbf{0}). \quad (\text{I.2.6.12})$$

Just as in the proof of the Proposition I.2.1.2 we see that (I.2.6.8) holds and so we can once again use Theorem [Swa17, Thm 3.2] (whose proof in the higher-dimensional case remains identical) to conclude that (I.2.1.67) holds. \blacksquare

I.2.6.3 Proof of Proposition I.2.1.3

The proof of Proposition I.2.1.3 depends on two lemmas for our processes of interest, and an abstract result for continuous-time Markov chains. Note that the space $\bar{\mathcal{H}}$ from (I.2.1.7) is countable, so our processes are continuous-time Markov chains with countable state spaces.

Lemma I.2.6.1 (The mean-field dual process is nonexplosive) *The continuous-time Markov chain $(\bar{Y}_t)_{t \geq 0}$ with state space $\bar{\mathcal{H}}$ and jump rates as in (I.2.1.20) is nonexplosive.*

Proof Let $(N_t)_{t \geq 0}$ be the continuous-time Markov chain with state space \mathbb{N} that jumps from k to $k+2$ at rate αk . I.e., N_t is the number of particles in a branching process where each particle gives birth to two new particles at rate α . It is well-known that $(N_t)_{t \geq 0}$ is nonexplosive (see for example [GS01], Section 6.8. Theorem (19)). For $E \in \mathcal{H}$, let $V_E := \bigcup E$ as in (I.2.1.5) denote the vertex set of E and let $\|E\| := |V_E|$ denote the cardinality of V_E . Similarly, let $\|\bar{E}\| := \|E\|$ denote the number of vertices of a hypergraph $\bar{E} \in \bar{\mathcal{H}}$. Since $\|\text{coop}_k^\bullet(E)\| = \|E\| + 2$ for each vertex k of E and since $\|\text{death}_k^\bullet(E)\| \leq \|E\|$, we can couple the processes $(N_t)_{t \geq 0}$ and $(\bar{Y}_t)_{t \geq 0}$ with initial states $\|\bar{Y}_0\| = N_0$ in such a way that $\|\bar{Y}_t\| \leq N_t$ ($t \geq 0$) a.s. In particular, $(\bar{Y}_t)_{t \geq 0}$ is nonexplosive. \blacksquare

Lemma I.2.6.2 (Convergence of the jump rates) *Let \bar{r}_N and \bar{r} be the rates defined in (I.2.1.12) and (I.2.1.20), respectively. Then*

$$\bar{r}_N(\bar{Y}, \bar{Z}) \xrightarrow{N \rightarrow \infty} \bar{r}(\bar{Y}, \bar{Z}) \quad (Y, Z \in \mathcal{H}, \bar{Y} \neq \bar{Z}). \quad (\text{I.2.6.13})$$

Remark For each $Y, Z \in \mathcal{H}$ there exists an M such that $\bar{Y}, \bar{Z} \in \bar{\mathcal{H}}_N$ for all $N \geq M$. In particular, the left-hand side of (I.2.6.13) is well-defined for all N large enough.

Proof of Lemma I.2.6.2 Write

$$\bar{r}_N(\bar{Y}, \bar{Z}) = \alpha \bar{r}_N^1(\bar{Y}, \bar{Z}) + \bar{r}^2(\bar{Y}, \bar{Z}) \quad \text{and} \quad \bar{r}(\bar{Y}, \bar{Z}) = \alpha \bar{r}^1(\bar{Y}, \bar{Z}) + \bar{r}^2(\bar{Y}, \bar{Z}), \quad (\text{I.2.6.14})$$

where

$$\begin{aligned} \bar{r}_N^1(\bar{Y}, \bar{Z}) &:= \frac{1}{(N-1)(N-2)} \sum_{ijk} 1_{\{\overline{\text{coop}}_{ijk}^\bullet(Y) = \bar{Z}\}}, \\ \bar{r}^1(\bar{Y}, \bar{Z}) &:= \sum_{k \in V_Y} 1_{\{\overline{\text{coop}}_k^\bullet(Y) = \bar{Z}\}}, \\ \bar{r}^2(\bar{Y}, \bar{Z}) &:= \sum_{k \in V_Y} 1_{\{\overline{\text{death}}_k^\bullet(Y) = \bar{Z}\}}. \end{aligned} \quad (\text{I.2.6.15})$$

Since the rates \bar{r}^2 do not depend on N , it suffices to show that

$$\bar{r}_N^1(\bar{Y}, \bar{Z}) \xrightarrow{N \rightarrow \infty} \bar{r}^1(\bar{Y}, \bar{Z}) \quad (Y, Z \in \mathcal{H}, \bar{Y} \neq \bar{Z}). \quad (\text{I.2.6.16})$$

Recall that we write $E \sim F$ when two elements $E, F \in \mathcal{H}$ are isomorphic and that $\Lambda_N = \{1, \dots, N\}$. Then, for each $Y, Z \in \mathcal{H}$ such that $\bar{Y} \neq \bar{Z}$ and N large enough such that $Y, Z \in \mathcal{H}(\Lambda_N)$,

$$\begin{aligned} \bar{r}_N^1(\bar{Y}, \bar{Z}) &= \frac{1}{(N-1)(N-2)} \sum_{ijk} 1_{\{\text{coop}_{ijk}^\bullet(Y) \sim Z\}} \\ &= \frac{1}{(N-1)(N-2)} |\{(i, j, k) \in \Lambda_N^3 : \text{coop}_{ijk}^\bullet(Y) \sim Z\}|. \end{aligned} \quad (\text{I.2.6.17})$$

Similarly,

$$\bar{r}^1(\bar{Y}, \bar{Z}) = |\{k \in V_Y : \text{coop}_{i_Y j_Y k}^\bullet(Y) \sim Z\}|, \quad (\text{I.2.6.18})$$

where $i_Y, j_Y \in \mathbb{N} \setminus V_Y$ are chosen in some arbitrary way as in (I.2.1.19). Clearly, both expressions are zero unless there exist $(i', j', k') \in \mathbb{N}^3$ such that $\text{coop}_{i'j'k'}^\bullet(Y) \sim Z$. Otherwise, we may fix such i', j', k' and write

$$\begin{aligned} \text{(i)} \quad \bar{r}_N^1(\bar{Y}, \bar{Z}) &= \frac{1}{(N-1)(N-2)} |\{(i, j, k) \in \Lambda_N^3 : \text{coop}_{ijk}^\bullet(Y) \sim \text{coop}_{i'j'k'}^\bullet(Y)\}|, \\ \text{(ii)} \quad \bar{r}^1(\bar{Y}, \bar{Z}) &= |\{k \in V_Y : \text{coop}_{i_Y j_Y k}^\bullet(Y) \sim \text{coop}_{i'j'k'}^\bullet(Y)\}|. \end{aligned} \quad (\text{I.2.6.19})$$

Since $\bar{Y} \neq \bar{Z}$, we must have $k' \in V_Y$, and each element (i, j, k) of the set in (I.2.6.19) (i) satisfies $k \in V_Y$.

If either $i' \in V_Y$ or $j' \in V_Y$ (or both), then $|V_Z| \leq |V_Y| + 1$ and each element (i, j, k) of the set in (I.2.6.19) (i) satisfies $i \in V_Y$ or $j \in V_Y$. In particular, this set has cardinality at most $2N|V_Y|^2$ so $\bar{r}_N^1(\bar{Y}, \bar{Z})$ is of order $1/N$ while $\bar{r}^1(\bar{Y}, \bar{Z}) = 0$.

On the other hand, if $i', j' \notin V_Y$, then $|V_Z| = |V_Y| + 2$ and each element (i, j, k) of the set in (I.2.6.19) (i) satisfies $i, j \notin V_Y$. Since for a given $k \in V_Y$, all choices of $i, j \notin V_Y$ yield after application of $\text{coop}_{ijk}^\bullet$ the same hypergraph (up to isomorphisms), we have in this case that

$$\bar{r}_N^1(\bar{Y}, \bar{Z}) = \frac{(N - |V_Y|)(N - |V_Y| - 1)}{(N-1)(N-2)} |\{k \in V_Y : \text{coop}_{i_Y j_Y k}^\bullet(Y) \sim \text{coop}_{i'j'k'}^\bullet(Y)\}|, \quad (\text{I.2.6.20})$$

which converges to $\bar{r}^1(\bar{Y}, \bar{Z})$ as $N \rightarrow \infty$. ■

In view of Lemmas I.2.6.1 and I.2.6.2, Proposition I.2.1.3 is implied by the following general result.

Proposition I.2.6.3 (Convergence of continuous-time Markov chains) *Let $(X_t^N)_{t \geq 0}$ ($N \geq 1$) and $(X_t)_{t \geq 0}$ be continuous-time Markov chains with countable state spaces S_N and S , respectively. Assume that $(X_t)_{t \geq 0}$ is nonexplosive. Assume moreover that $S_N \subset S$ and the S_N approximate S in the sense that*

$$\forall x \in S \exists M \text{ s.t. } x \in S_N \forall N \geq M. \quad (\text{I.2.6.21})$$

For $x \neq y$, let $r_N(x, y)$ resp. $r(x, y)$ denote the rate at which $(X_t^N)_{t \geq 0}$ resp. $(X_t)_{t \geq 0}$ jumps from x to y . Assume that for all $x \in S$

$$r_N(x, y) \xrightarrow{N \rightarrow \infty} r(x, y) \quad \text{uniformly in } y \in S, y \neq x \quad (\text{I.2.6.22})$$

and that the law of X_0^N converges weakly to the law of X_0 as $N \rightarrow \infty$. Then

$$\mathbb{P}[(X_t^N)_{0 \leq t \leq T} \in \cdot] \xrightarrow{N \rightarrow \infty} \mathbb{P}[(X_t)_{0 \leq t \leq T} \in \cdot] \quad (T \geq 0), \quad (\text{I.2.6.23})$$

where \rightarrow denotes convergence of probability measures in total variation norm distance on the space of piecewise constant, right-continuous functions from $[0, \infty)$ into S , equipped with the Skorohod topology.

Remark In view of (I.2.6.21), for each $x, y \in S$ such that $x \neq y$, there exists an M such that $x, y \in S_N$ for all $N \geq M$. In particular, the left-hand side of (I.2.6.22) is well-defined for all N large enough.

Proof of Proposition I.2.6.3 For each $x, y \in S$ with $x \neq y$, let $\pi(x, y) \subset [0, \infty)^2$ be a Poisson point set whose intensity measure is the Lebesgue measure on $[0, \infty)^2$. Extend r_N to S^2 by setting $r_N(x, y) := 0$ if x or y are not elements of S_N . Set

$$\begin{aligned} \omega_N(x, y) &:= \{t \geq 0 : (t, r) \in \pi(x, y), r \leq r_N(x, y)\}, \\ \omega(x, y) &:= \{t \geq 0 : (t, r) \in \pi(x, y), r \leq r(x, y)\}. \end{aligned} \quad (\text{I.2.6.24})$$

Then we can construct $(X_t^N)_{t \geq 0}$ from the Poisson processes $(\omega_N(x, y))_{x \neq y}$ in such a way that the Markov chain jumps only at times of these Poisson processes and at each time $t \in \omega_N(x, y)$, if $X_{t-}^N = x$ (i.e., just prior to time t the process is in x), then $X_t^N = y$ (i.e., the process jumps to y). Likewise, we can construct $(X_t)_{t \geq 0}$ from $(\omega(x, y))_{x \neq y}$. We claim that for each $T \geq 0$, there exists a random $M < \infty$ such that

$$(X_t^N)_{0 \leq t \leq T} = (X_t)_{0 \leq t \leq T} \quad \forall N \geq M. \quad (\text{I.2.6.25})$$

In particular, since $\lim_{K \rightarrow \infty} \mathbb{P}[M > K] = \mathbb{P}[M = \infty] = 0$, this implies (I.2.6.23). It remains to show that (I.2.6.25) holds. For a given $x \in S$ and $y \in S$, it holds for a time $t \geq 0$ that $t \in \omega_N(x, y) \setminus \omega(x, y)$ (resp. $t \in \omega(x, y) \setminus \omega_N(x, y)$) if and only if there exists a point $(s, r) \in \pi(x, y)$ such that $r \in (r(x, y), r_N(x, y)]$ (resp. $r \in (r_N(x, y), r(x, y)]$). By (I.2.6.22), for any $T \geq 0$ and any realization of $\pi(x, y)$ we can find an $N(x, y)$ such that

$$\omega_N(x, y) \cap [0, T] = \omega(x, y) \cap [0, T] \quad \forall N \geq N(x, y). \quad (\text{I.2.6.26})$$

Since the convergence (I.2.6.22) is uniform in $y \in S$, we can also find an $N(x)$ which does not depend on $y \in S$ such that for all $x \in R_T$ we have

$$\omega_N(x, y) \cap [0, T] = \omega(x, y) \cap [0, T] \quad \forall N \geq N(x), y \in S \quad (\text{I.2.6.27})$$

where the uniformity of convergence (I.2.6.22) guarantees that $N(x)$ does not depend on $y \in S$. We note that by our assumption that $(X_t)_{t \geq 0}$ is nonexplosive, the random set

$$R_T := \{x \in S : X_t = x \text{ for some } t \in [0, T]\} \quad (\text{I.2.6.28})$$

of points visited by $(X_t)_{t \geq 0}$ until time T is a.s. finite and so $M := \max_{x \in R_T} N(x) < \infty$ almost surely. Since (I.2.6.27) holds for all $x \in R_T$ with $N(x) := M$, this implies (I.2.6.25). \blacksquare

This completes the proof of Proposition I.2.1.3. \blacksquare

In the next section we will prove Proposition I.2.1.4 and Theorem I.2.1.19. Once again, Theorem I.2.1.19 is just a multidimensional version of Proposition I.2.1.4 and both proofs are very similar.

I.2.6.4 Proofs of Proposition I.2.1.4 and Theorem I.2.1.19

Proof of Proposition I.2.1.4 Assuming that $\bar{\phi}$ is continuous (which we will show later, see (I.2.6.34)), if $\bar{Y}_0^N \xrightarrow[N \rightarrow \infty]{} \bar{Y}_0$ and $\bar{X}_0^N \xrightarrow[N \rightarrow \infty]{} u_0$ almost surely, we see by (I.2.1.18) that we have

$$\mathbb{E}[\bar{\phi}(\bar{X}_t^N, \bar{Y}_0^N)] \xrightarrow[N \rightarrow \infty]{} \mathbb{E}[\bar{\phi}(u_t, \bar{Y}_0)] \quad (\text{I.2.6.29})$$

and by (I.2.1.22) we obtain that also

$$\mathbb{E}[\bar{\phi}(\bar{X}_0^N, \bar{Y}_t^N)] \xrightarrow[N \rightarrow \infty]{} \mathbb{E}[\bar{\phi}(u_0, \bar{Y}_t)]. \quad (\text{I.2.6.30})$$

If we can show that

$$\mathbb{E}[\bar{\phi}^N(\bar{X}_t^N, \bar{Y}_0^N)] \xrightarrow[N \rightarrow \infty]{} \mathbb{E}[\bar{\phi}(u_t, \bar{Y}_0)] \quad (\text{I.2.6.31})$$

and

$$\mathbb{E}[\bar{\phi}^N(\bar{X}_0^N, \bar{Y}_t^N)] \xrightarrow[N \rightarrow \infty]{} \mathbb{E}[\bar{\phi}(u_0, \bar{Y}_t)], \quad (\text{I.2.6.32})$$

then the statement follows by I.2.6.29, I.2.6.30 and Proposition I.2.1.1.

For (I.2.6.31), it is enough to show that for any $\varepsilon > 0$ there exists an $N_\varepsilon \in \mathbb{N}$ such that

$$\left| \mathbb{E}[\bar{\phi}(\bar{X}_t^N, \bar{Y}_0^N)] - \mathbb{E}[\bar{\phi}^N(\bar{X}_t^N, \bar{Y}_0^N)] \right| \leq \varepsilon, \quad N \geq N_\varepsilon. \quad (\text{I.2.6.33})$$

For an $\bar{E} \in \bar{\mathcal{H}}$ and E an arbitrary fixed representative of \bar{E} , we denote by $|\bar{E}| = |E| := \#\{e : e \in E\}$ the number of hyperedges and by $\|\bar{E}\| = \|E\| := \max\{|e| : e \in E\}$ the maximum number of occupied sites in an hyperedge of \bar{E} . Let $\bar{E} \in \bar{\mathcal{H}}$ and let E be its arbitrary representative and recall that by $V_E = \{i \in \Lambda_N; \exists e \in E \ i \in e\}$ we denote the set of vertices of E . Since $\phi(p, E)$ does not depend on the choice of the representative E and the event $\{\text{Thin}_{\tilde{x}}(E) \neq \emptyset\}$ only depends on the values $\tilde{x}(i)$ for $i \in V_E$ we can write for any $p \in [0, 1]$

$$\begin{aligned} \bar{\phi}(p, \bar{E}) &= \phi(p, E) = \sum_{\tilde{x} \in \mathcal{S}(\Lambda_N)} \mathbb{P}[(B_p(i), i \in \Lambda_N) = \tilde{x}] \mathbf{1}_{\{\text{Thin}_{\tilde{x}}(E) \neq \emptyset\}} \\ &= \sum_{x \in \mathcal{S}(V_E)} p^{|x|} (1-p)^{|V_E| - |x|} \mathbf{1}_{\{\text{Thin}_x(E) \neq \emptyset\}}. \end{aligned} \quad (\text{I.2.6.34})$$

In words, we sum over all the probabilities of all possible configurations of open and closed vertices such that thinning with that configuration leaves E nonempty. Here, for any $i \in \Lambda_N$ we have that $B_p(i) = 1$ with probability p independently of $B_p(j)$, $j \neq i$. In particular, (I.2.6.34)

shows that $\bar{\phi}$ is continuous, so our initial assumption is justified. Similarly for any $N \in \mathbb{N}$ and any $\bar{E} \in \bar{\mathcal{H}}_N$ let E be an arbitrary fixed representative of \bar{E} . Since the event $\{\text{Thin}_{\tilde{x}}(E) \neq \emptyset\}$ only depends on the values $\tilde{x}(i)$ for $i \in V_E$, for $k \in \{0, \dots, N\}$ we can write

$$\begin{aligned} \phi^N \left(\frac{k}{N}, E \right) &= \sum_{\tilde{x} \in \mathcal{S}(\Lambda_N)} \mathbb{P}[B_p^N = \tilde{x}] \mathbf{1}_{\{\text{Thin}_{\tilde{x}}(E) \neq \emptyset\}} \\ &= \sum_{x \in \mathcal{S}(V_E)} \prod_{i=0}^{|x|-1} \binom{k-i}{N-i} \vee 0 \prod_{j=0}^{|V_E|-|x|-1} \binom{N-k-j}{N-|x|-j} \vee 0 \mathbf{1}_{\{\text{Thin}_x(E) \neq \emptyset\}}. \end{aligned} \quad (\text{I.2.6.35})$$

Here $\prod_{i=0}^{|x|-1} \binom{k-i}{N-i} \vee 0 \prod_{j=0}^{|V_E|-|x|-1} \binom{N-k-j}{N-|x|-j} \vee 0$ is the probability that $(B_p^N(i), i \in V_E) = x$ if the ratio of occupied sites on Λ_N is $\frac{k}{N}$. In words, (I.2.6.35) is the same as (I.2.6.34), except that the number of open sites in Λ_N is fixed and so the probabilities correspond to sampling occupied and unoccupied sites without replacement.

Let $p = \frac{k}{N}$ for some $k \leq N \in \mathbb{N}$. Then it holds for all $i \leq \min(k, |V_E|)$ that

$$p - \frac{k-i}{N-i} = \frac{i(N-k)}{N(N-i)} < \frac{|V_E|}{N-|V_E|} = O(1/N)$$

and for $k \leq |V_E|$ we have that $\frac{k}{N} < \frac{N_E}{N} = O(1/N)$. That means in particular that

$$\left| p - \binom{k-i}{N-i} \vee 0 \right| = O(1/N). \quad (\text{I.2.6.36})$$

Similarly, for $m \leq |V_E|$, $j \leq |V_E| - m - 1$ and $k \leq N$ we obtain

$$\left| (1-p) - \binom{N-k-j}{N-m-j} \vee 0 \right| = \frac{O(N)}{O(N^2)} = O(1/N). \quad (\text{I.2.6.37})$$

Since $\bar{\phi}(p, \bar{E})$ is continuous in p for any $\bar{E} \in \mathcal{H}$, we see from (I.2.6.36) and (I.2.6.37) that also

$$\left| \bar{\phi}^N \left(\frac{k}{N}, \bar{E} \right) - \bar{\phi} \left(\frac{k}{N}, \bar{E} \right) \right| = O(1/N), \quad (\text{I.2.6.38})$$

where the right-hand side can be chosen so that it only depends on $|V_E|$ and N .

If for any $t \geq 0$ the quantity $|V_{Y_t^N}|$ were uniformly bounded over all $N \in \mathbb{N}$, we would be done. Generally, that is not the case. However, we can find a bound for the number of vertices of the hypergraph, at least on an event of an arbitrarily high probability and that will suffice to show the convergence. The process $(\bar{Y}_t)_{t \geq 0}$ takes values on finite subsets of finite configurations on \mathbb{N} . Therefore, for any $\varepsilon > 0$ we can find a $k_\varepsilon^0 \in \mathbb{N}$ and $l_\varepsilon^0 \in \mathbb{N}$ such that the event $A_\varepsilon^0 := \{|\bar{Y}_0| < k_\varepsilon^0, |\bar{Y}_0| < l_\varepsilon^0\}$ has probability at least $1 - \varepsilon$. Since both $\bar{\phi}$ and $\bar{\phi}^N$ are bounded (uniformly for all $N \in \mathbb{N}$), it is then enough to show that for any $\varepsilon > 0$ and $N_\varepsilon \in \mathbb{N}$ large enough we have

$$\left| \mathbf{1}_{A_\varepsilon^0} \left(\bar{\phi}(\bar{X}_t^N, \bar{Y}_0^N) - \bar{\phi}^N(\bar{X}_t^N, \bar{Y}_0^N) \right) \right| \leq \varepsilon \text{ almost surely, } N \geq N_\varepsilon. \quad (\text{I.2.6.39})$$

Denote by $\bar{\mathcal{H}}_N^\varepsilon := \{\bar{E} \in \bar{\mathcal{H}}_N, |\bar{E}| < k_\varepsilon, |\bar{E}| < l_\varepsilon\}$. Since for any $N \in \mathbb{N}$, \bar{Y}_0^N is independent of $(\bar{X}_t^N)_{t \geq 0}$, the left-hand side of (I.2.6.39) can be bounded from above by

$$\left| \max_{\bar{E} \in \bar{\mathcal{H}}_N^\varepsilon} \left(\bar{\phi}(\bar{X}_t^N, \bar{E}) - \bar{\phi}^N(\bar{X}_t^N, \bar{E}) \right) \right|,$$

which is well-defined as for each $\varepsilon > 0$ as \bar{Y}_0^N takes finitely many values (independently of N) on the event A_ε^0 . Now (I.2.6.39) follows by (I.2.6.38) and the fact that on the event A_ε and for any representative Y_0^N of \bar{Y}_0^N , $|V_{Y_0^N}|$ is bounded by $n_\varepsilon = k_\varepsilon l_\varepsilon$ for all $N \in \mathbb{N}$. From this, (I.2.6.33) follows and so by (I.2.6.29) we see that (I.2.6.31) holds as well.

It remains to show (I.2.6.32). For any $N \in \mathbb{N}$, processes $(|\bar{Y}_t^N|)_{t \geq 0}$, $(\|\bar{Y}_t^N\|)_{t \geq 0}$, $(|\bar{Y}_t|)_{t \geq 0}$ and $(\|\bar{Y}_t\|)_{t \geq 0}$ are all integrable since each of them can be coupled with a pure birth process $(M_t)_{t \geq 0}$ (which is nonexplosive, see for example [GS01], Section 6.8. Theorem (19)) so that $\|\bar{Y}_t\| \vee \|\bar{Y}_t^N\| \leq M_t$ and $|\bar{Y}_t| \vee |\bar{Y}_t^N| \leq M_t$ for all $N \in \mathbb{N}$. Hence, for any $\varepsilon > 0$ and $t \geq 0$ we can find a k_ε^t and l_ε^t such that $A_\varepsilon^t := \{|\bar{Y}_t| < k_\varepsilon^t, |\bar{Y}_t^N| < l_\varepsilon^t\}$ has probability at least $1 - \varepsilon$. Now, we can obtain (I.2.6.32) in the same way we obtained (I.2.6.31), noting that also \bar{X}_0^N is independent of $(\bar{Y}_t^N)_{t \geq 0}$. This completes the proof of Proposition I.2.1.4. \blacksquare

Proof of Theorem I.2.1.19 The proof is almost identical to that of Proposition I.2.1.4. For $N \in \mathbb{N}$, $\bar{E} \in \mathcal{H}_N$ and $p \in \mathcal{M}_n$ we can write

$$\bar{\phi}^n(p, \bar{E}) = \sum_{x \in \mathcal{S}(V_E)^n} \prod_{\sigma \in \{0,1\}^n} \mathbf{p}(\sigma)^{|x(\sigma)|} \mathbf{1}_{\{\text{Thin}_{x_i}(E) \neq \emptyset \ \forall i=1, \dots, n\}}, \quad (\text{I.2.6.40})$$

where E is an arbitrary representative of \bar{E} . Let $\mathbf{p} \in \mathcal{M}_n^N$. Then we can write $\mathbf{p} = \frac{k}{N}$ where $k : \{0, 1\}^n \mapsto \{0, \dots, N\}$ such that $\sum_{\sigma} \in \{0, 1\}^n k(\sigma) = N$. We have

$$\bar{\phi}^n\left(\frac{k}{N}, \bar{E}\right) = \sum_{x \in \mathcal{S}(V_E)^n} \prod_{i=0}^{|V_E|-1} \frac{1}{N-i} \prod_{\sigma \in \{0,1\}^n} \prod_{j=0}^{|x(\sigma)|-1} ((k(\sigma) - j) \vee 0) \mathbf{1}_{\{\text{Thin}_{x_i}(E) \neq \emptyset \ \forall i=1, \dots, n\}}. \quad (\text{I.2.6.41})$$

We can apply the estimates (I.2.6.36) and (I.2.6.37) to $\bar{\phi}_N^n$ and the remainder of the proof is the same as in the proof of the Proposition I.2.1.4. \blacksquare

In the next subsection, we will prove the existence and uniqueness of solutions of the equation (I.2.1.33) and the duality between those solutions and the mean-field dual process.

I.2.6.5 Proofs of Propositions I.2.1.6 and I.2.1.8, Lemma I.2.1.7 and Theorem I.2.1.9

Let \mathcal{W} be the space of all words over the alphabet $\{1, 2, 3\}$, i.e., elements of \mathcal{W} are strings of the form $a_1 \cdots a_n$ with $n \geq 0$ and $a_i \in \{1, 2, 3\}$ for all $i = 1, \dots, n$. In particular, \emptyset denotes the empty word of length $n = 0$. For any set \mathcal{S} , we also let $\mathcal{S} \times \mathcal{W}$ denote the set of all words of the form $a_0 \cdots a_n$ with $a_0 \in \mathcal{S}$ and $a_1 \cdots a_n \in \mathcal{W}$. If $w \in \mathcal{S} \times \mathcal{W}$ and $a \in \{1, 2, 3\}$, then wa denotes the word w with the letter a appended on the right. Similarly, for $v, w \in \mathcal{W}$, we let vw denote the word obtained by appending w on the right of v . We write $v \prec w$ if $w = vu$ for some $u \in \mathcal{W}$ and we say that w is a descendant of v . We set

$$\mathcal{Q}(\mathcal{S} \times \mathcal{W}) := \{V \in \mathcal{P}_{\text{fin}}(\mathcal{S} \times \mathcal{W}) : v \not\prec w \ \forall v, w \in V \text{ with } v \neq w\}, \quad (\text{I.2.6.42})$$

that is $\mathcal{Q}(\mathcal{S} \times \mathcal{W})$ is the set of all finite sets V of words such that no word which is an element of V is a descendant of any other element of V (other than of itself).

Let \mathcal{S} be a countable set and let (V_0, E_0) be a finite hypergraph with $V_0 \subset \mathcal{S}$.² In our present setting, we do not require that $\bigcup E_0 = V_0$, i.e., (V_0, E_0) may contain isolated vertices.

²We allow for the case that $\mathcal{S} = \mathcal{W}$ and hence $\mathcal{S} \times \mathcal{W} = \mathcal{W}$. In this case we assume that $V_0 \in \mathcal{Q}(\mathcal{W})$. In particular, $v \not\prec w$ for all $v, w \in V_0$ with $v \neq w$.

We will define a Markov process $(V_t, E_t)_{t \geq 0}$ such that at each time, $V_t \in \mathcal{Q}(\mathcal{S} \times \mathcal{W})$ and (V_t, E_t) is a finite hypergraph. Moreover, we construct the process in such a way that $(E_t)_{t \geq 0}$ is up to isomorphisms equal in distribution to our mean-field dual process $(\bar{Y}_t)_{t \geq 0}$.

We construct $(V_t)_{t \geq 0}$ as a continuous-time branching process with the following description:

- (i) Each element $v \in V_t$ is with rate α replaced by three new elements $v1, v2, v3$.
- (ii) Each element $v \in V_t$ dies with rate 1.

We observe that the whole family structure of V_t can be read off from the labeling of its vertices and that $V_t \in \mathcal{Q}(\mathcal{S} \times \mathcal{W})$ for all $t \geq 0$.

To define $(E_t)_{t \geq 0}$, we need some definitions. For any $e, f \in \mathcal{Q}(\mathcal{S} \times \mathcal{W})$, we write $e \rightarrow f$ and say that e is a direct descendant of f if one of the following two conditions is satisfied:

- (i) $v1 \in e$ and $f = (e \setminus \{v1\}) \cup \{v\}$,
- (ii) $v2, v3 \in e$ and $f = (e \setminus \{v2, v3\}) \cup \{v\}$.

We write $e \rightsquigarrow f$ and say that e is a descendant of f if there exist $n \geq 0$ and $e_0, \dots, e_n \in \mathcal{Q}(\mathcal{S} \times \mathcal{W})$ such that $e = e_0 \rightarrow e_1 \rightarrow \dots \rightarrow e_n = f$ (in particular, $e \rightsquigarrow e$). With these definitions, we set

$$E_t := \{e \subset V_t : \exists f \in E_0 \text{ s.t. } e \rightsquigarrow f\}, \quad (\text{I.2.6.43})$$

i.e. E_t is the set of all subsets e of the vertex set V_t such that e is a descendant of some set of words f which is an element of E_0 .

Lemma I.2.6.4 (Branching representation of the mean-field dual) *There exists a bijection $m : \mathcal{S} \times \mathcal{W} \rightarrow \mathbb{N}$ such that for every $t \geq 0$, there exists a representative Y_t of \bar{Y}_t such that $m(E_t) = Y_t$, where for $e \in \mathcal{Q}(\mathcal{S} \times \mathcal{W})$ we let $m(e) := \{i \in \mathbb{N} : i = m(v) \text{ for some } v \in e\}$ be the image of e under the map m and for $E \in \mathcal{P}(\mathcal{Q}(\mathcal{S} \times \mathcal{W}))$ we denote by $m(E) := \{m(e) : e \in E\}$ the image of E under this image map.*

Proof We want to show that there exists a bijection m such that $m(E_t)$ belongs to the equivalence class \bar{Y}_t for every $t \geq 0$. The set $\mathcal{S} \times \mathcal{W}$ is countable and therefore isomorphic to the set \mathbb{N} . In particular, there exists a bijection $m : \mathcal{S} \times \mathcal{W} \rightarrow \bar{\mathbb{N}}$. Unfortunately, we cannot simply associate a particular representative $(Y_t)_{t \geq 0}$ of $(\bar{Y}_t)_{t \geq 0}$ with $(m(E_t))_{t \geq 0}$ since in a branching event, the "parent" word v is relabelled to $v1$ in the process $(E_t)_{t \geq 0}$ whereas no relabeling of sites during a birth-death event occurs in our original representation of the mean field-dual. However, m is a bijection and the set $\mathcal{Q}(\mathcal{S} \times \mathcal{W})$ is defined in such a way that $v1 \notin f$ for all $f \in E_t$ whenever $v \in e$ for some $e \in E_t$. Therefore, roughly speaking, at each branching event we will switch from one representative of \bar{Y}_t to another where the change in the representative corresponds to the relabelling of the parent vertex v of the branching event to $v1$. We will give a construction for the process E_t and show that $m(E_t)$ belongs to the equivalence class \bar{Y}_t for every $t \geq 0$.

