Recursive tree processes and the mean-field limit of stochastic flows

Tibor Mach [*]   Anja Sturm [†]   Jan M. Swart [*]

Abstract

Interacting particle systems can often be constructed from a graphical representation, by applying local maps at the times of associated Poisson processes. This leads to a natural coupling of systems started in different initial states. We consider interacting particle systems on the complete graph in the mean-field limit, i.e., as the number of vertices tends to infinity. We are not only interested in the mean-field limit of a single process, but mainly in how several coupled processes behave in the limit. This turns out to be closely related to recursive tree processes as studied by Aldous and Bandyopadyay in discrete time. We here develop an analogue theory for recursive tree processes in continuous time. We illustrate the abstract theory on an example of a particle system with cooperative branching. This yields an interesting new example of a recursive tree process that is not endogenous.

Contents

[*]The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod vodárenskou věží 4, 18208 Praha 8, Czech Republic; E-mail: swart@utia.cas.cz

[†]Institute for Mathematical Stochastics, Georg-August-Universität Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany; E-mail: asturm@math.uni-goettingen.de

1 Introduction and main results

1.1 Introduction

Let Ω and S be Polish spaces, let $\mathbf{r}$ be a finite measure on Ω with total mass $|\mathbf{r}| := \mathbf{r}(\Omega) > 0$, and let $\gamma : \Omega \times S^{\mathbb{N}_+} \to S$ be measurable, where $\mathbb{N}_+ := \{1, 2, \dots\}$. Let T be the operator acting on probability measures on S defined as

$$T(\mu) := \text{ the law of } \gamma[\boldsymbol{\omega}](X_1, X_2, \dots), \tag{1.1}$$

where $\boldsymbol{\omega}$ is an Ω-valued random variable with law $|\mathbf{r}|^{-1}\mathbf{r}$ and $(X_i)_{i \geq 1}$ are i.i.d. with law μ. We assume that there exists a measurable function $\kappa : \Omega \to \mathbb{N} \cup \{\infty\}$ such that

$$\gamma[\omega](x_1, x_2, \dots) = \gamma[\omega](x_1, \dots, x_{\kappa(\omega)}) \qquad (\omega \in \Omega, \ x \in S^{\mathbb{N}_+}) \tag{1.2}$$

depends only on the first $\kappa(\omega)$ coordinates. In some of our results, we can allow for the case that $\kappa(\omega) = \infty$ (i.e., $\gamma[\omega]$ depends on all its coordinates), but most of the time the assumption

$$\int_\Omega \mathbf{r}(\mathrm{d}\omega)\, \kappa(\omega) < \infty \tag{1.3}$$

will be in place. We will be interested in the differential equation

$$\tfrac{\partial}{\partial t}\mu_t = |\mathbf{r}|\big\{T(\mu_t) - \mu_t\big\} \qquad (t \geq 0). \tag{1.4}$$

In Theorem 1.1 below, we will prove existence and uniqueness of solutions to (1.4) under the assumption (1.3). Our interest in equation (1.4) stems from the fact that, as we will prove in Theorem 1.5 below, the *mean-field limits* of a large class of interacting particle systems are described by equations of the form (1.4). In view of this, we call (1.4) a *mean-field equation*. The analysis of this sort of equations is commonly the first step towards understanding a given interacting particle system. Some illustrative examples of mean-field equations in the literature are [DN97, (1.1)], [NP99, (1.2)], and [FL17, (4)].

In the special case that $\kappa(\omega) = 1$ for all $\omega \in \Omega$, we observe that $T(\mu) = \int_S \mu(\mathrm{d}x)K(x, \cdot)$, where K is the probability kernel on S defined as

$$K(x, A) := \mathbb{P}\big[\gamma[\boldsymbol{\omega}](x) \in A\big] \qquad (x \in S, \ A \subset S \text{ measurable}). \tag{1.5}$$

In view of this, if $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots$ are i.i.d. with law $|\mathbf{r}|^{-1}\mathbf{r}$ and X_0 has law μ, then setting $X(k) := \gamma[\boldsymbol{\omega}_k](X(k-1))$ $(k \geq 1)$ inductively defines a Markov chain with transition kernel K, such that $X(k)$ has law $T^k(\mu)$, where T^k denotes the k-th iterate of the map T. Also, (1.4) describes the forward evolution of a continuous-time Markov chain where random maps $\gamma[\omega]$ are applied with Poisson rate $\mathbf{r}(\mathrm{d}\omega)$. A representation of a probability kernel K in terms of a random map $\gamma[\boldsymbol{\omega}]$ as in (1.5) is called a *random mapping representation*.

More generally, when the function κ is not identically one, Aldous and Bandyopadhyay [AB05] have shown that the iterates T^k of the map T from (1.1) can be represented in terms of a *Finite Recursive Tree Process* (FRTP), which is a generalization of a discrete-time Markov chain where time has a tree-like structure. More precisely, they construct a finite tree of depth k where the state of each internal vertex is a random function of the states of its offspring. If the states of the leaves are i.i.d. with law μ, they show that the state at the root has law $T^k(\mu)$. They are especially interested in fixed points of T, which generalize the concept of an invariant law of a Markov chain. They show that each such fixed point ν gives rise to a *Recursive Tree Process* (RTP), which is a process on an infinite tree where the state of each vertex has law ν. One can think of such an RTP as a generalization of a stationary backward Markov chain $(\ldots, X_{-2}, X_{-1}, X_0)$. A fixed point equation of the form $T(\nu) = \nu$ is called a *Recursive Distributional Equation* (RDE). Studying RDEs and their solutions is of independent interest as they appear naturally in many applications, see for example [AB05, Als12].

In the present paper, we develop an analogue theory in continuous time, generalizing the concept of a continuous-time Markov chain to chains where time has a tree-like structure. Let $(T_t)_{t \geq 0}$ be the semigroup defined by

$$T_t(\mu) := \mu_t \quad \text{where } (\mu_t)_{t \geq 0} \text{ solves (1.4) with } \mu_0 = \mu. \tag{1.6}$$

In Theorem 1.6, we show that T_t has a representation similar to (1.1), namely

$$T_t(\mu) = \text{ the law of } G_t\big((X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}\big), \tag{1.7}$$

where $\mathbb{T}$ is a countable set, $G_t : S^{\mathbb{T}} \to S$ is a random map, and the $(X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ are i.i.d. with law μ and independent of G_t. Similar to what we have in (1.2), the map G_t does not depend on all coordinates in $\mathbb{T}$ but only on a finite subcollection $(X_{\mathbf{i}})_{\mathbf{i} \in \nabla \mathbb{S}_t}$. Here $(\nabla \mathbb{S}_t)_{t \geq 0}$ turns out to be a branching process and condition (1.3) (which is not needed in the discrete-time theory) guarantees that the offspring distribution of this branching process has finite mean. Similarly to (1.5), we can view (1.7) as a random mapping representation of the operator in (1.6).

We note that in the RTPs studied by Aldous and Bandyopadhyay [AB05], the underlying tree is always deterministic. In the continuous-time setting, one has to work with

the random subtree $\mathbb{S}$ defined in (1.43) below, which is the family tree of a branching process. RTPs on random trees generated by branching processes have recently also been studied in [MS19, JPS20, BDF20].

As we have already mentioned, in Theorem 1.5 below, we prove that the mean-field limits of a large class of interacting particle systems are described by equations of the form (1.4). These interacting particle systems are constructed by applying local maps at the times of associated Poisson processes, which are introduced in detail in Section 1.3.

We are not only interested in the mean-field limit of a single process, but mainly in the mean-field limit of n coupled processes that are constructed from the same Poisson processes. For each $n \geq 1$, a measurable map $g : S^k \to S$ gives rise to n-*variate map* $g^{(n)} : (S^n)^k \to S^n$ defined as

$$g^{(n)}\big(x_1, \dots, x_k\big) = g^{(n)}\big(x^1, \dots, x^n\big) := \big(g(x^1), \dots, g(x^n)\big) \qquad (x^1, \dots, x^n \in S^k), \qquad (1.8)$$

where we denote an element of $(S^n)^k$ as $(x_i^m)_{i=1,\dots,k}^{m=1,\dots,n}$ with $x_i = (x_i^1, \dots, x_i^n)$ and $x^m = (x_1^m, \dots, x_k^m)$. Note that since $(S^n)^k \cong (S^k)^n$, we can view $g^{(n)}$ either as a function of $(x_1, \dots, x_k)$ or $(x^1, \dots, x^n)$. Let $\mathcal{P}(S)$ denote the space of probability measures on a space S. Letting $\gamma^{(n)}[\omega]$ denote the n-variate map associated with $\gamma[\omega]$ then, in analogy to (1.1),

$$T^{(n)}(\mu) := \text{ the law of } \gamma^{(n)}[\boldsymbol{\omega}](X_1, \dots, X_{\kappa(\boldsymbol{\omega})}), \qquad (1.9)$$

defines an n-*variate map* $T^{(n)} : \mathcal{P}(S^n) \to \mathcal{P}(S^n)$, which as in (1.4) gives rise to an n-*variate mean-field equation*, which describes the mean-field limit of n coupled processes.

If X is an S-valued random variable whose law $\nu := \mathbb{P}[X \in \cdot]$ is a fixed point of T, then $\overline{\nu}^{(n)} := \mathbb{P}[(X, \dots, X) \in \cdot]$ is a fixed point of $T^{(n)}$ that describes n perfectly coupled processes. We will be interested in the stability (or instability) of $\overline{\nu}^{(n)}$ under the n-variate mean-field equation. In other words, for our mean-field interacting particle systems, we fix the Poisson processes used in the construction and want to know if small changes in the initial state lead to small (or large) changes in the final state. Aldous and Bandyopadhyay [AB05] define an RTP to be *endogenous* if the state at the root is a measurable function of the random maps attached to all vertices of the tree. They showed, in some precise sense (see Theorem 1.11 below), that endogeny is equivalent to stability of $\overline{\nu}^{(n)}$. In Theorem 1.12, we generalize their result to the continuous-time setting.

The n-variate map $T^{(n)}$ is well-defined even for $n = \infty$, and $T^{(\infty)}$ maps the space of all exchangeable probability laws on $S^{\mathbb{N}_+}$ into itself. Let ξ be a $\mathcal{P}(S)$-valued random variable with law $\rho \in \mathcal{P}(\mathcal{P}(S))$, and conditional on ξ, let $(X^m)^{m=1,2,\cdots}$ be i.i.d. with common law ξ. Then the unconditional law of $(X^m)^{m=1,2,\cdots}$ is exchangeable, and by De Finetti, each exchangeable law on $S^{\mathbb{N}_+}$ is of this form. In view of this, $T^{(\infty)}$ naturally gives rise to a map $\check{T} : \mathcal{P}(\mathcal{P}(S)) \to \mathcal{P}(\mathcal{P}(S))$ which is the *higher-level map* defined in [MSS18], and which analogously to (1.4) gives rise to a *higher-level mean-field equation*. For any $\nu \in \mathcal{P}(S)$, let $\mathcal{P}(\mathcal{P}(S))_\nu$ denote the set of all $\rho \in \mathcal{P}(\mathcal{P}(S))$ with mean $\int \rho(\mathrm{d}\mu)\mu = \nu$. In [MSS18] it is shown that if ν is a fixed point of T, then the corresponding higher-level map $\check{T}$ has two fixed points $\underline{\nu}$ and $\overline{\nu}$ in $\mathcal{P}(\mathcal{P}(S))_\nu$ that are minimal and maximal with respect to the *convex order*, defined in Theorem 1.15 below. Moreover, $\underline{\nu} = \overline{\nu}$ if and only if the RTP corresponding to ν is endogenous.

We will apply the theory developed here as well as in [MSS18] to the higher-level mean-field equation for a particular interacting particle system with cooperative branching and deaths; see also [SS15, Mac17, BCH18] for several different variants of the model. To formulate this properly, it is useful to introduce some more general notation. Recall that for each $\omega \in \Omega$, $\gamma[\omega]$ is a map from $S^{\kappa(\omega)}$ into S. We let

$$\mathcal{G} := \big\{ \gamma[\omega] : \omega \in \Omega \big\} \qquad (1.10)$$

denote the set of all maps that can be obtained by varying ω. Here, elements of $\mathcal{G}$ are measurable maps $g : S^k \to S$ where $k = k_g \in \mathbb{N} \cup \{\infty\}$ may depend on g. If $k = 0$, then S^0 is defined to be a set with just one element, which we denote by $\varnothing$ (the empty sequence, which we distinguish notationally from the empty set $\emptyset$). We equip $\mathcal{G}$ with the final σ-field for the map $\omega \mapsto \gamma[\omega]$ (i.e., the strongest σ-field on $\mathcal{G}$ for which this map is measurable) and let π denote the image of the measure $\mathbf{r}$ under this map. Then the mean-field equation (1.4) can be rewritten as

$$\frac{\partial}{\partial t}\mu_t = \int_{\mathcal{G}} \pi(\mathrm{d}g)\{T_g(\mu_t) - \mu_t\} \qquad (t \geq 0), \tag{1.11}$$

where for any measurable map $g : S^k \to S$,

$$T_g(\mu) := \text{ the law of } g(X_1, \ldots, X_k), \quad \text{where } (X_i)_{i=1,\ldots,k} \text{ are i.i.d. with law } \mu. \tag{1.12}$$

In the concrete example that we are interested in, $S := \{0, 1\}$ and $\mathcal{G} := \{\mathtt{cob}, \mathtt{dth}\}$ each have just two elements. Here $\mathtt{cob} : S^3 \to S$ and $\mathtt{dth} : S^0 \to S$ are maps defined as

$$\mathtt{cob}(x_1, x_2, x_3) := x_1 \vee (x_2 \wedge x_3) \quad \text{and} \quad \mathtt{dth}(\varnothing) := 0. \tag{1.13}$$

We choose

$$\pi(\{\mathtt{cob}\}) := \alpha \geq 0 \quad \text{and} \quad \pi(\{\mathtt{dth}\}) := 1. \tag{1.14}$$

Then the mean-field equation (1.11) takes the form

$$\frac{\partial}{\partial t}\mu_t = \alpha\{T_{\mathtt{cob}}(\mu_t) - \mu_t\} + \{T_{\mathtt{dth}}(\mu_t) - \mu_t\}. \tag{1.15}$$

which describes the mean-field limit of a particle system with cooperative branching (with rate α) and deaths (with rate 1). We will see that for $\alpha > 4$, (1.11) has two stable fixed points $\nu_{\mathrm{low}}, \nu_{\mathrm{upp}}$, and an unstable fixed point ν_{mid} that separates the domains of attraction of the stable fixed points.

In Theorem 1.18 below, we find all fixed points of the corresponding higher-level mean-field equation, and determine their domains of attraction. Note that solutions of the higher-level mean-field equation take values in the probability measures on $\mathcal{P}(\{0, 1\})$ $\cong [0, 1]$. As mentioned before, each fixed point ν of the original mean-field equation gives rise to two fixed points $\underline{\nu}, \overline{\nu}$ of the higher-level mean-field equation, which are minimal and maximal in $\mathcal{P}(\mathcal{P}(S))_\nu$ with respect to the convex order. Moreover, $\underline{\nu} = \overline{\nu}$ if and only if the RTP corresponding to ν is endogenous. In our example, we find that the stable fixed points $\nu_{\mathrm{low}}, \nu_{\mathrm{upp}}$ give rise to endogenous RTPs, but the RTP associated with ν_{mid} is not endogenous. The higher-level equation has no other fixed points in $\mathcal{P}(\mathcal{P}(S))_{\nu_{\mathrm{mid}}}$ except for $\underline{\nu}_{\mathrm{mid}}$ and $\overline{\nu}_{\mathrm{mid}}$, of which the former is stable and the latter unstable. Numerical data for the nontrivial fixed point $\underline{\nu}_{\mathrm{mid}}$ (viewed as a probability measure on $[0, 1]$) are plotted in Figure 2.

1.2 The mean-field equation

In this subsection, we collect some basic results about the mean-field equation (1.4) that form the basis for all that follows. We interpret (1.4) in the following sense: letting $\langle \mu, \phi \rangle := \int \phi \, \mathrm{d}\mu$, we say that a process $(\mu_t)_{t \geq 0}$ solves (1.4) if for each bounded measurable function $\phi : S \to \mathbb{R}$, the function $t \mapsto \langle \mu_t, \phi \rangle$ is continuously differentiable and

$$\frac{\partial}{\partial t}\langle \mu_t, \phi \rangle = \int_\Omega \mathbf{r}(\mathrm{d}\omega)\{\langle T_{\gamma[\omega]}(\mu_t), \phi \rangle - \langle \mu_t, \phi \rangle\} \qquad (t \geq 0). \tag{1.16}$$

Our first result gives sufficient conditions for existence and uniqueness of solutions to (1.4). We note that in view of Lemma 3.3 below, in order to check that a process $(\mu_t)_{t \geq 0}$ solves (1.4), it suffices to check (1.16) for a measure-determining class of functions ϕ in the sense defined below (3.13).

Theorem 1.1 (Mean-field equation). *Let S and Ω be Polish spaces, let $\mathbf{r}$ be a nonzero finite measure on Ω, and let $\gamma : \Omega \times S^{\mathbb{N}_+} \to S$ be measurable. Assume that there exists a measurable function $\kappa : \Omega \to \mathbb{N}$ such that (1.2) and (1.3) hold. Then the mean-field equation (1.4) has a unique solution $(\mu_t)_{t \geq 0}$ for each initial state $\mu_0 \in \mathcal{P}(S)$.*

Theorem 1.1 allows us to define a semigroup $(T_t)_{t \geq 0}$ of operators $T_t : \mathcal{P}(S) \to \mathcal{P}(S)$ as in (1.6). It is often useful to know that solutions to (1.4) are continuous as a function of their initial state. The following proposition gives continuity w.r.t. the total variation norm $\| \cdot \|$ and moreover shows that if the constant K from (1.18) is negative, then the operators $(T_t)_{t \geq 0}$ form a contraction semigroup.

Proposition 1.2 (Continuity in total variation norm). *Under the assumptions of Theorem 1.1, one has*

$$\big\| T_t(\mu) - T_t(\nu) \big\| \leq e^{Kt} \| \mu - \nu \| \qquad (\mu, \nu \in \mathcal{P}(S), \ t \geq 0), \tag{1.17}$$

where

$$K := \int_\Omega \mathbf{r}(\mathrm{d}\omega) \big(\kappa(\omega) - 1 \big). \tag{1.18}$$

Continuity w.r.t. weak convergence needs an additional assumption.

Proposition 1.3 (Continuity w.r.t. weak convergence). *Assume that*

$$\mathbf{r}\big(\{\omega : \kappa(\omega) = k, \ \gamma[\omega] \text{ is discontinuous at } x\}\big) = 0 \qquad (k \in \mathbb{N} \cup \{\infty\}, \ x \in S^k). \tag{1.19}$$

Then the operator T in (1.1) is continuous w.r.t. the topology of weak convergence. Under the additional assumption (1.2) and (1.3), the same is true for the operators T_t $(t \geq 0)$ in (1.6).

The condition (1.19) is considerably weaker than the condition that $\gamma[\omega]$ is continuous for all $\omega \in \Omega$. A simple example is $\Omega = S = [0, 1]$, $\mathbf{r}$ the Lebesgue measure, $\kappa \equiv 1$, and $\gamma[\omega](x) := 1_{\{x \geq \omega\}}$.

1.3 The mean-field limit

In this subsection, we show that equations of the form (1.4) arise as the mean-field limits of a large class of interacting particle systems. In order to be reasonably general, and in particular to allow for systems in which more than one site can change its value at the same time, we will introduce quite a bit of notation that will not be needed anywhere else in Section 1, so impatient readers can just glance at Theorem 1.5 and the discussion surrounding (1.36) and skip the rest of this subsection.

Let S be a Polish space as before, and let $N \in \mathbb{N}_+$. We will be interested in continuous-time Markov processes taking values in S^N, where N is large. Denoting an element of S^N by $x = (x_1, \ldots, x_N)$, we will focus on processes with a high degree of symmetry, in the sense that their dynamics are invariant under a permutation of the coordinates. It is instructive, though not necessary for what follows, to view $\{1, \ldots, N\}$ as the vertex set of a complete graph, where all vertices are neighbors of each other. The basic ingredients we will use to describe our processes are:

(i) a Polish space Ω' equipped with a finite nonzero measure $\mathbf{q}$,
(ii) a measurable function $\lambda : \Omega' \to \mathbb{N}_+$,

as well as, for each $\omega \in \Omega'$ and $1 \leq i \leq \lambda(\omega)$,

(iii) a function $\gamma_i[\omega] : S^{\lambda(\omega)} \to S$,
(iv) a finite set $K_i(\omega) \subset \{1, \ldots, \lambda(\omega)\}$ such that $\gamma_i[\omega](x_1, \ldots, x_{\lambda(\omega)}) = \gamma_i[\omega]\big((x_j)_{j \in K_i(\omega)}\big)$ depends only on the coordinates in $K_i(\omega)$.

Setting $\Omega_l' := \{\omega \in \Omega' : \lambda(\omega) = l\}$, we assume that the functions

$$\Omega_l' \times S^l \ni (\omega, x) \mapsto \gamma_i[\omega](x) \quad \text{and} \quad \Omega_l' \ni \omega \mapsto 1_{\{j \in K_i(\omega)\}} \quad \text{are measurable} \qquad (1.20)$$

for each $1 \le i, j \le l$. We let $\vec{\gamma}[\omega] : S^{\lambda(\omega)} \to S^{\lambda(\omega)}$ denote the function

$$\vec{\gamma}[\omega](x) := \big(\gamma_1[\omega](x), \ldots, \gamma_{\lambda(\omega)}[\omega](x) \big) \qquad (1.21)$$

and let $\kappa_i(\omega) := |K_i(\omega)|$ denote the cardinality of the set $K_i(\omega)$.

The space Ω', measure $\mathbf{q}$, and functions λ and $\vec{\gamma}$ play roles similar, but not quite identical to $\Omega, \mathbf{r}, \kappa$, and γ from Subsection 1. We can use $\Omega', \mathbf{q}, \lambda$, and $\vec{\gamma}$ to define the following mean-field equation:

$$\frac{\partial}{\partial t} \mu_t = \int_{\Omega'} \mathbf{q}(\mathrm{d}\omega) \sum_{i=1}^{\lambda(\omega)} \big\{ T_{\gamma_i[\omega]}(\mu_t) - \mu_t \big\}. \qquad (1.22)$$

The following lemma says that (1.22) is really a mean-field equation of the form we have already seen in (1.4). This is why in subsequent sections we will only work with equations of this form.

Lemma 1.4 (Simplified equation). *Assume that*

$$\text{(i)} \int_{\Omega'} \mathbf{q}(\mathrm{d}\omega)\, \lambda(\omega) < \infty \quad \text{and} \quad \text{(ii)} \int_{\Omega'} \mathbf{q}(\mathrm{d}\omega) \sum_{i=1}^{\lambda(\omega)} \kappa_i(\omega) < \infty. \qquad (1.23)$$

Then equation (1.22) can be cast in the simpler form (1.4) for a suitable choice of $\Omega, \mathbf{r}, \kappa$, and γ, where (1.23) (i) guarantees that $\mathbf{r}$ is a finite measure and (1.23) (ii) implies that (1.3) holds. If

$$\mathbf{q}\big(\{\omega : \lambda(\omega) = l,\ \gamma_i[\omega] \text{ is discontinuous at } x\}\big) = 0 \qquad (1 \le i \le l,\ x \in S^l), \qquad (1.24)$$

then moreover (1.19) can be satisfied.

We now use the ingredients $\Omega', \mathbf{q}, \lambda$, and $\vec{\gamma}$ to define the class of Markov processes we are interested in. We construct these processes by applying local maps, that affect only finitely many coordinates, at the times of associated Poisson processes. In the context of interacting particle systems, such constructions are called *graphical representations*.

For any $N \in \mathbb{N}_+$ we set $[N] := \{1, \ldots, N\}$. We let $[N]^{\langle l \rangle}$ denote the set of all sequences $\mathbf{i} = (i_1, \ldots, i_l)$ for which $i_1, \ldots, i_l \in [N]$ are all different. Note that $[N]^{\langle l \rangle}$ has $N^{\langle l \rangle} := N(N-1) \cdots (N-l+1)$ elements. We will consider Markov processes $X = (X(t))_{t \ge 0}$ with values in S^N that evolve in the following way:

(i) At the times of a Poisson process with intensity $|\mathbf{q}| := \mathbf{q}(\Omega')$, an element $\omega \in \Omega'$ is chosen according to the probability law $|\mathbf{q}|^{-1}\mathbf{q}$.

(ii) If $\lambda(\omega) > N$, nothing happens.

(iii) Otherwise, an element $\mathbf{i} \in [N]^{\langle \lambda(\omega) \rangle}$ is selected according to the uniform distribution on $[N]^{\langle \lambda(\omega) \rangle}$, and the previous values $\big(X_{i_1}(t-), \ldots, X_{i_{\lambda(\omega)}}(t-)\big)$ of X at the coordinates $i_1, \ldots, i_{\lambda(\omega)}$ are replaced by

$$\big(X_{i_1}(t), \ldots, X_{i_{\lambda(\omega)}}(t)\big) = \vec{\gamma}[\omega]\big(X_{i_1}(t-), \ldots, X_{i_{\lambda(\omega)}}(t-)\big).$$

More formally, we can construct our Markov process $X = (X(t))_{t \ge 0}$ as follows. For each $\omega \in \Omega'$ with $\lambda(\omega) \le N$, and for each $\mathbf{i} \in [N]^{\langle \lambda(\omega) \rangle}$, define a map $m_{\omega,\mathbf{i}} : S^N \to S^N$ by

$$m_{\omega,\mathbf{i}}(x)_j := \begin{cases} \gamma_j[\omega](x_{i_1}, \ldots, x_{i_{\lambda(\omega)}}) & \text{if } j \in \{i_1, \ldots, i_{\lambda(\omega)}\}, \\ x_j & \text{otherwise,} \end{cases} \qquad (x \in S^N). \qquad (1.25)$$

Let Π be a Poisson point set on

$$\big\{(\omega, \mathbf{i}, t) : \omega \in \Omega', \ \mathbf{i} \in [N]^{\langle\lambda(\omega)\rangle}, \ t \in \mathbb{R}\big\} \tag{1.26}$$

with intensity

$$\mathbf{q}(\mathrm{d}\omega)\frac{1_{\{\lambda(\omega)\leq N\}}}{N^{\langle\lambda(\omega)\rangle}}\,\mathrm{d}t. \tag{1.27}$$

Since $\mathbf{q}$ is a finite measure, the set $\Pi_{s,u} := \{(\omega, \mathbf{i}, t) \in \Pi : s < t \leq u\}$ is a.s. finite for each $-\infty < s \leq u < \infty$, so we can order its elements as

$$\Pi_{s,u} = \big\{(\omega_1, \mathbf{i}_1, t_1), \ldots, (\omega_n, \mathbf{i}_n, t_n)\big\} \quad \text{with} \quad t_1 < \cdots < t_n \tag{1.28}$$

and use this to define

$$\mathbf{X}_{s,u} = m_{\omega_n, \mathbf{i}_n} \circ \cdots \circ m_{\omega_1, \mathbf{i}_1} \tag{1.29}$$

In words, Π is a list of triples $(\omega, \mathbf{i}, t)$. Here ω represents some external input that tells us that we need to apply the map $\vec{\gamma}[\omega]$. The coordinates where and the time when this map needs to be applied are given by $\mathbf{i}$ and t, respectively. It is easy to see that the random maps $(\mathbf{X}_{s,u})_{s\leq u}$ form a *stochastic flow*, i.e.,

$$\mathbf{X}_{s,s} = 1 \quad \text{and} \quad \mathbf{X}_{t,u} \circ \mathbf{X}_{s,t} = \mathbf{X}_{s,u} \qquad (s \leq t \leq u), \tag{1.30}$$

where 1 denotes the identity map. Moreover $(\mathbf{X}_{s,u})_{s\leq u}$ has *independent increments* in the sense that

$$\mathbf{X}_{t_1,t_2}, \ldots, \mathbf{X}_{t_{k-1},t_k} \quad \text{are independent} \tag{1.31}$$

for each $t_1 < \cdots < t_k$. It is well-known (see, e.g., [SS18, Lemma 1]) that if $X(0)$ is an S^N-valued random variable, independent of the Poisson set Π, then setting

$$X(t) := \mathbf{X}_{0,t}\big(X(0)\big) \qquad (t \geq 0) \tag{1.32}$$

defines a Markov process $X = (X(t))_{t\geq 0}$ with values in S^N. Note that $(X(t))_{t\geq 0}$ has piecewise constant sample paths, which are right-continuous because of the way we have defined $\Pi_{s,u}$.

We now formulate our result about the mean-field limit of Markov processes as defined in (1.32). For any $x \in S^N$, we define an *empirical measure* $\mu\{x\}$ on S by

$$\mu\{x\} := \frac{1}{N}\sum_{i\in[N]}\delta_{x_i}. \tag{1.33}$$

Below, $\mu^{\otimes n} := \mu \otimes \cdots \otimes \mu$ denotes the product measure of n copies of μ. The expectation $\mathbb{E}[\mu]$ of a random measure μ on a Polish space S is defined in the usual way, i.e., $\mathbb{E}[\mu]$ is the deterministic measure defined by $\int\phi\,\mathrm{d}\mathbb{E}[\mu] := \mathbb{E}[\int\phi\,\mathrm{d}\mu]$ for any bounded measurable $\phi : S \to \mathbb{R}$.

Theorem 1.5 (Mean-field limit). *Let S be a Polish space, let Ω', $\mathbf{q}$, λ, and $\vec{\gamma}$ be as above, and assume (1.23). For each $N \in \mathbb{N}_+$, let $(X^{(N)}(t))_{t\geq 0}$ be Markov processes with state space S^N as defined in (1.32), and let $\mu_t^N := \mu\{X^{(N)}(t)\}$ denote their associated empirical measures. Let d be any metric on $\mathcal{P}(S)$ that generates the topology of weak convergence. Fix some (deterministic) $\mu_0 \in \mathcal{P}(S)$ and assume that (at least) one of the following two conditions is satisfied.*

(i) $\mathbb{P}\big[d(\mu_0^N, \mu_0) \geq \varepsilon\big] \underset{N\to\infty}{\longrightarrow} 0$ *for all $\varepsilon > 0$, and (1.24) holds.*

(ii) $\big\|\mathbb{E}[(\mu_0^N)^{\otimes n}] - \mu_0^{\otimes n}\big\| \underset{N\to\infty}{\longrightarrow} 0$ *for all $n \geq 1$, where $\|\cdot\|$ denotes the total variation norm.*

Then

$$\mathbb{P}\Big[\sup_{0\le t\le T} d(\mu^N_{Nt}, \mu_t) \ge \varepsilon\Big] \underset{N\to\infty}{\longrightarrow} 0 \qquad (\varepsilon > 0,\ T < \infty), \tag{1.34}$$

where $(\mu_t)_{t\ge0}$ is the unique solution to the mean-field equation (1.22) with initial state μ_0.

Condition (ii) is in particular satisfied if $X^N_1(0), \dots, X^N_N(0)$ are i.i.d. with common law μ_0. Note that in (1.34), we rescale time by a factor N.

It is instructive to demonstrate the general set-up on our concrete example of a particle system with cooperative branching and deaths. As before, we have $S = \{0, 1\}$. We choose for Ω' a set with just two elements, say $\Omega' = \{1, 2\}$, and we set $\mathbf{q}(\{1\}) := \alpha \ge 0$ and $\mathbf{q}(\{2\}) := 1$. We let $\lambda(1) := 3$, $\lambda(2) := 1$, and define $\vec{\gamma}[1] : S^3 \to S^3$ and $\vec{\gamma}[2] : S \to S$ by

$$\vec{\gamma}[1](x_1, x_2, x_3) := \big(x_1 \vee (x_2 \wedge x_3), x_2, x_3\big) \quad \text{and} \quad \vec{\gamma}[2](x_1) := 0. \tag{1.35}$$

Then the particle system in (1.32) has the following description. Let us say that a site i is occupied at time t if $X_i(t) = 1$. Then, with rate α, three sites $(i_1, i_2, i_3) \in [N]^{\langle 3 \rangle}$ are selected at random. If the sites i_2 and i_3 are both occupied, then the particles at these sites cooperate to produce a third particle at i_1, provided this site is empty. In addition, with rate 1, a site i is selected at random, and any particle that is present there dies.

It is not hard to see that for our choice of $\Omega', \mathbf{q}, \lambda$, and $\vec{\gamma}$, the mean-field equation (1.22) simplifies to (1.15), Note that since $\gamma_2[1]$ and $\gamma_3[1]$ are the identity map, they drop out of (1.22), so only $\gamma_1[1] = \mathtt{cob}$ and $\gamma_1[2] = \mathtt{dth}$ remain. Since $\gamma_1[2](x_1) = 0$ regardless of the value of x_1, we can choose for $K_1(1)$ the empty set and view $\gamma_1[2] = \mathtt{dth}$ as a function $\mathtt{dth} : S^0 \to S$.

Solutions of (1.15) take values in the probability measures on $S = \{0, 1\}$, which are uniquely characterized by their value at 1. Rewriting (1.15) in terms of $p_t := \mu_t(\{1\})$ yields the equation

$$\tfrac{\partial}{\partial t} p_t = \alpha p_t^2 (1 - p_t) - p_t \qquad (t \ge 0). \tag{1.36}$$

This equation can also be found in [Nob92, (1.11)], [Neu94, (1.2)], [BW97, (3.1)], [FL17, (4)], and [BCH18, (2.1)]. It is not hard to check that for $\alpha < 4$, the only fixed point of (1.36) is $z_{\mathrm{low}} := 0$, while for $\alpha \ge 4$, there are additional fixed points

$$z_{\mathrm{mid}} := \tfrac{1}{2} - \sqrt{\tfrac{1}{4} - \tfrac{1}{\alpha}} \quad \text{and} \quad z_{\mathrm{upp}} := \tfrac{1}{2} + \sqrt{\tfrac{1}{4} - \tfrac{1}{\alpha}}. \tag{1.37}$$

If $\alpha < 4$, then solutions to (1.36) converge to z_{low} regardless of the initial state. On the other hand, for $\alpha \ge 4$, solutions to (1.36) with $p_0 > z_{\mathrm{mid}}$ converge to the upper fixed point z_{upp} while solutions to (1.36) with $p_0 < z_{\mathrm{mid}}$ converge to the lower fixed point z_{low}. In particular, if $\alpha > 4$, then z_{low} and z_{upp} are stable fixed points while z_{mid} is an unstable fixed point separating the domains of attraction of z_{low} and z_{upp}.

1.4 A recursive tree representation

In this subsection we formally introduce Finite Recursive Tree Processes (FRTPs) and state the random mapping representation of solutions to the mean-field equation (1.4) anticipated in (1.7).

For $d \in \mathbb{N}_+$, let $\mathbb{T}^d$ denote the space of all finite words $\mathbf{i} = i_1 \cdots i_n$ $(n \in \mathbb{N})$ made up from the alphabet $\{1, \dots, d\}$, and define $\mathbb{T}^\infty$ similarly, using the alphabet $\mathbb{N}_+$. If $\mathbf{i}, \mathbf{j} \in \mathbb{T}^d$ with $\mathbf{i} = i_1 \cdots i_m$ and $\mathbf{j} = j_1 \cdots j_n$, then we define the concatenation $\mathbf{ij} \in \mathbb{T}^d$ by $\mathbf{ij} := i_1 \cdots i_m j_1 \cdots j_n$. We denote the length of a word $\mathbf{i} = i_1 \cdots i_n$ by $|\mathbf{i}| := n$ and let $\varnothing$ denote the word of length zero. We view $\mathbb{T}^d$ as a tree with root $\varnothing$, where each vertex $\mathbf{i} \in \mathbb{T}^d$ has d children $\mathbf{i}1, \mathbf{i}2, \dots$, and each vertex $\mathbf{i} = i_1 \cdots i_n$ except the root has precisely one ancestor $\overleftarrow{\mathbf{i}} := i_1 \cdots i_{n-1}$. For each *rooted subtree* of $\mathbb{T}^d$, i.e., a subtree $\mathbb{U} \subset \mathbb{T}^d$ that contains $\varnothing$, we let $\partial\mathbb{U} := \{\mathbf{i} \in \mathbb{T}^d : \overleftarrow{\mathbf{i}} \in \mathbb{U},\ \mathbf{i} \notin \mathbb{U}\}$ denote the *boundary* of $\mathbb{U}$ relative to $\mathbb{T}^d$.

We write

$$\mathbb{T}^d_{(n)} := \{\mathbf{i} \in \mathbb{T}^d : |\mathbf{i}| < n\} \quad \text{and} \quad \partial\mathbb{T}^d_{(n)} = \{\mathbf{i} \in \mathbb{T}^d : |\mathbf{i}| = n\} \qquad (n \geq 1), \qquad (1.38)$$

and use the convention $\partial\emptyset := \{\varnothing\}$, so that (1.38) holds also for $n = 0$.

We return to the set-up of Subsection 1.1, i.e., S and Ω are Polish spaces, $\mathbf{r}$ is a nonzero finite measure on Ω, and $\gamma : \Omega \times S^{\mathbb{N}_+} \to S$ and $\kappa : \Omega \to \mathbb{N}$ are measurable functions such that (1.2) holds. We fix some $d \in \mathbb{N}_+ \cup \{\infty\}$ such that $\kappa(\omega) \leq d$ for all $\omega \in \Omega$ and set $\mathbb{T} := \mathbb{T}^d$. Let $(\boldsymbol{\omega}_\mathbf{i})_{\mathbf{i}\in\mathbb{T}}$ be an i.i.d. collection of Ω-valued r.v.'s with common law $|\mathbf{r}|^{-1}\mathbf{r}$. Fix $n \geq 1$ and assume that

   (i)   the $(X_\mathbf{i})_{\mathbf{i}\in\partial\mathbb{T}_{(n)}}$ are i.i.d. with common law μ and independent of $(\boldsymbol{\omega}_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_{(n)}}$,

   (ii) $X_\mathbf{i} := \gamma[\boldsymbol{\omega}_\mathbf{i}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\boldsymbol{\omega}_\mathbf{i})}) \qquad (\mathbf{i} \in \mathbb{T}_{(n)})$.

$$(1.39)$$

Then it is easy to see that the law of $X_\varnothing$ is given by $T^n(\mu)$, where T^n is the n-th iterate of the operator in (1.1). We call the collection of random variables

$$\big((\boldsymbol{\omega}_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_{(n)}}, (X_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_{(n)}\cup\partial\mathbb{T}_{(n)}}\big) \qquad (1.40)$$

a *Finite Recursive Tree Process* (FRTP). We can think of $(X_\mathbf{i})_{\mathbf{i}\in\mathbb{T}_{(n)}\cup\partial\mathbb{T}_{(n)}}$ as a generalization of a Markov chain, where time has a tree-like structure.

We now aim to give a similar representation of the semigroup $(T_t)_{t\geq 0}$ from (1.6). To do this, we let $(\sigma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}}$ be i.i.d. exponentially distributed random variables with mean $|\mathbf{r}|^{-1}$. We interpret $\sigma_\mathbf{i}$ as the lifetime of the individual with index $\mathbf{i}$ and let

$$\tau^*_\mathbf{i} := \sum_{m=1}^{n-1} \sigma_{i_1\cdots i_m} \quad \text{and} \quad \tau^\dagger_\mathbf{i} := \tau^*_\mathbf{i} + \sigma_\mathbf{i} \qquad (\mathbf{i} = i_1 \cdots i_n). \qquad (1.41)$$

denote the times when the individual $\mathbf{i}$ is born and dies, respectively. Then

$$\mathbb{T}_t := \big\{\mathbf{i} \in \mathbb{T} : \tau^\dagger_\mathbf{i} \leq t\big\} \quad \text{and} \quad \partial\mathbb{T}_t = \big\{\mathbf{i} \in \mathbb{T} : \tau^*_\mathbf{i} \leq t < \tau^\dagger_\mathbf{i}\big\} \qquad (t \geq 0) \qquad (1.42)$$

are the random subtrees of $\mathbb{T}$ consisting of all individuals that have died before time t and its boundary, which consists of all individuals that are alive at time t. If the function κ from (1.2) is bounded, then we can choose $\mathbb{T} := \mathbb{T}^d$ with $d < \infty$. Now it is easy to check that $(\partial\mathbb{T}_t)_{t\geq 0}$ is a continuous-time branching process where each particle is with rate $|\mathbf{r}|$ replaced by d new particles. In particular, $\mathbb{T}_t$ is a.s. finite for each $t > 0$. On the other hand, when κ is unbounded, we need to choose $\mathbb{T} := \mathbb{T}^\infty$, and this has the consequence that $\mathbb{T}_t$ is a.s. infinite for each $t > 0$. Nevertheless, under the assumption (1.3), it turns out that only a finite subtree of $\mathbb{T}_t$ is relevant for the state at the root $X_\varnothing$, as we explain now.

Let $\mathbb{S}$ be the random subtree of $\mathbb{T}$ defined as

$$\mathbb{S} := \big\{i_1 \cdots i_n \in \mathbb{T} : i_m \leq \kappa(\boldsymbol{\omega}_{i_1\cdots i_{m-1}}) \,\forall 1 \leq m \leq n\big\}, \qquad (1.43)$$

and for each subtree $\mathbb{U} \subset \mathbb{S}$, let $\nabla\mathbb{U} := \{\mathbf{i} \in \mathbb{S} : \overleftarrow{\mathbf{i}} \in \mathbb{U}, \mathbf{i} \notin \mathbb{U}\}$ denote the outer boundary of $\mathbb{U}$ relative to $\mathbb{S}$, where again we use the convention that $\nabla\mathbb{U} := \{\varnothing\}$ if $\mathbb{U}$ is the empty set. Then, under condition (1.3),

$$\mathbb{S}_t := \mathbb{T}_t \cap \mathbb{S} \quad \text{and} \quad \nabla\mathbb{S}_t = \big\{\mathbf{i} \in \mathbb{S} : \tau^*_\mathbf{i} \leq t < \tau^\dagger_\mathbf{i}\big\} \qquad (t \geq 0) \qquad (1.44)$$

are a.s. finite for all $t \geq 0$. Indeed, $(\nabla\mathbb{S}_t)_{t\geq 0}$ is a branching process where for each individual $\mathbf{i}$, with Poisson rate $\mathbf{r}(d\omega)$, an element $\omega \in \Omega$ is selected and $\mathbf{i}$ is replaced by new individuals $\mathbf{i}1, \ldots, \mathbf{i}\kappa(\omega)$. The condition on the rates (1.3) guarantees that this branching process has finite mean and in particular does not explode, so that $\mathbb{S}_t$ is a.s. a finite subtree of $\mathbb{S}$.
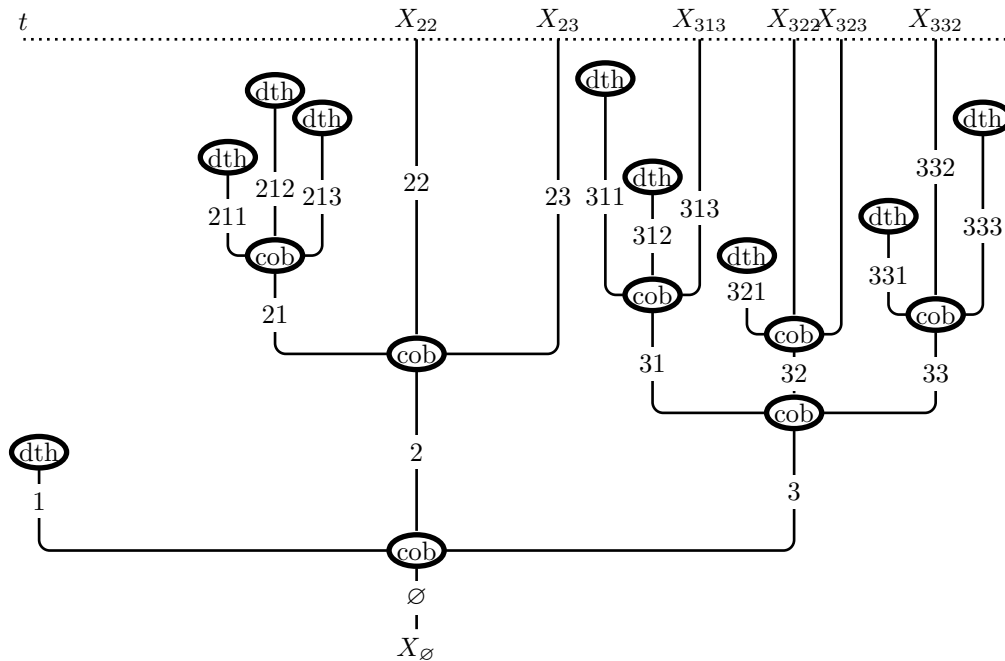
Figure 1: A particular realization of the branching process $\nabla\mathbb{S}_t$ for a system with cooperative branching and deaths. The random map $G_{\mathbb{S}_t}$ is the concatenation of random maps attached to the vertices of the family tree $\mathbb{S}_t$ of the individuals alive at time t. In this example, $\nabla\mathbb{S}_t = \{22, 23, 313, 322, 323, 332\}$ and the maps cob and dth are as defined in (1.13).

Let $(\boldsymbol{\omega_i})_{\mathbf{i}\in\mathbb{T}}$ be i.i.d. with common law $|\mathbf{r}|^{-1}\mathbf{r}$, independent of the lifetimes $(\sigma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}}$. For any finite rooted subtree $\mathbb{U} \subset \mathbb{S}$ and for each $(x_\mathbf{i})_{\mathbf{i}\in\nabla\mathbb{U}} = x \in S^{\nabla\mathbb{U}}$, we can inductively define $x_\mathbf{i}$ for $\mathbf{i} \in \mathbb{U}$ by

$$x_\mathbf{i} := \gamma[\boldsymbol{\omega_i}](x_{\mathbf{i}1}, \ldots, x_{\mathbf{i}\kappa(\boldsymbol{\omega_i})}) \qquad (\mathbf{i} \in \mathbb{U}). \tag{1.45}$$

Then the value $x_\varnothing$ we obtain at the root is a function of $(x_\mathbf{i})_{\mathbf{i}\in\nabla\mathbb{U}}$. Let us denote this function by $G_\mathbb{U} : S^{\nabla\mathbb{U}} \to S$, i.e.,

$$G_\mathbb{U}\big((x_\mathbf{i})_{\mathbf{i}\in\nabla\mathbb{U}}\big) := x_\varnothing, \quad \text{with } (x_\mathbf{i})_{\mathbf{i}\in\mathbb{U}} \text{ defined as in (1.45).} \tag{1.46}$$

We can think of $G_\mathbb{U}$ as the "concatenation" of the maps $(\gamma[\boldsymbol{\omega_i}])_{\mathbf{i}\in\mathbb{U}}$. We will in particular be interested in the random maps

$$G_t := G_{\mathbb{S}_t} \qquad (t \geq 0) \tag{1.47}$$

with $\mathbb{S}_t$ as in (1.44). For our running example of a system with cooperative branching and deaths, these definitions are illustrated in Figure 1.

Let $(\mathcal{F}_t)_{t\geq 0}$, defined as

$$\mathcal{F}_t := \sigma\big(\nabla\mathbb{S}_t, (\boldsymbol{\omega_i}, \sigma_\mathbf{i})_{\mathbf{i}\in\mathbb{S}_t}\big) \qquad (t \geq 0). \tag{1.48}$$

be the natural filtration associated with our evolving marked tree, that contains information about which individuals are alive at time t, as well as the random elements $\boldsymbol{\omega_i}$ and lifetimes $\sigma_\mathbf{i}$ associated with all individuals that have died by time t. In particular, G_t is measurable w.r.t. $\mathcal{F}_t$. The following theorem is a precise formulation of the random mapping representation of solutions of the mean-field equation (1.4), anticipated in (1.7).

Theorem 1.6 (Recursive tree representation). *Let S and Ω be Polish spaces, let $\mathbf{r}$ be a nonzero finite measure on Ω, and let $\gamma : \Omega \times S^{\mathbb{N}_+} \to S$ and $\kappa : \Omega \to \mathbb{N}$ be measurable functions satisfying (1.2) and (1.3). Let $(\boldsymbol{\omega}_\mathbf{i})_{\mathbf{i} \in \mathbb{S}}$ be i.i.d. with common law $|\mathbf{r}|^{-1}\mathbf{r}$ and let $(\sigma_\mathbf{i})_{\mathbf{i} \in \mathbb{S}}$ be an independent i.i.d. collection of exponentially distributed random variables with mean $|\mathbf{r}|^{-1}$. Fix $t \geq 0$ and let G_t and $\mathcal{F}_t$ be defined as in (1.47) and (1.48). Conditional on $\mathcal{F}_t$, let $(X_\mathbf{i})_{\mathbf{i} \in \nabla \mathbb{S}_t}$ be i.i.d. S-valued random variables with common law μ. Then*

$$T_t(\mu) = \text{ the law of } G_t\big((X_\mathbf{i})_{\mathbf{i} \in \nabla \mathbb{S}_t}\big), \tag{1.49}$$

where T_t is defined in (1.6).

Recalling the definition of G_t, we can also formulate Theorem 1.6 as follows. With $(\boldsymbol{\omega}_\mathbf{i}, \sigma_\mathbf{i})_{\mathbf{i} \in \mathbb{S}}$ as above, fix $t > 0$ and let $(X_\mathbf{i})_{\mathbf{i} \in \mathbb{S}_t \cup \nabla \mathbb{S}_t}$ be random variables such that

(i) Conditional on $\mathcal{F}_t$, the r.v.'s $(X_\mathbf{i})_{\mathbf{i} \in \nabla \mathbb{S}_t}$ are i.i.d. with common law μ,

(ii) $X_\mathbf{i} = \gamma[\boldsymbol{\omega}_\mathbf{i}]\big(X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\boldsymbol{\omega}_\mathbf{i})}\big) \qquad (\mathbf{i} \in \mathbb{S}_t)$. $\tag{1.50}$

Then (1.49) says that the state at the root $X_\varnothing$ has law $T_t(\mu)$. This is a continuous-time analogue of the FRTP (1.39).

In our proofs, we will first prove Theorem 1.6 and then use this to prove Theorem 1.5 about the mean-field limit of interacting particle systems. Recall that these particle systems are constructed from a stochastic flow $(\mathbf{X}_{s,t})_{s \leq t}$ as in (1.32). To find the empirical measure of $X(t) = \mathbf{X}_{0,t}(X(0))$, we pick a site $i \in [N]$ at random and ask for its type $X_i(t)$ which via $\mathbf{X}_{0,t}$ is a function of the initial state $X(0)$. When N is large, $X_i(t)$ does not depend on all coordinates $(X_j(0))_{j \in [N]}$ but only on a random subset of them, and indeed one can show that the map that gives $X_i(t)$ as a function of these coordinates approximates the map G_t from Theorem 1.6, in an appropriate sense. The heuristics behind this are explained in some more detail in Subsection 4.1 below.

Remark 1.7. Another way to write (1.49) is

$$\mu_t = \mathbb{E}\big[T_{G_t}(\mu_0)\big] \qquad (t \geq 0), \tag{1.51}$$

where T_{G_t} is defined as in (1.12) for the random map G_t and $(\mu_t)_{t \geq 0}$ is a solution to (1.4). One can check that $(\nabla \mathbb{S}_t, G_t)_{t \geq 0}$ is a Markov process. Let us informally denote this process by $(G_t)_{t \geq 0}$ and its state space by $\mathcal{G}$. Then equation (1.49) can be understood as a (generalized) duality relationship between $(G_t)_{t \geq 0}$ and $(\mu_t)_{t \geq 0}$ with (generalized) duality function $H : \mathcal{G} \times \mathcal{P}(S) \to \mathcal{P}(S)$ given by

$$H(g, \mu) := T_g(\mu). \tag{1.52}$$

With this definition, using the fact that G_0 is the identity map, (1.51) reads

$$\mu_t = H(G_0, \mu_t) = \mathbb{E}[H(G_t, \mu_0)] \tag{1.53}$$

and we can obtain a family of usual (real-valued) dualities by integrating against a test function ϕ.

1.5 Recursive tree processes

Recall the definition of the operator T in (1.1) and the semigroup $(T_t)_{t \geq 0}$ in (1.6). It is clear from (1.4) that for a measure $\nu \in \mathcal{P}(S)$, the following two conditions are equivalent:

(i) $T_t(\nu) = \nu$ $(t \geq 0)$ and   (ii) $T(\nu) = \nu$. $\tag{1.54}$

We call such a measure ν a *fixed point* of the mean-field equation (1.4). Condition (ii) is equivalent to saying that a random variable X with law ν satisfies

$$X \overset{\mathrm{d}}{=} \gamma[\omega](X_1, \ldots, X_{\kappa(\omega)}), \tag{1.55}$$

where $\overset{\mathrm{d}}{=}$ denotes equality in distribution, $X_1, X_2, \ldots$ are i.i.d. copies of X, and ω is an independent Ω-valued random variable with law $|\mathbf{r}|^{-1}\mathbf{r}$. Equations of this type are called *Recursive Distributional Equations* (RDEs).

FRTPs as in (1.39) are consistent in the sense that if $(X_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(n)}}$ are as in (1.39), then for any $1 \leq m \leq n$,

(i) the $(X_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{T}_{(m)}}$ are i.i.d. with common law $T^{n-m}(\mu)$
   and independent of $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}_{(m)}}$, $\qquad\qquad$ (1.56)

(ii) $X_{\mathbf{i}} := \gamma[\boldsymbol{\omega}_{\mathbf{i}}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\boldsymbol{\omega}_{\mathbf{i}})}) \qquad (\mathbf{i} \in \mathbb{T}_{(m)})$.

The following lemma states a similar consistency property in the continuous-time setting.

Lemma 1.8 (Consistency). *Fix $t > 0$ and let $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{S}_t\cup\nabla\mathbb{S}_t}$ be as in (1.50). Then, for each $s \in [0, t]$,*

(i) *conditional on $\mathcal{F}_s$, the r.v.'s $(X_{\mathbf{i}})_{\mathbf{i}\in\nabla\mathbb{S}_s}$ are i.i.d. with common law $T_{t-s}(\mu)$,*

(ii) $X_{\mathbf{i}} = \gamma[\boldsymbol{\omega}_{\mathbf{i}}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\boldsymbol{\omega}_{\mathbf{i}})}) \qquad (\mathbf{i} \in \mathbb{S}_s),$

(1.57)

where $(T_t)_{t\geq 0}$ is defined in (1.6).

Using the consistency relation (1.56) and Kolmogorov's extension theorem, it is not hard to see that if ν solves the RDE (1.54), then it is possible to define a stationary recursive process on an infinite tree such that each vertex has law ν. This was already observed in [AB05]. The following lemma is a slight reformulation of their observation.

Lemma 1.9 (Recursive Tree Process). *Let ν be a solution to the RDE (1.54). Then there exists a collection $(\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ of random variables whose joint law is uniquely characterized by the following requirements.*

(i) $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ *is an i.i.d. collection of Ω-valued r.v.'s with common law $|\mathbf{r}|^{-1}\mathbf{r}$.*

(ii) *For each finite rooted subtree $\mathbb{U} \subset \mathbb{T}$, the r.v.'s $(X_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{U}}$ are i.i.d. with common law ν and independent of $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i}\in\mathbb{U}}$.*

(iii) $X_{\mathbf{i}} = \gamma[\boldsymbol{\omega}_{\mathbf{i}}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\boldsymbol{\omega}_{\mathbf{i}})}) \qquad (\mathbf{i} \in \mathbb{T})$.

(1.58)

We call a collection of random variables $(\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ as in Lemma 1.9 the *Recursive Tree Process* (RTP) corresponding to the map γ and the solution ν of the RDE (1.54). We can view such an RTP as a generalization of a stationary backward Markov chain. For most purposes, we will only need the random variables $\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}}$ with $\mathbf{i} \in \mathbb{S}$, the random subtree defined in (1.43). The following proposition shows that by adding independent exponential lifetimes to an RTP, we obtain a stationary version of (1.57).

Proposition 1.10 (Continuous-time RTP). *Let $(\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an RTP corresponding to a solution ν of the RDE (1.54), and let $(\sigma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an independent i.i.d. collection of exponentially distributed random variables with mean $|\mathbf{r}|^{-1}$. Then, for each $t \geq 0$,*

(i) *conditional on $\mathcal{F}_t$, the r.v.'s $(X_{\mathbf{i}})_{\mathbf{i}\in\nabla\mathbb{S}_t}$ are i.i.d. with common law ν,*

(ii) $X_{\mathbf{i}} = \gamma[\boldsymbol{\omega}_{\mathbf{i}}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\boldsymbol{\omega}_{\mathbf{i}})}) \qquad (\mathbf{i} \in \mathbb{S})$.

(1.59)

At the end of Subsection 1.3 we have seen that in our example of a system with cooperative branching, the RDE (1.54) has three solutions when the branching rate satisfies $\alpha > 4$, two solutions for $\alpha = 4$, and only one solution for $\alpha < 4$. For $\alpha > 4$, the solutions to the RDE are $\nu_{\mathrm{low}}, \nu_{\mathrm{mid}}$, and ν_{upp}, where we let $\nu_{\ldots}$ denote the probability measure on $\{0, 1\}$ with mean $\nu_{\ldots}(\{1\}) = z_{\ldots}$ $(\ldots = \mathrm{low}, \mathrm{mid}, \mathrm{upp})$ as defined around (1.37). By Lemma 1.9, each of these solutions to the RDE defines an RTP.

1.6 Endogeny and bivariate uniqueness

In [AB05, Def 7], an RTP $(\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ corresponding to a solution ν of the RDE (1.54) is called *endogenous* if $X_\varnothing$ is a.s. measurable w.r.t. the σ-field generated by the random variables $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$. In Lemma 5.3 below, we will show that this is equivalent to $X_\varnothing$ being a.s. measurable w.r.t. the σ-field generated by the random variables $\$$ and $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{S}}$, where $\$$ is the random tree defined in (1.43). Aldous and Bandyopadhyay have shown that endogeny is equivalent to *bivariate uniqueness*, which we now explain.

Let $\mathcal{P}_{\mathrm{sym}}(S^n)$ denote the space of probability measures on S^n that are symmetric with respect to permutations of the coordinates. Let $\pi_m : S^n \to S$ denote the projection on the m-th coordinate, i.e., $\pi_m(x^1, \ldots, x^n) := x_m$, and let $\mu^{(n)} \circ \pi_m^{-1}$ denote the m-th marginal of a measure $\mu^{(n)} \in \mathcal{P}(S^n)$. For any $\mu \in \mathcal{P}(S)$, we define

$$\mathcal{P}(S^n)_\mu := \left\{ \mu^{(n)} \in \mathcal{P}(S^n) : \mu^{(n)} \circ \pi_m^{-1} = \mu \ \forall 1 \le m \le n \right\} \qquad (1.60)$$

to be the set of probability measures on S^n whose one-dimensional marginals are all equal to μ, and we denote $\mathcal{P}_{\mathrm{sym}}(S^n)_\mu := \mathcal{P}_{\mathrm{sym}}(S^n) \cap \mathcal{P}(S^n)_\mu$. Finally, we define a "diagonal" set

$$S_{\mathrm{diag}}^n := \left\{ (x_1, \ldots, x_n) \in S^n : x_1 = \cdots = x_n \right\} \qquad (1.61)$$

and given a measure $\mu \in \mathcal{P}(S)$, we let $\overline{\mu}^{(n)}$ denote the unique element of $\mathcal{P}(S^n)_\mu \cap \mathcal{P}(S_{\mathrm{diag}}^n)$, i.e.,

$$\overline{\mu}^{(n)} := \mathbb{P}\left[(X, \ldots, X) \in \cdot \right], \text{ where } X \text{ has law } \mu. \qquad (1.62)$$

Recall the definition of the n-variate map $T^{(n)}$ in (1.9). The following theorem has been proved in [MSS18, Thm 1], and in a slightly weaker form in [AB05, Thm 11]. Below, $\Rightarrow$ denotes weak convergence of probability measures.

Theorem 1.11 (Endogeny and n-variate uniqueness)**.** *Let ν be a solution of the RDE (1.54). Then the following statements are equivalent.*

(i) *The RTP corresponding to ν is endogenous.*
(ii) $(T^{(n)})^m(\mu) \underset{m \to \infty}{\Longrightarrow} \overline{\nu}^{(n)}$ *for all $\mu \in \mathcal{P}(S^n)_\nu$ and $n \ge 1$.*
(iii) $\overline{\nu}^{(2)}$ *is the only fixed point of $T^{(2)}$ in the space $\mathcal{P}_{\mathrm{sym}}(S^2)_\nu$.*

We remark that bivariate uniqueness as introduced in [AB05] refers to $\overline{\nu}^{(2)}$ being the only fixed point of $T^{(2)}$ in the space $\mathcal{P}(S^2)_\nu$. The equivalences in the above theorem tells us that bivariate uniqueness already follows from the weaker condition (iii) since it implies (ii), which implies n-variate uniqueness for any $n \ge 1$.

We will prove a continuous-time extension of Theorem 1.11, relating endogeny to solutions of the *n-variate mean-field equation*

$$\tfrac{\partial}{\partial t} \mu_t^{(n)} = |\mathbf{r}| \left\{ T^{(n)}(\mu_t^{(n)}) - \mu_t^{(n)} \right\} \qquad (t \ge 0), \qquad (1.63)$$

where we have replaced T in (1.4) by $T^{(n)}$ and we write $\mu_t^{(n)}$ to remind ourselves that this is a measure on S^n, rather than on S.

This equation has the following interpretation. As in Subsection 1.3, let $(\mathbf{X}_{s,u})_{s \le u}$ be a stochastic flow on S^N constructed from a Poisson point set Π. Let $(X^1(0), \ldots, X^n(0))$ be a random variable with values in S^n, independent of $(\mathbf{X}_{s,u})_{s \le u}$. Then setting

$$\left(X^1(t), \ldots, X^n(t) \right) := \left(\mathbf{X}_{0,t}(X^1(0)), \ldots, \mathbf{X}_{0,t}(X^n(0)) \right) \qquad (t \ge 0) \qquad (1.64)$$

defines a Markov process $(X^1(t), \ldots, X^n(t))_{t \ge 0}$ that consists of n Markov processes with initial states $X^1(0), \ldots, X^n(0)$ that are coupled in such a way that they are constructed using the same stochastic flow. Applying Theorem 1.5 to this *n-variate Markov process*, we see that the mean-field equation for the n-variate process takes the form (1.63).

We note that if $\mu_t^{(n)}$ solves the n-variate mean-field equation, then any m-dimensional marginal of $\mu_t^{(n)}$ solves the m-variate mean-field equation. Also, solutions to (1.63) started in an initial condition $\mu_0^{(n)} \in \mathcal{P}_{\mathrm{sym}}(S^n)$ satisfy $\mu_t^{(n)} \in \mathcal{P}_{\mathrm{sym}}(S^n)$ for all $t \geq 0$. Finally, it is easy to see that $\mu_0^{(n)} \in \mathcal{P}(S_{\mathrm{diag}}^n)$ implies $\mu_t^{(n)} \in \mathcal{P}(S_{\mathrm{diag}}^n)$ for all $t \geq 0$.

We now formulate a continuous-time extension of Theorem 1.11. Note that in view of (1.54), a measure $\nu^{(2)}$ is a fixed point of the *bivariate mean-field equation* (i.e., (1.63) with $n = 2$) if and only if it is a fixed point of $T^{(2)}$. Therefore, the equivalence of points (i) and (iii) from Theorem 1.11 immediately implies an analogue statement in the continuous-time setting.

Theorem 1.12 (Endogeny and the n-variate mean-field equation)**.** *Under the assumptions of Theorem 1.11, the following conditions are equivalent.*

  (i) *The RTP corresponding to ν is endogenous.*

 (ii) *For any $\mu_0^{(n)} \in \mathcal{P}(S^n)_\nu$ and $n \geq 1$, the solution $(\mu_t^{(n)})_{t \geq 0}$ to the n-variate equation (1.63) started in $\mu_0^{(n)}$ satisfies $\mu_t^{(n)} \underset{t \to \infty}{\Longrightarrow} \overline{\nu}^{(n)}$.*

Theorem 1.12 motivates us to study the bivariate mean-field equation in our example of a particle system with cooperative branching. Recall that in this example, $\mathcal{G} :=$ $\{\mathtt{cob}, \mathtt{dth}\}$ with $\mathtt{cob}$ and $\mathtt{dth}$ as in (1.13), and π is defined in (1.14). In line with (1.15) we write the bivariate mean-field equation as

$$\tfrac{\partial}{\partial t}\mu_t^{(2)} = \alpha\big\{T_{\mathtt{cob}^{(2)}}(\mu_t^{(2)}) - \mu_t^{(2)}\big\} + \big\{T_{\mathtt{dth}^{(2)}}(\mu_t^{(2)}) - \mu_t^{(2)}\big\}. \tag{1.65}$$

For simplicity, we restrict ourselves to symmetric solutions, i.e., solutions that take values in $\mathcal{P}_{\mathrm{sym}}(\{0,1\}^2)$. For any probability measure $\mu^{(2)} \in \mathcal{P}_{\mathrm{sym}}(\{0,1\}^2)$, we let $\mu^{(1)}$ denote its one-dimensional marginals, which are equal by symmetry. We let $\nu_{\mathrm{low}}, \nu_{\mathrm{mid}}, \nu_{\mathrm{upp}}$ denote the probability measures on $\{0,1\}$ with mean $\nu_{...}(\{1\}) = z_{...}$ $(... = \mathrm{low}, \mathrm{mid}, \mathrm{upp})$ as defined around (1.37).

Proposition 1.13 (Bivariate equation for cooperative branching)**.** *For $\alpha > 4$, the bivariate mean-field equation (1.65) has precisely four fixed points in $\mathcal{P}_{\mathrm{sym}}(\{0,1\}^2)$, namely*

$$\overline{\nu}_{\mathrm{low}}^{(2)}, \quad \underline{\nu}_{\mathrm{mid}}^{(2)}, \quad \overline{\nu}_{\mathrm{mid}}^{(2)}, \quad \text{and} \quad \overline{\nu}_{\mathrm{upp}}^{(2)}. \tag{1.66}$$

which are uniquely characterized by their respective marginals $\nu_{\mathrm{low}}, \nu_{\mathrm{mid}}, \nu_{\mathrm{mid}}, \nu_{\mathrm{upp}}$, as well as the fact that $\overline{\nu}_{\mathrm{low}}^{(2)}, \overline{\nu}_{\mathrm{mid}}^{(2)}$, and $\overline{\nu}_{\mathrm{upp}}^{(2)}$ are concentrated on $\{0,1\}_{\mathrm{diag}}^2 = \{(0,0),(1,1)\}$, but $\underline{\nu}_{\mathrm{mid}}^{(2)}$ is not.

For any $\mu_0^{(2)} \in \mathcal{P}_{\mathrm{sym}}(\{0,1\}^2)$, the solution to (1.65) started in $\mu_0^{(2)}$ converges as $t \to \infty$ to one of the fixed points in (1.66), the respective domains of attraction being

$$\begin{aligned} &\big\{\mu_0^{(2)} : \mu_0^{(1)}(\{1\}) < z_{\mathrm{mid}}\big\}, \quad \big\{\mu_0^{(2)} : \mu_0^{(1)}(\{1\}) = z_{\mathrm{mid}}, \ \mu_0^{(2)} \neq \overline{\nu}_{\mathrm{mid}}^{(2)}\big\}, \\ &\big\{\overline{\nu}_{\mathrm{mid}}^{(2)}\big\}, \quad \text{and} \quad \big\{\mu_0^{(2)} : \mu_0^{(1)}(\{1\}) > z_{\mathrm{mid}}\big\}. \end{aligned} \tag{1.67}$$

For $\alpha = 4$, there are two fixed points $\overline{\nu}_{\mathrm{low}}^{(2)}$ and $\overline{\nu}_{\mathrm{upp}}^{(2)}$ with respective domains of attraction

$$\big\{\mu_0^{(2)} : \mu_0^{(1)}(\{1\}) < z_{\mathrm{mid}}\big\} \quad \text{and} \quad \big\{\mu_0^{(2)} : \mu_0^{(1)}(\{1\}) \geq z_{\mathrm{mid}}\big\}, \tag{1.68}$$

while for $\alpha < 4$ all solutions converge to $\overline{\nu}_{\mathrm{low}}^{(2)}$.

Combining Proposition 1.13 with Theorem 1.12, we see that the RTPs corresponding to ν_{low} and ν_{upp} are endogenous, but for $\alpha > 4$, the RTP corresponding to ν_{mid} is not. As is

clear from [AB05, Table 1], few examples of nonendogenous RTPs were known at the time. Contrary to what is stated in [AB05, Table 1], frozen percolation on the binary tree is now known to be nonendogenous [RST19], but until recently few "natural" examples of nonendogenous RTPs have appeared in the literature. In fact, the RTP corresponding to ν_{mid} seems to be one of the simplest nontrivial examples of a nonendogenous RTP discovered so far. Another nice class of nonendogenous RTPs has recently been described in [MS19].

1.7 The higher-level mean-field equation

Following [MSS18, formula (1.1)], if S is a Polish space and $g : S^k \to S$ is a measurable map, then we define a measurable map $\check{g} : \mathcal{P}(S)^k \to \mathcal{P}(S)$ by

$$\begin{aligned} \check{g} := \ & \text{the law of } g(X_1, \ldots, X_k), \\ & \text{where } (X_1, \ldots, X_k) \text{ are independent with laws } \mu_1, \ldots, \mu_k. \end{aligned} \tag{1.69}$$

Note that in this notation, the map $T_g : \mathcal{P}(S) \to \mathcal{P}(S)$ from (1.12) is given by $T_g(\mu) = \check{g}(\mu, \ldots, \mu)$. As in [MSS18, formula (4.2)], we define a *higher-level map* $\check{T} : \mathcal{P}(\mathcal{P}(S)) \to \mathcal{P}(\mathcal{P}(S))$ by

$$\check{T}(\rho) := \ \text{the law of } \tilde{\gamma}[\boldsymbol{\omega}](\xi_1, \xi_2, \ldots), \tag{1.70}$$

where $\boldsymbol{\omega}$ is an Ω-valued random variable with law $|\mathbf{r}|^{-1}\mathbf{r}$ and $(\xi_i)_{i \geq 1}$ are i.i.d. $\mathcal{P}(S)$-valued random variables with law ρ. Iterates of the map $\check{T}$ have been studied in [MSS18, Section 4]. We will be interested in the *higher-level mean-field equation*

$$\tfrac{\partial}{\partial t}\rho_t = |\mathbf{r}|\big\{\check{T}(\rho_t) - \rho_t\big\} \qquad (t \geq 0). \tag{1.71}$$

A measure $\rho \in \mathcal{P}(\mathcal{P}(S))$ is the law of a random probability measure ξ on S. We denote the *n-th moment measure* of such a random measure ξ by

$$\rho^{(n)} := \mathbb{E}\big[\xi \otimes \cdots \otimes \xi\big] \quad \text{where } \xi \text{ has law } \rho. \tag{1.72}$$

(Here $\mathbb{E}[\,\cdot\,]$ denotes the expectation of a random measure; see the remark above Theorem 1.5.) Our notation for moment measures is on purpose similar to our earlier notation for solutions to the n-variate equation, because of the following proposition.

Proposition 1.14 (Moment measures). *If $(\rho_t)_{t \geq 0}$ solves the higher-level mean-field equation (1.71), then its n-th moment measures $(\rho_t^{(n)})_{t \geq 0}$ solve the n-variate equation (1.63).*

Similarly to Proposition 1.14, it has been shown in [MSS18, Lemma 2] that $\check{T}(\rho)^{(n)} = T^{(n)}(\rho^{(n)})$, and this formula holds even for $n = \infty$. In view of this, as discussed in Subsection 1.1, the higher-level map $\check{T}$ is effectively equivalent to the ∞-variate map $T^{(\infty)} : \mathcal{P}_{\mathrm{sym}}(S^\infty) \to \mathcal{P}_{\mathrm{sym}}(S^\infty)$. It follows from Proposition 1.14 that if ρ solves the *higher-level RDE*

$$\check{T}(\rho) = \rho, \tag{1.73}$$

then its n-th moment measures solve the n-variate RDE $T^{(n)}(\rho^{(n)}) = \rho^{(n)}$, with $T^{(n)}$ as in (1.9).

If X is an S-valued random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{H} \subset \mathcal{F}$ is a sub-σ-field, then $\mathbb{P}[X \in \cdot\, | \mathcal{H}]$ is a random probability measure[1] on S. As a consequence, the law of $\mathbb{P}[X \in \cdot\, | \mathcal{H}]$ is an element of $\mathcal{P}(\mathcal{P}(S))$. In the following theorem, which is based on [Str65, Thm 2] and which in its present form we cite from [MSS18, Thm 13], we use the fact that each Polish space S has a metrizable compactification $\overline{S}$ [Bou58, §6 No. 1, Theorem 1]. Moreover, we naturally identify $\mathcal{P}(S)$ with the space of all probability measures on $\overline{S}$ that are concentrated on S.

Theorem 1.15 (The convex order for laws of random probability measures). *Let S be a Polish space, let $\overline{S}$ be a metrizable compactification of S, and let $\mathcal{C}_{\mathrm{cv}}\big(\mathcal{P}(\overline{S})\big)$ denote the*

---

[1]Here we use that since S is Polish, regular versions of conditional expectations exist.

space of all convex continuous functions $\phi : \mathcal{P}(\overline{S}) \to \mathbb{R}$. Then, for $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$, the following statements are equivalent.

(i) $\int \phi \, d\rho_1 \leq \int \phi \, d\rho_2$ for all $\phi \in \mathcal{C}_{cv}\big(\mathcal{P}(\overline{S})\big)$.

(ii) There exists an S-valued random variable X defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and sub-σ-fields $\mathcal{H}_1 \subset \mathcal{H}_2 \subset \mathcal{F}$ such that $\rho_i = \mathbb{P}\big[\mathbb{P}[X \in \cdot | \mathcal{H}_i] \in \cdot\big]$ $(i = 1, 2)$.

If $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$ satisfy the equivalent conditions of Theorem 1.15, then we say that they are ordered in the *convex order* and denote this as $\rho_1 \leq_{cv} \rho_2$. It follows from [MSS18, Lemma 15] that $\leq_{cv}$ is a partial order; in particular, $\rho_1 \leq_{cv} \rho_2$ and $\rho_2 \leq_{cv} \rho_1$ imply $\rho_1 = \rho_2$.

Recall that in Subsection 1.1, we defined $\mathcal{P}(\mathcal{P}(S))_\mu$, which is $\big\{\rho \in \mathcal{P}(\mathcal{P}(S)) : \rho^{(1)} = \mu\big\}$. We define $\overline{\mu} \in \mathcal{P}(\mathcal{P}(S))_\mu$ by $\overline{\mu} := \mathbb{P}[\delta_X \in \cdot]$, where X has law μ. It is easy to see that the n-th moment measures of $\overline{\mu}$ are given by (1.62), so our present notation is consistent with earlier notation introduced there. By [MSS18, formula (4.7)], the measures $\delta_\mu, \overline{\mu}$ are the extremal elements of $\mathcal{P}(\mathcal{P}(S))_\mu$ w.r.t. the convex order, i.e.,

$$\delta_\mu \leq_{cv} \rho \leq_{cv} \overline{\mu} \qquad \big(\rho \in \mathcal{P}(\mathcal{P}(S))_\mu\big). \tag{1.74}$$

The following proposition is a continuous-time version of [MSS18, Prop 3].

Proposition 1.16 (Extremal solutions in the convex order). *If $(\rho_t^i)_{t \geq 0}$ $(i = 1, 2)$ are solutions to the higher-level mean-field equation (1.71) such that $\rho_0^1 \leq_{cv} \rho_0^2$, then $\rho_t^1 \leq_{cv} \rho_t^2$ for all $t \geq 0$. If ν solves the RDE (1.54), then $\overline{\nu}$ solves the higher-level RDE (1.73) and there exists a solution $\underline{\nu}$ of (1.73) such that*

$$\rho_t \underset{t \to \infty}{\Longrightarrow} \underline{\nu} \qquad \text{where } (\rho_t)_{t \geq 0} \text{ solves (1.71) with } \rho_0 = \delta_\nu. \tag{1.75}$$

Here $\Rightarrow$ denotes weak convergence of measures on $\mathcal{P}(S)$, equipped with the topology of weak convergence. Any solution $\rho \in \mathcal{P}(\mathcal{P}(S))_\nu$ to the higher-level RDE (1.73) satisfies

$$\underline{\nu} \leq_{cv} \rho \leq_{cv} \overline{\nu}. \tag{1.76}$$

The following result, which we cite from [MSS18, Prop. 4], describes the higher-level RTPs associated with the solutions $\underline{\nu}$ and $\overline{\nu}$ of the higher-level RDE.

Proposition 1.17 (Higher-level RTPs). *Let ν be a solution of the RDE (1.54) and let $\underline{\nu}$ and $\overline{\nu}$ as in (1.76) be the corresponding minimal and maximal solutions to the higher-level RDE, with respect to the convex order. Let $(\omega_\mathbf{i}, X_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ be an RTP corresponding to γ and ν and set*

$$\xi_\mathbf{i} := \mathbb{P}\big[X_\mathbf{i} \in \cdot \,|\, (\omega_{\mathbf{ij}})_{\mathbf{j} \in \mathbb{T}}\big] \qquad (\mathbf{i} \in \mathbb{T}). \tag{1.77}$$

Then $(\omega_\mathbf{i}, \xi_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ is an RTP corresponding to $\check{\gamma}$ and $\underline{\nu}$. Also, $(\omega_\mathbf{i}, \delta_{X_\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ is an RTP corresponding to $\check{\gamma}$ and $\overline{\nu}$.

Proposition 1.17 gives a more concrete interpretation of the solutions $\underline{\nu}$ and $\overline{\nu}$ to the higher-level RDE from (1.76). Indeed, if $(\omega_\mathbf{i}, X_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ is an RTP corresponding to ν, then

$$\overline{\nu} = \mathbb{P}\big[\delta_{X_\varnothing} \in \cdot\big], \tag{1.78}$$

which corresponds to "perfect knowledge" about the state $X_\varnothing$ of the root, while

$$\underline{\nu} = \mathbb{P}\big[\mathbb{P}\big[X_\varnothing \in \cdot \,|\, (\omega_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}\big] \in \cdot\big] \tag{1.79}$$

corresponds to the knowledge about $X_\varnothing$ that is contained in the random variables $(\omega_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$. Since $X_\varnothing$ is a measurable function of $(\omega_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ if and only if its conditional law given $(\omega_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ equals $\delta_{X_\varnothing}$, it follows from (1.78) and (1.79) that the RTP corresponding to ν is endogenous if and only if $\underline{\nu} = \overline{\nu}$.