First, we choose a particular representative Y_0 of \bar{Y}_0 such that $m(E_0) = Y_0$. Next, we let $D(v), v \in \mathcal{S} \times \mathcal{W}$ be a collection of independent Poisson processes with rate 1 and $C(v), v \in \mathcal{S} \times \mathcal{W}$ a collection of independent Poisson processes with rate α which are also independent of $D(v), v \in \mathcal{S} \times \mathcal{W}$.

For any $t > 0$ such that $t \in D(v)$ if $v \in V_{t-}$ we put $V_t := V_{t-} \setminus \{v\}$ and $Y_t := \text{death}_{m(v)}^\bullet(Y_{t-})$. The case of branching events is a bit more complicated. For $v \in \mathcal{S} \times \mathcal{W}$ and $E \in \mathcal{H}(\mathbb{N})$ let

$a^m(E, v)$ be the hypergraph E with the vertex $m(v)$ replaced by $m(v1)$ in every $e \in E$. In other words, $a^m(E, v) := \{a^m(e, v), e \in E\}$, where

$$a^m(e, v) := \begin{cases} e & \text{if } m(v) \notin e, \\ e \setminus \{m(v)\} \cup \{m(v1)\} & \text{if } m(v) \in e. \end{cases} \quad (\text{I.2.6.44})$$

Note that if $m(v1) \notin E$ then $a^m(E, v)$ and E both belong to the same equivalence class $\bar{E} \in \bar{\mathcal{H}}(\mathbb{N})$. For $t > 0$ such that $t \in C(v)$ and $v \in V_{t-}$ put $V_t := (V_{t-} \cup \{v1, v2, v3\}) \setminus \{v\}$ and $Y_t := \text{coop}_{m(v2)m(v3)m(v1)}^\bullet(a^m(Y_{t-}, v))$. Otherwise, we set $V_t := V_{t-}$ and $Y_t := Y_{t-}$.

We want to show that with this coupling, $Y_t = m(E_t)$. It is enough to show this in case when $m(E_{t-}) = Y_{t-}$. As long as $V_t = V_{t-}$ this is trivially true, since then also $Y_t = Y_{t-}$. From now on we will therefore assume that $V_t \neq V_{t-}$. If $v \in V_{t-}$ is an isolated vertex (i.e. if there exists no $e \in E_{t-}$ such that $v \in e$) and $t \in D(v) \cup C(v)$, then $E_t = E_{t-}$ since $v \notin \bigcup E_{t-}$. Likewise, $a^m(Y_{t-}, v) = Y_{t-}$, $\text{coop}_{ijm(v1)}^\bullet(Y_{t-}) = Y_t$, $i, j \in \mathbb{N}$ and $\text{death}_{m(v)} Y_{t-} = Y_t$, since $m(v) \notin \bigcup Y_{t-}$ and so $Y_t = Y_{t-}$. Next, we consider the case when $v \in V_{t-}$ is not isolated and $t \in D(v)$. Here

$$\begin{aligned} E_t &= \{e \subset V_{t-} \setminus \{v\} : \exists f \in E_0 \text{ s.t. } e \rightsquigarrow f\} \\ &= \{e \subset V_{t-} \setminus \{v\} : \exists f \in E_{t-} \text{ s.t. } e \rightarrow f\} = \{e \in E_{t-} : v \notin e\} \end{aligned}$$

and

$$Y_t = \text{death}_{m(v)}^\bullet(Y_{t-}) = \{y \in Y_{t-} : m(v) \notin y\} = \{m(e) \in m(E_{t-}) : m(v) \notin m(e)\} = m(E_t).$$

Finally, if $v \in V_{t-}$ is not an isolated vertex and $t \in C(v)$ then

$$\begin{aligned} E_t &= \{e \subset (V_{t-} \cup \{v1, v2, v3\}) \setminus \{v\} : \exists f \in E_0 \text{ s.t. } e \rightsquigarrow f\} \\ &= \{e \subset (V_{t-} \cup \{v1, v2, v3\}) \setminus \{v\} : \exists f \in E_{t-} \text{ s.t. } e \rightarrow f\} \\ &= \{e \in E_{t-} : v \notin e\} \cup \{e \setminus \{v\} \cup \{v1\}, e \in E_{t-} : v \in e\} \\ &\quad \cup \{e \setminus \{v\} \cup \{v2, v3\}, e \in E_{t-} : v \in e\} \end{aligned}$$

and

$$\begin{aligned} Y_t &= \text{coop}_{m(v2)m(v3)m(v1)}^\bullet(a^m(Y_{t-}, v)) = a^m(m(E_{t-}), v) \cup \text{bran}_{m(v1)m(v2)m(v3)}(a^m(m(E_{t-}), v)) \\ &= m(\{e \in E_{t-} : v \notin e\} \cup \{e \setminus \{v\} \cup \{v1\}, e \in E_{t-} : v \in e\}) \\ &\quad \cup m(\{e \setminus \{v\} \cup \{v2, v3\}, e \in E_{t-} : v \in e\}) = m(E_t). \end{aligned} \quad (\text{I.2.6.45})$$

■

Remark I.2.6.5 *Lemma I.2.6.4 tells us in particular, that in order to prove Theorem I.2.1.9 it is enough to show the equivalent statement for the branching representation $(E_t)_{t \geq 0}$.*

For any $v \in V_0$, we let

$$V_t^v := \{w \in V_t : v \prec w\} \quad (\text{I.2.6.46})$$

denote the descendants of v , and we set

$$E_t^v := \{e \subset V_t^v : e \rightsquigarrow \{v\}\}. \quad (\text{I.2.6.47})$$

Then $(V_t^v, E_t^v)_{t \geq 0}$ has the same dynamics as $(V_t, E_t)_{t \geq 0}$ and its initial state is $(V_0^v, E_0^v) = (\{v\}, \{\{v\}\})$. By the definition of $(V_t)_{t \geq 0}$ it is clear that $V_t = \bigcup_{v \in V_0} V_t^v$. Since $V_t^v \cap V_t^w = \emptyset$ for any $v \neq w \in V_0$, it follows that

$$E_t = \{e \subset V_t : \exists f \in E_0 \text{ s.t. } e \cap V_t^v \in E_t^v \forall v \in f\}. \quad (\text{I.2.6.48})$$

Let $E \in \mathcal{P}(\mathcal{Q}(\mathcal{S} \times \mathcal{W}))$ and let $\omega : \mathcal{S} \times \mathcal{W} \rightarrow [0, 1]$ be a function. Then we define $\psi(\omega, E) \in [0, 1]$ by

$$\psi(\omega, E) := \mathbb{P}[\text{Thin}_{B_\omega}(E) \neq \emptyset], \quad (\text{I.2.6.49})$$

where $B_\omega := (B_\omega(v))_{v \in \mathcal{S} \times \mathcal{W}}$ is a collection of independent Bernoulli random variables with $\mathbb{P}[B_\omega(v) = 1] = \omega(v)$. Note that in this definition, E and ω are both deterministic.

Let $(\{1\}, \{\{1\}\})$ denote the hypergraph with a single vertex 1 and a single hyperedge $\{1\}$, and let $(V_t^1, E_t^1)_{t \geq 0}$ denote the process given by (I.2.6.43) started in $(V_0^1, E_0^1) := (\{1\}, \{\{1\}\})$. For a given $\mu_0 \in \mathcal{M}_1[0, 1]$, we define a $\mathcal{M}_1[0, 1]$ -valued function $(\mu_t)_{t \geq 0}$ by

$$\mu_t := \mathbb{P}[\psi(\omega_0, E_t^1) \in \cdot] \quad (t \geq 0), \quad (\text{I.2.6.50})$$

where the random variables $(\omega_0(v))_{v \in \mathcal{S} \times \mathcal{W}}$ are i.i.d. with common law μ_0 . Note that since

$$\psi(\omega, \{\{1\}\}) = \omega \quad (\omega \in [0, 1]), \quad (\text{I.2.6.51})$$

setting $t = 0$ in our definition of μ_t yields μ_0 .

Lemma I.2.6.6 (Independent branches) *Let $(E_t)_{t \geq 0}$ be the process given by (I.2.6.43) started in a deterministic initial state E_0 , and let μ_0 be a probability law on $[0, 1]$. Fix $t \geq 0$ and let $(\omega_0(v))_{v \in \mathcal{S} \times \mathcal{W}}$ be i.i.d. with common law μ_0 and independent of $(E_t)_{t \geq 0}$. Then the random variables $(\omega_t(v))_{v \in \bigcup E_0}$ defined by*

$$\omega_t(v) := \psi(\omega_0, E_t^v) \quad (v \in \bigcup E_0) \quad (\text{I.2.6.52})$$

are i.i.d. with common law μ_t given by (I.2.6.50). Moreover

$$\psi(\omega_0, E_t) = \psi(\omega_t, E_0) \quad \text{a.s.}, \quad (\text{I.2.6.53})$$

where we put $\omega_t(v) := 0$, $v \notin \bigcup E_0$.

Proof The $E_t^v, v \in V_0$ are i.i.d. by the construction given in the proof of Lemma I.2.6.4 and independent of $\omega_v, v \in \mathcal{S} \times \mathcal{W}$, which are i.i.d. by definition. It remains to note that $\text{Thin}_{B_{\omega_0}}(E)$ is independent of $\omega_0(v), v \notin \bigcup E$ for any $E \in \mathcal{P}(\mathcal{Q}(\mathcal{S} \times \mathcal{Q}))$ and that by its definition $V_t^v \cap V_t^u = \emptyset$ for any $u \neq v, u, v \in V_0$ and so also $(\bigcup E_t^u) \cap (\bigcup E_t^v) = \emptyset$.

From (I.2.6.52) we see that for any $v \in \bigcup E_0$ we have

$$\psi(\omega_t, \{\{v\}\}) = \mathbb{P}[\text{Thin}_{B_{\omega_t}}(\{\{v\}\}) \neq \emptyset] = \omega_t(v) = \psi(\omega_0, E_t^v) \quad (\text{I.2.6.54})$$

In particular, this means that (I.2.6.53) holds when $E_0 = \{\{v\}\}$ for any $v \in \mathcal{S} \times \mathcal{W}$. For a general E_0 , we see from (I.2.6.48) that

$$\begin{aligned} \psi(\omega_0, E_t) &= \mathbb{P}(\exists e \subseteq V_t \exists f \in E_0 \text{ s.t. } e \cap V_t^v \in E_t^v \forall v \in f, B_{\omega_0}(u) = 1 \forall u \in e) \\ &= \mathbb{P}[\exists f \in E_0 \text{ s.t. } \text{Thin}_{B_{\omega_0}}(E_t^v) \neq \emptyset \forall v \in f], \end{aligned} \quad (\text{I.2.6.55})$$

and

$$\psi(\omega_t, E_0) = \mathbb{P}(\exists f \in E_0 \text{ s.t. } B_{\omega_t}(v) = 1 \forall v \in f).$$

Now, (I.2.6.53) follows from the fact that for any $v \in \bigcup E_0$,

$$\mathbb{P}[B_{\omega_t}(v) = 1] = \omega_t(v) = \mathbb{P}[\text{Thin}_{B_{\omega_0}}(E_t^v) \neq \emptyset].$$

■

Proposition I.2.6.7 (Differential equation) *For each $\mu_0 \in \mathcal{M}_1[0, 1]$, the function $(\mu_t)_{t \geq 0}$ defined in (I.2.6.50) solves (I.2.1.33).*

Proof The statement of the proposition follows from the fact that for any $t \geq 0$ and $\varepsilon > 0$, it holds for the law $\mu_{t+\varepsilon}$ of $\omega_{t+\varepsilon}^1$ that

$$\mu_{t+\varepsilon} = \mu_t e^{-(\alpha+1)\varepsilon} + \delta_0(1 - e^{-\varepsilon}) + \psi(\mu_t)(1 - e^{-\alpha\varepsilon}) + o(\varepsilon).$$

To show this, we note that the process $(V_t^1)_{t \geq 0}$ admits to a graphical representation via independent Poisson processes which have independent increments and that E_t^1 is defined as in (I.2.6.47). Using that and (I.2.6.52), we see that

$$\mu_{t+\varepsilon} = \mu_t \mathbb{P}[E_\varepsilon = E_0 = \{\{1\}\}] + \delta_0 \mathbb{P}[E_\varepsilon = \{\emptyset\}] + \psi(\mu_t) \mathbb{P}[E_\varepsilon = \{\{11\}, \{12, 13\}\}] + o(\varepsilon),$$

where $\psi(\mu_t)$ is defined by (I.2.1.32). Indeed, $E_\varepsilon = \{\{1\}\}$ if and only if there are no branching or death events on the interval $(0, \varepsilon)$.

Shifting by time ε the graphical representation with which we constructed the process $(V_t)_{t \geq 0}$, we construct the process $(V_t')_{t \geq 0}$, where we put $V_0' := \{1\}$. Since the processes which form the graphical representation of $(V_t)_{t \geq 0}$ have independent increments, we have conditioned on the event that there are no branchings or deaths on the interval $(0, \varepsilon)$ that $\omega_{t+\varepsilon} = \omega_t'$, where $\omega_t' = \psi(E_t', \omega_0)$ and $(E_t')_{t \geq 0}$ is obtained from $(V_t')_{t \geq 0}$ as in (I.2.6.47).

The probability that there is more than one event on the interval $(0, \varepsilon)$ is of order $o(\varepsilon)$. If death appears on $(0, \varepsilon)$, $V_\varepsilon = \emptyset$ and so also $E_\varepsilon = \{\emptyset\}$. Conditioned on that (and again using the that the increments of processes of the graphical representation are independent), $\omega_{t+\varepsilon} = 0$. Finally, if there is a branching event on $(0, \varepsilon)$, then $E_\varepsilon = \{\{11\}, \{12, 13\}\}$. Once again, we can use the graphical representation shifted by time ε to construct branching processes $(V_t')_{t \geq 0}$, $(V_t'')_{t \geq 0}$ and $(V_t''')_{t \geq 0}$ started in $V_0' = \{11\}$, $V_0'' = \{12\}$ and $V_0''' = \{13\}$ and by the independence of the increments of the graphical representation and (I.2.6.52), we have that $\omega_{t+\varepsilon} = \omega_t' + (1 - \omega_t')\omega_t''\omega_t'''$, where $\omega_t' = \psi(E_t', \omega_0)$ and ω_t'' and ω_t''' are independent copies of ω_t' .

Since $\mathbb{P}[E_\varepsilon = E_0 = \{\{1\}\}] = e^{-(\alpha+1)\varepsilon}$, $\mathbb{P}[E_\varepsilon = E_0 = \{\{1\}\}] = (1 - e^{-\varepsilon})$ and $\mathbb{P}[E_\varepsilon = \{\{11\}, \{12, 13\}\}] = (1 - e^{-\alpha\varepsilon})$, we are finished. ■

Proof of Proposition I.2.1.6 Proposition I.2.6.7 shows that there exists a solution of (I.2.1.33) which is $\mathcal{M}_1([0, 1])$ -valued by (I.2.6.50). ■

Proof of Lemma I.2.1.7 For any i.i.d. $X_i, i = 1, 2, 3$ with a distribution function F let $\psi(F)$ be the distribution function of $X_1 + (1 - X_1)X_2X_3$. Let F_t be the distribution function of μ_t and F_t^ψ the distribution function of $\psi(\mu_t)$. For the k -th moment $p_t(k)$ we have $p_t(k) = k \int_0^1 x^{k-1}(1 - F_t(x))dx$ and $\frac{\partial}{\partial t} p_t(k) = -k \int_0^1 x^{k-1} \frac{\partial}{\partial t} F_t(x)dx$. Hence, for $X_i, i = 1, 2, 3$ i.i.d. with $\mathcal{L}(X_1) = \mu_t$ we obtain from (I.2.1.34) that

$$\begin{aligned}
\frac{\partial}{\partial t} p_t(k) &= -k \int_0^1 x^{k-1} \left((1_{[x \geq 0]} - F_t(x)) + \alpha \left(F_t^\psi(x) - F_t(x) \right) \right) dx \\
&= -p_t(k) + \alpha \left(-p_t(k) + \sum_{l=0}^k \binom{k}{l} \mathbb{E}[X_1^{k-l} (1 - X_1)^l X_2^l X_3^l] \right) \\
&= -p_t(k) + \alpha \left(\sum_{l=1}^k \binom{k}{l} \mathbb{E}[X_1^{k-l} (1 - X_1)^l] p_t^2(l) \right) \\
&= -p_t(k) + \alpha \left(\sum_{l=1}^k \binom{k}{l} \sum_{j=0}^l \binom{l}{j} (-1)^j p_t(n-l+j) p_t^2(l) \right),
\end{aligned}$$

where in the second equality we use the binomial theorem to see that the k -th moment of $\psi(\mu_t)$ is equal to $\sum_{l=0}^k \binom{k}{l} \mathbb{E}[X_1^{k-l} (1 - X_1)^l X_2^l X_3^l]$ and in the last equality we use it again to expand the term $\mathbb{E}[X_1^{k-l} (1 - X_1)^l] p_t^2(l)$. \blacksquare

Proof of Proposition I.2.1.8 Since μ_t is concentrated on $[0, 1]$ for all $t \geq 0$, the moments $p_t(k)$, $k \in \mathbb{N}$ determine the distribution of μ_t uniquely (see for example [Shi96], Chapter II, §12., Theorem 7). Therefore there exists a unique solution to the equation (I.2.1.33) with initial state μ_0 if the equation (I.2.1.35) has a unique solutions for each $k \in \mathbb{N}$ whenever $p_0(1), \dots, p_n(k)$ are the first k moments of a probability distribution on $[0, 1]$. Note that the existence of such a solution follows immediately from Proposition I.2.1.6, so it remains to show uniqueness. We will take advantage of the fact that for each $k \in \mathbb{N}$, the right-hand side of (I.2.1.35) only depends on functions $(p_t(l))_{t \geq 0}$, $l \leq k$. Fix $k \in \mathbb{N}$ and for each $m \in \mathbb{N}$, $m \leq k$ let $f_m : [0, 1]^k \rightarrow [0, 1]$ be defined as

$$f_m(x_1, \dots, x_k) := -x_k + \alpha \left(\sum_{l=1}^k \binom{k}{l} \sum_{j=0}^l \binom{l}{j} (-1)^j x_{k-l+j} x_l^2 \right), \quad m = 1, \dots, k,$$

where we set $x_0 := 1$. We can then rewrite the equations (I.2.1.35) for $p_t(1), \dots, p_t(k)$ as a k -dimensional differential equation

$$\left(\frac{\partial}{\partial t} p_t(1), \dots, \frac{\partial}{\partial t} p_t(k) \right) = f(p_t(1), \dots, p_t(k)),$$

where we put $f(x_1, \dots, x_k) := (f_1(x_1, \dots, x_k), \dots, f_m(x_1, \dots, x_k))$. The function f is a polynomial and therefore globally Lipschitz continuous, so the uniqueness follows for example from Theorem I-1-4 in [HS99]. \blacksquare

Let $E \in \mathcal{H}$ and let μ be a measure on $[0, 1]$. Then we define a measure $\rho(\mu, E)$ on $[0, 1]$ by

$$\rho(\mu, E) := \mathbb{P}[\psi(\omega, E) \in \cdot], \quad (\text{I.2.6.56})$$

where $\omega := (\omega(v))_{v \in \mathcal{S} \times \mathcal{W}}$ is a collection of i.i.d. random variables with common law μ . Note that in this definition, E and μ are deterministic. By (I.2.6.51),

$$\rho(\mu, \{\{1\}\}) = \mu \quad (\mu \in \mathcal{M}_1[0, 1]). \quad (\text{I.2.6.57})$$

Proof of Theorem I.2.1.9 If \bar{Y}_0 is deterministic, then Lemma I.2.6.6 together with Remark I.2.6.5 imply that

$$\mathbb{E}[\bar{\rho}(\mu_0, \bar{Y}_t)] = \mathbb{P}[\psi(\omega_0, \bar{Y}_t) \in \cdot] = \mathbb{P}[\psi(\omega_t, \bar{Y}_0) \in \cdot] = \rho(\mu_t, \bar{Y}_0). \quad (\text{I.2.6.58})$$

The general statement follows by integrating over the law of \bar{Y}_0 . \blacksquare

I.2.6.6 Proof of Proposition I.2.1.10

This is a simple application of Theorem I.2.1.9. It is easy to see from (I.2.1.30) that

$$\rho(\delta_p, \bar{E}) = \delta_{\bar{\phi}(p, \bar{E})} \quad (p \in [0, 1], \mu \in \mathcal{M}_1[0, 1]) \quad (\text{I.2.6.59})$$

and

$$\rho(\mu, \bar{1}) = \mu \quad (\mu \in \mathcal{M}_1[0, 1]). \quad (\text{I.2.6.60})$$

It follows that for any $f \in B_b[0, 1]$,

$$\mathbb{E}^{\bar{1}}[f(\bar{\phi}(p, \bar{Y}_t))] = \mathbb{E}^{\bar{1}}[\langle \delta_{\bar{\phi}(p, \bar{Y}_t)}, f \rangle] = \mathbb{E}^{\bar{1}}[\langle \rho(\delta_p, \bar{Y}_t), f \rangle] = \langle \rho(\mu_t, \bar{1}), f \rangle = \langle \mu_t, f \rangle, \quad (\text{I.2.6.61})$$

proving (I.2.1.40). ■

I.2.6.7 Proof of Theorem I.2.1.13

It is easy to see from the equation (I.2.1.33) that (I.2.1.53) are fixed points of (I.2.1.33). Since ν_0 is a degenerate distribution and ν_1 and ν_2 are Bernoulli distributions, it is sufficient to study the first two moments of μ_t to obtain their domains of attraction as the Bernoulli distribution on $\{0, 1\}$ is the distribution with the largest variance of all distributions on $[0, 1]$ with a given mean.

Whenever we have for the mean $p_0(1)$ of μ_0 that $0 \leq p_0(1) < x_1$, we see from (I.2.1.35) (or (I.2.1.17)) that $p_t(1) \rightarrow 0$ as $t \rightarrow \infty$ and therefore also $\mu_t \rightarrow \nu_0$ since μ is concentrated on the interval $[0, 1]$ and so for all $k \in \mathbb{N}$ we have $p_t(k) \geq p_t(k+1)$.

If $x_1 < p_0(1)$, then we see from the proof of Proposition I.2.1.5 that $p_t(1)$ converges to the stable nontrivial fixed point x_2 . We will show that in this case, it also holds that $p_t(2) \rightarrow x_2$ for all $p_0(2) \in [p_0(1)^2, p_0(1)]$ (note that if $p_0(2) < p_0(1)^2$ then μ_0 cannot be a probability distribution on $[0, 1]$). To do that we will first prove that the second moment of any solution of (I.2.1.33) for which $p_0(1) = x_2$ can be bounded by the second moments of two particular solutions and use that to obtain the domains of attraction of the fixed points.

By Proposition I.2.1.11 and martingale convergence, when $\mu_0 = \delta_{x_2}$, we have that $\mu_t \rightarrow \nu_2$ since ν_2 is a distribution of a Bernoulli random variable with success probability x_2 . In particular, we have that $p_t(2) \rightarrow x_2$ in this case. Since δ_{x_2} is the distribution on $[0, 1]$ with the smallest second moment among all distributions with mean x_2 , it follows that the second moment $p_t^1(2)$ of the solution of (I.2.1.33) with $\mu_0 = \delta_{x_2}$ is smaller or equal than the second moment of any other solution of (I.2.1.33) with the first moment equal to x_2 (which follows for example from Theorem A.3.5). On the other hand the second moment $p_t^2(2)$ of the solution of (I.2.1.33) with $\mu_0 = \nu_2$ is the distribution with the largest second moment among all distributions on $[0, 1]$ with mean x_2 . Hence we see that for any initial distribution μ_0 with mean x_2 and the second moment $p_0(2) \in [x_2^2, x_2]$ we have that $p_t^1(2) \leq p_t(2) \leq p_t^2(2)$. Since both $p_t^1(2) \rightarrow x_2$ and $p_t^2(2) \rightarrow x_2$ as $t \rightarrow \infty$ (in fact $p_t^2(2) = x_2$ for all $t \geq 0$), it must also hold that $p_t(2) \rightarrow x_2$.

We now consider the case that μ_0 is such that $p_0(1) > x_1$ but $p_0(1) \neq x_2$. By Lemma I.2.1.7 we know that the second moment $p_t(2)$ of μ_t solves the equation

$$\frac{\partial}{\partial t} p_t(2) = \alpha p_t^3(2) + \alpha p_t^2(2)(1 - 2p_t(1)) - p_t(2)(1 + 2\alpha p_t^2(1)) + 2\alpha p_t^3(1). \quad (\text{I.2.6.62})$$

Define $F : [0, 1] \rightarrow [0, 1]$ by $F(p) := \alpha p^3 + \alpha p_t^2(1 - 2x_2) - p(1 + 2\alpha x_2^2) + 2\alpha x_1^3$. Since we already know that $p_t(1) \rightarrow x_2$ whenever $p_0(1) > x_1$, we can rewrite (I.2.6.62) as

$$\frac{\partial}{\partial t} p_t(2) = F(p_t(2)) + \varepsilon(t), \quad (\text{I.2.6.63})$$

where $\varepsilon(t) \rightarrow 0$ as $t \rightarrow \infty$. From the previous paragraph, we know that $p_t^2(2) \rightarrow x_2$ as $t \rightarrow \infty$ whenever μ_0 is such that $p_0(1) = x_2$. It follows that $F(p_t(2)) \rightarrow 0$ as $t \rightarrow \infty$ and hence also $F(p_t(2)) + \varepsilon(t) \rightarrow 0$ as $t \rightarrow \infty$. Therefore there exists a limit $z = \lim_{t \rightarrow \infty} p_t(2)$. Since $p_t(1) \rightarrow x_2$ and since $\nu_2 = (1 - x_2)\delta_0 + x_2\delta_1$ is the only fixed-point of (I.2.1.33) with mean x_2 it follows that $z = x_2$.

Finally, we consider the case that the mean $p_0(1)$ is equal to the unstable fixed point x_1 . In the proof of Lemma I.2.1.12 we saw in (I.2.1.52) that x_1 is also an unstable fixed point of the equation (I.2.1.48) and so $\mu_t \rightarrow \nu_1$ if and only if μ_0 is already equal to ν_1 , i.e. if μ_0 is a Bernoulli distribution with parameter x_1 .

In summary of the proof so far, we have shown (I.2.1.56) and so we see that any other fixed point of (I.2.1.33) has to have the first moment equal to x_1 . By (I.2.1.52) it is also necessary that the second moment of such a fixed point is equal to m_2 . Finally, the existence of an additional fixed point ν_3 follows immediately from (I.2.1.42) and (I.2.1.43) since we obtain ν_3 as the distribution of the limit H_∞^1 of H_t^1 . The particular form of ν_3 follows from the fact that x_2 is the survival probability of the mean-field dual process $(\bar{Y}_t)_{t \geq 0}$ as we showed in Proposition I.2.1.11 and the fact that $H_\infty^1 = \lim_{t \rightarrow \infty} \bar{\phi}(x_1, \bar{Y}_t)$ which is zero on the event that $(\bar{Y}_t)_{t \geq 0}$ dies out. \blacksquare

I.2.6.8 Proof of Proposition I.2.1.20

In Proposition I.2.1.18 we have already shown the convergence (I.2.1.67) of the process $(\bar{\mathbf{X}}_t^N)_{t \geq 0}$ to the solution $(\mathbf{p}_t)_{t \geq 0}$ of (I.2.1.66). It remains to show that $(\bar{p}_t)_{t \geq 0}$ solves (I.2.1.35) where \bar{p}_t is as in (I.2.1.70) for any $t \geq 0$.

For any $n \in \mathbb{N}$ let $\mathbf{1}_n = (1, \dots, 1) \in \{0, 1\}^n$. We note that $J(\mathbf{1}_n, \sigma) = K(\mathbf{1}_n, \sigma)$ for all $\sigma \in \{0, 1\}^n$, where J is as in (I.2.1.63) and K as in (I.2.1.64) and therefore also $O^p(\mathbf{1}_n, \sigma) = \hat{O}^p(\mathbf{1}_n, \sigma)$. For each $l \in \{1, \dots, n\}$ we have that

$$\sum_{\sigma \in \{0, 1\}_{n-l}^n} \mathbf{p}_t(\sigma) = \sum_{\sigma \in \{0, 1\}_{n-l}^n} \bar{p}_t(l, n) = \binom{n}{l} \bar{p}_t(l, n) \quad (\text{I.2.6.64})$$

is the limit as $N \rightarrow \infty$ of the probability that a given site is occupied in $n - l$ processes (at time t).

For all $k \leq n$, $\binom{n-k}{l} p_t(l, n)$ is the total ratio of sites which are occupied in a fixed set of k processes (out of n processes in total) and such that any combination l sites of the remaining $n - k$ sites are unoccupied. This gives us

$$\bar{p}_t(k) = \sum_{l=0}^{n-k} \binom{n-k}{l} \bar{p}_t(l, n), \quad k, n \in \mathbb{N}, \quad k \leq n. \quad (\text{I.2.6.65})$$

From (I.2.6.65) and (I.2.6.64) we get that for every $1 \leq l \leq n$, the quantity $\binom{n}{l} \bar{p}_t(l, n) (\bar{p}_t(l))^2$ is the limit as $N \rightarrow \infty$ of the probability that at time t a given site is empty in a set of l coupled processes and two other sites are occupied in the same set of l processes. Hence, we see from (I.2.1.66), (I.2.6.64) and (I.2.6.65) that

$$\frac{\partial}{\partial t} \bar{p}_t(n) = -\bar{p}_t(n) + \alpha \sum_{l=1}^n \binom{n}{l} \bar{p}_t(l, n) (\bar{p}_t(l))^2. \quad (\text{I.2.6.66})$$

Formula (I.2.6.65) also yields that

$$\bar{p}_t(l, n) = \sum_{j=0}^l \binom{l}{j} (-1)^j \bar{p}_t(n-l+j). \quad (\text{I.2.6.67})$$

Plugging (I.2.6.67) into (I.2.6.66), we see that (I.2.6.66) is the same as (I.2.1.35). \blacksquare

I.2.6.9 Proof of Lemma I.2.3.1

Set

$$f_y(x) := \chi(x, \{y\}) = 1_{\{y \leq x\}} \quad (x \in \mathcal{S}(\Lambda), y \in \mathcal{S}_{\text{fin}}(\Lambda)). \quad (\text{I.2.6.68})$$

We claim that the functions $(f_y)_{y \in \mathcal{S}_{\text{fin}}}$ separate points. Indeed, if $x \neq x'$, then there is some $i \in \Lambda$ such that $x(i) \neq x'(i)$. Without loss of generality $x(i) = 1$ and $x'(i) = 0$. Set $y(i) := 1$ and $y(j) := 0$ for all $j \neq i$. Then $y \leq x$ but not $y \leq x'$. We moreover observe that $f_{y_1} f_{y_2} = f_{y_1 \vee y_2}$. It follows that the linear span \mathcal{F} of $(f_y)_{y \in \mathcal{S}_{\text{fin}}}$ is an algebra that separates points. Since \mathcal{F} moreover contains the constant function $f_0 = 1$, by the Stone-Weierstrass theorem, \mathcal{F} is dense in the space $\mathcal{C}(\{0, 1\}^\Lambda)$ of all continuous real functions on the compact space $\mathcal{S}(\Lambda) = \{0, 1\}^\Lambda$, equipped with the supremum norm. Now (I.2.3.1) implies $\mathbb{E}[f(X)] = \mathbb{E}[f(X')]$ for all $f \in \mathcal{C}(\{0, 1\}^\Lambda)$ and hence X and X' are equal in distribution. \blacksquare

I.2.6.10 Proof of Lemma I.2.3.2

It suffices to prove the statement when Λ has exactly two elements. Set

$$g_x(Y) := \chi(x, Y) = 1_{\{\exists y \in Y \text{ s.t. } y \leq x\}} \quad (x \in \mathcal{S}(\Lambda), Y \in \mathcal{H}^\circ(\Lambda)). \quad (\text{I.2.6.69})$$

Since Λ has two elements, the set $\{g_x : x \in \mathcal{S}(\Lambda)\}$ has four elements and hence spans a linear space of dimension at most four. On the other hand, since

$$\mathcal{H}^\circ(\Lambda) = \{\emptyset, \{00\}, \{10\}, \{01\}, \{01, 10\}, \{11\}\} \quad (\text{I.2.6.70})$$

has six elements, the space of all probability laws on $\mathcal{H}^\circ(\Lambda)$ is a five-dimensional simplex. In particular, it is not possible to determine a general probability law μ on $\mathcal{H}^\circ(\Lambda)$ by specifying the values of just four linear functionals of μ . \blacksquare

I.2.6.11 Proof of Lemma I.2.3.3

It is not hard to see that the functions $(g_x)_{x \in \mathcal{S}_{\text{fin}}(\Lambda)}$ defined in (I.2.6.69) separate points, i.e., for deterministic $Y, Y' \in \mathcal{H}^\circ(\Lambda)$ such that $Y \neq Y'$, we can find some $x \in \mathcal{S}_{\text{fin}}(\Lambda)$ such that either $y \leq x$ for some $y \in Y$ while $y \not\leq x$ for all $y \in Y'$, or the same holds with the roles of Y and Y' reversed. Let \mathcal{F} be the linear span of the constant function 1 and all functions of the form $\prod_{i=1}^n g_{x_i}$. Then \mathcal{F} is an algebra that separates points. If Λ is finite, then by the Stone-Weierstrass theorem, \mathcal{F} is the space of all real functions on $\mathcal{H}^\circ(\Lambda)$. In particular, if Λ is finite, (I.2.3.5) implies that Y and Y' are equal in distribution. \blacksquare

Chapter I.3

Cooperative branching on other lattices

I.3.1 Cooperative branching on regular trees

In this section, we study the cooperative branching process (I.1.2.16) on regular trees \mathbb{T}_d with $d \geq 2$. In this case each vertex has degree $D = d + 1$. Our aim is to obtain a lower bound that β_{surv} when $\alpha = \gamma = 0$, i.e. when the individuals do not perform random walks and when there are no births with both parent sites being the nearest neighbours of the offspring sites. For simplicity, we will once again set $\delta = 1$, i.e. the deaths will happen at rate 1. As we already noted in Chapter I.2, by time-rescaling we can reduce any $\delta > 0$ to this case.

Lemma I.3.1.1 (Sufficient conditions for extinction) *It holds for the model in (I.1.2.16) on the regular tree \mathbb{T}_d with $\alpha = \gamma = 0$ and $\delta = 1$ that*

$$\beta_{\text{surv}} \geq \frac{d}{d-1} \tag{I.3.1.1}$$

The proof of Lemma I.3.1.1 can be found in subsection I.3.3.1.

I.3.2 The model on the integer lattice

I.3.2.1 Survival and nontriviality of the upper invariant law

In this subsection, we consider the model (I.1.2.16) on the integer lattice \mathbb{Z}^d with nearest neighbour edges. First, we consider the case when $\gamma = \beta = 0$ but $\alpha, \delta > 0$. In other words, deaths happen at a positive rate, there is no random walk dynamics and branching only occurs if both parent sites in a branching event are the nearest neighbours of the offspring site. It is easy to see that the process on \mathbb{Z}^d with these parameters can never survive. Indeed, on \mathbb{Z} , since new particles are placed only between existing particles, the distance between the right-most and left-most particle is a.s. nonincreasing. In a similar way, on \mathbb{Z}^d , the process can never escape from a cube of sites that contains all particles, since all sites outside such a cube border at most one site in the cube.

In the case when $\gamma = 0$, $\delta \geq 0$, $\alpha \geq 0$ and $\beta \geq 0$ Bezuidenhout and Gray show in the Corollary 2.6 of [BG94] that the process dies out when $\beta = \beta_{\text{surv}}$ (and other parameters held constant). We also note that the model can be coupled with a so-called contact process with double deaths as Sturm and Swart did in the Section 2.2. of [SS15] in order to show that

$\beta_{\text{surv}} < \infty$. The coupling in [SS15] is done for the process on \mathbb{Z}^1 and with $\delta = 0$, $\gamma = 1/2$. In the case $\delta > 0$ and $\gamma = 0$ we can still couple the cooperative branching process on \mathbb{Z}^1 with the contact process with double deaths, with the choice of a high enough death rate of the contact process with double deaths. It remains to note that $\beta_{\text{surv}} < \infty$ on \mathbb{Z}^1 implies that also $\beta_{\text{surv}} < \infty$ on \mathbb{Z}^d , $d > 1$. We will show that the methods of [BG94] can moreover be used to show the following.

Proposition I.3.2.1 *It holds for the process with generator (I.1.2.16) on \mathbb{Z}^d with $\gamma = 0$, $\alpha \geq 0$ and $\delta > 0$, that $\beta_{\text{surv}} \geq \beta_{\text{upp}}$.*

The proof of Proposition I.3.2.1 can be found in Subsection I.3.3.2.

I.3.2.2 Random walk dynamics and deaths

In this subsection, we will consider the model (I.1.2.16) on \mathbb{Z} with nearest-neighbour edges with $\alpha = 0$, $\gamma = 1$ and $\delta > 0$. We first prove Proposition I.1.3.3 and also provide lower bounds for the probabilities appearing there. We will denote the cooperative branching process with these parameters by $X = (X_t)_{t \geq 0}$. Put $\mathcal{S} := \mathcal{S}(\mathbb{Z}) = \{0, 1\}^{\mathbb{Z}}$. For $i \in \mathbb{Z}$ let $1_i \in \mathcal{S}$ be zero everywhere except i , i.e. $1_i(i) = 1$ and $1_i(j) = 0$ for all $j \neq i$ and denote by $\text{pair} = \text{pair}_{i,j} := 1_i + 1_j$ the configuration with only the sites i and j occupied. In particular, we will put $\text{pair} := \text{pair}_{0,1}$. Note that since the process is translation invariant, it does not matter for the existence of a nontrivial upper invariant law or the survival of the process whether the initial state is $X_0 = \text{pair}_{0,1}$ or $X_0 = \text{pair}_{i,i+1}$ for any $i \in \mathbb{Z}$.

Proposition I.3.2.2 *Let X be the process with generator (I.1.2.16) on \mathbb{Z} with $\alpha = 0$, $\gamma = 1$ and $\delta > 0$. Then it holds for every $t \geq 0$ that*

$$\mathbb{P}^{\text{pair}}[|X_t| \geq 1] \geq e^{-\delta t} \quad (\text{I.3.2.1})$$

The proof of Proposition I.3.2.2 can be found in subsection I.3.3.3.

Recall that by $1_{\mathbb{Z}}$ we denote the configuration with all sites $i \in \mathbb{Z}$ occupied.

Proposition I.3.2.3 *For any $t_0 > 0$ and any $\delta' > \delta$ there exists a constant $c = c(t_0, \delta')$ such that for all $t \geq t_0$ it holds that*

$$\mathbb{P}^{1_{\mathbb{Z}}}[X_t(0) = 1] \geq ce^{-\delta' t}. \quad (\text{I.3.2.2})$$

The proof of Proposition I.3.2.3 can be found in Subsection I.3.3.4.

Proposition I.3.2.4 *Assume that $\beta \leq 1$. Then it holds for all $t \geq 0$ that*

$$\mathbb{P}^{1_{\mathbb{Z}}}[X_t(0) = 1] \leq e^{-\delta t}, \quad (\text{I.3.2.3})$$

where $1_{\mathbb{Z}} \in \{0, 1\}^{\mathbb{Z}}$ is the configuration with all sites occupied.

The proof of Proposition I.3.2.4 can be found in Subsection I.3.3.5.

Proposition I.3.2.5 *Let X be the cooperative branching process with generator (I.1.2.16) on \mathbb{Z} and assume that $\beta \leq 1$. Then it holds for all $t \geq 0$ that*

$$\mathbb{P}^{\text{pair}}[|X_t| \geq 1] \leq 2e^{-\delta t}. \quad (\text{I.3.2.4})$$

The proof of Proposition I.3.2.5 can be found in Subsection I.3.3.5.

Remark I.3.2.6 In Propositions I.3.2.4 and I.3.2.5 we need to assume that $\beta \leq 1$ in order for our proofs to work. On the other hand, the statement should probably hold for any subcritical β (which is, based on simulations done by Sturm and Swart in [SS15], likely to be at least any $\beta \lesssim 2.45$).

I.3.3 Proofs

I.3.3.1 Proof of Lemma I.3.1.1

We need to show that the process dies out almost surely if $\beta \leq \frac{d}{d-1}$. First, we will prove that for the process started in any deterministic finite initial state $X_0 = x$, one has

$$\frac{\partial}{\partial t} \mathbb{E}[|X_t|] \Big|_{t=0} \leq 0. \quad (\text{I.3.3.1})$$

Let $N := |x|$. We will consider such $X_0 = x$, which maximizes $\frac{\partial}{\partial t} \mathbb{E}[|X_t|] \Big|_{t=0}$. If particles neighbour each other, then by the tree structure of the lattice, these particles form $N - 1$ nearest neighbour pairs. At rate β , each of these pairs produces offspring at a site chosen uniformly from the $2d$ neighbours of the pair, which is successful only if such a site is not blocked by being occupied already.

Building up our particle configuration one by one, we add new particles so that they always neighbour exactly one particle already in the configuration. This way, we create $N - 1$ nearest neighbour pairs such that each of them has $2d$ neighbouring sites and starting from the second pair we form, with each pair we create, we block one site bordering an already existing pair, and also one site bordering the newly created pair is blocked. The number of sites blocked in this way is then $2(N - 2)$. Thus, of the $2d(N - 1)$ neighbouring sites of the $(N - 1)$ pairs, counting the number of empty neighbouring site for each pair separately and then summing up, only $2d(N - 1) - 2(N - 2)$ are free. Note that some of the unoccupied neighbours are shared between two different pairs, so we count them twice. We do that on purpose however, as this also means that the birth rate at those sites is twice as high as at the other neighbouring sites. Since particles die at rate one, there are $N - 1$ pairs in the configuration with a total of $2d(N - 1)$ neighbouring sites, it follows that

$$\frac{\partial}{\partial t} \mathbb{E}[|X_t|] \Big|_{t=0} = \beta(N - 1) \frac{2d(N - 1) - 2(N - 2)}{2d(N - 1)} - N = \beta \left(\frac{d-1}{d} N - \frac{d-2}{d} \right) - N, \quad (\text{I.3.3.2})$$

which is non-positive provided that $\beta \leq \frac{d}{d-1}$. Given a fixed number $N \geq 2$ of occupied particles, a configuration constructed in such a way indeed maximizes the $\frac{\partial}{\partial t} \mathbb{E}[|X_t|] \Big|_{t=0}$. We can see that as follows. For any set of sites on \mathbb{T}_d , let σ_S denote the total number of unoccupied neighbours of sites in S counted for each pair of nearest neighbours in S separately (so we count some of them twice, as we did above). First, assume that a configuration $X_0 = x$ such that $|x| = N$ is not connected in the sense that the set S of sites $i \in \Lambda$ such that $x(i) = 1$ is not a connected set. Assume that $S = S_1 \cup S_2$, where S_1 and S_2 are connected sets. If either S_1 or S_2 contains only one element, it is clear that x does not maximize $\frac{\partial}{\partial t} \mathbb{E}[|X_t|] \Big|_{t=0}$. If both S_1 and S_2 contain at least two elements, we obtain a connected set S' by creating an edge between sites $i \in S_1$ and $j \in S_2$ such that both i and j share an edge with only one other site in S_1 and S_2 , respectively. In the resulting set S' , two additional neighbours are blocked by putting an edge between i and j but the additional pair (i, j) has $(2d - 2)$ unoccupied neighbouring sites, hence $\sigma_{S'} = \sigma_S - 2 + (2d - 2)$ so that $\sigma_{S'} \geq \sigma_S$ for all $d \geq 2$. Therefore we see that if x maximizes $\frac{\partial}{\partial t} \mathbb{E}[|X_t|] \Big|_{t=0}$, it has to be connected. It remains to note that the construction above minimizes σ_S among all connected sets S . Indeed, it is easy to see that if S is a connected set on \mathbb{T}_d , there are at most two pairs of nearest neighbours in S with $2d - 1$ neighbouring sites which are not elements of S and the remaining $N - 3$ pairs each neighbour at least two sites in S .

Using (I.3.3.1), it is now easy to show that also $\mathbb{E}[|X_t|] \leq |X_0|$ for any $t \geq 0$: For all $s \geq 0$ and $\varepsilon \geq 0$ put $f(s, \varepsilon) := \frac{\mathbb{E}^{X_s}[|X_{s+\varepsilon}| - |X_s|]}{\varepsilon}$. For all $s \geq 0$ we write

$$\frac{\partial}{\partial t} \mathbb{E}[|X_t|] \Big|_{t=s} = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{E}[|X_{s+\varepsilon}| - |X_s|]}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \mathbb{E} \left[\frac{\mathbb{E}^{X_s}[|X_{s+\varepsilon}| - |X_s|]}{\varepsilon} \right] = \lim_{\varepsilon \downarrow 0} \mathbb{E}[f(s, \varepsilon)], \quad (\text{I.3.3.3})$$

where in the second equality we used the Markov property of X . Since by coupling X with a pure death process with rate 1 and a pure birth process with rate β and using that on a tree, in a configuration with N particles there are at most $(N - 1)$ pairs, we get that $|X_s|e^{-\varepsilon} \leq \mathbb{E}^{X_s}[|X_{s+\varepsilon}|] \leq |X_s|e^{\beta\varepsilon}$ and so for all $\varepsilon < 1$ we see that almost surely

$$|f(s, \varepsilon)| \leq |X_s| \max\left(\left|\frac{e^{-\varepsilon} - 1}{\varepsilon}\right|, \left|\frac{e^{\beta\varepsilon} - 1}{\varepsilon}\right|\right) < \infty.$$

Hence, using the fact that $\mathbb{E}[|X_s|] < \infty$ we get by the Lebesgue Theorem that $\lim_{\varepsilon \downarrow 0} \mathbb{E}[f(s, \varepsilon)] = \mathbb{E}[\lim_{\varepsilon \downarrow 0} f(s, \varepsilon)]$. Finally, we get from (I.3.3.1) and (I.3.3.3) that

$$\frac{\partial}{\partial t} \mathbb{E}[|X_t|] \Big|_{t=s} \leq \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}^{X_s}[|X_t|] \Big|_{t=0} \right] \leq 0.$$

Since $\mathbb{E}[|X_t|] = |X_0| + \int_0^t \frac{\partial}{\partial s} \mathbb{E}[|X_s|] ds$, it follows that

$$\mathbb{E}[|X_t|] \leq |X_0| \quad \text{for all } t \geq 0. \quad (\text{I.3.3.4})$$

Showing that X almost surely dies out now follows as in the proof of Lemma 6 in [SS15]. We include this here for completeness. Fix $x \in S$ arbitrarily and put $\mathcal{F}_t^x := \sigma(X_s^x, 0 \leq s \leq t)$, where $(X_t^x)_{t \geq 0}$ is the process $X = (X_t)_{t \geq 0}$ started in $X_0 = x$. Using (I.3.3.4) and the Markov Property of X we can write

$$\mathbb{E}[|X_t^x| | \mathcal{F}_s^x] = \mathbb{E}[|X_{t-s}^{X_s^x}| | X_s^x] = |X_s^x| + \int_0^{t-s} \frac{\partial}{\partial u} \mathbb{E}[|X_u^{X_s^x}| | X_s^x] du \leq |X_s^x|, \quad (\text{I.3.3.5})$$

so we see that $|X_t^x|$ is a supermartingale with respect to \mathcal{F}_t^x . By supermartingale convergence, we get that there exists an \mathbb{N}_0 -valued random variable N such that $|X_t^x| \xrightarrow{t \rightarrow \infty} N$ almost surely. By $\mathcal{A}_T := \{|X_{t-}^x| \neq |X_t^x| \text{ for some } t \geq T\}$ we denote the event that the number of sites occupied by the process X changes at some time $t \geq T$. Furthermore, let $\rho(x)$ be the probability of \mathcal{A}_0 as a function of the initial state A . By the continuity of conditional probabilities with respect to the σ -algebra (for example [Bil86], Theorems 3.5.5 and 3.5.7), we see that for each $S \leq T$, it holds almost surely that

$$\rho(X_T^x) = \mathbb{P}[\mathcal{A}_T | \mathcal{F}_T^x] \leq \mathbb{P}[\mathcal{A}_S | \mathcal{F}_T^x] \xrightarrow{t \rightarrow \infty} \mathbb{P}[\mathcal{A}_S | \mathcal{F}_\infty^x] = 1_{\mathcal{A}_S}.$$

Therefore $\lim_{T \rightarrow \infty} \rho(X_T^x) = 0$ almost surely on the complement of the event $(\bigcap_{S \geq 0} \mathcal{A}_S)$, that is the event

$$\left\{ \lim_{T \rightarrow \infty} \rho(X_T^x) = 0 \right\} \cup \left\{ \forall S \geq 0 \exists t \geq S \text{ s.t. } |X_{t-}^x| \neq |X_t^x| \right\}$$

has probability 1. By the supermartingale convergence we conclude that $\lim_{T \rightarrow \infty} \rho(X_T^x) = 0$ almost surely, which implies that $N = 0$ and X dies out almost surely. ■

I.3.3.2 Proof of Proposition I.3.2.1

In this proof we will use an alternative (and equivalent) notation, used also by Bezuidenhout and Gray in [BG94] for the process (I.1.2.16). Instead of configurations $x \in S^\Lambda$ we will now consider the process to take values in $\mathcal{P}(\Lambda)$, i.e. the space of all subsets of Λ . To differentiate between the notations, we will denote the version of the process on $\mathcal{P}(\Lambda)$ by $(\xi_t)_{t \geq 0}$ instead of $(X_t)_{t \geq 0}$ as we do in the rest of this thesis. The state ξ_t of the process at time $t \geq 0$ then consists of the set of all sites in Λ which are occupied at time t . Formally, we define the set-valued interacting particle system as a $\mathcal{P}(\Lambda)$ -valued continuous time Markov process $\xi = (\xi_t)_{t \geq 0}$ with the generator of the form

$$Gf(\xi) = \sum_{m \in \mathcal{G}} r_m (f(m(\xi)) - f(\xi)), \quad \xi \in \mathcal{P}(\Lambda), \quad (\text{I.3.3.6})$$

where \mathcal{G} is again the set of local maps $m : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$ with the definition of local maps changed correspondingly in an obvious way. We define an *attractive spin system* on \mathbb{Z}^d as a $\mathcal{P}(\mathbb{Z}^d)$ -valued Markov process in continuous time with the following dynamics. For a fixed range $r \geq 0$ we let $N_r = \{y \in \mathbb{Z}^d : \max_i |y_i| \leq r\}$ be the r -neighbourhood of 0 and put $N'_r = N_r \setminus \{0\}$. Finally, for $\xi \in \mathcal{P}(\mathbb{Z}^d)$ and $i \in \mathbb{Z}^d$ we denote by $\xi - i := \{j - i; j \in \xi\}$ the set ξ shifted by i . Now we define

$$\begin{aligned} \beta &:= (\beta(\eta); \eta \subseteq N'_r) \\ \delta &:= (\delta(\eta); \eta \subseteq N'_r) \end{aligned} \quad (\text{I.3.3.7})$$

where $\beta(\eta) \geq 0$ and $\delta(\eta) \geq 0$. For any $\eta \subseteq N'_r$ we will call $\beta(\eta)$ a birth rate as it is the rate at which an empty site $i \notin \xi$ becomes occupied when $(\xi - i) \cap N'_r = \eta$, and we will call $\delta(\eta)$ the death rate, since it is the rate at which an occupied site $i \in \xi$ becomes empty when $(\xi - i) \cap N'_r = \eta$. Furthermore we require that the parameter sets (β, δ) be *attractive*, that is

$$\beta(\eta) \leq \beta(\eta') \quad \text{and} \quad \delta(\eta) \geq \delta(\eta'), \quad \eta \subseteq \eta' \subseteq N'_r. \quad (\text{I.3.3.8})$$

As for example Liggett notes in the Theorem 2.2. of Chapter III in [Lig85], the property (I.3.3.8) is equivalent to the spin system being monotone. Both the contact process and the (nearest neighbour) voter model on \mathbb{Z}^d are attractive spin systems and the same is true for the process with generator (I.1.2.16) on \mathbb{Z}^d with $\gamma = 0$. Notice that when $\gamma > 0$ the cooperative branching process cannot be described by the parameters of the form (I.3.3.7) since the random walk changes the local state of the process at two sites simultaneously and so the process with random walks is not a spin system.

In order to prove Proposition I.3.2.1 we will first need to introduce some terminology used in [BG94]. For the most part we will prove our results for spin systems in \mathbb{Z}^d with translation-invariant symmetric finite-range attractive rates, a class of processes which model (I.1.2.16) on \mathbb{Z}^d belongs to if $\gamma = 0$. For $w = (w_1, \dots, w_d) \in \mathbb{R}^d$, $h > 0$ and $\lambda = (\lambda_1, \dots, \lambda_d)$ we define

$$B(w, h; \lambda) := \{(x, t) \in \mathbb{R}^d \times \mathbb{R}_+ : 0 \leq t \leq h, -w_i \leq x_i - t \tan \lambda_i \leq w_i, i = 1, \dots, d\} \quad (\text{I.3.3.9})$$

as a space-time box with of width $2w_i$ in the i -th spatial coordinate, with height h and angles λ_i of inclination from the vertical direction in the i -th spatial coordinate. For brevity, we will also put $B(w, h) := B(w, h; 0)$. By $\tilde{\xi}^{A,t}(B) = (\tilde{\xi}_s^{A,t}(B))_{s \geq 0}$ we denote the process ξ started at time $t \in \mathbb{R}^+$ in the configuration $A \subseteq \mathbb{Z}^d$ and restricted to the space-time set $B \in \mathbb{Z}^d \times \mathbb{R}^+$ so

¹We could define attractive spin systems on more general lattices, but in this thesis we only need to consider this particular choice of the underlying lattice.

that if $(x, t) \notin B$, then $x \notin \xi_t$ and any event in the process which affects such x at time t is ignored.

Let $\text{pair}_i := \{0, 1_i\}$, where $1_i \in \mathbb{Z}^d$ such that $1_i(i) = 1$ and $1_i(j) = 0$ for all $j \neq i$ and $0 \in \mathbb{Z}^d$ is the origin. For any $A, B \subset \mathbb{Z}^d$ and $x \in \mathbb{Z}^d$ we put $A+x = \{a+x, a \in A\}$ and $A+B = \{a+b, a \in A, b \in B\}$. In Proposition 5.3. of [BG94] Bezuidenhout and Gray have shown the following Proposition with (I.3.3.10) replaced by the assumption that $\xi^{\{0\}} = (\xi_t^{\{0\}})_{t \geq 0}$ survives (or that it is viable as they call this property), i.e. that $\mathbb{P}[\xi_t^{\{0\}} \neq \emptyset \forall t \geq 0] > 0$. The proof of that proposition remains essentially unchanged if we instead assume that the process survives when started in a pair of particles.

Proposition I.3.3.1 *Suppose that $\{\xi_t^A : t \geq 0, A \subseteq \mathbb{Z}^d\}$ is a spin system with symmetric translation-invariant attractive rates (β, δ) with range r such that $\delta(N'_r) > 0$. Further suppose that*

$$\xi^{\text{pair}_i} = (\xi_t^{\text{pair}_i})_{t \geq 0} \text{ survives for all } i = 1, \dots, d, \quad (\text{I.3.3.10})$$

i.e. $P[\xi_t^{\text{pair}_i} \neq \emptyset \forall t \geq 0] > 0$ for all $i = 1, \dots, d$. Then for every $\varepsilon > 0$ and $k \geq 1$ there exists a finite set $D \subseteq \mathbb{Z}^d$, constants $w, w_i > 0$ and $h > 0$ such that, if we put

$$B^0 := B(0, \dots, 0, w_{d-1}, w_d, h),$$

$$B_k^\pm := B\left(3kw_1, \dots, 3kw_{d-2}, 2w_{d-1}, 2w_d, (k+1)h; 0, \dots, 0, \arctan\left(\frac{\pm w_d}{3h}\right)\right),$$

then it holds for every $(x, s) \in B^0$ that with probability at least $1 - \varepsilon$, after a suitable linear change of space-time variables which leaves the time coordinate fixed, the set $\{(y, t) : y \in \tilde{\xi}_t^{D+x, s}(B_k^\pm)\}$ contains a translate of $D \times \{0\}$ lying in the set

$$B^0 + (0, \dots, 0, \pm k \frac{w_d}{3}, kh) + R(k, w),$$

where

$$R(k, w) := \{(x, 0) : x_i \in [-3kw, 3kw] \forall i = 1, \dots, d-2, x_{d-1} = x_d = 0\}. \quad (\text{I.3.3.11})$$

Proof The proof stays the same as the proof of Proposition 5.3. in [BG94]. Proposition I.3.3.1 differs from it only in that we weaken the condition on the survival of the process ξ , namely we only assume that we need to have at least one pair of neighbouring particles to survive, whereas [BG94] require that the process started with a single occupied site does as well. With some minor changes, the proofs of the results leading up to Proposition 5.1. of [BG94] also hold in our setting. Specifically, Lemma 3.14 in [BG94] holds (and the proof stays exactly the same) when we replace the assumption that the process started with a single occupied site survives by the assumption that the process started in any occupied pair does. In Lemma 4.1. the proof stays the same when we replace the requirement that $\xi^{\{0\}}$ survives by the condition (I.3.3.10).

Finally, in the proof of Proposition 5.1. in [BG94] a sequence of space-time points $(x_k, t_k)_{k=0}^{N'}$, where N' is finite and random and (x_k, t_k) is chosen in a particular way prescribed in that proof. A set of events $(A_k(x_k, t_k))_{k=0}^{N'}$, is then defined such that A_k corresponds to the process started at time t_k with only the site x_k occupied produces an occupied translate $D + x_k$ of D within a fixed constant time. Since in our case at least one pair of neighbouring

particles is needed for branching, we start the process in the state $\{x_k, x_k + 1_d\}$ in the event A_k for each k instead of just $\{x_k\}$. ■

Proposition I.3.3.1 tells us that if the process ξ survives with a positive probability for certain parameters β and δ , then for any value of the scaling parameter $k \in \mathbb{N}$, there exists a set D such that (possibly after an appropriate relabelling of coordinates) with a high probability the process ξ , started with the set D occupied and restricted to a box which is tilted in the d -th coordinate and which is narrow in the $(d-1)$ -th and d -th coordinate, will contain a translate of $D \times \{0\}$ in a region at the top of B_k^\pm which is like B^0 , but wider in the first $d-2$ coordinates (see Figure I.3.1).

In the following proposition, following the construction in the Section 6.1. of [BG94], we will compare the process ξ to an oriented site percolation. Let us first define the percolation (on \mathbb{Z}^2 which will be enough for our purposes) and some related terms which we will need in the proof of Proposition I.3.3.5. Let $\mathcal{V} := \{(x, n) \in \mathbb{Z}^2 : x + n \text{ is even}\}$. We fix a so-called percolation parameter $p \in [0, 1]$ and let $(\omega_p^{(x,n)})_{(x,n) \in \mathcal{V}}$ be a collection of i.i.d. Bernoulli random variables with parameter p . We say there exists an open path from site $(x, n) \in \mathcal{V}$ to a site $(y, m) \in \mathcal{V}$ if there exists an $l \in \mathbb{N}$ and a sequence $(x, n) = (x_0, n_0), (x_1, n_1), \dots, (x_{l-1}, n_{l-1}), (x_l, n_l) = (y, m)$ of sites in \mathcal{V} such that $|(x_k, n_k) - (x_{k+1}, n_{k+1})| = 1$ and $\omega_p^{(x_k, n_k)} = 1$ for all $k = 0, \dots, l-1$.

By $i \rightsquigarrow j$ we denote that there exists an open path from i to j and we denote by $0 \rightsquigarrow \infty$ the existence of an infinite open path from the origin $0 \in \mathbb{Z}^2$, that is the existence of a path of length $l \in \mathbb{N}$ from $(0, 0)$ for every $l \in \mathbb{N}$. Using the Bernoulli random variables $(\omega_0^{(x,n)})_{(x,n) \in \mathcal{V}}$ we define the so called oriented percolation process $(W_n)_{n \in \mathbb{N}_0}$ which is a Markov chain given for each initial state $W_0 = A \in Z_{\text{even}} := \{x \in \mathbb{Z} : x \text{ is even}\}$ by

$$W_n^A := \{x \in \mathbb{Z} : (x, n) \in \mathcal{V}, \exists y \in A \text{ s.t. } (y, 0) \rightsquigarrow (x, n)\}. \quad (\text{I.3.3.12})$$

Furthermore, for $(x, n) \in \mathcal{V}$ we denote by $-\infty \rightsquigarrow (x, n)$ the event that there exists an infinite open path starting at time $-\infty$ and ending at (x, n) . Using the same Bernoulli random variables as in the construction of the process (I.3.3.12) we define the process

$$\overline{W}_n := \{x \in \mathbb{Z} : (x, n) \in \mathcal{V}, -\infty \rightsquigarrow (x, n)\}, \quad n \in \mathbb{Z}. \quad (\text{I.3.3.13})$$

Then we obtain the following

Theorem I.3.3.2 *For any $K \geq 0$ and $x \in Z_{\text{even}}$*

$$\lim_{n \rightarrow \infty} \mathbb{P}[W_n^{\{x\}} \neq \emptyset \text{ and } \overline{W}_n \cap [-K, K] \not\subseteq W_n^{\{x\}}] = 0. \quad (\text{I.3.3.14})$$

Proof See page 96 of [SS08] which itself refers to Theorem 5 in [BG90]. ■

For our comparison with the oriented percolation we will obtain a set of so called K -dependent random variables. Let $K, d \in \mathbb{N}$. Then a collection $(X_i)_{i \in \mathbb{Z}^d}$ of random variables is called K -dependent if for any $A, B \subset \mathbb{Z}^d$ with $\inf\{|i - j| : i \in A, j \in B\} > K$ we have that the collections $(X_i)_{i \in A}$ and $(X_j)_{j \in B}$ are independent of each other. We will use the following two well-known theorems.