It is instructive to demonstrate the general theory on our concrete example of a system with cooperative branching and deaths. Recall that for $\alpha > 4$, the mean-field equation (1.15) has three fixed points $\nu_{\mathrm{low}}, \nu_{\mathrm{mid}}, \nu_{\mathrm{upp}}$. We denote the corresponding minimal and maximal solutions to the higher-level RDE in the sense of (1.76) by $\underline{\nu}_{\ldots}$ and $\overline{\nu}_{\ldots}$ ($\ldots = \mathrm{low}, \mathrm{mid}, \mathrm{upp}$). The following theorem lifts the results from Proposition 1.13 about the bivariate equation to a higher level. Indeed, using the theorem below, it is easy to see that the measures $\overline{\nu}_{\mathrm{low}}^{(2)}, \underline{\nu}_{\mathrm{mid}}^{(2)}, \overline{\nu}_{\mathrm{mid}}^{(2)}$ and $\overline{\nu}_{\mathrm{upp}}^{(2)}$ from Proposition 1.13 are in fact the second moment measures of the measures $\overline{\nu}_{\mathrm{low}}, \underline{\nu}_{\mathrm{mid}}, \overline{\nu}_{\mathrm{mid}}$ and $\overline{\nu}_{\mathrm{upp}}$.

Theorem 1.18 (Higher-level equation for cooperative branching). *Let $\nu_{\mathrm{low}}, \nu_{\mathrm{mid}}$, and ν_{upp} denote the fixed points of the mean-field equation (1.15) defined above Proposition 1.13. Then we have for the corresponding minimal and maximal solutions to the higher-level RDE that*

$$\underline{\nu}_{\mathrm{low}} = \overline{\nu}_{\mathrm{low}}, \quad \underline{\nu}_{\mathrm{upp}} = \overline{\nu}_{\mathrm{upp}}, \quad but \quad \underline{\nu}_{\mathrm{mid}} \neq \overline{\nu}_{\mathrm{mid}} \quad (\alpha > 4). \tag{1.80}$$

For $\alpha > 4$, the higher-level RDE (1.73) has four solutions, namely

$$\overline{\nu}_{\mathrm{low}}, \quad \underline{\nu}_{\mathrm{mid}}, \quad \overline{\nu}_{\mathrm{mid}}, \quad and \quad \overline{\nu}_{\mathrm{upp}}. \tag{1.81}$$

Any solution $(\rho_t)_{t \geq 0}$ to the higher-level mean-field equation (1.71) converges as $t \to \infty$ to one of the fixed points in (1.81), the respective domains of attraction being

$$\big\{\rho_0 : \rho_0^{(1)}(\{1\}) < z_{\mathrm{mid}}\big\}, \quad \big\{\rho_0 : \rho_0^{(1)}(\{1\}) = z_{\mathrm{mid}}, \ \rho_0 \neq \overline{\nu}_{\mathrm{mid}}\big\}, \tag{1.82}$$

$$\big\{\overline{\nu}_{\mathrm{mid}}\big\}, \quad and \quad \big\{\rho_0 : \rho_0^{(1)}(\{1\}) > z_{\mathrm{mid}}\big\}.$$

For $\alpha = 4$, there are two fixed points $\overline{\nu}_{\mathrm{low}}$ and $\overline{\nu}_{\mathrm{upp}}$ with respective domains of attraction

$$\big\{\rho_0 : \rho_0^{(1)}(\{1\}) < z_{\mathrm{mid}}\big\} \quad and \quad \big\{\rho_0 : \rho_0^{(1)}(\{1\}) \geq z_{\mathrm{mid}}\big\}, \tag{1.83}$$

while for $\alpha < 4$ all solutions converge to $\overline{\nu}_{\mathrm{low}}$.

Since a probability measure $\mu \in \mathcal{P}(\{0,1\})$ is uniquely characterized by $\mu(\{1\}) \in [0,1]$, there is a natural identification $\mathcal{P}(\{0,1\}) \cong [0,1]$. Let $\widehat{\mathtt{cob}}$ and $\widehat{\mathtt{dth}}$ denote the higher-level maps $\check{g}$ corresponding to $g = \mathtt{cob}, \mathtt{dth}$, which using the identification $\mathcal{P}(\{0,1\}) \cong [0,1]$ we view as maps $\widehat{\mathtt{cob}} : [0,1]^3 \to [0,1]$ and $\widehat{\mathtt{dth}} : [0,1]^0 \to [0,1]$. One can check that

$$\widehat{\mathtt{dth}}(\varnothing) = 0 \quad and \quad \widehat{\mathtt{cob}}(\eta_1, \eta_2, \eta_3) = \eta_1 + (1 - \eta_1)\eta_2\eta_3 \quad (\eta_1, \eta_2, \eta_3 \in [0,1]). \tag{1.84}$$

Identifying $\mathcal{P}(\mathcal{P}(\{0,1\})) \cong \mathcal{P}[0,1]$, we can identify the measures $\overline{\nu}_{\mathrm{low}}, \underline{\nu}_{\mathrm{mid}}, \overline{\nu}_{\mathrm{mid}}$, and $\overline{\nu}_{\mathrm{upp}}$ with probability laws on $[0,1]$. Letting η denote a random variable with law $\nu \in \mathcal{P}[0,1]$, the higher-level RDE, written in the form (1.55), then reads

$$\eta \overset{\mathrm{d}}{=} \chi \cdot \big(\eta_1 + (1 - \eta_1)\eta_2\eta_3\big), \tag{1.85}$$

where η_1, η_2, η_3 are independent copies of η and χ is an independent Bernoulli random variable with $\mathbb{P}[\chi = 1] = \alpha/(\alpha + 1)$. Theorem 1.18 says that for $\alpha > 4$, this equation has four solutions. Three "trivial" solutions $\overline{\nu}_{\mathrm{low}}, \overline{\nu}_{\mathrm{mid}}, \overline{\nu}_{\mathrm{upp}}$ that correspond to Bernoulli η with parameters

$$\mathbb{P}[\eta = 1] = z, \quad \mathbb{P}[\eta = 0] = 1 - z \quad with \quad z = z_{\mathrm{low}}, z_{\mathrm{mid}}, z_{\mathrm{low}}, \tag{1.86}$$

and a "nontrivial" solution $\underline{\nu}_{\mathrm{mid}}$ for which $\mathbb{P}[0 < \eta < 1] > 0$. In view of Proposition 1.17, we can interpret this nontrivial solution (viewed as a probability law on $[0,1]$) as

$$\underline{\nu}_{\mathrm{mid}} = \mathbb{P}\big[\mathbb{P}[X_\varnothing = 1 \,|\, (\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}] \in \cdot\,\big], \tag{1.87}$$

where $(\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ is the RTP corresponding to $\underline{\nu}_{\mathrm{mid}}$. The following lemma summarizes some elementary facts about the law $\underline{\nu}_{\mathrm{mid}}$. We note that by solving the n-variate RDE for $n \geq 3$, one should in principle be able to calculate higher moments of $\underline{\nu}_{\mathrm{mid}}$, although the formulas quickly become unwieldy.
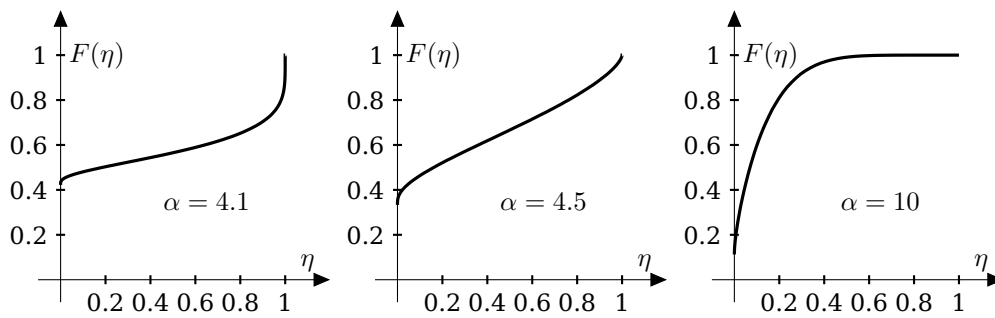
Figure 2: Numerical data for the distribution function $F(\eta) := \underline{\nu}_{\mathrm{mid}}([0, \eta])$ of the nontrivial solution $\underline{\nu}_{\mathrm{mid}}$ to the RDE (1.85).

Lemma 1.19 (Nontrivial solution of the higher-level RDE)**.** *Let $\alpha > 4$ and let η be a random variable with law $\underline{\nu}_{\mathrm{mid}}$. Then*

$$\begin{aligned}
\mathbb{E}[\eta] &= z_{\mathrm{mid}} = \tfrac{1}{2} - \sqrt{\tfrac{1}{4} - \tfrac{1}{\alpha}}, \\
\mathbb{E}[\eta^2] &= \tfrac{1}{2} z_{\mathrm{mid}} - \tfrac{1}{2} + \tfrac{1}{2}\sqrt{13 z_{\mathrm{mid}}^2 - 6 z_{\mathrm{mid}} + 1 + \tfrac{4}{\alpha}}.
\end{aligned} \tag{1.88}$$

Moreover,

$$\mathbb{P}[\eta = 0] = z_{\mathrm{mid}} \quad \text{and} \quad \mathbb{P}[\eta = 1] = 0. \tag{1.89}$$

It is not too hard to obtain numerical data for $\underline{\nu}_{\mathrm{mid}}$, see Figure 2. These data suggest that apart from the atom in 0, the measure $\underline{\nu}_{\mathrm{mid}}$ has a smooth density with respect to the Lebesgue measure, but we have no proof for this. We have tried to find an explicit formula for the density but have not been successful.

1.8 Lower and upper solutions

In this and the next subsection we collect a few further results on endogeny and the uniqueness of solutions to RDEs. In the present subsection, we show that the endogeny of the RTPs corresponding to ν_{low} and ν_{upp} follows from a general principle, discovered in [AB05], that says that RDEs that are defined by monotone maps always have a minimal and maximal solution with respect to the stochastic order, and that the RTPs corresponding to these solutions are always endogenous.

Let S be a compact metrizable space that is equipped with a partial order $\leq$ that is *closed* in the sense that

$$\big\{ (x, y) \in S^2 : x \leq y \big\} \tag{1.90}$$

is a closed subset of S^2, equipped with the product topology. Recall that a function f from one partially ordered space into another is monotone if $x \leq y$ implies $f(x) \leq f(y)$, and a subset A of a partially ordered space is increasing if $A \ni x \leq y$ implies $y \in A$. It is known that for two probability measures $\mu_1, \mu_2 \in \mathcal{P}(S)$, the following statements are equivalent:

(i) $\mu_1(A) \leq \mu_2(A)$ for all closed increasing $A \subset S$.

(ii) $\int f \, \mathrm{d}\mu_1 \leq \int f \, \mathrm{d}\mu_2$ for all bounded continuous monotone $f : S \to \mathbb{R}$.

(iii) Two random variables X_1, X_2 with laws μ_1, μ_2 can be coupled such that $X_1 \leq X_2$ a.s.

The equivalence of (ii) and (iii) is proved in [Lig85, Thm II.2.4]. The equivalence of (i) and (iii) holds more generally for Polish spaces, see [KKO77, Thm 1 (ii) and (vi)]. In

the general setting of Polish spaces, the implications (iii)⇒(i) and (iii)⇒(ii) are trivial, but the implication (ii)⇒(i) needs the additional assumption of *monotone normality*, see [HLL18, Prop. 3.6 and 3.11].

If μ_1, μ_2 satisfy the above conditions, then one says that they are *stochastically ordered*, denoted as $\mu_1 \leq \mu_2$. This defines a partial order on $\mathcal{P}(S)$; in particular, by Lemma 6.2 below, $\mu_1 \leq \mu_2 \leq \mu_1$ implies $\mu_1 = \mu_2$.

The proposition below is a variant of [AB05, Lemma 15]. As in our usual setting, we assume that S and Ω are Polish spaces, $\mathbf{r}$ is a nonzero finite measure on Ω, and $\gamma : \Omega \times S^{\mathbb{N}_+} \to S$ and $\kappa : \Omega \to \mathbb{N}$ are measurable functions such that (1.2) and (1.3) hold. If S is equipped with a partial order, then we equip S^k with the product partial order. Recall that Proposition 1.3 gives sufficient conditions for T to be continuous w.r.t. the topology of weak convergence.

Proposition 1.20 (Lower and upper solutions to RDE). *Assume that S is compact and equipped with a closed partial order. Assume that S has minimal and maximal elements, denoted by 0 and 1. Assume $\gamma[\omega]$ is monotone for each $\omega \in \Omega$ and that the operator T in (1.1) is continuous w.r.t. the topology of weak convergence. Then there exists solutions $\nu_{\mathrm{low}}, \nu_{\mathrm{upp}}$ to the RDE (1.54) that are minimal and maximal with respect to the stochastic order, in the sense that any solution ν to the RDE (1.54) must satisfy*

$$\nu_{\mathrm{low}} \leq \nu \leq \nu_{\mathrm{upp}}, \tag{1.91}$$

where $\Rightarrow$ denotes weak convergence. Moreover, if $(\mu_t^{\mathrm{low}})_{t \geq 0}$ and $(\mu_t^{\mathrm{upp}})_{t \geq 0}$ denote the solutions to the mean-field equation (1.4) with initial states $\mu_0^{\mathrm{low}} = \delta_0$ and $\mu_0^{\mathrm{upp}} = \delta_1$, then

$$\mu_t^{\mathrm{low}} \underset{t \to \infty}{\Longrightarrow} \nu_{\mathrm{low}} \quad and \quad \mu_t^{\mathrm{upp}} \underset{t \to \infty}{\Longrightarrow} \nu_{\mathrm{upp}}. \tag{1.92}$$

Finally, the RTPs corresponding to ν_{low} and ν_{upp} are endogenous.

We can view the solutions ν_{low} and ν_{upp} to the RDE (1.54) as mean-field versions of the lower and upper invariant laws of monotone particle systems; compare [Lig85, Thm III.2.3].

In our example of a system with cooperative branching, the maps `cob` and `dth` are monotone, so Proposition 1.20 is applicable. Since the measures we called ν_{low} and ν_{upp} before are the $t \to \infty$ limits of the solutions of the mean-field equation started in δ_0 and δ_1, our earlier notation agrees with the more general notation of Proposition 1.20. The endogeny of the RTPs corresponding to ν_{low} and ν_{upp}, which before we proved based on an analysis of the bivariate equation, using Proposition 1.13 and Theorem 1.11, alternatively follows from Proposition 1.20.

1.9 Conditions for uniqueness

In the present subsection, we prove some results of varying generality that allow one to conclude that a given RDE has a unique solution. In our example of a system with cooperative branching and deaths, this happens if and only if $\alpha < 4$. We will see that there are some general results that can be applied to prove uniqueness in the whole regime $\alpha < 4$. We also make a connection with a general duality for monotone particle systems described in [SS18]. Although duality plays only a minor role in our paper, the original motivation for the work that led to it was to understand this duality in the mean-field limit.

We return to our usual set-up from Subsection 1.1 with S and Ω Polish spaces and γ, κ and $\mathbf{r}$ satisfying (1.2) and (1.3). We also recall the random subtrees $\mathbb{S}_t \subset \mathbb{S} \subset \mathbb{T}$ defined in (1.43) as well as the fact that $\mathbb{S}_t$ for any $t \geq 0$ are a.s. finite by (1.3). The tree $\mathbb{S}$ is the family tree of the branching process $(\nabla \mathbb{S}_t)_{t \geq 0}$. In view of this, by well-known facts about

branching processes, $\mathbb{S}$ is a.s. finite if and only if

$$\int_\Omega \mathbf{r}(\mathrm{d}\omega)\,(\kappa(\omega)-1) < 0 \quad\text{or}\quad \int_\Omega \mathbf{r}(\mathrm{d}\omega)\,(\kappa(\omega)-1) = 0 \text{ while } \mathbf{r}\big(\{\omega:\kappa(\omega)\neq 1\}\big) > 0. \quad (1.93)$$

Recall that $G_t = G_{\mathbb{S}_t}$, where for any finite subtree $\mathbb{U} \subset \mathbb{S}$ that contains the root, $G_\mathbb{U} : S^{\nabla\mathbb{U}} \to S$ is the map defined in (1.46). If $\mathbb{S}$ is a.s. finite, then $\nabla\mathbb{S}_t = \emptyset$ for t sufficiently large and hence $G_t : S^{\nabla\mathbb{S}_t} \to S$ is eventually constant.

More generally, if $\mathbb{U}$ is finite subtree of $\mathbb{S}$ that contains the root $\varnothing$, then we say that $\mathbb{U}$ is a *root determining subtree* if the map $G_\mathbb{U} : S^{\nabla\mathbb{U}} \to S$ is constant. Note that this can happen even if $\nabla\mathbb{U} \neq \emptyset$. It is easy to see that if $\mathbb{V} \subset \mathbb{U}$ and $\mathbb{V}$ is root determining, then the same is true for $\mathbb{U}$. We say that $\mathbb{U}$ is a *minimal root determining subtree* if $\mathbb{U}$ is root determining but there exists no $\mathbb{V} \subset \mathbb{U}$ with $\mathbb{V} \neq \mathbb{U}$ that is root determining. By our previous remark, it suffices to check this for such $\mathbb{V}$ that differ from $\mathbb{U}$ by a single element.

Lemma 1.21 (Root determining subtrees). *The following conditions are equivalent:*

 (i) *There a.s. exists a $t < \infty$ such that G_s is constant for all $s \geq t$.*
 (ii) $\mathbb{S}$ *a.s. contains a root determining subtree.*
 (iii) $\mathbb{S}$ *a.s. contains a minimal root determining subtree.*

If $\mathbb{U}$ is a subtree of $\mathbb{S}$, then we denote by $\Xi_\mathbb{U}$ the set of all $x = (x_\mathbf{i})_{\mathbf{i}\in\mathbb{U}\cup\nabla\mathbb{U}}$ that satisfy (1.45). We say that $\mathbb{U}$ is *uniquely determined* if $x, y \in \Xi_\mathbb{U}$ imply $x_\mathbf{i} = y_\mathbf{i}$ ($\mathbf{i} \in \mathbb{U}$). The following lemma is inspired by [AB05, Lemma 14] who showed that (i) implies that the RDE (1.54) has a unique solution and the corresponding RTP is endogenous.

Lemma 1.22 (Uniquely determined subtrees). *Between the following four conditions, one has the implications (i)$\Rightarrow$(ii)$\Rightarrow$(iii)$\Rightarrow$(iv) and (ii)$\Rightarrow$(v). If S is finite, then moreover (iii)$\Rightarrow$(ii), and if $S = \{0,1\}$, then (ii)$\Rightarrow$(i).*

 (i) $\mathbb{S}$ *a.s. contains a finite, uniquely determined subtree that contains the root $\varnothing$.*
 (ii) *The equivalent conditions of Lemma 1.21 are satisfied.*
 (iii) $\mathbb{S}$ *is a.s. uniquely determined.*
 (iv) *The RDE (1.54) has at most one solution and any corresponding RTP is endogenous.*
 (v) *The RDE (1.54) has a solution ν that is globally attractive in the sense that any solution $(\mu_t)_{t\geq 0}$ to (1.4) satisfies $\|\mu_t - \nu\| \xrightarrow[t\to\infty]{} 0$, where $\|\cdot\|$ denotes the total variation norm.*

The following lemma illustrates these ideas on our example of a system with cooperative branching and deaths. Below, $|\mathbb{U}\cap\{\mathbf{i}2,\mathbf{i}3\}|$ denotes the cardinality of $\mathbb{U}\cap\{\mathbf{i}2,\mathbf{i}3\}$. See Figure 3 for an example.

Lemma 1.23 (The uniqueness regime). *Let $S = \{0,1\}$ and $\mathcal{G} := \{\mathtt{cob},\mathtt{dth}\}$, and let π be as in (1.14). Then (1.93) is satisfied if and only if $\alpha \leq \frac{1}{2}$, while conditions (i)–(iii) of Lemma 1.22 are satisfied if and only if $\alpha < 4$. Moreover, a finite subtree $\mathbb{U} \subset \mathbb{S}$ is a minimal root determining subtree if and only if*

$$\mathbf{i}1 \in \mathbb{U} \quad\text{and}\quad \big|\mathbb{U}\cap\{\mathbf{i}2,\mathbf{i}3\}\big| = 1 \quad\text{for each } \mathbf{i} \in \mathbb{U} \text{ with } \gamma[\boldsymbol{\omega}_\mathbf{i}] = \mathtt{cob}. \quad (1.94)$$

Lemma 1.23 shows that in our example of a system with cooperative branching and deaths, the conditions of Lemma 1.21 are in fact equivalent to uniqueness of solutions to the RDE. As the next lemma shows, this is a consequence of monotonicity.

Lemma 1.24 (Uniqueness for monotone systems). *Assume that S is a finite partially ordered set that contains a minimal and maximal element, and assume that $\gamma[\omega]$ is monotone for each $\omega \in \Omega$. Then the RDE (1.54) has a unique solution if and only if the equivalent conditions of Lemma 1.21 are satisfied.*
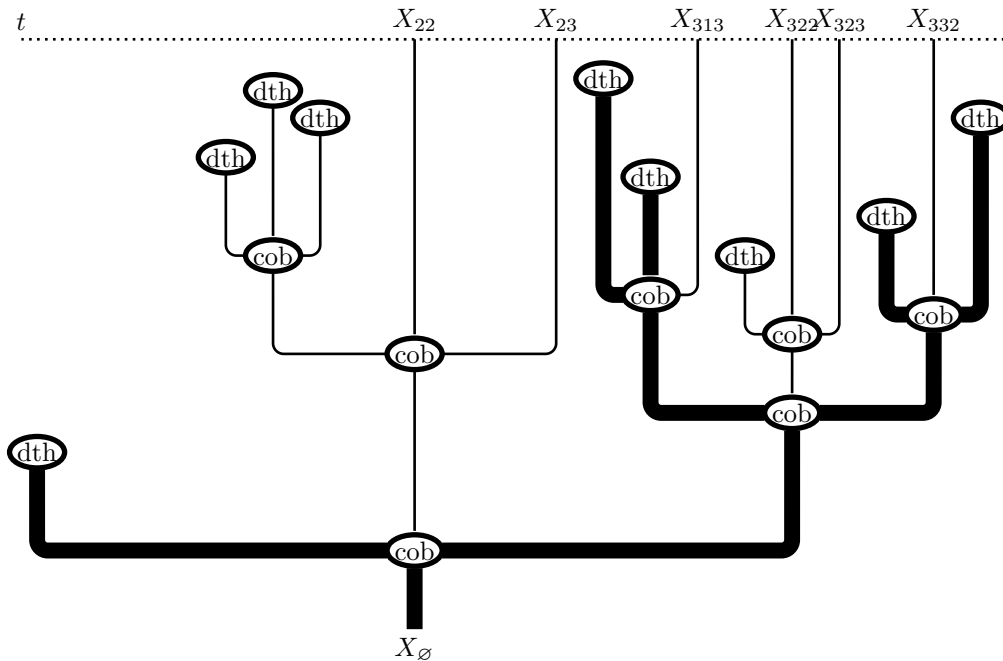
Figure 3: A minimal root determining subtree. In this example, $X_\varnothing = 0$ regardless of the values of $X_{22}, X_{23}, X_{313}, X_{322}, X_{323}, X_{332}$.

In the remainder of this subsection, we focus on the case that $S = \{0, 1\}$ and $\gamma[\omega]$ is monotone for all $\omega \in \Omega$, which allows us to make a connection to a general duality for monotone particle systems described in [SS18]. Recall that a set $A \subset \{0, 1\}^k$ is increasing if $A \ni x \leq y$ implies $y \in A$. A *minimal element* of A is an $y \in A$ such that $A \ni x \leq y$ implies $x = y$. If K is a nonempty finite set and $G : \{0, 1\}^K \to \{0, 1\}$ is a monotone map, then the inverse image $G^{-1}(\{1\})$ is an increasing set. We set

$$Y_G := \big\{ y \in \{0, 1\}^K : y \text{ is a minimal element of } G^{-1}(\{1\}) \big\}. \tag{1.95}$$

Then

$$G(x) = 1 \quad \text{if and only if} \quad x \geq y \text{ for some } y \in Y_G. \tag{1.96}$$

These formulas remain true when $K = \emptyset$, provided we define $\{0, 1\}^\emptyset := \{\varnothing\}$ and we let $Y_G := \{\varnothing\}$ if $G(\varnothing) = 1$ and $Y_G := \emptyset$ if $G(\varnothing) = 0$.

Recall from Section 1.4 that $(\nabla \$_t, G_t)_{t \geq 0}$ is a Markov process. If $S = \{0, 1\}$ and $\gamma[\omega]$ is monotone for all $\omega \in \Omega$, then the random map $G_t : \{0, 1\}^{\nabla \$_t} \to \{0, 1\}$ is monotone for each $t \geq 0$. In view of this, by (1.96), G_t is uniquely characterized by Y_{G_t} and hence $(\nabla \$_t, Y_{G_t})_{t \geq 0}$ is a Markov process too. For a system with cooperative branching and deaths, this process has been defined before in [Mac17, Section I.2.1.2]. As explained in more detail there, it can be seen as the mean-field limit of a general dual for monotone particle systems described in [SS18, Section 5.2].

Let $S = \{0, 1\}$, let $\gamma[\omega]$ be monotone for all $\omega \in \Omega$, and let $\mathbb{U}$ be a subtree of $\$$ that contains the root $\varnothing$. Borrowing terminology from percolation theory, we say that $\mathbb{O}$ is a *open subtree of* $\mathbb{U}$ if $\varnothing \in \mathbb{O} \subset \mathbb{U} \cup \nabla \mathbb{U}$ and

$$A_{\mathbf{i}} := \big\{ j \in [\kappa(\boldsymbol{\omega}_{\mathbf{i}})] : \mathbf{i}j \in \mathbb{O} \big\} \quad \text{satisfies} \quad 1_{A_{\mathbf{i}}} \in Y_{\gamma[\boldsymbol{\omega}_{\mathbf{i}}]} \quad \forall \mathbf{i} \in \mathbb{O} \cap \mathbb{U}, \tag{1.97}$$

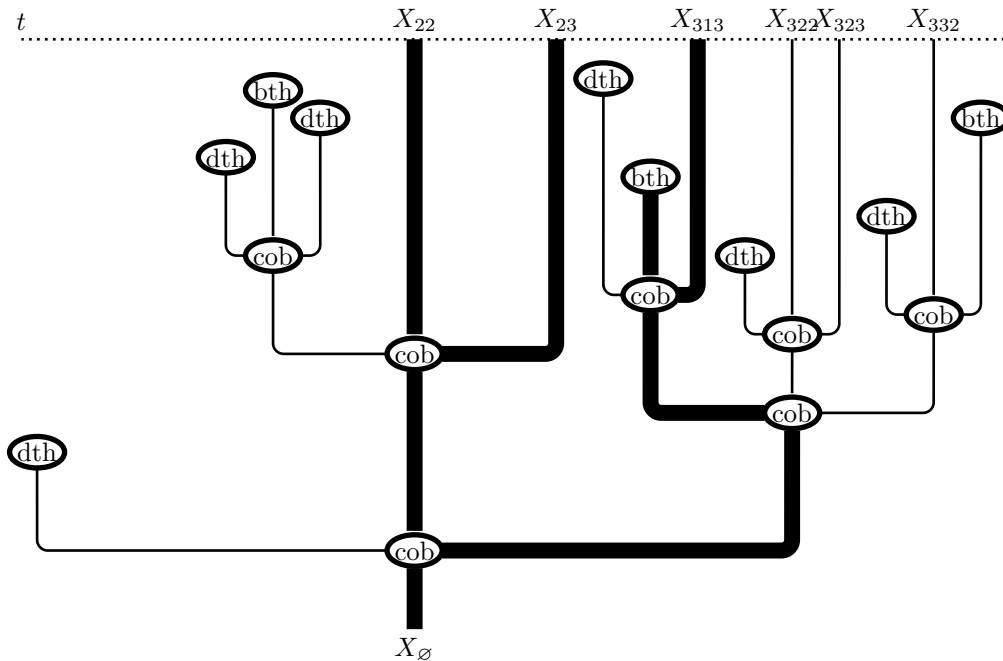where we use the convention that $1_{A_{\mathbf{i}}} := \varnothing$ if $\kappa(\boldsymbol{\omega}_{\mathbf{i}}) = 0$.

Figure 4: An open subtree of $\mathbb{S}_t$ for a system with cooperative branching, deaths, and births, represented by the maps $\mathtt{cob}$ and $\mathtt{dth}$ from (1.13) and $\mathtt{bth}$ from (1.100). In this example, if X_{22}, X_{23}, and X_{313} are 1, then the state at the root $X_\varnothing$ is also 1.

Lemma 1.25 (Open subtrees). *Assume that $S = \{0, 1\}$ and $\gamma[\omega]$ is monotone for all $\omega \in \Omega$. Then*

$$\nu_{\mathrm{upp}}(\{1\}) = \mathbb{P}\big[\text{there exists an open subtree of } \mathbb{S}\big],$$
$$\nu_{\mathrm{low}}(\{1\}) = \mathbb{P}\big[\text{there exists a finite open subtree of } \mathbb{S}\big]. \tag{1.98}$$

If moreover $\gamma[\omega](0, \ldots, 0) = 0$ for each $\omega \in \Omega$, then

$$Y_{G_t} = \big\{ y \in \{0, 1\}^{\nabla \mathbb{S}_t} : \exists \text{ open subtree } \mathbb{O} \text{ of } \mathbb{S}_t \text{ s.t. } y = 1_{\mathbb{O} \cap \nabla \mathbb{S}_t} \big\}. \tag{1.99}$$

We note that formula (1.98) can be generalized to more general finite partially ordered sets S, see Lemma 6.16 below. Again, it will be useful to illustrate our definitions on the concrete example of a system with cooperative branching and death. To make the example more interesting, we add a birth map $\mathtt{bth} : S^0 \to S$, which is defined similarly to the death map as

$$\mathtt{bth}(\varnothing) := 1. \tag{1.100}$$

The following lemma describes open subtrees for a system described by the maps $\mathtt{cob}, \mathtt{dth}, \mathtt{bth}$; see Figure 4 for an illustration.

Lemma 1.26 (Systems with cooperative branching, deaths, and births). *Let $S = \{0, 1\}$, $\mathcal{G} := \{\mathtt{cob}, \mathtt{dth}, \mathtt{bth}\}$, with*

$$\pi\big(\{\mathtt{cob}\}\big) := \alpha \geq 0, \quad \pi\big(\{\mathtt{dth}\}\big) := 1, \quad \text{and} \quad \pi\big(\{\mathtt{bth}\}\big) := \beta \geq 0. \tag{1.101}$$

Let $\mathbb{U}$ be a subtree of $\mathbb{S}$ that contains the root and let $\mathbb{O} \subset \mathbb{U} \cup \nabla \mathbb{U}$ satisfy $\varnothing \in \mathbb{O}$. Then $\mathbb{O}$ is an open subtree of $\mathbb{U}$ if and only if for all $\mathbf{i} \in \mathbb{O} \cap \mathbb{U}$,

$$\gamma[\boldsymbol{\omega_i}] \neq \mathtt{dth} \quad \text{and} \quad \{\mathbf{i}1, \mathbf{i}2, \mathbf{i}3\} \cap \mathbb{O} = \{\mathbf{i}1\} \text{ or } \{\mathbf{i}2, \mathbf{i}3\} \text{ if } \gamma[\boldsymbol{\omega_i}] = \mathtt{cob}. \tag{1.102}$$

We can think of open subtrees as a generalization of the open paths from oriented percolation. Outside of a mean-field setting, using ideas from [SS18, Section 5.2], one can characterize the upper invariant law of quite general monotone particle system in terms of "open structures" that in general are neither paths nor trees.

2 Discussion

This section is divided into four subsections. In Subsection 2.1, we discuss the relation of our work to [BCH18], who in parallel to our work have studied Moran models that generalize our running example of a system with cooperative branching and deaths. In Subsection 2.2, we compare our results and methods with the existing literature on mean-field limits. In Subsection 2.3, we state open problems and we conclude in Subsection 2.4 with an outline of the proofs.

2.1 A Moran model with frequency-dependent selection

Let $\mathtt{bra} : \{0,1\}^2 \to \mathtt{bra}$ be the branching map defined as

$$\mathtt{bra}(x_1, x_2) := x_1 \vee x_2 \qquad (x_1, x_2 \in \{0,1\}). \tag{2.1}$$

Consider a system with $S = \{0,1\}$, $\mathcal{G} := \{\mathtt{cob}, \mathtt{bra}, \mathtt{dth}, \mathtt{bth}\}$, with rates

$$\pi(\{\mathtt{cob}\}) := \gamma, \quad \pi(\{\mathtt{bra}\}) := s, \quad \pi(\{\mathtt{dth}\}) := u\nu_0, \quad \text{and} \quad \pi(\{\mathtt{bth}\}) := u\nu_1. \tag{2.2}$$

$\gamma \geq 0$, $s > 0$, $\nu_0, \nu_1 \geq 0$ with $\nu_0 + \nu_1 = 1$, and $u > 0$. If $(\mu_t)_{t \geq 0}$ solves the corresponding mean-field equation (1.11), then $p_t := \mu_t(\{1\})$ solves the ODE (compare (1.36))

$$\frac{\partial}{\partial t} p_t = \gamma p_t^2 (1 - p_t) + s p_t (1 - p_t) - u\nu_0 p_t + u\nu_1 (1 - p_t) \qquad (t \geq 0). \tag{2.3}$$

This equation has an interpretation in terms of a Moran model describing a fixed population of N individuals which can be of two types, 0 and 1, where type 1 is fitter than type 0. The parameter γ is the *frequency dependent selection rate*, s is the *selection rate*, u is the *mutation rate*, and ν_0, ν_1 are mutation probabilities. The frequency dependent selection is of a type that is especially appropriate to describe an advantageous, (partially) recessive gene in a diploid population.

In parallel to our work, Moran models of this form have been studied by Ellen Baake, Fernando Cordero, and Sebastian Hummel in [BCH18]. A notational difference between their work and the discussion here is that they denote the fitter type by 0, so their [BCH18, formula (2.1)] is our (2.3) rewritten in terms of $y(t) = 1 - p_t$ and with the roles of ν_0 and ν_1 reversed. They prove that (2.3) describes the mean-field limit of a class of Moran models [BCH18, Prop. 2.7] and that in the limit $N \to \infty$, the genealogy of a single individual is described by an *Ancestral Selection Graph* (ASG) $\mathcal{A}_t$, which in our notation corresponds to

$$\mathcal{A}_t := \big(\nabla \mathbb{S}_t, \mathbb{S}_t, (\boldsymbol{\omega}_\mathbf{i}, \sigma_\mathbf{i})_{\mathbf{i} \in \mathbb{S}_t}\big), \tag{2.4}$$

i.e., this is the random tree with maps attached to its branch points depicted in Figure 1.

The authors of [BCH18] define a duality function $H(\mathcal{A}_t, p)$ which corresponds to the duality function in (1.52) after the identification $\mu(\{1\}) = p$. (Here we have slightly rephrased things compared to the different conventions in [BCH18], where 0 denotes the fitter type and y is the frequency of the unfit type.) In [BCH18, Lemma 4.1], they show that $H(\mathcal{A}_t, p)$ can be calculated by concatenating the higher-level maps $\check{\gamma}[\boldsymbol{\omega}_\mathbf{i}]$ with $\mathbf{i} \in \mathbb{S}_t$. For example, the equation in [BCH18, Lemma 4.1 (4)] can be rewritten in terms of $p = 1 - y$ as $p = \widehat{\mathtt{cob}}(p_1, p_2, p_3)$ with $\widehat{\mathtt{cob}}$ as in (1.84).

In [BCH18, Section 2.5], it is shown that the ASG $\mathcal{A}_t$ can be simplified a lot, while retaining all information necessary to calculate the duality function $H(\mathcal{A}_t, p)$. This is done in two steps.

In the first step, the ASG is *pruned*. This is a process in which parts of the tree that are irrelevant for the map G_t are cut off. In particular, if the function G_t is constant, then the pruned G_t consists of a single edge ending in one of the maps dth or bth. In the remaining case, the pruned ASG is a finite tree where each branch point is marked with one of the maps cob and bra.

In the second step, the pruned ASG is *stratified*. The tree is divided into regions, which are the connected components after cutting the tree at the branch points marked with cob. These regions are replaced by single nodes marked by an integer indicating how many leaves they contain.

The result of this is a simplified process, the *stratified ASG* $\mathcal{T}_t$, which contains all necessary information about the ASG $\mathcal{A}_t$ in the sense that there exists a function $\mathcal{H}(\mathcal{T}_t, p)$ such that $\mathcal{H}(\mathcal{T}_t, p) = H(\mathcal{A}_t, p)$ [BCH18, Thm 2.22]. In particular, solutions of (2.3) can be represented as $p_t = \mathbb{E}[\mathcal{H}(\mathcal{T}_t, p_0)]$ ([BCH18, Thm 2.25] (compare (1.51)).

One can now check (compare Lemma 5.5 below) that $\rho_t := \mathbb{P}[\mathcal{H}(\mathcal{T}_t, p) \in \cdot]$ solves the higher-level mean-field equation with initial state $\rho_0 = \delta_p$, where we use the identification $\mathcal{P}(\{0, 1\}) \cong [0, 1]$. In [BCH18, Lemma 6.1], it is observed that $M_t := \mathcal{H}(\mathcal{T}_t, p_0)$ is a bounded sub- or supermartingale for each $p_0 \in [0, 1]$ and hence converges to an a.s. limit $\mathcal{H}_\infty(p_0)$. In [BCH18, Thm 2.27], it is proved that if p_0 is not an unstable fixed point of (2.3), then $\mathcal{H}_\infty(p_0)$ is a Bernoulli random variable with parameter $\lim_{t \to \infty} p_t$.

Our Propositions 1.16 and 1.17 imply that if p_0 is a fixed point of (2.3), then $\mathcal{H}_\infty(p_0)$ is a Bernoulli random variable if and only if the RTP corresponding to p_0 is endogenous. Thus, [BCH18, Thm 2.27] implies that for the model in (2.2), RTPs corresponding to stable fixed points are always endogenous. Since all stable fixed points of (2.3) are in fact lower or upper solutions, this alternatively also follows from our Proposition 1.20.

2.2 Mean-field limits

If N Markov processes interact in a way that is symmetric under permutations of the N coordinates, then it is frequently possible to obtain a nontrivial limit as $N \to \infty$. Such limits are generally called *mean-field limits*. In the mean-field limit, the individual processes behave asymptotically independently, but with transition probabilities that depend on the average behavior of all processes. For systems of interacting diffusions, this principle was demonstrated by McKean in his analysis of the Vlasov equation [McK66]. Consequently, mean-field limits are also called McKean-Vlasov limits. There exists an extensive literature on the topic. Most work has focused on interacting diffusions, but jump processes have also been studied [ST85, ADF18]. An elementary introduction to mean-field limits for interacting particle systems is given in [Swa17, Chapter 3].

In a biological setting, well-mixing populations converge in the mean-field limit to the solution of a deterministic ODE. Similarly, spatial populations with strong local mixing can be expected to converge, after an appropriate rescaling, to the solution of a determinstic PDE. For interacting particle systems whose dynamics have an exclusion process component with a large rate, this intuition was made rigorous by De Masi, Ferrari and Lebowitz [DFL86, Thm 2]. They state their theorem only for processes whose state space S consists of two points, and only prove the theorem for one particular one-dimensional example, but sketch how the proof should be adapted to the general case. In [DN94, Thm 1], a version of the theorem is stated where S can be any finite set; it is claimed that the proof is again the same.

In our running example of a particle system with cooperative branching and deaths, the limiting PDE takes the form

$$\tfrac{\partial}{\partial t} p_t(x) = \tfrac{1}{2} \tfrac{\partial^2}{\partial x^2} p_t(x) + \alpha p_t(x)^2 \big(1 - p_t(x)\big) - p_t(x) \qquad (t \geq 0, \ x \in \mathbb{R}). \tag{2.5}$$

This PDE was used in [Nob92] to derive asymptotic properties of the associated spatial particle system with strong mixing. We can view (2.5) as a spatial version of the ODE (1.36); in particular, if $p_0(x) = p_0$ does not depend on x, then $p_t = p_t(x)$ solves (1.36).

The intuition behind (2.5), and more general PDEs of this type, is easily explained. In the strong mixing limit, the genealogy of a single site should be described by a branching process as in Figure 1 where in addition, each particle has a position in $\mathbb{R}$, which moves according to an independent Brownian motion. Convergence to the PDE should then follow from, on the one hand, convergence of the genealogies to a system of branching Brownian motions with random maps attached to their branching events, and, on the other hand, a representation in the spirit of Theorem 1.6 of solutions of the PDE (2.5) in terms of such a system of branching Brownian motions.

The proof of [DFL86, Thm 2] is indeed based on this sort of dual approach, although one would wish that they had given a more explicit statement of the stochastic representation of solutions of their general PDE. Our proof of Theorem 1.5 follows the same strategy, i.e., we first prove the stochastic representation of solutions to the mean-field equation (Theorem 1.6) and then use this to prove our convergence result (Theorem 1.5).

2.3 Open problems

In the present paper, we have adapted results from [AB05, MSS18] about discrete-time Recursive Tree Processes and endogeny to the continuous-time setting, and applied our general results on a concrete system with cooperative branching and deaths. Among other things, we proved that for $\alpha > 4$, the RTPs corresponding to ν_{low} and ν_{upp} are endogenous but the RTP corresponding to ν_{mid} is not. The proof was based on an analysis of the bivariate mean-field equation. Here, it was convenient to be able to analyse a differential equation, as an analysis of the associated discrete-time bivariate evolution would have been possible, but more messy.

Our work leaves a number of questions unanswered, both in the general setting and more specifically for our running example with $\mathcal{G} := \{\mathtt{cob}, \mathtt{dth}\}$ and π as in (1.14). Concerning the latter, we pose the following questions.

Open Problem 1 Not every measure $\mu^{(n)} \in \mathcal{P}_{\mathrm{sym}}(\{0,1\}^n)$ is the n-th moment measure of a measure $\rho \in \mathcal{P}(\mathcal{P}(\{0,1\}))$. Determine all symmmetric solutions of the n-variate RDE, for general $n \geq 3$, and their domains of attraction.

Open Problem 2 Same as Open Problem 1 but without the symmetry assumption and for general $n \geq 2$.

Open Problem 3 Prove that apart from the atom at zero, the law $\underline{\nu}_{\mathrm{mid}}$, viewed as a probability law on $[0,1]$, has a smooth density with respect to the Lebesgue measure.

Open Problem 4 Determine the asymtotics of the distribution function F of $\underline{\nu}_{\mathrm{mid}}$ near 0 and 1.

Open Problem 5 For the more general model in (2.2), is it true that unstable fixed points of the mean-field equation that separate the domains of attraction of two stable fixed points correspond to nonendogenous RTPs? Is the picture for the higher-level RDE the same?

Partly inspired by our concrete example, we ask the following problems in the general setting.

Open Problem 6 Can (1.3) be relaxed to allow for branching processes $(\nabla \$_t)_{t \geq 0}$ that are nonexplosive but have infinite mean?

Question 7 Are there general results linking the (in)stability of fixed points of the mean-field equation to (non)endogeny of the related RTP?

Question 8 In our example, the higher-level RDE has two solutions $\underline{\nu}_{\mathrm{mid}}$ and $\overline{\nu}_{\mathrm{mid}}$ with mean ν_{mid}, of which the former is stable and the latter is unstable. Is this a general phenomenon in the nonendogenous case? Can one prove nonendogeny of an RTP corresponding to a solution ν of the RDE by showing that $\overline{\nu}$ is unstable?

Question 9 Are there examples of higher-level RDEs that have solutions $\rho \notin \{\underline{\nu}, \overline{\nu}\}$?

Open Problem 10 Is the higher-level RTP $(\boldsymbol{\omega}_{\mathbf{i}}, \xi_{\mathbf{i}})_{\mathbf{i} \in \mathbb{S}}$ from Proposition 1.17 always endogenous?

We note that Jon Warren [War06] has given a partial answer to Question 8 in the special case of RTPs on the binary tree with the property that $\gamma[\omega](x_1, x_2)$ is a symmetric function of x_1 and x_2 for each ω. Warren's methods are closely related to the recent work of Johnson, Podder, and Skerman [JPS20], who characterize endogeny in terms of a multitype branching process of pivotal vertices.

2.4 Outline of the proofs

In the remainder of the paper, we prove all results stated so far, except for Theorems 1.11 and 1.15 as well as Proposition 1.17, which we cite from [MSS18, Thm 1, Thm 13, and Prop 4].

In Section 3 we prove Theorem 1.1, Propositions 1.2 and 1.3, and Lemma 1.4, which state elementary properties of solutions of the mean-field equations (1.22) and (1.4), as well as Theorem 1.6, which gives a stochastic representation of solutions of the mean-field equation in terms of finite recursive tree processes. In Section 4, we use this stochastic representation to prove Theorem 1.5 about convergence of finite systems to a solution of the mean-field equation.

In Section 5, we prove our main results about RTPs with continuous time, which are largely analogous to known results from the discrete-time setting. Basic results are Lemma 1.8 and Proposition 1.10, as well as Lemma 1.9 which deals with discrete time and is a slight reformulation of known results. Following [AB05], Theorem 1.12 links the n-variate equation to endogeny, while Propositions 1.14 and 1.16 are concerned with the higher-level equation, and closely follow ideas from [MSS18].

In Section 6 we prove some additional results about RTPs, first Proposition 1.20, which generalizes [AB05, Lemma 15] and shows that upper and lower solutions of a monotonous RDE are always endogenous, and then Lemmas 1.21, 1.22, 1.24, and 1.25 which give conditions for uniqueness in a general setting and then more specifically for monotone systems.

In Section 7, finally, we have collected all proofs that deal specifically with our running example of a system with cooperative branching and deaths. The first such result is Proposition 1.13 about the bivariate equation, which is a two dimensional ODE for which by elementary means we find all fixed points and their domains of attraction. By combining Proposition 1.13 with ideas involving the convex order we then prove the much stronger Theorem 1.18 which gives all fixed points and domains of attraction for the higher-level equation. The picture is then completed by the proofs of Lemma 1.19, which gives some properties of the nontrivial fixed point of the higher-level equation, as well as Lemmas 1.23 and 1.26 which illustrate ideas from Section 6 in the concrete set-up of our example.

3 The mean-field equation

In this section, we prove Theorems 1.1 and 1.6, which state that the mean-field equation (1.4) has a unique solution and can be represented in terms of a random tree generated by a branching process, with random maps attached to its vertices. In addition, we also prove Propositions 1.2 and 1.3, as well as Lemma 1.4.

In Subsection 3.1, we start with some preliminaries, showing, in particular, that the integral in (1.16) is well-defined, and Lemma 1.4, which says that mean-field equations of the form (1.22) can be rewritten in the simpler form (1.4).

Next, in Subsection 3.2, we prove uniqueness of solutions of (1.4), which yields the uniqueness parts of Theorem 1.1. To prove existence, in Subsection 3.3, we show that the right-hand side of (1.49) solves (1.4), which not only completes the proof of Theorem 1.1 but also yields the stochastic representation that is Theorem 1.6.

The proofs of Propositions 1.2 and 1.3, finally, can be found in Subsection 3.4.

3.1 Preliminaries

Recall that we interpret the mean-field equation (1.4) as in (1.16), where, by (1.12),

$$\langle T_{\gamma[\omega]}(\mu), \phi \rangle = \int_S \mu(\mathrm{d}x_1) \cdots \int_S \mu(\mathrm{d}x_k)\, \phi\big(\gamma[\omega](x_1, \ldots, x_k)\big) \qquad (\omega \in \Omega_k,\ k \geq 1). \tag{3.1}$$

Since by assumption, $\gamma[\omega](x_1, \ldots, x_k)$ is jointly measurable in ω and $x_1, \ldots, x_k$, the right-hand side of (3.1) is measurable as a function of ω and hence the integral in (1.16) is well-defined.

Proof of Lemma 1.4. Recall from Subsection 1.3 that the basic ingredients that go into the equation (1.22) are the measure space $(\Omega', \mathbf{q})$ and function λ, as well as, for each $\omega \in \Omega'$ and $1 \leq i \leq \lambda(\omega)$, the function $\gamma_i[\omega]$ and set $K_i(\omega)$. Also, $\kappa_i(\omega) := |K_i(\omega)|$. In terms of these basic ingredients we need to define $\Omega, \mathbf{r}, \kappa$, and γ as in Subsection 1.1 so that (1.22) takes the simpler form (1.4).

Since we want to replace the integral and sum in (1.22) by a single integral, we put

$$\Omega := \bigcup_{l=1}^{\infty} \Omega'_l \times [l], \tag{3.2}$$

where as before $\Omega'_l := \{\omega \in \Omega' : \lambda(\omega) = l\}$ and $[l] := \{1, \ldots, l\}$, and we equip Ω with the measure

$$\mathbf{r}\big(A \times \{k\}\big) := \mathbf{q}(A) \qquad (A \subset \Omega'_l \text{ measurable},\ 1 \leq k \leq l). \tag{3.3}$$

In general, Ω need not be a Polish space, as required in Subsection 1.1. We will fix this problem at the end of our proof, but for the sake of the presentation we neglect it for the moment being. We define $\kappa : \Omega \to \mathbb{N}$ as in Subsection 1.1 by $\kappa(\omega, i) := \kappa_i(\omega)$, where the right-hand side is the function from Subsection 1.3. We write

$$K_i(\omega) = \big\{j_1, \ldots, j_{\kappa_i(\omega)}\big\} \quad \text{with} \quad j_1 < \cdots < j_{\kappa_i(\omega)}. \tag{3.4}$$

Since $\gamma_i[\omega](x_1, \ldots, x_{\lambda(\omega)})$ depends only on coordinates in $K_i(\omega)$, there exists a function $\gamma[\omega, i] : S^{\kappa(\omega, i)} \to S$ such that

$$\gamma_i[\omega](x_1, \ldots, x_{\lambda(\omega)}) = \gamma[\omega, i](x_{j_1}, \ldots, x_{j_{\kappa(\omega,i)}}) \qquad (x \in S^{\lambda(\omega)}). \tag{3.5}$$

Note that $T_{\gamma_i[\omega]} = T_{\gamma[\omega,i]}$ by (1.12). As in (1.2), we can associate $\gamma[\omega, i]$ with a function that is defined on $S^{\mathbb{N}_+}$ but depends only on the first $\kappa(\omega, i)$ coordinates. We take this as our definition of the function $\gamma : \Omega \times S^{\mathbb{N}_+} \to S$ from Subsection 1.1. It follows from (1.20) that $\gamma[\omega, i](x)$ is jointly measurabe as a function of (ω, i) and x.

Replacing the integral and sum in (1.22) by a single integral over $\mathbf{r}$ as defined in (3.3), using the fact that $T_{\gamma_i[\omega]} = T_{\gamma[\omega,i]}$ we see that (1.22) can be rewritten as

$$\frac{\partial}{\partial t}\mu_t = \int_\Omega \mathbf{r}(\mathrm{d}\omega)\big\{T_{\gamma[\omega,i]}(\mu_t) - \mu_t\big\}, \tag{3.6}$$

which coincides with (1.4). The condition that $\mathbf{r}$ should be a finite measure translates to (1.23) (i), while the condition (1.3), written in terms of $\mathbf{q}$, becomes (1.23) (ii). Moreover, if $\mathbf{q}$ satisfies (1.24), then $\mathbf{r}$ satisfies (1.19).

We still have to fix the problem that Ω, as defined in (3.2), is in general not a Polish space. There are several possible ways to fix this.[2] The solution we will choose is to replace Ω by the Polish space

$$\overline{\Omega} := \bigcup_{l=1}^{\infty} \overline{\Omega'_l} \times [l], \tag{3.7}$$

were $\overline{\Omega'_l}$ denotes the closure of Ω'_l in Ω'. We view $\mathbf{r}$ as a measure on $\overline{\Omega}$ that is concentrated on Ω and extend κ and γ in a measurable way to the larger space, which is possible since Ω is a measurable subset of $\overline{\Omega}$. Since $\mathbf{r}$ is concentrated on Ω, it does not matter how we extend κ and γ as this has no effect on (3.6). $\qquad\square$

3.2 Uniqueness

In the present section, we prove that under the assumption (1.3), solutions to (1.4) are unique, which settles the uniqueness part of Theorem 1.1.

Below, we let $\mathcal{M}(S)$ denote the space of all finite signed measures on S. The total variation norm has already been mentioned several times. There are two conventional definitions, which differ by a factor 2. We will use the definition

$$\|\mu\| := \tfrac{1}{2} \sup_{|f| \leq 1} \left| \int f \, \mathrm{d}\mu \right| \qquad \left(\mu \in \mathcal{M}(S)\right), \tag{3.8}$$

where the supremum runs over all measurable functions $f : S \to [-1,1]$. If X, Y are S-valued random variables, then it is easy to see that $\|\mu - \nu\| \leq \mathbb{P}[X \neq Y]$. Conversely, it is well-known [Lin92, page 19] that if $\mu, \nu \in \mathcal{P}(S)$, then it is possible to couple S-valued random variables X, Y in such a way that

$$\|\mu - \nu\| = \mathbb{P}[X \neq Y]. \tag{3.9}$$

Lemma 3.1 (Lipschitz continuity)**.** *Let* $g : S^k \to S$ *be measurable and let* T_g *be defined as in (1.12). Then*

$$\left\| T_g(\mu) - T_g(\nu) \right\| \leq k \|\mu - \nu\| \qquad \left(\mu, \nu \in \mathcal{P}(S)\right). \tag{3.10}$$

Moreover, if T *is defined as in (1.1), then*

$$\left\| T(\mu) - T(\nu) \right\| \leq \left(|\mathbf{r}|^{-1} \int_{\Omega} \mathbf{r}(\mathrm{d}\omega) \, \kappa(\omega) \right) \|\mu - \nu\| \qquad \left(\mu, \nu \in \mathcal{P}(S)\right). \tag{3.11}$$

Proof. By (3.9) we can find an S^2-valued random variable (X, Y) such that $\|\mu - \nu\| = \mathbb{P}[X \neq Y]$. Let $(X_1, Y_1), \ldots, (X_k, Y_k)$ be i.i.d. copies of (X, Y). Then, by (1.12),

$$\begin{aligned}
\left\| T_g(\mu) - T_g(\nu) \right\| &= \left\| \mathbb{P}\big[g(X_1, \ldots, X_k) \in \cdot\,\big] - \mathbb{P}\big[g(Y_1, \ldots, Y_k) \in \cdot\,\big] \right\| \\
&\leq \mathbb{P}\big[g(X_1, \ldots, X_k) \neq g(Y_1, \ldots, Y_k)\big] \\
&\leq \sum_{i=1}^{k} \mathbb{P}[X_i \neq Y_i] = k \|\mu - \nu\|.
\end{aligned} \tag{3.12}$$

This proves (3.10). Formula (3.11) follows by integrating over ω. $\qquad\square$

---

[2]For example, we can strengthen our assumptions on λ in the sense that $\{\omega : \lambda(\omega) = l\}$ is a G_δ-set for each $l \in \mathbb{N}_+$, or we can relax our assumptions on Ω allowing it to be a Lusin space, instead of just a Polish space, throughout.

Our next lemma gives equivalent formulations of the mean-field equation (1.4), that will also be useful in the next subsection where we prove existence of solutions. Below, we interpret an integral of a measure-valued integrand in the usual way, i.e., $\int_0^t \mathrm{d}s \, \mu_s$ denotes the measure defined by

$$\left\langle \int_0^t \mathrm{d}s \, \mu_s, \phi \right\rangle := \int_0^t \mathrm{d}s \, \langle \mu_s, \phi \rangle \tag{3.13}$$

for any bounded measurable $\phi : S \to \mathbb{R}$. We say that a class Φ of functions is *measure-determining* if two finite measures μ, ν that satisfy $\langle \mu, \phi \rangle = \langle \nu, \phi \rangle$ for all $\phi \in \Phi$ must be equal.

Lemma 3.2 (Equivalent formulations of the mean-field equation). *Assume (1.3). Let Φ be a measure-determining class of bounded measurable functions $\phi : S \to \mathbb{R}$. Let $[0, \infty) \ni t \mapsto \mu_t \in \mathcal{P}(S)$ be measurable. Then of the following conditions, (i) implies (ii) and (iii). If $[0, \infty) \ni t \mapsto \mu_t \in \mathcal{P}(S)$ is continuous with respect to the total variation norm, then all three conditions are equivalent.*

(i) *For each $\phi \in \Phi$, the function $t \mapsto \langle \mu_t, \phi \rangle$ is continuously differentiable and*
$\frac{\partial}{\partial t} \langle \mu_t, \phi \rangle = |\mathbf{r}| \{ \langle T(\mu_t), \phi \rangle - \langle \mu_t, \phi \rangle \} \ (t \geq 0).$

(ii) $\mu_t = \mu_0 + |\mathbf{r}| \int_0^t \mathrm{d}s \, \{ T(\mu_s) - \mu_s \} \ (t \geq 0).$

(iii) $\mu_t = e^{-|\mathbf{r}|t} \mu_0 + |\mathbf{r}| \int_0^t \mathrm{d}s \, e^{-|\mathbf{r}|s} T(\mu_{t-s}) \ (t \geq 0).$

Proof. Integrating the equation in (i) from time 0 until time t, we see that (i) implies $\langle \mu_t, \phi \rangle = \langle \mu_0, \phi \rangle + |\mathbf{r}| \int_0^t \mathrm{d}s \, \{ \langle T(\mu_s), \phi \rangle - \langle \mu_s, \phi \rangle \} \ (t \geq 0)$ for all $\phi \in \Phi$. Since Φ is measure-determining, this must hold for all bounded measurable ϕ, proving the implication (i)$\Rightarrow$(ii). Also, we can equivalently write the equation in (i) as

$$\frac{\partial}{\partial t} \left(e^{|\mathbf{r}|t} \langle \mu_t, \phi \rangle \right) = |\mathbf{r}| e^{|\mathbf{r}|t} \langle T(\mu_t), \phi \rangle \qquad (t \geq 0). \tag{3.14}$$

Integrating from time 0 until time t and using that Φ is measure-determining now yields

$$e^{|\mathbf{r}|t} \mu_t = \mu_0 + |\mathbf{r}| \int_0^t \mathrm{d}s \, e^{|\mathbf{r}|s} T(\mu_s). \tag{3.15}$$

Multiplying by $e^{-|\mathbf{r}|t}$ and substituting $s \mapsto t - s$ in the integral then yields the equation in (iii).

If $t \mapsto \mu_t \in \mathcal{P}(S)$ is continuous with respect to the total variation norm, then Lemma 3.1 together with (1.3) imply that also $t \mapsto T(\mu_t) \in \mathcal{P}(S)$ is continuous with respect to the total variation norm. It follows that $t \mapsto \langle \mu_t, \phi \rangle$ and $t \mapsto \langle T(\mu_t), \phi \rangle$ are continuous for each bounded measurable $\phi : S \to \mathbb{R}$. As a result, the right-hand side of (ii), integrated against any bounded measurable ϕ, is continuously differentiable as a function of t, and (ii) implies (i). By the same argument, rewriting (iii) as (3.15) and differentiating, we see that (iii) implies (i). $\qquad \square$

We now prove the promised uniqueness of solutions to (1.4). Proposition 1.2, which will be proved in Subsection 3.4 below, shows that the constant L from (3.17) is not optimal and can be replaced by the constant K from (1.18).

Lemma 3.3 (Uniqueness). *Let Φ be a measure-determining class of bounded measurable functions $\phi : S \to \mathbb{R}$. Let $(\mu_t)_{t \geq 0}$ and $(\nu_t)_{t \geq 0}$ be functions taking values in the probability measures on S such that for each $\phi \in \Phi$, the functions $t \mapsto \langle \mu_t, \phi \rangle$ and $t \mapsto \langle \nu_t, \phi \rangle$ are continuously differentiable and solve (1.16). Then*

$$\|\mu_t - \nu_t\| \leq e^{Lt} \|\mu_0 - \nu_0\| \qquad (t \geq 0), \tag{3.16}$$

where

$$L := \int_\Omega \mathbf{r}(\mathrm{d}\omega)\, \big(\kappa(\omega) + 1\big). \tag{3.17}$$

Proof. Equation (ii) of Lemma 3.2 implies that

$$
\begin{aligned}
\|\mu_t - \nu_t\| &\leq \|\mu_0 - \nu_0\| + |\mathbf{r}| \left\| \int_0^t \mathrm{d}s\, \{T(\mu_s) - \mu_s\} - \int_0^t \mathrm{d}s\, \{T(\nu_s) - \nu_s\} \right\| \\
&\leq \|\mu_0 - \nu_0\| + |\mathbf{r}| \int_0^t \mathrm{d}s\, \|T(\mu_s) - T(\nu_s)\| + |\mathbf{r}| \int_0^t \mathrm{d}s\, \|\mu_s - \nu_s\| \\
&\leq \|\mu_0 - \nu_0\| + L \int_0^t \mathrm{d}s\, \|\mu_s - \nu_s\|,
\end{aligned}
\tag{3.18}
$$

where $L = |\mathbf{r}| + \int_\Omega \mathbf{r}(\mathrm{d}\omega)\, \kappa(\omega)$ using (3.11) of Lemma 3.1. The claim now follows from Gronwall's lemma [EK86, Thm A.5.1]. $\square$

3.3 The stochastic representation

In this section, we prove the following proposition, that settles the existence part of Theorem 1.1. Together with Lemma 3.3, this completes the proof of Theorem 1.1 and at the same time also proves Theorem 1.6.

We work in our usual set-up where S and Ω are Polish spaces, $\kappa : \Omega \to \mathbb{N}$ is measurable, γ is as in Subsection 1.1, and $\mathbf{r}$ is a nonzero finite measure on Ω satisfying (1.3). We fix $\mathbb{T}$ as in Section 1.4 and let $(\boldsymbol{\omega_i})_{\mathbf{i} \in \mathbb{T}}$ be i.i.d. with common law $|\mathbf{r}|^{-1}\mathbf{r}$. We let $(\sigma_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ be an independent i.i.d. collection of exponentially distributed random variables with mean $|\mathbf{r}|^{-1}$ and define $\$$, $\$_t$, $\nabla\$_t$, and G_t as in (1.43), (1.44), and (1.47).

Proposition 3.4 (Recursive tree representation)**.** *For any $\mu_0 \in \mathcal{P}(S)$, setting*

$$\mu_t := \mathbb{E}\big[T_{G_t}(\mu_0)\big] \qquad (t \geq 0) \tag{3.19}$$

defines a solution $(\mu_t)_{t \geq 0}$ to the mean-field equation (1.4). Moreover, $[0, \infty) \ni t \mapsto \mu_t$ is continuous with respect to the total variation norm.

To prepare for the proof of Proposition 3.4, we need one lemma. Recall that $|\mathbf{i}|$ denotes the length of a word $\mathbf{i}$, i.e., $|i_1 \cdots i_n| := n$. Let

$$\$_{t,(n)} := \$_t \cap \$_{(n)} \quad \text{with} \quad \$_{(n)} := \{\mathbf{i} \in \$: |\mathbf{i}| < n\} \qquad (n \geq 1). \tag{3.20}$$

Fix $\mu_0 \in \mathcal{P}(S)$ and using notation as in (1.46), set

$$\mu_{t,(n)} := \mathbb{E}\big[T_{G_{t,(n)}}(\mu_0)\big] \quad \text{with} \quad G_{t,(n)} := G_{\$_{t,(n)}} \qquad (t \geq 0, n \geq 1), \tag{3.21}$$

and set $\mu_{t,(0)} := \mu_0$ $(t \geq 0)$. The following lemma is a "cut-off" version of Proposition 3.4.

Lemma 3.5 (Representation with cut-off)**.** *The measures $(\mu_{t,(n)})_{t \geq 0}$ defined in (3.21) satisfy*

$$\mu_{t,(n)} = e^{-|\mathbf{r}|t}\mu_0 + |\mathbf{r}| \int_0^t \mathrm{d}s\, e^{-|\mathbf{r}|s}\, T(\mu_{t-s,(n-1)}) \qquad (n \geq 1,\ t \geq 0). \tag{3.22}$$

Proof. Let $(Y_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ be i.i.d. with common law μ_0, independent of $(\boldsymbol{\omega_i}, \sigma_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$. Set

$$X_\mathbf{i} := Y_\mathbf{i} \qquad (\mathbf{i} \in \nabla\$_{t,(n)}), \tag{3.23}$$

and define $(X_\mathbf{i})_{\mathbf{i} \in \$_{t,(n)}}$ inductively by

$$X_\mathbf{i} := \gamma[\boldsymbol{\omega_i}](X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}k_\mathbf{i}}) \qquad (\mathbf{i} \in \$_{t,(n)}). \tag{3.24}$$

Then $X_\varnothing = G_{t,(n)}\big((X_{\mathbf{i}})_{\mathbf{i} \in \nabla \mathbb{S}_{t,(n)}}\big)$ and hence, in the same way as (1.49) is equivalent to (1.51),

$$\mu_{t,(n)} = \mathbb{P}[X_\varnothing \in \cdot]. \tag{3.25}$$

Conditioning on $\sigma_\varnothing$ and then also on $\boldsymbol{\omega}_\varnothing$, we see that

$$
\begin{aligned}
\mathbb{P}[X_\varnothing \in \cdot] &= \int_0^\infty \mathbb{P}\big[\sigma_\varnothing \in \mathrm{d}s\big]\, \mathbb{P}\big[X_\varnothing \in \cdot \,\big|\, \sigma_\varnothing = s\big] \\
&= \int_t^\infty |\mathbf{r}| e^{-|\mathbf{r}|s}\, \mathrm{d}s\, \mu_0 + \int_0^t |\mathbf{r}| e^{-|\mathbf{r}|s}\, \mathrm{d}s\, P\big[X_\varnothing \in \cdot \,\big|\, \sigma_\varnothing = s\big] \\
&= e^{-|\mathbf{r}|t} \mu_0 + \int_0^t |\mathbf{r}| e^{-|\mathbf{r}|s}\, \mathrm{d}s\, |\mathbf{r}|^{-1} \int_\Omega \mathbf{r}(\mathrm{d}\omega)\, P\big[X_\varnothing \in \cdot \,\big|\, \sigma_\varnothing = s,\, \boldsymbol{\omega}_\varnothing = \omega\big],
\end{aligned}
\tag{3.26}
$$

where we have used that $\boldsymbol{\omega}_\varnothing$ is independent of $\sigma_\varnothing$ with law $|\mathbf{r}|^{-1}\mathbf{r}$. We see from this that

$$
\begin{aligned}
\mathbb{P}[X_\varnothing \in \cdot] &= e^{-|\mathbf{r}|t}\mu_0 + \int_0^t e^{-|\mathbf{r}|s}\, \mathrm{d}s \int_\Omega \mathbf{r}(\mathrm{d}\omega)\, P\big[\gamma[\omega](X_1,\ldots,X_{\kappa(\omega)}) \in \cdot \,\big|\, \sigma_\varnothing = s,\, \boldsymbol{\omega}_\varnothing = \omega\big] \\
&= e^{-|\mathbf{r}|t}\mu_0 + \int_0^t e^{-|\mathbf{r}|s}\, \mathrm{d}s\, |\mathbf{r}|\, T(\mu_{t-s,(n-1)}),
\end{aligned}
\tag{3.27}
$$

where we have used (1.1) and the observation that conditional on $\sigma_\varnothing = s$ and $\boldsymbol{\omega}_\varnothing = \omega$, the random variables $X_1,\ldots,X_{\kappa(\omega)}$ are i.i.d. with common law $\mu_{t-s,(n-1)}$. $\qquad\square$

Proof of Proposition 3.4. The condition (1.3) guarantees that $(\nabla \mathbb{S}_t)_{t \geq 0}$ is a finite mean branching process; more precisely, by standard theory,

$$\mathbb{E}\big[|\nabla \mathbb{S}_t|\big] = e^{Kt} \quad \text{with} \quad K := \int_\Omega \mathbf{r}(\mathrm{d}\omega)\, (\kappa(\omega) - 1). \tag{3.28}$$

Fix $\mu_0 \in \mathcal{P}(S)$ and define μ_t and $\mu_{t,(n)}$ as in (3.19) and (3.22). Then the total variation norm distance between these measures can be bounded by

$$\big\|\mu_{t,(n)} - \mu_t\big\| \leq \mathbb{P}\big[\mathbb{S}_{t,(n)} \neq \mathbb{S}_t\big], \tag{3.29}$$

which tends to zero as $n \to \infty$ since $\mathbb{S}_t$ is a.s. finite by (3.28). In fact, since $\mathbb{P}[\mathbb{S}_{s,(n)} \neq \mathbb{S}_s] \leq \mathbb{P}[\mathbb{S}_{t,(n)} \neq \mathbb{S}_t]$ for all $s \leq t$, we have that

$$\sup_{0 \leq s \leq t} \big\|\mu_{s,(n)} - \mu_s\big\| \underset{n \to \infty}{\longrightarrow} 0 \qquad \forall t < \infty. \tag{3.30}$$

Using this and the Lipschitz continuity of T with respect to the total variation norm (Lemma 3.1), we can let $n \to \infty$ in (3.22) to obtain

$$\mu_t = e^{-|\mathbf{r}|t}\mu_0 + |\mathbf{r}| \int_0^t \mathrm{d}s\, e^{-|\mathbf{r}|t}\, T(\mu_{t-s}) \qquad (t \geq 0). \tag{3.31}$$

Since

$$\big\|\mu_{t+\varepsilon} - \mu_t\big\| \leq \mathbb{P}\big[\mathbb{S}_{t+\varepsilon} \neq \mathbb{S}_t\big], \tag{3.32}$$

using the fact that the branching process $(\nabla \mathbb{S}_t)_{t \geq 0}$ a.s. does not jump at deterministic times, we see that $[0, \infty) \ni t \mapsto \mu_t$ is continuous with respect to the total variation norm. Using this and (3.31), we see from Lemma 3.2 that $(\mu_t)_{t \geq 0}$ solves the mean-field equation (1.4). $\qquad\square$

3.4 Continuity in the initial state

In this subsection, we prove Propositions 1.2 and 1.3.

Proof of Proposition 1.2. It follows from Theorem 1.6 and Lemma 3.1 that

$$
\begin{aligned}
\big\| T_t(\mu) - T_t(\nu) \big\| &= \big\| \mathbb{E}[T_{G_t}(\mu) - T_{G_t}(\nu)] \big\| \\
&\leq \mathbb{E}\big[\big\| \mathbb{E}[T_{G_t}(\mu) - T_{G_t}(\nu) \,|\, \mathcal{F}_t] \big\| \big] \leq \mathbb{E}\big[|\nabla\$_t| \,\|\mu - \nu\| \big] = e^{Kt} \|\mu - \nu\|,
\end{aligned}
\tag{3.33}
$$

where $\mathcal{F}_t$ is the filtration defined in (1.48). $\qquad\square$

Proposition 1.3 follows from the following two lemmas.

Lemma 3.6 (Continuity of T)**.** *Under the condition (1.19), the operator T in (1.1) is continuous w.r.t. the topology of weak convergence.*

Proof. If $\mu_n \in \mathcal{P}(S)$ converge weakly to a limit μ_∞, then by Skorohod's representation theorem there exists random variables X^n with laws μ_n that converge a.s. to a limit X^∞ with law μ_∞. Let $\big((X_i^n)_{n\in\mathbb{N}\cup\{\infty\}}\big)_{i\geq 1}$ be i.i.d. copies of such a sequence $(X^n)_{n\in\mathbb{N}\{\infty\}}$ and let $\boldsymbol{\omega}$ be an independent random variable with law $|\mathbf{r}|^{-1}\mathbf{r}$. Then by (1.19),

$$
\gamma[\boldsymbol{\omega}](X_1^n, \ldots, X_{\kappa(\boldsymbol{\omega})}^n) \underset{n\to\infty}{\longrightarrow} \gamma[\boldsymbol{\omega}](X_1^\infty, \ldots, X_{\kappa(\boldsymbol{\omega})}^\infty) \quad \text{a.s.}
\tag{3.34}
$$

and hence $T(\mu_n)$ converges weakly to $T(\mu_\infty)$ by (1.1). $\qquad\square$

Lemma 3.7 (Continuity in the initial state)**.** *Assume that the operator T in (1.1) is continuous w.r.t. the topology of weak convergence. Then the same is true for the operators T_t $(t \geq 0)$ defined in (1.6).*

Proof. We need to show that solutions of the mean-field equation (1.4) are continuous in their initial state, in the sense that if $(\mu_t^k)_{t\geq 0}$ $(k \in \mathbb{N} \cup \{\infty\})$ are started in initial states such that $\mu_0^k \Rightarrow \mu_0^\infty$, then $\mu_t^k \Rightarrow \mu_t^\infty$ for all $t \geq 0$.

To see this, inductively define $\mu_{t,(n)}^k$ as in (3.22) with μ_0 replaced by μ_0^k. Using the continuity of T, by induction, we see that $\mu_{t,(n)}^k \Rightarrow \mu_{t,(n)}^\infty$ as $k \to \infty$ for all $n \geq 1$ and $t \geq 0$. By (3.29), for each bounded continuous $\phi : S \to \mathbb{R}$, the quantity $\langle \mu_{t,(n)}^k, \phi \rangle$ converges to $\langle \mu_t^k, \phi \rangle$ uniformly in $k \in \mathbb{N} \cup \{\infty\}$, which allows us to conclude that $\langle \mu_t^k, \phi \rangle \to \langle \mu_t^\infty, \phi \rangle$ as $k \to \infty$ for all $t \geq 0$. $\qquad\square$

4 Approximation by finite systems

4.1 Main line of the proof

In this section, we prove Theorem 1.5. The basic idea, which already goes back to [DFL86], is that in the mean-field limit, the genealogy of a site converges to a branching process, and sites are independent in the limit. More precisely, consider n sites, sampled uniformly at random from $[N]$. To find out what their states are at time t, we follow the sites back until the last time when a random map is applied that has the potential to change the state of one of our sites. At this point, we stop following that given site but replace it by the sites that are relevant for the outcome of the map at the given site, and we continue in this way. When N is large, the new sites that are added in each step are with high probability sites we have not been following before, so that in the limit we obtain a branching process with random maps attached to its branch points. Making this idea precise yields the following proposition, that will be proved in Subsection 4.2 below.

Proposition 4.1 (State at sampled sites). *For each $N \in \mathbb{N}_+$ let $(X^{(N)}(t))_{t \geq 0}$ be a process as in Theorem 1.5 started in a deterministic initial state $X^{(N)}(0)$. Fix $t \geq 0$ and let T_t be defined as in (1.6) but with the mean-field equation (1.4) replaced by (1.22). Fix $n \geq 1$ and let $I_1, \ldots, I_n$ be i.i.d. uniformly distributed on $[N]$ and independent of $X^N(t)$. Then*

$$\left\| \mathbb{P}\big[\big(X^{(N)}_{I_1}(Nt), \ldots, X^{(N)}_{I_n}(Nt)\big) \in \cdot\,\big] - \underbrace{T_t(\mu_0^N) \otimes \cdots \otimes T_t(\mu_0^N)}_{n\ times} \right\| \xrightarrow[N \to \infty]{} 0, \qquad (4.1)$$

where $\|\cdot\|$ denotes the total variation norm, and the convergence in (4.1) is uniform w.r.t. the initial state $X^{(N)}(0)$.

Proposition 4.1 allows us to control the mean and variance of μ_{Nt}^N, which is enough to prove the convergence of μ_{Nt}^N to μ_t for fixed times t. To boost this up to pathwise convergence, we use the following lemma, that will be proved in Subsection 4.3 below.

Lemma 4.2 (Tightness in total variation). *For each $N \in \mathbb{N}_+$ let $(X^{(N)}(t))_{t \geq 0}$ be a process as in Theorem 1.5 started in a deterministic initial state $X^{(N)}(0)$, and let $\mu_t^N := \mu\{X^{(N)}(t)\}$ denote the empirical measure of $X^{(N)}(t)$. Then there exist random processes $(\tau_t^N)_{t \geq 0}$ such that $\tau^N : \mathbb{R} \to \mathbb{R}$ is a.s. nondecreasing with $\tau_0^N = 0$ and*

(i) $\mathbb{P}\big[\sup_{0 \leq t \leq T} |\tau_t^N - t| \geq \varepsilon\big] \xrightarrow[N \to \infty]{} 0 \quad (\varepsilon > 0,\ T < \infty),$

(ii) $\|\mu_{Nt}^N - \mu_{Ns}^N\| \leq L(\tau_t^N - \tau_s^N) \quad (0 \leq s \leq t)$ a.s.,

where $\|\cdot\|$ denotes the total variation norm and $L := \int_\Omega \mathbf{q}(\mathrm{d}\omega)\, \lambda(\omega)$.

In Subsection 4.4, we will derive Theorem 1.5 from Proposition 4.1, Lemma 4.2, and some abstract considerations.

4.2 The state at sampled sites

In this subsection we prove Proposition 4.1. We start with two preparatory lemmas.

Let $(\mathbf{X}_{s,t}^N)_{s \leq t}$ be the stochastic flow defined in (1.29), where we have made the dependence on N explicit. Let I be uniformly distributed on $[N]$ and independent of $(\mathbf{X}_{s,t}^N)_{s \leq t}$. For each $t \geq 0$, let $\tilde{M}_t^N : S^{[N]} \to S$ be defined as (note the factor N rescaling the speed of time):

$$\tilde{M}_t^N(x) := \mathbf{X}_{-Nt,0}^N(x)_I. \qquad (4.2)$$

Let $G_t : S^{\nabla \mathbb{S}_t} \to S$ be the random map defined in (1.47), where $\Omega, \mathbf{r}, \gamma$ and κ from Subsection 1.1 are defined in terms of the "ingredients" $\Omega', \mathbf{q}, \gamma_i[\omega]$ and $K_i(\omega)$ from Subsection 1.3, see the proof of Lemma 1.4 in Subsection 3.1. Let $(I_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ be i.i.d. uniformly distributed on $[N]$ and independent of $(\nabla \mathbb{S}_t, G_t)_{t \geq 0}$. For each $t \geq 0$, let $M_t^N : S^{[N]} \to S$ be defined as

$$M_t^N(x) := G_t\big((x_{I_\mathbf{i}})_{\mathbf{i} \in \nabla \mathbb{S}_t}\big). \qquad (4.3)$$

The following lemma says that for large N, the map in (4.2) can be approximated by the map in (4.3).

Lemma 4.3 (Coupling of maps). *For each $t \geq 0$, it is possible to couple the random maps $\tilde{M}_t^N$ and M_t^N with $N \in \mathbb{N}_+$ in such a way that*

$$\mathbb{P}\big[\tilde{M}_t^N \neq M_t^N\big] \xrightarrow[N \to \infty]{} 0. \qquad (4.4)$$

Proof. The essence of the proof can be summarized as follows: since for large N, sampling with or without replacement from $[N]$ is almost the same, the genealogy of a given site is approximately given by a branching process. In spite of this simple idea,

the proof is quite long, mainly because we have to take care of a lot of definitions, such as the way $\Omega, \mathbf{r}, \gamma$ and κ are defined in terms of $\Omega', \mathbf{q}, \gamma_i[\omega]$ and $K_i(\omega)$ in the proof of Lemma 1.4.

We start by recalling that the random map G_t from (1.47) can be seen as the concatenation of random maps assigned to the branch points of a branching process. We then embed this branching process in the set $[N]$ and prove that what we obtain is a good approximation for the genealogy of a given site.