Theorem I.3.3.3 *There exists a critical parameter $p_c \in [\frac{1}{2}, \frac{80}{81}]$ such that $\mathbb{P}[0 \rightsquigarrow \infty] = 0$ for all $p < p_c$ and $\mathbb{P}[0 \rightsquigarrow \infty] > 0$ for all $p > p_c$.*

Proof See for example [Dur88], Chapter 5a, pages 85-86. ■

Theorem I.3.3.4 *Let $p \in (0, 1)$ and $K < \infty$. Assume that $(X_i)_{i \in \mathbb{Z}^d}$ are K -dependent Bernoulli random variables such that $\mathbb{P}[X_i = 1] \geq p$ for all $i \in \mathbb{Z}^d$. Then we can couple $(X_i)_{i \in \mathbb{Z}^d}$ to a collection of independent Bernoulli random variables $(\tilde{X}_i)_{i \in \mathbb{Z}^d}$ for which*

$$\mathbb{P}[\tilde{X}_i = 1] = (1 - (1 - p)^{1/K})^2, \quad i \in \mathbb{Z}^d$$

in such a way that $\tilde{X}_i \leq X_i$ for all $i \in \mathbb{Z}^d$.

Proof See for example Theorem 7.4 in [Swa17]. ■

Proposition I.3.3.5 *Let $\{\xi_t^A : t \geq 0, A \subseteq \mathbb{Z}^d\}$ fulfill the conditions of Proposition I.3.3.1. Then for all $k \geq 15$ there exists a constant $h > 0$ and a constant c , which is independent of h and k such that with positive probability for infinitely many $n \in \mathbb{N}$,*

$$\{(x, t) \in \mathbb{Z}^d \times \mathbb{R}^+ : x \in \xi_t^{\mathbb{Z}^d}\} \cap A_2 \text{ contains a translate of } \text{pair}_d \times \{0\},$$

where

$$A_2 = A_2^n(c, k, h) := \mathbb{Z}^{d-2} \times [-c, c]^2 \times [nkh, (nk + 1)h]. \quad (\text{I.3.3.15})$$

Proof

We will use Proposition I.3.3.1 to compare the process ξ to an oriented percolation. Fix an $\varepsilon > 0$ and for a given $k \in \mathbb{N}$ choose w, h, D so that the statement of the Proposition I.3.3.1 holds for $(\beta', \delta') = (\beta, \delta)$. We will now define a discrete-time Markov process $\Xi = (\Xi_n)_{n \in \mathbb{N}} := (I_n, P_n)$ which takes values in $\{\{0, 1\} \times (\mathbb{Z}^d \times \mathbb{R}^+)\}^{\mathbb{Z}^{d-1}}$. We note that the spatial dimension of the process Ξ is one lower than the spatial dimension of the original process. Here, the first $d - 2$ spatial coordinates of each process correspond to each other, the $(d - 1)$ st coordinate in Ξ corresponds to the d -th coordinate in ξ and the $(d - 1)$ st coordinate direction of the original process will be restricted to the bounded interval $[-w_{d-1}, w_{d-1}]$ so we no longer keep track of it in the process Ξ . The I_n and P_n (which are defined below) should be thought of as follows. Points $(x, n) \in \mathbb{Z}^{d-1} \times \mathbb{N}$ correspond to space-time boxes in the original process $\{\xi_t^A : t \geq 0, A \subseteq \mathbb{Z}^d\}$. Then if $I_n(x) = 1$, it indicates that an event happened in that space-time box in the original process ξ during which the set D is translated like in Proposition I.3.3.1. If $I_n(x) = 1$, $P_n(x)$ gives us additional information about the location of the translate of D . For $(i, n) \in \mathbb{Z} \times \mathbb{N}_0$ such that $i + n$ is even we put

$$V(i, n) := \left(0, ik \frac{w_d}{3}, nkh\right) + [-w_{d-1}, w_{d-1}] \times [-w_d, w_d] \times [0, h].$$

Let $x = (x_1, \dots, x_{d-1}) \in \mathbb{Z}^{d-1}$. The way we will define $P_n(x)$, if $x_{d-1} + n$ is even and if $I_n(x) = 1$, we will have for $x = (x_1, \dots, x_{d-1}) \in \mathbb{Z}^{d-1}$ that $P_n(x) \in \{x_1\} \times \dots \times \{x_{d-2}\} \times V(x_{d-1}, n)$.

For $(u, t) \in \mathbb{Z}^d \times \mathbb{R}^+$ and $(u_{d-1}, u_d, t) \in V(i, n)$ for some i and n , we define

$$\chi^\pm(u, t) = \left\{ (v, s) \in \mathbb{Z}^d \times \mathbb{R}^+ : v \in \tilde{\xi}_s^{D+u, t} \left((u_1, \dots, u_{d-2}, 0, ik \frac{w_d}{3}, nkh) + B_k^\pm \right) \right\},$$

that is $\chi^\pm(u, t)$ are the sets of all space-time points that occur in the process restricted to a space-time slab which has the same form as B_k^\pm . By $G^\pm(u, t)$ we denote the event that $\chi^\pm(u, t)$ contains a translate of $D \times \{0\}$ lying inside

$$(u_1, \dots, u_{d-2}, (i \pm 1)k \frac{w_d}{3}, (n + 1)kh) + B^0 + R(k, w),$$

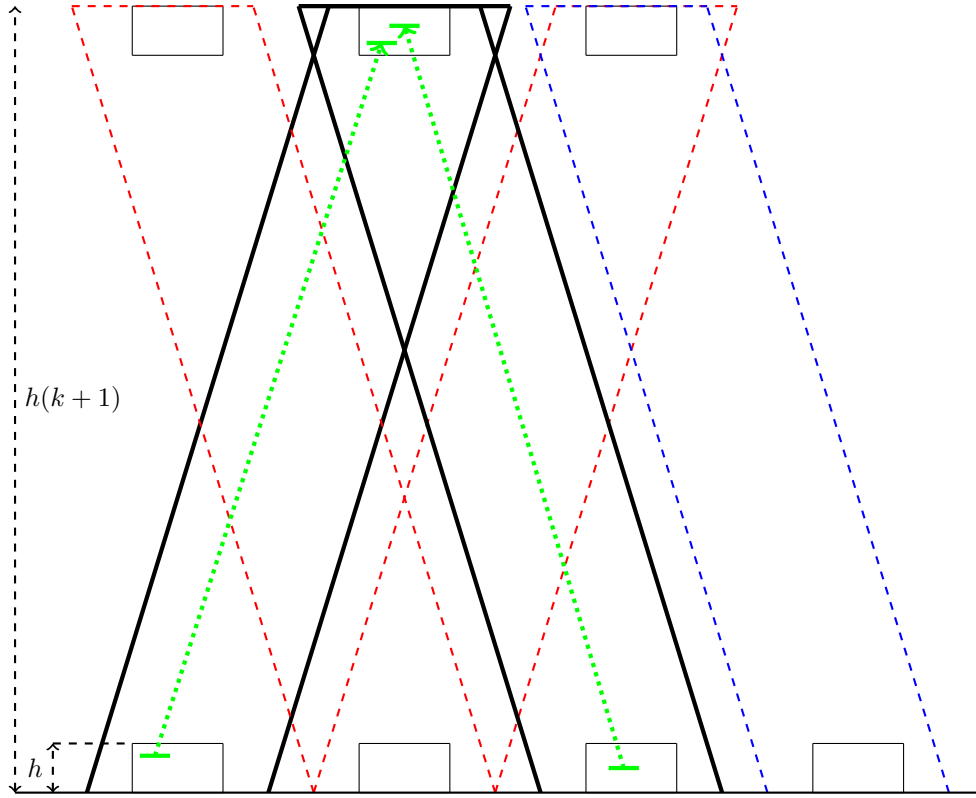


Figure I.3.1: Comparison with oriented percolation (projected to the d -th spatial dimension). Each small rectangle represents a translate of the box B_0 (while the ones at the top of the picture are wider in the first $d - 2$ dimensions) and the long tilted rectangles correspond to space-time slabs B_k^\pm . We denote the translates of $D \times \{0\}$ with a full green line. The events G^\pm are represented by the dashed green arrows. If k is at least 15, the space-time slabs which at the bottom contain different translates of B_0 which are more than two boxes away from each other have an empty intersection and therefore the processes restricted to each of those slabs are independent of each other.

where $R(k, w)$ was defined in (I.3.3.11).

By translation invariance, Proposition I.3.3.1 and the choice of w, h and D , we have for every $(u, t) \in \mathbb{Z}^d \times \mathbb{R}^+$, where $(u_{d-1}, u_d, t) \in \bigcup_{i,n} V(i, n)$, that

$$\mathbb{P}(G^\pm(u, t)) \geq 1 - \varepsilon_0. \quad (\text{I.3.3.16})$$

We will put $\Xi_0 := (I_0, P_0)$, where $I_0(x) = 1_{[x=0]}$ and $P_0(x) = (0, 0) \in \mathbb{Z}^d \times \mathbb{R}_+$ for all $x \in \mathbb{Z}^{d-1}$ (recall that the $(d - 1)$ st coordinate of x corresponds to the d -th coordinate direction in ξ). Now we assume that we have defined the random variables $\Xi_k(x)$, $k = 0, \dots, n$ and $x \in \mathbb{Z}^{d-1}$. Let $x \in \mathbb{Z}^{d-1}$ be such that $x_{d-1} + n + 1$ is even. Unless there exists a $y \in \mathbb{Z}^{d-1}$ such that

$$\begin{aligned} |x_l - y_l| &\leq 3w_l \quad \text{for } l = 1, \dots, d - 2 \\ |x_{d-1} - y_{d-1}| &= 1, \end{aligned} \quad (\text{I.3.3.17})$$

and $I_n(y) = 1$, we define $I_{n+1}(x) := 0$ and $P_{n+1}(x) = (0, 0)$. If such a y does exist and if also

- (i) $y_{d-1} = x_{d-1} \pm 1$,

(ii) The event $G^\mp(P_n(y))$ occurs,

(iii) The lowest translate of $D \times \{0\}$ lying inside

$$\chi^\pm((P_n(y)) \cap [(y_1, \dots, y_{d-2}, 0, x_{d-1}kw_d, (n+1)kh) + B^0 + R(k, w)])$$

is $(D + (x_1, \dots, x_{d-2}, \tilde{x}_{d-1}, \tilde{x}_d)) \times \{\tilde{t}\}$ for some $(\tilde{x}_{d-1}, \tilde{x}_d, \tilde{t}) \in V(x_{d-1}, n+1)$,

then we define $I_{n+1}(x) := 1$ and we choose $P_{n+1}(x) = (z, t)$ so that the lowest translate of $D \times \{0\}$ contained in

$$\{x_1\} \times \dots \times \{x_{d-2}\} \times V(x_{d-1}, n) \cap [\chi^+(P_n(y)) \cup \chi^-(P_n(y))] \quad (\text{I.3.3.18})$$

for some y which satisfies (I.3.3.17) is at $(D + z) \times \{t\}$.

For $x \in \mathbb{Z}$ and $n \in \mathbb{N}$ such that $x+n$ is even we now define random variables $Z_n(x)$ as follows. We put $Z_n(x) = 1$ if $I_n(x_1, \dots, x_{d-2}, x) = 1$ for some $(x_1, \dots, x_{d-2}) \in \mathbb{Z}^{d-2}$ and $Z_n(x) = 0$ otherwise. We note that the conditional distribution of $\Xi_{n+1}(x)$ given $\mathcal{F}_n := \sigma(\Xi_m, 0 \leq m \leq n)$ depends only on the quantities $\Xi_n(y)$ for y such that (I.3.3.17) holds. It follows that for $i \in \{0, 1\}$

$$\mathbb{P}(Z_{n+1}(x) = i | \mathcal{F}_n) = \mathbb{P}(Z_{n+1} = i | \{\Xi_n(y) : |y_{d-1} - x| = 1\}).$$

From (I.3.3.16) we then see that

$$\mathbb{P}(Z_{n+1}(x) = 1 | \mathcal{F}_n) \geq p(Z_n(x-1), Z_n(x+1)),$$

where $p(0, 0) = 0$ and $p(1, 0) = p(0, 1) = p(1, 1) = 1 - \varepsilon$. Let $Q_0(x) := 1$ for all $x \in \mathbb{Z}$. For each $n \in \mathbb{N}_0$ and $x \in \mathbb{Z}$ we now define a random variable $Q_{n+1}(x)$ as follows. If $\max\{Z_n(x-1), Z_n(x+1)\} = 1$, then $Q_{n+1}(x) := Z_{n+1}(x)$ and $Q_{n+1}(x) := 1$ otherwise. The parameter k "rescales" the distances between boxes in which the translation of $D \times \{0\}$ can be found with a high probability. Choosing k large enough, we ensure that if $(u_{d-1}, u_d, t) \in V(i, n)$ and $(v_{d-1}, v_d, r) \in V(j, n)$ for some $n \in \mathbb{N}$ and $i, j \in \mathbb{Z}$ such that $|i-j| \geq 3$ then the space-time slabs to which the process is restricted in the definitions of $\chi^+(u, t)$ and $\chi^-(v, r)$ have an empty intersection (see Figure I.3.1). From the definition of B_k^\pm we obtain that we need to choose $k \geq 15$. Hence, for any n , conditioned on \mathcal{F}_n , the random variables $Q_{n+1}(x), x \in \mathbb{Z}$ are at most 1-dependent. By Theorem I.3.3.4 we can find a collection of independent Bernoulli random variables $(\tilde{Q}_n(x))_{(x,n) \in \mathcal{V}}$ such that

$$\tilde{Q}_n(x) \leq Q_n(x), \quad n \in \mathbb{N}_0, x \in \mathbb{Z} \quad (\text{I.3.3.19})$$

almost surely and $P[\tilde{Q}_n(x)] = (1 - \varepsilon_0)^2$. Choosing ε_0 smaller if necessary, we see by Theorem I.3.3.3 that in an oriented site percolation with $(\omega^{(x,n)})_{(x,n) \in \mathcal{V}} := (\tilde{Q}_n(x))_{(x,n) \in \mathcal{V}}$ it holds that

$$\mathbb{P}[0 \rightsquigarrow \infty] > 0. \quad (\text{I.3.3.20})$$

We now construct the oriented percolation process $(W_n^{\{0\}})_{n \in \mathbb{N}_0}$ as in (I.3.3.12) using Bernoulli random variables $(\omega^{(x,n)})_{(x,n) \in \mathcal{V}} := (\tilde{Q}_n(x))_{(x,n) \in \mathcal{V}}$ and we couple it with the process $(\bar{W}_n)_{n \in \mathbb{N}_0}$ using the same random variables. From (I.3.3.19) and the way we constructed the process $(W_n^{\{0\}})_{n \in \mathbb{N}_0}$ we see that also

$$Z_n(x) = 1 \text{ implies } x \in W_n^{\{0\}} \quad \text{a.s. } \forall n \in \mathbb{N}, x \in \mathbb{Z}. \quad (\text{I.3.3.21})$$

From (I.3.3.20) and Theorem I.3.3.2 we see that $\lim_{n \rightarrow \infty} \mathbb{P}[0 \in W_n^{\{0\}}] > 0$ and by (I.3.3.21) we obtain that also

$$\mathbb{P}[Z_n(0) = 1 \text{ for infinitely many } n \in \mathbb{N}] > 0. \quad (\text{I.3.3.22})$$

We observe that if $Z_n(0) = 1$, then there exists a translate $D'_n \times \{t_n\}$ of $D \times \{0\}$ in A_2 . Since we assume that the process started with pair_d occupied survives and $\text{pair}_d \in D$, there also exists a translate of pair_d in A_2 . \blacksquare

Proposition I.3.3.5 tells us that if we start the process ξ with all sites occupied, there is a positive probability that we can infinitely often find a space-time box centred around $0 \in \mathbb{Z}^d$ and bounded in the d -th spatial dimension and the time dimension, such that each of the space-time boxes contains a translate of $\text{pair}_d \times \{0\}$ and all these boxes are of equal size.

Proposition I.3.3.6 *Let ξ meet the conditions of Proposition I.3.3.1. Then the distribution of $\xi_t^{\mathbb{Z}^d}$ converges to a non-trivial invariant law as $t \rightarrow \infty$.*

Proof It is sufficient to show that we can find a sequence $\{\tau_n\}_{n \geq 0}$ of times where $\tau_n \rightarrow \infty$ as $n \rightarrow \infty$ such that

$$\mathbb{P}[0 \in \xi_{\tau_n} \text{ for all } n \in \mathbb{N}] > a > 0 \quad (\text{I.3.3.23})$$

for some constant a . Indeed, from (I.3.3.23) it follows that the process ξ has a nontrivial upper-invariant law. Indeed, since ξ is an attractive spin system, we already know that the distribution of ξ_t , started at $\xi_0 = \mathbb{Z}^d$ converges to the upper invariant law. If (I.3.3.23) holds, then the upper invariant law cannot be trivial.

Fix $T > 0$. We would like to show that there is a positive probability that at times $\tau_n := t_n + T$ the origin is occupied and this probability is bounded away from zero. By Proposition I.3.3.5 this holds if $d = 2$, since then $A_2^n = A_2^n(c, k, h)$ (with A_2^n as in (I.3.3.15)) is bounded and

$$\max\{|x|; (x, t) \in A_2^n \text{ for some } t \in \mathbb{R}^+ \text{ and } n \in \mathbb{N}\} \leq \sqrt{2c^2}$$

and so there exists a $K \in \mathbb{N}$ independent of n , such that at most K independent events, all of which have a positive probability, are necessary in order for the translate of $\text{pair}_d \times \{0\}$ in A_2^n to "produce" a particle at the origin after time $T > 0$. To prove the same when $d \geq 3$, we have to show that A_2^n can be replaced by

$$A_d^n = A_d^n(c, k, h) := [-c, c]^d \times [nkh, (nk + 1)h], \quad (\text{I.3.3.24})$$

for which

$$\max\{|x|; (x, t) \in A_d^n \text{ for some } t \in \mathbb{R}^+ \text{ and } n \in \mathbb{N}\} \leq \sqrt{dc^2}.$$

It can be checked (for details see the proof of Theorem 2.8. in [BG94]) that the construction leading up to (I.3.3.22) also works if we consider the process $I = (I_n)_{n \in \mathbb{N}}$ from the proof of Proposition I.3.3.5 instead of ξ . More precisely, we obtain a process $\Xi^{(2)} = (\Xi_n^{(2)})_{n \in \mathbb{N}} := (I_n^{(2)}, P_n^{(2)})$ which takes values in $[\{0, 1\} \times (\mathbb{Z}^{d-1} \times \mathbb{R}^+)]^{\mathbb{Z}^{d-2}}$ and random variables $Z_n^{(2)}(x)$, $x \in \mathbb{Z}$ defined as $Z_n^{(2)}(x) = 1$ if $I_n^{(2)}(x_1, \dots, x_{d-3}, x) = 1$ for some $(x_1, \dots, x_{d-3}) \in \mathbb{Z}^{d-3}$ and $Z_n^{(2)}(x) = 0$ otherwise, such that (I.3.3.22) holds with $Z_n(0)$ replaced by $Z_n^{(2)}(0)$ and so we can replace A_2^n in (I.3.3.15) by

$$A_3^n := \mathbb{Z}^{d-3} \times [-c, c]^3 \times [nkh, (nk + 1)h].$$

Since $I^{(2)}$ is the same type of process as I , we can repeat the same construction for $I^{(2)}$ and by induction we see that A_2^n in (I.3.3.15) can indeed be replaced by A_d^n as defined in (I.3.3.24). \blacksquare

Finally, the process with generator (I.1.2.16) with $\delta > 0$, $\beta > 0$, $\gamma = 0$ and $\alpha \geq 0$ is an attractive spin system with symmetric translation-invariant rates which fulfils condition

(I.3.3.10) and so it meets all the assumptions of Proposition I.3.3.1. Therefore, by Proposition I.3.3.6 it has a nontrivial upper invariant law for any choice of β such that $\beta > \beta_{\text{surv}}$. It follows immediately that $\beta_{\text{surv}} \geq \beta_{\text{upp}}$, showing Proposition I.3.2.1. \blacksquare

I.3.3.3 Proof of Proposition I.3.2.2

In Proposition I.3.2.2 we consider the process X with generator as in I.1.2.16 with $\alpha = 0$, $\gamma = 1$ and $\delta > 0$. This process can be coupled with a cooperative branching process $\hat{X} = (\hat{X}_t)_{t \geq 0}$ which has the same parameters α, γ and δ as X and for which $\beta = 0$, so that $X_t \geq \hat{X}_t$ almost surely. It holds then that

$$\mathbb{P}^{\text{pair}}[|X_t| \geq 1] \geq \mathbb{P}^{\text{pair}}[|\hat{X}_t| \geq 1] = \mathbb{P}^{\text{pair}}[|\hat{X}_t| = 2] + \mathbb{P}^{\text{pair}}[|\hat{X}_t| = 1].$$

If $\delta = 0$, the probability $\mathbb{P}^{\text{pair}}[|\hat{X}_t| = 2]$ would correspond to the probability that two independent random walks started next to each other do not meet until time t . Sturm and Swart show in Lemma 9 of [SS15] that this probability is equal to $kt^{-1/2}$, where k is a constant. Since we assume that $\delta > 0$, the process \hat{X} survives with at least two particles only if moreover neither of the two random walkers dies before time t . Since the paths of the random walkers prior to their coalescence are independent, the death events on their paths are also and we may therefore write

$$\mathbb{P}^{\text{pair}}[|\hat{X}_t| = 2] \geq kt^{-1/2}e^{-2\delta t}. \quad (\text{I.3.3.25})$$



(a) Particles coalesce before any deaths occur and before time t

(b) Coalescence happens after the death of one particle and before time t



(c) One particle dies before the time of coalescence which happens before time t

(d) One particle dies before time t and coalescence happens after t

Figure I.3.2: Events that lead to the survival of exactly 1 particle in $\bar{\eta}$ at time t

Since the Poisson processes that form the graphical representation of \hat{X} are all independent and identically distributed and the branching rate of \hat{X}_t is zero, the survival of exactly one

individual at time t has the same probability as the survival of a simpler model with only two sites 0 and 1, where the deaths occur at each site as in \hat{X}_t and the two individuals coalesce after time $\tau_{0,1}$ either by the individual at site 0 moving to 1 or vice versa. We define $\tau_{0,1}$ as the first meeting time of two independent random walkers on \mathbb{Z} which are started next to each other. By τ_i^\dagger we denote the time of the first death event at site $i, i = 0, 1$ in this simpler model. There are four possible type of events that lead to the survival of exactly one individual at time t . Namely, these events are as follows. First possibility is that the two particles coalesce before either of them dies and then the single remaining particle survives until time t . This is shown in Figure I.3.2a. In the second and third possible event the coalescence would have only happened after one of the particles has already died and the second particle survives until time t as shown in Figures I.3.2b and I.3.2c. The difference between the second and third event is in that in I.3.2b it is the surviving particle that jumps (which would have resulted in a coalescence if both particles were still alive) and in the third case the particle which dies is the one that would have jumped and coalesced with the surviving one. Finally, the last type of event is one in which the coalescence would have only happened after time t , one of the two particles dies before time t and the other one survives as in Figure I.3.2d. In each of the pictures, red lines represent the time during which the individual has to stay alive before time t while thick black lines represent the presence of a living individual otherwise, coalescence events are denoted by horizontal arrows (dashed if they have no influence on survival) and crosses represent death events. Each of these events can also happen with the roles of 0 and 1 reversed with the same probability (since τ_0^\dagger and τ_1^\dagger are independent and identically distributed). One of the first three possible events shown in Figures I.3.2a, I.3.2b and I.3.2c occurs if there is a coalescence before time t and if there also exists a path of length t with no death events on it. This has the probability

$$\mathbb{P}[\tau_0^\dagger > t > \tau_{0,1}].$$

The last possibility, corresponding to the Figure I.3.2d is that coalescence only happens after time t and therefore one individual has to survive until time t , whereas the other has to die before that time. This has the probability

$$2\mathbb{P}[\tau_0^\dagger > t, \tau_1^\dagger < t < \tau_{0,1}].$$

where the factor 2 appears because in this case we do not care whether the individual from site 1 moves to 0 or the other way around (and we still can have either the individual at 0 or the individual at 1 survive with equal probabilities). Going back to process \hat{X} , we get that

$$\begin{aligned} \mathbb{P}^{\text{pair}}[\hat{X}_t = 1] &= \mathbb{P}[\tau_0^\dagger > t > \tau_{0,1}] + 2\mathbb{P}[\tau_0^\dagger > t, \tau_1^\dagger < t < \tau_{0,1}] \\ &= e^{-\delta t}(1 - kt^{-1/2}) + 2e^{-\delta t}(1 - e^{-\delta t})kt^{-1/2} \\ &= e^{-\delta t}(1 + kt^{-1/2} - 2e^{-\delta t}kt^{-1/2}) \end{aligned}$$

Together with (I.3.3.25) we finally obtain

$$\mathbb{P}^{\text{pair}}[\hat{X}_t \geq 1] = e^{-\delta t} + e^{-\delta t}kt^{-1/2} - e^{-2\delta t}kt^{-1/2} \geq e^{-\delta t}.$$

■

I.3.3.4 Proof of Proposition I.3.2.3

By Theorem 4, [SS15] we see that if $\delta = 0$, then the left-hand side of (I.3.2.2) could be estimated from below by $kt^{-1/2}$, where k is a constant dependent on t_0 . Let us first consider the process

X without cooperative branching and deaths, i.e. the process with generator (I.1.2.16) on \mathbb{Z} with $\beta = \delta = 0$ in addition to $\alpha = 0$ and $\gamma = 1$. This process is then just a process of (symmetric) coalescing random walks on \mathbb{Z} . Let $N_i^{\rightarrow}, i \in \mathbb{Z}$ and $N_i^{\leftarrow}, i \in \mathbb{Z}$ all be independent Poisson processes with intensity $1/2$ and denote by ω^{\rightarrow} (resp. ω^{\leftarrow}) the set of Poisson events corresponding to N_i^{\rightarrow} (resp. N_i^{\leftarrow}). Then the Poisson processes $N_i^{\rightarrow}, i \in \mathbb{Z}$ and $N_i^{\leftarrow}, i \in \mathbb{Z}$ define the graphical representation of X as follows. If $s \in \omega_i^{\rightarrow}$ for some $s > 0$ and $i \in \mathbb{Z}$ then we set

$$X_s(j) := \begin{cases} 0 & \text{if } j = i, \\ X_s(i) \vee X_s(j) & \text{if } j = i + 1, \\ X_{s-}(j) & \text{otherwise} \end{cases} \quad (\text{I.3.3.26})$$

and analogously for N_i^{\leftarrow} . If $s \notin \omega_i^{\rightarrow} \cup \omega_i^{\leftarrow}$ for any i , then we put $X_s := X_{s-}$. For a fixed $t \geq 0$ we define an open path on an interval $I \subset \mathbb{R}$ as a càdlàg function $\xi : I \mapsto \mathbb{Z}$ which satisfies the following conditions:

- (i) If $t \in \omega^{\leftarrow}$ (resp. $t \in \omega^{\rightarrow}$) for some $t \in I$ then $\xi_t = \xi_{t-} - 1$ (resp. $\xi_t = \xi_{t-} + 1$)
- (ii) If $t \in I \setminus (\omega^{\leftarrow}(\xi_{t-}) \cup \omega^{\rightarrow}(\xi_{t-}))$ then $\xi_t = \xi_{t-}$.

In the process of coalescing random walks (i.e. the model with $\alpha = \beta = \delta = 0$ and $\gamma = 1$) the probability $\mathbb{P}^{1\mathbb{Z}}[X_t(0) = 1]$ corresponds to the probability that there exists an open path $\xi : [0, t] \mapsto \mathbb{Z}$ such that $\xi_t = 0$. Sturm and Swart show in Section 3.2. of [SS15] that this probability can be estimated from below by $kt^{-1/2}$. Since the cooperative branching process is monotone, this is also the lower estimate for the same probability in the case when $\beta > 0$.

If $\delta > 0$, we also need that there exist a path $\xi' : [0, t] \mapsto \mathbb{Z}$ with no death events on it, that is whenever $\xi'_s = i$ for some $s \geq 0$ and $i \in \mathbb{Z}$, no death events occur at time s at site i . Since the death events occur independently at every site at a rate δ , the probability that there are no deaths on ξ' is $e^{-\delta t}$. We can therefore bound the left hand side of (I.3.2.2) from below by $kt^{-1/2}e^{-\delta t}$ which can then be further estimated from below by $ce^{-\delta' t}$ for any $\delta' > \delta$ and a constant $c = c(\delta', t_0)$. ■

I.3.3.5 Proofs of Propositions I.3.2.4 and I.3.2.5

In both proofs we will use the following simple Lemma.

Lemma I.3.3.7 *Let $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ be a continuously differentiable function such that*

$$\frac{\partial}{\partial t} f(t) \leq -af(t), \quad t \geq 0 \quad (\text{I.3.3.27})$$

for some $a > 0$. Then

$$f(t) \leq f(0)e^{-at}, \quad t \geq 0.$$

Proof This is a special case of Gronwall's inequality. Since the proof is very short, we provide it here for completeness. From (I.3.3.27) we get that that

$$\frac{\partial}{\partial t} (e^{at} f(t)) = e^{at} \left(\frac{\partial}{\partial t} f(t) + af(t) \right) \leq 0, \quad t \geq 0.$$

Hence $\int_0^t \frac{\partial}{\partial s} (e^{as} f(s)) ds \leq 0$ and so also $f(t)e^{at} - f(0) \leq 0$ from which the claim follows. ■

Proof of Proposition I.3.2.4 For $x_0, \dots, x_n \in \{0, 1\}$ put $p_t(x_0 x_1 \cdots x_n) := \mathbb{P}^{1\mathbb{Z}}[\{X_t(i) = x_0, X_t(i+1) = x_1, \dots, X_t(i+n) = x_n\}]$. The definition does not depend on the choice of $i \in \mathbb{Z}$

since the process and the initial law $\delta_{1_{\mathbb{Z}}}$ are translation invariant. From the generator (I.1.2.16) of X we obtain that

$$\begin{aligned}
\frac{\partial}{\partial t} p_t(1) &= -p_t(1) + \frac{1}{2}p_t(10) + \frac{1}{2}p_t(01) + \frac{1}{2}\beta p_t(110) + \frac{1}{2}\beta p_t(011) - \delta p_t(1) \\
&= -p_t(11) + \beta(p_t(11) - p_t(111)) - \delta p_t(1) \\
&= (\beta - 1)p_t(11) - \beta p_t(111) - \delta p_t(1) \\
&\leq -\delta p_t(1),
\end{aligned} \tag{I.3.3.28}$$

where the second equality follows from the fact that $p_t(110) = p_t(011) = p_t(11) - p_t(111)$ and $p_t(10) = p_t(01) = p_t(1) - p_t(11)$ and the inequality from the fact that $\beta \leq 1$. The rest of the proof follows from Lemma I.3.3.7 and the fact that $\mathbb{P}^{1_{\mathbb{Z}}}[X_t(0) = 1] = 1$. \blacksquare

Proof of Proposition I.3.2.5 For $\beta \leq 1$ we get from (I.1.2.16) that

$$\begin{aligned}
\frac{\partial}{\partial t} \mathbb{E}^{\text{pair}}[|X_t|] &= (\beta - 1) \sum_{i \in \mathbb{Z}} \mathbb{P}^{\text{pair}}[X_t(i) = X_t(i+1) = 1] \\
&\quad - \beta \sum_{i \in \mathbb{Z}} \mathbb{P}^{\text{pair}}[X_t(i) = X_t(i+1) = X_t(i+2) = 1] - \delta \sum_{i \in \mathbb{Z}} \mathbb{P}^{\text{pair}}[X_t(i) = 1] \\
&\leq -\delta \sum_{i \in \mathbb{Z}} \mathbb{P}^{\text{pair}}[X_t(i) = 1].
\end{aligned} \tag{I.3.3.29}$$

Here, the inequality follows from the fact that $\beta \leq 1$. Since $\sum_{i \in \mathbb{Z}} \mathbb{P}^{\text{pair}}[X_t(i) = 1] = \mathbb{E}^{\text{pair}}[|X_t|]$, we then get by Lemma I.3.3.7 that $\mathbb{E}^{\text{pair}}[|X_t|] \leq \mathbb{E}^{\text{pair}}[|X_0|]e^{-\delta t}$ for all $t \geq 0$. Since $\mathbb{P}^{\text{pair}}[|X_t| \geq 1] \leq \mathbb{E}^{\text{pair}}[|X_t|]$, and $\mathbb{E}^{\text{pair}}[|X_0|] = 2$, this completes the proof. \blacksquare

Part II

Genealogy of neutral loci in a random environment

Chapter II.1

Overview and previous results

In the second part of the thesis, we will be concerned with a generalization of a particular model from population genetics which was studied by Barton, Etheridge and Sturm in the paper [BES04]. They consider a population of individuals which reproduce and die. Their model then describes a genealogy of a sample of neutral genes which are in each individual encoded on a chromosome in close vicinity to a gene which has two variants and which is influenced by selection. This genealogy is further influenced by a phenomenon called *recombination* in which, roughly speaking, the chromosomes of the parents are split and rearranged into a new one which then forms the chromosome of their offspring. The difference between the model studied in [BES04] and the one considered here is that in [BES04] each individual is considered to carry only one of these neutral genes, whereas we will study a model in which there are K neutral genes for some fixed $K \in \mathbb{N}$. This changes somewhat the structure of the genealogy.