We observe that in order to construct the map $G_t : S^{\nabla \mathbb{S}_t} \to S$ from (1.47), it suffices to know

$$\big(\mathbb{S}_t, (\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{S}_t}\big), \tag{4.5}$$

where $\mathbb{S}_t$ is defined in (1.44). Indeed, from the information in (4.5) we can determine $\nabla \mathbb{S}_t$, since

$$\nabla \mathbb{S}_t = \big\{\mathbf{i}k : \mathbf{i} \in \mathbb{S}_t, \ 1 \leq k \leq \kappa(\boldsymbol{\omega}_{\mathbf{i}}), \ \mathbf{i}k \notin \mathbb{S}_t\big\}, \tag{4.6}$$

and the map $G_t : S^{\nabla \mathbb{S}_t} \to S$ is obtained by concatenating the maps $\gamma[\boldsymbol{\omega}_{\mathbf{i}}]$ with $\mathbf{i} \in \mathbb{S}_t$ according to the tree structure of $\mathbb{S}_t$.

The object in (4.5) is in fact a Markov chain as a function of t. Starting from the initial state $\mathbb{S}_0 = \emptyset$ and $\nabla \mathbb{S}_0 = \{\varnothing\}$, its evolution is as follows: Independently for each $\mathbf{i} \in \nabla \mathbb{S}_t$, with rate $|\mathbf{r}|$, we add $\mathbf{i}$ to $\mathbb{S}_t$ and assign to it a value $\boldsymbol{\omega}_{\mathbf{i}}$ chosen according to the probability law $|\mathbf{r}|^{-1}\mathbf{r}$.

We will be interested in the process in (4.5) in the special case when $\Omega, \mathbf{r}, \kappa$, and γ are defined in terms of $\Omega', \mathbf{q}, \lambda, K_i(\omega)$, and $\gamma_i[\omega]$ as in the proof of Lemma 1.4. In this case, elements of Ω are pairs (ω, n) where $\omega \in \Omega'$ and $1 \leq n \leq \lambda(\omega)$, so we denote the process in (4.5) as

$$\big(\mathbb{S}_t, (\boldsymbol{\omega}_{\mathbf{i}}, n_{\mathbf{i}})_{\mathbf{i} \in \mathbb{S}_t}\big), \tag{4.7}$$

where $\boldsymbol{\omega}_{\mathbf{i}} \in \Omega'$ and $1 \leq n_{\mathbf{i}} \leq \lambda(\boldsymbol{\omega}_{\mathbf{i}})$. The set $\nabla \mathbb{S}_t$ is now given by

$$\nabla \mathbb{S}_t = \big\{\mathbf{i}k : \mathbf{i} \in \mathbb{S}_t, \ 1 \leq k \leq \kappa_{n_{\mathbf{i}}}(\boldsymbol{\omega}_{\mathbf{i}}), \ \mathbf{i}k \notin \mathbb{S}_t\big\}. \tag{4.8}$$

Defining $\mathbf{r}$ as in (3.3), the process in (4.5) now evolves in such a way that independently for each $\mathbf{i} \in \nabla \mathbb{S}_t$, with rate $|\mathbf{r}|$, we add $\mathbf{i}$ to $\mathbb{S}_t$ and assign values $(\boldsymbol{\omega}_{\mathbf{i}}, n_{\mathbf{i}})$ to it that are chosen according to the probability law $|\mathbf{r}|^{-1}\mathbf{r}$.

Let $\alpha \in [N]$ be fixed. Our next aim is to "embed" the process from (4.7) in the set $[N]$, in such a way that it approximates the genealogy of the site α. To this aim, we define, for each time, a random function $\psi_t^N : \mathbb{S}_t \cup \nabla \mathbb{S}_t \to [N]$. Initially, we set $\psi_0^N(\varnothing) := \alpha$. We let the function ψ_t^N evolve in a Markovian way together with the process in (4.7) in the following way. Recall that when we add an element $\mathbf{i}$ to $\mathbb{S}_t$ and assign values $(\boldsymbol{\omega}_{\mathbf{i}}, n_{\mathbf{i}})$ to it, this element is at the same time removed from $\nabla \mathbb{S}_t$ and replaced by new elements $\mathbf{i}1, \ldots, \mathbf{i}\kappa_{n_{\mathbf{i}}}(\boldsymbol{\omega}_{\mathbf{i}})$. We assign labels $\psi_t^N(\mathbf{i}k)$ $(k = 1, \ldots, \kappa_{n_{\mathbf{i}}}(\boldsymbol{\omega}_{\mathbf{i}}))$ to these new elements as follows. First, we choose $(I_l)_{l=1,\ldots,\lambda(\boldsymbol{\omega}_{\mathbf{i}})}$ in such a way that $I_{n_{\mathbf{i}}} := \psi_t^N(\mathbf{i})$ and

$$(I_l)_{l \neq n_{\mathbf{i}}} \quad \text{are i.i.d. uniformly chosen from } [N], \tag{4.9}$$

and next, we set $\psi_t^N(\mathbf{i}k) := I_{j_k}$, where as in (3.4), we order the elements of $K_{n_{\mathbf{i}}}(\boldsymbol{\omega}_{\mathbf{i}}) \subset \{1, \ldots, \lambda(\boldsymbol{\omega}_{\mathbf{i}})\}$ as

$$K_{n_{\mathbf{i}}}(\boldsymbol{\omega}_{\mathbf{i}}) = \big\{j_1, \ldots, j_{\kappa_{n_{\mathbf{i}}}(\boldsymbol{\omega}_{\mathbf{i}})}\big\} \quad \text{with} \quad j_1 < \cdots < j_{\kappa_{n_{\mathbf{i}}}(\boldsymbol{\omega}_{\mathbf{i}})}. \tag{4.10}$$

Note that this has the effect that if $n_{\mathbf{i}}$ is an element of $K_{n_{\mathbf{i}}}(\boldsymbol{\omega}_{\mathbf{i}})$, say $n_{\mathbf{i}} = j_k$, then the corresponding element $\mathbf{i}k$ gets the same label as $\mathbf{i}$, i.e., $\psi_t^N(\mathbf{i}k) = \psi_t^N(\mathbf{i})$. Otherwise, we assign new i.i.d. labels to all new elements of $\nabla \mathbb{S}_t$.

Using the function ψ_t^N that embeds the process in (4.7) in the set $[N]$, we define a function $\phi_t^N : S^N \to S^{\nabla \mathbb{S}_t}$ by

$$\phi_t^N(x)_{\mathbf{i}} := x_{\psi_t^N(\mathbf{i})} \qquad (\mathbf{i} \in \nabla \mathbb{S}_t). \tag{4.11}$$

We now consider the maps

$$\tilde{M}_t^N(x) := \mathbf{X}_{-Nt,0}(x)_\alpha \quad \text{and} \quad M_t^N(x) := G_t \circ \phi_t^N(x) \qquad (x \in S^N), \tag{4.12}$$

where $\alpha = \psi_0^N(\varnothing) \in [N]$ is the label initially assigned to the root. We claim that

$$\left\| \mathbb{P}\big[\tilde{M}_t^N \in \cdot \, \big] - \mathbb{P}\big[M_t^N \in \cdot \, \big] \right\| \xrightarrow[N \to \infty]{} 0, \tag{4.13}$$

where $\| \cdot \|$ denotes the total variation norm. In particular, if α is chosen uniformly distributed in $[N]$ and independent of everything else, then $(\psi_t^N(\mathbf{i}))_{\mathbf{i} \in \nabla \mathbb{S}_t}$ are i.i.d. uniformly distributed in $[N]$ and independent of the map G_t, so (4.13) implies (4.4).

To prove (4.13), we construct a process similar to the process in (4.7), together with an embedding in $[N]$, that describes the true genalogy of the site α, and show that the error we make by replacing this true genealogy by the process we had before is small. We denote this process as

$$\big(\tilde{\mathbb{S}}_t, (\boldsymbol{\omega}_{\mathbf{i}}, n_{\mathbf{i}})_{\mathbf{i} \in \tilde{\mathbb{S}}_t}, \tilde{\psi}_t^N \big). \tag{4.14}$$

At each time, $\nabla \tilde{\mathbb{S}}_t$ is defined in terms of this process in the same way as $\nabla \mathbb{S}_t$ is defined in (4.8). We also define $\tilde{G}_t$ and $\tilde{\phi}_t^N : S^N \to S^{\nabla \tilde{\mathbb{S}}_t}$ as before, i.e., $\tilde{G}_t$ is the concatenation of the random maps $\gamma_{n_{\mathbf{i}}}[\omega_{\mathbf{i}}]$ with $\mathbf{i} \in \tilde{\mathbb{S}}_t$ according to the tree structure of $\tilde{\mathbb{S}}_t$, and $\tilde{\phi}_t^N : S^N \to S^{\nabla \tilde{\mathbb{S}}_t}$ is defined in terms of $\tilde{\psi}_t^N$ as in (4.11).

Recall that $\tilde{M}_t^N(x) = \mathbf{X}_{-Nt,0}^N(x)_\alpha$. As for our previous process we start with $\tilde{\mathbb{S}}_0 = \emptyset$, $\nabla \tilde{\mathbb{S}}_0 = \{\varnothing\}$, and $\tilde{\psi}^N(\varnothing) = \alpha$. In Subsection 1.3, the stochastic flow $(\mathbf{X}_{s,t}^N)_{s \leq t}$ is constructed from a Poisson point set Π. We will construct the process in (4.14) in terms of Π in such a way that

$$\tilde{G}_t \circ \tilde{\phi}_t^N(x) = \tilde{M}_t^N(x) \qquad (t \geq 0), \tag{4.15}$$

which expresses the fact that the process in (4.14) describes the "true genealogy" of the site α.

The Poisson set Π consists of triples $(\omega, \mathbf{i}, t)$ which express the fact that at time t the random map $\vec{\gamma}[\omega]$ should be applied to the coordinates $\mathbf{i} = (i_1, \ldots, i_{\lambda(\omega)})$. Note that we are interested in $\mathbf{X}_{-Nt,0}^N(x)_\alpha$, which means that we look at negative times and need to rescale time by a factor N. For each $(\omega, \mathbf{i}, -Nt) \in \Pi$ and $\mathbf{j} \in \nabla \tilde{\mathbb{S}}_t$ such that $\tilde{\psi}_t^N(\mathbf{j}) = i_l$ for some $1 \leq l \leq \lambda(\omega)$, we update the process in (4.14) as follows:

(i) We remove $\mathbf{j}$ from $\nabla \tilde{\mathbb{S}}_t$ and add it to $\tilde{\mathbb{S}}_t$.

(ii) We set $\omega_{\mathbf{j}} := \omega$ and $n_{\mathbf{j}} := l$.

(iii) We add $\mathbf{j}1, \ldots, \mathbf{j}\kappa_l(\omega)$ to $\tilde{\mathbb{S}}_t$.

(iv) We define $\tilde{\psi}_t^N(\mathbf{j}k) := i_{j_k}$ $(k = 1, \ldots, \kappa_l(\omega))$, where $K_l(\omega) = \{j_1, \ldots, j_{\kappa_l(\omega)}\}$ as in (3.4).

It is straightforward to check that these rules guarantee that (4.15) holds and hence the process in (4.14) describes the true genealogy of the site α. As some more explanation, we can add the following: we follow a site β back in time till the first time when a map is applied that has the possibility to change the value of β. From that moment on, we follow back all sites that are relevant for the outcome of the map at β, and we number them according to the convention in (3.4). This defines a family structure, i.e., $\mathbf{i} = i_1 i_2 i_3$ is the i_3-th child of the i_2-th child of the i_1-th child of the original site α. The map $\tilde{\psi}_t^N$ applied to $\mathbf{i}$ tells us where this ancestor lives in the set $[N]$. There may be some overlap, i.e., it is possible that $\tilde{\psi}_t^N(\mathbf{i}) = \tilde{\psi}_t^N(\mathbf{j})$ for some $\mathbf{i}, \mathbf{j} \in \tilde{\mathbb{S}}_t \cup \nabla \tilde{\mathbb{S}}_t$. For $\mathbf{i}, \mathbf{j} \in \nabla \tilde{\mathbb{S}}_t$,

however, the probability that two ancestors live at the same site in $[N]$ tends to zero as $N \to \infty$, as we will see in a moment.

In view of (4.15), to prove (4.13), it suffices to prove that the Markov process in (4.14) is close in total variation distance to the process with $\tilde{\mathbb{S}}_t$ and $\tilde{\psi}_t^N$ replaced by $\mathbb{S}_t$ and ψ_t^N. Since the latter process is nonexplosive by (3.28), it suffices to prove convergence for the processes stopped at the first time when the cardinality of $\nabla \tilde{\mathbb{S}}_t$ resp. $\nabla \mathbb{S}_t$ exceeds a certain value, and then at the end send this value to infinity. We will prove convergence of the stopped processes in a number of steps, by making small changes in the jump rates. Here we use the fact that if the transition kernels of two continuous-time Markov chains are close in total variation norm, uniformly in the starting point, then by standard arguments the two processes can be coupled so that their laws at fixed time are close in total variation norm.

Let $\tilde{\Psi}_t^N := \tilde{\psi}_t^N(\nabla \tilde{\mathbb{S}}_t)$ denote the image of $\nabla \tilde{\mathbb{S}}_t$ under the map $\tilde{\psi}_t^N$. As a first step, we change the dynamics of the (stopped) process from (4.14) in such a way that elements $(\omega, \mathbf{i}, -Nt) \in \Pi$ have no effect if $\{i_1, \ldots, i_{\lambda(\omega)}\}$ intersects $\tilde{\Psi}_t^N$ in more that one point. Then the modified process is still Markovian; we claim the change in jump rates compared to the original process is of order N^{-1}. Indeed, for fixed l, if $i_1, \ldots, i_l$ are chosen uniformly without replacement from $[N]$, then the probability that one, resp. two or more of them lie in a set A of fixed cardinality is of order N^{-1} resp. N^{-2} as $N \to \infty$. Taking into account the fact that we rescale time by a factor N, as well as the summability condition (1.23) (i), this translates into a change in jump rates of order N^{-1} for the modified process, stopped at the first time when the cardinality of $\nabla \tilde{\mathbb{S}}_t$ exceeds a fixed value.

Recall that by (4.15), $\mathbf{X}_{-Nt,0}^N(x)_\alpha$ is a function only of $(x_\beta)_{\beta \in \tilde{\Psi}_t^N}$. The modified process we have just constructed has the property that $\tilde{\psi}_t^N : \nabla \tilde{\mathbb{S}}_t \to \tilde{\Psi}_t^N$ is a bijection, i.e., each element $\beta \in \tilde{\Psi}_t^N$ corresponds only to a single place $(\psi_t^N)^{-1}(\beta)$ in the family tree. The dynamics of the modified process can be described as follows:

(i) Independently for each $\beta \in \tilde{\Psi}_t^N$, with rates described by the measure $\mathbf{r}$ from (3.3), we choose a pair (ω, n) with $1 \le n \le \lambda(\omega)$.

(ii) If $\lambda(\omega) > N$, we do nothing.

(iii) Otherwise, we choose $(\beta_k')_{k=1,\ldots,\lambda(\omega)}$ such that $\beta_n' := \beta$ and $(\beta_k')_{k \ne n}$ are drawn from $[N] \setminus \{\beta\}$ without replacement.

(iv) If some of the $(\beta_k')_{k \ne n}$ are elements of $\tilde{\Psi}_t^N$, we do nothing.

(v) Otherwise, we remove β from $\tilde{\Psi}_t^N$ and add $(\beta_{j_k})_{1 \le k \le \kappa_n(\omega)}$ to $\tilde{\Psi}_t^N$, where $K_n(\omega) = \{j_1, \ldots, j_{\kappa_n(\omega)}\}$ with $j_1 < \cdots < j_{\kappa_n(\omega)}$.

(vi) If $\mathbf{j} = (\tilde{\psi}_{t-}^N)^{-1}(\beta)$ is the place of β in the family tree immediately prior to time t, then we assign to each new element of $\tilde{\Psi}_t^N$ a place in the family tree by setting $(\tilde{\psi}_{t-}^N)^{-1}(\beta_{j_k}) := \mathbf{j}k$.

Note that the measure $\mathbf{r}$ from (3.3) occurs naturally here, since each $\lambda(\omega)$-tuple of sites in $[N]$ can contain a given site β in $\lambda(\omega)$ different ways, as its 1st, 2nd,..., $\lambda(\omega)$-th member.

Removing the restrictions in points (ii) and (iv) above, and performing sampling without replacement instead of sampling with replacement in point (iii), we only make changes in the transition rates of order N^{-1}, and arrive at a process whose family tree evolves as the process in (4.7) and where to new members of the family tree, sites in $[N]$ are assigned chosen uniformly with replacement, as described by the process ψ_t^N. $\quad\square$

In the proof of Lemma 4.3, we have seen that in the mean-field limit $N \to \infty$, the genealogy of a single site can be approximated by a branching process with random maps attached to its branch points. Similarly, the genealogy of n randomly chosen sites can be approximated by n independent branching processes, which leads to the following extension of Lemma 4.3.

Lemma 4.4 (The genealogy of multiple sites). *Let $(\mathbf{X}_{s,t}^N)_{s\le t}$ be the stochastic flow defined in (1.29) and let $I_1,\ldots,I_n$ be i.i.d. uniformly distributed on $[N]$, independent of $(\mathbf{X}_{s,t}^N)_{s\le t}$. Let $\Omega, \mathbf{r}, \gamma$ and κ be defined in terms of $\Omega', \mathbf{q}, \gamma_i[\omega]$ and $K_i(\omega)$ as in the proof of Lemma 1.4. Fix $t\ge 0$ and let $(\nabla\mathbb{S}_t^i, G_t^i)$ $(i=1,\ldots,n)$ be i.i.d. copies of the random set and map defined in (1.44) and (1.47). Conditional on $(\nabla\mathbb{S}_t^i, G_t^i)_{i=1,\ldots,n}$, let $(I_\mathbf{j}^i)_{\mathbf{j}\in\nabla\mathbb{S}_t^i}^{i=1,\ldots,n}$ be i.i.d. uniformly distributed on $[N]$. Define $\tilde{M}_t^N : S^N \to S^n$ and $M_t^N : S^N \to S^n$ by*

$$\tilde{M}_t^N(x)_i := \mathbf{X}_{-Nt,0}^N(x)_{I_i} \quad \text{and} \quad M_t^N(x)_i := G_t^i\big((x_{I_\mathbf{j}^i})_{\mathbf{j}\in\nabla\mathbb{S}_t^i}\big). \tag{4.16}$$

Then $\tilde{M}_t^N$ and M_t^N can be coupled such that $\mathbb{P}\big[\tilde{M}_t^N \ne M_t^N\big] \underset{N\to\infty}{\longrightarrow} 0$.

Proof. The proof is the same as the proof of Lemma 4.3, except that instead of following back the genealogy of one site, one follows the genealogies of n sites. By the same arguments as given in the proof of Lemma 4.3, when N is large, with high probability, the genealogies do not intersect, and hence can be approximated by independent branching processes. Although writing down all objects involved is notationally complicated, no new ideas are needed so we omit the details. $\qquad\square$

Proof of Proposition 4.1. Let $x := X^{(N)}(0)$ be the (deterministic) initial state and using notation as in (1.33) let $\mu_0^N = \mu\{x\}$ denote its empirical measure. Define maps $\tilde{M}_t^N$ and M_t^N as in Lemma 4.4. Then $\big(X_{I_1}^{(N)}(Nt),\ldots,X_{I_n}^{(N)}(Nt)\big)$ has law $\tilde{M}_t^N(x)$ while the coordinates of $M_t^N(x)$ are i.i.d. with a law that by Theorem 1.6 equals $T_t(\mu_0^N)$. In view of this, the claim follows from Lemma 4.4. $\qquad\square$

4.3 Tightness in total variation

In this subsection we prove Lemma 4.2.

Proof of Lemma 4.2. The process $(X^{(N)}(t))_{t\ge 0}$ is defined in (1.32) in terms of a stochastic flow which is in turn defined in terms of a Poisson set Π. Elements of Π are triples $(\omega, \mathbf{i}, s)$ which tell us that at time s the map $\vec{\gamma}[\omega]$ should be applied to the coordinates $\mathbf{i} = (i_1,\ldots,i_{\lambda(\omega)})$. We let

$$\tau_t^N := \frac{1}{NL} \sum_{(\omega,\mathbf{i},s)\in\Pi:\, 0<s\le Nt} \lambda(\omega) \qquad (t\ge 0), \tag{4.17}$$

where $L := \int_\Omega \mathbf{q}(\mathrm{d}\omega)\,\lambda(\omega)$, which is finite by (1.23). Then (i) follows from a functional law of large numbers. Since for any $s \le t$, the fraction of sites in $[N]$ that changes its type is bounded from above by $L(\tau_t^N - \tau_s^N)$, in view of (3.9), we obtain also (ii). $\qquad\square$

4.4 Convergence to the mean-field equation

In this subsection, we prove Theorem 1.5. The proof is split into a number of lemmas. We start by proving convergence at fixed times. This part of the proof is based on Proposition 4.1. At the end of the proof, we use Lemma 4.2 to obtain pathwise convergence.

Lemma 4.5 (Expectation of test functions). *Let $\Omega', \mathbf{q}, \lambda,$ and $\vec{\gamma}$ be as in Subsection 1.3, and assume (1.23). Let $(T_t)_{t\ge 0}$ denote the semigroup defined as in (1.6) but with the mean-field equation (1.4) replaced by (1.22). For each $N \in \mathbb{N}_+$, let $(X^{(N)}(t))_{t\ge 0}$ be Markov processes with state space S^N as defined in (1.32), and let $\mu_t^N = \mu\{X^{(N)}(t)\}$ denote their associated empirical measures. Then*

$$\sup_{|\phi|\le 1} \mathbb{P}\big[\big|\langle \mu_{Nt}^N, \phi\rangle - \langle T_t(\mu_0^N), \phi\rangle\big| \ge \varepsilon\big] \underset{N\to\infty}{\longrightarrow} 0 \qquad (\varepsilon > 0,\ t\ge 0), \tag{4.18}$$

where the supremum runs over all measurable functions $\phi : S \to [-1,1]$.

Proof. Fix $t \geq 0$. Let $\phi : S \to [-1, 1]$ be measurable. Let I_1 and I_2 be uniformly distributed on $[N]$ and independent of each other and of $X^N(t)$. Since

$$\langle \mu_{Nt}^N, \phi \rangle = \frac{1}{N} \sum_{i=1}^N \phi\big(X_i^{(N)}(Nt)\big), \tag{4.19}$$

we see that

$$\begin{aligned}
\mathbb{E}\big[\langle \mu_{Nt}^N, \phi \rangle\big] &= \mathbb{E}\big[\phi\big(X_{I_1}^{(N)}(Nt)\big)\big], \\
\mathbb{E}\big[\langle \mu_{Nt}^N, \phi \rangle^2\big] &= \mathbb{E}\big[\phi\big(X_{I_1}^{(N)}(Nt)\big)\phi\big(X_{I_2}^{(N)}(Nt)\big)\big].
\end{aligned} \tag{4.20}$$

Assume for the moment that $X^{(N)}(0)$ is deterministic. Then applying Proposition 4.1 with $n = 1, 2$ we find that

$$\begin{aligned}
\sup_{|\phi| \leq 1} \Big| \mathbb{E}\big[\langle \mu_{Nt}^N, \phi \rangle\big] - \langle T_t(\mu_0^N), \phi \rangle \Big| &\xrightarrow[N \to \infty]{} 0, \\
\sup_{|\phi| \leq 1} \Big| \mathbb{E}\big[\langle \mu_{Nt}^N, \phi \rangle^2\big] - \langle T_t(\mu_0^N), \phi \rangle^2 \Big| &\xrightarrow[N \to \infty]{} 0,
\end{aligned} \tag{4.21}$$

where we take the supremum over all measurable $\phi : S \to [-1, 1]$. It follows that

$$\sup_{|\phi| \leq 1} \mathrm{Var}\big(\langle \mu_{Nt}^N, \phi \rangle\big) = \sup_{|\phi| \leq 1} \Big(\mathbb{E}\big[\langle \mu_{Nt}^N, \phi \rangle^2\big] - \mathbb{E}\big[\langle \mu_{Nt}^N, \phi \rangle\big]^2 \Big) \xrightarrow[N \to \infty]{} 0, \tag{4.22}$$

and hence (4.18) follows by Chebyshev's inequality. To obtain (4.18) more generally when $X^{(N)}(0)$ is random, we condition on the initial state to get, for each $\varepsilon > 0$ and measurable $\psi : S \to [-1, 1]$.

$$\begin{aligned}
\int &\mathbb{P}[X^{(N)}(0) \in \mathrm{d}x]\, \mathbb{P}\big[\big|\langle \mu_{Nt}^N, \psi \rangle - \langle T_t(\mu_0^N), \psi \rangle\big| \geq \varepsilon \,\big|\, X^{(N)}(0) = x\big] \\
&\leq \int \mathbb{P}[X^{(N)}(0) \in \mathrm{d}x] \sup_{|\phi| \leq 1} \mathbb{P}\big[\big|\langle \mu_{Nt}^N, \phi \rangle - \langle T_t(\mu_0^N), \phi \rangle\big| \geq \varepsilon \,\big|\, X^{(N)}(0) = x\big].
\end{aligned} \tag{4.23}$$

Since the integrand on the right-hand side does not depend on ψ and tends to zero in a bounded pointwise way as a function of $x \in S^N$, (4.18) follows. $\quad\square$

Our next aim is to prove that if in addition to the assumptions of Lemma 4.5, condition (i) or (ii) of Theorem 1.5 is satisfied, then

$$\mathbb{P}\big[d(\mu_{Nt}^N, T_t(\mu_0)) \geq \varepsilon\big] \xrightarrow[N \to \infty]{} 0 \qquad (\varepsilon > 0,\ t \geq 0), \tag{4.24}$$

where d is any metric on $\mathcal{P}(S)$ that generates the topology of weak convergence. Applying the following well-known fact to the Polish space $\mathcal{P}(S)$, we see that if (4.24) holds for one such metric, then it holds for all of them.

Lemma 4.6 (Convergence in probability). *Let X_n be random variables taking values in a Polish space S, let $x \in S$ be deterministic, and let d be a metric generating the topology on S. Then one has*

$$\mathbb{P}\big[d(X_n, x) \geq \varepsilon\big] \xrightarrow[n \to \infty]{} 0 \qquad (\varepsilon > 0), \tag{4.25}$$

if and only if

$$\mathbb{P}\big[X_n \in \cdot\,\big] \underset{n \to \infty}{\Longrightarrow} \delta_x, \tag{4.26}$$

where $\Rightarrow$ denotes weak convergence of probability measures on S.

Proof. It is easy to see that (4.25) implies $\mathbb{E}[\phi(X_n)] \to \phi(x)$ for all bounded continuous $\phi : S \to \mathbb{R}$, so (4.25) implies (4.26). Conversely, if (4.26) holds, then by Skorohod's representation theorem it is possible to couple the random variables X_n such that $X_n \to x$ a.s., which implies (4.25). $\quad\square$

The following lemma gives sufficient conditions for the type of convergence of (4.24).

Lemma 4.7 (Convergence to a deterministic measure). *Let S be a Polish space, let $\mu \in \mathcal{P}(S)$ be deterministic, and let μ^N be random variables with values in $\mathcal{P}(S)$. Let d be a metric on $\mathcal{P}(S)$ generating the topology of weak convergence. Then the following conditions are equivalent.*

(i) $\mathbb{P}\big[d(\mu^N, \mu) \geq \varepsilon\big] \underset{N\to\infty}{\longrightarrow} 0$ *for all* $\varepsilon > 0$.

(ii) $\mathbb{P}\big[|\langle \mu^N, \phi\rangle - \langle \mu, \phi\rangle| \geq \varepsilon\big] \underset{N\to\infty}{\longrightarrow} 0$ *for all* $\varepsilon > 0$ *and bounded continuous* $\phi : S \to \mathbb{R}$.

(iii) $\mathbb{E}\big[\prod_{i=1}^{n}\langle \mu^N, \phi_i\rangle\big] \underset{N\to\infty}{\longrightarrow} \prod_{i=1}^{n}\langle \mu, \phi_i\rangle$ *for all bounded continuous functions* $\phi_1, \ldots, \phi_n$ ($n \geq 1$).

Proof. We equip $\mathcal{P}(S)$ with the topology of weak convergence, making it into a Polish space. Then by Lemma 4.6, condition (i) is equivalent to

(i)′ $\mathbb{P}[\mu^N \in \,\cdot\,] \underset{N\to\infty}{\Longrightarrow} \delta_\mu$.

We will prove (i)′⇒(ii)⇒(iii)⇒(i)′.

(i)′⇒(ii). By Skorohod's representation theorem, (i)′ implies that the μ^N can be coupled such that $\mu^N \underset{N\to\infty}{\Longrightarrow} \mu$ a.s., which implies (ii).

(ii)⇒(iii). Without loss of generality we may assume that the ϕ_i's take values in $[-1, 1]$. Since the function $(x_1, \ldots, x_n) \mapsto \prod_{i=1}^{n} x_i$ is continuous, (ii) implies that

$$\mathbb{P}\big[|\prod_{i=1}^{n}\langle \mu^N, \phi_i\rangle - \prod_{i=1}^{n}\langle \mu, \phi_i\rangle| \geq \varepsilon\big] \underset{N\to\infty}{\longrightarrow} 0 \tag{4.27}$$

for all $\varepsilon > 0$ and bounded continuous $\phi_1, \ldots, \phi_n$. Since moreover $|\prod_{i=1}^{n}\langle \mu^N, \phi_i\rangle| \leq 1$, this implies (iii).

(iii)⇒(i)′. Since S is Polish, it has a metrizable compactification, i.e., there exists a compact metrizable space $\overline{S}$ such that S is a dense subset of $\overline{S}$ and the topology on S is the induced topology from $\overline{S}$ [Cho69, Theorem 6.3]. It is known that this implies that S is a G_δ-subset of $\overline{S}$ [Bou58, §6 No. 1, Theorem. 1]. In particular, S is a Borel measurable subset of $\overline{S}$ and we can identify $\mathcal{P}(S)$ with the space of probability measures on $\overline{S}$ that are concentrated on S. If we equip $\mathcal{P}(\overline{S})$ with the topology of weak convergence, then the induced topology on $\mathcal{P}(S)$ is also the topology of weak convergence (this follows, e.g., from [EK86, Thm 3.3.1]), and in fact $\mathcal{P}(\overline{S})$ (being compact by Prohorov's theorem) is a metrizable compactification of $\mathcal{P}(S)$.

We view μ^N and μ as probability measures on $\overline{S}$. Since $\overline{S}$ is compact, so are $\mathcal{P}(\overline{S})$ and $\mathcal{P}(\mathcal{P}(\overline{S}))$, so by going to a subsequence if necessary, we can assume that the laws $\mathbb{P}[\mu^N \in \,\cdot\,]$ converge weakly to some limit $\rho \in \mathcal{P}(\mathcal{P}(\overline{S}))$. Since the restriction to S of a continuous function $\phi : \overline{S} \to \mathbb{R}$ is a bounded continuous function on S, condition (ii) implies that

$$\int_{\mathcal{P}(\overline{S})} \rho(\mathrm{d}\nu) \prod_{i=1}^{n}\langle \nu, \phi_i\rangle = \prod_{i=1}^{n}\langle \mu, \phi_i\rangle \tag{4.28}$$

for general $n \geq 1$ and continuous functions $\phi_i : \overline{S} \to \mathbb{R}$ ($i = 1, \ldots, n$). By the Stone-Weierstrass theorem, the linear span of functions of the form $\nu \mapsto \prod_{i=1}^{n}\langle \mu, \phi_i\rangle$ is dense in the space of continuous functions on $\mathcal{P}(\overline{S})$, and hence (4.28) implies $\rho = \delta_\mu$. $\square$

We now prove (4.24) under either of the conditions (i) and (ii) of Theorem 1.5.

Lemma 4.8 (Continuity argument). *In addition to the assumptions of Lemma 4.5, assume that condition (i) of Theorem 1.5 is satisfied. Then (4.24) holds.*

Proof. Fix $t \geq 0$. In view of Lemma 4.7 (ii), it suffices to show that

$$\mathbb{P}\big[\big|\langle \mu_{Nt}^N, \phi \rangle - \langle T_t(\mu_0), \phi \rangle\big| \geq \varepsilon\big] \underset{N \to \infty}{\longrightarrow} 0 \qquad (\varepsilon > 0) \tag{4.29}$$

for any bounded continuous $\phi : S \to \mathbb{R}$. By Lemma 4.5, it suffices to show that

$$\mathbb{P}\big[\big|\langle T_t(\mu_0^N), \phi \rangle - \langle T_t(\mu_0), \phi \rangle\big| \geq \varepsilon\big] \underset{N \to \infty}{\longrightarrow} 0 \qquad (\varepsilon > 0). \tag{4.30}$$

By the second part of condition (i), Lemma 1.4, and Proposition 1.3, the operator T_t is continuous w.r.t. weak convergence. In view of this, (4.30) is implied by the first part of condition (i). $\qquad \square$

Lemma 4.9 (Moment argument)**.** *In addition to the assumptions of Lemma 4.5, assume that condition (ii) of Theorem 1.5 is satisfied. Then (4.24) holds.*

Proof. Fix $t \geq 0$. In view of Lemma 4.7 (iii), it suffices to show that

$$\mathbb{E}\big[\prod_{i=1}^n \langle \mu_{Nt}^N, \phi_i \rangle\big] \underset{N \to \infty}{\longrightarrow} \prod_{i=1}^n \langle T_t(\mu_0), \phi_i \rangle \tag{4.31}$$

for all $n \geq 1$ and bounded continuous functions $\phi_i : S \to \mathbb{R}$, $i = 1, \ldots, n$. Without loss of generality we may assume that the ϕ_i's take values in $[-1, 1]$. Let $X^{(N)}(t)$ be as in Theorem 1.5 and let $I_1, \ldots, I_n$ be i.i.d. uniformly distributed on $[N]$ and independent of $X^N(t)$. Then

$$\mathbb{E}\big[\prod_{i=1}^n \langle \mu_{Nt}^N, \phi_i \rangle\big] = \mathbb{E}\big[\prod_{i=1}^n \phi_i\big(X_{I_i}^{(N)}(Nt)\big)\big]. \tag{4.32}$$

By Proposition 4.1 applied to the process conditioned on $X^{(N)}(0)$, there exist $\varepsilon_N \to 0$ such that

$$\Big\|\mathbb{P}\Big[\big(X_{I_1}^{(N)}(Nt), \ldots, X_{I_n}^{(N)}(Nt)\big) \in \cdot \,\big|\, X^{(N)}(0)\Big] - T_t(\mu_0^N)^{\otimes n}\Big\| \leq \varepsilon_N. \tag{4.33}$$

In view of (3.8), it follows that

$$\Big|\mathbb{E}\Big[\prod_{i=1}^n \phi_i\big(X_{I_i}^{(N)}(t)\big)\,\Big|\, X^{(N)}(0)\Big] - \prod_{i=1}^n \langle T_t(\mu_0^N), \phi_i \rangle\Big| \leq \varepsilon_N. \tag{4.34}$$

Combining this with (4.32), taking the expectation, we obtain that

$$\Big|\mathbb{E}\big[\prod_{i=1}^n \langle \mu_{Nt}^N, \phi_i \rangle\big] - \mathbb{E}\big[\prod_{i=1}^n \langle T_t(\mu_0^N), \phi_i \rangle\big]\Big| \leq \varepsilon_N. \tag{4.35}$$

In view of this, to prove (4.31), it suffices to show that

$$\mathbb{E}\Big[\prod_{i=1}^n \langle T_t(\mu_0^N), \phi_i \rangle\Big] \underset{N \to \infty}{\longrightarrow} \prod_{i=1}^n \langle T_t(\mu_0), \phi_i \rangle. \tag{4.36}$$

If $\mu \in \mathcal{P}(S)$ is deterministic, then Theorem 1.6 tells us that

$$\langle T_t(\mu), \phi \rangle = \mathbb{E}\big[\phi \circ G_t\big((X_{\mathbf{j}})_{\mathbf{j} \in \nabla \mathbb{S}_t}\big)\big], \tag{4.37}$$

where $\nabla \mathbb{S}_t$ and G_t are as in (1.44) and (1.47) and $(X_{\mathbf{j}})_{\mathbf{j} \in \mathbb{T}}$ are i.i.d. with law μ. Conditional on μ_0^N, let $(X_{\mathbf{j}}^i)_{\mathbf{j} \in \mathbb{T}}^{i=1,\ldots,n}$ be i.i.d. with common law μ_0^N. Let $(\nabla \mathbb{S}_t^i, G_t^i)$ $(i = 1, \ldots, n)$ be i.i.d.

and distributed as the random variables in (1.44) and (1.47), independent of μ_0^N and $(X_{\mathbf{j}}^i)_{\mathbf{j}\in\mathbb{T}}^{i=1,\ldots,n}$. Then (4.37) implies that

$$
\mathbb{E}\Big[\prod_{i=1}^n \langle T_t(\mu_0^N),\phi_i\rangle\Big] = \mathbb{E}\Big[\prod_{i=1}^n \phi_i \circ G_t^i\big((X_{\mathbf{j}}^i)_{\mathbf{j}\in\nabla\mathbb{S}_t^i}\big)\Big]. \tag{4.38}
$$

If we replace the expectation on the right-hand side by a conditional expectation given $(\nabla\mathbb{S}_t^i, G_t^i)_{i=1,\ldots,n}$, then this is the integral of a measurable $[-1,1]$-valued function with respect to the expectation of a product measure of the form $(\mu_0^N)^{\otimes m}$, where $m = \sum_{i=1}^n |\nabla\mathbb{S}_t^i|$. Condition (ii) of Theorem 1.5 allows us to replace the integral w.r.t. $\mathbb{E}[(\mu_0^N)^{\otimes m}]$ by the integral w.r.t. $\mu_0^{\otimes m}$ at the cost of a small error. Thus,

$$
\begin{aligned}
&\mathbb{E}\Big[\prod_{i=1}^n \phi_i \circ G_t^i\big((X_{\mathbf{j}}^i)_{\mathbf{j}\in\nabla\mathbb{S}_t^i}\big) \,\Big|\, (\nabla\mathbb{S}_t^i, G_t^i)_{i=1,\ldots,n}\Big]\\
&= \mathbb{E}\Big[\prod_{i=1}^n \phi_i \circ G_t^i\big((\tilde{X}_{\mathbf{j}}^i)_{\mathbf{j}\in\nabla\mathbb{S}_t^i}\big) \,\Big|\, (\nabla\mathbb{S}_t^i, G_t^i)_{i=1,\ldots,n}\Big] + R^N,
\end{aligned} \tag{4.39}
$$

where the $(\tilde{X}_{\mathbf{j}}^i)_{\mathbf{j}\in\mathbb{T}}^{i=1,\ldots,n}$ are i.i.d. with common law μ_0 and independent of $(\nabla\mathbb{S}_t^i, G_t^i)_{i=1,\ldots,n}$, and R^N is a random error term that by condition (ii) can be estimated as

$$
|R^N| \le \varepsilon_m^N \quad \text{with} \quad m = \sum_{i=1}^n |\nabla\mathbb{S}_t^i|, \tag{4.40}
$$

where $\lim_{N\to\infty}\varepsilon_m^N = 0$ for each m. Note that moreover $|R^N| \le 2$ since the ϕ_i's take values in $[-1,1]$. Integrating over the randomness of $(\nabla\mathbb{S}_t^i, G_t^i)_{i=1,\ldots,n}$, using bounded convergence, (4.37) and (4.38), (4.36) follows. $\qquad\square$

With Lemmas 4.8 and 4.9 proved, most of the work needed for proving Theorem 1.5 is done. The only remaining task is to improve the convergence at fixed times in (4.24) to pathwise convergence as in (1.34). Our first aim is to show that the condition (1.34) does not depend on the choice of the metric d. This follows from the following lemma, applied to the Polish space $\mathcal{P}(S)$.

Lemma 4.10 (Convergence in path space). *Let S be a Polish space and let d be a metric generating the topology on S. Let $\mathcal{D}_S[0,\infty)$ be the space of cadlag functions $x : [0,\infty) \to S$, equipped with the Skorohod topology. Let $X_n = (X_n(t))_{t\ge 0}$ be random variables with values in $\mathcal{D}_S[0,\infty)$ and let $x : [0,\infty) \to S$ be a continuous function. Then one has*

$$
\mathbb{P}\Big[\sup_{0\le t\le T} d\big(X_n(t), x(t)\big) \ge \varepsilon\Big] \underset{n\to\infty}{\longrightarrow} 0 \qquad (\varepsilon > 0,\, T < \infty), \tag{4.41}
$$

if and only if

$$
\mathbb{P}\big[X_n \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \delta_x, \tag{4.42}
$$

where $\Rightarrow$ denotes weak convergence of probability measures on $\mathcal{D}_S[0,\infty)$.

Proof. It is well-known that $\mathcal{D}_S[0,\infty)$ is a Polish space [EK86, Sect. 3.5]. Let d_{S} be the metric generating the topology on $\mathcal{D}_S[0,\infty)$ defined in [EK86, (5.2) of Chapter 3]. Then it is easy to see that for all $\delta > 0$ there exist $\varepsilon > 0$ and $T < \infty$ such that

$$
\sup_{0\le t\le T} d\big(y(t), x(t)\big) \le \varepsilon \quad \text{implies} \quad d_{\mathrm{S}}(y,x) \le \delta \qquad (x,y \in \mathcal{D}_S[0,\infty)). \tag{4.43}
$$

In view of this, (4.41) implies

$$
\mathbb{P}\big[d_{\mathrm{S}}(X_n, x) \ge \varepsilon\big] \underset{n\to\infty}{\longrightarrow} 0 \qquad (\varepsilon > 0), \tag{4.44}
$$

which by Lemma 4.6 implies (4.42). Conversely, if (4.42) holds, then by Skorohod's representation theorem it is possible to couple the random variables X_n such that $d_{\mathrm{S}}(X_n, x) \to 0$ a.s. By the continuity of x and [EK86, Lemma 3.10.1], this implies that

$$\sup_{0 \leq t \leq T} d\big(X_n(t), x(t)\big) \underset{n \to \infty}{\longrightarrow} 0 \quad \text{a.s.} \quad (T < \infty), \tag{4.45}$$

which implies (4.41). $\qquad\square$

Before the proof of Theorem 1.5 we need one more lemma.

Lemma 4.11 (Weak convergence and convergence in total variation norm)**.** *Let S be a Polish space. Then there exists a metric d on $\mathcal{P}(S)$ such that d generates the topology of weak convergence and $d(\mu, \nu) \leq \|\mu - \nu\|$ $(\mu, \nu \in \mathcal{P}(S))$, where $\|\cdot\|$ denotes the total variation norm.*

Proof. Let r be a metric generating the topology on S. Replacing $r(x, y)$ by $r(x, y) \wedge 1$ if necessary we can assume without loss of generality that $r \leq 1$. Let $\mathcal{L}$ be the space of all functions $\phi : S \to \mathbb{R}$ such that $|\phi(x) - \phi(y)| \leq r(x, y)$ $(x, y \in S)$, i.e., these are Lipschitz continuous functions with Lipschitz constant ≤ 1. Then

$$d(\mu, \nu) := \sup_{\phi \in \mathcal{L}} \left| \int \phi \, \mathrm{d}\mu - \int \phi \, \mathrm{d}\nu \right| \tag{4.46}$$

is the 1-Wasserstein metric on $\mathcal{P}(S)$, which is known to generate the topology of weak convergence. Let $\mathcal{L}' := \{\phi \in \mathcal{L} : \sup_{x \in S} |\phi(x)| \leq 1\}$. Since $r \leq 1$, each function $\phi \in \mathcal{L}$ can be written as $\phi = \frac{1}{2}\phi' + c$ with $\phi' \in \mathcal{L}'$ and $c \in \mathbb{R}$. In view of this and (3.8),

$$d(\mu, \nu) = \tfrac{1}{2} \sup_{\phi \in \mathcal{L}'} \left| \int \phi \, \mathrm{d}\mu - \int \phi \, \mathrm{d}\nu \right| \leq \tfrac{1}{2} \sup_{|\phi| \leq 1} \left| \int \phi \, \mathrm{d}\mu - \int \phi \, \mathrm{d}\nu \right| = \|\mu - \nu\|. \tag{4.47}$$

$\qquad\square$

Proof of Theorem 1.5. Lemmas 4.8 and 4.9 show that either of the conditions (i) and (ii) implies (4.24). We will use Lemma 4.2 to improve (4.24) to pathwise convergence as in (1.34). By Lemma 4.10 it suffices to prove (1.34) for one particular metric d on $\mathcal{P}(S)$ that generates the topology of weak convergence. We will choose a metric d as in Lemma 4.11.

Set $\mu_t := T_t(\mu_0)$ $(t \geq 0)$ denote the solution to the mean-field equation (1.22) with initial state μ_0. Lemma 4.2 implies that

$$\mathbb{P}\big[\|\mu_{Ns}^N - \mu_{Nt}^N\| \geq L|t - s| + \varepsilon\big] \underset{N \to \infty}{\longrightarrow} 0 \quad (\varepsilon > 0, \ s, t \geq 0). \tag{4.48}$$

Taking the limit $N \to \infty$, using the fact that $d(\mu, \nu) \leq \|\mu - \nu\|$ and (4.24), it follows that

$$d(\mu_s, \mu_t) \leq L|t - s| \quad (s, t \geq 0). \tag{4.49}$$

Since for any $s, t \geq 0$,

$$d(\mu_{Ns}^N, \mu_s) \leq \|\mu_{Ns}^N - \mu_{Nt}^N\| + d(\mu_{Nt}^N, \mu_t) + d(\mu_t, \mu_s), \tag{4.50}$$

using Lemma 4.2, (4.24), and (4.49), we see that for each $T > 0$ and $t \in [0, T]$,

$$\mathbb{P}\big[\exists s \in [0, T] \text{ s.t. } d(\mu_{Ns}^N, \mu_s) \geq 2L|t - s| + \varepsilon\big] \underset{N \to \infty}{\longrightarrow} 0 \quad (\varepsilon > 0). \tag{4.51}$$

Combining this with the fact that by (4.24), for any $n \geq 1$,

$$\mathbb{P}\Big[\sup_{k=0,\ldots,n} d(\mu_{N(k/n)T}^N, \mu_{(k/n)T}) \geq \varepsilon\Big] \underset{N \to \infty}{\longrightarrow} 0 \quad (\varepsilon > 0), \tag{4.52}$$

we find that

$$\mathbb{P}\Big[\sup_{t \in [0, T]} d(\mu_{Nt}^N, \mu_t) \geq L/n + \varepsilon\Big] \underset{N \to \infty}{\longrightarrow} 0 \quad (\varepsilon > 0, \ n \geq 1). \tag{4.53}$$

Since ε and n are arbitrary, this implies (1.34). $\qquad\square$

5 Recursive tree processes

In this section, we prove our main results about RTPs with continuous time. For completeness, we also prove Lemma 1.9 which deals with discrete time and says that each solution to the RDE (1.54) gives rise to an RTP. This is done in Subsection 5.1

Our basic results about continuous-time RTPs are Lemma 1.8 and Proposition 1.10. Lemma 1.8 describes the evolution of the law of the process

$$\{X_{\mathbf{i}} : \mathbf{i} \in \nabla \mathbb{S}_{t-s}\}_{0 \le s \le t}, \tag{5.1}$$

that is constructed by assigning independent values $X_{\mathbf{i}}$ to elements $\mathbf{i} \in \nabla \mathbb{S}_t$ and then calculating backwards. Proposition 1.10 says that adding exponential lifetimes to the elements of an RTP yields a stationary version of the process in (5.1). These results are proved in Subsection 5.2.

In Subsection 5.3, we prove continuous-time analogues of known discrete-time results related to endogeny. Following [AB05], Theorem 1.12 links the n-variate mean-field equation to endogeny, while Propositions 1.14 and 1.16 are concerned with the higher-level mean-field equation, and closely follow ideas from [MSS18].

5.1 Construction of RTPs

Proof of Lemma 1.9. For each finite subtree $\mathbb{U} \subset \mathbb{T}$ that contains the root, we can construct random variables $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U}}$ and $(X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U} \cup \partial \mathbb{U}}$ such that the $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U}}$ are independent with common law $|\mathbf{r}|^{-1}\mathbf{r}$, the $(X_{\mathbf{i}})_{\mathbf{i} \in \partial \mathbb{U}}$ are i.i.d. with common law ν and independent of the $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U}}$, and the $(X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U}}$ are inductively defined by

$$X_{\mathbf{i}} = \gamma[\boldsymbol{\omega}_{\mathbf{i}}]\big(X_{\mathbf{i}1}, \ldots, X_{\mathbf{i}\kappa(\boldsymbol{\omega}_{\mathbf{i}})}\big) \qquad (\mathbf{i} \in \mathbb{U}). \tag{5.2}$$

The joint law of $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U}}$ and $(X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U} \cup \partial \mathbb{U}}$ is a probability law $\mathbb{P}_{\mathbb{U}}$ on $\Omega^{\mathbb{U}} \times S^{\mathbb{U} \cup \partial \mathbb{U}}$. Since Ω and S are Polish spaces, we can apply Kolmogorov's extension theorem. The statement of the lemma then follows provided we can show that the laws $\mathbb{P}_{\mathbb{U}}$ are consistent in the sense that if $\mathbb{V} \subset \mathbb{U}$ is another subtree that contains the root, then the projection of $\mathbb{P}_{\mathbb{U}}$ on $\Omega^{\mathbb{V}} \times S^{\mathbb{V} \cup \partial \mathbb{V}}$ equals $\mathbb{P}_{\mathbb{V}}$. It suffices to prove this when $\mathbb{U}$ and $\mathbb{V}$ differ by one element only, say $\mathbb{U} = \mathbb{V} \cup \{\mathbf{i}\}$ where $\mathbf{i} \in \nabla \mathbb{V}$. It follows from (5.2) and the fact that ν solves the RDE (1.54) that $X_{\mathbf{i}}$ has law ν and is independent of $(X_{\mathbf{j}})_{\mathbf{j} \in \nabla \mathbb{V} \setminus \{\mathbf{i}\}}$, and from this we see that the projection of $\mathbb{P}_{\mathbb{U}}$ is indeed $\mathbb{P}_{\mathbb{V}}$. $\qquad\square$

It will be useful in what follows to have a somewhat stronger version of Lemma 1.9 that applies also to certain random subtrees $\mathbb{U} \subset \mathbb{T}$. Let $\mathcal{T}$ denote the set of all finite subtrees $\mathbb{U} \subset \mathbb{T}$ such that either $\varnothing \in \mathbb{U}$ or $\mathbb{U} = \emptyset$. Let us define a *stopping tree* to be a random variable $\mathbb{U}$ with values in $\mathcal{T}$ such that

$$\{\mathbb{U} = \mathbb{V}\} \text{ is measurable w.r.t. } \sigma\big((\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{V}}\big) \quad \text{for all } \mathbb{V} \in \mathcal{T}. \tag{5.3}$$

In the special case that $\kappa \equiv 1$ and $\mathbb{T} = \mathbb{N}$, a stopping tree is just a stopping time w.r.t. the filtration generated by $\boldsymbol{\omega}_{\varnothing}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_{11}, \ldots$.

Lemma 5.1 (RTPs and stopping trees). *Let $(\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ be an RTP corresponding to a map γ and a solution ν to the RDE (1.54), and let $\mathbb{U} \subset \mathbb{T}$ be a stopping tree. Then conditional on $\mathbb{U}$, the random variables $(X_{\mathbf{i}})_{\mathbf{i} \in \partial \mathbb{U}}$ are i.i.d. with common law ν and independent of $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U}}$.*

Proof. For each fixed $\mathbb{V} \in \mathcal{T}$, by Lemma 1.9, conditional on $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{V}}$, the random variables $(X_{\mathbf{i}})_{\mathbf{i} \in \partial \mathbb{V}}$ are i.i.d. with common law ν. By (5.3), it follows that conditional on the event $\{\mathbb{U} = \mathbb{V}\}$ and $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{V}}$, the random variables $(X_{\mathbf{i}})_{\mathbf{i} \in \partial \mathbb{V}}$ are i.i.d. with common law ν. Since this holds for all $\mathbb{V} \in \mathcal{T}$, and since $\mathbb{U} \in \mathcal{T}$ a.s., the claim follows. $\qquad\square$

5.2 Continuous-time RTPs

In this subsection, we prove Lemma 1.8 and Proposition 1.10. We work in our usual set-up as described above Proposition 3.4. We start with a preparatory lemma that says that if we condition on the σ-field $\mathcal{F}_t$ defined in (1.48), then the subtrees of $\mathbb{S}$ rooted at $\mathbf{i} \in \nabla\mathbb{S}_t$ are i.i.d. with the same distribution as $\mathbb{S}$. To formulate this properly, we need some notation.

We call the object

$$\big(\mathbb{S}, (\boldsymbol{\omega}_\mathbf{i}, \sigma_\mathbf{i})_{\mathbf{i}\in\mathbb{S}}\big). \tag{5.4}$$

a *marked branching tree*. For each $\mathbf{i} \in \mathbb{S}$, let $\mathbb{S}^\mathbf{i}$ describe the subtree of $\mathbb{S}$ that is rooted at $\mathbf{i}$, i.e.,

$$\mathbb{S}^\mathbf{i} := \{\mathbf{j} \in \mathbb{T} : \mathbf{ij} \in \mathbb{S}\}. \tag{5.5}$$

We set $\omega_\mathbf{j}^\mathbf{i} := \omega_{\mathbf{ij}}$ $(\mathbf{i}, \mathbf{j} \in \mathbb{T})$, so that $\omega_\mathbf{j}^\mathbf{i}$ is the random element of Ω that "belongs" to $\mathbf{j} \in \mathbb{S}^\mathbf{i}$. Fix $t \geq 0$. For each $\mathbf{i} \in \nabla\mathbb{S}_t$, let $\sigma_\mathbf{j}^{\mathbf{i},t}$ describe the lifetime of an individual $\mathbf{j} \in \mathbb{S}^\mathbf{i}$ after time t, i.e.,

$$\sigma_\varnothing^{\mathbf{i},t} := \sigma_\mathbf{i} - (t - \tau_\mathbf{i}^*) \quad \text{and} \quad \sigma_\mathbf{j}^{\mathbf{i},t} := \sigma_{\mathbf{ij}} \quad (\varnothing \neq \mathbf{j} \in \mathbb{S}^\mathbf{i}), \tag{5.6}$$

where $t - \tau_\mathbf{i}^*$ is the age of the individual $\mathbf{i}$ at time t.

Lemma 5.2 (Memoryless property). *For each $t \geq 0$, conditional on the σ-field $\mathcal{F}_t$, the marked branching trees*

$$\big(\mathbb{S}^\mathbf{i}, (\omega_\mathbf{j}^\mathbf{i}, \sigma_\mathbf{j}^{\mathbf{i},t})_{\mathbf{j}\in\mathbb{S}^\mathbf{i}}\big)_{\mathbf{i}\in\nabla\mathbb{S}_t} \tag{5.7}$$

are i.i.d. with the same distribution as the marked branching tree in (5.4).

Proof. Let $\mathcal{T}$ be as defined above (5.3). Then, for each $\mathbb{V} \in \mathcal{T}$, the event $\{\mathbb{S}_t = \mathbb{V}\}$ is measurable w.r.t. the σ-field generated by the random variables

$$(\boldsymbol{\omega}_\mathbf{i}, \sigma_\mathbf{i})_{\mathbf{i}\in\mathbb{V}} \quad \text{and} \quad (1_{\{\sigma_\mathbf{i} > t - \tau_\mathbf{i}^*\}})_{\mathbf{i}\in\nabla\mathbb{V}}. \tag{5.8}$$

Note that here $\nabla\mathbb{V} = \{\mathbf{i}j : \mathbf{i} \in \mathbb{V}, \ j \leq \kappa(\boldsymbol{\omega}_\mathbf{i})\}$ is measurable w.r.t. the σ-field generated by $(\boldsymbol{\omega}_\mathbf{i})_{\mathbf{i}\in\mathbb{V}}$ while for each $\mathbf{i} \in \nabla\mathbb{V}$, the random variable $\tau_\mathbf{i}^*$ is measurable w.r.t. the σ-field generated by $(\sigma_\mathbf{i})_{\mathbf{i}\in\mathbb{V}}$.

Conditional on $\{\mathbb{S}_t = \mathbb{V}\}$ and the random variables in (5.8), the random variables $(\boldsymbol{\omega}_\mathbf{i})_{\mathbf{i}\in\mathbb{T}\backslash\mathbb{V}}$ are still i.i.d. with their original law and independent of $(\sigma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}\backslash\mathbb{V}}$. The latter are also still independent of each other and the $(\sigma_\mathbf{i})_{\mathbf{i}\in\mathbb{T}\backslash(\mathbb{V}\cup\nabla\mathbb{V})}$ still have their original law, but the laws of $(\sigma_\mathbf{i})_{\mathbf{i}\in\nabla\mathbb{V}}$ are changed since conditioning on $\{\mathbb{S}_t = \mathbb{V}\}$ entails conditioning on $\sigma_\mathbf{i} > t - \tau_\mathbf{i}^*$ for each $\mathbf{i} \in \nabla\mathbb{V}$.

Since this holds for each $\mathbb{V} \in \mathbb{S}$, we see that if we condition on $\mathcal{F}_t$ as in (1.48), then under the conditional law the random variables $\boldsymbol{\omega}_\mathbf{i}$ and $\sigma_\mathbf{i}$ with $\mathbf{i} \in \mathbb{T}\backslash\mathbb{S}_t$ are still independent, and all of these random variables still have their original law, except the $\sigma_\mathbf{i}$ with $\mathbf{i} \in \nabla\mathbb{S}_t$, whose laws are conditioned on the events $\sigma_\mathbf{i} > t - \tau_\mathbf{i}^*$. From this observation, using the memoryless property of the exponential distribution, the claim of the lemma follows. $\square$

For each $s \geq 0$ and $\mathbf{i} \in \nabla\mathbb{S}_s$, within the marked branching tree $\big(\mathbb{S}^\mathbf{i}, (\omega_\mathbf{j}^\mathbf{i}, \sigma_\mathbf{j}^{\mathbf{i},s})_{\mathbf{j}\in\mathbb{S}^\mathbf{i}}\big)$ rooted at $\mathbf{i}$, we define the birth and death times $\tau_\mathbf{j}^{\mathbf{i},*}$ and $\tau_\mathbf{j}^{\mathbf{i},\dagger}$ as in (1.41), with $\sigma_\mathbf{j}$ replaced by $\sigma_\mathbf{j}^{\mathbf{i},s}$, and we use this to define $\mathbb{S}_t^{\mathbf{i},s}$ and $\nabla\mathbb{S}_t^{\mathbf{i},s}$ $(t \geq 0)$ as in (1.44). Finally, we define $G_t^{\mathbf{i},s} = G_{\mathbb{S}_t^{\mathbf{i},s}}$ as in (1.46) and (1.47).

Proof of Lemma 1.8. We fix a marked branching tree as in (5.4) and times $0 \leq s \leq t$. Conditional on $\mathcal{F}_t$, we assign i.i.d. $(X_\mathbf{i})_{\mathbf{i}\in\nabla\mathbb{S}_t}$ with common law μ_0 to the leaves of $\mathbb{S}_t$ and define $(X_\mathbf{i})_{\mathbf{i}\in\mathbb{S}_t}$ inductively as in (1.50).

We observe that $\nabla\mathbb{S}_t$ is given by the disjoint union

$$\nabla\mathbb{S}_t = \bigcup_{\mathbf{i}\in\nabla\mathbb{S}_s} \{\mathbf{ij} : \mathbf{j} \in \nabla\mathbb{S}_{t-s}^{\mathbf{i},s}\}. \tag{5.9}$$

Conditioning on $\mathcal{F}_t$ is the same as first conditioning on

$$\big(\nabla\mathbb{S}_s, (\boldsymbol{\omega_j}, \sigma_{\mathbf{j}})_{\mathbf{j}\in\mathbb{S}_s}\big), \tag{5.10}$$

and then conditioning on

$$\big(\nabla\mathbb{S}_{t-s}^{\mathbf{i},s}, (\omega_{\mathbf{j}}^{\mathbf{i}}, \sigma_{\mathbf{j}}^{\mathbf{i},s})_{\mathbf{j}\in\mathbb{S}_{t-s}^{\mathbf{i},s}}\big)_{\mathbf{i}\in\nabla\mathbb{S}_s}, \tag{5.11}$$

which by Lemma 5.7 are conditionally independent given the random variable in (5.10). Set

$$X_{\mathbf{j}}^{\mathbf{i}} := X_{\mathbf{ij}} \qquad (\mathbf{j} \in \mathbb{S}_{t-s}^{\mathbf{i},s} \cup \nabla\mathbb{S}_{t-s}^{\mathbf{i},s},\ \mathbf{i}\in\nabla\mathbb{S}_s). \tag{5.12}$$

Then

$$X_{\mathbf{i}} = X_{\varnothing}^{\mathbf{i}} = G_{t-s}^{\mathbf{i},s}\big((X_{\mathbf{j}}^{\mathbf{i}})_{\mathbf{j}\in\nabla\mathbb{S}_{t-s}^{\mathbf{i},s}}\big). \tag{5.13}$$

In view of this, by Theorem 1.6, conditional on the the random variable in (5.10), i.e., conditional on $\mathcal{F}_s$, the random variables $(X_{\mathbf{i}})_{\mathbf{i}\in\nabla\mathbb{S}_s}$ are i.i.d. with common law μ_{t-s}, where $(\mu_s)_{s\geq 0}$ denotes the solution of the mean-field equation (1.4) with initial state μ_0. $\qquad\square$

Proof of Proposition 1.10. Since $(\sigma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ and $(\boldsymbol{\omega_i}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are independent, the conditional law of $(\boldsymbol{\omega_i}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ given $(\sigma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is the same as the unconditional law. We claim that under the conditional law given $(\sigma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$, the random finite subtree $\mathbb{S}_t$ is a stopping tree in the sense of (5.3). Indeed, $\mathbb{S}_t = \mathbb{V}$ if and only if for each $\mathbf{i} \in \mathbb{V}$ and $j \in \mathbb{N}_+$ (resp. $j \in [d]$, depending on how $\mathbb{T}$ is chosen), one has $\mathbf{i}j \in \mathbb{V}$ if and only if

$$\text{(i)}\ j \leq \kappa(\mathbf{i}) \quad\text{and}\quad \text{(ii)}\ \tau_{\mathbf{i}j}^{\dagger} \leq t. \tag{5.14}$$

Here the event in (i) is clearly measurable w.r.t. $\sigma((\boldsymbol{\omega_i})_{\mathbf{i}\in\mathbb{V}})$ while under the conditional law given $(\sigma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$, (ii) is just a deterministic condition. We can therefore apply Lemma 5.1 to conclude that conditional on $(\sigma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$, $\mathbb{S}_t$, and $(\boldsymbol{\omega_i})_{\mathbf{i}\in\mathbb{S}_t}$, the random variables $(X_{\mathbf{i}})_{\mathbf{i}\in\partial\mathbb{S}_t}$ are i.i.d. with common law ν.

We observe that

$$\nabla\mathbb{S}_t = \big\{\mathbf{i}j : \mathbf{i} \in \mathbb{S}_t,\ \mathbf{i}j \notin \mathbb{S}_t,\ j \leq \kappa(\boldsymbol{\omega_i})\big\} \tag{5.15}$$

is a function of $\mathbb{S}_t$, and $(\boldsymbol{\omega_i})_{\mathbf{i}\in\mathbb{S}_t}$. Therefore, if we condition on $\mathcal{F}_t = \sigma(\nabla\mathbb{S}_t, (\boldsymbol{\omega_i}, \sigma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{S}_t})$, the random variables $(X_{\mathbf{i}})_{\mathbf{i}\in\nabla\mathbb{S}_t}$ are i.i.d. with common law ν. This proves (1.59) (i). Condition (1.59) (ii) is also clearly fulfilled by the definition of an RTP. $\qquad\square$

5.3 Endogeny, bivariate uniqueness, and the higher-level equation

In this subsection, we prove Theorem 1.12 and Propositions 1.14 and 1.16.

Recall that an RTP $(\boldsymbol{\omega_i}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ is endogenous if $X_{\varnothing}$ is measurable with respect to the σ-field generated by the random variables $(\boldsymbol{\omega_i})_{\mathbf{i}\in\mathbb{T}}$. In general, if X is a random variable taking values in a Polish space and $\mathcal{F}$ is a sub-σ-field, then it is not hard to see that X is a.s. equal to a $\mathcal{F}$-measurable function if and only if the conditional law $\mathbb{P}[X \in \cdot\,|\mathcal{F}]$ is a.s. a delta-measure. In view of this, the following lemma implies that an RTP is endogenous if and only if $X_{\varnothing}$ is a.s. measurable w.r.t. the σ-field generated by the random variables $\mathbb{S}$ and $(\boldsymbol{\omega_i})_{\mathbf{i}\in\mathbb{S}}$.

Lemma 5.3 (Relevant randomness). *Let $(\boldsymbol{\omega_i}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an RTP corresponding to a solution ν of the RDE (1.54). Let $\overline{\mathcal{F}}$ be the σ-field generated by the random variables $(\boldsymbol{\omega_i})_{\mathbf{i}\in\mathbb{T}}$ and let $\mathcal{F}$ be the σ-field generated by the random variables $\mathbb{S}$ and $(\boldsymbol{\omega_i})_{\mathbf{i}\in\mathbb{S}}$. Then*

$$\mathbb{P}[X_{\varnothing} \in \cdot\,|\overline{\mathcal{F}}] = \mathbb{P}[X_{\varnothing} \in \cdot\,|\mathcal{F}] \quad\text{a.s.} \tag{5.16}$$

Proof. Since $\overline{\mathcal{F}}$ is generated by $\mathcal{F}$ and the random variables $(\boldsymbol{\omega_i})_{i \in \mathbb{T} \setminus \mathbb{S}}$, formula (5.16) says that conditional on on $\mathcal{F}$, the random variables $(\boldsymbol{\omega_i})_{i \in \mathbb{T} \setminus \mathbb{S}}$ are independent of $X_\varnothing$. Let $\mathbb{U}^{(n)}$ be deterministic finite rooted subtrees of $\mathbb{T}$ that increase to $\mathbb{T}$. Let $\overline{\mathcal{F}}^{(n)}$ be the σ-field generated by $(\boldsymbol{\omega_i})_{i \in \mathbb{U}^{(n)}}$ and let $\mathcal{F}^{(n)}$ be the σ-field generated by $\mathbb{S} \cap \mathbb{U}^{(n)}$ and $(\boldsymbol{\omega_i})_{i \in \mathbb{S} \cap \mathbb{U}^{(n)}}$. Conditional on $\mathcal{F}^{(n)}$, the state at the root $X_\varnothing$ is a deterministic function of $(X_i)_{i \in \nabla(\mathbb{S} \cap \mathbb{U}^{(n)})}$. Therefore, by point (ii) in the definition of an RTP in Lemma 1.9, $X_\varnothing$ is conditionally independent of $(\boldsymbol{\omega_i})_{i \in (\mathbb{T} \setminus \mathbb{S}) \cap \mathbb{U}^{(n)}}$ given $\mathcal{F}^{(n)}$, or equivalently,

$$\mathbb{P}[X_\varnothing \in A \,|\, \overline{\mathcal{F}}^{(n)}] = \mathbb{P}[X_\varnothing \in A \,|\, \mathcal{F}^{(n)}] \quad \text{a.s.} \tag{5.17}$$

for each measurable $A \subset S$. Letting $n \to \infty$, using martingale convergence, we arrive at (5.16). $\qquad\square$

The following lemma prepares for the proof of Theorem 1.12.

Lemma 5.4 (Successful coupling). *Let $(\boldsymbol{\omega_i}, X_i)_{i \in \mathbb{T}}$ be an endogenous RTP corresponding to a solution ν of the RDE (1.54) and let $(\sigma_i)_{i \in \mathbb{T}}$ be an independent i.i.d. collection of exponential random variables with mean $|\mathbf{r}|^{-1}$. Furthermore, let $(Y_i)_{i \in \mathbb{T}}$ be an i.i.d. collection of S-valued random variables with common law ν, independent of $(\boldsymbol{\omega_i}, X_i, \sigma_i)_{i \in \mathbb{T}}$. For each $t > 0$, define random variables $(X_i^t)_{i \in \mathbb{S}_t \cup \nabla \mathbb{S}_t}$ by*

$$\begin{aligned}
&\text{(i)} \quad X_i^t := Y_i \quad (i \in \nabla \mathbb{S}_t) \\
&\text{(ii)} \quad X_i^t = \gamma[\boldsymbol{\omega_i}](X_{i1}^t, \ldots, X_{i\kappa(\boldsymbol{\omega_i})}^t) \qquad (i \in \mathbb{S}_t).
\end{aligned} \tag{5.18}$$

Then

$$X_\varnothing^t \xrightarrow[t \to \infty]{} X_\varnothing \text{ in probability}. \tag{5.19}$$

Proof. The following argument is a continuous-time version of the proofs of [AB05, Thm 11 (c)] and [MSS18, Lemma 6]. Let $\mathcal{F}_t$ be the filtration defined in (1.48). We add a final element $\mathcal{F}_\infty := \sigma(\bigcup_{t \geq 0} \mathcal{F}_t)$ to the filtration, which is the σ-algebra generated by the random tree $\mathbb{S}$ and the random variables $(\boldsymbol{\omega_i}, \sigma_i)_{i \in \mathbb{S}}$. Let $f, g : S \to \mathbb{R}$ be bounded and measurable functions. Since $X_\varnothing$ and $X_\varnothing^t$ are conditionally independent and identically distributed given $\mathcal{F}_t$, we have

$$\begin{aligned}
\mathbb{E}[f(X_\varnothing)g(X_\varnothing^t)] &= \mathbb{E}\big[\mathbb{E}[f(X_\varnothing)|\mathcal{F}_t]\mathbb{E}[g(X_\varnothing^t)|\mathcal{F}_t]\big] = \mathbb{E}\big[\mathbb{E}[f(X_\varnothing)|\mathcal{F}_t]\mathbb{E}[g(X_\varnothing)|\mathcal{F}_t]\big] \\
&\xrightarrow[t \to \infty]{} \mathbb{E}\big[\mathbb{E}[f(X_\varnothing)|\mathcal{F}_\infty]\mathbb{E}[g(X_\varnothing)|\mathcal{F}_\infty]\big] = \mathbb{E}[f(X_\varnothing)g(X_\varnothing)],
\end{aligned} \tag{5.20}$$

where we used the martingale convergence and in the last equality also endogeny and Lemma 5.3. Since (5.20) holds in particular for any bounded continuous f and g, we conclude that the law of $(X_\varnothing, X_\varnothing^t)$ converges weakly to the law of $(X_\varnothing, X_\varnothing)$, which implies (5.19). $\qquad\square$

Proof of Theorem 1.12. If (ii) holds, then $\overline{\nu}^{(2)}$ is the only fixed point in $\mathcal{P}(S^2)_\nu$ of the bivariate mean-field equation. Since a measure is a fixed point of the bivariate mean-field equation if and only if it is a fixed point of the map $T^{(2)}$, by Theorem 1.11, it follows that the RTP corresponding to ν is endogenous.

Assume, conversely, that the RTP corresponding to ν is endogenous. Consider a collection $(Y_i^1, \ldots, Y_i^n)_{i \in \mathbb{T}}$ of i.i.d. S^n-valued random variables with common law $\mu_0^{(n)}$, independent of the RTP $(\boldsymbol{\omega_i}, X_i)_{i \in \mathbb{T}}$ and the exponential lifetimes $(\sigma_i)_{i \in \mathbb{T}}$. For each $1 \leq m \leq n$ and $t > 0$, define random variables $(X_i^{m,t})_{i \in \mathbb{S}_t \cup \nabla \mathbb{S}_t}$ by

$$\begin{aligned}
&\text{(i)} \quad X_i^{m,t} := Y_i^m \quad (i \in \nabla \mathbb{S}_t) \\
&\text{(ii)} \quad X_i^{m,t} = \gamma[\boldsymbol{\omega_i}](X_{i1}^{m,t}, \ldots, X_{i\kappa(\boldsymbol{\omega_i})}^{m,t}) \qquad (i \in \mathbb{S}_t).
\end{aligned} \tag{5.21}$$

Then, by Theorem 1.6 applied to the n-variate map $\gamma^{(n)}$, we see that $(X_\varnothing^{1,t}, \ldots, X_\varnothing^{n,t})$ has law $\mu_t^{(n)}$. By endogeny we get from Lemma 5.4 that

$$(X_\varnothing^{1,t}, \ldots, X_\varnothing^{n,t}) \underset{t \to \infty}{\longrightarrow} (X_\varnothing, \ldots, X_\varnothing) \text{ in probability.} \tag{5.22}$$

This completes the proof since the right-hand side of (5.22) has law $\overline{\nu}^{(n)}$ as defined in (1.62). $\qquad\square$

Proof of Proposition 1.14. The fact that $(\rho_t)_{t \geq 0}$ solves the higher-level mean-field equation (1.71) means that

$$\tfrac{\partial}{\partial t} \langle \rho_t, \phi \rangle = \int_\Omega \mathbf{r}(\mathrm{d}\omega) \big\{ \langle T_{\check{\gamma}[\omega]}(\rho_t), \phi \rangle - \langle \rho_t, \phi \rangle \big\} \qquad (t \geq 0), \tag{5.23}$$

for any bounded measurable $\phi : \mathcal{P}(S) \to \mathbb{R}$. In particular, we can apply this to functions of the form

$$\phi(\mu) = \int_S \mu(\mathrm{d}x^1) \cdots \int_S \mu(\mathrm{d}x^n) \, f(x^1, \ldots, x^n), \tag{5.24}$$

where $f : S^n \to \mathbb{R}$ is bounded and measurable. Then

$$\langle \rho_t, \phi \rangle = \int_{S^n} \rho_t^{(n)}(\mathrm{d}x) \, f(x) \quad \text{and} \quad \langle T_{\check{\gamma}[\omega]}(\rho_t), \phi \rangle = \int_{S^n} \big(T_{\check{\gamma}[\omega]}(\rho_t) \big)^{(n)}(\mathrm{d}x) \, f(x), \tag{5.25}$$

where $(T_{\check{\gamma}[\omega]}(\rho_t))^{(n)}$ denotes the n-th moment measure of $T_{\check{\gamma}[\omega]}(\rho_t)$. By [MSS18, Lemma 2],

$$\big(T_{\check{\gamma}[\omega]}(\rho_t) \big)^{(n)} = T_{\gamma^{(n)}[\omega]}(\rho_t^{(n)}). \tag{5.26}$$

Inserting this into (5.23), we see that $(\rho_t^{(n)})_{t \geq 0}$ solves the n-variate mean-field equation. $\qquad\square$

The following lemma prepares for the proof of Proposition 1.16.

Lemma 5.5 (Conditional law of the root). *Let $(\omega_\mathbf{i}, X_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ be an RTP corresponding to a solution ν of the RDE (1.54), let $(\sigma_\mathbf{i})_{\mathbf{i} \in \mathbb{T}}$ be an independent i.i.d. collection of exponentially distributed random variables with mean $|\mathbf{r}|^{-1}$, and let $(\mathcal{F}_t)_{t \geq 0}$ be the filtration defined in (1.48). Then the measures*

$$\rho_t := \mathbb{P}\big[\mathbb{P}[X_\varnothing \in \cdot | \mathcal{F}_t] \in \cdot \big] \qquad (t \geq 0) \tag{5.27}$$

solve the higher-level mean-field equation (1.71) with initial state $\rho_0 = \delta_\nu$.

Proof. Conditional on $\mathcal{F}_t$, the map $G_t : S^{\nabla \mathbb{S}_t} \to S$ is a deterministic map, and $(X_\mathbf{i})_{\mathbf{i} \in \nabla \mathbb{S}_t}$ are i.i.d. with common law ν. Therefore, applying [MSS18, Lemma 8] to the case that the σ-fields $\mathcal{H}_k$ there are all trivial and the probability measure $\mathbb{P}$ there is replaced by the conditional law given $\mathcal{F}_t$, we see that

$$\begin{aligned} T_{\check{G}_t}(\delta_\nu) &= \mathbb{P}\big[\mathbb{P}[G_t\big((X_\mathbf{i})_{\mathbf{i} \in \nabla \mathbb{S}_t}\big) \in \cdot \,\big|\, \mathcal{F}_t] \in \cdot \,\big|\, \mathcal{F}_t \big] \\ &= \mathbb{P}\big[\mathbb{P}[X_\varnothing \in \cdot \,\big|\, \mathcal{F}_t] \in \cdot \,\big|\, \mathcal{F}_t \big] \end{aligned} \tag{5.28}$$

Now by Theorem 1.6,

$$\rho_t := \mathbb{E}[T_{\check{G}_t}(\delta_\nu)] = \mathbb{P}\big[\mathbb{P}[X_\varnothing \in \cdot \,\big|\, \mathcal{F}_t] \in \cdot \big] \qquad (t \geq 0) \tag{5.29}$$

solves the higher-level mean-field equation (1.71) with initial state $\rho_0 = \delta_\nu$. $\qquad\square$

Proof of Proposition 1.16. Let $(\rho_t^i)_{t \geq 0}$ $(i = 1, 2)$ be solutions to the higher-level mean-field equation (1.71) such that $\rho_0^1 \leq_{\mathrm{cv}} \rho_0^2$. Define $\rho_{t,(n)}^i$ as in (3.22), with T replaced by the

higher-level map $\tilde{T}$ from (1.73). It has been shown in [MSS18, Prop 3] that $\tilde{T}$ is monotone w.r.t. the convex order, so by induction we obtain from (3.22) that $\rho^1_{t,(n)} \leq_{\mathrm{cv}} \rho^2_{t,(n)}$ for all $n \geq 1$ and $t \geq 0$. Letting $n \to \infty$, using (3.30), we see that $\rho^1_t \leq_{\mathrm{cv}} \rho^2_t$ for all $t \geq 0$.

Let ν be a solution of the RDE (1.54). It has been shown in [MSS18, Prop. 3] that $\overline{\nu}$ solves the higher-level RDE (1.73) and there exists a (necessarily unique) solution $\underline{\nu}$ of (1.73) such that (1.76) holds. It has moreover been shown in [MSS18, Prop. 4] that $\underline{\nu}$ is given by (1.79). In view of this, to complete the proof, it suffices to show that the solution $(\rho_t)_{t \geq 0}$ to the higher-level mean-field equation (1.71) with initial state $\rho_0 = \delta_\nu$ converges to the measure in (1.79).