We will first define more precisely the biological terms we mentioned above as well as some others that we will be using in the thesis. We say that an organism is *haploid* if it only has a single chromosome and *diploid* if it has two chromosomes (other numbers of chromosomes also occur in nature). In our model, we will consider a population of haploid individuals. We can also interpret the model as an approximation for a diploid population with half as many individuals in the sense that in the haploid model we forget the way chromosomes are paired together within the diploid organisms and consider each chromosome separately as an independent individual.

Different variants of a single gene are called *alleles*, so that for example a gene coding for the eye colour can have several alleles each of which manifests itself in a particular eye colour of the animal with that allele.

A particular location on the chromosome where a gene is encoded is called a *locus*. We refer to a locus as *selective* if the gene at that locus is influenced by natural selection in the sense that changing the allele of the gene at that locus changes the fitness of the individual carrying it. We say that a locus is neutral if the allele has no influence on the *fitness* of the individual. In our model we will keep track of a single selective locus on a chromosome with a fixed number of neutral loci which lie close to it.

In sexual reproduction in diploid organisms, the individuals form haploid cells called gametes (eggs and sperm) which are obtained from the organism's diploid cells (i.e. cells which contain both of its chromosomes) via *recombination*. We can picture recombination as follows. The two chromosomes are lined up next to each other and cut in several places (the same places on both chromosomes). After that each cut segment of a chromosome switches places with the corresponding segment on the other chromosome, so that two new chromosomes consisting of

parts of each of the two original chromosomes are created by this process. Two gametes are therefore produced in this way, each carrying a single chromosome. As was mentioned above, in our model we treat the diploid population as a haploid population with twice as many individuals. From this point of view, the two chromosomes in the recombination can be seen as the "parents" of the gamete "offspring" and the recombination can happen between any two individuals instead of between the two chromosomes of each individual as in the diploid case. In our setting, where we want to study the genealogy of the neutral loci in a finite sample of individuals, the main consequence of recombination is that some or all of the neutral loci might not necessarily trace their ancestry to the same parent as the selective locus. This is important since the allele at the selective locus influences the fitness of the individual and it might be different in each parent. In our model, we will call the parent which is ancestral to the gene at the selective locus the *first parent* and we will call the other one the *second parent*.

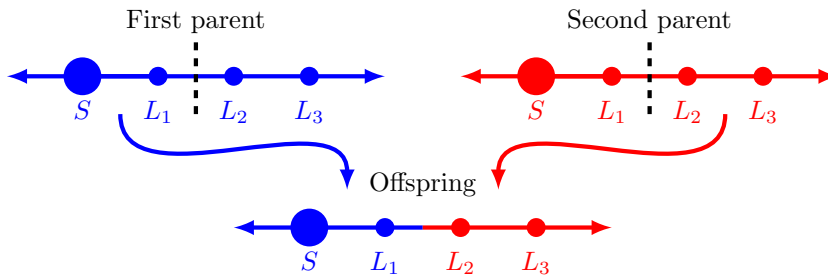


Figure II.1.1: An example of a recombination. Here, the big ball denoted by the letter S represents a selective locus and the smaller balls denoted by L_i , $i = 1, 2, 3$ represent three neutral loci which are close to the selective locus. The two "parent" chromosomes are cut at the same location indicated by the dashed line and the "offspring" is created by combining the part of the chromosome to the left of the cut in one parent and the rest from the other parent.

Finally, we assume that the gene at the selective locus undergoes *mutation*. This means that the gene at the selective locus of the offspring mutates, i.e. its allele changes to one which is different from the allele of the first parent. In our model, we will consider a simple case where the gene at the selective locus only has two alleles called P and Q and a mutation always either changes P to Q or the other way around.

Before we introduce the full model with selective and neutral loci, recombination and mutation, we will describe a simple population model which our model is based on, namely the so called *Moran model*. Assume we have a population of $N \in \mathbb{N}$ individuals such that each individual is either of type P or of type Q . We assume that the population evolves (in continuous time) in the following way. Each individual dies at rate 1 independently of all other individuals. When that happens, the individual is replaced by a new individual whose type is chosen uniformly from the current population. In other words, at rate N a so called *birth-death event* happens in which two individuals are chosen uniformly at random from the population. One individual from the pair is then chosen with equal probability. This individual dies and is replaced by a copy of the other individual. We will refer to the individual which copied itself (reproduced) as the parent and to its copy as the offspring of the birth-death event.

We notice that the Moran model is just a special case of the voter model (I.1.2.8) which we defined in Part I of the thesis. Indeed, the Moran model above can be formally defined as the nearest neighbour voter model on a complete graph with $2N$ vertices and a local state $S := \{P, Q\}$ or more precisely as this process modulo permutation of sites, since we do not

care about the particular labelling of the individuals.

In our model we will consider a version of the Moran model where the population is subject to mutation and selection which can depend on the ratio of each allele in the population. This no longer corresponds to the voter model since even if the entire population has the same type at a particular birth-death event, the offspring of the birth-death event can end up with the other type. In contrast to that, in the voter model the process remains constant once the local state is the same at every site.

In what follows, we will be studying a fixed size population of individuals who reproduce according to the Moran model with selection and mutation. The individuals consist of a selective locus with alleles P and Q , which influence the individual's chance to reproduce during a birth-death event, and K neutral loci which do not affect the reproduction chances and which are situated close to the selective locus. We assume that the alleles at the selective locus of the offspring of a birth-death event can mutate to the allele different from its parent with a positive probability for both alleles and that there can be a recombination between the parent and the deceased individual so that some (or all) of the neutral loci of the offspring trace their genealogy to the dead individual instead of the parent.

We assume that the population has evolved for a long time and the process which keeps track of the ratio of the alleles in the population has reached stationarity. We then take a sample of a finite number of individuals from the population and study the genealogy of their neutral loci. To do that, we will construct a process which keeps track of the distribution of the ancestry of the neutral sample and the allele ratio backwards in time. The main result of part II is Theorem II.2.3.1 where we show the convergence of this process to a diffusion approximation obtained in the limit when the total size of the population grows beyond all bounds and time is appropriately rescaled, and we describe the limiting process by the means of its generator. Note that while we assume the population to evolve according to the Moran model (with selection and mutation), previous results in the simple setting without selection and mutation show that the same approximation will be obtained for a large class of models, see for example [MS98]. One such model is the Wright-Fisher model in which the entire population is replaced at each birth-death event and in which each individual "chooses" its parent uniformly at random from the previous generation.

Barton, Etheridge and Sturm study this model in [BES04] in the case when $K = 1$, i.e. when there is only a single neutral locus. The main difference between $K = 1$ and $K \geq 2$ is that when $K = 1$ the neutral locus in each individual from the sample can trace their ancestry to only a single individual, whereas if $K \geq 2$ the neutral loci in an individual can "split" due to recombination so that some of the neutral loci trace their ancestry to one individual and a different part to a different one. In other words one can say that if $K = 1$, then backwards in time the neutral individuals carrying the neutral sample can only coalesce (when two individuals from the sample and so also their neutral loci trace their genealogy to the same parent), whereas in the case when $K \geq 2$ each individual may also split in the sense above. From the point of view of applications, keeping track of multiple neutral loci also allows us to study certain additional characteristics, such as pairwise differences between the sampled individuals or the correlation between coalescence times of different neutral loci. Unfortunately, while it is theoretically possible to obtain some analytical results about these characteristics the same way as in [BES04], it becomes exceedingly technically difficult as K increases. Given that, we limit ourselves to simulations in the applications we consider in Section II.2.4. For a further discussion of the results of [BES04], particularly the equations

of identities, see the accompanying paper [BE04]. There is a number of related results in the literature. Taylor [Tay07] constructs the so-called common ancestor process which keeps track of the background of an individual who eventually becomes the common ancestor of the entire population, and discusses the model with multiple (i.e. more than two) alleles at the selective locus. Kluth, Hustedt and Baake [KHB13] expand on [Tay07] as well as [Fea02]. Pokalyuk and Pfaffelhuber [PP13] study the fixation of a beneficial allele in the regime of a strong directional selection (and no mutation). In the case when $K = 2$, an analytical treatment of the characteristics considered in Section II.2.4 in a similar model can be found in [BBE13]. The model of that paper differs from ours most notably in two ways. It does not assume a random environment (i.e. the selective locus), but on the other hand it has a more complicated birth-death mechanics where in addition to the birth-death events which we consider, there is also a possibility of a "large" birth-death event in which a certain (possibly random) portion of the population is replaced at once.

Similar models are often studied with the assumption of a selective sweep, see for example [DS05], [EPW06], [PHW06], [PS07]. [HP08] or the dissertation [BS14]. In that setting, a beneficial mutation is assumed to have happened in the past at the selective locus which spreads through the population and eventually fixates (which never happens in our model). [IS03] provides a method to distinguish between background selection with selective sweeps in practice.

Part II is organized as follows. In Section II.2.1 we describe our model with multiple neutral loci, mutation and recombination in detail. In Section II.2.2 we obtain the transition rates of the model which keeps track of the allele ratio and the sampled individuals backwards in time and in Section II.2.3 we show the convergence to the diffusion approximation whose generator is given in Theorem II.2.3.1. Section II.2.4 discusses possible applications of this approximation and illustrates some of its behaviour obtained by simulations. Finally, in Section II.2.5 the proofs of the results from previous sections are provided.

Chapter II.2

Multiple neutral loci

II.2.1 Model

Let $N \in \mathbb{N}$ be the (fixed) size of a population of diploid individuals which we will from here on treat as a population of $2N$ haploid individuals instead. We will denote the two alleles at the selective locus by letters P and Q , respectively. We will say that an individual comes from the *background* P (or that it is a P -background individual) if it has the P allele present at its selective locus. Similarly, will say that it comes from the background Q if it has the Q allele at the selective locus.

We now define p_t , $t \geq 0$ as the proportion at time t of the P -background individuals in the whole population. Similarly, $q_t = 1 - p_t$, $t \geq 0$ will be the proportion of Q -background individuals. We put $p := (p_t)_{t \geq 0}$ and $q := (q_t)_{t \geq 0}$ and call p the process of *allele ratios*.

Let $S : [0, 1] \mapsto [-1, 1]$ be a function. If two individuals from different backgrounds are chosen in a birth-death event and the ratio of the P -background in the population is p , we will assume that with the probability $\frac{1+S(p)}{2}$ it is the P -background individual which reproduces and the Q -background individual which dies. We will call S the *selection coefficient*.

Due to mutation, the offspring of a birth-death event can have a different allele at the selective locus than its parent. More precisely, an offspring of a P -background parent will be a Q -background individual with the probability $\mu^P > 0$ and similarly an offspring of a Q -background individual will have the allele P at its selective locus with probability $\mu^Q > 0$. We will call μ^P and μ^Q the *mutation probabilities* or *mutation rates*. Note that since the mutation rates are both assumed to be positive, neither allele ever fixates in the population.

Finally, to use same version of the Moran model as in [BES04], we will assume that branching-death events only happen at rate N instead of $2N$ as would be the case in the simple version of the Moran model we introduced in II.1 with a population of size $2N$.

As we mentioned in Chapter II.1, we want to study the genealogy of a neutral sample taken from a population after reaching stationarity. To do that, we will need to keep track of the allele ratio p and of those individuals (in each background and through time backwards) to which at least one of the sampled neutral loci trace their genealogy. As a shorthand, we will refer to these individuals as the individuals that come from the sample and we will refer to those of its neutral loci which are ancestral to the neutral sample at present as *active* loci. We assume there are K neutral loci situated close to the selective locus and we denote them by L_1, \dots, L_K . To differentiate between individuals from the sample with different active (neutral) loci, we will need the notion of a neutral type. Let $\mathcal{I} := \mathcal{P}(\{1, \dots, K\}) \setminus \emptyset$. We say that an individual from the sample is of the (*neutral*) *type* $I \in \mathcal{I}$ if and only if its neutral

loci with indices in I are active and no other neutral loci are. Denote by n_I^P (resp. n_I^Q) the number of individuals from the sample that are of type I and from background P (resp. Q). For example if $I = \{2, K\}$, then n_I^P is the number of individuals from the sample that are in the background P and whose set of active loci consists of the second and K -th neutral locus.

We need to keep track of the numbers of all possible kinds of individuals from the sample. We define $\mathfrak{n}^P := \{n_I^P\}_{I \in \mathcal{I}}$ as the collection of numbers of individuals from the sample in background P and corresponding to each neutral type. Analogously we define $\mathfrak{n}^Q := \{n_I^Q\}_{I \in \mathcal{I}}$ as the collection of numbers of individuals from the sample in background Q corresponding to each neutral type.

We will study the (backwards in time) process $\mathbb{X}^{(1)} = (p_t; \mathfrak{n}^P(t), \mathfrak{n}^Q(t))_{t \geq 0}$ which keeps track of the allele ratio p and the distribution of the ancestral sample in the population as well as the background of all individuals which carry active loci.

In a birth-death event, viewed backwards in time, the neutral loci in an individual from the sample either "move" to (i.e. trace their genealogy to) the parent if there was no recombination, or some or all of them move to the individual who died during the event. Whenever there is a recombination, we will call the parent of the selective locus the *first parent* and we will refer to the individual who dies in the birth-death event and to whom a part of the neutral sample traces its genealogy as the *second parent*. Since we will not differentiate between two individuals from the same background and with the same neutral type, a recombination which moves all the active neutral loci from the offspring to the second parent affects the process $\mathbb{X}^{(1)}$ only when the backgrounds of the offspring and of the second parent differ. However, when a recombination occurs, which causes some but not all of the active loci to move to the second parent, the ancestry of the neutral sample splits and we have to keep track of both the first and the second parent regardless of their backgrounds.

We need to define the recombination probabilities. If $\hat{I} \in \mathcal{I}$ then by $r_{\hat{I}}$ we denote the probability of a recombination which results in (viewed backwards in time) moving the neutral loci with indices in \hat{I} to the second parent. We assume that $\sum_{\hat{I} \in \mathcal{I}} r_{\hat{I}} \leq 1$. Naturally, the recombination only has an effect on the genealogy of the sample if at least some of the neutral loci that are to move to the second parent are active. Therefore, we define the *recombination probabilities* for an individual of the neutral type I as

$$r_{I, \hat{I}} := \sum_{J \in \mathcal{I}} r_J \mathbf{1}_{[I \cap J = \hat{I}]}, \quad \hat{I} \in \mathcal{I}$$

and by $r_{I, \emptyset}$ we denote the probability that there is no recombination in an individual of (neutral) type I , that is

$$r_{I, \emptyset} := 1 - \sum_{\substack{\hat{I} \in \mathcal{I} \\ \hat{I} \subseteq I}} r_{\hat{I}} = 1 - \sum_{\hat{I} \in \mathcal{I}} r_{I, \hat{I}}. \quad (\text{II.2.1.1})$$

A natural way to define the recombination probabilities (for a given geometry) is given by the following example.

Example II.2.1.1 Assume that all K loci are all aligned to the right of the selective locus. We will also assume that they are ordered by their distance from the selective locus, i.e. L_i is closer to the selective locus than L_{i+1} for all $i \in [K-1]$. For $i \in [K] := \{1, \dots, K\}$ put $c_i \in [0, 1]$ arbitrarily and denote by L_0 the selective locus. These will be the probabilities of a chromosome being cut in between two neutral loci, so that a cut between L_{i-1} and L_i occurs with probability c_i . Assume that these events are independent and for notational purposes

put $c_{K+1} := 1$. Ordered by their distance from the selective locus, every odd cut causes all the neutral loci to the right of it until the next cut to move to the second parent and every even cut causes everything to the right of it, up to the next cut, to move to the first one. For example, if three cuts appear during a single recombination event, such that their indices are i, j and k where $i < j < k$, then all neutral loci with indices in the set $\{i, \dots, j-1\}$ move to the second parent, all neutral loci with indices in $\{j, \dots, k-1\}$ move to the first one and all those with indices in $\{k, \dots, K\}$ again move to the second parent. Set $\inf(\emptyset) := \infty$. Now for $I \in \mathcal{I}$, we put

$$\begin{aligned}\hat{\beta}_0^I &:= 0 \\ \hat{\alpha}_0^I &:= \min\{i \in I\}, \\ \hat{\beta}_j^I &:= \inf\{i > \hat{\alpha}_{j-1}^I, i \notin I\}, \quad j = 1, \dots, K, \\ \hat{\alpha}_j^I &:= \inf\{i > \hat{\beta}_j^I, i \in I\}, \quad j = 1, \dots, K\end{aligned}$$

and for $j = 0, 1, \dots, K$

$$\alpha_j^I := \begin{cases} \hat{\alpha}_j^I, & \hat{\alpha}_j^I < \infty \\ K+1, & \text{otherwise,} \end{cases} \quad \beta_j^I := \begin{cases} \hat{\beta}_j^I, & \hat{\beta}_j^I < \infty \\ K+1, & \text{otherwise.} \end{cases}$$

Then we obtain the recombination probabilities r_I as

$$r_I = \prod_{i=0}^K \left(c_{\alpha_i^I} c_{\beta_{i+1}^I} \prod_{\substack{j=\beta_i^I+1 \\ j \neq \alpha_i^I}}^{\beta_{i+1}^I-1} (1 - c_j) \right), \quad I \in \mathcal{I}.$$

Remark II.2.1.2 The situation is simpler in the case where we only have one neutral locus (i.e. $K = 1$). There we observe the (backwards in time) process $X = (p_t; n_t^P, n_t^Q)_{t \geq 0}$, where n_t^P is the number of individuals at time t whose neutral locus is ancestral to the sample and similarly for n_t^Q . With just one neutral locus there is only one possible recombination event which affects the process X , namely the one which moves the active locus (there can only be one in this case) of an individual from the sample to the second parent. We assume that the recombination happens in a birth-death event with a constant probability r . This case is treated in [BES04].

II.2.2 Finite population size process

We would like to obtain the generator of the process $\mathbb{X}^{(1)}$, but because of the abundance of possible events that affect the process, we will not attempt to write down the generator in a closed form (we will do that later with a generator of the diffusion approximation) and limit ourselves mostly to describing those events which will also play a role later in the diffusion limit and noting that all the others will not appear there. We will proceed the same way as in the Lemma 2.4. in [BES04]. The argument is based on the reversibility of the Moran model.

Denote by p_m and p_{m+1} the P allele frequency immediately (backwards in time) before and after a birth-death event and let us define $P_{p, \tilde{p}} := \mathbb{P}[p_{m+1} = \tilde{p} | p_m = p]$. It is easy to see

that it holds for these probabilities that

$$\begin{aligned}
P_{p,p} &= p^2(1 - \mu^P) + (1 - p)^2(1 - \mu^Q) + 2p(1 - p) \left(\frac{1 + S}{2} \mu^P + \frac{1 - S}{2} \mu^Q \right), \\
P_{p,p-1/(2N)} &= p \left(p\mu^P + 2(1 - p) \frac{1 - S}{2} (1 - \mu^Q) \right), \\
P_{p,p+1/(2N)} &= (1 - p) \left((1 - p)\mu^P + 2p \frac{1 + S}{2} (1 - \mu^P) \right).
\end{aligned}$$

In each birth-death event, there are three individuals involved - the individual who died, the parent and the offspring. We will write (D, A, O) with $D, A, O \in \{P, Q\}$ for the event that type of the deceased individual is D , type of the parent is A ("ancestor") and type of the offspring is O . The individuals of the birth-death event are chosen with replacement, so the parent can also be the individual who died. We will denote such events by (D, O) instead of (D, A, O) . Let p_m^f and p_{m+1}^f be the forwards in time process immediately before and after a birth-death event. Using Bayes rule and reversibility of the stationary distribution of the Moran model, we obtain

$$\begin{aligned}
\mathbb{P}[(D, A, O) | p_{m+1} = \tilde{p}, p_m = p] &= \mathbb{P}[(D, A, O) | p_m^f = \tilde{p}, p_{m+1}^f = p] \\
&= \frac{\mathbb{P}[(D, A, O) \cap [p_{m+1}^f = p] | p_m^f = \tilde{p}]}{P_{\tilde{p},p}} \\
&= \frac{\mathbb{P}[(D, A, O) | p_m^f = \tilde{p}]}{P_{\tilde{p},p}} \mathbf{1}_{(D,A,O),\tilde{p},p},
\end{aligned}$$

where $\mathbf{1}_{(D,A,O),\tilde{p},p}$ is one if the event (D, A, O) results in a forward in time change in the proportion of type P from \tilde{p} to p and zero otherwise. Denote

$$\begin{aligned}
p_+ &= p + \frac{1}{2N}, & p_- &= p - \frac{1}{2N}, \\
q_+ &= q + \frac{1}{2N}, & q_- &= q - \frac{1}{2N}, \\
S_+ &= S(p_+), & S_- &= S(p_-),
\end{aligned}$$

Then the events in which a P -background individual dies have the following probabilities

$$\begin{aligned}
\mathbb{P}[(P, P, P) | p_{m+1} = p, p_m = p] &= \frac{1}{P_{p,p}} pp_-(1 - \mu^P), \\
\mathbb{P}[(P, Q, Q) | p_{m+1} = p_+, p_m = p] &= \frac{1}{P_{p_+,p}} p_+ q_-(1 - S_+)(1 - \mu^Q), \\
\mathbb{P}[(P, Q, P) | p_{m+1} = p, p_m = p] &= \frac{1}{P_{p,p}} pq(1 - S(p))\mu^Q, \\
\mathbb{P}[(P, P, Q) | p_{m+1} = p_+, p_m = p] &= \frac{1}{P_{p_+,p}} p_+ p \mu^P, \\
\mathbb{P}[(P, P) | p_{m+1} = p, p_m = p] &= \frac{1}{P_{p,p}} p \frac{1}{2N} (1 - \mu^P), \\
\mathbb{P}[(P, Q) | p_{m+1} = p_+, p_m = p] &= \frac{1}{P_{p_+,p}} p_+ \frac{1}{2N} \mu^P,
\end{aligned} \tag{II.2.2.1}$$

and similarly we could obtain the probabilities of the events in which it is the Q -background individual who dies. For a full list of these probabilities, see the proof of Lemma 2.4. in [BES04].

Denote by $\bar{n}^P = \sum_{I \in \mathcal{I}} n_I^P$ and $\bar{n}^Q = \sum_{I \in \mathcal{I}} n_I^Q$ the total number of individuals from the sample in background P and Q , respectively. We now consider the probabilities of events on the neutral sample conditioned on the type of the birth-death event and the state of the process $\mathbb{X}^{(1)}$ just before (backwards in time) the occurrence of the birth-death event. There are several types of events which can happen. We only down the probabilities for those events in which the offspring is background P . The probabilities of events where the offspring is in the background Q are obtained simply by exchanging P and Q and p and q in the list below.

- (i) *A coalescence of two individuals from the sample with identical neutral types and backgrounds (without recombination).* Let I be the type of both the parent and the offspring. Then the probability of this event, conditioned on either (Q, P, P) or (P, P, P) occurring, is

$$\frac{\binom{n_I^P}{2}}{\binom{2Np}{2}} r_{I, \emptyset}. \quad (\text{II.2.2.2})$$

Here $2\binom{2Np}{2}$ is the total number of ordered pairs of individuals in background P and $2\binom{n_I^P}{2}$ is the number of such pairs where both individuals have the same neutral type I . This event results in a decrease of n_I^P to $n_I^P - 1$.

- (ii) *A coalescence of two individuals from the sample with identical backgrounds and with neutral types such that one is a subset of the other (without recombination).* Let I be the neutral type of the offspring and J the neutral type of the parent and let $I \subsetneq J$. Then the probability of this event, conditioned on (Q, P, P) or (P, P, P) , is

$$\frac{1}{2} \frac{n_I^P n_J^P}{\binom{2Np}{2}} r_{I, \emptyset}. \quad (\text{II.2.2.3})$$

Here, $n_I^P n_J^P$ is the number of ordered pairs of individuals from background P and from the sample such that the first individual has neutral type I and the second has neutral type J . This again results in a decrease of n_I^P by one.

- (iii) *A coalescence of two individuals from the sample with identical backgrounds and different neutral types with a non-empty symmetric difference (without recombination).* Again, let I be the neutral type of the offspring and J the neutral type of the parent and let $I \Delta J \neq \emptyset$. The probability of such an event, conditioned on (Q, P, P) or (P, P, P) , is again

$$\frac{1}{2} \frac{n_I^P n_J^P}{\binom{2Np}{2}} r_{I, \emptyset}. \quad (\text{II.2.2.4})$$

This event results in a decrease of n_I^P and n_J^P by 1 and an increase of $n_{I \cup J}^P$, also by 1.

- (iv) *A coalescence of two individuals from the same background with recombination and with only the offspring in the sample.* Let I be the neutral type of the offspring individual. Assume that there is a recombination in which the neutral loci with indices in $J \subsetneq I$ move to the second parent (i.e. the individual who died in the birth-death event). Then the probability of this event conditioned on (Q, P, P) or (P, P, P) is

$$\frac{1}{2} \frac{n_I^P (2Np - \bar{n}^P)}{\binom{2Np}{2}} r_{I, J}. \quad (\text{II.2.2.5})$$

Here, $n_I^P(2Np - \bar{n}^P)$ is the number of ordered pairs of individuals from background P such that the first individual comes from the sample and has neutral type I and the second individual does not come from the sample and $r_{I,J}$ is the probability of a recombination described above. This leads to a decrease of n_I^P by 1 and depending on whether the background of the second parent is P or Q , to an increase of n_J^P and $n_{I \setminus J}^P$, both also by 1, or to an increase of n_J^Q and $n_{I \setminus J}^P$ by 1.

- (v) *A coalescence of two individuals from different backgrounds and with only the offspring in the sample (without recombination).* Let I be the neutral type of the offspring. Then conditioned on (Q, Q, P) or (P, Q, P) this event has the probability

$$\frac{n_I^P(2Nq - \bar{n}^Q)}{4N^2pq} r_{I, \emptyset}. \quad (\text{II.2.2.6})$$

Here $n_I^P(2Nq - \bar{n}^Q)$ is the number of ordered pairs such that the first individual is from the background P and the sample with neutral type I and the second individual is from background Q and does not come from the sample and $4N^2pq = 2Np(2N - 2Np)$ is the number of all ordered pairs of individuals such that one is from background P and the other from background Q . This yields a decrease of n_I^P by 1 and an increase of n_I^Q , also by 1.

The following two types of events appear in process $\mathbb{X}^{(1)}$, but vanish later on in the diffusion limit, which is our main point of interest. Therefore, they are described somewhat briefly.

- *A coalescence of two individuals from the sample with different backgrounds and neutral types (without recombination).* Let I be the neutral type of the offspring and J the neutral type of the parent. Conditioned on (Q, Q, P) or (P, Q, P) this event has the probability

$$\frac{n_I^P n_J^Q}{4N^2pq} r_{I, \emptyset}. \quad (\text{II.2.2.7})$$

The effect on the process depends on the exact relationship of I and J similarly like in (II.2.2.3) and (II.2.2.4) above.

- *The parent dies.* Let I be the neutral type of the offspring. Conditioned on (P, P) or (Q, P) , the probability of this event is

$$\frac{n_I^P}{2Np}. \quad (\text{II.2.2.8})$$

This only has an effect on the process if there was a mutation (i.e. in the birth-death event (Q, P)), otherwise nothing changes.

Any other event on the neutral loci either has a zero probability, does not affect the process $\mathbb{X}^{(1)}$ at all, or, as we will see, also vanishes in the diffusion limit.

The birth-death events that do not affect the individuals from the sample give rise to the terms

$$T_+ = (NP_{p,p+1/(2N)} - R_+)(f(p_+, \mathfrak{n}^P, \mathfrak{n}^Q) - f(p, \mathfrak{n}^P, \mathfrak{n}^Q)) \quad (\text{II.2.2.9})$$

$$T_- = (NP_{p,p-1/(2N)} - R_-)(f(p_-, \mathfrak{n}^P, \mathfrak{n}^Q) - f(p, \mathfrak{n}^P, \mathfrak{n}^Q)), \quad (\text{II.2.2.10})$$

in the generator of $\mathbb{X}^{(1)}$, where R_+ (resp. R_-) is the sum of the rates of all events which affect the sample and which result in an increase (resp. a decrease) of the allele ratio p and the

factor N is the rate at which birth-death events appear in the process. From the law of total probability we get that all other generator terms have the form

$$\begin{aligned} N \Big(& cP_{p^*,p} \mathbb{P}[(D, A, O) | \mathbb{X}_{t-}^{(1)} = (p, \mathfrak{m}^P, \mathfrak{m}^Q), \mathbb{X}_t^{(1)} = (p^*, (\mathfrak{m}^P)^*, (\mathfrak{m}^Q)^*)] \\ & \times \mathbb{P}[\mathcal{N} | (D, A, O), \mathbb{X}_{t-}^{(1)} = (p, \mathfrak{m}^P, \mathfrak{m}^Q), \mathbb{X}_t^{(1)} = (p^*, (\mathfrak{m}^P)^*, (\mathfrak{m}^Q)^*)] \Big) \\ & \times (f(p^*, (\mathfrak{m}^P)^*, (\mathfrak{m}^Q)^*) - f(p, \mathfrak{m}^P, \mathfrak{m}^Q)), \end{aligned} \quad (\text{II.2.2.11})$$

where $c = \frac{P_{p,p^*}}{P_{p^*,p}}$, \mathcal{N} denotes an event affecting the neutral sample and $\mathbb{X}_{t-}^{(1)}$ and $\mathbb{X}_t^{(1)}$ are the values of process $\mathbb{X}^{(1)}$ (backwards in time) before and after an arbitrary birth-death event.

II.2.3 Convergence to the diffusion limit

We now define the process $\mathbb{X}^{(N)}$ as the process $\mathbb{X}^{(1)}$ rescaled by the factor of N , that is the process $\mathbb{X}^{(1)}$ with mutation rates $(\mu^P)^{(N)} := \frac{\mu^P}{N}$ and $(\mu^Q)^{(N)} := \frac{\mu^Q}{N}$, recombination rates $(r_I^{(N)}, I \in \mathcal{I})$ where $r_I^{(N)} := \frac{r_I}{N}$, $I \in \mathcal{I}$ and time fast-forwarded by the factor of N . In the process $\mathbb{X}^{(N)}$ we will define the selection coefficient through a Lipschitz continuous function $s : [0, 1] \mapsto \mathbb{R}$ in the following way. We let $s^{(N)} : [0, 1] \rightarrow [-1, \infty)$ be defined as

$$s^{(N)}(p) := \begin{cases} \frac{s(p)}{N} & \text{if } s(p) \in [-N, \infty) \\ -1, & \text{otherwise,} \end{cases}$$

Then we define the rescaled selection coefficient for $p \in [0, 1]$ as

$$S^{(N)}(p) := \frac{s^{(N)}(p)}{2 + s^{(N)}(p)} = \frac{s(p)}{2N} + o(1/N).$$

In Theorem II.2.3.1 we will need the following notation. For $I \in \mathcal{I}$ let $\mathbb{1}_I := \{\mathbb{1}_I(J)\}_{J \in \mathcal{I}}$, where

$$\mathbb{1}_I(J) := \begin{cases} 1 & \text{if } I = J \\ 0 & \text{otherwise} \end{cases}$$

and let $\mathbb{1}_\emptyset(J) := 0$ for all $J \in \mathcal{I}$. Recall that by $\mathfrak{n}^P(t) = \{n_I^P(t)\}_{I \in \mathcal{I}}$ and $\mathfrak{n}^Q(t) = \{n_I^Q(t)\}_{I \in \mathcal{I}}$ we denote the collection of numbers of individuals from the sample corresponding to each neutral type in background P and Q at time $t \geq 0$, respectively. For $t \geq 0$ denote by $\bar{\mathfrak{m}}^P(t) = \sum_{I \in \mathcal{I}} n_I^P(t)$ and $\bar{\mathfrak{m}}^Q(t) = \sum_{I \in \mathcal{I}} n_I^Q(t)$ the total number of individuals from the sample at time t in background P and Q , respectively. Note that while the sample might split into more individuals and so $n_I^P(0) + n_I^Q(0)$ might be smaller than $n_I^P(t)$ for some $t > 0$ and $I \in \mathcal{I}$, each neutral locus is present at most $n_0 := \bar{\mathfrak{m}}^P(0) + \bar{\mathfrak{m}}^Q(0)$ times, since n_0 is the same as the number of individuals sampled at time 0 and so each individual neutral locus is present at most n_0 times. Therefore n_0 is the upper bound for $n_I^P(t)$ and $n_I^Q(t)$ for all $I \in \mathcal{I}$ and $t \geq 0$.

Theorem II.2.3.1 *Let*

$$E := [0, 1] \times \prod_{I \in \mathcal{I}} \{0, \dots, n_0\} \times \prod_{I \in \mathcal{I}} \{0, \dots, n_0\}$$

and let $\mathcal{C}_1^2(E)$ be the space of all functions $f : E \mapsto \mathbb{R}$ which are twice continuously differentiable with respect to p . Then there exists a process \mathbb{X} such that the processes $\mathbb{X}^{(N)}$ converge weakly

in $D_E[0, \infty)$ as $N \rightarrow \infty$ to \mathbb{X} and \mathbb{X} is generated by A , where $A \in \mathcal{L}(\mathcal{C}_1^2(E))$ is defined as follows.

$$\begin{aligned} Af(p, \mathfrak{n}^P, \mathfrak{n}^Q) &= A^P f(p, \mathfrak{n}^P, \mathfrak{n}^Q) + A^Q f(p, \mathfrak{n}^P, \mathfrak{n}^Q) \\ &+ (-\mu^P p + \mu^Q q + s(p)pq) \frac{1}{2} \frac{\partial}{\partial p} f(p, \mathfrak{n}^P, \mathfrak{n}^Q) + pq \frac{1}{4} \frac{\partial^2}{\partial p^2} f(p, \mathfrak{n}^P, \mathfrak{n}^Q) \end{aligned} \quad (\text{II.2.3.1})$$

where

$$A^P f(p, \mathfrak{n}^P, \mathfrak{n}^Q) = \frac{1}{2p} \sum_{I \in \mathcal{I}} \left[\binom{n_I^P}{2} (f(p, \mathfrak{n}^P - \mathbb{1}_I, \mathfrak{n}^Q) - f(p, \mathfrak{n}^P, \mathfrak{n}^Q)) \right] \quad (\text{II.2.3.2})$$

$$+ \frac{1}{2p} \sum_{I \in \mathcal{I}} \sum_{\substack{J \in \mathcal{I} \\ J \supseteq I}} \left[\frac{n_I^P n_J^P}{2} (f(p, \mathfrak{n}^P - \mathbb{1}_I, \mathfrak{n}^Q) - f(p, \mathfrak{n}^P, \mathfrak{n}^Q)) \right] \quad (\text{II.2.3.3})$$

$$+ \frac{1}{2p} \sum_{I \in \mathcal{I}} \sum_{\substack{J \in \mathcal{I} \\ I \Delta J \neq \emptyset}} \left[\frac{n_I^P n_J^P}{2} (f(p, \mathfrak{n}^P + \mathbb{1}_{I \cup J} - \mathbb{1}_I - \mathbb{1}_J, \mathfrak{n}^Q) - f(p, \mathfrak{n}^P, \mathfrak{n}^Q)) \right] \quad (\text{II.2.3.4})$$

$$+ \frac{1}{2} \frac{q}{p} \mu^Q \sum_{I \in \mathcal{I}} \left[n_I^P (f(p, \mathfrak{n}^P - \mathbb{1}_I, \mathfrak{n}^Q + \mathbb{1}_I) - f(p, \mathfrak{n}^P, \mathfrak{n}^Q)) \right] \quad (\text{II.2.3.5})$$

$$+ \frac{p}{2} \sum_{I \in \mathcal{I}} \sum_{\substack{\hat{I} \in \mathcal{I} \\ \hat{I} \subsetneq I}} \left[r_{I, \hat{I}} n_I^P (f(p, \mathfrak{n}^P + \mathbb{1}_{I \setminus \hat{I}} + \mathbb{1}_{\hat{I}} - \mathbb{1}_I, \mathfrak{n}^Q) - f(p, \mathfrak{n}^P, \mathfrak{n}^Q)) \right] \quad (\text{II.2.3.6})$$

$$+ \frac{q}{2} \sum_{I \in \mathcal{I}} \sum_{\substack{\hat{I} \in \mathcal{I} \\ \hat{I} \subsetneq I}} \left[r_{I, \hat{I}} n_I^P (f(p, \mathfrak{n}^P + \mathbb{1}_{I \setminus \hat{I}} - \mathbb{1}_I, \mathfrak{n}^Q + \mathbb{1}_{\hat{I}}) - f(p, \mathfrak{n}^P, \mathfrak{n}^Q)) \right] \quad (\text{II.2.3.7})$$

and $A^Q f(p, \mathfrak{n}^P, \mathfrak{n}^Q)$ is $A^P f(p, \mathfrak{n}^P, \mathfrak{n}^Q)$ with P and Q and p and q reversed.

The proof of Theorem II.2.3.1 can be found in section II.2.5.

Remark II.2.3.2 The terms of the generator correspond to the possible events that affect the process \mathbb{X} . The first three terms describe coalescence. In particular, (II.2.3.2) corresponds to a coalescence of two individuals of the same type and the same combination of neutral loci from the sample, (II.2.3.3) corresponds to a coalescence of two individuals of the same type of which one contains all the sampled neutral loci carried by the other and also some which the other individual does not carry and (II.2.3.4) describes coalescences where the two individuals each carry neutral loci that the other does not. The term (II.2.3.5) corresponds to an individual moving to a different background due to mutation. The last two terms describe recombination events in which the sampled loci in the individual split (backwards in time) and one part of them moves to the second parent who is either in the same background (see (II.2.3.6)) or the other background (see (II.2.3.7)) than the offspring. Note that in and (II.2.3.7) we also include the scenario in which all sampled loci move to the other background, whereas such an event has no effect on the process if the second parent comes from the same background, so it is not included in (II.2.3.6).

II.2.4 Applications

Denote $[K] := \{1, \dots, K\}$. We will be interested in estimating some characteristics of the process \mathbb{X} , namely the distribution functions of T^l , $l \in [K]$, which we define as the first time at least l of the neutral loci from the sample coalesce to a common ancestor, the correlation between the coalescence times of different neutral loci and the expected number of pairwise differences, i.e. the expected number of different mutations of neutral loci between two sampled individuals. More precisely, we put $n_I(t) := n_I^P(t) + n_I^Q(t)$, $I \in \mathcal{I}$, $t \geq 0$ and we define T^1 as

$$T_1 := \inf\{t \geq 0; \exists i \in [K] \text{ s.t. } \sum_{\substack{J \in \mathcal{I} \\ i \in J}} n_J(t) = 1\}.$$

In other words, T_1 is the first time at which the ancestry of the sample of at least one neutral locus is concentrated in one individual (who can possibly also carry other ancestors from the sample). The distribution function of T^1 , conditioned on $\mathbb{X}_0 = (p_0, \mathfrak{n}^P(0), \mathfrak{n}^Q(0))$, can be then written as follows.

$$F_1^{\mathbb{X}_0}(t) = \mathbb{P}^{\mathbb{X}_0}[\bigcup_{i \in [K]} \{\sum_{\substack{J \in \mathcal{I} \\ i \in J}} n_J(t) = 1\}]. \quad (\text{II.2.4.1})$$

Generally, for $l \in [K]$ we put

$$T_l := \inf\{t \geq 0; \exists A \subseteq [K], |A| \geq l, \text{ s.t. } \sum_{\substack{J \in \mathcal{I} \\ i \in J}} n_J(t) = 1 \forall i \in A\}$$

and the corresponding distribution function is then

$$F_l^{\mathbb{X}_0}(t) = \mathbb{P}^{\mathbb{X}_0}[\bigcup_{\substack{A \subseteq [K] \\ |A|=l}} \bigcap_{i \in A} \{\sum_{\substack{J \in \mathcal{I} \\ i \in J}} n_J(t) = 1\}]. \quad (\text{II.2.4.2})$$

We will also be interested in the correlation of the times of coalescence of different neutral loci. Let us denote by \tilde{T}_i the first time the ancestral sample at the l -th neutral locus coalesces, i.e.

$$\tilde{T}_i := \inf\{t \geq 0; \sum_{\substack{J \in \mathcal{I} \\ i \in J}} n_J(t) = 1\}. \quad (\text{II.2.4.3})$$

For each $i, j \in [K]$ let \tilde{F}_i denote the distribution function of \tilde{T}_i and by $\tilde{F}_{i,j}$ the joint distribution function of \tilde{T}_i and \tilde{T}_j . Then we have

$$\tilde{F}_{i,j}^{\mathbb{X}_0}(s, t) = \mathbb{P}^{\mathbb{X}_0}[\left\{\sum_{\substack{J \in \mathcal{I} \\ i \in J}} n_J(t) = 1\right\} \cap \left\{\sum_{\substack{J \in \mathcal{I} \\ j \in J}} n_J(t) = 1\right\}]. \quad (\text{II.2.4.4})$$

From this we could calculate the correlation between the two times of coalescence. Assume now that the neutral loci are also affected by mutation. Since neutral loci do not affect the reproductive chances of an individual we can assume that these mutations occur at each locus independently of each other and independently of the process of allele ratios. Denote by D the number of pairwise identities between two sampled individuals, that is the number neutral loci which have the same allele in both individuals (so then $K - D$ is the number of pairwise differences). We see that $\mathbb{E}^{\mathbb{X}_0}[D] = \sum_{i=1}^K \mathbb{P}^{\mathbb{X}_0}[\text{The neutral alleles at locus } i \text{ are the same}]$. Letting

ν_j be the mutation rate to a novel allele at locus j and rescaling these rates accordingly as in the Section 3, we get that

$$\mathbb{E}^{\times_0}[D] = \sum_{i=1}^K \int_0^\infty e^{-2\nu_i t} \frac{\partial \tilde{F}_i^{\times_0}(t)}{\partial t} dt. \quad (\text{II.2.4.5})$$

The summands can be then calculated exactly as in Section 6 of [BES04]. We can also obtain the distribution of D as follows. Let $\text{Var}(\binom{[K]}{k})$ denote the set of subsets of $[K]$ with k elements. Then the distribution of D is given by

$$\mathbb{P}[D = k] = \sum_{V \in \text{Var}(\binom{[K]}{k})} \int_{\mathbb{R}_+^K} \left(\prod_{i \in V} e^{-2\nu_i t_i} \prod_{j \notin V} (1 - e^{-2\nu_j t_j}) \right) \frac{\partial^K \tilde{F}^{\times_0}(t_1, \dots, t_K)}{\partial t_1 \cdots \partial t_K} dt_1 \cdots dt_K, \quad k \in [K] \quad (\text{II.2.4.6})$$

where \tilde{F}^{\times_0} is the distribution function of $\tilde{T} := (\tilde{T}_1, \dots, \tilde{T}_K)$.

In estimating (II.2.4.6) as well as other characteristics mentioned in this section, we will limit ourselves to simulation results. The rationale for this is that in order to obtain the distribution function of the time T_1 of the first coalescence, one needs to solve 18 coupled differential equations even in the simplest interesting case where $K = 2$, we start with a sample of two individuals with both neutral loci from the sample.

Remark II.2.4.1 To obtain the number of equations for $N \in \mathbb{N}$ sampled individuals with $K \in \mathbb{N}$ neutral loci, we need to introduce a few concepts. We call *multiset* a generalization of a set in which elements may appear more than once. In particular, every set is a multiset. The *multiplicity* $\chi_A(x)$ of an element x of a multiset A is the number of times that element appears in the multiset A . A *partition* \mathcal{A} of a set A is a system of disjoint subsets $A' \in \mathcal{A}$ such that their union forms the set A . By $\text{Part}(A)$ we will denote the set of all partitions of A . Finally we let \mathbb{M}_K^N be the system of all multisets \mathcal{M} such that \mathcal{M} is a multiunion of N partitions of $[K]$, where we define the multiunion $A \uplus B$ of sets A and B as a multiset which consists of all elements in A and B and such that $\chi_{A \uplus B}(x) = 2$ for all $x \in A \cap B$. The system of multisets \mathbb{M}_K^N contains all possible ways the NK sampled neutral loci can be distributed between individuals if we do not distinguish between individuals with the same set of active loci. To obtain the number of equations we would have to solve, we still need to take into account the backgrounds of these individuals. Each individual can be either in background P or Q and there are $n + 1$ unique combinations of backgrounds among n individuals who are only differentiated by their background. The number of coupled differential equations that need to be solved corresponds to all possible ways that the sample of NK neutral loci can be distributed among individuals of different backgrounds so that each copy of a single neutral locus is present in at least two individuals (i.e. that neutral locus has not yet entirely coalesced to a common ancestor). Therefore we obtain, that the number of equations we need to solve to calculate the distribution of T_1 when we sample N individuals with K neutral loci is

$$C_K^N \geq \sum_{\mathcal{M} \in \mathbb{M}_K^N} \prod_{A \in \mathcal{M}} (\chi_{\mathcal{M}}(A) + 1). \quad (\text{II.2.4.7})$$

Here the right-hand side corresponds to all possible distributions of the sample in which none of the N copies of the K loci are still distributed between N different individuals. In particular, if $N = 2$ then we have an equality in (II.2.4.7). The sets $A \in \mathcal{M}$ each correspond to a neutral type of of an individual in the sample when the NK neutral loci are distributed in the way

given by \mathcal{M} . The number of elements of \mathbb{M}_K^N is $\binom{N+B_K-1}{N}$, where B_K (also referred to as the Bell number) is the number of distinct partitions of $[K]$. We can get a (rather low) lower bound for C_K^N by noting that $\chi_{\mathcal{M}}(A) + 1 \geq 2$ and so $C_K^N \geq 2 \binom{N+B_K-1}{N}$. For the Bell number it holds that $B_K = \frac{1}{e} \sum_{n=0}^{\infty} \frac{n^K}{n!}$ (see for example [Rot64]). We note that already for $K = 5$, $B_5 = 52$ which means that $C_5^2 \geq 2756$ and for $K = 10$, we get $B_{10} = 115975$ and $C_{10}^2 \geq 13450316600$.

So far we have not made any assumptions about the way we define our recombination probabilities or how the neutral loci are aligned around the selective locus (their geometry). In what follows, we will commit ourselves to the following choice of recombination probabilities and the geometry. In the discrete-time processes $\mathbb{X}^{(N)}$, we will define the recombination probabilities and the alignment of the loci as in the Example II.2.1.1, only with $c_i, i \in [K]$ (the probabilities of cuts) rescaled to $c_i \mapsto c_i/N$. Then all recombination events which consist of more than one simultaneous cut are of order $O(1/N^2)$ and so do not appear in the limit. Thus in the process \mathbb{X} we only observe "single" recombinations, that is those which separate the chromosome into two continuous parts so that the part which contains the selective locus moves to the first parent and the part which does not moves to the second parent. A recombination event which separates the chromosome between the $(i-1)$ -st locus L_{i-1} and i -th locus L_i (where we again denote by L_0 the selective locus as in the Example II.2.1.1) happens at rate c_i . We could also assume a more general geometry where the selective locus is located between two neutral loci. However, qualitatively, the only result which differs from the results in the one-sided alignment is in the correlations of the times \tilde{T}_k , so we will only discuss the two-sided geometry there.

The following discussion is based on simulations of the process \mathbb{X} . As stated before, we limit ourselves to the case where the process starts with two sampled individuals with all K loci active. The selection s is always assumed to be balancing, more precisely it has the form $s(x) := s_0(p_b - x)$. This is motivated by the fact that the balancing selection is a common type of frequency-dependent selection in nature, the standard example of which is the sickle-cell allele in humans which helps protect against malaria but which causes sickle cell anemia when its copy is present on both chromosomes. In all simulations we choose $s_0 = 0.16$ as the selection strength and $p_b = \frac{1}{2}$ as the balancing allele ratio. We also start the process p in $p_0 = \frac{1}{2}$ and choose the value 0.3 for the mutation rates μ^Q and μ^P . We simulate the process via a discrete approximation on a finite time interval divided into 10000 steps and using the rates given in Theorem II.2.3.1. We are above all interested in the way the recombination rates influence the distribution of $T_l, l \in [K]$ and the correlations of $\tilde{T}_i, i \in [K]$. Unless stated otherwise we assume that all recombination events happen at the same rate r , i.e. we set $c_i := r$ for all $i \in [K]$.

In Figures II.2.1a and II.2.1b we set $K = 7$ and observe the plotted the expected times $\mathbb{E}[T_k], k = 1, \dots, 7$ as a function of the recombination rate. The way recombination affects T_k is very different when we start with both individuals in the same background as in Figure II.2.1b (which we denote in the captions as a "PP start") and when we start with each individual in a different background as in Figure II.2.1a (which the captions refer to as a "PQ start"). Heuristically, the reason this happens can be explained using a much simpler model of two coalescing individuals which approximates some of the dynamics of our process.

Let $Y = (Y_t)_{t \geq 0}$ be a Markov chain on $S := \{0, 1, 2\}$. Here, the state 0 corresponds to the two individuals being each in a different background and therefore unable to coalesce. State 1 corresponds to them being in the same background from which coalescence is possible and finally, state 2 is an absorbing state and the presence of the process in it indicates that the two individuals have coalesced. Y moves from 0 to 1 and from 1 to 0 at a constant rate $m > 0$

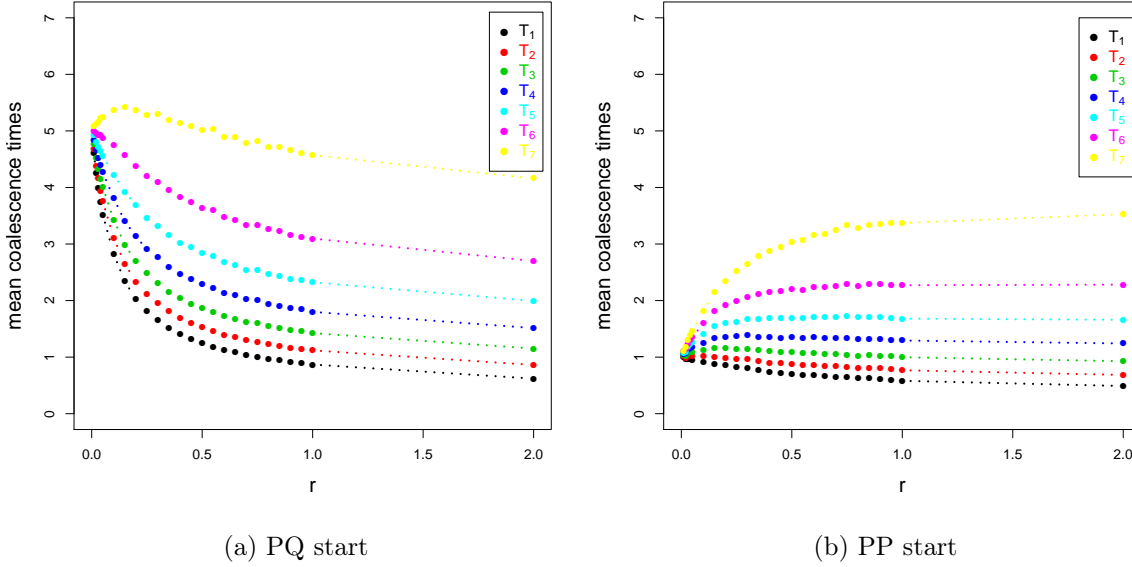


Figure II.2.1: Expected times T_l of coalescence of the first $l = 1, \dots, 7$ (out of 7 in total) neutral loci plotted as a function of the recombination rate.

and it moves from the state 1 to the absorbing state 2 at rate $c > 0$. The rate m of the process Y corresponds to the rate at which the sampled loci change their background which in the original process \mathbb{X} can be caused either by mutation or by recombination. We also assume that either $Y_0 = 0$ almost surely or $Y_0 = 1$ almost surely. The distribution of the time T of absorption of the process Y in 2 (Distributions of similar absorption times are sometimes referred to as phase-type distributions) can be easily obtained by calculating the transition probability matrix of the Markov chain Y . Depending on whether the process Y starts in 0 or 1 we get the following distribution functions of T .

$$F_0^T(t) := \frac{w^-(2 \exp(e^+t) - \exp(e^-t)) - w^+ \exp(e^-t)}{w^- - w^+} \quad (\text{II.2.4.8})$$

if $Y_0 = 0$ almost surely and

$$F_1^T(t) := \frac{w^+(-2 \exp(e^+t) + \exp(e^-t)) + w^- \exp(e^-t)}{w^- - w^+} \quad (\text{II.2.4.9})$$

if $Y_0 = 1$ almost surely, where $w^\pm := -\frac{c \pm \sqrt{4\mu^2 + c^2}}{2\mu}$, and $e^\pm := \frac{-2\mu - c \pm \sqrt{4\mu^2 + c^2}}{2}$.

Figure II.2.2 shows the mean time of absorption $\mathbb{E}[T]$ as a function of m . We can see that as the rate m increases, the expected absorption times converge to the same value when $Y_0 = 0$ and when $Y_0 = 1$. However, the effects of the faster transitions between states 0 and 1 on the absorption time is exactly opposite. This gives us a heuristic explanation for the difference between Figures II.2.1b and II.2.1a. Increasing the recombination rate if we start with two individuals in different backgrounds causes them to come to the same background faster and therefore speeds up the coalescence. On the other hand, if they already start in the same background, increasing the recombination rate increases the average time spent in different backgrounds and so the coalescence is slower. If the migration rate is high enough, then the individuals spend on average the same amount of time in each background regardless of the starting configuration.

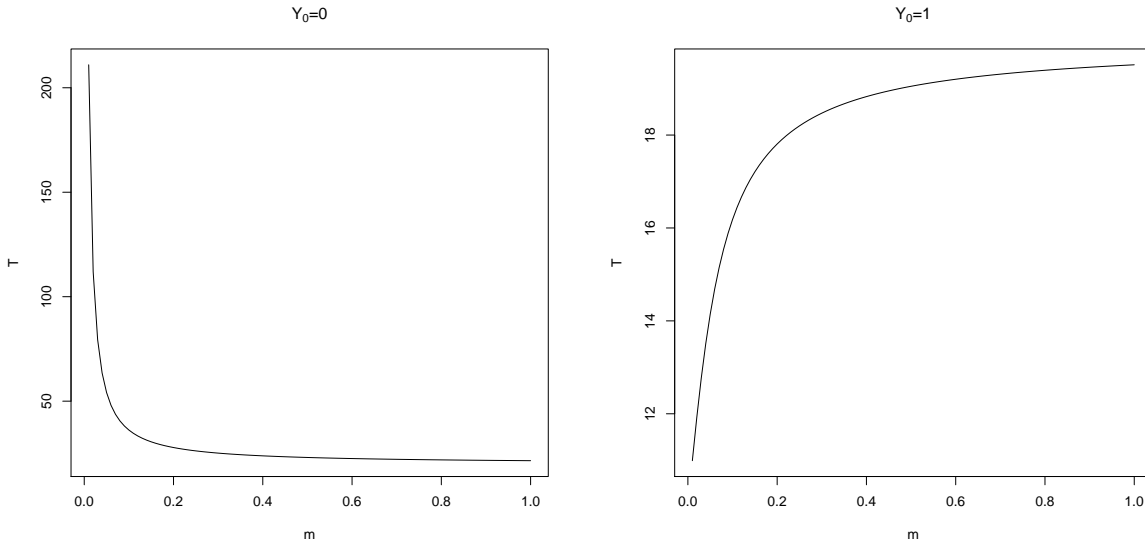


Figure II.2.2: The expected absorption time T of process Y

Recombination in the process \times (at least as long as $K > 1$) not only causes the sampled individuals to switch backgrounds but can also split the neutral sample and move a part of it to the first parent and another part to the second parent. If in a recombination event a part of the active loci of an individual move to the first parent, we will call that a *split*. We will call a *migration* the event when some or all active loci of an individual move to a different background. This can either be caused by mutation or by a recombination with a second parent which is in a different background than the offspring individual. Recombination can either have the same effect on the neutral loci as a mutation if there is no split or it can cause a migration of only a subset of the active neutral loci of the individual (the one which moves to the second parent).

The rates at which migration and splits occur both increase with the increasing recombination rate. However, these two events affect the coalescence times T_k differently and the exact effect also depends on k and K . We can see from the generator A in the Theorem II.2.3.1 that the coalescence rate is not influenced by the configuration of the neutral loci, i.e. two individuals will coalesce at the same rate regardless of which neutral loci are active in each of them. Because of that, it is clear that T_K will increase and T_1 will decrease as the rate at which splits occur increases, since the more splits happen, the faster some of the neutral loci will coalesce but the longer it will take for all of them to do so. For k between 1 and K the situation is less clear. Intuitively, the increase in the rate of splits should cause T_k to decrease for k close to 1 and increase for k close to K , while the effect on T_k should grow weaker as $\min\{|K - k|, |1 - k|\}$ increases. It is possible to obtain (at least conditioned on the process p of allele ratios) the upper and lower bounds of the times T_1 and T_K by assuming the process \times starts and stays with all K neutral loci either carried by $2K$ individuals (one pair of individuals for each locus) or that it starts as usually with two individuals with all their loci active and stays that way until coalescence. In that case we can easily use the framework given in [BES04] in Section 6. The reason we do not do it here is that the bounds obtained this way will be very crude, since in both cases we ignore the splitting effect of recombination (which is essentially the only thing which makes the model different for $K > 1$). We cannot obtain bounds on T_k for k other than K or 1 in the same way, since there is no single distribution

of active neutral loci between individuals which would at all times lead to the slowest or the fastest coalescence rates.

We would like to separate the migration and splitting effects of recombination to see how changing the rate of one of those affect the process when the rate of the other does not change. Since both are affected by the migration rate, we cannot separate the effects directly. We can however at least observe a rather artificially constructed process where the migration caused by recombination does not happen (even though we still observe migrations between backgrounds due to mutation).

We modify the process \mathbb{X} in such a way, that the rate of all recombinations which cause a migration is set to zero. The generator of such a process, which we will call $\hat{\mathbb{X}}$ will look as the generator A of the Theorem II.2.3.1 without the line (II.2.3.7). Since splits and migration can happen simultaneously, this means that we also reduce the rate of splits. However, the effect of the increasing recombination rate in the process $\hat{\mathbb{X}}$ is strictly that of an increased rate of splits so by observing how the expected values of T_k change as we change the recombination rate, we can approximately see how the increased rate of splits affects the process \mathbb{X} .

The results from simulating the process $\hat{\mathbb{X}}$ are shown in Figures II.2.3a and II.2.3b. We observe that unlike in the case where the recombination also causes migration, in the process $\hat{\mathbb{X}}$ increasing the recombination rate affects the process in the same way regardless of whether we start the process with two individuals in the same background or in different backgrounds. Overall the effect on the times T_K of increasing the recombination rate is a mixture of an effect corresponding to that which we saw in the simpler model Y and Figure II.2.2 and that observed in the process \mathbb{X} and Figures II.2.3a and II.2.3b. As we already noted, the splitting effect of recombination leads to a larger T_K and so we observe in Figure II.2.1a that the mean of T_7 first increases with the increasing recombination rate and then decreases because of the increasing rate of migration. In the case when we start in the same background, both the splits and the migration make T_K larger, so we do not observe such a change in the Figure II.2.1b.

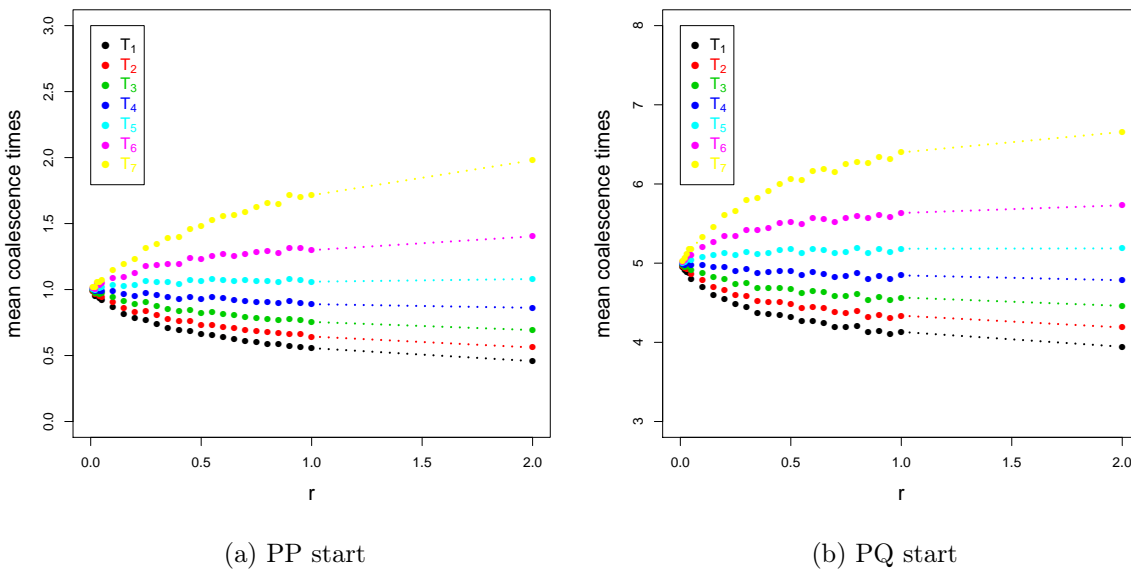


Figure II.2.3: Expected times T_l of coalescence of the first $l = 1, \dots, 7$ (out of 7 in total) neutral loci in the process $\hat{\mathbb{X}}$ plotted as a function of the recombination rate

Another effect of recombination is that the correlation between the coalescence times \tilde{T}_k and \tilde{T}_{k+l} increases as a function of k , i.e. with the distance from the selective locus. This is an effect of the particular choice of recombination rates of the one-sided alignment given in Example II.2.1.1. The further away a locus is from the selective locus, the more likely it is to shift backgrounds due to recombination since there are more possible recombination events that cause its migration to the other background. Hence if we start with both individuals in the same background, the further the neutral loci are from the selective locus, the slower they will coalesce and the opposite is true for an initial configuration in which each individual is in a different background. Again, looking at the distribution of the absorption time of the simple model Y , we can explain the observed increase of correlation by noting that the distribution of T becomes flatter as the rate m increases.

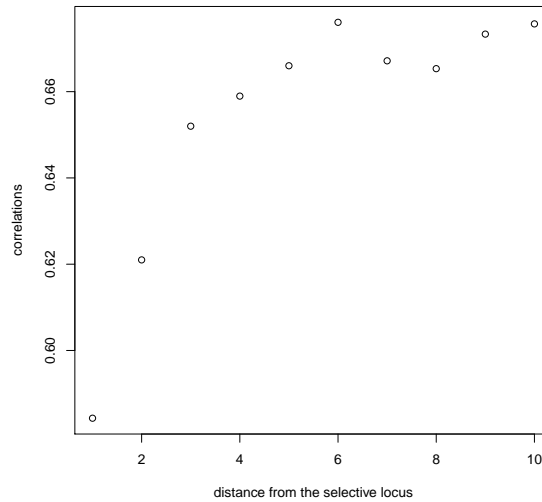


Figure II.2.4: The effect of the distance from the selective locus with $r = 0.5$ (PP start)

In Figure II.2.4 we can see the empirical correlation between two neutral loci in a simulation with $K = 2$ and a recombination rates r and r_0 where r is the rate of recombinations which split the chromosome between the first and the second neutral locus and $r_0 := cr, c \in \mathbb{N}$ is the rate of recombinations which split it between the selective locus and the first neutral locus, causing both neutral loci to move to the second parent. Here $c \in \mathbb{N}$ should be thought of as the distance from the selective locus and the plot shows the correlation of \tilde{T}_1 and \tilde{T}_2 as a function of c . Note that the correlations in Figure II.2.4 appear as close to but not quite monotone, this is most likely an artefact of the simulation. Comparing the Figure II.2.4 to the Figure II.2.5a of correlations between neighbouring loci and also to the Figure II.2.5b of the same correlations in the process \hat{X} , we see that the increase of the correlation of \tilde{T}_i and \tilde{T}_{i+j} as a function of i is caused entirely by the distance from the selective locus.

In Figure II.2.6 we see the correlations between 6 neutral loci in the model where the selective locus is located between L_3 and L_4 and it holds for all $i = 1, \dots, 5$ that L_i is located left of L_{i+1} . The recombination is defined as before - only those recombination which split the chromosome in two parts are allowed and they happen between any two neighbouring loci (including the selective locus) at a constant rate $r = 0.1$. The part which includes the selective locus again moves to the first parent and the part which does not moves to the second parent. Again, we observe that the correlations of coalescence times of the individual neutral

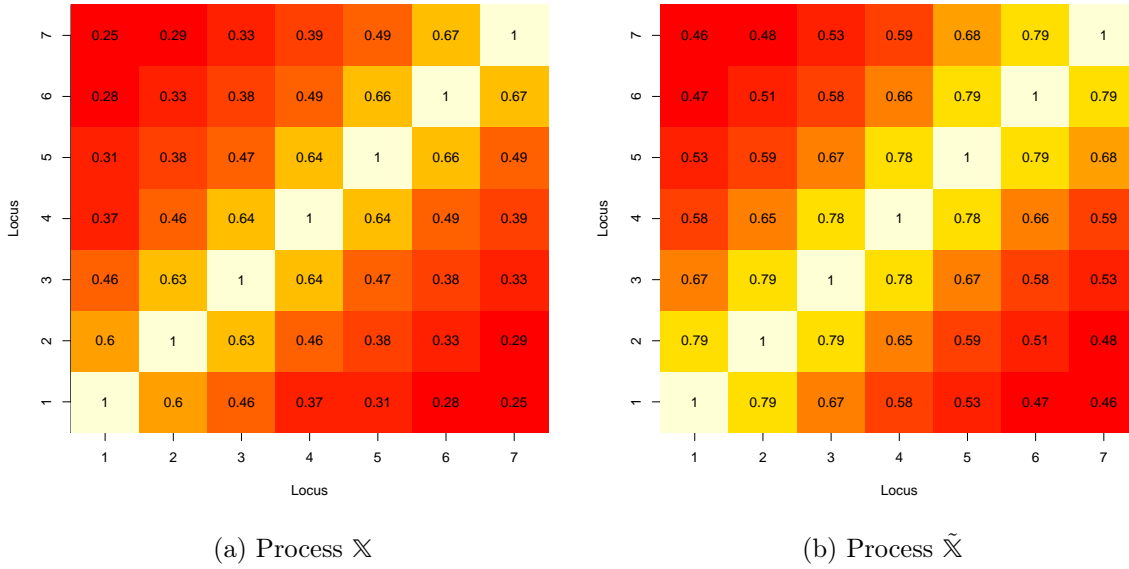


Figure II.2.5: Correlations of the \tilde{T}_i and \tilde{T}_j , $i, j = 1, \dots, 7$ with $r = 0.5$ (PP start)

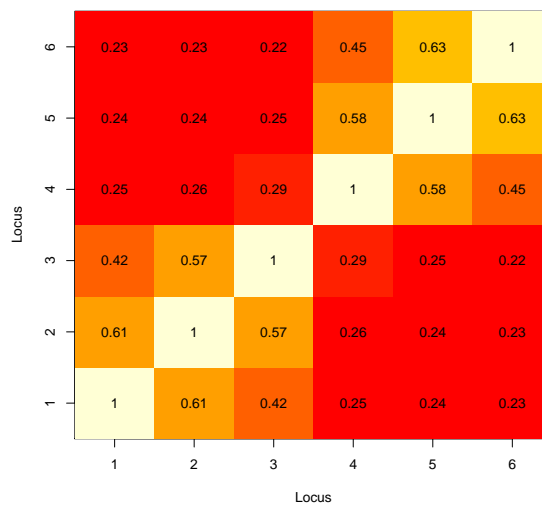
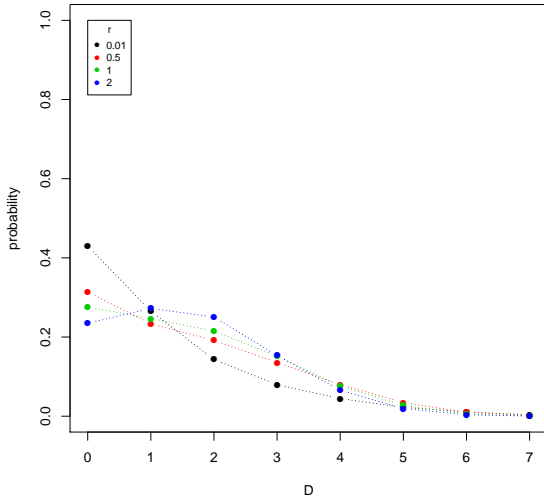


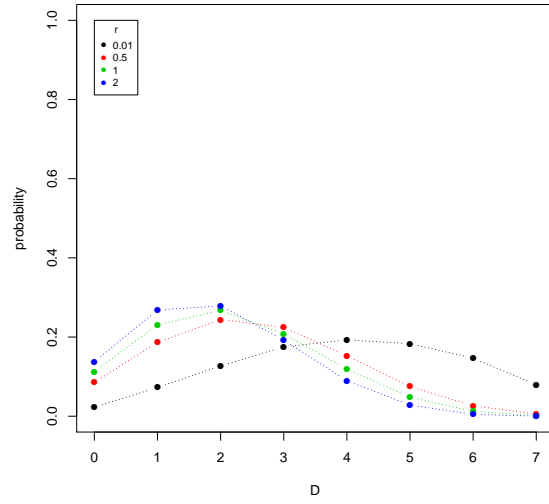
Figure II.2.6: Correlations of \tilde{T}_i and \tilde{T}_j , $i, j = 1, \dots, 6$ in the model where the selective locus is located between L_3 and L_4 with $r = 0.5$ (PP start)

loci increase with their distance from the selective locus. Between pairs of loci in which both are on the same side of the selective locus, we see the same pattern as in the model with one-sided alignment. The correlation of those \tilde{T}_i and \tilde{T}_j , where L_i and L_j are each on a different side of the selective locus are much weaker. This is partly because (assuming that $i < j$) in addition to the neutral loci L_{i+1}, \dots, L_{j-1} , there is the selective locus between L_i and L_j and therefore they are further away from each other on the chromosome. It is also because, unlike for a pair of loci on one side of the selective locus, there are no recombinations which could simultaneously move both loci to the second parent.

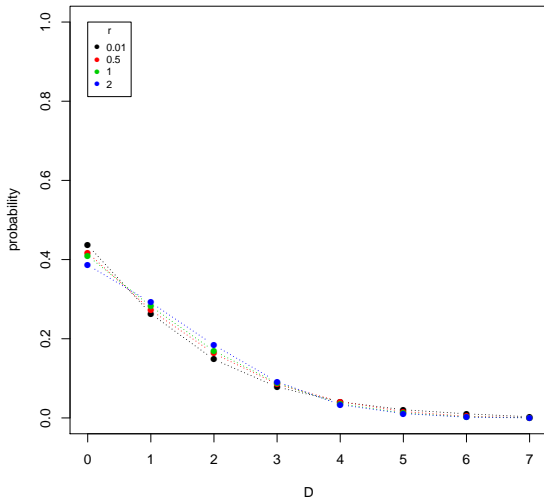
Finally, we take a look at the distribution of the number D of pairwise identities between individuals which we defined above the formula (II.2.4.5). Comparing the Figures II.2.7a and II.2.7b and II.2.7c, the effect of increasing the recombination rate appears to be almost entirely driven by the increased migration and increasing it results in a decrease or increase in the individual coalescence times, depending on whether the process starts with two individuals in the same background or in different ones.



(a) PP start



(b) PQ start



(c) Process \hat{X} (PP start)

Figure II.2.7: The empirical distribution of D

II.2.5 Proof of Theorem II.2.3.1

II.2.5.1 Existence of the limiting process

We follow (and at one point clarify) the proof of existence from [BES04] where it is done for the case when $K = 1$. Since some of the rates in the generator A from Theorem II.2.3.1 tend to infinity as the gene frequency p tends either to 0 or 1, it is not immediately clear that there exists a process with generator A since such a process could possibly jump an infinite number of times in a finite time interval. To prove that this does not happen, we will first define "good" regions in the state space E where this cannot occur. Recall that we defined $n_0 = \bar{n}^P(0) + \bar{n}^Q(0)$. We define the "good regions" as

$$\begin{aligned} U^{(k)} &:= \left[0, \frac{1}{k}\right] \times \{\mathbb{0}\} \times \prod_{I \in \mathcal{I}} \{0, 1, \dots, n_0\} \\ &\cup \left[1 - \frac{1}{k}, 1\right] \times \prod_{I \in \mathcal{I}} \{0, 1, \dots, n_0\} \times \{\mathbb{0}\} \\ &\cup \left[\frac{1}{k}, 1 - \frac{1}{k}\right] \times \prod_{I \in \mathcal{I}} \{0, 1, \dots, n_0\} \times \prod_{I \in \mathcal{I}} \{0, 1, \dots, n_0\}, \end{aligned}$$

where $\mathbb{0} := \prod_{I \in \mathcal{I}} \{0\}$. First, we will show that a process with generator A exists at least until its first exit time from $U^{(k)}$ for each $k \in \mathbb{N}$. Let

$$A_S^{(k)} f(p, \mathfrak{m}^P, \mathfrak{m}^Q) := \mathbb{1}_{U^{(k)}}((p, \mathfrak{m}^P, \mathfrak{m}^Q)) A_S f(p, \mathfrak{m}^P, \mathfrak{m}^Q)$$

and

$$A_N^{(k)} f(p, \mathfrak{m}^P, \mathfrak{m}^Q) := \mathbb{1}_{U^{(k)}}((p, \mathfrak{m}^P, \mathfrak{m}^Q)) A_N f(p, \mathfrak{m}^P, \mathfrak{m}^Q),$$

where A_S is the part of operator A described by (II.2.3.1) and A_N is the sum of A^P and A^Q .

Lemma II.2.5.1 *Let $C^2(E)$ be the space of bounded functions $f : E \mapsto \mathbb{R}$ which are twice continuously differentiable with respect to p . Then the closure of*

$$\{(f, A^{(k)} f) : f \in C^2(E)\} = \{(f, A_S^{(k)} f + A_N^{(k)} f) : f \in C^2(E)\} \quad (\text{II.2.5.1})$$

generates a Feller semigroup on $C(E)$.

Proof Since the selection coefficient s is Lipschitz continuous, we get from Lemma A.3.1 that the closure of $\{(f, A_S f) : f \in C^2([0, 1])\}$ generates a Feller semigroup on $C([0, 1])$. Since for a fixed p , the process generated by $A_N^{(k)}$ is continuous-time Markov chain with bounded rates and on a finite state space, it follows that $A_N^{(k)}$ generates a Feller semigroup on continuous functions (which in this case means simply all functions) on $\mathcal{N} := \prod_{I \in \mathcal{I}} \{0, 1, \dots, n_0\}$.

Denote $\bar{n} := \sum_{I \in \mathcal{I}} (n_I^P(0) + n_I^Q(0)) |I|$ and $r := \max_{I, \hat{I} \in \mathcal{I}} r_{I, \hat{I}}$. Both $A_N^{(k)}$ and $A_S^{(k)}$ can be seen as acting on the whole E and since

$$\|A_N^{(k)} f\| \leq (k\bar{n}^2(2^K + 2^{2K}) + \bar{n}(k2^K + r(2^K + 2^{2K}))) \|f\|,$$

we can apply Lemma A.3.2 to see that the closure of their sum generates a strongly continuous contraction semigroup. From the Trotter formula A.3.3 we see, that the positivity and conservative property also hold for the semigroup generated by the sum. \blacksquare

For any $k \in \mathbb{N}$ let $X^{(k)}$ be the process whose generator is the closure of (II.2.5.1). We can see the process $X^{(k)}$ as a solution to a stopped martingale problem. Put

$$\tau_k := \inf\{t \geq 0 : (p_t, \mathfrak{m}^P(t), \mathfrak{m}^Q(t)) \notin U_k\}. \quad (\text{II.2.5.2})$$

Then

$$\begin{aligned} & f((p_t, \mathfrak{n}^P(t), \mathfrak{n}^Q(t))) - \int_0^{t \wedge \tau_k} A^{(k)} f((p_s, \mathfrak{n}^P(s), \mathfrak{n}^Q(s))) ds \\ &= f((p_t, \mathfrak{n}^P(t), \mathfrak{n}^Q(t))) - \int_0^{t \wedge \tau_k} A f((p_s, \mathfrak{n}^P(s), \mathfrak{n}^Q(s))) ds \end{aligned}$$

is a martingale. Since $A^{(k)}$ generates a Feller semigroup, Theorem 4.1. in Chapter 5 of [EK86] ensures, that the stopped martingale problem is well-posed. Our task is now to show the existence of a process corresponding to the generator A on the whole E . We will use the following result.

Theorem II.2.5.2 *Let (E, r) be a complete and separable metric space and let $A \subset C_b(E) \times B(E)$, where $C_b(E)$ are bounded continuous functions on E and $B(E)$ are bounded Borel measurable functions on E . Let $U_1 \subset U_2 \subset \dots$ be open subsets of E . Fix $\nu \in \mathcal{P}(E)$ and suppose that for each K there exists a unique solution X_k of the stopped martingale problem for (A, ν, U_k) (where ν denotes the initial distribution of the solution) with sample paths in $D_E[0, \infty)$. Setting*

$$\tau_k := \inf\{t : X_k(t) \notin U_k \text{ or } X_k(t-) \notin U_k\}, \quad (\text{II.2.5.3})$$

suppose that for each $t > 0$,

$$\lim_{k \rightarrow \infty} \mathbb{P}[\tau_k \leq t] = 0. \quad (\text{II.2.5.4})$$

Then there exists a unique solution of the $D_E[0, \infty)$ martingale problem for (A, ν) .

Proof [EK86], Chapter 4, Theorem 6.3. ■

To use the Theorem II.2.5.2 we need to show that (II.2.5.4) holds for $U_k = U^{(k)}$. First, we will need some preliminary results. For $a \in [0, 1]$ let $\gamma_{p_0}(a)$ be the first hitting time of a by the process $p = (p_t)_{t \geq 0}$ with the initial state $p_0 \in [0, 1]$. See A.3.4 for the definition of an (in)accessible boundary of a one-dimensional diffusion on a bounded interval.

Lemma II.2.5.3 *(i) Point 0 is an accessible boundary for the process of allele frequencies p if and only if $\mu_2 < 1/2$. Point 1 is an accessible boundary for p if and only if $\mu_1 < 1/2$.*

(ii) Let $\mu_2 < 1/2$ resp. $\mu_1 < 1/2$. Then for any fixed $p_0 \in (0, 1)$ and any $K > 0$ we have

$$\lim_{k \rightarrow \infty} \mathbb{P}^{p_0} \left[\int_0^{\gamma(1/k)} \frac{1}{p(s)} ds > K \right] = 1, \quad (\text{II.2.5.5})$$

respectively

$$\lim_{k \rightarrow \infty} \mathbb{P}^{p_0} \left[\int_0^{\gamma(1-1/k)} \frac{1}{1-p(s)} ds > K \right] = 1. \quad (\text{II.2.5.6})$$

Proof For the proof of (i) see [BES04], Lemma 4.4. As noted by Jesse Taylor, the second part of the proof of Lemma 4.4. in [BES04] which corresponds to the proof of (ii) is wrong. For the correct version of the proof see [Tay07], Lemma 2.1. ■

Lemma II.2.5.4 *Let τ_k be as in (II.2.5.2). Then it holds for all $t > 0$ that*

$$\lim_{k \rightarrow \infty} \mathbb{P}[\tau_k \leq t] = 0.$$

Proof The proof is almost the same as in the case when $K = 1$. However, since some details were omitted in the proof of the corresponding Proposition 4.3. in [BES04], we provide the proof here in full.

Choose $t > 0$ arbitrarily and assume that 0 is an inaccessible boundary for p . Then $\mathbb{P}[\gamma(0) \leq t] = 0$ for all $k \in \mathbb{N}$, where $\gamma(0)$ is the first hitting time of p at 0. Let $\mathfrak{n} \in \mathbb{N}_0^{\mathcal{I}}$ and $0 \neq \mathfrak{i} \in \mathbb{N}_0^{\mathcal{I}}$. Then it holds for $\times^{(k)}$ that the probability that the process hits $\{1/k\} \times \{\mathfrak{i}\} \times \{\mathfrak{n}\}$ before time t tends to zero as k tends to infinity. A symmetrical argument applies for the case when 1 is an inaccessible boundary.

Now, consider the case when 0 is an accessible boundary. For $\mathfrak{i} \in \mathbb{N}_0^{\mathcal{I}}$ define $|\mathfrak{i}| = \sum_{I \in \mathcal{I}} \mathfrak{i}_I \cdot |2^I|$, where $|2^I|$ denotes the number of nonempty subsets of I and let $r := \max\{r_{I,\hat{I}}; I, \hat{I} \in \mathcal{I}\}$. Then for a fixed $\mathfrak{n}^P(0)$ and $\mathfrak{n}^Q(0)$ and $n_0 = \bar{\mathfrak{n}}^P(0) + \bar{\mathfrak{n}}^Q(0)$ it holds that $\frac{n_0 r p}{2}$ is the upper bound for the rate of arrivals of the sampled individuals to background P due to recombination and $\frac{n_0 \mu^P p}{2(1-p)}$ is an upper bound for the rate of arrivals due to mutation. Choose $t > 0$ and $\delta > 0$ arbitrarily. For $p \in (0, 1)$ define

$$\lambda_u(p) := n_0 \left(\frac{p \mu^P}{2(1-p)} + \frac{r p}{2} \right) \quad (\text{II.2.5.7})$$

and

$$\lambda_l(p) := \frac{1-p}{p} \frac{\mu^Q}{2}. \quad (\text{II.2.5.8})$$

Then λ_u is an upper bound for the rate at which sampled individuals arrive at P and λ_l is a lower bound for the rate at which they jump away, provided that $\mathfrak{n}^P \neq \mathbb{0}$. Now we would like to estimate $\mathbb{P}^{p_0}[\mathfrak{n}^P(\gamma(1/k)) = \mathbb{0}]$, that is we want to estimate the probability that for an arbitrary $p_0 \in [0, 1]$, when p_t hits $1/k$ for the first time the number of sampled individuals in background P is zero.

For $0 \leq a < x < b \leq 1$ we have that

$$\mathbb{P}^x[\gamma(a) < \gamma(b)] = \frac{n(b) - n(x)}{n(b) - n(a)},$$

where $n(x)$ is the scale function of (A.3.2) (see for example [Kal06], Theorem 23.7, page 456). In our case the scale function is

$$n(x) = \int_c^x \exp\left(-\int_c^y \frac{-\mu^P z + \mu^P(1-z) + s(z)z(1-z)}{\frac{1}{2}z(1-z)} dz\right) dy,$$

which can be approximated for small x by $\int_0^x y^{-2\mu^Q} dy$ and for all N there exists an $\tilde{\epsilon}(N)$ such that for all $\epsilon < \tilde{\epsilon}(N)$

$$\left| \frac{n(N\epsilon) - n(\epsilon)}{n(N\epsilon) - n(0)} - \frac{\int_{\epsilon}^{N\epsilon} y^{-2\mu^Q} dy}{\int_0^{N\epsilon} y^{-2\mu^Q} dy} \right| < \delta/16.$$

If we choose N large enough so that

$$\frac{\int_{\epsilon}^{N\epsilon} y^{-2\mu^Q} dy}{\int_0^{N\epsilon} y^{-2\mu^Q} dy} = \frac{N^{1-2\mu^Q} - 1}{N^{1-2\mu^Q}} > 1 - \delta/16$$

then

$$\mathbb{P}^{\epsilon}[\gamma(0) < \gamma(N\epsilon)] > 1 - \delta/8 \quad (\text{II.2.5.9})$$

for all $\epsilon < \tilde{\epsilon}(N)$. We use (II.2.5.9) to restrict our attention to the event that between the time p first hits ϵ and then $1/k$ we always have $p < N\epsilon$. Let $S_{N\epsilon}$ be an exponentially distributed random variable with rate $\lambda_u(N\epsilon)$. Next, choose ϵ small enough so that

$$\mathbb{P}[S_{N\epsilon} > t] > 1 - \delta/8. \quad (\text{II.2.5.10})$$

Equation (II.2.5.10) ensures that there are no individuals in the sample which come to the background P between the first hitting of 0 and of $N\epsilon$ since $\lambda_u(p)$ is increasing in p , so the actual arrival rate to the background P is always below $\lambda_u(N\epsilon)$. Let Y be a Poisson random variable with mean K . Choose K large enough so that

$$\mathbb{P}[Y > n_0] > 1 - \delta/8. \quad (\text{II.2.5.11})$$

Suppose that $p_0 \geq \epsilon$ and using Lemma II.2.5.3 choose $k_0 = k_0(\delta)$ large enough that for $k > k_0$,

$$\mathbb{P}^{p_0} \left[\int_{\gamma(\epsilon)}^{\gamma(1/k)} \lambda_l(p_s) ds > K \right] > 1 - \delta/8. \quad (\text{II.2.5.12})$$

From (II.2.5.12) we have that the emigration rate in that time interval is at least K (at least as long as there are any individuals still in P) and (II.2.5.11) ensures that all individuals actually jump out of P before the process hits $1/k$. Putting together (II.2.5.9)-(II.2.5.12), we get that with probability at least $1 - \delta/2$ when the process p hits $1/k$, $\mathfrak{m}^P = \mathbb{0}$.

Starting from $p_0 = 1/k$ and $\mathfrak{m}^P(0) = \mathbb{0}$, we now let the process run until the first time T that $\mathfrak{m}^P \neq \mathbb{0}$. Let $p_t^{1/k}$ be the process of allele frequencies started at $p_0^{1/k} = 1/k$. We would like to find a lower bound for $T = T(1/k)$, at least for k high enough, which is independent of ϵ , δ and \aleph_0 . To do that, we first see from the Comparison Theorem A.3.5 that

$$\mathbb{P}[p_t^{1/2} \geq p_t^{1/k}, t \geq 0] = 1, \quad k > 1.$$

Next, we construct a Cox process $\tilde{M}(t)$ with rate

$$\tilde{\lambda}(t) = \frac{p_t^{1/2}}{q_t^{1/2}} \mu^P \frac{n_0}{2} + r \frac{n_0 p_t^{1/2}}{2}, \quad (\text{II.2.5.13})$$

where $q_t^{1/k} := 1 - p_t^{1/k}$ for $k \in \mathbb{N}$. This process, conditioned on the process of allele frequencies, is a time inhomogeneous Poisson process. It is clear that for $k > 1$ it holds for the rates

$$\lambda_{1/k}(t) = \frac{1}{2} \frac{p_t^{1/k}}{q_t^{1/k}} \mu^P \sum_{I \in \mathcal{I}} n_I^Q(t) + \frac{1}{2} p_t^{1/k} \sum_{I \in \mathcal{I}} \sum_{\hat{I} \in \mathcal{I}} r_{I, \hat{I}} n_I^Q(t), \quad k \in \mathbb{N} \quad (\text{II.2.5.14})$$

of the arrivals of sampled individuals to P in process X that $\lambda_{1/k}(t) \leq \tilde{\lambda}(t)$. Let $N_1(t)$ and $N_2(t)$ two independent Poisson processes with rate 1. Then

$$M_{1/k}(t) := N_1 \left(\int_0^t \lambda_{1/k}(s) ds \right)$$

is a Cox process with rate $\lambda_{1/k}(t)$ which is also the process of arrivals of sampled individuals to the background P . If we put

$$\tilde{M}(t) := M_{1/k}(t) + N_2 \left(\int_0^t (\tilde{\lambda}(s) - \lambda_{1/k}(s)) ds \right),$$

then it clearly holds for all $k > 1$ that

$$\mathbb{P}[\tilde{M}(t) \geq M_{1/k}(t), t \geq 0] = 1$$

and the distribution of $\tilde{T} := \inf\{t \geq 0, \tilde{M}(t) \neq 0\}$ is a lower bound of $T = T(1/k)$ for all $k > 1$, independent of δ and ϵ , that is $\tilde{T} \stackrel{\text{as}}{\leq} T$ for any $k \in \mathbb{N}$.

Next, we need to make sure that at time T the probability that the process p is greater than ϵ is at least $1 - \delta/2$ so that we can use the machinery above. It is possible to find an ϵ_0 which satisfies this for $p_0 = 0$. If it were not, then we would have $\mathbb{P}^0[p_T = 0] > 0$. However, letting $\lambda_0(t)$ be defined as in (II.2.5.14) but with $p_0 = 0$ we see that

$$\mathbb{P}^0[p_T = 0] = \mathbb{E}^0[\mathbb{P}^0[p_T = 0 | (p_t)_{t \geq 0}]] = \mathbb{E}^0 \left[\int_0^\infty \lambda_0(t) e^{-\int_0^t \lambda_0(s) ds} 1_{[p_t=0]} dt \right] = 0,$$

since $\lambda_0(t) = 0$ when $p_t = 0$. Therefore we can find an $\epsilon_0 > 0$ small enough, so that

$$\mathbb{P}^0[p_T \geq \epsilon_0] \geq 1 - \delta/12. \quad (\text{II.2.5.15})$$

Denote by V the first time the configuration of active loci in background Q changes, i.e. $V := \inf\{t > 0; n^Q(t) \neq n^Q(0)\}$. Since if $p_0 = 0$ then $\gamma(1/k) \rightarrow 0$ as $k \rightarrow \infty$ almost surely and $\mathbb{P}(V = 0) = 0$, we see that we can find a $k_1 \in \mathbb{N}$ large enough so that

$$\mathbb{P}^0(V > \gamma(1/k)) \geq 1 - \delta/12, \quad k \geq k_1. \quad (\text{II.2.5.16})$$

From (II.2.5.15) we also obtain that $\mathbb{P}^0[p_T > 1/k] \geq 1 - \delta/12$ for all $k > 1/\epsilon_0$ and so it holds for all $k \geq \max(1/\epsilon_0, k_1)$ and all $\epsilon > 0$ that

$$\begin{aligned} \mathbb{P}^0[p_T > \epsilon] &\leq \mathbb{P}^0[p_T > \epsilon, T > \gamma(1/k), V > \gamma(1/k)] + \delta/4 \\ &\leq \delta/4 + \mathbb{P}^0[p_T > \epsilon | T > \gamma(1/k), V > \gamma(1/k)] \mathbb{P}^0[T > \gamma(1/k), V > \gamma(1/k)] \\ &\leq \delta/4 + \mathbb{P}^0[p_T > \epsilon | T > \gamma(1/k), V > \gamma(1/k)], \end{aligned} \quad (\text{II.2.5.17})$$

where the first inequality comes from the fact that if $p_0 = 0$ then $\{T \leq \gamma(1/k)\} \subseteq \{p_T \leq 1/k\}$. We can write $T = \inf\{t > 0; \int_0^t \int_0^{\lambda(s)} N(dr, ds) > 0\}$, where $\lambda = \lambda_{p_0}$ and N is a Poisson point process on \mathbb{R}_+^2 with intensity 1 and independent of \mathbb{X} . Denoting by $(\theta_t)_{t \geq 0}$ the shift operator, we have on the event $\{T > \gamma(1/k)\}$ that

$$T = \gamma(1/k) + \inf\{t > \gamma(1/k); \int_{\gamma(1/k)}^t \int_0^{\lambda(\gamma(1/k)+s)} N(dr, ds) > 0\} = \gamma(1/k) + T \circ \theta_{\gamma(1/k)}.$$

Hence putting $\mathbb{X}_0 := (0, \mathbb{0}, n^Q(0))$ we get

$$\begin{aligned} \mathbb{P}^{\mathbb{X}_0}[p_T > \epsilon | T > \gamma(1/k), V > \gamma(1/k)] &= \mathbb{P}^{\mathbb{X}_0}[p_{\gamma(1/k)+T \circ \theta_{\gamma(1/k)}} > \epsilon | X_{\gamma(1/k)} = (1/k, \mathbb{0}, n^Q(0))] \\ &= \mathbb{P}^{(1/k, \mathbb{0}, n^Q(0))}[p_T > \epsilon]. \end{aligned} \quad (\text{II.2.5.18})$$

Combining (II.2.5.15), (II.2.5.17) and (II.2.5.18) we get for $\epsilon_1 := \min(\epsilon_0, 1/k_1)$ that

$$1 - \delta/4 \leq \mathbb{P}^0[p_T > \epsilon_1] \leq \delta/4 + \mathbb{P}^{1/k}[p_T > \epsilon_1]$$

and so we have shown that we have for all $k > 1/\epsilon_0$ that $\mathbb{P}^{1/k}[p_T > \epsilon_1] > 1 - \delta/2$.

Putting all of the above together we see that the process p hits $p = 1/k$ with $n^P \neq \mathbb{0}$ only after a geometric number of hits of p at $1/k$ with $n^P = \mathbb{0}$. The success probability of this geometric random variable is at most δ . Each failure adds a waiting time bounded below by an independent copy of \tilde{T} . Since δ was arbitrary, the proof is complete. The case when 1 is an accessible boundary is treated symmetrically. \blacksquare

II.2.5.2 Convergence

We still need to prove that the processes $\mathbb{X}^{(N)}$ indeed do converge to \mathbb{X} . We will start with proving a pointwise convergence of their generators.

Lemma II.2.5.5 *For $p \in (0, 1)$, $A^{(N)}f(p, \mathfrak{n}^P, \mathfrak{n}^Q) \xrightarrow{N \rightarrow \infty} Af(p, \mathfrak{n}^P, \mathfrak{n}^Q)$, where $A^{(N)}$ is the generator of process $\mathbb{X}^{(N)}$.*

Proof The terms of the generator $A^{(N)}$ have the same form as (II.2.2.11), except that they are multiplied by a further factor of N due to the time-rescaling and all mutation and recombination rates as well as the selection coefficient are rescaled as at the beginning of Section II.2.3.

We see that the events of the type described in (II.2.2.8) and (II.2.2.7) do not appear in the limit. Indeed, (II.2.2.7) is of order $O(1/N^2)$ and the corresponding event only happens if there is a mutation. With the rescaled mutation this shows that the corresponding term in the generator $A^{(N)}$ is of order $O(1/N^3)$ and since birth-death events happen at rate N^2 these type of events only happen at rate $O(1/N)$ and so they vanish in the limit. Similarly, (II.2.2.8) is of order $O(1/N)$ and the corresponding event only occurs at rate $O(1/N^3)$ since it only happens in a birth-death event of the form (P, Q) or (P, P) whose probabilities are of order $O(1/N^2)$ (see (II.2.2.1)). All other neutral loci events which include recombination are also of rate $O(1/N)$ and disappear in the limit with the exception for the event in which two individuals of the same background coalesce with only the offspring in the sample with the corresponding probability (II.2.2.5).

In the generator $A^{(N)}$ the difference between c (which was defined below the formula (II.2.2.11)) and 1 will be of order $O(1/N)$ and so the term c disappears in the limit. If we then sum the terms of the generator $A^{(N)}$ which contain (II.2.2.2), sum the result over all possible $I \in \mathcal{I}$ then in the limit we obtain (II.2.3.2). The terms (II.2.3.3) and (II.2.3.4) again arise in a similar fashion from those terms in $A^{(N)}$ which contain (II.2.2.3) (or (II.2.2.4)). We only need to differentiate between the case when the neutral type of one individual is a subset of the neutral type of the other and when their neutral types have a non-empty symmetric difference just as we did in the finite population case, since each event affects the process differently. Similarly, we get (II.2.3.5) from the terms containing (II.2.2.6). The last two terms (II.2.3.6) and (II.2.3.7) come from the terms in $A^{(N)}$ containing (II.2.2.5). Here, the process is affected differently in the case when $\hat{I} = I$ and when $\hat{I} \neq I$ and the outcome in the latter case depends on whether the individual that died in the selective event was from the background P or Q . Finally, the terms (II.2.2.9) and (II.2.2.10) lead to (II.2.3.1) by Taylor expansion just as in the proof of Lemma 3.1. in [BES04]. \blacksquare

We want to show the convergence of the process $\mathbb{X}^{(N)}$ generated by $A^{(N)}$ to the process X generated by A . First, we will show for all $k \in \mathbb{N}$ the convergence of the stopped processes $\mathbb{X}^{(N,k)}$ corresponding to the generators $A^{(N,k)} := \mathbf{1}_{U^{(k)}}A^{(N)}$ to the process \mathbb{X}_k generated by A^k . To that end, we will use the following theorem.

Theorem II.2.5.6 *Let (E, d) be a complete separable metric space and A a Feller generator on E corresponding to the Markov process X . For each $N \geq 1$ let further $X^{(N)}$ be progressively measurable E -valued process with full generators $\hat{A}^{(N)}$ and such that $X^{(N)}(0)$ converges weakly to $X(0)$ as $N \rightarrow \infty$. Suppose that $\overline{D(A)}$ separates points and that the compact containment condition holds for $\{X^{(N)}\}_{N \geq 1}$, that is, for every $\epsilon > 0$ and every $T > 0$ there exists a compact*

set $\Gamma_{\epsilon,T} \subseteq E$ for which

$$\inf_N \mathbb{P}(X^{(N)}(t) \in \Gamma_{\epsilon,T} \text{ for } 0 \leq t \leq T) \geq 1 - \epsilon.$$

Suppose that for each $(f, g) \in A$ and $T > 0$ there exist $(f^{(N)}, g^{(N)}) \in \hat{A}^{(N)}$ and $G^{(N)} \subseteq E$ such that

$$\lim_{N \rightarrow \infty} \mathbb{P}(X^{(N)}(t) \in G^{(N)}, 0 \leq t \leq T) = 1, \quad (\text{II.2.5.19})$$

$\sup_N \|f^{(N)}\|_\infty < \infty$ and

$$\lim_{N \rightarrow \infty} \sup_{x \in G^{(N)}} |f(x) - f^{(N)}(x)| = \lim_{N \rightarrow \infty} \sup_{x \in G^{(N)}} |g(x) - g^{(N)}(x)| = 0. \quad (\text{II.2.5.20})$$

Then $X^{(N)}$ converges weakly to X as $N \rightarrow \infty$.

Proof This is a special case of Corollary 8.3, Chapter 4 in [EK86]. ■

We now fix $k \in \mathbb{N}$. The compact containment condition of Theorem II.2.5.6 holds, since E itself is compact. In condition (II.2.5.19) we can simply choose $G^{(N)} = E$ for all $N \in \mathbb{N}$ and $f^{(N)}$ in (II.2.5.20) can be chosen as $f^{(N)} := f|_{E^{(N)}}$. This proves the convergence of $\mathbb{X}^{(N,k)}$ to \mathbb{X}^k for any $k \in \mathbb{N}$ and combining Theorem II.2.5.4 with the following lemma gives us the convergence of $\mathbb{X}^{(N)}$ to \mathbb{X} .

Lemma II.2.5.7 *Let $\{\mathbb{P}^{(N)}\}_{N \geq 1}$ be a sequence of probability measures on the space $D_E[0, \infty)$ and suppose that $T^{(k)}$ is a nondecreasing sequence of stopping times (with respect to the natural filtration) increasing to infinity almost surely. For each $k \geq 1$, let $\{\mathbb{P}^{(N,k)}\}_{N \geq 1}$ be a relatively compact sequence of probability measures such that $\mathbb{P}^{(N,k)}$ is equal to $\mathbb{P}^{(N)}$ on $\mathcal{F}_{T^{(k)}}$. If the probability measure \mathbb{P} has the property that for any $k \geq 1$ any limit point of $\{\mathbb{P}^{(N,k)}\}_{N \geq 1}$ agrees with \mathbb{P} on $\mathcal{F}_{T^{(k)}}$, then $\mathbb{P}^{(N)}$ converges to \mathbb{P} as $N \rightarrow \infty$.*

Proof See [SV79], Lemma 11.1.1 ■

Appendix A

A.1 A few definitions from graph theory

Let $G = (V, E)$ be a graph with a countable vertex set V and the set of edges E with each edge $e \in E$ of the form $\{v, w\}$ for some $v, w \in V$.

We say that G connected if there exists a path between each two vertices of G , that is if for all $v, w \in V$ there exists a finite sequence $(v_i)_{i \in \{1, \dots, N\}}$ of $N \in \mathbb{N}$ vertices such that $v = v_0$, $w = v_N$ and $\{v_i, v_{i+1}\} \in E$ for all $i \in \{1, \dots, N\}$.

An automorphism of G is defined as a permutation π of V such that for all $v, w \in V$ it holds that $\{v, w\} \in E$ if and only if $\{\pi(v), \pi(w)\} \in E$. We say that G is vertex transitive if for any two vertices $v, w \in V$ there exists an automorphism π of G such that $\pi(v) = w$.

We say that G is locally finite if every vertex $v \in V$ has a finite degree, i.e. if $\deg_G(v) := |\{w \in W : \{v, w\} \in E\}| < \infty$ for all $v \in V$.

A.2 Pathwise duality for monotone interacting particle systems

Let $S = (S, \leq)$ be a partially ordered set. For any subset $A \subseteq S$ we put

$$A^\uparrow := \{x \in S : y \leq x \text{ for some } y \in A\}.$$

Clearly $A \subseteq A^\uparrow$. We say that A is *increasing* if also $A^\uparrow \subseteq A$. We define A^\downarrow and *decreasing* sets analogously with the order reversed. We define the dual of a partially ordered set S as a partially ordered set S' together with a bijection $S \ni x \rightarrow x' \in S'$ where for all $x, y \in S$ it holds that

$$x \leq y \quad \text{if and only if} \quad y' \leq' x'. \quad (\text{A.2.1})$$

For any set $A \subseteq S$ we write $A' := \{x' : x \in A\}$ and we denote by $A_{max} := \{x \in A : \nexists y \in A, y \neq x \text{ such that } x \leq y\}$ the set of maximal elements of A and analogously we denote by A_{min} the set of its minimal elements.

We now assume that S is a finite partially ordered set bounded from above by an element denoted by 1 and that Λ is a countable set. We equip the set S^Λ with a product order

$$x \leq y \text{ if and only if } x(i) \leq y(i) \text{ for all } i \in \Lambda. \quad (\text{A.2.2})$$

Then also S^Λ is bounded from above with the upper bound 1 given by $1(i) := 1$ for all $i \in \Lambda$. If S' is the dual of S then $(S')^\Lambda$ is the dual of S^Λ where for any $x \in S^\Lambda$ and $i \in \Lambda$ we put $x'(i) := (x(i))'$. Since S and S^Λ are bounded from above, the sets S' and $(S')^\Lambda$ are bounded from below by $0 := 1'$ and $0 := 1'$ respectively. Now, for any $x \in (S')^\Lambda$ and $B \in \mathcal{P}((S')^\Lambda)$ we

put

$$\begin{aligned} \text{supp}(x) &:= \{i \in \Lambda : x(i) \neq 0\}, \\ \text{supp}(B) &:= \{i \in \Lambda : x(i) \neq 0 \text{ for some } x \in B\} = \bigcup_{x \in B} \text{supp}(x) \end{aligned} \quad (\text{A.2.3})$$

and we let $(S')_{loc}^\Lambda := \{x \in (S')^\Lambda : |\text{supp}(x)| < \infty\}$ be the set of finitely supported $x \in (S')^\Lambda$. By $P_* := \mathcal{P}_{\text{fin}}((S')_{loc}^\Lambda)$ we denote the set of finite subsets of $(S')_{loc}^\Lambda$. Finally, we define a duality function $\phi : S^\Lambda \times P_* \rightarrow \{0, 1\}$ by

$$\phi(x, B) := \mathbf{1}_{\{x \leq y' \text{ for some } y' \in B\}}, \quad x \in S^\Lambda, B \subseteq P_*. \quad (\text{A.2.4})$$

Let (S, \leq) and (T, \leq') be partially ordered sets. We say that a map $m : S \rightarrow T$ is *monotone* if it is a (\leq, \leq') -homomorphism, that is if

$$m(x) \leq' m(y) \quad \text{whenever} \quad x \leq y, \quad x, y \in S. \quad (\text{A.2.5})$$

A particle system is called *monotonically representable* if it can be represented using monotone maps only, that is if all $m \in \mathcal{G}$ in (I.1.2.2) are monotone maps. A monotonically representable particle system is monotone and while the opposite is necessarily true in general, all processes we are studying in Part I of this thesis are constructed using a random mapping representation consisting of only monotone maps.

For monotone interacting particle systems we obtain the following theorem which appears in [SS16] as Proposition 30 in a slightly different formulation.

Theorem A.2.1 (Pathwise duality for monotone systems, [SS16]) *Let S be a finite partially ordered set which is bounded from above, Λ a countable set and let X be a S^Λ -valued monotone interacting particle system with a generator of the form (I.1.2.2) where all elements of \mathcal{G} are monotone local maps which satisfy (I.1.2.3). Then the P_* -valued process Y^* and the $P_\dagger := \{B \in P_* : B = B_{\min}\}$ -valued Markov process Y^\dagger are pathwise dual to X with respect to the duality function ϕ as defined in (A.2.4), where the generators H_\dagger of Y^\dagger and H_* of Y^* are*

$$\begin{aligned} H_\dagger f(B) &:= \sum_{m \in \mathcal{G}} r_m (f(m^\dagger(B)) - f(B)), \\ H_* f(B) &:= \sum_{m \in \mathcal{G}} r_m (f(m^*(B)) - f(B)), \end{aligned} \quad (\text{A.2.6})$$

respectively, and for each $m \in \mathcal{G}$ the maps $m^\dagger : \mathcal{P}(S') \rightarrow \mathcal{P}(S')$ and $m^* : \mathcal{P}(S') \rightarrow \mathcal{P}(S')$ are dual to m (with respect to ϕ) and defined as

$$m^\dagger(B)' := (m^{-1}(B^\downarrow))_{\max} \quad \text{and} \quad m^*(B)' := \bigcup_{x \in B} (m^{-1}(\{x'\}^\downarrow))_{\max}. \quad (\text{A.2.7})$$

The proof of Theorem (A.2.1) is based on the fact that the subspace of decreasing subsets is invariant with respect to an inverse image of a monotone map and so we can encode a decreasing set $A \subset S$ in terms of a set $B \subset S'$ such that $A = \{y' : y \in B\}^\downarrow$. As Sturm and Swart noted, a monotone map $m : S \rightarrow S$ stays monotone if we reverse the order on S . Therefore, the inverse map m^{-1} also maps increasing subsets into increasing subsets and using completely analogous arguments as in the proof of Theorem (A.2.1), we can obtain the following alternative duality.

Theorem A.2.2 (Alternative monotone duality, [SS16]) *Let S , X and Λ be as in the previous theorem. Then the P_* -valued process Y^\bullet and the $P_\diamond := \{B \in P_* : B = B_{\max}\}$ -valued Markov process Y^\diamond are pathwise dual to X with respect to the duality function*

$$\tilde{\phi}(x, B) := \mathbf{1}_{\{x \geq y' \text{ for some } y \in B\}}, \quad x \in S^\Lambda, \quad B \subseteq P_*, \quad (\text{A.2.8})$$

where the generators H_\diamond of Y^\diamond and H_\bullet of Y^\bullet are

$$\begin{aligned} H_\diamond f(B) &:= \sum_{m \in \mathcal{G}} r_m (f(m^\diamond(B)) - f(B)), \\ H_\bullet f(B) &:= \sum_{m \in \mathcal{G}} r_m (f(m^\bullet(B)) - f(B)) \end{aligned} \quad (\text{A.2.9})$$

respectively, and for each $m \in \mathcal{G}$ the maps $m^\diamond : \mathcal{P}(S') \rightarrow \mathcal{P}(S')$ and $m^\bullet : \mathcal{P}(S') \rightarrow \mathcal{P}(S')$ are dual to m (with respect to $\tilde{\phi}$) and defined as

$$m^\diamond(B)' := (m^{-1}(B'^\uparrow))_{\min} \quad \text{and} \quad m^\bullet(B)' := \bigcup_{x \in B} (m^{-1}(\{x'\}^\uparrow))_{\min}. \quad (\text{A.2.10})$$

For more details of the construction of the duals and the proof of the Theorem (A.2.1) see [SS16].

A.3 Appendix to Section II.2.5

Lemma A.3.1 *Let $a : [0, 1] \mapsto \mathbb{R}_+$, $a(x) = x(1-x)$ and let $b : [0, 1] \mapsto \mathbb{R}$ be Lipschitz continuous for which $b(0) \geq 0$ and $b(1) \leq 0$. Then the closure of $\{(f, Gf) : f \in C^2([0, 1])\}$, where*

$$G := \frac{1}{2}a(x)\frac{d^2}{dx^2} + b(x)\frac{d}{dx},$$

generates a Feller semigroup on $C([0, 1])$.

Proof See [EK86], page 375. ■

Lemma A.3.2 (Perturbation) *Let A be a generator of a strongly continuous contraction semigroup on $\mathcal{C}(E)$ and B a dissipative linear operator on $\mathcal{C}(E)$, where $\mathcal{D}(A) = \mathcal{D}(B)$. If B is a bounded operator, then the closure of $A + B$ generates a strongly continuous contraction semigroup on $\mathcal{C}(E)$.*

Proof See [EK86], page 37. ■

Lemma A.3.3 (Trotter formula) *Let T, S and U be strongly continuous contraction semigroups on $\mathcal{C}(E)$ with generators A, B and C , respectively, so that $\mathcal{D}(A) = \mathcal{D}(B) = \mathcal{D}(C)$ and $A = B + C$. Then for all $f \in \mathcal{C}(E)$ it holds that*

$$\lim_{n \rightarrow \infty} \left[S\left(\frac{t}{n}\right) U\left(\frac{t}{n}\right) \right]^n f = T(t)f$$

for all $t \geq 0$.

Proof See [EK86], page 33. ■

Definition A.3.4 (Scale and speed, accessible boundary) Let $b_1 < b_2 \in \mathbb{R}$ and let X be a one-dimensional diffusion process on $[b_1, b_2]$ with generator

$$L = \frac{1}{2}a(x)\frac{d^2}{dx^2} + b(x)\frac{d}{dx}. \quad (\text{A.3.1})$$

Then we define the scale $n(x)$ and speed $m(x)$ of the process for $x \in [b_1, b_2]$ by

$$n(x) = \int_c^x \exp\left(-\int_c^y \frac{2b(z)}{a(z)}dz\right) dy, \quad (\text{A.3.2})$$

and

$$m(x) = \int_c^x \frac{2}{a(y)} \exp\left(\int_c^y \frac{2b(z)}{a(z)}dz\right) dy, \quad (\text{A.3.3})$$

where $c \in (b_1, b_2)$ is fixed arbitrarily.

We say that b_i , $i = 1, 2$ is an accessible boundary if it holds for $u(x) = \int_c^x m(y)dn(y)$ that $u(b_i) < \infty$. Otherwise, we say that b_i is inaccessible.

Theorem A.3.5 (Yamada, Comparison Theorem) Consider the equations

$$dX_t^1 = b_1(X_t)dt + \sigma(X_t)dW_t, \quad (\text{A.3.4})$$

$$dX_t^2 = b_2(X_t)dt + \sigma(X_t)dW_t, \quad (\text{A.3.5})$$

where b_1 and b_2 are continuous and σ is Lipschitz continuous on $[0, 1]$ with a Lipschitz constant L . Further, assume that

$$b_1(x) \leq b_2(x), \quad x \in [0, 1]. \quad (\text{A.3.6})$$

Let X_t^1 and X_t^2 be solutions of (A.3.4) and (A.3.5) respectively, such that

$$X_0^1 \leq X_0^2 \text{ almost surely}$$

Then

$$\mathbb{P}(X_t^1 \leq X_t^2 \text{ for all } t \geq 0) = 1.$$

Proof We will first prove the theorem for

$$b_1(x) < b_2(x), \quad x \in [0, 1]. \quad (\text{A.3.7})$$

Define time $\xi_0 := \inf\{t \geq 0; X_t^2 \leq X_t^1\}$ until which $X_t^1 \leq X_t^2$ almost surely and put $\tau_0 := \inf\{t \geq \xi_0; b(X_t^2) - b(X_t^1) \leq 0\}$. The drifts b_1 and b_2 are continuous, X_t^1 and X_t^2 have almost surely continuous paths and $b(X_{\xi_0}^1) < b(X_{\xi_0}^2)$ almost surely, so clearly $\mathbb{P}(\xi_0 = 0) = 0$. Define $t' := t \wedge \tau_0$. We see that

$$\begin{aligned} \mathbb{E}[X_{t'}^2 - X_{t'}^1] &= \mathbb{E}\left[\int_0^{t'} (b(X_s^2) - b(X_s^1))ds\right] + \mathbb{E}[X_0^2 - X_0^1] \\ &\quad + \mathbb{E}\left[\int_0^{t'} (\sigma(X_s^2) - \sigma(X_s^1))dW_s\right], \end{aligned} \quad (\text{A.3.8})$$

where the last integral on the right is zero.

We would like to estimate $\mathbb{E}[|X_{t'}^2 - X_{t'}^1|]$. We will make a use of the Itô formula and we construct a twice continuously differentiable approximation of the absolute value for that purpose. To do that, first define $\{a_n\}_{n \in \mathbb{N}}$ such that $1 = a_0 > a_1 > \dots > a_n \xrightarrow{n \rightarrow \infty} 0$ and

$$\int_{a_n}^{a_{n-1}} \frac{1}{L^2 x^2} dx = n, \quad n \in \mathbb{N}$$

which is possible since $\int_{0+}^1 1/x^2 := \lim_{\epsilon \rightarrow 0+} \int_{\epsilon}^1 1/x^2 = \infty$. Then for each $n \in \mathbb{N}$ let $f_n \in \mathcal{C}^2([-1, 1])$ be as follows. For each $n \in \mathbb{N}$ put $f_n(0) = 0$ and construct the first and second derivatives f_n' and f_n'' in such a way that

$$f_n'(x) \begin{cases} = 0, & 0 \leq x < a_n \\ \in [0, 1], & a_n \leq x \leq a_{n-1} \\ = 1, & x > a_{n-1}, \end{cases}$$

and

$$f_n''(x) \begin{cases} = 0, & 0 \leq x < a_n \\ \in [0, \frac{2}{n} \frac{1}{L^2 x^2}], & a_n \leq x \leq a_{n-1} \\ = 0, & x > a_{n-1}. \end{cases} \quad (\text{A.3.9})$$

Next, make f_n symmetric by putting $f_n(-x) := f_n(x)$, $x \in [0, 1]$. For functions f_n we now have that

$$f_n(x) \uparrow |x|, \quad n \rightarrow \infty. \quad (\text{A.3.10})$$

From the Itô formula we obtain

$$\begin{aligned} f_n(X_{t'}^2 - X_{t'}^1) &= f_n(X_0^2 - X_0^1) + \int_0^{t'} f_n'(X_s^2 - X_s^1)(b_2(X_s^2) - b_1(X_s^1)) ds \\ &\quad + \int_0^{t'} f_n'(X_s^2 - X_s^1)(\sigma(X_s^2) - \sigma(X_s^1)) dW_s \\ &\quad + \frac{1}{2} \int_0^{t'} f_n''(X_s^2 - X_s^1)(\sigma(X_s^2) - \sigma(X_s^1))^2 ds, \\ &= f_n(X_0^2 - X_0^1) + I_1 + I_2 + I_3. \end{aligned} \quad (\text{A.3.11})$$

Obviously, $\mathbb{E}[I_2] = 0$. Since $b_1(X_s^1) \leq b_2(X_s^2)$ almost surely for $s \in [0, t']$ and $|f'(x)| \leq 1$ for all x ,

$$\mathbb{E}[I_1] \leq \mathbb{E} \left[\int_0^{t'} (b_2(X_s^2) - b_1(X_s^1)) ds \right].$$

Using (A.3.9) we get that

$$\begin{aligned} \mathbb{E}[I_3] &\leq \frac{1}{2} \mathbb{E} \left[\int_0^{t'} f_n''(|X_s^2 - X_s^1|) \frac{1}{L^2 (X_s^2 - X_s^1)^2} \right] \\ &\leq \frac{t}{2} \cdot \max_{a_n \leq x \leq a_{n-1}} f_n''(x) \frac{1}{L^2 x^2} \leq \frac{t}{2} \frac{2}{n} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

From (A.3.8), (A.3.10), (A.3.11) and (A.3.7) we see that

$$\mathbb{E}[|X_{t'}^2 - X_{t'}^1|] \leq \mathbb{E}[|X_0^2 - X_0^1|] + \mathbb{E} \left[\int_0^{t'} (b_2(X_s^2) - b_1(X_s^1)) ds \right] = \mathbb{E}[X_{t'}^2 - X_{t'}^1] \quad (\text{A.3.12})$$

and so

$$\mathbb{E}[|X_{t'}^2 - X_{t'}^1|] = \mathbb{E}[X_{t'}^2 - X_{t'}^1].$$

This implies that $X_t^1 \leq X_t^2$ almost surely and therefore $X_t^1 \leq X_t^2$ for all $t \in [0, \tau_0)$ almost surely. Now, for $k \in \mathbb{N}$ put $\xi_k := \inf\{t \geq \tau_{k-1}; X_t^1 \leq X_t^2\}$ and $\tau_k := \inf\{t \geq \xi_k; b_2(X_t^2) \leq b_1(X_t^1)\}$ and proceed similarly as above. Since $\lim_{k \rightarrow \infty} \tau_k = \infty$ almost surely, we are done for the case when (A.3.7) holds. To see that this is true, assume that there exists an almost surely finite limit τ_∞ . Put $\xi_\infty := \inf\{t \geq \tau_\infty; X_t^2 \leq X_t^1\}$. We see from (A.3.6) that τ_∞ cannot be finite.

Generally for the case (A.3.6), put for $n \in \mathbb{N}$ $b_1^n := b_1 - \frac{1}{n}$ and $b_2^n := b_2 + \frac{1}{n}$. Let $X^{1,n}$ and $X^{2,n}$ be the solutions of (A.3.4) and (A.3.5), respectively, with b_1 replaced by b_1^n and b_2 by b_2^n . We have already shown that the theorem holds for $X^{1,n}$ and $X^{2,n}$ for all $n \in \mathbb{N}$ and since the drifts are continuous, we have that $X^{1,n} \rightarrow X^1$ and $X^{2,n} \rightarrow X^2$ as $n \rightarrow \infty$ almost surely. ■

Bibliography

- [AB05] D. Aldous and A. Bandyopadhyay. *A survey of max-type recursive distributional equations*. *Annals of Applied Probability* (2005), 1047-1110.
- [BBE13] M. Birkner, J. Blath, B. Eldon. *An ancestral recombination graph for diploid populations with skewed offspring distribution*. *Genetics* 193.1 (2013): 255-290.
- [BE04] N. Barton and A. Etheridge. *The effect of selection on genealogies*. *Genetics* 166.2 (2004): 1115-1131.
- [BES04] N. Barton, A. Etheridge, A. Sturm. *Coalescence in a random background*. *The Annals of Applied Probability* 2004, Vol. 14, No.2, pp. 758-785
- [Bil86] P. Billingsley. *Probability and Measure*. Wiley, New York, 1986
- [BG90] C. Bezuidenhout and G. Grimmett. *The critical contact process dies out*. *The Annals of Probability* (1990): 1462-1482.
- [BG94] C. Bezuidenhout and L. Gray. *Critical attractive spin systems*. *The Annals of Probability* (1994): 1160-1194.
- [BS14] R. Brink-Spalink. *Stochastic Models in Population Genetics: The Impact of Selection and Recombination*. (2015).
- [CR84] J. Cox and U. Roesler. *A duality relation for entrance and exit laws for Markov processes*. *Stochastic processes and their applications* 16.2 (1984): 141-156.
- [DN08] R. Darling and J. Norris. *Differential equation approximations for Markov chains*. *Probability surveys* 5 (2008): 37-79.
- [Dob71] R. Dobrushin. *Markov processes with many locally interacting components—the reversible case and some generalizations*. *Problemy Peredachi Informatsii* 7.3 (1971): 57-66.
- [DS05] R. Durrett and J. Schweinsberg. *A coalescent model for the effect of advantageous mutation on the genealogy of a population*. *Stochastic processes and their applications* 115.10 (2005): 1628-1657.
- [Dur88] R. Durrett. *Lecture notes on particle systems and percolation*. Brooks/Cole Pub Co, 1988.
- [EK86] S. Ethier and T. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.

- [EPW06] A. Etheridge, P. Pfaffelhuber, A. Wakolbinger. *An approximate sampling formula under genetic hitchhiking*. The Annals of Applied Probability 16.2 (2006): 685-729.
- [Fea02] P. Fearnhead. *The Common Ancestor at a Nonneutral Locus*. Journal of Applied Probability, Vol. 39, No. 1 (Mar., 2002 Appendix of part II2), pp. 38-54
- [For84] O. Forster. *Analysis 2: Differentialrechnung im Rn-Gewöhnliche Differentialgleichungen*. Vieweg Studium, 1984.
- [GS01] G. Grimmett and D. Stirzaker. *Probability and random processes*. Oxford university press, 2001.
- [HP08] J. Hermisson and P. Pfaffelhuber. *The pattern of genetic hitchhiking under recurrent mutation*. Electronic Journal of Probability 13.68 (2008): 2069-2106.
- [HS99] P. Hsieh and Y. Sibuya. *Basic theory of ordinary differential equations*. Springer-Verlag, 1999
- [IS03] H. Innan and W. Stephan. *Distinguishing the hitchhiking and background selection models*. Genetics 165.4 (2003): 2307-2312.
- [JK14] S. Jansen and N. Kurt. *On the notion (s) of duality for Markov processes*. Probab. Surveys 11 (2014).
- [Kal06] O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media. (2006)
- [KHB13] S. Kluth, T. Hustedt, E. Baake. *The Common Ancestor Process Revisited*. Bulletin of mathematical biology (2013), 75(11), 2003-2027.
- [Lig85] T.M. Liggett. *Interacting Particle Systems*. Springer-Verlag, New York, 1985.
- [Lig99] T.M. Liggett. *Stochastic interacting systems: Contact, Voter and Exclusion processes*. Springer Science and Business Media, 1999.
- [MS98] M. Möhle and S. Sagitov. *A characterization of ancestral limit processes arising in haploid population genetics models*. Preprint. Johannes Gutenberg-Universität Mainz (1998).
- [PHW06] P. Pfaffelhuber, B. Haubold, A. Wakolbinger. *Approximate genealogies under genetic hitchhiking*. Genetics 174.4 (2006): 1995-2008.
- [PP13] C. Pokalyuk and P. Pfaffelhuber. *The ancestral selection graph under strong directional selection*. Theoretical population biology 87 (2013): 25-33.
- [PS07] P. Pfaffelhuber, and A. Studeny. *Approximating genealogies for partially linked neutral loci under a selective sweep*. Journal of mathematical biology 55.3 (2007): 299-330.
- [Rot64] G. Rota (1964). *The number of partitions of a set*. The American Mathematical Monthly, 71(5), 498-504.
- [Shi96] A.N. Shiryaev. *Probability*. Springer-Verlag, New York (1996).
- [Spi69] F. Spitzer. *Random processes defined through the interaction of an infinite particle system*. Probability and Information Theory 89 (1969): 201-223.

- [SS08] A. Sturm and J. Swart. *Voter models with heterozygosity selection*. The Annals of Applied Probability (2008): 59-99.
- [SS15] A. Sturm and J. Swart. *A particle system with cooperative branching and coalescence*. The Annals of Applied Probability 25.3 (2015): 1616-1649.
- [SS16] A. Sturm and J. Swart. *Pathwise duals of monotone and additive Markov processes*. Journal of Theoretical Probability (2016): 1-52.
- [SV79] D. Stroock and S. Varadhan. *Multidimensional Diffusion Processes*. Springer-Verlag, New York (1979).
- [Swa17] J. Swart. *A Course in Interacting Particle Systems*. Lecture notes (2017), ArXiv:1703.10007.
- [Tay07] J. Taylor. (2007). *The common ancestor process for a Wright-Fisher diffusion*. Electronic Journal of Probability 12 (2007): 808-847.

1989 Born in Pilsen

2008-2011

Bachelor degree, Mathematics

Charles University, Faculty of Mathematics and Physics, Prague

Subject Field: General Mathematics

2011-2013

Master's degree, Mathematics

Charles University, Faculty of Mathematics and Physics, Prague

Subject Field: Probability, Mathematical Statistics and Econometry

Concentration: Probability theory

2013-2017

Ph.D., Mathematics (advisor Prof. Dr. Anja Sturm)

Georg-August University Göttingen, Faculty of Mathematics and Computer Science,

Institute for Mathematical Stochastics, Göttingen