We apply Lemma 5.5. As in the proof of Lemma 5.4, we add a final element $\mathcal{F}_\infty := \sigma(\bigcup_{t \geq 0} \mathcal{F}_t)$ to the filtration, which is the σ-algebra generated by the random tree $\mathbb{S}$ and the random variables $(\boldsymbol{\omega}_{\mathbf{i}}, \sigma_{\mathbf{i}})_{\mathbf{i} \in \mathbb{S}}$. Then, by martingale convergence,

$$\mathbb{P}[X_\varnothing \in \cdot | \mathcal{F}_t] \underset{t \to \infty}{\Longrightarrow} \mathbb{P}[X_\varnothing \in \cdot | \mathcal{F}_\infty] \quad \text{a.s.,} \tag{5.30}$$

and hence the measures ρ_t in (5.27) satisfy

$$\rho_t \underset{t \to \infty}{\Longrightarrow} \mathbb{P}\big[\mathbb{P}[X_\varnothing \in \cdot | \mathcal{F}_\infty] \in \cdot \big], \tag{5.31}$$

where $\Rightarrow$ denotes weak convergence of probability measures on $\mathcal{P}(S)$, which is in turn equipped with the topology of weak convergence of probability measures on S. Since the exponentially distributed random variables $(\sigma_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ are independent of the RTP $(\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$, we have

$$\mathbb{P}[X_\varnothing \in \cdot | \mathcal{F}_\infty] = \mathbb{P}[X_\varnothing \in \cdot | \mathcal{F}] = \mathbb{P}[X_\varnothing \in \cdot | (\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}], \tag{5.32}$$

where as in Lemma 5.3 $\mathcal{F}$ denotes the σ-field generated by the random variables $\mathbb{S}$ and $(\boldsymbol{\omega}_{\mathbf{i}})_{\mathbf{i} \in \mathbb{S}}$ and the last equality follows from that lemma. Inserting this into (5.31) we see that ρ_t converges weakly to $\underline{\nu}$ as defined in (1.79). $\qquad \square$

6 Further results

In this section, we prove some additional results about RTPs. In Subsection 6.1, we prove Proposition 1.20 about the upper and lower solutions of a monotonous RDE. In Subsection 6.2 we prove Lemmas 1.21, 1.22, and 1.24 which give conditions for uniqueness of solutions to an RDE. Subsection 6.3 is devoted to the proof of Lemma 1.25.

6.1 Monotonicity

In this subsection, we prove Proposition 1.20. We start with a number of simple lemmas.

Lemma 6.1 (A continuous monotone function)**.** *Let S be a compact metrizable space that is equipped with a closed partial order in the sense of (1.90), and let d be a metric that generates the topology. Then*

$$f(x, y) := \inf \big\{ d(x', y') : x' \leq x, \ y' \geq y \big\} \qquad (x, y \in S) \tag{6.1}$$

defines a continuous function $f : S^2 \to [0, \infty)$ such that $f(x, y) = 0$ if and only if $x \geq y$ and moreover $f(x, y)$ is decreasing in x and increasing in y.

Proof. Since for any $(x, y), (x', y') \in S^2$,

$$\big| d(x, y) - d(x', y') \big| \leq d(x, x') + d(y, y'), \tag{6.2}$$

the function $d : S^2 \to [0, \infty)$ is continuous. Assume that $(x_n, y_n) \in S^2$ converge to a limit (x, y). Since the infimum of a family of continuous functions is upper semi-continuous,

we have

$$\limsup_{n \to \infty} f(x_n, y_n) \leq f(x, y). \tag{6.3}$$

To prove that f is actually continuous, assume the converse. Then there exists a sequence such that

$$\liminf_{n \to \infty} f(x_n, y_n) \leq f(x, y) - \varepsilon. \tag{6.4}$$

for some $\varepsilon > 0$. By the definition of f, there exist $x'_n \leq x_n$ and $y'_n \geq y_n$ such that $d(x'_n, y'_n) \leq f(x_n, y_n) + \varepsilon/2$. Since S is compact, we can select a subsequence such that (6.4) still holds and the (x'_n, y'_n) converge to a limit (x', y'). Since the partial order is closed in the sense of (1.90), we have $x' \leq x$ and $y' \geq y$, so

$$f(x, y) \leq d(x', y') = \lim_{n \to \infty} d(x'_n, y'_n) \leq \varepsilon/2 + \liminf_{n \to \infty} f(x_n, y_n), \tag{6.5}$$

which contradicts (6.4). We conclude that $f : S^2 \to [0, \infty)$ is continuous.

If $x \geq y$, then setting $(x', y') = (x, x)$ shows that $f(x, y) = 0$. Conversely, if $f(x, y) = 0$ then there exist $x_n \leq x$ and $y_n \geq y$ such that $d(x_n, y_n) \to 0$. Using the compactness of S, by going to a subsequence, we can assume that the (x_n, y_n) converge to a limit (z, z). Since the partial order is closed in the sense of (1.90), $y \leq z \leq x$ and hence $x \geq y$.

If $x \leq x_*$ and $y \geq y_*$, then

$$f(x_*, y_*) := \inf \big\{ d(x', y') : x' \leq x_*, \ y' \geq y_* \big\} \leq \inf \big\{ d(x', y') : x' \leq x, \ y' \geq y \big\} = f(x, y) \tag{6.6}$$

since the second infimum is taken over a smaller set, showing that $f(x, y)$ is decreasing in x and increasing in y. $\square$

Lemma 6.2 (Comparison principle). *Let S be a compact metrizable space that is equipped with a partial order that is closed in the sense of (1.90). Let X, Y be S-valued random variables such that $X \leq Y$ a.s. and $\mathbb{P}[X \in \cdot] \geq \mathbb{P}[Y \in \cdot]$. Then $X = Y$ a.s.*

Proof of Lemma 6.2. Set $f_z(x) := f(z, x)$ with f as in Lemma 6.1. Then, for each $z \in S$, $f_z : S \to [0, \infty)$ is continuous and monotone increasing, and $f_z(x) = 0$ if and only if $x \leq z$. Let

$$S^2_< := \big\{ (x, y) \in S^2 : x \leq y, \ x \neq y \big\}. \tag{6.7}$$

We will prove the lemma by showing that if X, Y are S-valued random variables such that $\mathbb{P}[(X, Y) \in S^2_<] > 0$, then $\mathbb{E}[f_z(X)] < \mathbb{E}[f_z(Y)]$ for some $z \in S$ contradicting $\mathbb{P}[X \in \cdot] \geq \mathbb{P}[Y \in \cdot]$. For each $z \in S$ and $\delta > 0$, we define an open set $O_{z,\delta} \subset S^2$ by

$$O_{z,\delta} := \{ (x, y) \in S^2 : f_z(y) - f_z(x) > \delta \}. \tag{6.8}$$

Since for each $(x, y) \in S^2_<$, one has $f_x(x) = 0$ but $f_x(y) > 0$, we see that

$$\bigcup \{ O_{z,\delta} : z \in S, \ \delta > 0 \} \supset S^2_<. \tag{6.9}$$

We now use the inner regularity of measures on Polish spaces w.r.t. compacta, which follows from the regularity and tightness of any probability measure on a Polish space [Par05, Thm. 1.2 and 3.2]. Thus, we can find a compact set $K \subset S^2_<$ such that $\mathbb{P}[(X, Y) \in K] > 0$. Since K is compact, it is covered by finitely many sets of the form (6.8), so there must exists a $z \in S$ and $\delta > 0$ such that $\mathbb{P}[(X, Y) \in O_{z,\delta}] > 0$. Since f_z is monotone increasing and $X \leq Y$ it follows that $\mathbb{E}[f_z(X)] < \mathbb{E}[f_z(Y)]$. $\square$

Lemma 6.3 (Compatibility of the stochastic order). *Assume that S is equipped with a partial order that is closed in the sense of (1.90). Then the stochastic order on $\mathcal{P}(S)$ is closed with respect to the topology of weak convergence.*

Proof. We need to show that if $\mu_n^1 \leq \mu_n^2$ for all $n \in \mathbb{N}$ and the $\mu_n^i \in \mathcal{P}(S)$ converge weakly as $n \to \infty$ to a limit μ_∞^i $(i = 1, 2)$, then $\mu_\infty^1 \leq \mu_\infty^2$. Since $\mu_n^1 \leq \mu_n^2$, for each n, we can

couple X_n^i with laws μ_n^i $(i = 1, 2)$ such that $X_n^1 \leq X_n^2$. Since μ_n^1 and μ_n^2 converge as $n \to \infty$, the joint laws of (X_n^1, X_n^2) are tight, so by going to a subsequence we may assume that they converge. Then, by Skorohod's representation theorem, we can couple the random variables (X_n^1, X_n^2) for different n in such a way that they converge a.s. to a limit (X_∞^1, X_∞^2). Since the partial order on S is closed, we have $X_\infty^1 \leq X_\infty^2$ a.s., proving that $\mu_\infty^1 \leq \mu_\infty^2$. $\qquad\square$

Lemma 6.4 (Monotonicity of T). *Assume that S is equipped with a partial order that is closed and that $\gamma[\omega]$ is monotone for all $\omega \in \Omega$. Then the operator T in (1.1) is monotone w.r.t. the stochastic order.*

Proof. If $\mu_1 \leq \mu_2$, then we can couple random variables X^1 and X^2 with laws μ_1, μ_2 such that $X^1 \leq X^2$. Let $(X_i^1, X_i^2)_{i \geq 1}$ be i.i.d. copies of (X^1, X^2). Then

$$\gamma[\omega](X_1^1, \ldots, X_{\kappa(\omega)}^1) \leq \gamma[\omega](X_1^2, \ldots, X_{\kappa(\omega)}^2) \tag{6.10}$$

for all $\omega \in \Omega$ and hence $T(\mu_1) \leq T(\mu_2)$ by (1.1). $\qquad\square$

In practice, Lemma 6.4 is the usual way to prove monotonocity of a map of the form (1.1). Nevertheless, it is known that there are maps of the form (1.1), in particular, probability kernels, that are monotone yet cannot be represented in terms of monotone maps [FM01, Example 1.1].

Lemma 6.5 (Monotonicity in the initial state). *Assume that S is equipped with a partial order that is closed and that the operator T in (1.1) is monotone w.r.t. the stochastic order. Then solutions $(\mu_t^i)_{t \geq 0}$ $(i = 1, 2)$ of the mean-field equation (1.4) started in initial states $\mu_0^1 \leq \mu_0^2$ satisfy $\mu_t^1 \leq \mu_t^2$ $(t \geq 0)$.*

Proof. Inductively define $\mu_{t,(n)}^i$ as in (3.22) with μ_0 replaced by μ_0^i $(i = 1, 2)$. Then $\mu_{t,(n)}^1 \leq \mu_{t,(n)}^2$ for all $n \geq 1$ and $t \geq 0$. Letting $n \to \infty$, we see as in the proof of Proposition 3.4 that $\mu_{t,(n)}^i \Rightarrow \mu_t^i$ as $n \to \infty$. By Lemma 6.3, we conclude that $\mu_t^1 \leq \mu_t^2$ $(t \geq 0)$. $\qquad\square$

In the next two lemmas we need to assume compactness of S.

Lemma 6.6 (Increasing limits). *Assume that S is a compact metrizable space equipped with a partial order that is closed. Then every increasing sequence in S converges to a limit.*

Proof. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in S such that $x_n \leq x_{n+1}$ for all $n \in \mathbb{N}$. By compactness, it suffices to prove that all subsequential limits are the same. Let $(x_m)_{m \in M}$ and $(x_k)_{k \in K}$ be subsequences that converge to limits x and x', respectively. For all $k \in K$, let $k_- := \sup\{m \in M : m \leq k\}$. Then $x_{k_-} \leq x_k$ for all $k \in K$ and letting $k \to \infty$, using the compatibility condition (1.90), we see that $x \leq x'$. The same argument gives $x' \leq x$ and hence $x = x'$. $\qquad\square$

Lemma 6.7 (Increasing limits in the stochastic order). *Assume that S is a compact metrizable space equipped with a partial order that is closed. Then every sequence in $\mathcal{P}(S)$ that is increasing in the stochastic order converges weakly to a limit.*

Proof. Let $(\mu_n)_{n \geq 0}$ be increasing in the stochastic order. Then, for each $n \geq 1$, we can couple random variables X_{n-1} and X_n with laws μ_{n-1} and μ_n such that $\mathbb{P}[X_{n-1} \leq X_n] = 1$. Let $K_n(x, \mathrm{d}y) := \mathbb{P}[X_n \in \mathrm{d}y \,|\, X_{n-1} = x]$ and let $(Y_n)_{n \geq 0}$ be a time-inhomogeneous Markov chain with initial law μ_0 and transition kernels $\mathbb{P}[Y_n \in \mathrm{d}y \,|\, Y_{n-1} = x] = K_n(x, \mathrm{d}y)$. Then $Y_{n-1} \leq Y_n$ a.s. for all $n \geq 1$ and hence the Y_n a.s. increase to a limit Y_∞ by Lemma 6.6. It follows that the μ_n converge weakly to the law of Y_∞. $\qquad\square$

We now turn to the proof of Proposition 1.20.

Lemma 6.8 (Lower and upper solutions)**.** *All conclusions of Proposition 1.20 except for the statement about endogeny hold when the assumption that $\gamma[\omega]$ is monotone for all $\omega \in \Omega$ is replaced by the weaker condition that T is monotone.*

Proof. The proof is similar to the proof of [AB05, Lemma 15], which in turn is based on well-known principles [Lig85, Thm III.2.3]. By symmetry, it suffices to prove the statement for ν_{low}.

Since for each $s \geq 0$, $(\mu_{s+t}^{\mathrm{low}})_{t\geq0}$ solves (1.4) with initial state $\mu_s^{\mathrm{low}} \geq \delta_0$, we conclude from Lemmas 6.4 and 6.5 that $\mu_{s+t}^{\mathrm{low}} \geq \mu_t^{\mathrm{low}}$ for each $s \geq 0$ and hence $t \mapsto \mu_t^{\mathrm{low}}$ is increasing w.r.t. the stochastic order. By Lemma 6.7, it follows that $\mu_t^{\mathrm{low}} \Rightarrow \nu_{\mathrm{low}}$ for some probability measure ν_{low} on S. Since $\mu_{s+t}^{\mathrm{low}} \Rightarrow \nu_{\mathrm{low}}$ for all $s \geq 0$, using Lemma 3.7 and the continuity of T, we see that ν_{low} is a fixed point of the mean-field equation (1.4) and hence solves the RDE (1.54).

If ν is any solution of the RDE (1.54), then $\mu_t^{\mathrm{low}} \leq \nu$ for all $t \geq 0$ by Lemma 6.5 and the fact that ν is a fixed point of (1.4). Letting $t \to \infty$, using Lemma 6.3, we see that $\nu_{\mathrm{low}} \leq \nu$. $\qquad\square$

Lemma 6.9 (Random maps applied to extremal elements)**.** *Under the assumptions of Proposition 1.20, if $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ are i.i.d. with common law $|\mathbf{r}|^{-1}\mathbf{r}$, then there exist random variables $X_{\varnothing}^{\mathrm{upp}}$ and $X_{\varnothing}^{\mathrm{low}}$ with laws ν_{upp} and ν_{low} that are given by the decreasing, resp. increasing limits*

$$X_{\varnothing}^{\mathrm{upp}} = \lim_{\mathbb{U}^{(n)}\uparrow\mathbb{S}} G_{\mathbb{U}^{(n)}}(1,\ldots,1) \quad and \quad X_{\varnothing}^{\mathrm{low}} := \lim_{\mathbb{U}^{(n)}\uparrow\mathbb{S}} G_{\mathbb{U}^{(n)}}(0,\ldots,0), \qquad (6.11)$$

where the limit does not depend on the choice of the sequence $\mathbb{U}^{(n)} \in \mathcal{T}$ such that $\mathbb{U}^{(n)} \uparrow \mathbb{S}$. Here $\mathcal{T}$ denotes the set of all finite subtrees $\mathbb{U} \subset \mathbb{T}$ such that either $\varnothing \in \mathbb{U}$ or $\mathbb{U} = \emptyset$, and for each $\mathbb{U} \in \mathcal{T}$, the random map $G_{\mathbb{U}} : S^{\nabla\mathbb{U}} \to S$ is defined in (1.46).

Proof. By symmetry, it suffices to prove the statement for $X_{\varnothing}^{\mathrm{low}}$. Since $\gamma[\omega]$ is monotone for each $\omega \in \Omega$, the map $G_{\mathbb{U}}$ is monotone for each $\mathbb{U} \in \mathcal{T}$. Define

$$X_{\varnothing}^{\mathbb{U}} := G_{\mathbb{U}}(0,\ldots,0) \qquad (\mathbb{U} \in \mathcal{T}). \qquad (6.12)$$

Then $X_{\varnothing}^{\mathbb{U}} \leq X_{\varnothing}^{\mathbb{V}}$ for all $\mathbb{U} \subset \mathbb{V}$ and hence if $\mathbb{U}^{(n)} \in \mathcal{T}$ increase to $\mathbb{S}$, then the $X_{\varnothing}^{\mathbb{U}^{(n)}}$ increase to a limit $X_{\varnothing}^{\mathrm{low}}$ that does not depend on the choice of the sequence $\mathbb{U}^{(n)}$.

Let $(\sigma_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an independent i.i.d. collection of exponential random variables with mean $|\mathbf{r}|^{-1}$. Define $\mathbb{S}_t \in \mathcal{T}$ as in (1.44). Then by Theorem 1.6, $X_{\varnothing}^{\mathbb{S}_t}$ has law μ_t^{low} while by what we have already proved $X_{\varnothing}^{\mathbb{S}_t}$ increases to $X_{\varnothing}^{\mathrm{low}}$. Since $\mu_t^{\mathrm{low}} \Rightarrow \nu_{\mathrm{low}}$, it follows that $X_{\varnothing}^{\mathrm{low}}$ has law ν_{low}. $\qquad\square$

Proof of Proposition 1.20. In view of Lemma 6.8, it only remains to prove the statement about endogeny. Let $(\omega_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$ be an RTP corresponding to γ and some solution ν to the RDE (1.54). Then

$$X_{\varnothing} := G_{\mathbb{U}}\big((X_{\mathbf{i}})_{\mathbf{i}\in\nabla\mathbb{U}}\big) \geq X_{\varnothing}^{\mathbb{U}} \qquad (\mathbb{U} \in \mathcal{T}), \qquad (6.13)$$

with $X_{\varnothing}^{\mathbb{U}}$ as in (6.12). So letting $\mathbb{U} \uparrow \mathbb{T}$, using the fact that the partial order is closed, we obtain that $X_{\varnothing} \geq X_{\varnothing}^{\mathrm{low}}$. In particular, if $\nu = \nu_{\mathrm{low}}$, then since $X_{\varnothing}^{\mathrm{low}}$ also has law ν_{low}, Lemma 6.2 tells us that $X_{\varnothing} = X_{\varnothing}^{\mathrm{low}}$ a.s. Since the latter is measurable w.r.t. the σ-field generated by the $(\omega_{\mathbf{i}})_{\mathbf{i}\in\mathbb{T}}$, this proves the endogeny of the RTP corresponding to γ and ν_{low}. $\qquad\square$

6.2 Conditions for uniqueness

In this subsection, we prove Lemmas 1.21, 1.22, and 1.24.

Proof of Lemma 1.21. If G_t is constant then $\mathbb{S}_t$ is a root determining subtree, proving the implication (i)$\Rightarrow$(ii). Conversely, if there a.s. exists a root determining subtree $\mathbb{U}$, then, since $\mathbb{S}_t \uparrow \mathbb{S}$, there a.s. exists a (random) $t < \infty$ such that $\mathbb{S}_t \supset \mathbb{U}$ and hence G_s is constant for all $s \geq t$. The implication (iii)$\Rightarrow$(ii) is trivial. Conversely, if $\mathbb{S}$ contains a root determining subtree $\mathbb{U}$, then by the finiteness of the latter we can keep removing elements from $\mathbb{U}$ as long as this is still possible while retaining the property that $\mathbb{U}$ is root determining. $\qquad\square$

Proof of Lemma 1.22. (i)$\Rightarrow$(ii): This is clear, since a finite uniquely determined subtree is root determining.

(ii)$\Rightarrow$(iii): For each $\mathbf{i} \in \mathbb{S}$, let $\mathbb{S}^{\mathbf{i}}$, defined in (5.5), denote the subtree of $\mathbb{S}$ that is rooted at $\mathbf{i}$. Since $\mathbb{S}^{\mathbf{i}}$ is equally distributed with $\mathbb{S}$, by (ii), for each $\mathbf{i} \in \mathbb{S}$, there a.s. exists a root determining subtree $\mathbb{U}^{\mathbf{i}} \subset \mathbb{S}^{\mathbf{i}}$. Since $x \in \Xi_{\mathbb{S}}$ implies $x_{\mathbf{i}} = G_{\mathbb{U}^{\mathbf{i}}}\big((x_{\mathbf{ij}})_{\mathbf{j} \in \nabla \mathbb{U}^{\mathbf{i}}}\big)$ and $G_{\mathbb{U}^{\mathbf{i}}}$ is constant, it follows that $\mathbb{S}$ is a.s. uniquely determined.

(ii)$\Rightarrow$(v): Since $G_{\mathbb{U}^{\mathbf{i}}}$ is constant, we can define

$$X_{\mathbf{i}} := G_{\mathbb{U}^{\mathbf{i}}}\big((x_{\mathbf{ij}})_{\mathbf{j} \in \nabla \mathbb{U}^{\mathbf{i}}}\big), \tag{6.14}$$

where the right-hand side does not depend on the choice of $(x_{\mathbf{ij}})_{\mathbf{j} \in \nabla \mathbb{U}^{\mathbf{i}}}$. It is straightforward to check that $(\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ satisfies conditions (i)–(iii) of Lemma 1.9 and hence is an RTP corresponding to γ. It follows that $\nu := \mathbb{P}[X_{\varnothing} \in \,\cdot\,]$ solves the RDE (1.54).

Let $(Y_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ be an independent i.i.d. collection of S-valued random variables with common law μ_0, let $(\sigma_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ be an independent i.i.d. collection of exponential random variables with mean $|\mathbf{r}|^{-1}$, and define $X_{\varnothing}^t := G_t\big((Y_{\mathbf{i}})_{\mathbf{i} \in \nabla \mathbb{S}_t}\big)$. Then $X_{\varnothing}^t$ has law μ_t by Theorem 1.6. Since $G_t = G_{\mathbb{S}_t}$ with $\mathbb{S}_t \uparrow \mathbb{S}$, we see from (6.14) that $\mathbb{P}[X_{\varnothing}^t \neq X_{\varnothing}] \to 0$ as $t \to \infty$, proving that $\|\mu_t - \nu\| \to 0$.

(iii)$\Rightarrow$(iv): We note the following general principle: if S_1, S_2, S_3 are Polish spaces and (X_1, X_2) and (X_1', X_3) are random variables taking values in $S_1 \times S_2$ resp. $S_1 \times S_3$ such that X_1 and X_1' are equal in law, then we can couple (X_1, X_2) and (X_1', X_3) such that $X_1 = X_1'$. To see this, let μ denote the law of X_1, let $K_i(x_1, \mathrm{d}x_i)$ denote a regular version of the conditional law of X_i given X_1 resp. X_1' ($i = 1, 2$), and define the joint law of X_1, X_2, X_3 as

$$\mathbb{P}[X_1 \in A_1, \ X_2 \in A_2, \ X_3 \in A_3] := \int_{A_1} \mu(\mathrm{d}x_1) \int_{A_2} K_2(x_1, \mathrm{d}x_2) \int_{A_3} K_3(x_1, \mathrm{d}x_3), \tag{6.15}$$

i.e., make X_2 and X_3 conditionally independent given X_1. Applying this general principle, we see that if ν_1, ν_2 are solutions to the RDE (1.54), then we can couple the associated RTPs $(\boldsymbol{\omega}_{\mathbf{i}}, X_{\mathbf{i}}^1)_{\mathbf{i} \in \mathbb{T}}$ and $(\boldsymbol{\omega}_{\mathbf{i}}', X_{\mathbf{i}}^2)_{\mathbf{i} \in \mathbb{T}}$ in such a way that $\boldsymbol{\omega}_{\mathbf{i}} = \boldsymbol{\omega}_{\mathbf{i}}'$ for all $\mathbf{i} \in \mathbb{T}$. Since $\mathbb{S}$ is a.s. uniquely determined, it follows that $X_{\varnothing}^1 = X_{\varnothing}^2$ a.s. and hence $\nu_1 = \nu_2$. The same argument also shows that any solution to the bivariate RDE is concentrated on the diagonal, which by Theorem 1.11 implies endogeny.

(iii) and S finite imply (ii): Since $\mathbb{V} \subset \mathbb{U}$ and $\mathbb{V}$ root determining imply that $\mathbb{U}$ is root determining, we see that $\mathbb{P}[G_t \text{ not constant}]$ decreases to $\mathbb{P}[G_t \text{ not constant } \forall t \geq 0]$. Assume that this event has positive probability and condition on it. Choose $t(n) \to \infty$. Then there exist $x^n, y^n \in \Xi_{\mathbb{S}_{t(n)}}$ such that $x_{\varnothing}^n \neq y_{\varnothing}^n$. Since S is finite, the sequences x^n and y^n have subsequences that converge pointwise for each $\mathbf{i} \in \mathbb{S}$ to limits x^∞, y^∞. It is easy to see that $x^\infty, y^\infty \in \Xi_{\mathbb{S}}$. Moreover, $x_{\varnothing}^\infty \neq y_{\varnothing}^\infty$. This shows that on the event $\{G_t \text{ not constant } \forall t \geq 0\}$, the tree $\mathbb{S}$ is not uniquely determined.

(ii) and $S = \{0,1\}$ imply (i): It suffices to show that each root determining subtree $\mathbb{U}$ of $\mathbb{S}$ contains a uniquely determined subtree. For any $\mathbf{i} \in \mathbb{T}$, let $\mathbb{T}^{(\mathbf{i})} := \{\mathbf{ij} : \mathbf{j} \in \mathbb{T}\}$ denote $\mathbf{i}$ and its descendants, and let $\Xi_{\mathbb{U},\mathbf{i}}$ denote the set of all $(x_{\mathbf{j}})_{\mathbf{j} \in (\mathbb{U} \cup \nabla\mathbb{U}) \cap \mathbb{T}^{(\mathbf{i})}}$ that satisfy

$$x_{\mathbf{j}} = \gamma[\boldsymbol{\omega}_{\mathbf{j}}](x_{\mathbf{j}1}, \ldots, x_{\mathbf{j}\kappa(\boldsymbol{\omega}_{\mathbf{j}})}) \qquad (\mathbf{j} \in \mathbb{U} \cap \mathbb{T}^{(\mathbf{i})}). \tag{6.16}$$

Define $(\chi_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U} \cup \nabla\mathbb{U}}$ by

$$\chi_{\mathbf{i}} := \{x_{\mathbf{i}} : x \in \Xi_{\mathbb{U},\mathbf{i}}\} \qquad (\mathbf{i} \in \mathbb{U} \cup \nabla\mathbb{U}). \tag{6.17}$$

We claim that

If $\mathbb{V} \subset \mathbb{U}$ is a subtree, then any $x \in \Xi_{\mathbb{V}}$ that satisfies $x_{\mathbf{i}} \in \chi_{\mathbf{i}}$ for all $\mathbf{i} \in \nabla\mathbb{V}$ can be extended to an $x \in \Xi_{\mathbb{U}}$. $\tag{6.18}$

Indeed, this follows from the fact that the sets $(\mathbb{U} \cup \nabla\mathbb{U}) \cap \mathbb{T}^{(\mathbf{i})}$ for different $\mathbf{i} \in \nabla\mathbb{V}$ are mutually disjoint, which allows us to choose $x \in \Xi_{\mathbb{U},\mathbf{i}}$ independently for each $\mathbf{i} \in \nabla\mathbb{V}$.

Note that $\chi_{\mathbf{i}} = S$ for $\mathbf{i} \in \nabla\mathbb{U}$ and $|\chi_{\varnothing}| = 1$ if $\mathbb{U}$ is root determining. Let $\mathbb{V}$ be the connected component of $\{\mathbf{i} \in \mathbb{U} : |\chi_{\mathbf{i}}| = 1\}$ that contains $\varnothing$. Since S has only two elements, $\chi_{\mathbf{i}} = S$ for $\mathbf{i} \in \nabla\mathbb{V}$. Using (6.18), it follows that $\mathbb{V}$ is uniquely determined. $\qquad\square$

For completeness, we give three examples to show that the implications (ii)$\Rightarrow$(i), (iii)$\Rightarrow$(ii), and (v)$\Rightarrow$(ii) do not hold in general. In all of these examples, $\kappa(\omega) = 1$ for all $\omega \in \Omega$, which means $\mathbb{S} = \{\varnothing, 1, 11, 111, \ldots\} = \{1_{(n)} : n \geq 0\}$, where $1_{(n)}$ denotes the word of length n made from the alphabet $\{1\}$. It follows that the operator T from (1.1) is just the linear operator associated with the transition kernel of a Markov chain. In all our examples, we take $\gamma[\omega] = g$ for all $\omega \in \Omega$, where $g : S \to S$ is a fixed map.

Example 6.10 ((ii)$\not\Rightarrow$(i))**.** Let $S = \{0,1,2\}$ and $g(x) := (x-1) \vee 0$. Then $\mathbb{S}$ a.s. contains a root determining subtree but $\mathbb{S}$ a.s. does not contain a uniquely determined subtree.

Proof. The subtree $\mathbb{U} := \{\varnothing, 1\}$ is root determining, since $G_{\mathbb{U}}(x) = (x-2) \vee 0 = 0$ for all $x \in S^{\nabla\mathbb{U}} = S$. On the other hand, if $\mathbb{V} = \{\varnothing, 1, 11, \ldots, 1_{(n)}\}$ is a finite subtree of $\mathbb{S}$ that contains the root, then there exist $x, y \in \Xi_{\mathbb{V}}$ with $x_{1_{(n)}} = 0$ and $y_{1_{(n)}} = 1$, which shows $\mathbb{V}$ is not uniquely determined. $\qquad\square$

Example 6.11 ((iii)$\not\Rightarrow$(ii))**.** Let $S = \mathbb{N}$ and

$$g(x) := \begin{cases} 0 & \text{if } x = 0, \\ x+1 & \text{if } x \geq 1. \end{cases} \tag{6.19}$$

Then $\mathbb{S}$ is a.s. uniquely determined but $\mathbb{S}$ a.s. contains no root determining subtree.

Proof. If $x \in \Xi_{\mathbb{S}}$ satisfies $x_{1_{(n)}} = m \neq 0$ for some n, then $x_{1_{(n+k)}} \neq 0$ and $x_{1_{(n+k)}} = m - k$ for all $k \geq 0$, which leads to a contradiction. It follows that $\Xi_{\mathbb{S}}$ contains a single element, which is given by $x_{1_{(n)}} = 0$ for all $n \geq 0$. In particular, $\mathbb{S}$ is uniquely determined. On the other hand, for each finite subtree $\mathbb{U} = \{\varnothing, 1, 11, \ldots, 1_{(n-1)}\}$ that contains the root, the function $G_{\mathbb{U}}$ is of the form $G_{\mathbb{U}}(0) = 0$ and $G_{\mathbb{U}}(x) = x + n$ $(x \geq 1)$, which is clearly not constant. $\qquad\square$

Example 6.12 ((v)$\not\Rightarrow$(ii))**.** Let $S = \{0,1\}$ and $g(x) := 1 - x$ with $\pi(\{g\}) > 0$. Then the RDE (1.54) has a solution ν that is globally attractive in the sense that any solution $(\mu_t)_{t \geq 0}$ to (1.4) satisfies $\|\mu_t - \nu\| \xrightarrow[t \to \infty]{} 0$, where $\|\cdot\|$ denotes the total variation norm. Nevertheless, $\mathbb{S}$ contains no root determining subtree.

Proof. Since the continuous-time Markov chain that jumps from x to $1 - x$ with rate $\pi(\{g\})$ is ergodic, the RDE (1.54) has a solution ν that is globally attractive. On the other hand, if $\mathbb{U} = \{\varnothing, 1, 11, \ldots, 1_{(n-1)}\}$ is a finite subtree that contains the root, then $G_{\mathbb{U}}(x) = x$ if n is even and $G_{\mathbb{U}}(x) = 1 - x$ if n is odd, so $G_{\mathbb{U}}$ is not constant. $\qquad\square$

Proof of Lemma 1.24. By Lemma 1.22, it suffices to prove that if the RDE (1.54) has a unique solution, then G_t is constant for t large enough. By Proposition 1.20, the RDE (1.54) has a unique solution if and only if $\nu_{\text{low}} = \nu_{\text{upp}}$. Let 0 and 1 denote the minimal and maximal elements of S. By Lemma 6.9, $G_t(0, \ldots, 0)$ and $G_t(1, \ldots, 1)$ converge as $t \to \infty$ to a.s. limits with laws ν_{low} and ν_{upp}, respectively. Since $\gamma[\omega]$ is monotone for each ω, the maps G_t are monotone, and hence

$$G_t(0, \ldots, 0) \leq G_t(x) \leq G_t(1, \ldots, 1) \tag{6.20}$$

for all $x \in S^{\nabla \mathbb{S}_t}$. Since S is finite, if the laws of the left- and right-hand sides of (6.20) converge to the same limit, then $\lim_{t \to \infty} \mathbb{P}[G_t(0, \ldots, 0) = G_t(1, \ldots, 1)] = 1$, proving that G_t is constant for t large enough. $\qquad\square$

6.3 Duality

In this subsection, we prove Lemma 1.25. For a start, we will generalize quite a bit and assume that S is a finite partially ordered set and that $\gamma[\omega] : S^{\kappa(\omega)} \to S$ is monotone for all $\omega \in \Omega$, where $S^{\kappa(\omega)}$ is equipped with the product partial order. As in Subsection 5.1, we let $\mathcal{T}$ denote the set of all finite subtrees $\mathbb{U} \subset \mathbb{T}$ such that either $\varnothing \in \mathbb{U}$ or $\mathbb{U} = \emptyset$. For each $\mathbb{U} \in \mathcal{T}$, we define $G_{\mathbb{U}} : S^{\nabla \mathbb{U}} \to S$ as in (1.46), where $\nabla \mathbb{U} := \{\varnothing\}$ if $\mathbb{U} = \emptyset$.

For any $\mathbb{U} \in \mathcal{T}$, we let $\Sigma_{\mathbb{U}}$ denote the set of all $(y_{\mathbf{i}})_{\mathbf{i} \in \mathbb{U} \cup \nabla \mathbb{U}}$ that satisfy

$$(y_{\mathbf{i}1}, \ldots, y_{\mathbf{i}\kappa(\boldsymbol{\omega}_{\mathbf{i}})}) \text{ is a minimal element of } \{x \in S^{\kappa(\boldsymbol{\omega}_{\mathbf{i}})} : \gamma[\boldsymbol{\omega}_{\mathbf{i}}](x) \geq y_{\mathbf{i}}\} \quad (\mathbf{i} \in \mathbb{U}). \tag{6.21}$$

Lemma 6.13 (Monotone duality). *For any $\mathbb{U} \in \mathcal{T}$, $x \in S^{\nabla \mathbb{U}}$, and $z \in S$, one has $G_{\mathbb{U}}(x) \geq z$ if and only if there exists a $y \in \Sigma_{\mathbb{U}}$ such that $y_{\varnothing} = z$ and $x \geq y$ on $\nabla \mathbb{U}$.*

Proof. Fix $z \in S$. For each $\mathbb{U} \in \mathcal{T}$, let us write

$$Y_{\mathbb{U}} := \{(y_{\mathbf{i}})_{\mathbf{i} \in \nabla \mathbb{U}} : y \in \Sigma_{\mathbb{U}}, \, y_{\varnothing} = z\}. \tag{6.22}$$

Then we need to show that

$$G_{\mathbb{U}}(x) \geq z \quad \Leftrightarrow \quad \exists y \in Y_{\mathbb{U}} \text{ s.t. } x \geq y. \tag{6.23}$$

The proof is by induction on the number of elements of $\mathbb{U}$. If $\mathbb{U} = \emptyset$, then $G_{\mathbb{U}}$ is the identity map, $Y_{\mathbb{U}} = \{z\}$, and the statement is trivial.

We will show that if the statement is true for $\mathbb{U}$ and if $\mathbf{j} \in \nabla \mathbb{U}$, then the statement is also true for $\mathbb{V} := \mathbb{U} \cup \{\mathbf{j}\}$. Let $x \in S^{\nabla \mathbb{V}}$ and inductively define $x_{\mathbf{i}}$ for $\mathbf{i} \in \mathbb{V}$ as in (1.45). By the induction hypothesis, $x_{\varnothing} \geq z$ if and only if

$$(x_{\mathbf{i}})_{\mathbf{i} \in \nabla \mathbb{U}} \geq y \text{ for some } y \in Y_{\mathbb{U}}. \tag{6.24}$$

Here $\nabla \mathbb{V} = (\nabla \mathbb{U} \backslash \{\mathbf{j}\}) \cup \{\mathbf{j}1, \ldots, \mathbf{j}\kappa(\boldsymbol{\omega}_{\mathbf{j}})\}$ and

$$Y_{\mathbb{V}} = \Big\{ y \in S^{\nabla \mathbb{V}} : \exists y' \in \Sigma^{\nabla \mathbb{U}} \text{ s.t. } y_{\mathbf{i}} = y'_{\mathbf{i}} \, \forall \mathbf{i} \in \nabla \mathbb{U} \backslash \{\mathbf{j}\}$$
$$\text{and} \quad (y_{\mathbf{j}1}, \ldots, y_{\mathbf{j}\kappa(\boldsymbol{\omega}_{\mathbf{j}})}) \text{ is a minimal element of } \{x \in S^{\kappa(\boldsymbol{\omega}_{\mathbf{j}})} : \gamma[\boldsymbol{\omega}_{\mathbf{j}}](x) \geq y_{\mathbf{j}}\} \Big\}. \tag{6.25}$$

It follows that (6.24) is equivalent to

$$(x_{\mathbf{i}})_{\mathbf{i} \in \nabla \mathbb{V}} \geq y \text{ for some } y \in Y_{\mathbb{V}}, \tag{6.26}$$

which completes the induction step of the proof. $\qquad\square$

Lemma 6.14 (Minimal elements). *Assume that for all $\omega \in \Omega$, there do not exist $z, z' \in S$ and minimal elements y, y' of $\{y : \gamma[\omega](y) \geq z\}$ resp. $\{y' : \gamma[\omega](y') \geq z'\}$ such that $z \not\leq z'$ but $y \leq y'$. Fix $z \in S$. For any $\mathbb{U} \in \mathcal{T}$, define $Y_{\mathbb{U}}$ as in (6.22) dependent on z. Then*

$$Y_{\mathbb{U}} = \big\{ y \in S^{\nabla\mathbb{U}} : y \text{ is a minimal element of } G_{\mathbb{U}}^{-1}(\{z\}) \big\}. \tag{6.27}$$

Proof. By Lemma 6.13,

$$G_{\mathbb{U}}^{-1}\big(\{z\}\big) = \big\{ x \in S^{\nabla\mathbb{U}} : x \geq y \text{ for some } y \in Y_{\mathbb{U}} \big\}. \tag{6.28}$$

In view of this, it suffices to prove that

$$Y_{\mathbb{U}} \text{ does not contain two elements } y, y' \text{ with } y \neq y' \text{ and } y \leq y'. \tag{6.29}$$

The proof is by induction on the number of elements of $\mathbb{U}$. If $\mathbb{U} = \emptyset$, then $\nabla\mathbb{U} = \{\varnothing\}$ and $Y_{\mathbb{U}}$ consists of a single element that has $y_{\varnothing} = z$, so (6.29) is satisfied. Assume that (6.29) holds for $\mathbb{U}$ and let $\mathbb{V} := \mathbb{U} \cup \{\mathbf{i}\}$ for some $\mathbf{i} \in \nabla\mathbb{U}$. Then (6.25) and the assumption of the lemma imply that (6.29) holds for $\mathbb{V}$. $\square$

Lemma 6.15 (Sets with two elements). *Assume that $S = \{0, 1\}$ and that $\gamma[\omega](0, \ldots, 0) = 0$ for all $\omega \in \Omega$. Then the assumption of Lemma 6.14 is satisfied.*

Proof. If $z \not\leq z'$ then we must have $z = 1$ and $z' = 0$, so we must show that there do not exist minimal elements y, y' of $\{y : \gamma[\omega](y) \geq 1\}$ resp. $\{y' : \gamma[\omega](y') \geq 0\}$ such that $y \leq y'$. Clearly, $\{y' : \gamma[\omega](y') \geq 0\} = \{0, 1\}^{\kappa(\omega)}$ has only one minimal element, which is the configuration $(0, \ldots, 0) \in \{0, 1\}^{\kappa(\omega)}$, so we must show that there does not exist a minimal element y of $\{y : \gamma[\omega](y) \geq 1\}$ such that $y \leq (0, \ldots, 0)$. Equivalently, this says that $\gamma[\omega](0, \ldots, 0) \not\geq 1$ which is satisfied since $\gamma[\omega](0, \ldots, 0) = 0$. $\square$

Lemma 6.16 (Lower and upper solutions). *Assume that S is a finite partially ordered set that contains minimal and maximal elements, denoted by 0 and 1. Assume that $\gamma[\omega]$ is monotone for all $\omega \in \Omega$. Then, for all $z \in S$,*

$$\begin{aligned}
\nu_{\mathrm{upp}}\big(\{x : x \geq z\}\big) &= \mathbb{P}\big[\exists y \in \Sigma_{\mathbb{S}} \text{ s.t. } y_{\varnothing} = z\big] \\
\nu_{\mathrm{low}}\big(\{x : x \geq z\}\big) &= \mathbb{P}\big[\exists y \in \Sigma_{\mathbb{S}} \text{ s.t. } y_{\varnothing} = z \text{ and } \{\mathbf{i} \in \mathbb{S} : y_{\mathbf{i}} \neq 0\} \text{ is finite}\big].
\end{aligned} \tag{6.30}$$

Proof. By Lemma 6.13, $G_{\mathbb{U}}(1, \ldots, 1) \geq z$ if and only if $\Sigma_{\mathbb{U}}^z := \{y \in \Sigma_{\mathbb{U}} : y_{\varnothing} = z\}$ is not empty. If $\mathbb{U} \subset \mathbb{V}$, then $\Sigma_{\mathbb{V}}^z \neq \emptyset$ implies $\Sigma_{\mathbb{U}}^z \neq \emptyset$, so the events $\{\Sigma_{\mathbb{U}(n)}^z \neq \emptyset\}$ decrease to a limit. We claim that this is the event $\{\Sigma_{\mathbb{S}}^z \neq \emptyset\}$. Since the restriction of an element $y \in \Sigma_{\mathbb{S}}^z$ to $\mathbb{U}$ yields an element of $\Sigma_{\mathbb{U}}^z$, it is clear that

$$\big\{ \Sigma_{\mathbb{U}(n)}^z \neq \emptyset \;\forall n \big\} \supset \big\{ \Sigma_{\mathbb{S}}^z \neq \emptyset \big\}. \tag{6.31}$$

Conversely, if for each n there exists some $y(n) \in \Sigma_{\mathbb{U}(n)}^z$, then by the finiteness of S we can select a subsequence of the $y(n)$ that converges pointwise to a limit y. Since $y \in \Sigma_{\mathbb{S}}^z$, this proves the other inclusion. By Lemma 6.9, it follows that

$$\mathbb{P}[X_{\varnothing}^{\mathrm{upp}} \geq z] = \mathbb{P}\big[\Sigma_{\mathbb{S}}^z \neq \emptyset\big]. \tag{6.32}$$

By Lemma 6.13, $G_{\mathbb{U}}(0, \ldots, 0) \geq z$ if and only if $\Sigma_{\mathbb{U}}^z$ contains an element y such that $y_{\mathbf{i}} = 0$ for all $\mathbf{i} \in \nabla\mathbb{U}$. Since for each $\omega \in \Omega$, the zero configuration $(0, \ldots, 0)$ is the unique minimal element of $\{x \in S^{\kappa(\omega)} : \gamma[\omega](x) \geq 0\}$, we observe that if $y \in \Sigma_{\mathbb{U}}$ satisfies $y_{\mathbf{i}} = 0$ for all $\mathbf{i} \in \nabla\mathbb{U}$, then y can uniquely by extended to an element of $\Sigma_{\mathbb{V}}$ for any $\mathbb{V} \supset \mathbb{U}$ by putting $y_{\mathbf{i}} := 0$ for $\mathbf{i} \in (\mathbb{V} \cup \nabla\mathbb{V}) \backslash (\mathbb{U} \cup \nabla\mathbb{U})$. In view of this, by Lemma 6.9,

$$\mathbb{P}[X_{\varnothing}^{\mathrm{low}} \geq z] = \mathbb{P}\big[\exists y \in \Sigma_{\mathbb{S}}^z \text{ s.t. } \{\mathbf{i} \in \mathbb{S} : y_{\mathbf{i}} \neq 0\} \text{ is finite}\big]. \tag{6.33}$$

$\square$

Proof of Lemma 1.25. Recall the definition of $\Sigma_{\mathbb{U}}$ in (6.21). We observe that $\mathbb{O}$ is an open subtree of $\mathbb{U}$ if and only if its indicator function $1_{\mathbb{O}}$ satisfies $1_{\mathbb{O}} \in \Sigma_{\mathbb{U}}$ and $1_{\mathbb{O}}(\varnothing) = 1$. In view of this, (1.98) is just a special case of Lemma 6.16. Formula (1.99) follows from Lemmas 6.14 and 6.15 applied to $\mathbb{U} = \mathbb{S}_t$. $\qquad\square$

7 Cooperative branching

In this section we prove all results that deal specifically with our running example of a system with cooperative branching and deaths. In Subsection 7.1, we prove Proposition 1.13 about the bivariate mean-field equation. In Subsection 7.2, we prove Theorem 1.18 and Lemma 1.19 about the higher-level mean-field equation. In Subsection 7.3, finally, we prove Lemmas 1.23 and 1.26 which illustrate the concepts of minimal root determining subtrees and open subtrees in the concrete set-up of our example.

7.1 The bivariate mean-field equation

In this subsection we prove Proposition 1.13. We identify a measure $\mu^{(2)}$ on $\{0,1\}^2$ with the function $\mu^{(2)} : \{0,1\}^2 \to \mathbb{R}$ defined as $\mu^{(2)}(0,0) := \mu(\{(0,0)\})$, $\mu^{(2)}(0,1) := \mu(\{(0,1)\})$, etc. We parametrize a measure $\mu^{(2)} \in \mathcal{P}_{\mathrm{sym}}(\{0,1\}^2)$ by the parameters

$$p := \mu^{(2)}(1,0) + \mu^{(2)}(1,1) \quad \text{and} \quad r := \mu^{(2)}(0,1) + \mu^{(2)}(1,0) + \mu^{(2)}(1,1). \tag{7.1}$$

We observe that $\mu^{(2)}(0,0) = 1 - r$, $\mu^{(2)}(1,0) = \mu^{(2)}(0,1) = r - p$, and hence $\mu^{(2)}(1,1) = 1 - (1-r) - 2(r-p) = 2p - r$. It follows that p and r determine $\mu^{(2)}$ uniquely and indeed, the map

$$\mathcal{P}_{\mathrm{sym}}(\{0,1\}^2) \ni \mu^{(2)} \mapsto (p,r) \in \mathcal{D} := \big\{(p,r) \in \mathbb{R}^2 : 0 \le p \le 1, \ p \le r \le 1 \wedge 2p\big\} \tag{7.2}$$

is a bijection. Moreover, $\mu^{(2)}$ is concentrated on the diagonal $\{(0,0),(1,1)\}$ if and only if $p = r$. A function $(\mu_t^{(2)})_{t \ge 0}$ with values in $\mathcal{P}_{\mathrm{sym}}(\{0,1\}^2)$ gives through (7.1) rise to a function $(p_t, r_t)_{t \ge 0}$ taking values in $\mathcal{D}$.

Lemma 7.1 (Change of parameters). *A function $(\mu_t^{(2)})_{t \ge 0}$ with values in $\mathcal{P}_{\mathrm{sym}}(\{0,1\}^2)$ solves (1.65) if and only if the associated function $(p_t, r_t)_{t \ge 0}$ solves*

$$\left.\begin{array}{ll} \text{(i)} & \frac{\partial}{\partial t} p_t = \alpha p_t^2 (1 - p_t) - p_t, \\[2mm] \text{(ii)} & \frac{\partial}{\partial t} r_t = \alpha\big[r_t^2 - 2(r_t - p_t)^2\big](1 - r_t) - r_t, \end{array}\right\} \ (t \ge 0). \tag{7.3}$$

Proof. As noted in Section 1.6, if $\mu_t^{(2)}$ solves the bivariate mean-field equation, then its one-dimensional marginals solve the mean-field equation (1.4). Since $\mu_t^{(2)}$ is symmetric, both marginals are the same. We denote these by μ_t. Then $p_t := \mu_t(\{1\})$ and the equation we find for p_t is the same as in (1.36).

We will now obtain the equation for the parameter r_t. By definitions (1.12) and (1.13) we have for any $\mu^{(2)} \in \mathcal{P}(\{0,1\}^2)$ that $T_{\mathsf{cob}^{(2)}}(\mu^{(2)})$ is the law of the random variable

$$\big(X_1^1 \vee (X_2^1 \wedge X_3^1), X_1^2 \vee (X_2^2 \wedge X_3^2)\big), \tag{7.4}$$

where (X_i^1, X_i^2) $(i = 1,2,3)$ are i.i.d. with law $\mu^{(2)}$. It follows that

$$T_{\mathsf{cob}^{(2)}}(\mu^{(2)})(0,0) = \mathbb{P}\big[(X_1^1, X_1^2) = (0,0)\big]\Big(1 - \mathbb{P}\big[(X_2^1, X_2^2) \ne (0,0)\big]\mathbb{P}\big[(X_3^1, X_3^2) \ne (0,0)\big]$$

$$+ \mathbb{P}\big[(X_2^1, X_2^2) = (0,1)\big]\mathbb{P}\big[(X_3^1, X_3^2) = (1,0)\big] + \mathbb{P}\big[(X_2^1, X_2^2) = (1,0)\big]\mathbb{P}\big[(X_3^1, X_3^2) = (0,1)\big]\Big)$$

$$= (1-r)\big(1 - r^2 + 2(r-p)^2\big).$$

$$\tag{7.5}$$

Similar, but simpler considerations give

$$T_{\mathsf{dth}^{(2)}}(\mu^{(2)})(0,0) = 1. \tag{7.6}$$

Equation (1.65) in the point $(0,0)$ now gives

$$\tfrac{\partial}{\partial t}(1 - r_t) = \tfrac{\partial}{\partial t}\mu_t^{(2)}(0,0) = \alpha\big\{(1 - r_t)\big(1 - r_t^2 + 2(r_t - p_t)^2\big) - (1 - r_t)\big\} + \big\{1 - (1 - r_t)\big\}, \tag{7.7}$$

which simplifies to the second equation in (7.3). $\qquad\square$

In view of Lemma 7.1 and the remarks that precede it, Proposition 1.13 follows from the following proposition.

Proposition 7.2 (Bivariate differential equation). *For $\alpha > 4$, the equation (7.3) has four fixed points in the space $\mathcal{D}$ defined in (7.2), which are of the form*

$$(z_{\mathrm{low}}, z_{\mathrm{low}}), \quad (z_{\mathrm{mid}}, z_{\mathrm{mid}}), \quad (z_{\mathrm{mid}}, r_{\mathrm{mid}}), \quad \text{and} \quad (z_{\mathrm{upp}}, z_{\mathrm{upp}}), \tag{7.8}$$

with $z_{\mathrm{low}}, z_{\mathrm{mid}}, z_{\mathrm{upp}}$ as in (1.37) and $z_{\mathrm{mid}} < r_{\mathrm{mid}}$. Solutions to (7.3) started in $\mathcal{D}$ converge to one of these fixed points, the domains of attraction being

$$\{(p, r) : p < z_{\mathrm{mid}}\}, \quad \{(z_{\mathrm{mid}}, z_{\mathrm{mid}})\}, \quad \{(z_{\mathrm{mid}}, r) : r > z_{\mathrm{mid}}\}, \quad \text{and} \quad \{(p, r) : p > z_{\mathrm{mid}}\}, \tag{7.9}$$

respectively. For $\alpha = 4$, the equation (7.3) has two fixed points in the space $\mathcal{D}$, which are

$$(z_{\mathrm{low}}, z_{\mathrm{low}}) \quad \text{and} \quad (z_{\mathrm{mid}}, z_{\mathrm{mid}}) = (z_{\mathrm{upp}}, z_{\mathrm{upp}}), \tag{7.10}$$

with domains of attraction

$$\{(p, r) : p < z_{\mathrm{mid}}\} \quad \text{and} \quad \{(p, r) : p \geq z_{\mathrm{mid}}\}. \tag{7.11}$$

For $\alpha < 4$, $(z_{\mathrm{low}}, z_{\mathrm{low}})$ is the only fixed point in $\mathcal{D}$ and its domain of attraction is the whole space $\mathcal{D}$.

Proof. In Section 1.3 we have found all fixed points of (7.3) (i) and determined their domains of attraction. It is clear from (7.3) that if z is a fixed point of (7.3) (i), then (z, z) is a fixed point of (7.3), so $(z_{\mathrm{low}}, z_{\mathrm{low}})$ and for $\alpha \geq 4$ also $(z_{\mathrm{mid}}, z_{\mathrm{mid}})$ and $(z_{\mathrm{upp}}, z_{\mathrm{upp}})$ are fixed points of (7.3).

If $\alpha \geq 4$ and $p_0 < z_{\mathrm{mid}}$ or if $\alpha < 4$ and p_0 is arbitrary, then we have seen in Section 1.3 that solutions to (7.3) (i) satisfy $p_t \to 0 = z_{\mathrm{low}}$ as $t \to \infty$. Since $0 \leq r_t \leq 2p_t$, it follows that also $r_t \to 0$. This proves the statements of the proposition about the domain of attraction of $(z_{\mathrm{low}}, z_{\mathrm{low}})$ for all values of α.

Let

$$P_\alpha(p) := \alpha p^2(1 - p) - p \quad \text{and} \quad R_{\alpha,p}(r) := \alpha\big[r^2 - 2(r - p)^2\big](1 - r) - r \tag{7.12}$$

denote the drift functions of p_t and r_t, respectively. We observe that $R_{\alpha,p}(r) \leq P_\alpha(r)$ $(p, r \in \mathbb{R})$ and $P_\alpha(r) < 0$ for all $z_{\mathrm{upp}} < r \leq 1$, which implies that

$$\sup_{p \in \mathbb{R}} R_{\alpha,p}(r) < 0 \qquad (z_{\mathrm{upp}} < r \leq 1). \tag{7.13}$$

It follows that solutions of (7.3) satisfy

$$\limsup_{t \to \infty} r_t \leq z_{\mathrm{upp}}. \tag{7.14}$$

If $\alpha > 4$ and $p_0 > z_{\mathrm{mid}}$ or if $\alpha = 4$ and $p_0 \geq z_{\mathrm{mid}}$, we have seen in Section 1.3 that solutions to (7.3) (i) satisfy $p_t \to z_{\mathrm{upp}}$ as $t \to \infty$. Combining this with (7.14) and the fact that $p_t \leq r_t$, we see that $(p_t, r_t) \to (z_{\mathrm{upp}}, z_{\mathrm{upp}})$.

To complete the proof, we must investigate the long-time behavior of solutions of (7.3) when $\alpha > 4$ and $p_0 = z_{\mathrm{mid}}$. In this case $p_t = z_{\mathrm{mid}}$ for all $t \geq 0$ and r_t takes values in $[z_{\mathrm{mid}}, 2z_{\mathrm{mid}}]$ and solves the differential equation

$$\tfrac{\partial}{\partial t} r_t = R_{\alpha, z_{\mathrm{mid}}}(r_t) \qquad (t \geq 0). \tag{7.15}$$

It is clear $r_t = z_{\mathrm{mid}}$ for all $t \geq 0$ is a solution. Since $z_{\mathrm{mid}} < 1/2$, in view of (7.2), we must prove that all solutions with $z_{\mathrm{mid}} < r_0 \leq 2z_{\mathrm{mid}}$ converge to a nontrivial fixed point. We write

$$R_{\alpha, z_{\mathrm{mid}}}(r) = P_\alpha(r) - 2\alpha(r - z_{\mathrm{mid}})^2(1 - r). \tag{7.16}$$

Since the first term has a positive slope at $r = z_{\mathrm{mid}}$ while the second term has zero slope, we conclude that $R_{\alpha, z_{\mathrm{mid}}}$ has a positive slope at $r = z_{\mathrm{mid}}$. Since solutions to (7.3) do not leave the domain $\mathcal{D}$, we must have $R_{\alpha, z_{\mathrm{mid}}}(2z_{\mathrm{mid}}) \leq 0$. Since $R_{\alpha, z_{\mathrm{mid}}}(r) = \alpha r^3 + O(r^2)$ as $r \to \infty$, we must have $R_{\alpha, z_{\mathrm{mid}}}(r) > 0$ for r sufficiently large. These observations imply that the cubic function $R_{\alpha, z_{\mathrm{mid}}}$ has three zeros $r_{\mathrm{low}} < r_{\mathrm{mid}} < r_{\mathrm{up}}$ with

$$r_{\mathrm{low}} = z_{\mathrm{mid}} < r_{\mathrm{mid}} \leq 2z_{\mathrm{mid}} < r_{\mathrm{upp}}, \tag{7.17}$$

and $R_{\alpha, z_{\mathrm{mid}}} > 0$ on $(z_{\mathrm{mid}}, r_{\mathrm{mid}})$ and $R_{\alpha, z_{\mathrm{mid}}} < 0$ on $(r_{\mathrm{mid}}, r_{\mathrm{upp}})$. It follows that solutions to (7.15) started with $z_{\mathrm{mid}} < r_0 \leq 2z_{\mathrm{mid}}$ satisfy $r_t \to r_{\mathrm{mid}}$ as $t \to \infty$. $\qquad \square$

7.2 The higher-level mean-field equation

In this subsection we prove Theorem 1.18 and Lemma 1.19. We start with two preparatory lemmas.

Lemma 7.3 (Convex order and second moments). *Let S be a Polish space S and let $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{P}(S))$ satisfy $\rho_1 \leq_{\mathrm{cv}} \rho_2$ and $\rho_1^{(2)} = \rho_2^{(2)}$. Then $\rho_1 = \rho_2$.*

Proof. This follows from [MSS18, Lemma 14]. $\qquad \square$

In the next lemma, we use the notation $\overline{\mu} := \mathbb{P}[\delta_X \in \cdot\,]$ defined in Subsection 1.7.

Lemma 7.4 (Maximal measure in convex order). *Let S be a Polish space and let $\mu \in \mathcal{P}(S)$. Then a measure $\rho \in \mathcal{P}(\mathcal{P}(S))$ satisfies $\rho = \overline{\mu}$ if and only if $\rho^{(2)} = \overline{\mu}^{(2)}$.*

Proof. The condition $\rho^{(2)} = \overline{\mu}^{(2)}$ implies that the first moment measure of ρ is μ. By (1.74), it follows that $\rho \leq_{\mathrm{cv}} \overline{\mu}$, so the statement follows from Lemma 7.3. $\qquad \square$

Proof of Theorem 1.18. It follows from their definition that the measures $\overline{\nu}_{\mathrm{low}}, \overline{\nu}_{\mathrm{mid}}, \overline{\nu}_{\mathrm{upp}}$ and $\underline{\nu}_{\mathrm{low}}, \underline{\nu}_{\mathrm{mid}}, \underline{\nu}_{\mathrm{upp}}$ solve the higher-level RDE and their first moment measures are $\nu_{\mathrm{low}}, \nu_{\mathrm{mid}}$ and ν_{upp}, respectively.

By [MSS18, Thm 5], one has $\underline{\nu} = \overline{\nu}$ if and only if the RTP corresponding to ν is endogenous. By Theorem 1.11, endogeny is equivalent to bivariate uniqueness, so we obtain from Proposition 1.13 that $\underline{\nu}_{\mathrm{low}} = \overline{\nu}_{\mathrm{low}}$, $\underline{\nu}_{\mathrm{mid}} \neq \overline{\nu}_{\mathrm{mid}}$, and $\underline{\nu}_{\mathrm{upp}} = \overline{\nu}_{\mathrm{upp}}$.

Since the second moment measures of $\overline{\nu}_{\mathrm{low}}, \overline{\nu}_{\mathrm{mid}}, \overline{\nu}_{\mathrm{upp}}$ are of the form (1.62), we see that the measures $\overline{\nu}_{\mathrm{low}}^{(2)}, \overline{\nu}_{\mathrm{mid}}^{(2)}, \overline{\nu}_{\mathrm{upp}}^{(2)}$ from Proposition 1.13 are indeed the second moment measures of $\overline{\nu}_{\mathrm{low}}, \overline{\nu}_{\mathrm{mid}}, \overline{\nu}_{\mathrm{upp}}$.

By Proposition 1.14, the second moment measure of $\underline{\nu}_{\mathrm{mid}}$ solves the bivariate RDE. Since $\underline{\nu}_{\mathrm{mid}} \neq \overline{\nu}_{\mathrm{mid}}$, Lemma 7.4 tells us that the second moment measure of $\underline{\nu}_{\mathrm{mid}}$ is different from $\overline{\nu}_{\mathrm{mid}}^{(2)}$. It follows that the measure $\underline{\nu}_{\mathrm{mid}}^{(2)}$ from Proposition 1.13 is indeed the second moment measure of $\underline{\nu}_{\mathrm{mid}}$.

Let $(\rho_t)_{t \geq 0}$ be a solution to the higher-level mean-field equation. Assume that $\alpha > 4$. Then Propositions 1.13 and 1.14 tell us that $\rho_t^{(2)}$ converges to one of the fixed points $\overline{\nu}_{\text{low}}^{(2)}, \underline{\nu}_{\text{mid}}^{(2)}, \overline{\nu}_{\text{mid}}^{(2)}, \overline{\nu}_{\text{upp}}^{(2)}$, depending on whether

$$
\begin{array}{ll}
\text{(i) } \rho^{(1)}(\{1\}) < z_{\text{mid}}, & \text{(ii) } \rho^{(1)}(\{1\}) = z_{\text{mid}} \text{ and } \rho^{(2)} \neq \overline{\nu}_{\text{mid}}^{(2)}, \\
\text{(iii) } \rho^{(2)} = \overline{\nu}_{\text{mid}}^{(2)}, & \text{(iv) } \rho^{(1)}(\{1\}) > z_{\text{mid}}.
\end{array}
\tag{7.18}
$$

By Lemma 7.4, these four cases correspond exactly to the four domains of attraction in (1.82). To prove that in fact ρ_t converges to $\overline{\nu}_{\text{low}}, \underline{\nu}_{\text{mid}}, \overline{\nu}_{\text{mid}}$, or $\overline{\nu}_{\text{upp}}$, respectively, in each of these cases, by the compactness of $\mathcal{P}(\mathcal{P}(\{0,1\}))$, it suffices to prove that if $\rho_{t_n} \Rightarrow \rho_*$ along a sequence of times $t_n \to \infty$, then ρ_* is the right limit point. In the cases (i), (iii) and (iv) this is clear from Lemma 7.4.

To prove the statement also in case (ii), let $(\rho_t')_{t \geq 0}$ be the solution to the higher-level mean-field equation started in $\rho_0' = \delta_{\nu_{\text{mid}}}$. Then (1.74) and Proposition 1.16 tells us that $\rho_t' \leq_{\text{cv}} \rho_t$ for all $t \geq 0$ and $\rho_t' \Rightarrow \underline{\nu}_{\text{mid}}$. Taking the limit $t_n \to \infty$, using condition (i) of Theorem 1.15, we conclude that $\underline{\nu}_{\text{mid}} \leq_{\text{cv}} \rho_*$. Since moreover $\underline{\nu}_{\text{mid}}^{(2)} = \rho_*^{(2)}$, we can apply Lemma 7.3 to conclude that $\underline{\nu}_{\text{mid}} = \rho_*$.

This completes the proof for $\alpha > 4$. The cases $\alpha = 4$ and $\alpha < 4$ are similar, but simpler. $\qquad\square$

Proof of Lemma 1.19. We note that if $\eta_1, \eta_2, \eta_3 \in [0,1]$, then

$$
\begin{aligned}
1_{\{\widehat{\text{cob}}(\eta_1,\eta_2,\eta_3)=1\}} &= 1_{\{\eta_1+(1-\eta_1)\eta_2\eta_3=1\}} = 1_{\{\eta_1=1\}} \vee \left(1_{\{\eta_2=1\}} \wedge 1_{\{\eta_3=1\}}\right), \\
1_{\{\widehat{\text{cob}}(\eta_1,\eta_2,\eta_3)>0\}} &= 1_{\{\eta_1+(1-\eta_1)\eta_2\eta_3>0\}} = 1_{\{\eta_1>0\}} \vee \left(1_{\{\eta_2>0\}} \wedge 1_{\{\eta_3>0\}}\right).
\end{aligned}
\tag{7.19}
$$

Combining this with (1.13) and Proposition 1.14, we see that if ρ solves the higher-level RDE (1.73), then

$$
\int \rho(\mathrm{d}\eta)\eta, \quad \rho(\{1\}), \quad \text{and} \quad \rho((0,1])
\tag{7.20}
$$

must all solve the RDE (1.54). Applying this to $\underline{\nu}_{\text{mid}}$ which has $\int \underline{\nu}_{\text{mid}}(\mathrm{d}\eta)\eta = z_{\text{mid}}$, we see that

$$
\underline{\nu}_{\text{mid}}(\{1\}) \in \{0, z_{\text{mid}}, z_{\text{upp}}\} \quad \text{and} \quad \underline{\nu}_{\text{mid}}((0,1]) \in \{0, z_{\text{mid}}, z_{\text{upp}}\}.
\tag{7.21}
$$

We first observe that since $\int \underline{\nu}_{\text{mid}}(\mathrm{d}\eta)\eta = z_{\text{mid}}$, we can have $\underline{\nu}_{\text{mid}}(\{1\}) \geq z_{\text{mid}}$ only if $\underline{\nu}_{\text{mid}} = \overline{\nu}_{\text{mid}}$, which we know is not the case, so we conclude that $\underline{\nu}_{\text{mid}}(\{1\}) = 0$. If $\underline{\nu}_{\text{mid}}((0,1]) \leq z_{\text{mid}}$ then $\int \underline{\nu}_{\text{mid}}(\mathrm{d}\eta)\eta = z_{\text{mid}}$ forces $\underline{\nu}_{\text{mid}}(\{1\}) = z_{\text{mid}}$, which we know is not the case, so we conclude that $\underline{\nu}_{\text{mid}}((0,1]) = z_{\text{upp}}$ and hence $\underline{\nu}_{\text{mid}}(\{0\}) = 1 - z_{\text{upp}} = z_{\text{mid}}$, where the last equality follows from (1.37).

To calculate $\int \underline{\nu}_{\text{mid}}(\mathrm{d}\eta)\eta^2$, we use that $1 - \underline{\nu}_{\text{mid}}^{(2)}(0,0) = r_{\text{mid}}$, where r_{mid} is the second largest solution r of the equation $R_{\alpha,z_{\text{mid}}}(r) = 0$, with $R_{\alpha,z_{\text{mid}}}$ defined as in (7.12). The smallest solution of the cubic equation $R_{\alpha,z_{\text{mid}}}(r) = 0$ is $r = z_{\text{mid}}$. Dividing by $(r - z_{\text{mid}})$ yields a quadratic equation of which r_{mid} is the smallest solution. Since these are straightforward, but tedious calculations, we omit them. $\qquad\square$

7.3 Root-determining and open subtrees

In this subsection we prove Lemmas 1.23 and 1.26.

Proof of Lemma 1.23. Since $\kappa(\omega) = 3$ if $\gamma[\omega] = \text{cob}$ and $\kappa(\omega) = 0$ if $\gamma[\omega] = \text{dth}$, we see that

$$
\int \mathbf{r}(\mathrm{d}\omega)(\kappa(\omega) - 1) = 2\alpha - 1,
\tag{7.22}
$$

which is ≤ 1 if and only if $\alpha \leq \frac{1}{2}$. At the end of Subsection 1.3, we have seen that in our example the RDE (1.54) has a unique solution if and only if $\alpha < 4$. By Lemma 1.24 this is equivalent to condition (ii) of Lemma 1.22. Since $S = \{0, 1\}$, Lemma 1.22 tells is that in our example, conditions (i)–(iii) are equivalent.

We claim that a finite subtree $\mathbb{U} \subset \mathbb{S}$ satisfying (1.94) is uniquely determined and in fact $x \in \Xi_\mathbb{U}$ implies $x_\mathbf{i} = 0$ for all $\mathbf{i} \in \mathbb{U}$. To prove this, let $A = \{\mathbf{i} \in \mathbb{U} : x_\mathbf{i} = 0 \ \forall x \in \Xi_\mathbb{U}\}$. Since $\mathbb{U}$ is finite, if $\mathbb{U} \backslash A$ is not empty then we can find some $\mathbf{i} \in \mathbb{U} \backslash A$ such that $\mathbf{i}j \notin A$ for $j = 1, 2, 3$. (Here we take $\mathbb{T}$ to be the set of all words made from the alphabet $\{1, 2, 3\}$.) If $\gamma[\boldsymbol{\omega}_\mathbf{i}] = \mathtt{dth}$, then $x_\mathbf{i} = 0$ for all $x \in \Xi_\mathbb{U}$ which contradicts the fact that $\mathbf{i} \in \mathbb{U} \backslash A$. But if $\gamma[\boldsymbol{\omega}_\mathbf{i}] = \mathtt{cob}$, then (1.94) and the fact that $\mathbf{i}j \notin A$ for $j = 1, 2, 3$ again imply $x_\mathbf{i} = 0$ for all $x \in \Xi_\mathbb{U}$, so we see that $\mathbb{U} \backslash A$ must be empty. In particular, this shows that $\mathbb{U}$ is root determining.

To see that $\mathbb{U}$ is a minimal root determining subtree, assume that $\mathbb{V} \subset \mathbb{U}$ is a smaller one. Then there must be be some $\mathbf{i} \in \mathbb{V}$ such that $\gamma[\boldsymbol{\omega}_\mathbf{i}] = \mathtt{cob}$ and either $\mathbf{i}1 \notin \mathbb{V}$ or $\mathbb{V} \cap \{\mathbf{i}2, \mathbf{i}3\} = \emptyset$. (Here we use that by definition, minimal root determining subtrees contain the root, so $\mathbb{V}$ is not empty.) But then either $\mathbf{i}1 \in \nabla\mathbb{V}$ or $\{\mathbf{i}2, \mathbf{i}3\} \subset \nabla\mathbb{V}$. Define $x \in \Xi_\mathbb{V}$ inductively by (1.45) with $x_\mathbf{j} = 1$ for all $\mathbf{j} \in \nabla\mathbb{V}$. Then $x_\mathbf{i} = 1$. Either $\mathbf{i}$ is the root or its predecessor $\overleftarrow{\mathbf{i}}$ satisfies $x_{\overleftarrow{\mathbf{i}}} = 1$ by (1.94), so by induction we see that $x_\mathbf{i} = 1$. Since the all-zero configuration is also an element of $\Xi_\mathbb{V}$, this proves that $\mathbb{V}$ is not root determining and hence $\mathbb{U}$ is minimal. $\qquad\square$

Proof of Lemma 1.26. We observe that

$$
\begin{aligned}
Y_\mathtt{cob} &= \big\{y \in \{0, 1\}^3 : y \text{ is a minimal element of } \mathtt{cob}^{-1}(\{1\})\big\} \\
&= \big\{y \in \{0, 1\}^3 : \mathtt{cob}(y) = 1 \text{ and } \mathtt{cob}(y') = 0 \ \forall y' \leq y, \ y' \neq y\big\} \qquad (7.23) \\
&= \big\{(1, 0, 0), (0, 1, 1)\big\}.
\end{aligned}
$$

Since the set $\mathtt{dth}^{-1}(\{1\}) = \{y \in \{0, 1\}^0 : \mathtt{dth}(y) = 1\}$ is empty, it has no minimal elements, and hence $Y_\mathtt{dth} = \emptyset$. On the other hand, $\mathtt{bth}(y) = 1$ for all $y \in \{0, 1\}^0$. In fact, $\{0, 1\}^0 = \{\varnothing\}$ is a set with only one element, the empty word, so $\mathtt{bth}^{-1}(\{1\}) = \{\varnothing\}$ and hence the set of its minimal elements is $Y_\mathtt{bth} = \{\varnothing\}$. Now (1.97) with the convention that $1_{A_\mathbf{i}} := \varnothing$ if $\kappa(\boldsymbol{\omega}_\mathbf{i}) = 0$ says that $\mathbb{O}$ is an open subtree of $\mathbb{U}$ if and only if:

(i) $\{j \in \{1, 2, 3\} : \mathbf{i}j \in \mathbb{O}\} = \{1\}$ or $= \{2, 3\}$ for each $\mathbf{i} \in \mathbb{O} \cap \mathbb{U}$ such that $\gamma[\boldsymbol{\omega}_\mathbf{i}] = \mathtt{cob}$,

(ii) $\varnothing \in \emptyset$ for each $\mathbf{i} \in \mathbb{O} \cap \mathbb{U}$ such that $\gamma[\boldsymbol{\omega}_\mathbf{i}] = \mathtt{dth}$,

(iii) $\varnothing \in \{\varnothing\}$ for each $\mathbf{i} \in \mathbb{O} \cap \mathbb{U}$ such that $\gamma[\boldsymbol{\omega}_\mathbf{i}] = \mathtt{bth}$,

which corresponds to the condition in (1.102). $\qquad\square$

References

[AB05] Aldous, D.J. and Bandyopadhyay, A.: A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* **15(2)** (2005), 1047–1110. MR-2134098

[ADF18] Andreis, L., Dai Pra, P. and Fischer, M.: McKean-Vlasov limit for interacting systems with simultaneous jumps. *Stoch. Anal. Appl.* **36(6)** (2018), 1–36. MR-3925147

[Ald00] Aldous, D.J.: The percolation process on a tree where infinite clusters are frozen. *Math. Proc. Cambridge Philos. Soc.* **128** (2000), 465–477. MR-1744108

[Als12] Alsmeyer, G.: Random recursive equations and their distributional fixed points. *Unpublished manuscript* (2012), available from https://www.uni-muenster.de/Stochastik/lehre/WS1112/StochRekGleichungenII/book.pdf.

[BCH18] Baake, E., Cordero, F. and Hummel, S.: Lines of descent in the deterministic mutation-selection model with pairwise interaction, arXiv:1812.00872v2.

[BDF20] Broutin, N., Devroye, L. and Fraiman, N.: Recursive functions on conditional Galton-Watson trees, Random Struct. Alg. (2020), 1–13.

[Bou58] Bourbaki, N.: Éléments de Mathématique. VIII. Part. 1: Les Structures Fondamentales de l'Analyse. Livre III: Topologie Générale. Chap. 9: Utilisation des Nombres Réels en Topologie Générale. 2iéme éd. *Actualités Scientifiques et Industrielles* **1045** Hermann & Cie, Paris, 1958. MR-0173226

[BW97] Baake, E. and Wiehe, T.: Bifurcations in haploid and diploid sequence space models. *J. Math. Biol.* **35** (1997) 321–343. MR-1478053

[Cho69] Choquet, G.: Lectures on Analysis. Volume I. Integration and Topological Vector Spaces. *Benjamin*, London, 1969. MR-0250011

[DFL86] De Masi, A., Ferrari, P.A. and Lebowitz, J.L.: Reaction-diffusion equations for interacting particle systems. *J. Stat. Phys.* **44** (1986), 589–644. MR-0857069

[DN94] Durrett, R. and Neuhauser, C.: Particle systems and reaction-diffusion equations. *Ann. Probab.* **22(1)** (1994), 289–333. MR-1258879

[DN97] Durrett, R. and Neuhauser, C.: Coexistence results for some competition models. *Ann. Appl. Probab.* **7(1)** (1997), 10–45. MR-1428748

[EK86] Ethier, S.N. and Kurtz, T.G.: Markov Processes; Characterization and Convergence. *John Wiley & Sons*, New York, 1986. MR-0838085

[FL17] Foxall, E. and Lanchier, N.: Survival and extinction results for a patch model with sexual reproduction. *J. Math. Biol.* **74(6)** (2017), 1299–1349. MR-3634786

[FM01] Fill, J.A. and Machida, M.: Stochastic monotonicity and realizable monotonicity *Ann. Probab.* **29(2)** (2001), 938–978. MR-1849183

[HLL18] Hiai, F., Lawson, J. and Lim, Y.: The stochastic order of probability measures on ordered metric spaces *J. Math. Anal. Appl.* **464(1)** (2018), 707–724. MR-3794112

[JPS20] Johnson, T., Podder, M. and Skerman, F.: Random tree recursions: which fixed points correspond to tangible sets of trees? *Random Struct. Alg.* **56(3)** (2020), 796–837.

[KKO77] Kamae, T., Krengel, U. and O'Brien, G.L.: Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5** (1977), 899–912. MR-0494447

[Lig85] Liggett, T.M.: Interacting Particle Systems. *Springer*, New York, 1985. MR-0776231

[Lin92] Lindvall, T.: Lectures on the Coupling Method. *John Wiley and Sons*, New York, 1992. MR-1180522

[Mac17] Mach, T.: Dualities and genealogies in stochastic population models. *PhD thesis, University of Göttingen*, 2017.

[MS19] Martin, J.B. and Stasiński, R.: Minimax functions on Galton-Watson trees, *Combinatorics, Probability and Computing* (2019), 1–30.

[MSS18] Mach, T., Sturm, A. and Swart, J.M.: A new characterization of endogeny. *Math. Phys. Anal. Geom.* **21(4)** (2018), paper no. 30, 19 pages. MR-3861389

[McK66] McKean, H.P.: A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **56(6)** (1966), 1907–1911. MR-0221595

[Neu94] Neuhauser, C.: A long range sexual reproduction process. *Stochastic Processes Appl.* **53** (1994), 193–220. MR-1302910

[Nob92] Noble, C.: Equilibrium behavior of the sexual reproduction process with rapid diffusion. *Ann. Probab.* **20(2)** (1992), 724–745. MR-1159570

[NP99] Neuhauser, C. and Pacala, S.W.: An explicitly spatial version of the Lotka-Volterra model with interspecific competition. *Ann. Appl. Probab.* **9(4)** (1999), 1226–1259. MR-1728561

[Par05] Parthasarathy, K.R.: Probability measures on metric spaces. *AMS Chelsea Publishing*, Providence, RI. (2005). MR-2169627

[RST19] Ráth, B., Swart, J.M., and Terpai, T.: Frozen percolation on the binary tree is nonendogenous, arXiv:1910.09213.

[SS15] Sturm, A. and Swart, J.M.: A particle system with cooperative branching and coalescence. *Ann. Appl. Probab.* **25(3)** (2015), 1616–1649. MR-3325283

[SS18] Sturm, A. and Swart, J.M.: Pathwise duals of monotone and additive Markov processes. *J. Theor. Probab.* **31(2)** (2018), 932–983. MR-3803920

[ST85] Shiga, T. and Tanaka, H.: Central limit theorem for a system of Markovian particles with mean field interactions. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **69** (1985), 439–459. MR-0787607

[Str65] Strassen, V.: The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** (1965), 423–439. MR-0177430

[Swa17] Swart, J.M.: A Course in Interacting Particle Systems, arXiv:1703.10007.

[War06] Warren, J.: Dynamics and endogeny for recursive processes on trees. *Stochastics* **78(5)** (2006), 327–341. MR-2270941

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm