Branching Processes – A General Concept

Andreas Greven, Thomas Rippl and Patrick Glöede

Friedrich-Alexander University Nuremberg-Erlangen,
Lehrstuhl fuer Math. Stochastik,
Cauerstr. 11,
91058 Erlangen, Germany.
E-mail address: greven@math.fau.de
URL: http://www.fau.de

University of Goettingen,
Institut fuer Math. Stochastik,
Goldschmidtstr. 7,
37077 Goettingen, Germany.
E-mail address: trippl@uni-goettingen.de
URL: http://www.uni-goettingen.de

The Faculty of Industrial Engineering and Management,
Technion-Israel Institute of Technology,
Technion City,
Haifa 3200003, Israel.
E-mail address: gloede@technion.ac.il
URL: http://www.technion.ac.il

Abstract. The paper has four goals. First, we want to generalize the classical concept of the branching property so that it becomes true for historical and genealogical processes, where the classical concept fails (here we use the description of genealogies by (V-marked) ultrametric measure spaces leading to state spaces $\mathbb{U}$ resp. $\mathbb{U}^V$). In particular we want to complement the corresponding concept of infinite divisibility developed in Glöde et al. (2019) for this context. The processes we consider are always defined by well-posed martingale problems. The point of the generalized branching property is that the state at times $t+s$ can at any time t be decomposed in a measurable function of the state at time t and an independent part which itself then decomposes in independent copies of the process evolving for time s. Secondly we want to find a corresponding characterization of the generators of branching processes both easy to apply and general enough to cover a wide range of mechanisms and state spaces, this is our first main result.

As a third goal we want to obtain the branching property for some important examples as the $\mathbb{U}$-valued Feller diffusion respectively $\mathbb{U}^V$-valued super random walk and the historical process on countable geographic spaces, the latter as two examples of a whole zoo of spatial processes we could treat. The fourth goal is to show the robustness of the method and to get the generalized branching property for genealogies marked with ancestral path, giving the line of descent moving through the ancestors and space, leading to path-marked ultrametric measure spaces. These new processes are constructed here and proved to have the generalized branching property, both together our second major result.

We develop an abstract framework covering above situations and questions, leading to a new generator criterion. The state spaces suitable for historical and genealogical processes are consistent collections of topological semigroups each enriched with a compatible collection of maps, the truncation maps. All objects are defined on the state space of the process. The method allows to treat every type of population model formulated as solution to a well-posed martingale problem. This framework in particular includes processes taking values in the space of marked ultrametric measure spaces and hence allows to treat historical information and genealogies of spatial population models both at once, if genealogies are described this way. Another example is a multitype population, more specific with genetic types under mutation.

Contents

1. Introduction

The problem and goals. Branching processes are one of the most important and best-studied class of stochastic processes covered in many books, starting with Harris (2002) and Athreya and Ney (2004) and later including *spatial* models (Dawson, 1993 and Etheridge, 2000). However new questions and aspects keep arising, concerning in particular the evolution of genealogies and histories in such branching models, as we shall see below in detail. The key point of this work is to define and get criteria for the *"generalized branching property"* in the face of new types of processes describing *genealogies or histories in spatial population models*, where the classical branching property does *not* hold. Some of these processes are introduced here, others are well known in the literature. The reader not familiar with the terms genealogy or history is invited to read Remark 1.1.

The focus in the theory of branching processes has now shifted to *historical* and *genealogical information* on the population as objects of interest in its own right. It is also a tool to understand the behaviour of population sizes or type frequencies in spatial models, see for example Neveu (1986), Aldous (1991a), Neveu and Pitman (1989), Le Gall (1989), Le Gall (1993), Le Gall (1999), Duquesne and Le Gall (2002), or in another perspective Bertoin and Le Gall (2000, 2003, 2005) to name a few examples for the use of genealogies modeled as *labeled trees*. Here also *multitype* models fit in as marked labeled trees, where individuals have genetic types which undergo mutation.

We focus here on the description of genealogical information more in the spirit of the description via historical processes (Dawson and Perkins, 1991) or $\mathbb{R}$-trees (Evans et al., 2006), by using here equivalence classes of *(marked) ultrametric measure spaces*. The latter approach is developed in Greven et al. (2009), Depperschmidt et al. (2011), applied to Fleming-Viot type models in Depperschmidt et al. (2012) and Greven et al. (2013) and it has been extended to branching in Glöde (2012); Depperschmidt and Greven (2019); Glöde et al. (2019). This particular description of genealogies is chosen in this paper to be able to work with the framework of *well-posed martingale problems* to characterize the models and second to be able to use the concept of infinite divisibility from Glöde et al. (2019).

For this approach using (equivalence classes of) ultrametric measure spaces there is a self-contained and detailed survey, see Depperschmidt and Greven (2020), where in particular one can find examples and references showing more generally the *usefulness of this abstract approach* to *evolving* (in time) *genealogies* of populations, which is based on the construction of the genealogy valued processes via *well-posed martingale problems*. The latter allows to use the tools which have been developed in Ethier and Kurtz (1986), Dawson (1993) in particular for processes relevant in the theory of spatial population models for example see Dawson and Greven (2014), but where genealogical questions arise, which can in the framework of $\mathbb{U}$-valued processes now be tackled.

If we record *genealogies* and *histories* for the time-t-population there are two forms in which information about the past may sit in the *present state*, for once in the ancestral relationships of individuals (genealogies) or in the marks as locations or types (histories). Examples are paths in geographic space or paths in type space, the latter storing information on the genome. Both effects may require some care in order to still distill a generalized branching property, since the classical branching property does not hold in that situation. Our examples we treat show the typical obstacles to the classical form of independence in the further evolution of subpopulations with states storing information about the past.

The main goals of this paper are:

- first to develop for a *generalized* branching property a suitable *abstract framework* and formulation,
- second to find and prove an *operator criterion* for a process specified by a well-posed martingale problem to check this property,
- thirdly to apply it to *historical processes* and *evolving genealogies* the latter described in a specific way namely as *equivalence classes of ultrametric measure spaces*, respectively their *marked* versions for *spatial* and for *multitype* population models and
- fourth but not least, construct *ancestral path marked genealogies*, generalizing and combining both *historical* processes as well as *genealogical* processes and which we introduce here as new processes and where we use the criterion to verify the generalized branching property.

Remark 1.1 (The word genealogy). As the word genealogy or history gives occasionally rise to misunderstanding we clarify here some points concerning the precise meaning of these words *genealogy - history*.

To clarify terminology consider a binary splitting Galton-Watson process in continuous time. We can draw the *labeled tree* of all individuals ever alive up to time t (labeled means individuals are identified and distinguished by a name). Then we observe how this object evolves further in time t. The vertices of the tree are the individuals at their birth time and edges representing the life time are attached, ending or giving rise to two descendants splitting the edge at a new vertex. This allows to define ancestors and descendants and we define the distance between two points in this tree as the sum of the two cumulative edge lengths back to the most recent common ancestor. This induces a tree-like labeled metric space representing the labeled genealogy. If we pass to the equivalence class under isometries we obtain an equivalence class of tree-like metric spaces (we forget the names of individuals).

Embedded in the labeled tree is the set of leaves at height t forming a *labeled ultrametric space*, describing the *labeled genealogy* of the *individuals alive at time*

t referred to as *current population*. The ultrametric space is always in one-to-one correspondence with a *unique minimal $\mathbb{R}$-tree* of which it is the set of *leaves* (compare Example 2.2). Again we might pass to the equivalence class under isometries and forget the names of individuals getting what we refer to as genealogy of the current population. The genealogy including the *fossils* is then the whole genealogical tree including everybody alive at some times before time t. The individuals of our population might have types or locations in some space, where they move according to some stochastic process. Then we get *marked* genealogies.

In such a genealogy we can recover historical information via the ancestral path, obtained by tracing backward type or location of an individual alive at time t till its birth time, continue tracing the type/location of the father etc. This way we get a collection of backward path in type or geographic space, the *ancestral path*. Give each path weight 1 and obtain a measure on paths, this is the state of the *historical process* a measure-valued process on the space of paths.

We are interested in the time *evolution* of *genealogies/histories* as the current time t evolves in particular in the genealogy of the population currently alive represented by the equivalence class of ultrametric spaces.

Since we are interested in *large populations* even the infinite population limit, such a genealogy can only be observed by taking *samples* from the current population. Therefore we add a probability measure on the individuals alive at time t to the genealogy. This results in a labeled genealogy, a triple (X, r, μ) with X set of individuals, r the genealogical distance (an ultrametric) and μ the sampling measure. The equivalence class under isometries which are measure preserving denoted $[X, r, \mu]$ represents for us the relevant information on the genealogy of the population alive at some time t. The genealogy of all individuals alive ever we call the *fossil process*. This is however not the object of this paper. To incorporate population sizes fluctuating we consider also μ which are finite measures. We return to this point.

Branching property. Important examples for the typical structure of branching processes are *continuous state branching processes* (Le Gall, 1999, Lambert, 2007) and *measure valued branching processes* (Dawson, 1993, Le Jan, 1991) or the *historical Dawson-Watanabe* process (Dawson and Perkins, 1991). The state spaces then are *linear* spaces, here $\mathbb{R}^d$ and $\mathcal{M}_f(E)$, respectively. Here E is a Polish space and $\mathcal{M}_f(E)$ is the space of finite measures on $(E, \mathcal{B}(E))$. All branching processes share the following characteristic *branching property*: given the state of the present population, the states of sub-populations descending from different "ancestors" in today's population evolve independently from each other and with identical laws. This has been formalized so far as follows.

The underlying state space S is a *semigroup* meaning it is endowed with the binary operation of addition (of reals or measures, respectively). Suppose that $X = (X_t)_{t \geq 0}$ is a Markov process with values in S.

The transition probabilities denoted by P_t, that is, $(P_t f)(x) = P_t(x, \cdot)[f] = \mathbb{E}[f(X_t)|X_0 = x]$ for $f \in b\mathcal{B}(S)$, the space of bounded and measurable functions on S. The semigroup $(P_t)_{t \geq 0}$ has the *branching property* if

$$P_t(x + y, \cdot)[f] = (P_t(x, \cdot) * P_t(y, \cdot))[f], \text{ for all } x, y \in S \text{ and } f \in b\mathcal{B}(S), \quad (1.1)$$

with $*$ the convolution. Then X is defined to have the *branching property* if $(P_t)_{t \geq 0}$ has the branching property.

If $b\mathcal{B}(S)$ contains a subset D of functions which are *multiplicative* w.r.t. the semigroup operation $+$ and D *separates* laws on S, then (1.1) holds if $P_t(x + y, \cdot)[f] = P_t(x, \cdot)[f] \, P_t(y, \cdot)[f]$ for all $x, y \in S$ and $f \in D$. Indeed, in some of the *classical cases* the branching property is related to the form of the *Laplace transform*. In the second example above we have $X_t \in S = \mathcal{M}_f(E)$. With the help of a *time evolution operator* $V_{s,t} : \mathrm{bp}\mathcal{B}(E) \to \mathrm{bp}\mathcal{B}(E)$, $0 \le s \le t$, $\mathrm{bp}\mathcal{B}(E)$ the non-negative, bounded, measurable functions on E, the Laplace transform can be represented as (set $\langle \mu, f \rangle = \int f \, d\mu$ for every $\mu \in \mathcal{M}_f(E)$ and $f \in b\mathcal{B}(E)$):

$$\mathbb{E}[\exp(-\langle X_t, \phi \rangle)|X_s] = \exp(-\langle X_s, V_{s,t}\phi \rangle), \quad \text{a.s.} \tag{1.2}$$

Then (1.1) can be read of from this rhs. Clearly, for $E = \{1, 2, \ldots, d\}$ this includes $\mathbb{R}^d$-valued processes.

Then first the question arises whether the processes describing features of the *history* or *genealogies* of individuals of a "branching" population have an analogue property and structure. The problem being that the *present state* contains information about the state at *past times*. In other words, can we define an abstract *generalized branching property* dealing with this problem. This complements the investigation of a concept of *infinite divisibility* for *genealogical* structures modeled as ultrametric measure spaces which is introduced in Glöde et al. (2019) which presents a generalized infinite divisibility since the classical one does not hold.

However even if we can define a branching property for the objects, in general it may be difficult to obtain analytical expressions for $P_t(x, \cdot)[f]$ to *check* in practice whether (1.1) holds. In particular often an analogue of (1.2) cannot be found. Instead of working with the Markov process' transition probabilities $(P_t)_{t \ge 0}$ or the martingale problem itself we may use its *operator* A. Here we hope to derive a statement in the following spirit: *Suppose the generator of a Markov process (or operator of a martingale problem) has a particular form, then the Markov process is a branching process* and similarly for the generalized branching property. Here one has the Kurtz criterion in Ethier and Kurtz (1986). The point however is that this particular form of the generator criterion needs to be such that it is *easily* verifiable. This is important for *general* state spaces and more complex evolution mechanisms for example for evolving genealogies. This means in the new cases we want to cover, we have to go *beyond* the *Kurtz criterion* in Ethier and Kurtz (1986) by using an *additional structure*. To find such a structure and criterion is our second goal.

Therefore we first *extend the concept* of branching processes to more general state spaces and evolution mechanisms with certain *algebraic* and fitting *topological* properties. Second we find the *characterization of operators* for processes which have the generalized branching property. In particular we will present simple criteria which the generator has to satisfy in order for the processes to have the (generalized) branching property. One of the criteria is such that it can be checked by explicit calculation for complicated processes *if we have checked that our setup applies*, which needs of course some work, since it requires that our process is given by a well-posed martingale problem.

We briefly explain below first using classical examples how the criterion works in a simpler context where also the classical branching property holds and this is verified by this criterion here. Then we come to our third goal and check the generalized branching property for the genealogy of *Fellers diffusion model* or the *super random walk*.

We finally come to our *fourth* goal where we use the criterion to get the generalized branching property in a spatial model where every individual alive today is equipped with its *ancestral path*. This is the line of descent moving geographic or type through space till the own birth time, that of the father etc. This process we define and characterize here via a *well-posed martingale problem*. We will argue why this new class of processes with values in marked ultrametric measure spaces, is in fact a *very general class* of population processes in this context of describing genealogies and histories. This process has some relation to the "snake" of LeGall Le Gall (1999) formulated via labeled trees. We construct our object in Section 3.2 and 3.3 combining the approach based on historical processes from Dawson and Perkins (1991) with the approach based on the ultrametric measure space valued description of genealogies Greven et al. (2013), Depperschmidt et al. (2013) and Greven et al. (2016). The new processes constitute the class of *(ancestral path)-marked $\mathbb{U}^V$-valued Feller resp. super random walk processes* and if we model the population *ever* alive we obtain processes also *including fossils*.

The term very general is here meant in the sense that a multitude of processes typically studied in population models can be embedded into this class. *Therefore we prove here the criterion works in this case* which is then one of the deeper reasons, that we have the branching property for all the embedded Markovian processes often studied in their own right. As there are functionals typically not one-to-one the proof of the branching property for them requires work (for example getting the Markov property). We come to this at the end of Section 3.

We cannot cover here everything of interest however. It would be nice to cover the case of the genealogy of the Dawson-Watanabe process. Here however at least in the $d \geq 2$ case the *martingale problem formulation* of the process has a different form on a continuum geographic space and preparing this needs some effort, which would take too much space in this paper and will be treated in forthcoming work. However we can define the $\mathbb{U}^{\mathbb{R}^d}$-valued process as a functional of the historical Dawson-Watanabe process to which we can apply our criterion so that we obtain the branching property of that functional as well. Similar issues arise also for continuum state branching processes not of the diffusive type.

A corresponding interesting *open problem* is to find other examples of processes fitting the abstract algebraic and topological framework, but are *not* arising from genealogies or histories.

The classical framework and the criterion of Kurtz. In classical situations the state space S has the following features, where we will always use S as a generic notation for a *Polish state space*. We will for now assume that S *is a topological semigroup with operation* $+$.

Let us recall the following observations from Ethier and Kurtz (1986)[Section 4.10]. Assume the following properties, first

$$D \subset D' = \{f \in \mathrm{b}\mathcal{B}(S) : f(w+z) = f(w)f(z), \text{ for all } w, z \in S\} \qquad (1.3)$$

is a set of multiplicative functions and $A : D \to \mathrm{b}\mathcal{B}(S)$ is a linear operator and X^x solves the martingale problem for (A, D, δ_x) where $x \in S$. Furthermore assume that X^x has the (classical) branching property meaning for every x (1.1) holds. Finally let X^y be a solution of the (A, D, δ_y) martingale problem, $y \in S$, which is independent of X^x.

Then for any $f \in D$, the process

$$f(X_t^y + X_t^y) - \int_0^t Af(X_s^x)f(X_s^y) + f(X_s^x)Af(X_s^y)\,\mathrm{d}s, \, t \geq 0 \qquad (1.4)$$

defines a *martingale*. This in turn implies for the *operator* A that

$$Af(x + y) = f(y)Af(x) + f(x)Af(y). \qquad (1.5)$$

Conversely, by uniqueness, if (1.5) holds this implies that X^{x+y} has the same distribution as $X^x + X^y$ where X^x and X^y are independent. This is the *branching property*.

Thus, one can indeed prove that whenever the martingale problem for (A, D, δ_x) is well-posed for all x and $D \subseteq D'$ then the solution process has the *branching property iff* (1.5) *holds*.

Obviously, for complicated operators A it will not be obvious whether or not (1.5) is satisfied. Hence we would like to have a criterion for a generator which is easy to check and which guarantees (1.5). That here the *linearity* of the rate may play an important role was discovered and used used for $\mathbb{R}^+$-valued processes in Caballero et al. (2009). This approach can be formulated more abstractly as follows.

Towards a new operator criterion. It is easy to check that (1.5) is satisfied if for any $f \in D \subseteq D'$ there exists a *semigroup homomorphism* $g_f : S \to (\mathbb{R}, +)$ (typically *depending on* f) such that for all $x \in S$,

$$Af(x) = g_f(x)f(x). \qquad (1.6)$$

This works very well for the classical branching processes with values in *linear spaces* (where we get the classical branching porperty), it is however not working in the case of genealogical or historical processes as we shall explain latter on. Therefore (1.6) will be the *starting point* for our criterion we develop below, which needs some *new* elements if the state space is *not* anymore a *linear space* as in the classical situation and also the mechanism of the process is of a different nature so that we get only the *generalized* branching property.

The new criterion in some classical cases (Examples). Some well-known examples for branching processes show that indeed (1.6) is a good criterion for the classical branching property as well and that g_f can be specified *explicitly*.

Example 1.2. First consider the simplest example, the classical continuous time *critical binary Galton-Watson* process with branching rate $b > 0$, meaning we have $exp(b)$-distributed splitting or death times. That is, $E = \mathbb{Z}_+$, $D = \{\mathbb{N}_0 \to \mathbb{R}, x \mapsto e^{-\lambda x} : \lambda > 0\}$, $Af(x) = \frac{1}{2}bx(f(x+1) + f(x-1) - 2f(x))$. The martingale problem for (A, D) is well-posed, see Section 8.3 in Ethier and Kurtz (1986) and for $f \in D$:

$$Af(x) = \frac{1}{2}bx(f(1) + f(-1) - 2f(0))f(x), \, x \in \mathbb{N}. \qquad (1.7)$$

So, we can choose the homomorphism as:

$$g_\lambda(x) = \frac{1}{2}bx(f(1) + f(-1) - 2f(0)) = \frac{1}{2}b(e^{-\lambda} + e^\lambda - 2)x, x \in \mathbb{N} \qquad (1.8)$$

and (1.6) is satisfied.

Example 1.3. Now, consider the class of *measure-valued continuous state branching processes* (CSBP). (We *specialize* later to the case where the description via a well-posed martingale problem has already been established.) They are spatial

analogues of the real-valued CSBP, see Le Gall (1999). More precisely, let E be a locally compact metric space and let $S = \mathcal{M}_f(E)$, the space of finite measures on E. We denote the process by $X = (X_t)_{t \geq 0}$. Intuitively, X locally behaves like a CSBP plus there is migration/mutation in E governed by a Feller process with generator $(B, \mathcal{D}(B))$, say. The branching dynamics is locally determined by the branching mechanism, where for each $x \in E$, $\alpha(x) \in \mathbb{R}$, $\beta(x) > 0$ for all $x \in E$

$$\psi(x, u) = \alpha(x)u + \beta(x)u^2 + \int_{(0, \infty)} (e^{-ru} - 1 + ru)\, \pi(x, \mathrm{d}r) \tag{1.9}$$

and the measure $\pi(x, \mathrm{d}y)$ satisfies $\sup_{x \in E}(\int_0^1 r^2 \pi(x, \mathrm{d}r) + \int_1^\infty r \pi(x, \mathrm{d}r)) < \infty$. The generator of X then takes the following form, see Dawson (1993, page 106):

with $D := \{F(\mu) = F_\phi(\mu) = f(\langle \mu, \phi \rangle) : \phi \in \mathcal{D}(B)\}$, choose $f(x) = e^{-x}$ (1.10)

and set for arbitrary $\phi \in \mathcal{D}(B)$

$$\Omega F_\phi(\mu) = f'(\langle \mu, \phi \rangle)\langle \mu, B\phi \rangle + f'(\langle \mu, \phi \rangle)\langle \mu, \alpha\phi \rangle + \frac{1}{2}f''(\langle \mu, \phi \rangle)\langle \mu, \beta\phi^2 \rangle \tag{1.11}$$

$$+ \int_E \mu(\mathrm{d}x) \int_0^\infty n(x, \mathrm{d}u)[f(\langle \mu, \phi \rangle) + u\phi(x))$$
$$- f(\langle \mu, \phi \rangle) - f'(\langle \mu, \phi \rangle)u\phi(x)] .$$

The martingale problem for (Ω, D, δ_μ) is well-posed for all $\mu \in \mathcal{M}_f(E)$. One can easily check that

$$\Omega F_\phi(\mu) = g_\phi(\mu) F_\phi(\mu), \tag{1.12}$$

with the linear function

$$g_\phi(\mu) \quad = -\langle \mu, B\phi \rangle - \langle \mu, \alpha\phi \rangle + \langle \mu, \frac{1}{2}\beta\phi^2 \rangle + \int_E \mu(\mathrm{d}x) \int_0^\infty n(x, \mathrm{d}u)[f(u\phi(x)) - 1$$
$$+ u(\phi(x))]$$
$$\tag{1.13}$$

Hence criterion (1.6) is satisfied. Note that if E is a point this covers the well-known $\mathbb{R}$-valued CSBP.

Example 1.4 (Shortlist of classical examples). We could add here *spatial* models like *super random walk* or *multitype branching with mutation* and many more as long as we have a characterization of the process as *well-posed martingale problem* of the form given by an operator A. The latter is a restriction, as we will see in the context of the Dawson-Watanabe process which we can only treat with our method so far in $d = 1$. Given such a characterization the criterion is easily checked again, by explicit calculation left to the reader.

These examples show that the criterion given in (1.6) is indeed *easy to verify*. For the classical processes mentioned above the branching property is typically proven by verifying (1.2), see for example Dawson (1993, Chapter 4) for the measure-valued setting and Le Gall (1999, Section II.1) for the real-valued setting.

However, the Laplace approach turns out to be difficult to apply to *more general classes* of processes like $\mathbb{U}$-*valued* branching processes while the *generator approach introduced above still works* to at least give a generalized branching property. One reason is that the classical approach of showing the branching property via Laplace transforms (evolution equations) depends on the underlying state space being *linear*, or at least a convex cone, which is not true for $\mathbb{U}$-valued processes. It turns out

however that the generator approach can be extended and the branching property *generalized* to cover *genealogy-valued branching processes* (in ultrametric measure space description) and its spatial versions as well as *historical processes*, where the classical property does not hold.

A generalized branching property. The classical concept of branching processes can be modified and then extended to much more *general mechanism* and with it state spaces, where the branching property in the classical sense *is false* as for example Markov processes of evolving *histories* and *genealogies* in particular also in the description with marked ultrametric measure spaces. The idea of the *generalized branching property* in that context is as follows.

We have to incorporate the possibility that sub-populations originating from the population at time s still at a later time t *share some common information about the population* and hence, *conditional on the information up to time s fail to be independent* and hence the classical branching property is false.

For *example*, consider a Galton-Watson branching dynamics on $\mathbb{N}_0$. Conditional on the information up to time s, the sizes of the subfamilies at a later time t originating from the individuals alive at time s are independent. However, if we incorporate the genealogical relationships into the state space *independence fails*. The reason is that unless the population at time s consisted of only one individual all individuals at time t are connected by ancestral lines going back beyond time s. Another example for the same problem occurs for instance if in addition to branching the individuals migrate independently from each other in geographic space or mutate in type space (genome). Then if we include information about their path in geographic or type space up to the present time s, then independence *fails*.

However we would like to consider both these processes as *generalized* branching processes. Since we have no branching property as a process with values in S in the classical sense we have to introduce some additional structure on S and to define a (new) *general branching property*.

The branching property is intimitely connected with the concept of *infinite divisibility*. In Glöde et al. (2019) we developed a new concept of infinite divisibility for genealogy valued processes, which used some abstract algebraic and fitting topological structures which allowed to prove basic facts on infinitely divisible random variables with states in genealogies. This allowed to prove the (generalized form of) Lévy -Khintchine representation of genealogy valued random variables. This is based on algebraic arguments which had been developed by Evans and Molchanov (2017). We also formulated a branching property for stochastic processes on $\mathbb{U}$ and $\mathbb{U}^V$. We therefore make here again use of abstract concepts in this spirit and use the tools developed in that paper.

However we move on here to a *more abstract approach* of the branching property which allows us to handle more examples as the historical process and path-marked genealogical processes as we shall see later on. This allows us then in particular also to obtain new examples for processes whose marginal distributions have Lévy -Khintchine representations and inhomogeneous Poison point process representations.

This we next first motivate and then introduce formally in Section 2.

Idea of formalization for a general branching property and a generator criterion: Intuitively, the solution of the problem from above is as follows. Here

since we have to distinguish past and the future the time parameter t will appear, the past is taken relative to t. We have to replace the $+$-operation we have in the measure-valued process by an abstract algebraic operation $\sqcup^t$, in fact depending on a time parameter t and which forms a semigroup for each $t \geq 0$. Furthermore we need maps T_t, which extracts the part of the state at time t, which contains information on the past and the part which arises newly in the future. For that we then need the set S_t of these pieces of the state and the map T_t and the "$\sqcup^t$"-operation has to be compatible.

The state space S contains subsets S_t for $t \geq 0$. Elements in S_u contain genealogical information (or historical one) which has been generated at times at most u back from the current time t. Formally we require $S_u \subseteq S_s$ for $u < s$.

Furthermore we need three assumptions. Consider times $u < s < t$. *First*, we assume that there are *"truncation maps"* T_s on S which allow us to remove at any time t the information about the states of the process before any time $s < t$. This is mathematically a function $T_{t-s} : S \to S_{t-s}$. *Second*, we assume that on each S_t there is a binary operation $\sqcup^t : S_t \times S_t \to S_t$ we call *concatenation*. This operations generate for each $t > 0$ a *topological semigroup structure*. This is the analogue of the addition in the classical settings with S a linear topological space. *Thirdly we will relate T_t and $\sqcup^t$* allowing to extend the latter naturally from S_t to $S \times S$ using T_t.

The idea is now that a process which at time t takes values in S_t is a *generalized branching process*, if for each $s < t$ conditional on time s, the $(t - s)$-truncation of a state in S_t has the same distribution as the concatenation of independent subfamilies originating from the individuals alive at time s. Then we *generalize* the generator criterion 1.6 to this structure working with the whole collection of semigroups and truncations.

Outline. The results on our four goals we present in two sections 2 and 3. We make these ideas described above rigorous in Section 2 and give the precise result, which characterizes the generalized branching property via the generator criterion. In Section 3 we give an application to processes of evolving genealogies. Namely to the $\mathbb{U}$-valued Feller diffusion in Section 3.1 and in Section 3.2 to the $\mathbb{U}^V$-valued super random walk and to historical processes or more general evolving populations with individuals carrying ancestral path. Altogether we have the four main results Theorems 2.11, 3.3, 3.14, 3.17, the first and the last the highlights.

The proof of the criterion is given in Section 4. The proofs of all other statements, those concerning applications, are in Sections 5, 6 and then in 7 we formulate the additional arguments needed for the extension of the claims to *spatial* models.

2. Results 1: Generator Characterization of Generalized Branching

We now carry out the last two points of the last sections rigorously with the key result Theorem 2.11.

Formal framework. The whole section works for processes with Polish *state space* S arising as solution of a *well-posed martingale problem*. We need the following assumptions on the state space S which formalizes the maps T_t, *truncation* and the binary operation $\sqcup^t$, *concatenation* as well as their relation, which is a form of *consistency* property, together with the further technical Assumption 2.6 below.

Assumption 2.1 (Collection of semigroups with consistent truncation maps). Let S be a Polish space. We use $\mathcal{B}(S)$ for the Borel σ-field on S and $b\mathcal{B}(S)$ to denote the bounded measurable functions on S.

Assume there are $S_t \subseteq S$, $t \geq 0$ with the following properties.

- $S_s \subseteq S_t$ for $0 \leq s \leq t$.
- S_t is closed in S for all $t \geq 0$.
- There is for every $t > 0$ a continuous mapping $T_t : S \mapsto S_t$, which is the identity if restricted to S_t: $T_t(x) = x$ for any $x \in S_t$.
- There is for all $t \geq 0$ a binary operation $\sqcup^t : S_t \times S_t \to S_t$ such that $(S_t, \sqcup^t)$ is a commutative topological semigroup with neutral element $0 \in S_0$.
- The extension of $\sqcup^t$ to all elements of S is defined for all $t \geq 0$ via

$$\sqcup^t : S \times S \to S_t,\ (x,y) \mapsto (T_t(x)) \sqcup^t (T_t(y)). \tag{2.1}$$

For simplicity, we drop the index t at $\sqcup^t$ if it's clear from the context.

There are many natural examples for this structure without the dependence on t, i.e. $S = S_t$ for all t, namely all classical case and the classical branching property, see Example 1.2, 1.3 and the list 1.4, where we saw they include $S = [0, \infty)$ and $S = \mathcal{M}_f(E)$, the space of finite Borel measures which both are semigroups when equipped with addition (of reals or measures, respectively).

However including *genealogical* or *historical* information requires $t-$dependence and the generalized branching property, here are examples for both these effects.

Example 2.2 (Ultrametric measure space $\mathbb{U}$-valued processes). Recall the setting of Glöde et al. (2019) and Depperschmidt and Greven (2019) using equivalence classes of ultrametric measure spaces and the semigroup of t-forests. In that paper the evolving *genealogy of the population* alive at time t was described via a set of individuals U_t, the genealogical distance between individuals $r_t(\cdot, \cdot)$ on $U_t \times U_t$, a population size $\bar{\mu}_t$ and a sampling (probability) measure $\hat{\mu}_t$ on U_t, altogether giving an *ultrametric measure space* $(U_t, r_t, \bar{\mu}_t\hat{\mu}_t)$ and finally with its isomorphy class $[U_t, r_t, \bar{\mu}_t\hat{\mu}_t]$ we get the elements of the *state space* $\mathbb{U}$ describing for us genealogies. (These objects are called often trees in some literature even though they describe the subset of the *leaves* of weighted $\mathbb{R}$-trees, since nevertheless for every ultrametric space there is an $\mathbb{R}$-tree such that the space can be mapped $1 - 1$ isometric in the set of leaves, see Greven et al. (2013), Remark 2.2 for details.)

Elements of $\mathbb{U}$ are $t-$*truncated* by truncating the metric at $2t$ and two such objects are *concatenated*, denoted $\sqcup^t$, by taking the disjoint union of the sets of individuals, keeping the metric in each population and setting the distance between individuals from different sub-populations equal to $2t$ and by adding the (extended) measures. Then $S = \mathbb{U}$, $S_t = \mathbb{U}(t)^\sqcup$, the latter the equivalence classes of ultrametric measure spaces of diameter at most $2t$ and $T_t\mathfrak{u} = \lfloor \mathfrak{u} \rfloor(t) = [U, r \wedge 2t, \mu]$ the t-truncation map, $t \geq 0$, $\mathfrak{u} = [U, r, \mu] \in \mathbb{U}$. Details are in Section 3.1.

Example 2.3 (Historical process: measures on ancestral paths). Consider $S_t = \mathcal{M}_f(D^t(\mathbb{R}, E))$, $t \geq 0$ where $D^t(\mathbb{R}, E) = \{f \in D(\mathbb{R}, E) | f(u) = f(0)\ \forall u \leq 0, f(u) = f(t)\ \forall u \geq t\}$, as subsets of measures on càdlàg functions on $\mathbb{R}$ with values in E, a locally compact metric space, for example E would be a geographic space as $\mathbb{Z}^d$ or $\mathbb{R}^d$ or E is a type space as finite sets (genome) embedded in $[0, 1]$. These paths could either represent the *geographic location* in the past along the ancestral line through *migration*, or of say a *genetic type* evolving under *mutation*.

Now there are two possibilities to proceed. This is due to the fact that of interest for us is the ancestral path *seen from the present individual*, i.e. all positions are relative to this present location or alternatively we can thin of the class of path differing only by time-shift as the key object.

(1) The set S_t consists then of the measures on the path constant except for an interval of at most length t. The equivalence classes are taken w.r.t. *translation in time* of the path. The truncation of the state is the push forward under the map: replace the path after time t by its time t value (this means after the $t-s$ truncation the equivalence class is a path which has evolved for time $t - s$ and is otherwise constant, the bracket indicates taking the equivalence classes. The binary operation is the one induced (on the equivalence classes $[\cdot]$) by the addition of measures.

(2) Alternatively we could at time t *shift all path by $-t$* in the time index to get path which are not constant only for times in $[-t, 0]$. The truncation and concatenation are defined analog to (1).

Indeed we shall later exploit mainly this second possibility. See Section 3.2.

Certain functions on the semigroup will play an important role.

Definition 2.4 (*t*-multiplicativity and *t*-additivity). Let $f : S \to \mathbb{R}$ measurable and $t \geq 0$. We say that f is *t-multiplicative on S* if

$$f(x_1 \sqcup^t \cdots \sqcup^t x_n) = f(x_1) \cdots f(x_n), \quad n \in \mathbb{N}, \, x_1, \ldots, x_n \in S. \qquad (2.2)$$

We say that f is *t-additive on S* if

$$f(x_1 \sqcup^t \cdots \sqcup^t x_n) = f(x_1) + \cdots + f(x_n), \quad n \in \mathbb{N}, \, x_1, \ldots, x_n \in S. \qquad (2.3)$$

Remark 2.5.   (1) In fact the previous definition means that multiplicative functions f are semigroup-homomorphisms $f : (S_t, \sqcup^t) \to (\mathbb{R}, \cdot)$ *except* for the usual continuity assumption for topological semigroup homomorphisms. Likewise for t-additive functions and the semigroup $(\mathbb{R}, +)$.

(2) The previous definition implicitly implies that $f(x) = f(T_t(x))$, $x \in S$ for a t-multiplicative (or t-additive) function. This can be seen in the case of a t-multiplicative function via

$$f(x) = f(x)f(0) = f(x \sqcup^t 0) = f(T_t(x) \sqcup^t 0) = f(T_t(x)). \qquad (2.4)$$

In order for the above functions being a rich enough set to work with later on we complement the Assumption 2.1 and require:

Assumption 2.6. For any $t \geq 0, b\mathcal{B}(S_t)$ contains a set $D_t \subset \{f \in b\mathcal{B}(S)|f(x + y) = f(x)f(y), \, x, \, y \in S_t\}$ of functions which are all strictly positive and the set separates points in S_t, i.e. for all $x \neq y \in S_t$ we can find $f \in D_t$ with $f(x) \neq f(y)$.

The examples 2.2 and 2.3 above will be shown to satisfy this further condition. **Key result.** Next we define the new concept of the *generalized* branching property which holds for a richer class of Markov processes then the Markov processes satisfying the classical one. This is formalizing the property that a process which at time t takes values in S_t is a *generalized branching process*, if for each $s < t$ conditional on time s, the $(t - s)$-truncation has the same distribution as the concatenation of independent subfamilies originating from the individuals alive at time s.

Definition 2.7 (Generalized branching property). Suppose Assumptions 2.1 and 2.6 hold. We are given furthermore a Markov process $(X_t)_{t \geq 0}$ with values in S

defined by a family $(P_t)_{t\geq 0}$ of probability kernels $P_t : S \times \mathcal{B}(S) \to [0,1]$, $t \geq 0$, given via

$$P_t(x, f) = \mathbb{E}[f(X_t)|X_0 = x], \ x \in S, f \in b\mathcal{B}(S), t \geq 0. \tag{2.5}$$

The process has the *generalized branching property* if

$$P_t(x_1 \sqcup^s x_2, h_t) = P_t(x_1, h_t)P_t(x_2, h_t), \quad x_1, x_2 \in S_s\,, \tag{2.6}$$

for any $s, t \geq 0$ and $h_t \in b\mathcal{B}(S)$ t-multiplicative on S_t.

This defines as well the generalized branching property for a solution of a well-posed martingale problem on $D([0, \infty), S)$.

We write h for $h(t, x) = h_t(x), (t, x) \in \mathbb{R} \times S$. Of course in the previous definitions it suffices to consider a *separating subset* of D_t functions in (2.6).

Remark 2.8 (Terminology). Note that we really have here a "$((S_t, \sqcup^t)_{t\geq 0}, (T_t)_{t\geq 0})$-branching property", but since we will specify these ingredients in examples we suppress this in the notation throughout.

Remark 2.9 (Time-homogenity of $(X_t)_{t\geq 0}$). If we want to treat the time-inhomogeneous case we would have to consider $((S_{u,t}, \sqcup^{(u,t)})_{t\geq u}, (T_t)_{t\geq u})_{u\geq 0}$ here setting of a cascade of indices which we want to avoid to focus on the key point.

Remark 2.10 (Measurable path). Recall at this point that solutions of a martingale problem must have a version with *measurable* path, since otherwise the defining relation for the martingale problem is not welldefined, since the compensator would not be properly defined. This property is implied by the stochastic continuity of the process, but is in infinite dimensional state spaces not equivalent to it, therefore requiring stochastic continuity is a restriction.

From our heuristic reasoning it should be clear that the time an evolution has run is of importance. Therefore we will even in the time-homogeneous case i.e. a time independent operator for our martingale problem generating the process X consider the time-space process $(t, X_t)_{t\geq 0}$ to have the time the evolution has run explicitly coded in the state.

We can now formulate the main result which gives the *characterization of operators* A of processes $(X)_{t\geq 0}$ which satisfy the *generalized branching property*. We consider here the time-space process $(t, X_t)_{t\geq 0}$ with state space $[0, \infty) \times S$ and operator $A + \frac{\partial}{\partial t}$. By Ethier and Kurtz (1986, Lemma 4.3.2) the *wellposedness of the (A, D)-martingale problem implies wellposedness of the $(\tilde{A}, \tilde{D})$ martingale problem*.

Theorem 2.11 (Criterion for operators of generalized branching processes). *Suppose Assumptions 2.1 and 2.6 hold and let D_t denote the set introduced there. Let*

$$\tilde{D} \subset \{(x, t) \mapsto \psi(t)h_t(x) : \ \psi \in C_b^1(\mathbb{R}, \mathbb{R}), \ h_t \in D_t, \ (t \mapsto h_t) \in C^1(\mathbb{R}, C_b(S))\} \tag{2.7}$$

and let furthermore be given $\tilde{A} = A + \partial_t : \tilde{D} \to B(\mathbb{R} \times S)$. Finally assume that for any $(x, 0) \in S \times \mathbb{R}$ the following holds.

For any two solutions $(t, X_t)_{t\geq 0}$ and $(t, X'_t)_{t\geq 0}$ of the martingale problem for $(\tilde{A}, \tilde{D}, \delta_{(x,0)})$ one has $T_t X_t \overset{\mathrm{d}}{=} T_t X'_t$ for every $t > 0$, and a solution $(t, X_t)_{t\geq 0}$ has a stochastically continuous version.

Under those conditions the family $(P_t)_{t\geq 0}$ as in (2.5) has the generalized branching property if and only if either of the following conditions is satisfied:

(1) *For $x_1, x_2 \in S_t$, $\psi h \in \tilde{D}$, $t \geq 0$:*

$$\tilde{A}\psi(t)h_t(x_1 \sqcup^t x_2) = \psi'(t)h_t(x_1 \sqcup^t x_2) + \psi(t)[h_t(x_2)\tilde{A}h_t(x_1) + h_t(x_1)\tilde{A}h_t(x_2)]. \quad (2.8)$$

(2) *For each $\psi h \in \tilde{D}$ there exists a function $g_{\psi,h} : \mathbb{R}_+ \times S \to \mathbb{R}$ such that $g_{\psi,h}(t, \cdot)$ is t-additive for each $t \geq 0$, the h_t are all strictly positive and, for all $(t, x) \in \mathbb{R}_+ \times S_t$,*

$$\tilde{A}\psi(t)h_t(x) = \psi'(t)h_t(x) + \psi(t)g_{\psi,h}(t, x)h_t(x). \quad (2.9)$$

Note that (a) is the generalized Kurtz criterion and it is (b) which we will apply in examples.

In the next section we apply part (b) of this result to $\mathbb{U}$-*valued* and *historical* processes to get *new* examples for the generalized branching property, where the "classical" one does *not* hold. In this section we conclude by explaining how the previous result simplifies if $S_t = S$ for all $t \geq 0$ as in our classical examples 1.2- 1.4, where then the classical branching property holds. This applies in particular to all classical branching processes mentioned in the introduction, hence it gives an alternative proof of the branching property of measure-valued CSBP in particular of all real-valued CSBP.

Indeed if $S = S_t$ for all $t \geq 0$ then multiplicativity of h_t on S_t just means that h_t is multiplicative on S. Assume $D \subseteq b\mathcal{B}(S)$ is a set of multiplicative functions and

$$\tilde{D} \subset \{(x,t) \mapsto \psi(t)h(x) : \psi(t) = e^{-\lambda t}, \lambda > 0, h \in D\}. \quad (2.10)$$

We get, recall from above $(A, D), (\tilde{A}, \tilde{D})$-well-posedness are equivalent, the following corollary to Theorem 2.11 giving the *classical examples* in (1.7)-(1.13) by choosing for D the functions specified there and for $\sqcup$ just $+$.

It is then immediate from Theorem 2.11, that we have the following.

Corollary 2.12 (Branching generator: classical case)**.** *Assume S is a Polish space. The test functions D satisfy, (i) $D \subset b\mathcal{B}(S)$ is a set of multiplicative functions on S, (ii) contains only strictly positive functions and is (iii) separating. Finally require that the (A, D)-martingale problem is well-posed and has a stochastically continuous solution $(X_t)_{t \geq 0}$.*

Then the semigroup associated to $(X_t)_{t \geq 0}$ has the branching property if and only if either of the following conditions is satisfied.

(1) *For all $x_1, x_2 \in S$, $h \in D$:*

$$Ah(x_1 \sqcup x_2) = h(x_2)Ah(x_1) + h(x_1)Ah(x_2). \quad (2.11)$$

(2) *For each $h \in D$ there exists a $\sqcup$-additive function $g_h : S \to \mathbb{R}$, i.e. $g_h(x_1 \sqcup x_2) = g_h(x_1) + g_h(x_2)$ for any $x_1, x_2 \in S$, with*

$$Ah(x) = g_h(x)h(x), \quad x \in S, h \in D. \quad (2.12)$$

3. Results 2: New Applications for some evolving genealogies or histories

This work was motivated by trying to generalize and prove the branching property for *evolving genealogies* modeled as marked ultrametric measure spaces, where the classical branching property does not hold. Of particular interest are *Feller's diffusion model* and the *super random walk*, see Depperschmidt and Greven (2019).

Note that we need in order to apply our theory that the martingale problem is well-posed. Hence we need to know this either from the literature or we have to prove this here. For reasons of keeping page numbers under control we give three examples and treat only one example, where we construct a new process and the others are chosen such that we can refer for the constructions of the process to existing literature. The corresponding results we have here are Theorems 3.3 , 3.14, 3.17.

Indeed we now discuss two classes of examples, each in a subsection which are processes of *evolving genealogies* respectively *histories*, arising as diffusion limits of classical Galton-Watson models. First we treat the prototype case which motivated the present paper, the $\mathbb{U}$-*valued Feller diffusion*.

Then next we adapt this to the world of *spatial* $\mathbb{U}^V$-valued branching processes covering in particular the *genealogical super random walk* or the corresponding *historical super random walk process*. This shows the potential of the criterion for the study of population models. There is a whole zoo of further examples. With the same method we could treat branching random walks, branching Brownian motions or multitype branching processes where individuals undergo mutation or more general continuous state branching processes.

Finally we formulate and construct a *new type* of example the *(ancestral-path)-marked* genealogical super random walk combining histories and genealogies. This shows the robustness of our concepts and methods for a *new* process of great interest in its own right.

Remark 3.1. The last model contains the other ones mentioned in the sense that they are functionals. However that does not mean that we get the branching property that easy from the general result, since first of all we have to still check the functionals are Markovian, which requires wellposedness of their martingale problems to obtain the Markov property of the functional. Furthermore we need to show that the ingredients $((S_t, \sqcup^t)_{t\geq 0}(T_t)_{t\geq 0})$ arise as projections. Therefore we do not loose much by building our examples from bottom to top.

Then in the last subsection we describe further examples where the criterion is probably applicable if one carries out some extensions of the theory of $\mathbb{U}$-valued processes to construct them rigorously via well-posed martingale problems. An example is the *continuum space* spatial models as the Dawson-Watanabe super process.

3.1. $\mathbb{U}$-*valued Feller Diffusion.*

Description of current population and its genealogy. We consider now the evolution of the genealogy of the *population currently alive* (i.e. at a time t) described by an equivalence class of ultrametric measure spaces, in a continuum mass version of a critical binary branching process. This is the $\mathbb{U}$-*valued Feller diffusion*. The *main result* of this subsubsection is *Theorem 3.3*. However first we have to explain the state space and the process. To explain the method of describing genealogical information consider first a critical Galton-Watson process.

The idea is to give two individuals in the population alive at time t a genealogical distance, which is twice the time one has to go back to the most recent common ancestor and to equip the population with the uniform distribution. Then we get an ultrametric measure space which describes the genealogy and whose evolution we follow. The exchangeability of names and with it the exchangeability of individuals with equal sampling weight then suggests to pass to *equivalence classes* of these

objects under *weight preserving isometries*. More precisely we proceed as follows and begin by shortly reviewing the state space of this process.

The genealogy is modeled as equivalence class of a ultrametric measure spaces $[U, r, \mu]$ where the set U describes the set of *individuals alive* at the current time, the ultrametric r on U the *genealogical distance* between individuals and $\mu = \bar{\mu}\widehat{\mu}, \bar{\mu} \in \mathbb{R}^+$ the *population size* and $\widehat{\mu} \in M_1(U, \mathcal{B}(U))$ the *sampling measure* specifying how to draw samples of typical individuals from the population alive at time t as for example the uniform distribution. The *equivalence class* is denoted, $\mathfrak{U} = [U, r, \mu]$, equivalence of representations is defined w.r.t. the *isometries* of the $supp(\mu) \subseteq U$ which are *measure preserving*. For $\bar{\mu} = 0$ the measure $\widehat{\mu}$ is not defined. The corresponding element we call the zero tree.

The space of all equivalence classes is denoted

$$\mathbb{U}. \tag{3.1}$$

Note that a finite set of points is as metric space characterized by the pairwise distances, i.e. the distance matrix. The set $\mathbb{U}$ is equipped with the *Gromov weak topology* and is with this topology a *Polish* space. The topology can be defined, see Greven et al. (2009),Glöde (2012) introducing a metric. It is well known however that this topology arises, if we require that a sequence of elements $\mathfrak{u}_n$ converges to an element $\mathfrak{u}$ iff the sequence of distance matrix distributions ν_n converges to the distance distribution ν of all orders $m \in \{2, 3, \ldots\}$. Here the distance distribution of order m arises (denoting with the Π^* the push forward of measures under a map Π) as $R^*(\mu^{\otimes m})$ with $R : U^m \to (\mathbb{R}_+)^{\binom{n}{2}}$ given by

$$R\left((x_i, x_j)_{1 \leq i < j \leq n}\right) = (r(x_i, x_j))_{1 \leq i < j \leq m}. \tag{3.2}$$

The description of genealogies by ultrametric *probability* measure spaces was introduced in Greven et al. (2009) extending ideas appearing in Evans (2000) and Evans et al. (2006). This was generalized to the case of *finite* measures in Glöde (2012) and to locally bounded measures in Greven et al. (2016).

The objects of Assumption 2.1 now arise as follows. Recall here that the ultrametric spaces can be represented by the set of leaves in an $\mathbb{R}$-tree. This motivates the following definitions. The state spaces are the equivalence classes $\mathfrak{u} = [U, r, \mu]$ of ultrametric measure spaces (U, r, μ) called $\mathbb{U}, \mathbb{U}(h)^{\sqcup}$ are the elements of $\mathbb{U}$ with diameter at most $h, \mathbb{U}(h)$ with diameter strictly less than h and one considers the *truncation of "trees"* of diameter t at height $t - h$ for $h \in (0, t]$ and the $h-concatenation$ of "trees" as the binary operation on $\mathbb{U}(h)^{\sqcup}$.

Formally define

$$\mathbb{U}(h)^{\sqcup} = \left\{\mathfrak{u} \in \mathbb{U} \mid \mu^{\otimes 2}(\{(x, y) \in U^2 \mid r(x, y) > 2h\}) = 0 \tag{3.3}$$

$$\mathbb{U}(h) = \left\{\mathfrak{u} \in \mathbb{U} \mid \mu^{\otimes 2}(\{(x, y) \in U^2 \mid r(x, y) \geq 2h\}) = 0. \tag{3.4}$$

Then define for $u, v \in \mathbb{U}(h)^{\sqcup}$ the *concatenation* (using $\uplus$ for disjoint union):

$$\mathfrak{u} \sqcup \mathfrak{v} = [U \uplus V, r_U \sqcup^h r_V, \widetilde{\mu} + \widetilde{v}], \text{ with} \tag{3.5}$$

$$r_U \sqcup^h r_V \mid_{U \times U} = r_U, \quad r_U \sqcup^h r_V \mid_{V \times V} = r_V, \tag{3.6}$$

$$r_U \sqcup r_V(x, y) = 2h, \ x \in U, y \in V \tag{3.7}$$

and $\widetilde{\nu}, \widetilde{\mu}$ the extension of ν and μ to the disjoint union which is zero on the respective other component.

The *h-top* $\lfloor \mathfrak{u} \rfloor(h)$ of $\mathfrak{u} \in \mathbb{U}$ is defined:

$$\lfloor \mathfrak{u} \rfloor(h) = [U, r \wedge 2h, \mu] \in \mathbb{U}(h)^{\sqcup}. \tag{3.8}$$

Then we define the *h*-truncation:

$$T_h(\mathfrak{u}) = \lfloor \mathfrak{u} \rfloor(h), \ S_h = \mathbb{U}(h)^{\sqcup}. \tag{3.9}$$

The process: $\mathbb{U}$-valued Feller diffusion. We consider the $\mathbb{U}$-valued Feller diffusion $(\mathfrak{U}_t)_{t \geq 0}$ the process which corresponds to the large population-rapid reproduction-small mass limit of the genealogy of the population alive at time t of the critical Galton-Watson process. The process is the $\mathbb{U}$-valued process related with its population size process $(X_t)_{t \geq 0}$ to the solution of the SDE $dX_t = aX_t \, dt + \sqrt{2bX_t} \, dB_t$, where $b > 0$ and $a \in \mathbb{R}$. For the construction and uniqueness of $\mathbb{U}$-valued processes already treated in the literature we refer to Greven et al. (2013); Glöde (2012) and for the present model and its spatial versions to Depperschmidt and Greven (2019) where the process is constructed via a well-posed martingale problem and many properties of its longtime behaviour are studied.

We recall its operator for the *martingale problem*. We need the concept of a polynomial to get the domain of the operator. Fix $n \in \mathbb{N}$ and $\phi \in C_b^1(\mathbb{R}^{\binom{n}{2}}, \mathbb{R})$. Then define for an equivalence class of an ultrametric measure space $[U, r, \mu]$ the function

$$\Phi^{n,\phi}\Big([U, r, \mu]\Big) = \int_{U^n} \phi\Big((r(x_i, x_j), 1 \leq i < j \leq n)\Big) \, \mu(dx_1) \ldots \mu(dx_n) \tag{3.10}$$

and the action of the operator is given as sum of two operators:

$$\Omega^{\uparrow} \Phi^{n,\phi}(\mathfrak{u}) = \Omega^{\uparrow,\mathrm{grow}} \Phi^{n,\phi}(\mathfrak{u}) + \Omega^{\uparrow,\mathrm{bran}} \Phi^{n,\phi}(\mathfrak{u}) \tag{3.11}$$

and $\Omega^{\uparrow} \Phi^{n,\phi}(0) = 0$. We need the notation

$$\bar{\mathfrak{u}} = \mu(U) \tag{3.12}$$

for the total mass of (X, r, μ), which is an invariant of $[X, r, \mu]$. The operators on the r.h.s. are given by

$$\Omega^{\uparrow,\mathrm{grow}} \Phi^{n,\phi}(\mathfrak{u}) = \Phi^{n, 2\overline{\nabla}\phi}(\mathfrak{u}), \quad \overline{\nabla}\phi = \sum_{1 \leq i < j \leq n} \frac{\partial \phi}{\partial r_{i,j}}, \tag{3.13}$$

$$\Omega^{\uparrow,\mathrm{bran}} \Phi^{n,\phi}(\mathfrak{u}) = an\Phi^{n,\phi}(\mathfrak{u}) + \frac{b}{\bar{\mathfrak{u}}} \sum_{1 \leq k < l \leq n} \Phi^{n, \phi \circ \theta_{k,l}}(\mathfrak{u}), \tag{3.14}$$

where

$$\big(\theta_{k,l}(\underline{r})\big)_{i,j} := r_{i,j} \mathbb{1}_{\{i \neq l, j \neq l\}} + r_{k,j} \mathbb{1}_{\{i=l\}} + r_{i,k} \mathbb{1}_{\{j=l\}}, \quad 1 \leq i < j. \tag{3.15}$$

Note that the *martingale problem* for $(\Omega^{\uparrow}, \Pi(C_b^1))$, where $\Pi(A)$ are the polynomials defined by ϕ chosen from the set A, has a *unique* solution, see Depperschmidt and Greven (2019). This result is presented in Proposition 5.6.

Interpretation of generator. This generator arises in this form in a scaling limit of an individual based model (binary critical Galton-Watson process) as the number of individuals tends to infinity. The first part correspond to the growth of the genealogical distance with time arising from the growth of the genealogical tree at the tree top only. The second part describes the effect of a splitting in two branches of the genealogical tree upon a surviving birth. The term $\bar{u}$ in (3.14) arises as rate of the operator which in fact is the resampling operator defining the $\mathbb{U}$-valued Fleming-Viot model. The background here is the fact that, if we have two groups of ancestors and follow the relative proportion of one among the complete population, we obtain a Fisher-Wright diffusion at a rate which is the inverse of the total mass.

See Depperschmidt and Greven (2020) for more explanation and Depperschmidt and Greven (2019) for the derivation from an individual based model and detailed information on the structure of the generator and its relation to the Fleming-Viot operator.

Definition 3.2 ($\mathbb{U}$-valued Feller diffusion)**.** We refer to the unique solution of the $(\Omega^{\uparrow}, \Pi(C_b^1))-$ martingale problem as $\mathbb{U}$-*valued Feller diffusion* and denote it by $\mathfrak{U} = (\mathfrak{U}_t)_{t \geq 0}$.

The result. We consider here as binary operation the *concatenation* $\sqcup^s$ of trees in $\mathbb{U}^{\sqcup}(s)$ and as operation T_t the *truncation* operation, which associates if we consider the $\mathbb{R}$-tree representation of the ultrametric space with such a tree its *t-tree top* of depth $t \leqq s$, with S_t the set of such objects.

We state the main result of this subsection obtained with our criterion.

Theorem 3.3 (Generalized branching property: $\mathbb{U}$-valued Feller)**.** *The $\mathbb{U}$-valued Feller diffusion $\mathfrak{U}$ has the generalized branching property.*

A concrete consequence: expected sum of squares of subfamily sizes. At first sight, Theorem 3.3 seems to be an abstract statement about an "extended" martingale problem on a complicated space. However if we use Lemma 5.2 from the proof section we can obtain interesting statements, since we can obtain differential equations for expectations of moments where the rhs. involves lower order moments. This allows in some case to obtain an *explicit solution* of such closed systems of equations. The following is an example. There are much more explicit representations coming out of the martingale problem, in particular via duality relations, see here Depperschmidt and Greven (2019) for details.

We consider the expected sum of the squares of the subfamily sizes.

Theorem 3.4 (Moment recursion for Feller diffusion)**.** *For the $\mathbb{U}$-valued Feller diffusion and $t > 0$, if $a \neq 0$:*

$$\mathbb{E}[\Phi^{2, \mathbb{1}(r_{12} < 2t)}(\mathfrak{U}_t)] = b\bar{u}_0 \frac{1}{a} \left(e^{2at} - e^{at} \right). \tag{3.16}$$

Note the rhs. is nothing else than the variance of the Feller diffusion on $\mathbb{R}^+$. It is known that if $\mathfrak{U}_t = [U_t, r_t, \mu_t]$ then $(\mu_t(U_t))_{t \geq 0}$ is the classical $\mathbb{R}^+$-valued Feller diffusion with parameter b, see Depperschmidt and Greven (2019). In the case $a = 0$ the right hand side of (3.16) is replaced by $2b\bar{u}_0$.

So far we are not able to derive from the martingale problem other equations which form a closed system and allow for an *explicit* solution. However the compensator of the martingale problem involves lower order expressions which gives some hope.

3.2. *Historical Branching Processes, evolving genealogies of super random walk and path-marked genealogies.* We consider now *stochastic evolutions* of the *genealogy* of populations distributed in a *geographic space* (or type space) E, more precisely a Polish space (E, r_E). In addition we have an evolution of the marks, think of migration if marks are locations and of mutation if they are types. Since the mechanisms are linear in a proper sense we can still hope for a generalized branching property. This modification of the process requires the generalization of the setup of the previous subsection from the state $\mathbb{U}$ we have to pass to the state space of marked genealogies $\mathbb{U}^V$.

Here E is separable, complete, metric space which is locally compact as $E = \mathbb{Z}^d$ or any other at most countably infinite abelian group with a metric or a continuum for example $E = \mathbb{R}^d$. We also consider the case where we have further *historical information* on the whole ancestral paths of the population as they evolve in time.

The *key results* on the generalized branching property are on versions of the model initially motivating the present paper, namely enrichments of the classical *super random walk* (sometimes called interacting Feller diffusions, see Dawson and Greven, 2003 for example). In particular the *genealogical super random walk* process, the *ancestral path marked genealogical super random walk* process and the *historical process of super random walk*, which are *Theorem 3.14, Theorem 3.17* and its *Corollary.*

However we have first to define these processes and develop the ingredients for our framework which needs some effort. We give here the concepts to handle E of the form mentioned even though we prove the main theorem assuming only that E is a countable abelian group, because the theory of such processes is for continuum space not yet well developed on the *genealogical level* i.e. as $\mathbb{U}^V$-valued process with $V = E$ to be introduced below. However as soon as we have settled the *existence and uniqueness problem* of our process we can readily verify the generalized branching property via our criterion.

We discuss first of all three important and related processes which are spatial versions of the case discussed in Subsection 3.1, among which is the *historical process* or the *location-marked genealogies* of the *super random walk* modeled as marked ultrametric measure spaces and show that their generators allow to read off the generalized branching property via our criterion and therefore get more and *new* examples, in particular ones not satisfying the classical branching property.

At the same time we introduce a third model (containing the above as Markovian functionals). Namely we take up and generalize from *historical processes*, see chapter 12 in Dawson (1993), the concept of *ancestral paths* associated with individuals currently alive and combine this with the genealogy described by elements of $\mathbb{U}$. Then we can define here a *(ancestral paths)- marked genealogy-valued*, precisely $\mathbb{U}^V$-valued, class of *super processes*. This class contains as functionals *many* processes. In particular all processes describing *historical* and *genealogical information* of an evolving population: location-marked $\mathbb{U}$-valued processes, $\mathbb{U}$-valued processes, historical processes, measure but also $\mathbb{R}^d$ or $\mathbb{R}$-valued processes for which the generator

criterion for the branching property works very well. However the case of ancestral path marked genealogies is conceptually and technically *quite demanding*.

Outline. We proceed now in four steps, first we develop the formal description of *genealogical* and *historical information*, second we rigorously define the involved stochastic processes via well-posed martingale problems and thirdly we identify the ingredients of the formal framework fitting Theorem 2.11 and finally we state in the fourth steps the main results on the generalized *branching property*. (For facts on relevant spaces of measures and Laplace methods see Section 3 and 4, 6, 8 of Dawson, 1993).

Step 1: Genealogical and historical information in spatial population models. We begin by introducing the *state space* of the involved processes systematically.

(1) Historical process. The *historical process*, a process with values in *measures on paths*, was invented in Dawson and Perkins (1991) to describe the *ancestral paths* of the population alive at time t in a spatial critical branching process. Here we assume that the population is observed from time 0 on and no further information on the past does exist.

Namely every individual alive has a path associated with the migration in space from his birth on, before this birth times the birth time of his direct ancestor, etc. so that the ancestral path of a currently alive individuals gives the *path of descent* (or ancestral path) and the motion through geographic space (or in type space under mutation).

(i) State space That is if E denotes the geographic space, which has to be at least a Polish space (most frequently with a specified metric, this the setup in chapter 12 of Dawson, 1993) one has an element of $D(\mathbb{R}, E)$ by continuing the path as constant beyond time t and before time 0. These paths merge of course and are in particular equal for all times before T if they belong to individuals descending from the same parent and T is the parents death time and the birth time of the two new descendants. All individuals are considered exchangeable if they have the same ancestral path. Hence for time t consider for the population currently alive simply the *"counting measure" on the space of the ancestral paths* to describe the state of the population.

(ii) Initial states As initial state in this description one uses typically initially *constant paths* so that the initial state corresponds to the locations of the individuals at the time 0 in E in a unique way. Note that in fact we have values in the closed subspace $D_{0,t}$ where $D_{0,t} \subseteq D$ are the càdlàg paths with values in E which are constant before time 0 and beyond t. This is a *closed* subset of D. To get a dynamically closed set of states we need to use $\underset{t>0}{\cup} D_{0,t}$. Similarly call $D_{0,\infty}$ the (topologically closed) set of paths constant for times ≤ 0. In order to then discuss infinitely old populations and equilibria we need $D_{-\infty,t}$ resp $D_{\infty,\infty} = D$. In particular we do have that

$$D_{s,t} \text{ with } s,t \in \bar{\mathbb{R}} \text{ is a } Polish \text{ state space for ancestral paths.} \qquad (3.17)$$

Traditionally one describes the state of these ancestral lines by a *locally finite measure*, i.e. a $\nu_t^* \in \mathcal{M}(D(\mathbb{R}, E))$ for a locally compact geographic space E (like $\mathbb{R}^d$ or some countable set E like $\mathbb{Z}^d$). In the general setup we have measures which are bounded on bounded sets. These measures generalize the *empirical measures* of a finite population of ancestral paths at varying times and contains in particular the

occupation measure of the population by projecting the measure ν on E by regarding from the path only the time-t position. This measure ν_t^* *"counts"* the ancestral path having some specified features, in particular projecting on the current position gives the *occupation measure* $\nu_t^{*,\downarrow}$ on E of the population at time t. Namely we define

$$\nu_t^{*,\downarrow}(A) = \nu_t^*(\{v \in D(\mathbb{R}, E) : v(t) \in A\}), \quad \forall A \in \mathcal{B}(E). \tag{3.18}$$

Now we say ν_t^* has the *"locally finite* property" if all its time-t projections satisfy:

$$\nu_t^{*,\downarrow}(A) < \infty, \text{ for all A bounded respectively finite in the discrete case.} \tag{3.19}$$

So we have altogether as *historical process* the following measure valued process:

$$(\nu_t^*)_{t \geq 0}, \ \nu_t^* \in \mathcal{M}(D_{0,t}(\mathbb{R}, E)). \tag{3.20}$$

Remark 3.5. The process X, as for example the process in (3.20) has typically a time *inhomogeneous* dynamics and therefore it is better to work with the *time-space process* (i.e. $(t, X(t))_{t \geq 0}$ instead of X) of the path process and on top of that also with the time-space process of the measure valued process i.e. we consider $(\mathbb{R} \times D(\mathbb{R}, E))$ respectively in (3.20) $\mathbb{R} \times \mathcal{M}(\mathbb{R} \times D_{0,t}(\mathbb{R}, E))$ as state space and consider measures on that space together with another *explicit time coordinate*. This way one obtains a *time-homogeneous* Markovian dynamics.

The historical process, better its law in its general form, see Dawson and Perkins (1991) can be described as solution to a Log-Laplace equation or for us more relevant as solution of a *martingale problem* (see Dawson, 1993, chapter 12), in the form which is specifying martingales for evaluations of the measure and giving their increasing processes. We will use here for our purposes the more traditional form of the (local) martingale problem for a given operator, which however is known to be *equivalent* to the above descriptions.

(2) Genealogical processes. The object in (3.20) describes the genealogy at least implicitly, only if independent copies of the migration paths do *not agree on any positive interval*, so that the time point before which the ancestral path agree must be the exact birth point. This problem can be avoided as follows, using the concept of *marked genealogy-valued processes* describing the genealogy of the population *currently alive*, which also allows to formulate and prove the generalized branching property with our criterion. In that concept it is possible to attach marks to the individuals as for example *types* or *locations*. This concept has been introduced in Depperschmidt et al. (2011) and has been successfully applied in the context of Fleming-Viot processes and their genealogies in Greven et al. (2013), Depperschmidt et al. (2012) and Greven et al. (2016) and we use it here for the branching world, see here for $\mathbb{U}$ or $\mathbb{U}^V$-valued processes of this form also Glöde et al. (2019); Depperschmidt and Greven (2019).

(i) State space: generalities We consider more precisely a random variable $\mathfrak{U}$ with values a *equivalence class* of $V-$*marked ultrametric measure spaces*

$$\mathfrak{U} = [(U \times V, r \otimes r_V, \nu)], \tag{3.21}$$

with (U, r) an ultrametric space, (V, r_V) is a metric space both complete and separable and ν is a Borel measure on $\mathcal{B}((U \times V, r \otimes r_V))$ which is *boundedly finite* meaning $\nu(U \times \cdot)$ is a boundedly finite (on bounded sets finite) measure on V. We denote by μ the measure on $(U, \mathcal{B}(U))$ given by $\mu(\cdot) = \nu(\cdot \times V)$, which is increasing limit of finite measures on that space. Here we have to allow measures μ now,

which are infinite, since on an infinite geographic space we typically want to allow a population with infinite total mass which is only *locally finite*.

Consider $(U \times V, r \otimes r_V, \nu)$ and $(U' \times V, r' \otimes r_{V'}, \nu')$. Call a map $\phi : U \times V \to U' \times V$ on *mark and measure preserving isometry* if $\phi(u, v) = (\widetilde{\phi}(u), v)$ for all $u \in supp(\nu)$ and $\widetilde{\phi}$ is isometric between $supp(\mu)$ and $supp(\mu')$ and measure preserving, i.e. $\phi_* \nu = \nu'$. The space of all equivalence classes, the latter denoted $[U \times V, r \otimes r_v, \nu]$, of $V-$marked ultrametric measure spaces w.r.t. *measure* and *mark and measure preserving isometries* of the support of μ of all *restrictions* to points with marks in bounded sets, equipped with the *marked Gromov weak topology*, we denote

$$\mathbb{U}^V, \text{ which is a } Polish\ space. \tag{3.22}$$

We may take equivalently any sequence of bounded sets exhausting the full space V.

See Depperschmidt et al. (2011) and Greven et al. (2016) for the concept of $V-$*marked* ultrametric measure spaces and basic topological facts. Roughly convergence amounts to convergence of all equivalence classes of marked finite subspaces spanned by n points sampled according to ν. (Alternatively we could say: convergence of all polynomials a concept we shall discuss in detail in (3.31).) Then define for $\mu \in \mathcal{M}(U)$ with $\mu(A) = \nu(A \times V)$ the kernel κ by

$$\nu = \mu \otimes \kappa, \tag{3.23}$$

called the *mark kernel*. This κ arises in the special case where a mark *function* exists as $\kappa(u, dv) = \delta_{\kappa(u)}(dv)$ for a measurable function κ. The second marginal of ν corresponding to V is denoted

$$\nu^*. \tag{3.24}$$

We want to choose the *current locations* or alternatively the *ancestral paths* as marks:

$$V = E \ \text{ or } \ V = D(\mathbb{R}, E). \tag{3.25}$$

The latter choice embeds then the historical process in the genealogical process on $\mathbb{U}^V$ which contains then the combined information of the $\mathbb{U}^E$ and the $\mathcal{M}(D(\mathbb{R}, E))$-valued process.

In (3.25) the first case is easier to handle (compare Greven et al. (2016) for the Fleming-Viot process in that case and Depperschmidt and Greven (2019); Glöde et al. (2019) for the branching case). For the second case we need some further *preparation* we focus on next, a reader only interested in the more classical situation, the first case (genealogical super random walk) might move on to the next step. Nevertheless, this second example is the highlight, despite the technicalities necessary.

(ii) State spaces: Path-marked genealogies   Indeed the spatial branching process of Dawson and Perkins (1991), the so called *historical process*, can be extended to a $\mathbb{U}^V$-valued process $\mathfrak{U}$ containing most of the relevant genealogical *and* historical information (see the discussion below) and then this ν^* of (3.24) is the measure state and with it we get a version of the historical process, formulated in the *"classical"* way, i.e. for the measurable Ψ on $\mathbb{U}^V$ arising from projecting an element of $U \times V$ on V and hence ν on ν^*, set:

$$(\nu_t^*)_{t \geq 0} = \Psi((\mathfrak{U}_t)_{t \geq 0}). \tag{3.26}$$

We formulate now a setup in which to apply the martingale problem techniques (see Remark 3.5). Introduce as mark space V:

$$D = \bigcup_{t \in \mathbb{R}} D_{[t,\infty)} \quad , \quad D^+ = \bigcup_{t \in [0,\infty)} D_{0,t}(\mathbb{R}, E). \tag{3.27}$$

We use parallel to the historical process also his time-space process (recall Remark 3.5), hence also as mark resp. state space:

$$V = \mathbb{R} \times D \text{ or } V = D^+ \text{ resp. } \mathbb{R} \times \mathbb{U}^V. \tag{3.28}$$

Note that we pass here both on the level of the ancestral path *and* on the $\mathbb{U}^V-$valued process to the time-space process because we want on both levels *time-homogeneous* processes (recall Theorem 2.11 is for time-homogeneous processes). In that case we choose as marks in points with the explicit time component t, paths in $D_{0,t}(\mathbb{R}, E)$, which are constant after time t and before time 0.

Step 2: The class of genealogical and historical models: Formal construction of $\mathbb{U}^V$-valued super random walk. We focus now on two cases, namely $V = E$ and $V = D$ where E is a metric space, which is Polish. But we restrict E to be a countable abelian group such that we can define *random walks*, which allows simplifications in construction and more is known about populations in such geographic spaces. The basic process is now the $\mathbb{U}^E$-*valued super random walk on* E, which we now introduce in (1) before we come to (2) to the case $V = D$ and in (3) to an alternative path-valued process. As preparation we begin with the classical case.

(0) A classical spatial process. The classical *super-random walk* is the following system of interacting diffusions $X = \left((x_i(t))_{i \in E}\right)_{t \geq 0}$ with parameter $b > 0$ and $a(\cdot, \cdot)$ a transition probability on $E \times E$ and state space contained in $[0, \infty)^E$:

$$dx_i(t) = \sum_{j \in E} a(i, j)(x_j(t) - x_i(t))dt + \sqrt{bx_i(t)} \, dw_i(t) \quad , i \in E, \tag{3.29}$$

where E is embedded in a continuum group E', which is Polish and with $a(., .)$ a transition probability on E some *discrete abelian group*, for example $E = \mathbb{Z}^d, E' = \mathbb{R}^d$. See Dawson and Greven (1996), Greven et al. (2002) for construction and properties of this process. Recall that in the approximating individual based model individuals migrate from i to j with probability $\bar{a}$ instead of a(here $\bar{a}(\xi, \xi') = a(\xi', \xi)$).

The associated *marked genealogy-valued*, i.e. $\mathbb{U}^V$-valued, dynamic has to be defined below rigorously by a *well-posed martingale problem* which can be show to arise as a *scaling limit* of the genealogy of a branching random walk on E, equipped with the appropriate marks in E resp. D, the walk with transition rate $a(\cdot, \cdot)$ with critical branching with many individuals-small mass and rapid branching.

Remark 3.6. It is also interesting to consider *continuum space* versions of the above process, as the Dawson-Watanabe process. The Dawson-Watanabe process as usual arises as spatial continuum limit from the systems indexed by scaled versions of E, for example $E' = \mathbb{R}^d$ and $E = \varepsilon \cdot \mathbb{Z}^d$, with $\varepsilon \to 0$, compare Section 3.3 part *(iii)* for a more detailed discussion of the arising problems.

If the underlying geographic space E is finite we work with finite measures. If however E is countable it is more natural to work with measures which are only *locally finite*. In that case the construction of the process of total masses (per site

population sizes) via the system of SDE's in (3.29) already requires to restrict the state space to configurations, where the local mass cannot explode in time due to the flow of migration. The appropriate tool here is the so called *Liggett-Spitzer space*, a subset of $[0, \infty)^E$ of the form

$$\{\overline{x} \in [0, \infty)^E \mid \sum_{i \in E} \overline{x}_i \gamma_i < \infty\} \tag{3.30}$$

for some summable positive γ satisfying $a\gamma \leq M\gamma$ for some $M < \infty$.

On this space the solution of the SSDE can be constructed and any random configuration which is translation invariant with $E \mid \overline{x}_0 \mid < \infty$ is almost surely in this space, regardless which γ we choose, and furthermore the solutions of the SSDE have almost surely *path* which take values in the Liggett-Spitzer space for all $t \geq 0$. We do not discuss this in more detail here, the reader may look at Liggett and Spitzer (1981), Dawson and Greven (1996) and Greven et al. (2005) for details.

(1) The $\mathbb{U}^E$-valued super random walk. We have to discuss now state spaces, test functions and operators.

(i) Marked genealogies of super random walk and their state space. For this process above we want to construct now the *genealogy* of the individuals alive at a given time t together with information on *locations*, which is described as an element of $\mathbb{U}^V$. Primarily this is a E-marked genealogy ($V = E$) which we have to define.

The next point needed are the genealogies which requires the extension of the theory of processes in *marked* ultrametric measure spaces from the state space of the previous subsection. Here we consider as marks on the genealogy the *current location* of the individual which is changing due to migration or more generally its *path of descent*, here the marks evolve according to the *path process*, see Dawson (1993). Processes on that space of marked genealogies have been introduced in Depperschmidt et al. (2012), for the case of finite sampling measures, we explain the basic concepts and facts we need here. A further point is to allow on infinite sets infinite population sizes in the form of *boundedly finite* measures. This works by considering the localization of the population to ones with marks in bounded subsets of the geographic space, defining the equivalence classes w.r.t. to all the restrictions and defining the topology by defining convergence by the convergence of all localizations to finite (bounded) sets. For detail we refer the reader to Greven et al. (2016) where also a class of genealogical processes is introduced, different from ours though, namely for the spatial genealogical Fleming-Viot process.

The $\mathbb{U}^V$-valued process is constructed as solution to a *well-posed martingale problem*. The martingale problem of the $\mathbb{U}$-valued process is treated in Depperschmidt and Greven (2019) together with its spatial $\mathbb{U}^E$-valued version in all detail for a survey see also Depperschmidt and Greven (2020).

As *starting points* for our dynamic we will allow states of a special nature by requiring that if $[U \times V, r, \nu]$ is such a state, then the restriction of ν to V satisfies (3.30), for $V = E$ and for the case of path that the further projection of the path onto the time t position satisfies this relation. This subset of $\mathbb{U}^V$ satisfying (3.30) is called then $\mathcal{E}$, a set defined by requiring the property of the projection of ν on V that is supported by the measurable subset.

(ii) Test functions, generators and the genealogical processes. We specify now the *generator* of the martingale problem and its *domain*. The basis is the operator

we had in (3.11)-(3.15) but we have to *lift* it to the marked case and we have to add the migration part giving the *dynamics of the marks* by specifying the *generator*.
(a) Test functions. The first step is to generalize the concept of a polynomial to cover marks. Choose now a function $\chi : V^n \to \mathbb{R}$, $\phi \in C_b(V^n, \mathbb{R})$ and consider *the monomial*:

$$\Phi^{n,\phi,\chi}([U \times V, r, \nu]) = \tag{3.31}$$

$$\int\limits_{(U \times V)^n} \phi\left((r(x_i, x_j))_{1 \le i < j \le n}\right) \chi(v_1, v_2, \ldots, v_n) \, \nu\left(d(x_1, v_1), \ldots \nu(d(x_n, v_n))\right),$$

where it suffices to consider χ of the form (these are still *separating*)

$$\chi(v_1, \cdots, v_n) = \chi^1(v_1) \cdots \chi^n(v_n). \tag{3.32}$$

Then consider the generated algebra the *polynomials*, denoted Π_E see Depperschmidt et al. (2011), Greven et al. (2016).

The simple case is where V is simply the geographic space then $V = E$ and $\chi^k : E \to \mathbb{R}$. Then we can consider again *Laplace functionals* $exp(-\Phi^{n,\phi,\chi})$ for $\phi, \chi \ge 0, n \in \mathbb{N}$. In this marked case where E is not a finite (bounded) set and where we work with populations which are not necessarily finite but are only *locally finite* we take $E_m \uparrow E, E_m$, finite (bounded) and restrict the χ to be of finite (bounded) support. We work with χ^k of the form that it specifies a single site of observation, i.e.

$$\chi^k(\cdot) = 1_{\{\xi^k\}}(\cdot), \xi^k \in E \text{ and write for } \chi \text{ in that case } \chi^{\underline{\xi}} \text{ for } \underline{\xi} = (\xi^1, \cdots, \xi^n) \in E^n. \tag{3.33}$$

Then the polynomial Φ depends only on the population in a finite number of sites in E. Then we can take the generated algebra and have a separating set (Greven et al. (2016)). The reader only interested in the super random walk $V = E$ may skip the next point and continue directly with the dynamics.
(b) The dynamics and the operator. Next the action of the *operator*. The growth and branching operator from (3.13) and (3.14) act now as follows. The *growth operator* acts only as before on ϕ and the *replacement operator* the same but only if the marks of the two chosen points have the *same current locations*, i.e. are marked with path with the *same* current *site*.

The branching operator $\Omega^{\uparrow, \mathrm{bra}}$ however is now a *sum* of operators $\Omega_\xi^{\uparrow, \mathrm{bra}}$ acting on the population at the site ξ, the sum over $\xi \in E$. We have

$$\Omega_\xi^{\uparrow, \mathrm{bra}} \Phi^{m,\phi,\chi} = \frac{2b}{\bar{u}_\xi} \, \Phi^{m,\phi',\chi}, \text{ with } \phi' = \sum_{\substack{k,i=1 \\ k \ne i}}^{m} (\theta_{k,i} \, \phi - \phi) 1_{\{v_i = v_k = \xi\}}. \tag{3.34}$$

Note that our polynomial depends only on a *finite* number of sites so that the sum is well defined.

Also the operator $\Omega_\xi^{\uparrow, \mathrm{bra}} \Phi^{m,\phi,\chi}$ is only non zero for such $\xi \in \{\xi^1, \ldots, \xi^{n(\chi)}\}$ which specify χ as in (3.33) and (3.41) such that $\bar{u}_{\xi^k} \ne 0$.

Note also at this point that now for $\bar{u}_\xi = 0$ for some ξ on which the polynomial depends we have a *singularity* and we set then $\Omega_\xi^\uparrow \Phi^{m,\phi,\chi} = 0$. Observe that this occurs now even if the population as a whole is *not* yet extinct. This requires restrictions on the test functions we can use in the martingale problem. Here we refer for the technicalities also to Dawson and Greven (2003).

Next we need the operator of the mark evolution the *migration* operator. We need here that the path of marks arises from a Markov process $(Y(t))_{t\geq 0}$ on E solving a *well-posed martingale problem*, with operator A. For example for the branching random walk the migration rate $\bar{a}$, where $\bar{a}(v,v') = a(v',v)$ for $v,v' \in E$ we have:

$$(Af)(v) = \Big(\sum_{v'\in E} \bar{a}(v,v')f(v') \Big) - f(v), \quad v \in E, f \in b\mathcal{B}(E,\mathbb{R}), \qquad (3.35)$$

representing the motion of individuals on E.

We define the operator $\Omega^{\uparrow,\mathrm{mig}}$ describing the evolution of the marks driven by the migration defined by A above and which acts on the polynomial $\Phi^{\phi,\chi}$ as follows:

$$\Omega^{\uparrow,\mathrm{mig}}\Phi^{n,\phi,\chi} = \sum_{k=1}^{n} \Phi^{n,\phi,A_k\chi} \ , A_k\chi = \Big(\prod_{i\neq k} \chi^{\xi^i} \Big) A\chi^{\xi^k}. \qquad (3.36)$$

Then summing the *migration operator* in (3.36), the *growth operator* and the *resampling operator* $\Omega^{\uparrow,\mathrm{bra}}$ (see (3.34) and above) results in an operator

$$(\Omega^{*,E}, \Pi_E). \qquad (3.37)$$

Definition 3.7 ($\mathbb{U}^E$-valued super random walk)**.** The well-posed $(\Omega^{*,E}, \Pi_E, \mathfrak{u})$-martingale problem (see the non-spatial case and the spatial case in Depperschmidt and Greven (2019) for the wellposedness result) specifies for $\mathfrak{u} \in \mathbb{U}$ with $\bar{\mathfrak{u}} \in \mathcal{E}$ a process :

$$(\mathfrak{U}_t^{*,E})_{t\geq 0} \qquad (3.38)$$

the $\mathbb{U}^E$-*valued super random walk on* E.

(2) The (ancestral path)-valued case: D^+ Next we have to focus on the case of *ancestral path as marks*. Recall here (3.27). We will give here first the process using as ancestral path elements in D^+ which are the ones directly given by as by the dynamics naturally and will only later introduce the ones marked with D^* using the knowledge of the D^+ case. First we discuss the test functions for path-valued marks, then the dynamics and operator.

Here we have to begin by introducing the test functions on V which is in our context a set of *E-valued càdlàg path*.

(i) Test functions Here we have the time-inhomogeneous and time-homogeneous case.

(a) The test function χ^k on the rhs. of (3.32) each evaluate the positions of the path at a tuple of time points and will be of the form of a product of functions on E which are applied to the position of the path at a specific time giving now a tableau of functions representing the χ^i of (3.32):

$$\chi^{\underline{\underline{\xi}}}(v) = \prod_{i=1}^{n} \prod_{k=1}^{m(i)} \chi_{i,k}^{\underline{\underline{\xi}}}\left(v(t_k^{(i)})\right) \ , v \in D(\mathbb{R}, E), \qquad (3.39)$$

for some $n \in \mathbb{N}$ and for $k = 1, \cdots, m(i)$ we have $0 \leq t_1^{(i)} < t_2^{(i)} < \ldots < t_{m(i)}^{(i)} < \infty$, with $i = 1, \cdots, n$. The $\chi_{i,k}^{\underline{\underline{\xi}}}$ are again indicators as in (3.33). Here $\underline{\xi}$ from above becomes $\underline{\underline{\xi}} = (\xi_k^i)_{\substack{i=1,\cdots,n \\ k=1,\cdots,m(i)}}$.

Note that $\chi^{\underline{\underline{\xi}}}$ above is *not* in $C_b(D(\mathbb{R}, E))$ as required in the basic setup of $\mathbb{U}^V$-valued martingale problems where we take usually functions from $C_b(\mathbb{U}^V, \mathbb{R})$ (fitting

the ones in the definition of the topology), but here we have functions being only in $b\mathcal{B}(\mathbb{U}^V, E)$). This technical point we have to tackle later in the proof section by giving either an equivalent martingale problem on continuous functions via a moving average of the present test functions or we have to use special properties of the path which are charged by the law.

Remark 3.8. One could try to work with occupation time integrals of the form:

$$\chi^{\underline{\xi},\underline{F}}(v) = \prod_{m=1}^{M} \chi_m(v) = \prod_{m=1}^{M} \int_{\mathbb{R}} \mathrm{d}s F_m(s)\mathbb{1}_{\xi_m}(v_s) \tag{3.40}$$

for $\xi_m \in E, F_m \in C^1(\mathbb{R})\cap L^1(\mathbb{R})$, $m = 1,\ldots,M$. This however does not fit together easily with the martingale problem for V which is the one which allows to access easily and directly important information about the f.d.d. of the ancestral path. We will use this idea in a different form namely once we work with truncations and we have to smoothen.

(b) For the *time homogeneous* set up of the path process (i.e. the time-space process) we consider

$$\widehat{\chi}^{\underline{\xi}}(t, v) = \prod_{i=1}^{n} \prod_{k=1}^{m(i)} \widehat{\chi}^{\underline{\xi}}_{i,k}\left(t, v(t \wedge t_k^{(i)})\right), \tag{3.41}$$

with $\widehat{\chi}^{\underline{\xi}}_{i,k}(t, v) = \Psi(t)\,\Psi^{\underline{\xi}}_{i,k}(t)\chi^{\underline{\xi}}_{i,k}(v)$ and $\Psi(t)$ and $\Psi^{\underline{\xi}}_{i,k} \in C_b^1(\mathbb{R}, \mathbb{R})$ as the functions in (3.31) to generate polynomials. Call the set of these polynomials $\widehat{\Pi}$.

We continue now the discussion of path-marked processes, with giving the dynamics.

(ii) Dynamics and operator Here we proceed as in the point (1) on $V = E$ except now that the operator of the mark evolution given in (3.35)-(3.37) has to be replaced by a new operator and in particular the test function χ from (3.41) instead of (3.33) has to be used.

Here we proceed in several steps. First introduce the operator which describes the change in the path process, then secondly based on this we define the operator of the mark evolution of the $\mathbb{U}^V$-valued process with $V = D(\mathbb{R}, E)$ acting on the polynomial Φ. The evolution of marks, i.e. of the ancestral path as time evolves is driven by an evolution of an element of $D(\mathbb{R}, E)$. This Markov process is called the *path process*. This process was introduced, for example in Section 12.2.2. in Dawson (1993). This evolution is time *inhomogeneous*.

The next step is to write down the generator of the path process. Since already the path process Y is time-inhomogeneous (recall the path is $\mathbb{R}-$indexed), we pass first on that level to the *time-space process* and need test functions on $\mathbb{R} \times E$ rather than just E.

Recall the generator of the motion process of a single individual $(Y(t))_{t\geq0}$ was called A. For the process Y the time-space process $(t, Y(t))_{t\geq0}$ then has generator $\widetilde{A} = \frac{\partial}{\partial t} + A$. The corresponding *path process generator* $\widehat{A}$ acts (see Section 12.2.2 in Dawson (1993)) on χ of the form (3.41) for $t_k \leq s < t_{k+1}$:

$$\widehat{A}\widehat{\chi}(s, v) = \prod_{\ell=1}^{k} \widehat{\chi}_\ell\left(s, v(s \wedge t_\ell)\right) \widetilde{A}\left(\prod_{\ell=k+1}^{m} \widehat{\chi}_\ell\left(s, v(s)\right)\right) \text{ and gives 0 for } s > t_m.$$

$$\tag{3.42}$$

This operator specifies a well-posed martingale problem on the spaces $D([0, \infty), \mathbb{R} \times D(\mathbb{R}, E))$ (Section 12.2.2 in Dawson, 1993). In a second step this has to be then lifted to an operator on the polynomials on $\mathbb{U}^V$, where it acts via the action of the path process generator on the function χ appearing in the polynomial.

Now turn to the $\mathbb{U}^D$-valued super random walk, the process $\mathfrak{U}^{*,\mathrm{anc}}$. In order to get a time-homogeneous Markov process we pass to the state space

$$\mathbb{R} \times \mathbb{U}^D \text{ and polynomials in } \widehat{\Pi} = \Pi_D, \tag{3.43}$$

which are products of a function $\psi \in C_b(\mathbb{R}, \mathbb{R})$ and a polynomial as in (3.31) with χ as in (3.41).

The operator $\Omega^{\uparrow,\mathrm{anc}}$ corresponding to the change of the marks now acts on the polynomial Φ as follows. Namely denoting this operator of the mark evolution of $\mathfrak{U}$ by $\Omega^{\uparrow,\mathrm{anc}}$ we have for each sampled marked individual the action of the path process generator $\widehat{A}$ but now acting on the corresponding factor $\widehat{\chi}^k$:

$$\Omega^{\uparrow,\mathrm{anc}}\Phi^{\phi,\widehat{\chi}} = \sum_{k=1}^{n} \Phi^{\phi, A_k^* \widehat{\chi}}, \text{ where } A_k^* \widehat{\chi} =: \Big(\prod_{\substack{\ell=1 \\ \ell \neq k}}^{n} \widehat{\chi}^{\ell} \Big) \widehat{A}\widehat{\chi}^k. \tag{3.44}$$

Taking the sum of the *mark* operator $\Omega^{\uparrow,\mathrm{anc}}$ and the adapted and *lifted* (from $\mathbb{U}$ and then from $\mathbb{U}^V$ to $\mathbb{R} \times \mathbb{U}^D$) *growth* and the *resampling* operator from point 1 then gives immediately

$$(\Omega^{*,\mathrm{anc}}, \widehat{\Pi}). \tag{3.45}$$

We now have to specify precisely the possible *initial states* for starting times s for our martingale problem. We choose here the subset of $\mathbb{U}^D$ which is characterized by a further restriction namely having marks only on a set $\widetilde{V}$ (i.e. $supp(\mu) \subseteq U \times \widetilde{V}$):

$$\widetilde{D} = \bigcup_{-\infty < s \leq t < \infty} \{t\} \times D_{s,t} \text{ or even } \bigcup_{s \in \mathbb{R}} \{s\} \times D_{s,s}. \tag{3.46}$$

The corresponding subset of $\mathbb{U}^{\widetilde{D}} \cap \mathcal{E}$ is then called $\widetilde{\mathcal{E}}$.

We have to show now that the martingale problem on the set of test functions above is *well-posed*.

Proposition 3.9 (Existence and uniqueness of the $\mathbb{U}^D$-valued super random walk)**.** *The $(\widehat{\Omega}^{*,\mathrm{anc}}, \widehat{\prod}, \mathfrak{u})$-martingale problem is well-posed.*

Here some technical points have to be addressed in case of an *infinite* geographic space in particular the potentially infinite total mass of the measure ν on $U \times D$. We address this in Section 7 in more detail as well as the martingale problem establishing wellposedness. We now have made sense of what we mean by the ancestral path marked process.

Definition 3.10 ($\mathbb{U}^D$-valued super random walk)**.** The solution of the well-posed $(\Omega^{*,\mathrm{anc}}, \widehat{\Pi}, \mathfrak{u})$-martingale problem, the $\mathbb{U}^D$-valued super random walk, is denoted

$$(\mathfrak{U}_t^{*,\mathrm{anc}})_{t \geq 0}. \tag{3.47}$$

Remark 3.11. We note that this process is different from the E-marked ultrametric Feller process who records genealogy and *current* position, which is a Markovian functional of $\mathfrak{U}^{*,\mathrm{anc}}$. We shall see later how we can verify for this process as well the generalized branching property, similarly for the historical process. This will in

particular show how these various processes and their $(T_t)_{t\geq 0}, (S_t)_{t\geq 0}$ and $(\sqcup^t)_{t\geq 0}$ are related.

(3) Pathmarked case: a sufficient variant of the standard state description via D^* We have to specify state space, test functions, operators and processes.

(i) State space   In this setup for the version we formulate next we can then show the *generalized branching property* since we use a more suitable concept of ancestral path which we introduce next. Namely we have to *modify* the description of the genealogy and mark a bit further to fit our framework of Assumption 2.1 and of the Theorem 2.11 to be applied here, but in which we are *keeping the really wanted information described above*. Recall Theorem 2.11 requires a *time-homogeneous* process and a semigroup setting.

The idea is that the ancestral path describes the situation looking back from the presently living individual and hence in particular positions *relative* to this current position give the interesting informations. We define therefore the set of path providing the desired information as follows:

$$D_{-t,0}(\mathbb{R}, E) = \{v \in D(\mathbb{R}, E) | v(s) = v(-t) \, \forall s < -t, \, v(u) = v(0) \, \forall u > 0\}, \, t \in [0, \infty] \tag{3.48}$$

and

$$\widehat{D}^* = \bigcup_{t \in [0,\infty)} D_{-t,0} \tag{3.49}$$

as paths which are constant before a specific time $-t$ and after time 0. To obtain a Polish mark space we have of course to take

$$D^* \text{ as the closure of } \widehat{D}^*. \tag{3.50}$$

The last space D^* defines now the *mark space* of interest and the resulting state space for the generalized branching property is

$$\mathbb{U}^{D^*}. \tag{3.51}$$

(ii) Test functions   The $V = D^*$-set up is handled as follows. This process arises for us later on as a functional of the process in (2) via a map $\mathfrak{R}$ in (3.55) and (3.56). In order to specify a martingale problem for this process which is again Markovian we have to begin by introducing the test functions on V which is in our context an *E-valued càdlàg path* constant after time 0 and before time $-t$. This means we get functions as (3.39) but with $t_1^{(i)} < t_2^{(i)} < \ldots < t_{m(i)}^{(i)} \leq 0, i = 1, \ldots, n$. The test function χ^k on the rhs. of (3.32) each evaluate the positions of the path at a tuple of time points and will be of the form of a product of functions on E which are applied to the $(-t)$-shifted version of the function $\chi^{\underline{\xi}}$ from above at time t, which means that we work with a function $\widetilde{\chi}^{\underline{\xi}}$ evaluating $D_{-\infty,0}$-functions at fixed times and locations, independent of the current time t as in (3.39).

(iii) The D^*-marked genealogical super random mark and the Ω^{shift} operator   From the process $\mathfrak{U}^{*,\mathrm{anc}}$ we obtain now another process. We observe that relevant for us is the information on the path viewed from the present position which we can identify with a path which moves only between time $-t$ and 0. Therefore we apply the $(-t)$-shift at time t to the marks, which are the ancestral path. The new path we call the *adjusted* path. This gives uniquely a new process for which we obtain the *well-posed martingale problem* (see Section 7) by adding the operator

corresponding to the $(-t)$-shift of the ancestral path, which then allows us to apply our theory. This additional operator we have to specify below. In addition we act now with the operator $\Omega^{\uparrow,\mathrm{anc}}$, acting on the marks on elements in $D_{-t,0}$ at time t only and call this restriction of $\Omega^{\uparrow,\mathrm{anc}}$ now $\Omega^{\uparrow,\mathrm{anc-sh}}$.

Observation: The interesting information on the path sits at time T of evolution starting in an element $\{s\} \times D_{s,s}$ in the piece $(T > t \geq s)$ i.e. a path in the time interval $[s, T]$.

For our purpose revealing the generalized branching property the following fact is important. Using the setup for the path space D^* we can formulate a *time-homogeneous* dynamics. This is different from the classical historical process as defined in Chapter 12 of Dawson (1993). We modify (3.42) by putting $s = 0$ and adding at time t in (3.36) the generator of the path shift by $-t$.

Before we can define the generator of our process we have to calculate the effect *of the shift*. Observe that the underlying path is that of a jump process with generator as in (3.7) which means that the path are piecewise constant with finitely many jumps in finite time intervals. At each jump time of the path coinciding with a time of evaluation of the path by our function we may get a contribution to the generator since the jumps then leads to a jump of the evaluation functional of the path as we shift. Such a situation does occur with probabilities we can control. Proceed as follows.

Calculation. Consider now a polynomial $\Phi^{\phi,\chi}$ where χ is as in (3.39). Let $-t_m^{(i)} < -t_{m-1}^{(i)} < \cdots < -t_1^{(i)} \leq 0$ be the times where such a path is evaluated for the n paths labelled $i = 1, \cdots, n$ and which is tested whether it is in points $\xi_k^i, k = 1, \cdots, m(i)$ at the times $t_k^{(i)}$. Furthermore consider the *jump times* of the n paths which we denote by $(s_k^{(i)})_{k=1,\cdots,m(i)}$ ordered from the left to the right in k, where the path jumps from the point ζ_k^i to $\zeta_k^{i,+}$.

A contribution arises if $\xi_k^i \in \{\zeta_k^{i,+}, \zeta_k^i\}$ since then in the concerned factor by a small shift a jump from 1 to 0 or 0 to 1 occurs and is then causing a change in the complete product of $+1, 0$, or -1 respectively. Note that this effect occurs along the whole path. The corresponding operator is the jump generator of the switch of $\chi_{i,k}$ to 0 resp. 1, if a jump from ξ_k^i away resp. into that point from outside at the rates given below in (3.52).

We have next to obtain the probability for this possibility to occur due to a small time shift Δt, which allows for one jump at order Δt probability at each of the possible time point $i = 1, \cdots, m(i)$ of one path. However the intensity for a jump from ξ_k^i *away* respectively *into* at time $t_k^{(i)}$ for some $i \in \{1, \cdots, m(i)\}$ is given by

$$\sum_{\xi \in \Omega \setminus \xi_i^k} a(\xi_k^i, \xi) \text{ respectively } \sum_{\xi \in \Omega} a(\xi, \xi_k^i) \tag{3.52}$$

at both expressions summed over $k \in \{1, \cdots, m(i)\}$. This can happen in one for each of the n path (corresponding to the index i).

We get therefore as additional generator term the expression:

$$\Omega_t^{\mathrm{shift}} \Phi^{\phi,\chi} = \sum_{i=1}^n \Phi^{\phi, A_i^{\mathrm{cum}}\chi}, \text{ with } A_i^{\mathrm{cum}}\chi_{i,k} = \sum_{k=1}^{m(i)} \widetilde{A}_k \chi_{i,k}, \tag{3.53}$$

where $\widetilde{A}_k$ acts on the k-th component of the samples. Note that the operator on the rhs. does in fact not depend on t.

Now add $\Omega^{\uparrow,\mathrm{bran}}, \Omega^{\uparrow,\mathrm{grow}}, \Omega^{\uparrow,\mathrm{anc-sh}}$ and Ω^{shift} to get $\overset{\leftarrow*,\mathrm{anc}}{\Omega}$.

(iii) Dynamics   This allows us now to define a new process as follows.

Definition 3.12 ($\mathbb{U}^{D^*}$-valued super random walk)**.** We call the solution of this well-posed $(\overset{\leftarrow*,\mathrm{anc}}{\Omega}, \widehat{\Pi}^{\leftarrow})$- martingale problem the *adjusted ancestral path-marked genealogy-valued super random walk* a process denoted

$$\overset{\leftarrow*,\mathrm{anc}}{\mathfrak{U}}, \tag{3.54}$$

with values in $\mathbb{U}^{D^*}$ as in (3.51).

What is the *relation* between states in $\mathbb{U}^{D^*}$ and $\mathbb{U}^D$? Why is the martingale problem on $\mathbb{U}^{D^*}$ wellposed? The key point is now that we have a collection of $1-1$ maps $\mathfrak{R}_t, t \geq 0$ between our process of interest with value in $\mathbb{U}^V$ the one with $V = D$ and the one with marks $V = D^*$. Namely for the path-valued process at time t, starting at time s with $t \geq s, t \geq 0$, we shift to the left all the paths by t in the time coordinate and we obtain *a time-homogeneous process with marks in D^**: i.e. for a functional of our process at some specific time t we have a map

$$\mathfrak{R}_t \left([U \times V, r \otimes r_V, \nu]\right) = [(U \times D^*, r \otimes r_{D^*}, \nu^*)] \tag{3.55}$$

induced by

$$(u, v) \to (u, v^*); \ v^*(r) = v(r - t), \ r \in \mathbb{R} \text{ and } v \text{ the path forward associated with } u. \tag{3.56}$$

This collection of mappings define on the *paths* $(\mathfrak{R}_t\mathfrak{U}_t)_{t\geq 0}$ of the stochastic process a map $\mathfrak{R}$ and it will be

$$\text{the process } \mathfrak{R}(\mathfrak{U}), \tag{3.57}$$

to which we apply our theorem on the generalized branching property. *However since this is a bijection we do not loose any information we coded in the state description initially, but have it in a technically more convenient form.*

The $\mathbb{U}^{D^*}$-valued process arises as a functional (namely (3.57)), which also solves a well-posed martingale problem in its own right. Here the wellposedness follows from the fact that $\mathfrak{R}$ is one-to-one and the wellposedness of the basic martingale problem on D^+ mentioned above.

From the construction of the process (since the martingale problem is well-posed) we have

$$\overset{\leftarrow*,\mathrm{anc}}{\mathfrak{U}} = \mathfrak{R}(\mathfrak{U}^{*,\mathrm{anc}}). \tag{3.58}$$

Remark 3.13. With the choice of test functions as in (3.40) we would get the generator actions: replace "$A\chi^{\xi^k}$" by $\tilde{A}\chi^{\underline{\xi}^k}$ where this new operator $\tilde{A}$ acts as follows on one test function as in (3.40):

$$\tilde{A}\chi^{\underline{\xi},\underline{F}}(v) = \sum_{m=1}^{M} \frac{\chi(v)}{\chi_m(v)} \left[-\int_{-\infty}^{0} \partial_s F_m(s) \mathbb{1}_{\xi_m}(u(s)) \, \mathrm{d}s + F_m(0)\mathbb{1}_{\xi_m})(u(0)) \right. \tag{3.59}$$

$$\left. + \int_{0}^{\infty} F_m(s) \, dx s \, A(x \mapsto \mathbb{1}_{\xi_m)(x)}|_{x=u(0)}) \right].$$

Step 3: The framework: Truncation, concatenation, subfamily decomposition and generalized branching property on the spatial level. We have to define now the *concatenation* and the *truncation* for V-*marked* ultrametric measure spaces, denoted $\mathbb{U}^V$ for complete separable metric V space (recall (3.21) and sequel). This means we have to extend the concepts in (3.3) - (3.15) from the previous subsection now to the *marked* case. Here we have to distinguish the marked case with marks being the current location from the one marked with ancestral path.

Case 1. Look first at $V = E$. Let ν_U denote the projection of ν on U. We first define $\mathbb{U}^V(h)$ as the elements $[U \times V, r, \nu]$ of $\mathbb{U}^V$ with (U, r) having ν_U-essential diameter strictly less than h, similarly $(\mathbb{U}^V(h))^{\sqcup}$ with less than or equal to h. In order to form a *concatenation* of $(\mathbb{U}^V(h))^{\sqcup}$-elements $\mathfrak{u}_1 = [U_1 \times V, r_{U_1} \otimes r_V, \nu_1]$ and $\mathfrak{u}_2 = [U_2 \times V, r_{U_2} \otimes r_V, \nu_2]$ define, recall $\sqcup$ abbreviates $\sqcup^h$, and (3.5)-(3.7)) for $r_{U_1 \sqcup U_2}$:

$$\mathfrak{u}_1 \sqcup \mathfrak{u}_2 = [(U_1 \uplus U_2) \times V, (r_{U_1 \sqcup U_2}) \otimes r_V, \widetilde{\nu}_1 + \widetilde{\nu}_2] \tag{3.60}$$

where $\widetilde{\nu}_1(A) = \nu_1(A \cap (U_1 \times V)), \widetilde{\nu}_2(A) = \nu_2(A \cap (U_2 \times V))$ for $A \in \mathcal{B}((U_1 \uplus U_2) \times V)$.

Next we need the *truncation*. The truncation affects only the distances and acts as before, hence we get again by lifting the operation from U to $U \times V$:

$$\lfloor \mathfrak{u} \rfloor(h), \mathfrak{u} \sqcup^h \mathfrak{u}', T_t(\mathfrak{u}). \tag{3.61}$$

Case 2. In the second case $V = D(\mathbb{R}, E)$ (respectively $\mathbb{R} \times D(\mathbb{R}, E)$ as we will pass later to the time-homogeneous formulation) the marks contain themselves some *information from the past* in particular they contain information about genealogies. We therefore have to extend the h-truncation to the marks so that the marks contain only the information about the ancestral path for some time h back analogue to the genealogy which we include till depth h.
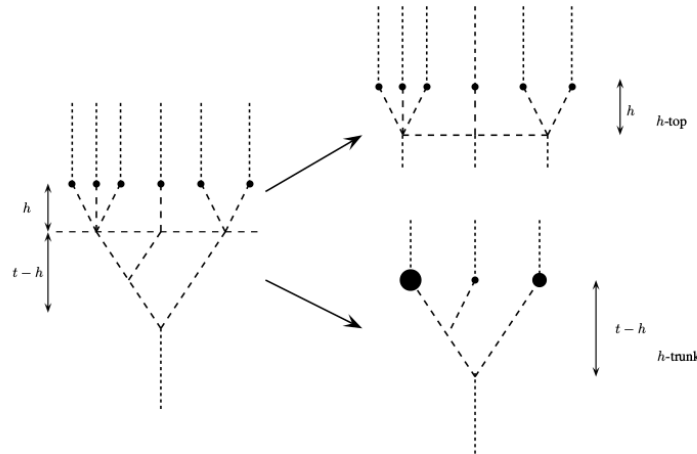

FIGURE 3.1. Example of path marked h-top and h-trunk, cut at height $t - h$. The $\cdots$ mark the constant parts of the path and $-$ the parts where different values might be assumed!

Fix a time horizon T. We define a collection of mark spaces $(V_t)_{t>0}$ with $V_t \subseteq V_s \subseteq V$ if $T \geq t > s$ where for a process evolved till time T we consider the path fluctuating only between time t and T:

$$V_t^T = \{v \in D(\mathbb{R}, E), \ v(u) = v(0), u < 0; \ v(u) = v(t), u \leq t; v(u) = v(T), u \geq T\},$$
$$V_t := V_t^\infty.$$
$$(3.62)$$

In our time-homogeneous system with marks in $\mathbb{R} \times V$, the first mark component is preserved and we have to act only on the second, then where at time T we have ancestral path constant before time 0 and after time T, we would *truncate the mark* as follows:

Set for $u \in \mathbb{R}$, $T \geq t \geq 0$,

$$T_t^V : V_0^T \longrightarrow V_t^T, \ (T_t^V(v))(u) = v(u)1_{[t,\infty)}(u) + v(t)1_{[u \leq t]}(u). \quad (3.63)$$

Note that truncation means here *truncations of the fluctuations* not the path as such and we look from the bottom up rather than from the top down keeping only randomness in the path beyond time t.

Next define the $h-truncated$ $V-$marked objects. Begin with the $t-$ truncation map T_t^V. The truncation map T_t^V of $[U \times V, r \otimes r_V, \nu]$ is defined considering first a map T_t^U on the genealogical part $[U, r, \mu]$. Define T_t^U acting on $[U, r, \mu]$ as before T_t. The resulting space we have to equip with the truncated marks, i.e. each point in U now is *marked* with $T_t^V(v)$ instead of $v \in V$. Finally we pass to the image measure of ν under this combined mapping on $[U \times V, r \otimes r_V, \nu]$.

We define $\mathbb{U}^V(h)$ as a marked ultra-metric measure space where all distances are strictly less than h and the mark kernel κ satisfies

$$\kappa(u, V \setminus V_{T-h}^T) = 0 \ , \ \forall \ u \in U. \quad (3.64)$$

Next $\mathbb{U}^V(h)^\sqcup$ has distances $\leq h$ and the mark kernel κ satisfies again (3.64).

Step 4: The setup for the generator criterion and results on the spatial level. The next task is to apply our criterion which is without problems if $V = E$ but the situation is more subtle for $V = D(\mathbb{R}, E)$.

The result for E-marked super random walk. This will be a consequence of our more general (and more complicated to formulate) result on ancestral path marked $\mathbb{U}^V$-valued super random walk, see Corollary 3.19, later on for that.

Theorem 3.14 ($\mathbb{U}^E$-valued super random walk)**.** *The process $\mathfrak{U}^{*,E}$ has the generalized branching property.*

Result for path-valued process. We now treat $V = D^+$ and $V = D^*$, which needs more efforts.

(1) The path-valued results: Introduction The situation is more subtle in the path-valued situation on which we have to focus here next. We first explain the idea.

Remark 3.15 (Basic idea of truncation with marks in D^*). The needed truncation can be understood best as follows using a measure $\mathbb{R}$-tree representation of the state in $\mathbb{U}^{D^*}$ even though in our proofs this viewpoint *is not used*.

Recall that for every *ultrametric measure space* of diameter $2t$ there is a unique (diameter $2t$) $\mathbb{R}-tree$ such that the points in the ultrametric space (better the

support of the sampling measure) are the *leafs* of the $\mathbb{R}-$tree. Consider this repre-
sentations for the states of our evolving genealogy. If in addition these $\mathbb{R}-$trees, see
Example 2.2 for explanation and more references, for varying t are all embedded in
an weighted $\mathbb{R}$-tree of all individuals alive at some time before time t, a property
of the dynamics (which we have in our case and which we will discuss more in
Depperschmidt and Greven, 2019 and we sketch some of it in part (ii) of the next
subsection) allowing that we can define a *E-valued ancestral path* in this object.
Then the ancestral path can be used to mark the points in the $\mathbb{R}-$tree such that
the *E*-marked *geodesic* between founding father and the current individuals of a
maximal subfamily (i.e. a current leaf descending from the founding father) is a
copy of the ancestral path. Then the point on the geodesic starting from a leaf
carries the current position of the ancestor at depth h in distance h of the tagged
leaf.

In this picture based as the associated weighted $\mathbb{R}$-tree we want for h-truncation
to *cut the tree at depth $s = t - h$* and then use the *cut marked geodesic* as the
ancestral path we associate with the leaf in the truncated state. This leads to a
replacement of the ancestral path by one which we continue beyond the piece *back
time h constant*, so that we still have an $\mathbb{R}$-*indexed path* but no information from
the past before time $t - h$ is retained. Note that this means introducing the one
root at depth h via the truncated and now depth-h $\mathbb{R}-$tree and at this root we have
a mark-kernel (even if we had before a mark function). Note that this new path if
evaluated at time positions gives different numbers then the one where we *restrict*
to path which are constant before time $t - h$. Note that the latter procedure would
include information on the past before time $t - h$.

We may go further and shift the obtained path by t to the left to obtain for
every t always a path from $D_{-\infty,0}$ which is in fact for the used initial state in $D_{-t,0}$
then for $t > 0$ also in $D_{-t,0}$ and after truncation in $D_{-h,0}$ and is for all $t \geq h \geq 0$
independent of t. This fact is the reason why we get for $\mathfrak{R}(\mathfrak{U})$ the generalized
branching property.

(2) A problem in the case of path-valued marks and its solution   This
construction above however does not quite fit our setup in Theorem 2.11 yet since
we now do not have images under truncation in the right set *since in* (3.62)-(3.64)
we have still T around (but recall our $\mathbb{R}-$indexed path have started evolving at time
0 up to the current time T). This time dependence we have to remove. However we
have already introduced a system with *reduced information* namely $\overset{\leftarrow *,\mathrm{anc}}{\mathfrak{U}}$, since we
are interested in the ancestral path looked backward from the present time. This
means if we code the present time in the state we are only interested in the element
in D^* we get! We set therefore for $t > 0$:

$$V_t = D_t^* = D_{-t,0} \subseteq D^*. \tag{3.65}$$

Remark 3.16. We may use as marks now D^* rather than D^+ but we note that we do
not loose information this way if we *know* that we start in initial conditions at time
0 as specified, namely constant before time 0, since then by piecing that constant
piece together with the piece from D^* shifted by t beyond time 0 we reconstruct
uniquely the original state at time t.

We saw above that we may pass via the map $\mathfrak{R}$ to the mark space $V = D^*$ if
we take the time-space process. This induces also a map on the truncated objects

where we now get with t-truncations elements in $V_t^* = D_t^*$. The same relations hold then for the truncated marked ultrametric measure spaces induced by D_t^*, D^* as we required for S_t and S.

Then we can now define the full truncation:

$$T_t^* ((U \times V, r_U \otimes r_V, \nu)) = (S_t \times D_t^*, (r_U \wedge 2t) \otimes r_{V^*}, (id \otimes T_t^{D_t^*})_*(\nu)), \quad (3.66)$$

which induces the map on the of the D^*-marked ultrametric spaces as well, *correspondingly the S_t are defined by*:

$$\left((\mathbb{U}(t))^{\sqcup}\right)^{D_t^*}, \ t \geq 0. \quad (3.67)$$

The $h-$concatenation of two D^*-marked forests is now defined by:

$$\text{in (3.60) we replace } V \text{ by } D^*. \quad (3.68)$$

This means for the corresponding set of polynomials we obtain the new elements

$$\widehat{\prod} \text{ and } \widehat{\prod}^{\leftarrow} \quad (3.69)$$

and in order to get the *corresponding* elements we replace χ by χ^* where we specify the time points where we evaluate path now in points of the left half axis. These objects again satisfy the conditions of Assumption 2.1.

Summarizing we have the setup of our Theorem 2.11 The semigroups $(S_h, \sqcup^h)$, $h \geq 0$, of Assumption 2.1 are as follows:

$$S = \mathbb{U}^{D^*}, \ S_h = \mathbb{U}^{D_h^*}. \quad (3.70)$$

For the definition of the mark spaces see (3.50).

Truncation is defined via the pull-back mappings. Therefore let $\mathfrak{u} = [U \times D^*, r \otimes r_V, \nu] \in \mathbb{U}^{V^*}$ and $h \geq 0$. Define the truncated space $T_h \mathfrak{u}$ as

$$T_h^* \mathfrak{u} = [U \times D^*, (r \wedge 2h) \otimes r_V, (id \otimes T_h^*)_* \nu], \quad (3.71)$$

where $T_h^* : D^* \mapsto D_t^*$ via

$$T_h^* v = \begin{cases} v(-t), & s < -t, \\ v(s), & s \geq -t. \end{cases} \quad (3.72)$$

Concatenation follows the same definition as (3.60).

(3) Results on (Adjusted path-)marked genealogical super random walk Having completed these preparations we state a key fact about the martingale problem. Recall from above truncation and $h-$concatenation. The truncation T_h at depth h now cuts the genealogical distance at $2h$ and the mark, i.e. the path is set constant up to a piece of length h and lies in $D_{-h,0}$. Therefore the map T_h now cuts of the h-top of the E-marked weighted $\mathbb{R}$-tree associated with $\mathfrak{U}$ and if we work with the process $(\mathfrak{U}_t^{\leftarrow *, \text{anc}})_{t \geq 0}$ our approach fits and Theorem 2.11 will apply and give the following.

Theorem 3.17 (Functional of $\mathfrak{U}_t^{*, \text{anc}}$ has generalised branching property). *The adjusted ancestral path marked $\mathbb{U}^{D^*}$-valued super random walk satisfies:*

$$\left(\left(\mathfrak{U}_t^{\leftarrow *, \text{anc}}\right)\right)_{t \geq 0} \text{ has the generalized branching property.} \quad (3.73)$$

If we make a *strong assumption* we can obtain the branching property of a functional.

Corollary 3.18 (Generalization). *Suppose we have a separable, complete, metric space E for which our process is solution of a well-posed martingale problem on $\mathbb{U}^V$ with $V = E$ as in* (3.25), *then the above theorem holds as well.*

Since it is known form the literature that the assumption above holds we get:

Corollary 3.19 ($\mathbb{U}^E$-valued super random walk). *The process* $(\mathfrak{U}_t^E)_{t \geq 0}$ *has the generalized branching property.*

There is another case where we know that the strong assumption of the well-posedness of the martingale problem through existing deep work (Dawson and Perkins, 1991) and we obtain:

Corollary 3.20 (Historical process of super random walk). *The historical process associated with* (3.29) *has the property that* $(\nu_t^*)_{t \geq 0}$ *of* (3.20) *satisfies that* $(\mathfrak{R}_t(\nu_t^*))_{t \geq 0}$, *the historical process of adjusted path, has the generalized branching property.*

Why does this all follow from our Theorem 3.17 ? The answer comes in two remarks.

Remark 3.21. (a) For the *E-marked* genealogy-valued process where we record only the present location , i.e. $V = E$ the corresponding truncation map will *not* change the mark. The concatenation operation is now as before (with the different V). A measurable function of $\mathfrak{U}^{*,\mathrm{anc}}$ projecting the mark on the value of the path at time t gives us a process which is a Markov process in its own right solving a well-posed martingale problem with the operator in (3.37), thus having the branching property. In other words, we may use Theorem 2.11 since those multiplicative functions we need to check for the generator criterion were already considered by the ancestral path marked genealogy-valued super random walk.

(b) The *historical process* is a functional of the V- marked genealogy, where now $V = D^*(\mathbb{R}, E)$ i.e. we map

$$[U \times V, r, \mu] \longrightarrow (\pi_V)_* \mu \tag{3.74}$$

where π_V is the projection from $U \times V \to V$. This functional is again a Markov process, namely a *modification* of the one known as the *historical process* introduced in Dawson and Perkins (1991) but now using *adjusted* path i.e. the *state space* D^*.

However there is a unique lifting to a $V-$marked case by shifting by t the path which are constant path before time 0, i.e. which "start" at 0 and end fluctuating at the current time t. For the historical process we will have a truncation map which is the map induced by the replacement of path with the path which is *before depth h kept constant* and the concatenation $\sqcup^h$ is the sum of the measures. Hence this process inherits the generalized branching property for these choices of truncation and concatenation.

Remark 3.22 (Infinite divisibility and Lévy -Khintchine formula). We can now use the result we obtained for the process $\overleftarrow{\mathfrak{U}}^{*,\mathrm{anc}}$ to show with results from Glöde et al. (2019) that the process has *infinitely divisible* marginal distributions if we start in a fixed initial state. In particular do we then have a *Lévy -Khintchine presentation for the marginals* and with it an *inhomogeneous Poisson point process representation* of the state at time t. For that we can generalize the definition and the proofs of the Theorem 1.37,2.44, Corollary 1.40 in Glöde et al. (2019) to the

object $[(S_t, \sqcup^t)_{t \geq 0}, (T_t)_{t \geq 0}]$ immediately, since we use only what we have postulated as Assumption 2.1 and 2.6. This allows us to apply this representation in our path-marked model and to study the subpopulations defined as individuals in maximal distance say h for $h \in (0, t]$. We can ask for the number of such families and their structure, see here also Depperschmidt and Greven (2019) for details on these objects, questions and results.

3.3. *Outlook and perspectives: open questions for genealogies.* We discuss here three directions of extensions for genealogical processes with values in equivalence classes of marked metric measure spaces which should be studied and resolved in the future:

- genealogies including the *fossils*,
- offspring laws with *fat tails*,
- genealogical processes in *continuum geographical space*.

The first point can be handled based on some work in progress the two others are open problem here some approaches we investigated but nothing complete exists so far.

(i) The genealogy including fossils and CRT

Another type of extension would be to consider the genealogy of all the individuals *ever* alive before the current time t. This is the population including the *fossils*. Then with t running through $(0, \infty)$ we obtain an evolving genealogy. This will be a subset of the equivalence classes of treelike metric measure spaces. This gives us an $\mathbb{M}$-valued stochastic process. Here again we have the (generalized) branching property, via the criterion. In fact we could consider $t \to \infty$ and take the complete genealogical tree of all individuals ever alive if the population becomes extinct we obtain a limit, see Depperschmidt and Greven (2019). This object has values in $\mathbb{M}$ the space of equivalence classes of metric measure spaces. For the Feller branching dynamic the equivalence class in $\mathbb{M}$ has a representation which is known as the CRT, for the latter see Aldous (1990, 1991a,b, 1993), Le Gall (1993).

In that case we work with metric measure spaces of a specific form instead of *ultra*metric ones, which requires some *new* elements. The topology on the state space has been treated in this general form, but there is the issue of the dynamic. First of all the corresponding processes have to be constructed with well-posed martingale problems and then the concatenation and truncation structures have to be introduced and then the criterion has to be checked.

The first point is treated in work in progress Greven et al. (2020) the state space is contained in the *rooted marked metric measure spaces*, for the state at time t with distance at most distance t from the root denoted $\mathbb{M}_t$ and we comment here on the second point.

The state of time t is in $\mathbb{M}_t$. Now the $h-$truncation on $\mathbb{M}_t$ removes all points in distance less than $t - h$ from the root, and truncates distances at $2h$. The S_h consist of subspaces with points which are in distance at least $t - h$ from the root and have at most the distance h. Accordingly $M_t(h)^{\sqcup}$ are now the metric measure spaces with a root, points at most in distance t from the root but at least $t - h$ from the root and other distances at most $2h$, analog only further than $t - h$ from the root and of the distances less than $2h$ for $\mathbb{M}_t(h)$. The $h-$ concatenation in $\mathbb{M}_t(h)^{\sqcup}$ is defined as before. For the dynamic the branching property follows from the criterion. This induces a generalized branching property on the $t \to \infty$ limit the $\mathbb{M}$-valued version of the CRT.

(ii) Branching with more general offspring distribution

A natural question is how we can treat the branching processes where we have as basis branching processes for an offspring distribution without higher than first moments.

In this case the total mass process does not have anymore all moments and we can not work as before with a martingale problem where the test functions are monomials, which was a key point in the proof of the previous results in Depperschmidt and Greven (2019), so that we are lacking at the moment a characterization by a well-posed martingale problem. Here we now have to work with *local* martingale problems or work with Laplace functionals.

Another point is that now the resampling operator has to be replaced by an operator where in a sample not only distances to one point in the sample change but a whole random set of individuals arises now from one ancestor, the Kingman coalescence mechanism in the dual is replaced by a $\Lambda-$coalescent type dual transition which would have to be identified, since in general this cannot be just a Λ-coalescent. This means it requires *very substantial work* to rigorously construct the $\mathbb{U}$-valued process via a well-posed martingale problem, even though no principal problems seem in the way.

(iii) Genealogical and historical processes in continuum space

In the literature (Dawson, 1977, 1993) one studies the *continuum space* analogue and limit of the super random walk respectively its historical process version the so called *(historical) Dawson-Watanabe process*. Here one would like to proceed similarly and introduce the *genealogy valued version* of these processes. Here some problems arise at the starting point, namely to establish the wellposedness of the martingale problem. To show uniqueness the most powerful tool is duality respectively Feynman-Kac duality as presented in Section of Ethier and Kurtz (1986), which works with our approach only in the case of *strongly recurrent migration*, for example in $d = 1$. Only in that case we can apply the technique of the *Feynman-Kac duality* (a duality where an exponential functinal, as in equation (5.40) or (7.17)-(7.20) appears in the dual expectation, see here Depperschmidt and Greven (2019) for details) to obtain the needed uniqueness of the solutions of the martingale problem. Furthermore due to the fact that the continuum space limits for the occupation measures are not given via SPDE's since the states are singular measures the existence problem is also more subtle since we cannot work with conditional dualities so easily (see Depperschmidt and Greven, 2019 or Dawson and Greven, 2003 for this concept).

In $d \geq 2$ the necessary uniform integrability in passing to the continuum space limit fails and the limit expression makes no sense since two path of the migration do not have a joint occupation time. This is related to the fact that the population is now supported by a set of Hausdorff dimension less than d (for $d = 2$, the picture being a bit more subtle) and here one would have to work with an *approximate duality*. Hence we need here a different approach which remains to be developed. However we can obtain from the historical Dawson-Watanabe process as a functional an $\mathbb{U}^V$-valued process which then has the generalized branching property. However it should be possible to get then directly via the $\mathbb{U}^V$-valued martingale problem. This is a general problem with continuum spaces and is addressed in forthcoming work in Greven et al. (2020).

Outline of the proof section. In Section 4 we prove the criterion and then prove in Section 6 that the example from Section 3.1 fits in our framework and satisfies the criterion; this proof is based on some key facts derived beforehand in Section 5.1- 5.3. In Section 7 we give the extensions of the proofs to the spatial models.

4. Proof of basic criterion: Theorem 2.11

We prove separately the two parts of the Theorem 2.11.

ad (1)**:** We saw in the introduction that the generalized branching property implies that the relation (2.8) holds so we need only the other direction.

Let $t > 0$ and $x = x_1 \sqcup^s x_2 \in S$. Let $(X_t^{x_i}, t)_{t \geq 0}$ be solutions to the martingale problem for $(\tilde{A}, \tilde{D}, \delta_{(0,x_i)})$, $i = 1, 2$ and $X^{x_1} \perp\!\!\!\perp X^{x_2}$. For $f \in \tilde{D}$, i.e. $f(s, x) = \psi(s)h_s(x)$, $s \geq 0$, $x \in S$ we will show that

$$\left(\psi(t)h_t(X_t^{x_1} \sqcup^t X_t^{x_2}) - \psi(0)h_0(X_0^{x_1} \sqcup^s X_0^{x_2}) - \int_0^t \tilde{A}\psi(r)h_r(X_r^{x_1} \sqcup X_r^{x_2}) \, dr \right)_{t \geq 0}$$

$$(4.1)$$

defines a martingale and, thus, $(X_t^{x_1} \sqcup^t X_t^{x_2}, t)_{t \geq 0}$ is a càdlàg solution to the $(\tilde{A}, \tilde{D}, \delta_{(x_1 \sqcup x_2, 0)})$ martingale problem. Due to the uniqueness assumption for the martingale problem, this implies $T_t X_t^{x_1 \sqcup x_2} \overset{d}{=} T_t(X_t^{x_1} \sqcup^t X_t^{x_2})$. Then, we get the branching property for any $t \geq 0$, $x_1, x_2 \in S$:

$$P_t(x_1 \sqcup^s x_2, h_t) \overset{\text{def}}{=} \mathbb{E}[h_t(X_t^{x_1 \sqcup^s x_2})] = \mathbb{E}[h_t(X_t^{x_1} \sqcup^t X_t^{x_2})]$$

$$\overset{\text{def}}{=} \left(P_t(x_1, \cdot) *^t P_t(x_2, \cdot) \right)(h_t), h_t \in D_t. \qquad (4.2)$$

It remains therefore verify to (4.1). First set $\psi \equiv 1$. After the argument it will become clear how to generalize. The proof follows that of Lemma 4.3.4 in Ethier and Kurtz (1986): By independence of the two processes for $t_2 > t_1 \geq 0$:

$$\mathbb{E}\left[\left(h_{t_2}(X_{t_2}^{x_1}) - h_{t_1}(X_{t_1}^{x_1}) - \int_{t_1}^{t_2} \tilde{A}h_r(X_r^{(x_1)}) \, dr \right) h_{t_2}(X_{t_2}^{x_2}) | \mathcal{F}_{t_1} \right] = 0 \qquad (4.3)$$

$$\mathbb{E}\left[\left(h_{t_2}(X_{t_2}^{x_2}) - h_{t_1}(X_{t_1}^{x_2}) - \int_{t_1}^{t_2} \tilde{A}h_r(X_r^{(x_2)}) \, dr \right) h_{t_1}(X_{t_1}^{x_1}) | \mathcal{F}_{t_1} \right] = 0, \qquad (4.4)$$

using the filtration $(\mathcal{F}_t)_{t \geq 0}$, the *joint* filtration of X^{x_1} and X^{x_2}. Combined we get

$$\mathbb{E}\left[h_{t_2}(X_{t_2}^{x_1})h_{t_2}(X_{t_2}^{x_2}) - h_{t_1}(X_{t_1}^{x_1})h_{t_1}(X_{t_1}^{x_2}) \right. \qquad (4.5)$$

$$\left. - \int_{t_1}^{t_2} h_{t_2}(X_{t_2}^{x_2})\tilde{A}h_r(X_r^{(x_1)}) + h_{t_1}(X_{t_1}^{x_1})\tilde{A}h_r(X_r^{(x_2)}) \, dr | \mathcal{F}_{t_1} \right] = 0. \qquad (4.6)$$

Using a partition of $[s, t]$, $s = t_0 < t_1 < \cdots < t_n = t$ we get

$$\mathbb{E}\left[h_t(X_t^{x_1})h_t(X_t^{x_2}) - h_s(X_s^{x_1})h_s(X_s^{x_2}) \right. \qquad (4.7)$$

$$\left. - \int_s^t \tilde{A}h_r(X_r^{(x_1)})h_r(X_r^{x_2}) + h_r(X_r^{x_1})\tilde{A}h_r(X_r^{(x_2)}) \, dr | \mathcal{F}_s \right] \qquad (4.8)$$

$$+ \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \mathbb{E}\left[(h_{t_i}(X_{t_i}^{x_2}) - h_r(X_r^{x_2}))\tilde{A}h_r(X_r^{x_1}) \right. \tag{4.9}$$

$$\left. + (h_{t_i}(X_{t_i}^{x_1}) - h_r(X_r^{x_1}))\tilde{A}h_r(X_r^{x_2})|\mathcal{F}_s \right] \mathrm{d}r = 0. \tag{4.10}$$

Denote by $F_1(r), F_2(r)$ the functions $r \to E[h_r(X_r^{x_2})\widetilde{A}h_r(X_r^{x_1}) \mid \mathcal{F}_s]$ respectively with x_1, x_2 interchanged. We have here the expression $h_r(X_s) - h_t(X_u)$, which is evaluated for $u = s$ and $r = t$ and hence we need properties of the two compared functions. Observe that for $\max|t_{k+1} - t_k| \to 0$ the continuity assumptions on $r \mapsto h_r$ and $r \mapsto X_r$ imply that the *last term vanishes* leading to (4.1) via (2.8). Namely we observe that for as $max_k \mid t_{k+1} - t_k \mid \to 0$ the functions F^1, F^2 are approximated in L_1 by $\widetilde{F}^1, \widetilde{F}^2$ given by replacing h_r by h_{t_i} in the interval $[t_i, t_{i+1}]$. This follows from the continuous differentiability of $r \to h_r$ we assumed and the stochastic continuity of X, together with the conditional independence of X^{x_1} and X^{x_2} which allows to rewrite the first expression of the integral in the third term now as: $\mathbb{E}[h(X_{t_i}^{x_2}) - h_r(X_r^{x_2}) \mid \mathcal{F}_s]\mathbb{E}[Ah_r(X_r^{x_2}) \mid \mathcal{F}_s]$ and similar for the second term making the claim immediate.

For general ψ the argument proceeds starting in (4.3) replacing $h.$ by $\psi(\cdot)h..$

ad (2)**:** This result is a corollary to (1) and it suffices to verify (2.8) having (2.9). Drop in $g_{\psi,h}$ the indices and calculate:

$$\tilde{A}\psi(t)h_t(x_1 \sqcup x_2) = \psi'(t)h_t(x_1 \sqcup x_2) + \psi(t)g(t, x_1 \sqcup x_2)h_t(x_1 \sqcup x_2) \tag{4.11}$$

$$= \psi'(t)h_t(x_1 \sqcup x_2) + \psi(t)(g(t, x_1) + g(t, x_2))h_t(x_1)h_t(x_2) \tag{4.12}$$

$$= \psi'(t)h_t(x_1)h_t(x_2) + \psi(t)\left(Ah_t(x_1)h_t(x_2) + h_t(x_1)Ah_t(x_2)\right). \tag{4.13}$$

On the other hand having the branching property we set $g_{\psi,h} = (Ah_t)/h_t \cdot \psi(t)$ to obtain a homomorphism using the multiplicity of h_t the only point to check is that the expression is well defined for a multiplicative function h_t, which is the case for $h_t(\cdot) > 0$.

5. Formulation and proofs of key facts to be used in Section 6

In this section we formulate and prove the statements which give the *key tools* used subsequently in our argument that the assumptions needed to apply the *criterion* indeed do hold, see the next section. In the following proofs we will use the notation $x_1 \sqcup x_2$ to denote the generic $x_1 \sqcup^s x_2$ which is required for the branching property.

5.1. *Formulation of the key tools. The key point is to verify the uniqueness property of the martingale problem, to choose $\widetilde{D}$ and to calculate the g in our criterion and show its $\sqcup-$ additivity* (Step 3). Everything is put together in Step 4. The truncation necessary to define the branching properties raises some technical problems in applying stochastic analysis tools. Therefore before carrying out the proof of Theorem 3.3 and working with the duality techniques in Section 5 to establish that we have a well-posed martingale problem we need some preparations (Step 1 and 2).

Step 1: Preparations. To prepare the proof of Theorem 3.3 we state some results which are proven later in Section 5.2. We recall the notation $\mathbb{D}_n = \{\underline{\underline{r}} \in \mathbb{R}^{\binom{n}{2}} : 0 \leq r_{ik} \leq r_{ij} + r_{jk}, 1 \leq i < j < k \leq n\}$ as the subset of mutual distances which can be realized by n points in a metric space, $n \in \mathbb{N}$. By convention $\mathbb{D}_1 = \{0\}$.

We will need in the sequel a function, $\varrho : \mathbb{R} \times \bigcup_n \mathbb{D}_n \to \mathbb{R}$, which generates a sliding window of truncation which filters out the information in the cut out pieces, like the truncation operator, but which has *smoothness properties* which allow for the calculus of martingale problems. We use the notation $\varrho_t(\underline{\underline{r}}) = \varrho^{(n)}(t, \underline{\underline{r}})$, $t \in \mathbb{R}$, $\underline{\underline{r}} \in \mathbb{D}_n$, $n \in \mathbb{N}$.

Lemma 5.1. *Assume that* $\varrho^{(n)} \in C^1(\mathbb{R} \times \mathbb{D}_n, [0, 1])$, $n \in \mathbb{N}$. *The process* $(t, \mathfrak{U}_t)_{t \geq 0}$ *with* $(\mathfrak{U}_t)_{t \geq 0}$ *from Definition 3.2 is a solution to the* $(\tilde{\Omega}, D_{\text{lin}})$ *martingale problem, where*

$$\tilde{\Omega} = \Omega^{\uparrow} + \partial_t, \tag{5.1}$$

$$D_{\text{lin}} = \{(t, \mathfrak{u}) \mapsto \psi(t)\Phi^{n, \phi \varrho_t}(\mathfrak{u}) : \psi \in C_b^1(\mathbb{R}, \mathbb{R}), \phi \in C_b^1\}. \tag{5.2}$$

Note that D_{lin} *depends on* ϱ, *but we do not explicitly state that dependence.*

The lemma will be proved in Section 5.2. The next lemma shows that another domain of the operator can be chosen giving an equivalent martingale problem.

Lemma 5.2. *The following are equivalent:*

(1) $(t, \mathfrak{U}_t)_{t \geq 0}$ *solves the* $(\tilde{\Omega}, D_{\text{lin}})$-*MGP*

(2) $(t, \mathfrak{U}_t)_{t \geq 0}$ *solves the* $(\tilde{\Omega}, \tilde{D})$-*MGP, where*

$$\tilde{D} = \{(t, \mathfrak{u}) \mapsto \psi(t)\exp(-\Phi^{n, \phi \varrho_t}(\mathfrak{u})) : \psi \in C_b^1(\mathbb{R}), \phi \in C_b^1(\mathbb{D}_n), n \in \mathbb{N}\}. \tag{5.3}$$

Note that $\tilde{D}$ *depends on* ϱ, *but we do not explicitly state that dependence.*

This lemma is important since it allows to work with multiplicative functions as required in Theorem 2.11.

Step 2: Sliding window of functions. Return to the functions ϱ from above and specialize to the present context.

We define the $t-$truncated polynomial:

$$\Phi_t^{n, \phi}(\mathfrak{u}) = \Phi^{n, \phi \cdot c_t}(\mathfrak{u}), \text{ where } c_t(\underline{\underline{r}}) = \prod_{1 \leq i < j \leq n} \mathbb{1}(r_{ij} < 2t) \tag{5.4}$$

for $\Phi^{n, \phi} \in \Pi$. Truncated polynomials are additive on $\mathbb{U}(t)^{\sqcup}$ (see Theorem 2.27 in Glöde et al., 2019) which makes $\tilde{D}$ a set of *multiplicative* functions on $\mathbb{U}(t)^{\sqcup}$. Unfortunately, truncated polynomials $\Phi_t^{n, \phi}$ do not have C^1-functions ϕc_t and that is why we use an *approximation argument* which makes use of the following assumptions on ϱ.

Assumption 5.3. The functions $\varrho^{(n)} : \mathbb{R} \times \mathbb{D}_n \to [0, 1]$ are in C^1 and of the form that for $t \in \mathbb{R}, \underline{\underline{r}} \in \mathbb{D}_n$:

$$\varrho^{(n)}(t, \underline{\underline{r}}) = 0 \text{ if and only if there is } 1 \leq i < j \leq n \text{ with } r_{ij} \geq 2t. \tag{5.5}$$

Moreover, we require that $\varrho^{(n)}(t, \cdot)$ is non-increasing in any coordinate for any $t \geq 0$.

Recall the notation $\underline{\underline{t}} = (t)_{1 \leq i < j \leq n} \in \mathbb{D}_n$, for an array with the constant entry t in all $\binom{n}{2}$ coordinates. We will not specify the dimension of the array in order not to overload notation.

Assumption 5.4. For any $n \in \mathbb{N}$, $c \geq 0$: $\varrho^{(n)}(t, \underline{r}) = \varrho^{(n)}(t+c, \underline{r}+2\underline{c})$, $t \in \mathbb{R}, \underline{r} \in \mathbb{D}_n$, $n \in \mathbb{N}$.

Under these assumptions we get the *approximation property* below.

Lemma 5.5. *Let $t > 0$. Suppose Assumptions 5.3 and 5.4 hold for functions $\varrho = \varrho^{(n)} : \mathbb{R} \times \mathbb{R}^{\binom{n}{2}} \to [0,1]$, $n \in \mathbb{N}$. Then for any truncated polynomial $\Phi_t^{n,\phi} \in \Pi(C^1(\mathbb{D}_n))$, we can find a sequence of polynomials $\Phi^{n,\phi_N \varrho_t}$ in the family of polynomials $\{\Phi^{n,\phi \varrho_t} : \phi \in C_b^1(\mathbb{D}_n), n \in \mathbb{N}\}$ such that for all $\mathfrak{u} \in \mathbb{U}$:*

$$\Phi^{n,\phi_N \varrho_t}(\mathfrak{u}) \nearrow \Phi_t^{n,\phi}(\mathfrak{u}), \ as \ N \to \infty. \tag{5.6}$$

Proof of Lemma 5.5: Recall that $(\phi \cdot \varrho_t)(\underline{r}) = \phi(\underline{r}) \cdot \hat{\varrho}(\underline{r} - 2\underline{t})$ by Assumption 5.4. Let $g_N \in C^1(\mathbb{R}, \mathbb{R})$ with $g_N|_{[-\infty,0]} \equiv 0$ and $g_N|_{[N^{-1},\infty)} \equiv 1$. Define

$$\phi_N(\underline{r}) = \frac{\phi(\underline{r})}{\varrho_t(\underline{r})} \cdot \prod_{1 \leq i < j \leq n} g_N(2t - r_{ij}). \tag{5.7}$$

Clearly, $\phi_N \in C_b^1(\mathbb{D}_n)$ for any $N \in \mathbb{N}$. This is the case since $\varrho^{(n)}(t, \cdot)$ is decreasing by Assumption 5.3. With a similar argument as that of *Lemma 3.12* in Glöde et al. (2019) we can see that $\Phi^{n,\phi_N \cdot \varrho_t}(\mathfrak{u}) \to \Phi^{n,\phi \cdot h_t}(\mathfrak{u})$ as $N \to \infty$). $\qquad \square$

This completes our technical preparations and we have to check later the basic assumptions to work with our approach in this model.

Proposition 5.6 (Feynman-Kac duality and uniqueness, Depperschmidt and Greven, 2019)**.** *Under Assumptions 5.3 and 5.4 there is a dual process for the process $(t, \mathfrak{U}_t)_{t \geq 0}$ with a Feynman-Kac duality relation. If the initial condition $\mathbb{P}[\mathfrak{U}_0 \in \cdot]$ is deterministic, then uniqueness holds for the $(\tilde{\Omega}, \tilde{D})$ martingale problem in the sense that for any other solution $(\mathfrak{U}_t', t)_{t \geq 0}$ we have $\mathfrak{U}_t(t) \overset{\mathrm{d}}{=} \mathfrak{U}_t'(t)$, for every $t > 0$.*

This means that any two solutions at time t have the same t-top. One can see that as a one-dimensional uniqueness result. It is not surprising that we do not obtain a finer result: Assumption 5.3 cuts off information beyond that level. The uniqueness result also holds more general if we require moment bounds on the initial conditions.

Next, the formula for g and the important property of $\sqcup^t$-additivity is stated in the next proposition. This is the key for the proof of Theorem 3.3.

Proposition 5.7 (Key formula for $\mathbb{U}$-valued branching)**.** *For $f = \psi e^{-\Phi} \in \tilde{D}$ of Lemma 5.2 with ϱ satisfying Assumptions 5.3 and 5.4:*

$$\tilde{\Omega}f(t, \mathfrak{u}) = \psi'(t)e^{-\Phi(\mathfrak{u})} + \psi(t)g_\phi(t, \mathfrak{u})e^{-\Phi(\mathfrak{u})}, \tag{5.8}$$

where

$$g_\phi(t, \mathfrak{u}) = \Omega^{\uparrow,\mathrm{grow}}\Phi^{n,\phi \varrho_t}(\mathfrak{u}) + \frac{bn}{2\bar{\mathfrak{u}}}\Phi^{2n,(\phi \varrho_t)\times(\phi \varrho_t)\circ\theta_{1,n+1}}(\mathfrak{u}) \tag{5.9}$$

is $\sqcup^t$-additive. Moreover, under Assumption 5.3 the function $\mathbb{U}(t)^\sqcup \mapsto \mathbb{R}$, $\mathfrak{u} \mapsto \Phi^{n,\phi \varrho_t}(\mathfrak{u})$ is additive for any $\phi \in C(\mathbb{R}^{\binom{n}{2}})$, $n \in \mathbb{N}$.

The proofs of all the previous results and in particular the *calculation of g_ϕ* are contained in Subsection 5.2 and 5.3.

5.2. *Proof of tools: Lemma 5.1, 5.2 and Proposition 5.7.*

Proof of Lemma 5.1: The proof checks the conditions of Lemma 4.3.4 in Ethier and Kurtz (1986) giving the claim. Define the following functions in our case not depending on ω, but we use the notation of the reference).

$$u : \begin{cases} [0,\infty) \times \mathbb{U} \times \Omega & \to \mathbb{R} \\ (t, \mathfrak{u}, \omega) & \mapsto \psi(t)\Phi^{n,\phi\varrho_t}(\mathfrak{u}), \end{cases} \tag{5.10}$$

$$v : \begin{cases} [0,\infty) \times \mathbb{U} \times \Omega & \to \mathbb{R} \\ (t, \mathfrak{u}, \omega) & \mapsto \psi'(t)\Phi^{n,\phi\varrho_t}(\mathfrak{u}) + \psi(t)\Phi^{n,\phi\partial_t\varrho_t}(\mathfrak{u}), \end{cases} \tag{5.11}$$

$$w : \begin{cases} [0,\infty) \times [0,\infty) \times \mathbb{U} \times \Omega & \to \mathbb{R} \\ (t, s, \mathfrak{u}, \omega) & \mapsto \psi(t)\Big(\Phi^{n,2\overline{\nabla}(\phi\varrho_t)}(\mathfrak{u}) + an\Phi^{n,\phi\varrho_t}(\mathfrak{u}) \\ & \qquad\qquad + \frac{b}{\underline{\mathfrak{u}}}\sum_{1\leq k<l\leq n}\Phi^{n,\phi\varrho_t\circ\theta_{k,l}}(\mathfrak{u})\Big). \end{cases} \tag{5.12}$$

The following Assumptions (4.3.10) and (4.3.11) of the reference are by the construction of the process (note we vary here either only in the explicit time coordinate or in the state of the genealogy) satisfied for $s \leq t$:

$$\mathbb{E}\left[u(t,\mathfrak{U}_t) - u(s,\mathfrak{U}_t) - \int_s^t v(r,\mathfrak{U}_t)\,\mathrm{d}r|\mathcal{F}_s\right] = 0, \tag{5.13}$$

$$\mathbb{E}\left[u(s,\mathfrak{U}_t) - u(s,\mathfrak{U}_s) - \int_s^t w(s,r,\mathfrak{U}_r)\,\mathrm{d}r|\mathcal{F}_s\right] = 0. \tag{5.14}$$

Additionally, $\mathfrak{U}$ is right-continuous by construction and $\mathfrak{u} \mapsto v(t,\mathfrak{u},\omega)$ is continuous for fixed t,ω, since $\phi\varrho_t$ and $\phi\partial_t\varrho_t$ are continuous and so the terms involved in v are classical polynomials. The left-continuity of $t \mapsto w(t,s,\mathfrak{u},\omega)$ is clear by continuity of $t \mapsto \overline{\nabla}\varrho_t$ and $t \mapsto \varrho_t$.

Note that our functions are *not bounded* as required in Ethier and Kurtz (1986). However, to obtain convergence in their equation (4.3.17), the necessary *integrability criteria* follow by a dominated convergence argument with moment conditions on $\sup_{s\leq t} \mathfrak{U}_s$ and the bounds on $\|\psi\|, \|\psi'\|, \|\phi\|, \|\partial_t\varrho_t\|, \|\overline{\nabla}\varrho\|$, etc. More concrete: boundedness of v can be replaced by $\mathbb{E}[\sup_{s\in[t_1,t_2]}|v(s,\mathfrak{U}_s)|] < \infty$ and similarly for w. This is a consequence of properties of the total mass process, the Feller diffusion, where all moments exist for all t starting in a fixed point and the total mass process is a semi-martingale.

We now calculate the generator action of the Markov process which was introduced in Lemma 5.1, namely for $f = \psi\Phi \in D_{\mathrm{lin}}$:

$$\Omega^\uparrow\psi(t)\Phi^{n,\phi\varrho_t}(\mathfrak{u}) = \psi(t)\Big(an\Phi^{n,\phi\varrho_t}(\mathfrak{u}) + \frac{b}{\underline{\mathfrak{u}}}\sum_{1\leq k<l\leq n}\Phi^{n,\phi\varrho_t\circ\theta_{k,l}}(\mathfrak{u}) \tag{5.15}$$
$$+ \Phi^{n,2(\overline{\nabla}\phi)\varrho_t}(\mathfrak{u}) + \Phi^{n,2\phi(\overline{\nabla}\varrho_t)}(\mathfrak{u})\Big),$$

$$\partial_t\psi(t)\Phi^{n,\phi\varrho_t}(\mathfrak{u}) = \psi'(t)\Phi^{n,\phi\varrho_t}(\mathfrak{u}) + \psi(t)\Phi^{n,\phi\partial_t\varrho_t}(\mathfrak{u}). \tag{5.16}$$

combining (5.10)-(5.16) gives the claim of the lemma. $\square$

Proof of Lemma 5.2: We start with showing (2). Before we start we have to use the definition of Ω on D to obtain now the action of the operator which is induced on other functions, namely $\tilde{\Omega}$ acting on $\tilde{D}$. This will give us then the compensator

for the process $exp(-\Phi^{n,\phi}(\mathfrak{U}_t))$, allowing us to show that the following process $(M_t^{n,\phi})_{t\geq 0}$ is a martingale (here $\phi \times \phi$ refers to the function ϕ evaluating two independent samples) if we set for $t \geq 0$:

$$M_t^{n,\phi} := \exp(-\Phi^{n,\phi}(\mathfrak{U}_t)) - \exp(-\Phi^{n,\phi}(\mathfrak{U}_0)) \tag{5.17}$$

$$- \int_0^t \exp(-\Phi^{n,\phi}(\mathfrak{U}_t)) \left[\Omega^{\uparrow,\mathrm{grow}} \Phi^{n,\phi}(\mathfrak{U}_s) + \frac{nb}{2\bar{\mathfrak{U}}_s} \Phi^{2n,\phi \times \phi \circ \theta_{1,n+1}}(\mathfrak{U}_s) \right] \,\mathrm{d}s. \tag{5.18}$$

Therefore, we obtain the operator $\widetilde{\Omega} = \Omega^{\uparrow} + \frac{\partial}{\partial t}$ on $\widetilde{D}$ by calculating:

$$\Omega^{\uparrow} \psi(t) \exp(-\Phi^{n,\phi\varrho_t})(\mathfrak{u}) = \psi(t) \exp(-\Phi^{n,\phi\varrho_t}(\mathfrak{u})) \Big(\Omega^{\uparrow,\mathrm{grow}} \Phi^{n,\phi\varrho_t}(\mathfrak{u}) \tag{5.19}$$

$$+ \frac{bn}{2\bar{\mathfrak{u}}} \Phi^{2n,\phi \times \phi \circ \theta_{1,n+1}}(\mathfrak{u}) \Big), \tag{5.20}$$

$$\partial_t \psi(t) \exp(-\Phi^{n,\phi\varrho_t})(\mathfrak{u}) = \psi'(t) \exp(-\Phi^{n,\phi\varrho_t})(\mathfrak{u}) \tag{5.21}$$

$$+ \psi(t) \exp(-\Phi^{n,\phi\varrho_t}(\mathfrak{u})) \cdot \Phi^{n,\phi\partial_t\varrho_t}(\mathfrak{u}). \tag{5.22}$$

Arguing as in the proof of Lemma 5.1 one gets then indeed,

$$N_t^{\psi,\phi} := \psi(t) \exp(-\Phi^{n,\phi\varrho_t}(\mathfrak{U}_t)) - \psi(0) \exp(-\Phi^{n,\phi\varrho_t}(\mathfrak{U}_s))$$

$$- \int_0^t \Omega^{\uparrow} \psi(t) \exp(-\Phi^{n,\phi\varrho_t})(\mathfrak{U}_s) + \partial_t \psi(t) \exp(-\Phi^{n,\phi\varrho_t})(\mathfrak{U}_s) \,\mathrm{d}s, \; t \geq 0 \tag{5.23}$$

defines a martingale. This is (2).

We therefore have to verify (5.19) and (5.21). The second is simply calculus. Next to the first. A standard calculation as in Corollary 2.13 of Greven et al. (2013) shows this via the Itô-formula. Namely we expand the exponential and calculate the compensator of $\Phi(\mathfrak{U}_t)$ and the quadratic variation of $(\Phi(\mathfrak{U}_t))_{t\geq 0}$, where the latter requires to calculate the compensator of $(\Phi^2(\mathfrak{U}_t))_{t\geq 0}$. Follow here Greven et al. (2013) equation (8.1) and (8.3) to get this.

Now (1) follows by differentiation of $N_t^{\psi,\lambda\phi}$ w.r.t. λ at $\lambda = 0$. □

Proof of Proposition 5.7: To get the action of $\tilde{\Omega}$ on $\tilde{D}$ we need to add up terms in (5.19) and (5.21). Using Assumption 5.4 as before this allows to eliminate the terms with $\nabla \varrho_t$ and $\partial_t \varrho_t$ to obtain:

$$\tilde{\Omega} \psi(t) \exp(-\Phi^{n,\phi\varrho_t})(\mathfrak{u}) = \psi(t) \exp(-\Phi^{n,\phi\varrho_t})(\mathfrak{u}) \Big(\Omega^{\uparrow,\mathrm{grow}} \Phi^{n,\phi\varrho_t} \tag{5.24}$$

$$+ \frac{bn}{2\bar{\mathfrak{u}}} \Phi^{2n,\phi\varrho_t \times \phi\varrho_t \circ \theta_{1,n+1}}(\mathfrak{u}) \Big) + \psi'(t) \exp(-\Phi^{n,\phi\varrho_t})(\mathfrak{u}). \tag{5.25}$$

Using the notation of the lemma we need to show that the expression below is $\sqcup^t$-additive:

$$g(t,\mathfrak{u}) = \Omega^{\uparrow,\mathrm{grow}} \Phi^{n,\phi\varrho_t}(\mathfrak{u}) + \frac{bn}{2\bar{\mathfrak{u}}} \Phi^{2n,(\phi\varrho_t) \times (\phi\varrho_t) \circ \theta_{1,n+1}}(\mathfrak{u}) =: g_1(t,\mathfrak{u}) + g_2(t,\mathfrak{u}). \tag{5.26}$$

We see that the expressions are truncated polynomials. It is elementary using Assumption 5.3 and Proposition 3.8 in Glöde et al. (2019) to establish that $g_1(t,\cdot)$ is $\sqcup^t$-additive. Similar reasoning applies for g_2. Again by Proposition 3.8 in the same

reference we see with Assumption 5.3 that $\mathfrak{u} \mapsto \exp(-\Phi^{n,\phi \varrho_t}(\mathfrak{u}))$ is multiplicative on $\mathbb{U}(t)^{\sqcup}$ and note here that the operator of our process relates via (3.13) to the operator of Fleming-Viot in the reference via $\bar{\mathfrak{u}} \Omega^{\uparrow} - Id$. $\qquad \square$

5.3. *Proof of duality: Proposition 5.6.* Here we will derive a duality related for the operator $(\tilde{\Omega}, D_{\mathrm{lin}})$. This allows to deduce *uniqueness* for the $(\tilde{\Omega}, \tilde{D})$ martingale problem.

It is easy to verify the following preparatory lemma.

Lemma 5.8. *Under Assumption 5.3 and 5.4 the function $\varrho^{(n)} \in C^1(\mathbb{R}_+ \times \mathbb{D}_n, [0,1])$ is of the form*

$$\varrho^{(n)}(t,\underline{r}) = \hat{\varrho}^{(n)}(\underline{r} - 2\underline{t}), \ t \geq 0, \underline{r} \in \mathbb{D}_n, n \in \mathbb{N}, \tag{5.27}$$

for a function $\hat{\varrho}^{(n)} \in C^1(\mathbb{D}_n, [0,1])$ with $\hat{\varrho}^{(n)}\big|_{\mathbb{D}_n \setminus (0,\infty)^{\binom{n}{2}}} \equiv 0$ and positive on $(0,\infty)^{\binom{n}{2}}$. Moreover $(\partial_t + 2\bar{\nabla})\varrho^{(n)}(t,\underline{r}) = 0$.

Using the two assumptions we can thus derive simpler expression statement for the action of $\tilde{\Omega}$:

$$\tilde{\Omega}\psi(t)\Phi^{n,\phi \varrho_t}(\mathfrak{u}) = \psi(t)\Big(bn\Phi^{n,\phi \varrho_t}(\mathfrak{u}) + \frac{b}{\bar{\mathfrak{u}}} \sum_{1 \leq k < l \leq n} \Phi^{n,(\phi \varrho_t) \circ \theta_{k,l}}(\mathfrak{u}) \tag{5.28}$$

$$+ \Phi^{n,2(\bar{\nabla}\phi)\varrho_t}(\mathfrak{u})\Big) + \psi'(t)\Phi^{n,\phi \varrho_t}(\mathfrak{u}). \tag{5.29}$$

This makes the function of the two assumptions clear .

We restrict ourselves to the case that $a = 0$, i.e. there is no drift; all calculations can be done without that restriction, see Section 5.1, Step 2 in Depperschmidt and Greven (2019) for details. All this is done for a fixed function $\hat{\varrho}^{(n)}$ as in Lemma 5.8, i.e. Assumptions 5.3 and 5.4 are fulfilled and (5.28) can be used.

Define the following set of functions:

$$H^{\phi,\psi} : \begin{cases} (\mathbb{U} \times \mathbb{R}_{\geq 0}) \times (\mathbb{S} \times \mathbb{R}^{\binom{\mathbb{N}}{2}} \times \mathbb{R}_{\geq 0}) & \to \mathbb{R} \\ ((\mathfrak{u},t),(p,\underline{r}',s)) & \mapsto \psi(t+s) \int \mu^{\otimes \#p}(\mathrm{d}\underline{x}_p) \, \phi\left(\underline{r}^p(\underline{x}_p) + \underline{r}'\right) \\ & \qquad \varrho(2(\underline{t}+\underline{s}) - \underline{r}^p(\underline{x}_p) - \underline{r}'). \end{cases} \tag{5.30}$$

Define for $p(n) = \{\{1\},\ldots,\{n-1\},\{n,n+1,\ldots\}\}$ the following sets of functions:

$$\mathcal{H} = \{H^{\phi,\psi}(\cdot,\cdot) : \ \phi \in C_b^1(\mathbb{R}^{\binom{\mathbb{N}}{2}}) \text{ with finite support, } \psi \in C_b^1(\mathbb{R}_{\geq 0})\}, \tag{5.31}$$

$$\mathcal{G}^{\uparrow} = \{H^{\phi,\psi}(0,(p(n),\underline{0})) : \ n \in \mathbb{N}, \psi \in C_b^1(\mathbb{R}_{\geq 0}), \phi \in C_b^1(\mathbb{R}^{\binom{\mathbb{N}}{2}}) \text{ with finite support}\} \tag{5.32}$$

and

$$\mathcal{G}^{\downarrow} = \{H^{\phi,\psi}((t,\mathfrak{u}),\cdot) : \ \mathfrak{u} \in \mathbb{U}, t \geq 0, \phi \in C_b^1(\mathbb{R}^{\binom{\mathbb{N}}{2}}) \text{ with finite support}, \tag{5.33}$$

$$\psi \in C_b^1(\mathbb{R}_{\geq 0})\}.$$

Then $D_{\mathrm{lin}} = \mathcal{G}^{\uparrow}$.

Next turn to the dual process. For a function $G : \mathbb{K} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ depending on only finitely many coordinates define

$$L^{\downarrow,\text{grow}}G(s, p, \underline{\underline{r}}') = \partial_s G(s, p, \underline{\underline{r}}') + \sum_{i \nsim_p j} \frac{\partial}{\partial r'_{ij}} G(s, p, \underline{\underline{r}}') \qquad \text{and} \qquad (5.34)$$

$$L^{\downarrow,\text{coal}}G(s, p, \underline{\underline{r}}') = b \sum_{\pi,\pi' \in p} \left(G(s, \kappa_p(\pi, \pi'), \underline{\underline{r}}') - G(s, p, \underline{\underline{r}}') \right), \qquad (5.35)$$

for $p \in \mathbb{S}$, $\underline{\underline{r}}' \in \mathbb{R}^{\binom{N}{2}}$ and $s \geq 0$. Here $\chi_p(\pi, \pi')$ is the partition where π and π' are replaced by their union.

The coalescent operator is now:

$$L^{\downarrow,\text{K}} = L^{\downarrow,\text{grow}} + L^{\downarrow,\text{coal}}. \qquad (5.36)$$

One obtains readily for the Kingman coalescent a marked ultrametric measure space which gives a solution to a martingale problem related to that operator.

Lemma 5.9. *Let $n \in \mathbb{N}$ and let $(s, k_s)_{s \geq 0}$ be the time-space genealogy-valued Kingman coalescent started in $(0, p(n), \underline{0})$ defined on page 809 of Greven et al. (2013). Then the process $(s, k_s)_{s \geq 0}$ is a solution of the martingale problem for $(\delta_{(0,p(n),\underline{0})}, L^{\downarrow,K}, \cdot, \mathcal{G}^{\downarrow})$.*

Proof: Follow Lemma 4.3.4 in Ethier and Kurtz (1986). Let for any fixed $(t, \mathfrak{u}) \in \mathbb{R}_{\geq 0} \times \mathbb{U}$:

$$u : \begin{cases} [0, \infty) \times \mathbb{S} \times \mathbb{R}^{\binom{\mathbb{N}}{2}} \times \Omega \to \mathbb{R} \\ (s, (p, \underline{\underline{r}}'), \omega) \mapsto \psi(t + s) \int \mu^{\otimes \#p}(\mathrm{d}\underline{x}_p) \, \phi\left(\underline{\underline{r}}^p(\underline{x}_p) + \underline{\underline{r}}'\right) \varrho(2(\underline{t} + \underline{s}) - \underline{\underline{r}}^p(\underline{x}_p) - \underline{\underline{r}}'), \end{cases}$$

$$(5.37)$$

$$v : \begin{cases} [0, \infty) \times \mathbb{S} \times \mathbb{R}^{\binom{\mathbb{N}}{2}} \times \Omega \to \mathbb{R} \\ (s, (p, \underline{\underline{r}}'), \omega) \mapsto \psi'(t + s) \int \mu^{\otimes \#p}(\mathrm{d}\underline{x}_p) \, \phi\left(\underline{\underline{r}}^p(\underline{x}_p) + \underline{\underline{r}}'\right) \\ \qquad\qquad\qquad \varrho(2(\underline{t} + \underline{s}) - \underline{\underline{r}}^p(\underline{x}_p) - \underline{\underline{r}}') \\ \qquad + \psi(t + s) \int \mu^{\otimes \#p}(\mathrm{d}\underline{x}_p) \, \phi\left(\underline{\underline{r}}^p(\underline{x}_p) + \underline{\underline{r}}'\right) \\ \qquad\qquad\qquad \partial_s \varrho(2(\underline{t} + \underline{s}) - \underline{\underline{r}}^p(\underline{x}_p) - \underline{\underline{r}}') \end{cases} \qquad (5.38)$$

$$w : \begin{cases} [0, \infty) \times [0, \infty) \times \mathbb{S} \times \mathbb{R}^{\binom{\mathbb{N}}{2}} \times \Omega \to \mathbb{R} \\ (s, \vartheta, (p, \underline{\underline{r}}'), \omega) \mapsto (L^{\downarrow,\text{coal}} + \sum_{i \nsim_p j} \frac{\partial}{\partial r'_{ij}}) H((t, \mathfrak{u}), (s, p, \underline{\underline{r}}')). \end{cases} \qquad (5.39)$$

To check (4.3.10) and (4.3.11) in Ethier and Kurtz (1986) use calculations as in Section 4 in Greven et al. (2013).

Additionally, $(k_s)_{s \geq 0} = (p_s, \underline{\underline{r}}'_s)_{s \geq 0}$ is right-continuous by construction and $(p, \underline{\underline{r}}') \mapsto v(s, p, \underline{\underline{r}}', \omega)$ is continuous for fixed s, ω: Continuity in p is obvious, since $\mathbb{S}$ is a discrete space. Continuity in $\underline{\underline{r}}'$ is true, $(\underline{\underline{r}}', \underline{\underline{r}}) \mapsto \phi(\underline{\underline{r}} + \underline{\underline{r}}')\varrho(2(\underline{t} - \underline{s}) - \underline{\underline{r}} - \underline{\underline{r}}')$ is continuous and bounded. This allows to apply dominated convergence to get the continuity in $\underline{\underline{r}}'$.

The left-continuity of $s \mapsto w(s, \vartheta, p, \underline{\underline{r}}', \omega)$ is clear by continuity of $s \mapsto \overline{\nabla}\varrho_s$ and $s \mapsto \varrho_s$. Note that now our functions involved are *not bounded* as required in Ethier

and Kurtz (1986). However, $(\#p_s)_{s \geq 0}$ is decreasing and so for any initial state the convergence in (4.3.17) can be shown. $\qquad \square$

Lemma 5.10 (Feynman-Kac Duality). *Let $k = (s, k_s)_{s \geq 0}$ be a solution of the $(\delta_{(0,p(n),\underline{0})}, L^{\downarrow,K}, \cdot, \mathcal{G}^{\downarrow})$ martingale problem. Let $(t, \mathfrak{U}_t)_{t \geq 0}$ be a solution of the $(\tilde{A}, D_{\mathrm{lin}})$ martingale problem started in $\mathbf{P}_0 \otimes \delta_0 \in \mathcal{M}_1(\mathbb{U} \times \mathbb{R})$. Then for all $H^{\phi,\psi} \in \mathcal{H}$ the following Feynman-Kac duality holds:*

$$\mathbb{E}_{\mathbf{P}_0}\left[H^{\phi,1}((t, \mathfrak{U}_t), (0, p, \underline{0}))\right] = \mathbb{E}_{(0,p,\underline{0})}\left[H^{\phi,1}(\mathfrak{u}, (t, p_t, \underline{r}'_t))e^{\int_0^t \binom{\#p_s}{2}\,ds}\right]. \quad (5.40)$$

Proof: We show the *generator criterion for duality* relations namely (4.4.41) in Ethier and Kurtz (1986):

$$(L^{\downarrow,\mathrm{grow}} + \partial_s)H^{\phi,\psi}((t,\mathfrak{u}), (s, p, \underline{r}')) = (\Omega^{\uparrow,\mathrm{grow}} + \partial_t)H^{\phi,\psi}((t,\mathfrak{u}), (s, p, \underline{r}')), \quad (5.41)$$

$$\left(L^{\downarrow,\mathrm{bran}} + \binom{\#p}{2}\right)H^{\phi,\psi}((t,\mathfrak{u}), (s, p, \underline{r}')) = \Omega^{\uparrow,\mathrm{bran}}H^{\phi,\psi}((t,\mathfrak{u}), (s, p, \underline{r}')). \quad (5.42)$$

The latter line holds as in the case without marks. To prove the first statement is the same as without marks. Note that the exponential term is bounded and no integrability problems arise. $\qquad \square$

Proof of Proposition 5.6: Let $(\mathfrak{U}_t, t)_{t \geq 0}$ and $(\mathfrak{U}'_t, t)_{t \geq 0}$ be two solutions of the $(\tilde{\Omega}, \tilde{D})$ martingale problem. By Lemma 5.2 it is then also a solution to the $(\tilde{\Omega}, D_{\mathrm{lin}})$ martingale problem. Fix $\phi \in C_b^1(\mathbb{D}_n)$. Then, by duality in Lemma 5.10 we have for $p = \{\{1\}, \ldots, \{m-1\}, \{m, m+1, \ldots\}\}$ and $\psi \equiv 1$:

$$\mathbb{E}_{\mathbf{P}_0}\left[H^{\phi,1}((t, \mathfrak{U}_t), (0, p, 0))\right] = \mathbb{E}_{(p,0,0)}\left[H^{\phi,1}(\mathfrak{u}, (t, p_t, \underline{r}'_t))e^{\int_0^t \binom{\#p_t}{2}\,ds}\right] \quad (5.43)$$

$$= \mathbb{E}_{\mathbf{P}_0}\left[H^{\phi,1}((t, \mathfrak{U}'_t), (0, p, 0))\right]. \quad (5.44)$$

Hence

$$\mathbb{E}_{\mathbf{P}_0}\left[\Phi^{m,\phi\varrho_t}(\mathfrak{U}_t)\right] = \mathbb{E}_{\mathbf{P}_0}\left[\Phi^{m,\phi\varrho_t}(\mathfrak{U}'_t)\right]. \quad (5.45)$$

Above holds for *any* $\phi \in C_b^1(\mathbb{D}_n)$. By Lemma 5.5 and monotone convergence this implies

$$\mathbb{E}_{\mathbf{P}_0}\left[\Phi_t^{m,\phi}(\mathfrak{U}_t)\right] = \mathbb{E}_{\mathbf{P}_0}\left[\Phi_t^{m,\phi}(\mathfrak{U}'_t)\right]. \quad (5.46)$$

But $\Phi_t^{m,\phi}(\mathfrak{U}_t) \leq \|\phi\|_\infty \bar{\mathfrak{U}}_t^m$ and the bound $\mathbb{E}[\bar{\mathfrak{U}}_t^m] \leq c(t)(\mathbf{P}_0[\bar{\mathfrak{U}}_0^m] + 1)$ holds by the classical estimates for the Feller diffusion for some function $t \mapsto c(t) < \infty$. Therefore we continue with the following statement following the standard argument combining Proposition 2.6. in Greven et al. (2009), Proposition 4.4.6. (page 115) in Ethier and Kurtz (1986) and the discussion after equation 4.4.21. therein:

Lemma 5.11. *The algebra generated by Π is separating on*

$$\tilde{\mathcal{M}} = \left\{P \in \mathcal{M}_1(\mathbb{U}) : \limsup_{K \to \infty} \frac{1}{K}\left(\int \bar{\mathfrak{u}}^{nK}\,P(d\mathfrak{u})\right)^{1/K} < \infty \ \forall n \in \mathbb{N}\right\}. \quad (5.47)$$

The algebra generated by Π is convergence determining, whenever the limit point is in $\tilde{\mathcal{M}}$.

We conclude with this Lemma that:

$$\lfloor \mathfrak{U}_t \rfloor(t) \stackrel{\mathrm{d}}{=} \lfloor \mathfrak{U}'_t \rfloor(t). \quad (5.48)$$

This means that the t-tops of the both processes coincide in law at each time $t \geq 0$. $\qquad \square$

6. Verification of the criterion in our models: Theorem 3.3

The application of our criterion requires to check whether the list of assumptions in Theorem 2.11 can be verified in a given situation. We first give a detailed proof for the $\mathbb{U}$-valued Feller diffusion and then later provide the needed modifications for the spatial case the $\mathbb{U}^E - valued$ super random walk in Section 7.

6.1. *Verification of criterion for $\mathbb{U}$-valued Feller: Theorem 3.3.*
Step 1: Why the setup applies. Due to Lemma 5.5 we make use of truncated polynomials in the martingale problem of Lemma 5.2, to get a *multiplicative domain*. Two things are missing for the application of Theorem 2.11: a uniqueness result and the linearity of the generator. Both these facts are provided by Proposition 5.6 and Proposition 5.7.

Step 2 *Proof of Theorem 3.3:* We want to use Theorem 2.11. Fix a function ϱ as in Lemma 5.1 which satisfies Assumptions 5.3 and 5.4. We have $S_t = \mathbb{U}(t)^{\sqcup}$, which is a semigroup by Proposition 2.13 in Glöde et al. (2019). For $s \leq t$ we have that $S_s \subset S_t$ trivially and the embedding is topologically consistent. Therefore we are in the setting of Section 2. Moreover, Proposition 5.7 says that $D_t = \{\mathfrak{u} \mapsto \exp(-\Phi^{n,\phi\varrho_t}(\mathfrak{u})) : \phi \in C(\mathbb{R}^{\binom{n}{2}}), n \in \mathbb{N}\}$ is a t-multiplicative family on S_t. The uniqueness of the martingale problem $(\tilde{A}, \tilde{D})$, where uniqueness is understood in the sense specified in Theorem 2.11 and $\tilde{D}$ is defined as in (5.3), was shown in Proposition 5.6. It remains to establish the additivity of the generator in (2.9) with respect to the multiplicative functions in D_t. But this is Proposition 5.7. $\square$

6.2. *Proof of Theorem 3.4.* We work here with the truncated martingale problem. First some preparations.

Assumption 6.1. There is a function $\hat{\varrho} \in C^1(\mathbb{R} \times \mathbb{R}, \mathbb{R})$ such that for all $n \in \mathbb{N}$:

$$\varrho^{(n)}(t, \underline{r}) = \prod_{1 \leq i < j \leq n} \hat{\varrho}(t, r_{i,j}) \,. \tag{6.1}$$

The following lemma is easy to verify.

Lemma 6.2. *Under Assumptions 5.3, 5.4 and 6.1 there is a function $\hat{\varrho} \in C^1(\mathbb{R}, \mathbb{R})$ with $\hat{\varrho}|_{[0,\infty)} = 0$ and*

$$\varrho^{(n)}(t, \underline{r}) = \prod_{1 \leq i < j \leq n} \hat{\varrho}(r_{i,j} - 2\underline{t}) \,, \text{ for all } t, \underline{r}, n \,. \tag{6.2}$$

Proof of Theorem 3.4: We will use Assumptions 5.3, 5.4 and 6.1 for the truncation functions $\varrho^{(n)}$. Using Assumption 5.4 and consider $\psi \equiv 1$, $\phi \equiv 1$ by (5.28)

$$\tilde{\Omega}\psi(t)\Phi^{n,\phi\varrho_t}(\mathfrak{u}) = an\Phi^{n,\varrho_t}(\mathfrak{u}) + \frac{b}{\bar{\mathfrak{u}}} \sum_{1 \leq k < l \leq n} \Phi^{n,\varrho_t \circ \theta_{k,l}}(\mathfrak{u}) \,. \tag{6.3}$$

Under Assumption 6.1 then we can make the following elementary calculation for $1 \leq k < l \leq n$:

$$\varrho_t^{(n)} \circ \theta_{k,l}(\underline{r}) = \prod_{1 \leq i < j \leq n} \hat{\varrho}\left((\theta_{k,l}(\underline{r}))_{i,j} - 2t\right) \tag{6.4}$$

$$= \prod_{1 \leq i < j \leq n, i \neq l, j \neq l} \hat{\varrho}(r_{i,j} - 2t) \cdot \prod_{i \neq l} \hat{\varrho}(r_{k \wedge i, k \vee i} - 2t) \,. \tag{6.5}$$

This is a function which is independent of r_{ij} where either $i = l$ or $j = l$. Thus, $\frac{b}{\bar{u}} \sum_{1 \leq k < l \leq n} \Phi^{n, \varrho_t \circ \theta_{k,l}}(\mathfrak{u})$ is a polynomial of degree $n - 1$. Using Proposition 5.6 we know that

$$\Phi^{n, \varrho_t}(\mathfrak{U}_t) - \Phi^{n, \varrho_0}(\mathfrak{U}_0) - \int_0^t b \binom{n}{2} \Phi^{n-1, T(\varrho_s)}(\mathfrak{U}_s) + na \Phi^{n, \varrho_s}(\mathfrak{U}_s) \, ds \qquad (6.6)$$

is a martingale, where $T(\varrho_s)$ is the transformation outlined in (6.4). More precisely

$$\sum_{1 \leq k < l \leq n} \Phi^{n, \varrho_t \circ \theta_{k,l}}(\mathfrak{u}) = \binom{n}{2} \bar{u} \int \nu^{n-1, \mathfrak{u}}(\mathrm{d}\underline{r}) \, \varrho_t^{n-1}(\underline{r}) \prod_{2 \leq i \leq m-1} \hat{\varrho}(r_{1,i} - 2t) \hat{\varrho}(-2t). \qquad (6.7)$$

In particular for $n = 2$ the following is a a martingale

$$\Phi^{2, \varrho_t}(\mathfrak{U}_t) - \Phi^{2, \varrho_0}(\mathfrak{U}_0) - \int_0^t b \hat{\varrho}(-2s) \Phi^{1, \varrho_s^{(1)}}(\mathfrak{U}_s) + 2a \Phi^{2, \varrho_s^{(2)}}(\mathfrak{U}_s) \, ds. \qquad (6.8)$$

Note that $\Phi^{1, \varrho_t^{(1)}}(\mathfrak{u}) = \hat{\varrho}(-2t) \bar{u}$.

Thus, we obtain the following ODE:

$$\partial_t \mathbb{E}[\Phi^{2, \varrho_t^{(2)}}(\mathfrak{U}_t)] = b \hat{\varrho}(-2t) \hat{\varrho}(-2t) \mathbb{E}[\bar{\mathfrak{U}}_t] + 2a \mathbb{E}[\Phi^{2, \varrho_t^{(2)}}(\mathfrak{U}_t)]$$
$$= b \left(\hat{\varrho}(-2t) \right)^2 e^{bt} \bar{u}_0 + 2a \mathbb{E}[\Phi^{2, \varrho_t^{(2)}}(\mathfrak{U}_t)]. \qquad (6.9)$$

Let $m(n, t) = \Phi^{n, \mathbb{1}(r_{i,j} < 2t \ \forall i < j)}(\mathfrak{U}_t)$ be the sum of the n-th power of the $2t$-families at time t. Claiming that $m(2, 0) = 0$ (which holds since $\hat{\varrho}(0) = 0$ and $\hat{\varrho}$ is a continuous function by Assumption 5.3) we obtain from (6.9) that

$$m(2, t) = e^{2at} a \bar{u}_0 \int_0^t e^{-as} \left(\hat{\varrho}(-2s) \right)^2 \, ds. \qquad (6.10)$$

Now take a sequence of $\hat{\varrho}$ such that the limit $\hat{\varrho} \nearrow \mathbb{1}_{\mathbb{R}_+}$ holds point-wise. The integral on the right hand side of (6.10) converges to $a^{-1}(1 - e^{-at})$. The expectation of the left hand side convergences by monotone convergence. Thus, we obtain for the expected sum of the squares of the subfamily sizes

$$\mathbb{E}[\Phi^{2, \mathbb{1}(r_{12} < 2t)}(\mathfrak{U}_t)] = b \bar{u}_0 \frac{1}{a} \left(e^{2at} - e^{at} \right). \qquad (6.11)$$

In fact the second moment of the branching process satisfies (6.9) with $\hat{\varrho}(-2t) = 1$ as well. □

7. Extensions to the marked case: Proofs of Theorem 3.14, 3.17 and their Tools

We have already in Section 3.2 generalized the basic concepts of the present work to the *marked* case, so that we now work with objects in the framework for our criterion. Next in Section 7.1 we show how to formulate some further concepts needed for the proofs, as *truncated* polynomials or *smooth* truncation and then make up a list of things still to be proven, in order to be able to apply the criterion to the extensions to the spatial case. This will conclude the proof of the main results. This involves to have the tools we used in the non-marked case available in a suitably modified form.

Later in Section 7.2 we have to show the tools can indeed be generalized to spatial versions as well and provide the necessary extensions of the tools lemma by

lemma, proposition by proposition. Then we can continue along the path of proof taken for the non-spatial model. Then in Section 7.3 we prove as the *key* element, namely the properties of the martingale problem, which is the spatial version of Proposition 5.6 (FK-duality) and Proposition 3.9 (wellposedness).

7.1. *Verification criterion for* $\mathbb{U}^E$*-valued super random walk: Theorem 3.14 and 3.17.* Now we are ready to check the generator criterion for the branching property of the $\mathbb{U}^{D^*}$*-marked genealogy-valued super random walk.* Note that the $E-$marked case does not pose problems since then the marks are *not* truncated. The work has to be done to include *historical information* on ancestral path so that we can lift the argument we gave before to the *spatial process* with marks which are paths. This *lifting* we address in the sequel. The key concepts we need are *truncated* and *smoothly truncated* polynomials.

Truncation. To study the truncated states we have introduced in Section 3.2 we need truncated polynomials. This means we first have to introduce the *truncated monomials* in the spatial context, recall here (3.31), by proceeding analog to (5.4) and adding now the indicator on distances as before and another truncation map acting on the *mark variables*. We begin by looking at the time-inhomogeneous situation, before we pass to the time-space process and then to the *adjusted paths*.

Consider now polynomials based on the function χ of the mark which evaluates the path at a tuple of time points $0 \leq t_1 < \cdots t_k < \infty$. For these polynomials we have to define now truncated polynomials which do not contain information about times before t, for some $t \geq 0$. Recall we have (recall (3.62), (3.63)) a truncation operation which has two components which separately act on the distances and on the marks. Therefore we will in the definition of *truncated polynomials* use polynomials build on ϕ and χ on which we can then let the truncation act *separately* on ϕ and namely of product form. This means the truncation has the form that the $t-$truncated polynomials $\Phi_t^{\phi,\chi}$ given via (ϕ, χ) is *truncated* by switching to a polynomial given by $(\phi^{(t)}, \chi^{(t)})$:

$$(\phi, \chi) \to (\phi^{(t)}, \chi^{(t)}), \text{ with } \phi \to \phi \cdot 1_{\{r_{i,j} \leq 2t\}} \tag{7.1}$$

and the transformation of χ is as follows.

In the mark function we "truncate" the monomial by replacing the χ by taking χ *applied to the truncated path* i.e. $\chi^{(t)}(\underline{v}) = \chi(\underline{v}^{(t)})$ (recall (3.63) and the comment afterwards. This is again a polynomial, but now with a different χ, which we call $\chi^{(t)}$, for a $t-$truncation.

Summarized this means that for marks we now consider:

$$\widehat{\chi}^{(t)}(t, v) = \widehat{\chi}(t, \underline{v}^{(t)}), \ \underline{v} = (v_i)_{i=1,\dots,n} \ , \quad v_t^{(t)} \text{ given by (3.63) .} \tag{7.2}$$

Remark 7.1. This function $\chi^{(t)}$ is not of the form $\chi \cdot \varrho$ as with ϕ and the truncation there. The reason is that this would not fit with the truncation extended to $\mathbb{U}^V$ according to the reasoning we explained in Remark 3.15. Note also we have individuals with small distances associated with the tree top which corresponds to the innovation part of the evolution, while for the path this is an increment. What we need is to just remove the information on the complementary part.

We have to *lift this truncation* now to the case of D^* marks and deal with the time-shift operation. Here of course we do this by just shifting the truncated path by $-t$ to obtain the truncated adjusted path, so that we need no new notation.

Smooth truncation. The truncation map on path $\chi \to \chi^{(t)}$ involves in general a change in value since the values evaluated before time t are replaced by the ones derived from $v(t)$ instead of $v(t_k)$ for $t_k < t$. This means if we vary t we get a change in $t \to \chi^{(t)}(v)$ in fact typically a jump at t if t is a jump point of the path v or a limit point of such jump points. Furthermore recall that these polynomials based on evaluation functionals at time points are *not* continuous on $D(\mathbb{R}, E)$.

We note that the path process generator involves a derivative w.r.t. the explicit time only so that we might think we need not smooth this part as in the case of ϕ. *However we use this by using the truncation level t being equal to current time, so* that we need the total differential so that we have for $\widehat{\chi}$ smoothness in both variables, explicit time and location. Note that for the case $V = E$ there is *no* need for truncation of the mark, this is different for $V = D(\mathbb{R}, E)$.

Explanation Fix a function $\widehat{\chi}$, i.e. we fix the sequence of the $(t_k)_{1,\cdots,m}$ where paths are evaluated. Note that these are numbers in $\mathbb{R}$ with $-\infty < t_1 < t_2 < \cdots < t_m < +\infty$. Therefore if t increases we keep evaluating path which satisfy that we evaluate for $t \geq 0$, points which observe for t_k with $t_k < t$ only values which after t-truncation are all equal to $v(t)$ for the truncated object. However before we might have had other positions. Therefore all positions change to the position at t, hence we get here a jump at the truncation time if the path at t is different from the position at the $t_k < t$ and this change jumps with a varying t if there are jumps at time t or close to it. This jump we have to dampen in a differentiable way in time, if $\underline{v}$ is such that such a jump occurs, i.e. if the v_k are not such that they are constant for $s \leq t$.

In this spirit above we have to define the *smooth truncation* operator, i.e. we have to extend now the map $\varrho : \mathbb{R} \times \mathbb{D}_n \longrightarrow [0,1]$ to a map $\varrho : \mathbb{R} \times (\mathbb{D}_n \times V^n) \to [0,1] \times V_t$ of the form $(s, \underline{r}, \underline{v}) \longrightarrow (s, \underline{r} \, \varrho_t^1, \varrho_t^2(\underline{v}^{(t)}))$ and then formulate the analogues of the Assumptions 5.3, 5.4. How to choose ϱ_t^2?

We note that ϱ_t^2 must then provide a sliding window of the ancestral path with a window defined w.r.t. the current time. As function of the distances we want to preserve the properties in the non spatial case and *add* properties as a function of the mark.

(1). Return to the martingale problem and the needed modification of Assumption 5.3. Here we have removed every information on the ancestral path time h or more back and this cutoff we now approach smoothly from the top. What smoothness properties do we have to impose now?

We adapt the smoothness property to the special form we have for the test functions of the mark. The function is in C^1 as function of the *explicit time variable*. As a function of the *marks* we have to dampen the jump explained above. We therefore have to interpolate here between the two values. We consider the *convex combination* of $\chi^{(s)}$ with $s \geq t$. We generate this convex combination by taking a function $\widetilde{\varrho}_t^2(s), s \geq t$ which takes values in $[0,1]$ is in C_b^1, monotone decreasing with $\int_t^\infty \widetilde{\varrho}_t^2(s)ds = 1$. Therefore we now have that $\chi^{(t)}$ is a function of $(\underline{v}^{(s)})_{s \geq t}$ which we denote by $\varrho_t^2(\chi)$ so that now

$$(\phi, \chi) \to (\phi\varrho_t^1, \varrho_t^2(\chi)). \tag{7.3}$$

This is easily lifted to the time-space process replacing χ by $\widehat{\chi}$, with $\widehat{\chi}(s, \cdot) = \psi(s)\chi(\cdot)$.

We have to pass now to *functions of the adjusted paths*. We achieve this by just considering for current time T, the $(-T)$-shifted function $\widetilde{\varrho}_t^2(\cdot)$. This now gives us the modification needed in Assumption 5.3.

(2). In order to get the marked version of Assumption 5.4 we proceed as follows. If we want to $(t+c)$-truncate instead of t-truncation, hence the basic function $\widetilde{\varrho}_t(\cdot)$ is shifted by $-c$, i.e.

$$\varrho_t(s) = \varrho_{t-s}(0) = \varrho_0(t-s). \tag{7.4}$$

Test-functions for martingale problems. Consider the *smoothly truncated* test function on $\mathbb{U}(t)^{\sqcup}$ of the forms:

$$h_t(\mathfrak{u}) = \ \exp\left(-\Phi^{n,\phi\varrho_t^1,\varrho_t^2}(\widehat{\chi})\right), \ \text{with } \phi, \chi \geq 0, \tag{7.5}$$

$$F(t,\mathfrak{u}) = \Psi(t)\, \exp\left(-\Phi^{n,\phi\varrho_t^1,\varrho_t^2}(\widehat{\chi})\right). \tag{7.6}$$

We have to calculate the generator now for a function for $n \in \mathbb{N}, \phi, \chi \geq 0$ and $\phi \in C_b^1(\mathbb{R}^{\binom{n}{2}}, \mathbb{R})$ and $\widehat{\chi} \in b\mathcal{B}(\mathbb{R} \times V, \mathbb{R})$ of the form as constructed in (3.41), which are built from functions $\chi_i^k \in b\mathcal{B}(E, \mathbb{R})$ and functions $\Psi_k \in C_b^1(\mathbb{R}, \mathbb{R})$, furthermore $\Psi \in C_b^1(\mathbb{R}, \mathbb{R})$ and ϱ_t satisfies for all $t \geq 0$ respectively $\varrho.(\cdot)$ the path-marked version of the Assumption 5.3 and 5.4. This will be given in Section 7.2.

We can now define $\mathbb{D}_{\text{lin}}$ for the spatial case as the set of truncated polynomials again, but now we have to include also the spatial test function in the specification (compare (3.31)).

The first task is to verify that we can choose the set $\widetilde{D}$ as the set of all functions of the form above, and as the set D_t the set of $t-$multiplicative functions on S_t the functions in (7.6).

This has to be lifted to D^* now. We see that we can lift the expression above easily by replacing $\widehat{\chi}$ respectively $\varrho_t^2(\widehat{\chi})$ by their associated function of the adjusted path, i.e. $(\widehat{\chi})^*$ resp. $(\varrho_t^2(\widehat{\chi}))^*$ shifted by $-T$ if T is the current time to obtain finally the truncated function:

$$F^*. \tag{7.7}$$

Remaining tasks. As we pointed out in order to lift the argument in Section 4, 5 to the spatial case we have given the recipe how to lift test functions which are truncated. What remains to be done? The function F^* has to be shown to be $t-$*multiplicative* and then we need to verify the properties w.r.t. to the *wellposedness* of the martingale problem. We formulate these two points and the other needed extensions next to prepare the conclusion of the proof as given earlier with the concepts introduced in this Section 7.1, which amounts to carrying out the arguments for the claims we used above to verify the criterion in the sequel.

(i) This is again the $t-$additivity of the *truncated* polynomials, which also holds including now polynomials as function of the mark, which is *truncated* as well, but it will turn out that the point is that we still never sample points from different $2t-$balls which contribute since we work with $\phi \cdot \chi$.

(ii) We need that with $\widetilde{D}$ as test function we get a solution of the $(\Omega^{\uparrow,\text{anc}}, \widetilde{D})$-martingale problem and that we have uniquely determined laws of the $t-$tops of the solutions.

For **(ii)** the ideas are as follows. For the first point we need the spatial version of Lemma 5.2 in combination with a version of Lemma 5.1. The uniqueness property

follows again if we establish the marked version of Proposition 5.6. This means we have to define the *dual process* first.

Here we replace the coalescent by the spatial coalescent which is a $E-$marked partition valued process equipped with a *distance matrix* and a *vector* recording the *path of the "individuals"*, which are the initial partition elements containing one element. In the Feynman-Kac term we have the joint occupation time of partition elements sitting in the same sites of E.

The final point is to show the $t-$ multiplicativity of h_t, here we need the spatial version of Theorem (2.27) in Glöde et al. (2019). Finally we have to *calculate g and show its $t-$additive*, i.e. the spatial analog of Proposition 5.7.

All these extensions to the spatial case of statements given in Section 6.1 to the non-spatial genealogy-valued Feller diffusion will be given in Section 7.2 and 7.3. Once we have the generalized statements the argument given above is closed.

7.2. *Verifying the key Lemmata, Propositions in their spatial version.* We recall Section 7.1 where we saw how to apply our criterion requires to verify the spatial versions of the Lemmata and Propositions of Section 6.1. Start with some preparations needed involving some important observations concerning the form of the action of the generator on smoothly truncated polynomials, which is basic. Then it remains to show why the spatial versions of the key Lemmata and Propositions hold, and then we have to explain why the flow of arguments given in the subsections above can be modified appropriately. We also need to show the ancestral path-marked process exists and is unique characterized by a *well-posed martingale problem* both which we defer to the next subsection.

Generator action on truncated polynomials. The key point was to operate with the generator action on *truncated* polynomials and we had to develop a *smoothed* version of this operation to be able to apply stochastic calculus. We consider first the time-inhomogeneous version, i.e. marks $V = D^+$, and then pass to $V = D^*$.

(i) If we consider a polynomial it observes the state on a geographically finite window. Therefore the measure restricted to this set is a finite measure and two of the operators, the operators $\Omega^{\uparrow,\mathrm{grow}}, \Omega^{\uparrow,bran}_\xi$ act as before on a polynomial and also for $exp(-\Phi^{m,\phi,\chi})$ we have the same expressions we dealt with in the previous sections and therefore we can work with the smoothed truncation as before after carrying out the lifting of the functions to ones on $\mathbb{U}^V$. The only point is to control the total mass in the window of observation depending of course on the potentially *infinite* mass on all of E (on finite E or with finite initial mass nothing changes). Therefore we focus below mainly on the *new* operator $\Omega^{\uparrow,\mathrm{mig}}$ or $\Omega^{\uparrow,\mathrm{anc}}$ in particular if they are acting on *truncated* polynomials and on the *behaviour of the total masses* in the spatial finite window of observation.

This latter point however is a feature of the total mass process which in itself is a Markov process, namely the super random walk for which these issues are by now well understood and we refer here to the needed facts in the literature, see Liggett and Spitzer (1981), Greven et al. (2005) for details of such arguments which we don't copy here. Hence it remains the first issue.

The generator is acting on truncated polynomials as follows. Here the reader should recall the actions in the case without truncation, the point 2 below (3.38). The new element in the generator calculations compared to Section 5.2 is now the

fact that we want to work with *truncated* polynomials, *truncated in addition in the mark* rather than the truncated polynomials we had before. Since we have smoothed the truncation via the ϱ_t we can again use calculus and obtain now as additional terms the ones which arise from the operator $\frac{\partial}{\partial t}$ on the level of the explicit time coordinate with which the $\mathbb{U}^V-$valued process was augmented as well as $\frac{\partial}{\partial s}$ which is part of the generator of the path process which was also augmented with an explicit time coordinate. Now we get:

$$\Omega^{*,anc}\Phi^{n,\phi\varrho_t^1,\varrho_t^2(\widehat{\chi})}(t,\mathfrak{u}) = \tag{7.8}$$

$$\left[\sum_{\xi\in E}\sum_{1\le k<l<n}\frac{b}{\overline{\mathfrak{u}}_\xi}1_{\{v_k(t)=v_\ell(t)\}}\,(\underline{v})\,\Phi^{n,(\phi\varrho_t^1,\varrho_t^2(\widehat{\chi}))o\theta_{k,l}}(t,\mathfrak{u})\right]$$

$$+\,\Phi^{n,2(\overline{\nabla}_1\phi)\varrho_t^1,\varrho_t^2(\widehat{\chi})}(t,\mathfrak{u})$$

$$+\,\Phi^{n,\phi(\frac{\partial}{\partial t}\varrho_t^1),\varrho_t^2(\widehat{\chi})}(t,\mathfrak{u})+\Omega^{\uparrow,anc}\,\Phi^{n,\phi\varrho_t^1,\varrho_t^2(\widehat{\chi})}(t,\mathfrak{u}),$$

where $\overline{\nabla}_1$ is w.r.t. to the distance variables. It remains to write out the *last term* explicitly.

Since we have now the ϱ_t^2 smooth truncation this leads to a convex combination of such product functions resulting after the action of $\Omega^{\uparrow,\mathrm{anc}}$ in a convex combination of product functions. We also have to take into account the contribution to the generator of the time change due to the weight function leading to the operator we call B and which is again resulting in a polynomial. We have to write this out now in formulas.

We note that $\varrho_t^2(\widehat{\chi})$ is a function which is an integral over functions which are a product of factors of evaluations at *specific times* of the path. Through $s-$truncation some of these factors are now evaluated at different times. This concerns factors to be evaluated at times $t_k\wedge s$ which are replaced effectively by the path at the truncation level if t_k lies before the truncation time. Nevertheless this is a new function of the same type. Now the same operator A_k^* acts on this *new function $\widehat{\chi}'$*. We saw in (3.44) how the operator acts on functions of the path.

Similarly the time evolution leads to a change in $\varrho_t^2(\chi)$ by changing the applied weight function which results in a change of the mixing measure of the truncations of $\widehat{\chi}$ leading to an integral with weights $(\widetilde{\varrho}_t^2)'$, defining an *operator B*. Define $(\varrho_t^2)'$ as the operation where the weight function $\widetilde{\varrho}_t^2$ is replaced by $\frac{\partial}{\partial t}\widetilde{\varrho}_t^2$. We get then for the last term in (7.8):

$$=\sum_{k=1}^n(\Phi^{n,\phi\varrho_t^1,A_k^*(\varrho_t^2(\widehat{\chi}))}+\Phi^{n,\phi\varrho_t^1,B\varrho_t^2\widehat{\chi}})\ (\text{recall }(3.44)). \tag{7.9}$$

(ii) From this expressions above we have to pass now to the one with marks in D^* and identify the action of the operator of the mark evolution. This action arises from evolving the path on $D_{0,t}$ and then shifting it back by t.

This means after evolving for time t replacing $\varrho_t^2(\widehat{\chi})$ by $(\varrho_t^2(\widehat{\chi}))^*$ (recall here (7.3),(7.4)) where we introduce for a function χ on $D(\mathbb{R},E)$ a map $*:\chi\to\chi^*$ where the value of χ^* on $D(\mathbb{R},E)^*$ is the number obtained on $D_{-t,0}(\mathbb{R},E)$ as value after shifting the path by $-t$.

To clarify this we need to explain especially the action of $\Omega^{\uparrow,\mathrm{anc}}$ on truncated polynomials. Our goal is now to write out the formula for (dealing with the explicit time variable in the path process):

$$\Omega^{\uparrow,\mathrm{anc-sh}}\left(\Phi^{m,\phi\varrho_t^1,\varrho_t^2(\widehat{\chi}^*)}\right) = \Omega^{\uparrow,\mathrm{anc}}\left(\Phi^{m,\phi\varrho_t^1,\varrho_t^2(\widehat{\chi})}\right). \tag{7.10}$$

Now we have to adapt the expressions and act on elements of D^*. Observe however that the expression in the quoted equation gives the evolution of a function from $D_{0,t}$ truncated at the present time. This action is now shifted back by t onto elements of $D_{-t,0}$.

We can write the generator of the mark evolution as (7.10) plus shift in the form (note in the time-space process A_k^* does not explicitly depend on t and furthermore the operators B and * commute):

$$= \sum_{k=1}^n (\Phi^{n,\phi\varrho_t^1,(A_k^*(\varrho_t^2(\widehat{\chi})))^*} + \Phi^{n,\phi\varrho_t^1,(B\varrho_t^2\widehat{\chi})^*}) + \Omega_t^{\mathrm{shift}}\left(\Phi^{m,\phi\varrho_t^1,\varrho_t^2(\widehat{\chi})}\right) \tag{7.11}$$

$$\text{(recall (3.44,3.53)),}$$

where A_k^* arises from transitions at the $k-$th sampled individual including actions of migration, recall that ϱ_t depends on both distances and marks as written out explicitly. Note at this point we see already that the term is a polynomial, however for its properties we need the following explicit formula:

$$(A_k^*(\varrho_t^2(\widehat{\chi})))^* = (A_k^*\widehat{\chi})^* := A_k^{*,*}(\widehat{\chi}), \quad B\varrho_t^2(\widehat{\chi}) = (\varrho_t^2)'(\widehat{\chi}). \tag{7.12}$$

We need now that the rhs. of (7.12) "is a truncation", i.e. is zero if two sampled points are from different $t-$balls. This is the case since the smoothly truncated case contains in each term the factor which is zero for distances above the truncation level and note that this is not affected by the $-t$ shift.

Note that the form of the generator implies that acting on polynomials we have a polynomial again with a new function ϕ, χ.

Remark 7.2. Recall the approach given in Remarks 3.8, 3.13. Then we would get the following. The truncated marked monomial is defined via

$$\Phi_t^{n,\phi,\chi}(\mathfrak{u}) = \Phi^{n,\phi\cdot c_t,\chi\cdot i_t}(\mathfrak{u}), \ \mathfrak{u} \in \mathbb{U}^{D^*}, \tag{7.13}$$

for $t \geq 0$ and D^* as in (3.51). Here, c_t as in (5.4) and $i_t : V^n \to \mathbb{R}$ via

$$i_t(v_1,\ldots,v_n) = \prod_{j=1}^n \mathbb{1}(v_j(s) = v_j(-t) \ \forall s < -t). \tag{7.14}$$

It is easy to see that monomials of the form (7.13) are additive on $S_t = \mathbb{U}^{D_t^*}$.

Spatial version of Lemma 5.1. Reviewing the proof in the non-spatial case, we see that we have to explain how to modify the function w from (5.2). Here the new element is the term arising via the generator of the *mark evolution* for which we have to give the argument. However this term is again a drift term i.e. a *first order operator*. Hence we get an additional term appearing in the expression, which however gives in the expression for w just another truncated polynomial of the order n as we saw above in (7.11) and (7.12).

Next we have to replace the *moment calculations* for the Feller diffusion by those for the super random walk to account for the fact that polynomials are not bounded. We have to obtain bounds on the expectations of polynomials of order n to conclude they remain finite after finite time and to show that they still determine laws in spatial models. These calculations are well known in the literature, see

Dawson (1993). If we want moments *of all orders* for positive times which are law determining we need stronger restrictions then just requiring a finite Liggett-Spitzer norm , namely $\sup_n \left((n!)^{-1} \sum_{i \in \Omega} (\overline{u}_i)^n \gamma_i \right) < \infty$ (recall here (3.30)), or we have to pass to *local* martingale problems on the Liggett-Spitzer space. Then the argument proceeds as before, we omit the standard details.

Therefore the argument goes through again in the spatial case.

Spatial version of Lemma 5.2. We deal now with test functions $\psi(t) \exp(-\Phi^{m,\phi \varrho_t^1,(\varrho_t^2(\widehat{\chi}))^*})$, $(\psi \in C_b^1(\mathbb{R}, \mathbb{R}))$ incorporating now the time derivate operator as well, recall (5.15) and (5.16). The only change is that on the rhs. of (5.17) now the generator of the mark motion appears as well besides $\Omega^{\uparrow,\mathrm{grow}}, \Omega^{\uparrow,\mathrm{bran}}$ and $\frac{\partial}{\partial t}$ which give the contribution exhibited in (7.8) and (7.11), (7.12).

The calculation from (5.15) - (5.23) carries over once we understand the new term namely the operator $\Omega^{\uparrow,\mathrm{anc}}$ since the other terms are truncated polynomials and are $t-$ additive.

Now the new term, which as a first order operator lowers the order by one, but since we have coefficients which are linear, i.e. first order monomials, we get a polynomial of the same order back again, where coefficients are given via the $a(\cdot,\cdot)$, see here formula (7.11) combined with (7.12) where this term is analyzed in detail. Then the argument works as before.

Spatial version of Lemma 5.5.

Lemma 7.3 (Polynomials: approximation by truncated polynomials)**.** *Let $t > 0$. Suppose Assumptions 5.3 and 5.4 hold for functions $\varrho = \varrho^{(n)} : \mathbb{R}^{\binom{n}{2}} \to [0,1]$, $n \in \mathbb{N}$. Then for any truncated polynomial $\Phi_t^{n,\phi} \in \Pi(C^1(\mathbb{D}_n))$, we can find a sequence of polynomials $(\Phi^{n,\phi_N \varrho_t,(\varrho_t^{2,N}\chi)^*})_{N \in \mathbb{N}}$ in the family of polynomials $\{\Phi^{n,\phi \varrho_t,(\varrho_t^{2,N}\chi)^*} : \phi \in C_b^1(\mathbb{D}_n, \mathbb{R}^+), \chi \in C_b(V^*, \mathbb{R}^+), n \in \mathbb{N}\}$ such that for all $\mathfrak{u} \in \mathbb{U}$:*

$$\Phi^{n,\phi_N \varrho_t,(\varrho_t^{2,N}\chi)^*}(\mathfrak{u}) \nearrow \Phi_t^{n,\phi,\chi}(\mathfrak{u}), \ as \ N \to \infty. \tag{7.15}$$

For a proof we have to deal only with the marked part, i.e. the approximation of the function χ by its smooth truncations. This is standard and follows from the right continuity of the path.

Spatial version of Proposition 5.7. Recall the formula for the homomorphism g_f of our criterion in the non-spatial case in (5.7),(5.8). The change we need to provide is the new form of the g_f of the criterion which is now a $g_{\phi,\widehat{\chi}}$ and of which we have to show that it is $t-additive$. We have here (recall that this object here depends on $(\phi \varrho_t^1, (\varrho_t^2(\widehat{\chi}))^*))$, but we suppose this in the formula:

$$g(t,\mathfrak{u}) = \Omega^{\uparrow}\Phi^{n,\phi \varrho_t^1,(\varrho_t^2(\widehat{\chi}))^*} + \Omega^{\uparrow,\mathrm{anc}}\Phi^{n,\phi \varrho_t^1,(\varrho_t^2(\widehat{\chi}))^*}, \ where \ \Omega^{\uparrow} = \Omega^{\uparrow,\mathrm{grow}} + \Omega^{\uparrow,\mathrm{bran}}. \tag{7.16}$$

We have to see here still that the second term on the rhs., which is the new term, is $t-additive$. Additivity follows however, since $\Omega^{\uparrow,\mathrm{anc}}\Phi^{n,\phi \varrho_t^1,(\varrho_t^2(\widehat{\chi}))^*}$ is a *mixture* of *truncated polynomials*. Namely in detail we show:

(i) we have again that truncated polynomials are additive and

(ii) we can wright out the second term explicitly to see it is a truncated polynomial.

This has been proved in Glöde et al. (2019) in Theorem 2.27 for the case without marks. The point is that a sample only contributes if all sampled points are in the same open $2h-$ball, which *still holds with marks* since the monomial is defined by $\phi \cdot \chi$.

7.3. *Basic properties of the martingale problem: Proof Proposition 3.9, spatial version of Proposition 5.6 (FK-duality).* Crucial for our criterion is the characterization of the process via an *operator specifying a well-posed martingale problem.* One issue which we treat first is the *uniqueness* for the solution of the martingale problem using duality, then we discuss the *existence* of a solution and prove Proposition 3.9, which in particular requires proving the spatial version of Proposition 5.6 establishing the Feynman-Kac duality.

For the required *uniqueness* we have to start as first point by introducing the *dual process* and proving a *duality relation* in the spatial and path marked context, which requires to pass to a spatial coalescent on E and to augment it for the path valued case by a vector describing the paths of the sampled individuals taken up to the present backward time.

The second point we have to deal with the question of the *existence* of a solution for the martingale problem which is well known (see Depperschmidt and Greven, 2019) except for the *path* marked case which we therefore discuss in more detail below. Existence follows via *approximation* by individual based models as we point out below however it is *not* standard for marks from $D(\mathbb{R}, E)$ and that is why we give more details for this point.

Step 1: Uniqueness and FK-duality We shall now argue that we have again a duality relation and the *analog of Proposition 5.6* holds in the marked case. A first important observation is that since migration and branching occur independently and the generator consists therefore of a sum of three operators growth and branching we had before already and each allowing a duality and in addition the *evolution of the mark* for which we will establish duality below. If we now take as test functions polynomials which are based on test functions on $\mathbb{D}_n$ and V^n in *product form* and we get nice expressions for the generator such that for each of these operators we can establish duality. For that reason we get again a *Feynman-Kac duality relation* for this spatial model. Here are the details.

We give now first explicitly the *dual process*, then the *duality function* for first the $E-$ marked and then the path-valued case and finally verify that the *duality* holds. All is based on the *generator criterion for duality* again. Recall the notation and setup for duality in Section 5.3, where the duality is derived for the case without marks. We distinguish two cases $V = E$ and $V = D$ or D^*.

(i) Case $V = E$.

Begin with the *dual process* which is based on an *E-marked partition valued* process enriched by a *distance matrix.* Here the partition elements are each marked by an element of E.

We specify the *initial state* as follows. Start the dual process with m individuals. The initial distance matrix between these individuals is $\underline{r}' = \underline{0}$. Furthermore by fixing a set of m *initial locations*, which might be assigned in multiplicity, $\underline{\xi} = (\xi^1, \cdots, \xi^m) \in \Omega^m$ we determine a function χ for the location-marked case, as in (3.32)-(3.33). Furthermore choose $\phi \in C_b(\mathbb{D}_n, \mathbb{R})$.

Then the dynamic is as follows, partition elements *migrate* with the kernel a from (3.29) and *coalesce* if they share a site. The elements of the distance matrix grow at rate 2 as long as the two respective elements are in *different* partition elements. This specifies the dual process

$$(\mathfrak{C}_t)_{t \geq 0} , \quad \mathfrak{C}_t = (p_t, r'_t, \underline{\xi}_t) \text{ for } V = E. \tag{7.17}$$

The *duality function* is given as

$$H^{n,\phi,\chi}(\mathfrak{u},(p,\underline{r}',\xi)) = \int\limits_{(U\times V)^n} \phi(\underline{r}^p + \underline{r}')\chi(\underline{\xi})\mu^{\otimes n}(d(\underline{u},v)), \qquad (7.18)$$

where $\underline{r}((\underline{u},v)) = (r(u_i,u_j)_{1\le i<j\le n})$ and $\underline{\xi}(\underline{u},v) = \underline{v}.$

Finally we need the *Feynman-Kac potential* suitably modified in the spatial case. Here we count the occupation time of partition elements *at the same site*. Denote by $\#p_{s,\xi}$ the number of partition elements at time s in ξ. Then the *Feynman-Kac potential* on the time interval $[s,t]$ is

$$\beta_{s,t} = \int\limits_s^t \sum_{\xi \in E} \left(\binom{\#p_{u,\xi}}{2}\right)du. \qquad (7.19)$$

Then we have the following *FK-duality relation*:

$$E_{\mathfrak{u}_0}\left[H^{n,\phi,\chi}\left(\mathfrak{U}_t,(p,\underline{r}',\underline{\xi})\right)\right] = E_{\mathfrak{k}_0}\left[H^{n,\phi,\chi}\left(u_0,(p_t,r_t',\underline{\xi}_t)\right)\exp(\beta_{0,t})\right] \qquad (7.20)$$

To *determine the f.d.d. via the dual*, we get an expression based on the *time-space dual* giving the f.d.d. formula of the E-marked process based on the Markov property following Fleischmann and Greven (1994), Greven et al. (2016) page 13 which we formulate next.

We consider now the *time-space process* of $\mathbb{U}^E$-valued super random walk and consider the so called *time-space coalescent* on E to derive the following duality formula determining the f.d.d. of the original process.

Consider time points $0 \le t_1 < t_2 < \cdots < t_\ell = t \quad, \ell \in \mathbb{N}$. Then consider the following functional of the forward process:

$$\widetilde{\Phi}\left((\mathfrak{U}_t)_{t\ge 0}\right) = \prod_{k=1}^\ell \Phi_k(\mathfrak{U}_{t_k}) \qquad (7.21)$$

where $\Phi_k \in \prod \quad, k = 1,\cdots,\ell$.

Next consider the spatial coalescent with *frozen* partition elements, $(\mathfrak{K}_s)_{s\in[0,t]}$.

Here we start with partition elements in the time-space points $\left((t_k,i_k^j)_{j=1,\cdots,\ell_k}\right)_{k=1,\cdots,\ell}$ and the partition elements in $i_k^j, j = 1,\cdots,\ell_k$ will be *frozen* for the dual evolution till times $t - t_k$, for $k = 1,\cdots,\ell$. (Precisely: the unfrozen, called *active* partition elements, evolve as before as spatial coalescent. Similarly the Feynman-Kac potential at time s of the backward evolution does include only the *active* partition elements at that time, i.e. $\#p_{s,\xi}$ is the number of *active partition elements*.

Similarly the distances between frozen particles is zero and between a frozen and active one is initially zero and grow at speed 1 till they are both active and the growth is 2.)

Define now

$$\phi(\underline{r}) = \prod_{k=1}^\ell \phi_k\left((r_{i,j})_{\ell_1+\cdots+\ell_{k-1}<i<j\le\ell_1+\cdots+\ell_k}\right), \qquad (7.22)$$

$$\chi\left(\underline{t},\underline{v}^\pi\right) = \prod_{k=1}^\ell g_k(t_k,\underline{v}^{\pi_k}),$$

which characterizes the polynomial $\widetilde{\Phi}$ in (7.21). We see that $\widetilde{\Phi}(\mathfrak{U}_t) = H(\mathfrak{U}_t, \mathfrak{K}_0)$ with H based on K_0 and $\widetilde{\Phi}$ as before for the classical objects.

Now following the argument for the Fleming-Viot case in Fleischmann and Greven (1994), Greven et al. (2016) we obtain the formula relating the *time-space forward* and the *time-space dual* process namely we get with expectation on the l.h.s. with respect to $\mathfrak{U}$ on the rhs. will respect to $\mathfrak{K}$:

Duality formula for f.d.d.

$$E\left[\widetilde{\Phi}\left((\mathfrak{U}_t)_{t \in [0,T]}\right)\right] = E\left[H\left((\mathfrak{U}_s)_{s \le T}, \mathfrak{K}_0\right)\right] = E\left[H(\mathfrak{U}_0, \mathfrak{K}_T)\exp(\beta_{0,T})\right]. \qquad (7.23)$$

Since we have here a FK-duality we have to argue here that the Feynman-Kac term behaves as claimed. However the argument works with the Markov property and applying the duality relation to the time pieces between the $(t_k)_{k=1,\cdots,\ell}$ so that this can be imitated with the FK-duality the result follows from the fact that the Feynman-Kac term is an additive functional.

(ii) Case $V = D([0,\infty), E)$ and $V = D^*$.
For the path-valued case we have the *path marked spatial* respectively *time-space coalescent* which is the following modification of the above. Only the *mark and their evolution is different* compared to the dual process in case (i) above. We let t be the time horizon for the duality relation. This means the original process started at time $s < t$ will evolve till time t. The dual process will evolve from time t backwards till time s, its time of evolution runs therefore for time $t - s$.

We focus on the time-space coalescent including the "simple" one. We enrich the coalescent analog to the $V = E$ case. In addition to the *distance matrix* $\underline{\underline{r'}}$ we record now the *vector of the paths of locations* $\underline{\xi} = \left(\left(\xi^1(u)\right)_{u \in \mathbb{R}}, \cdots, \left(\xi^m(u)\right)_{u \in \mathbb{R}}\right)$ of all initial individuals recall here the description around (7.21), with $m = \sum_1^\ell \ell_k$, which also enters the duality relation the same way as in the forward evolution. The coalescent is denoted $\widetilde{\mathfrak{C}}$ and the time-space coalescent is denoted $(\widetilde{\mathfrak{K}}_s)_{s \ge 0}$.

Note that here we keep the path of descent of every of the initial individuals, even though they may be piecewise joint path beyond some backward time. Note that the dual path evolves backwards from t to s, rather than forward from s to t. On the other hand the input in the duality function is, as we shall see, the same over the full interval $[s, t]$.

Next we come to the *duality function*. To write down the duality function we need the sampling measure restricted to the population at a site, called μ_i and given by

$$\mu_i(A) = \nu(A \times \{i\}), A \in \mathcal{B}(U). \qquad (7.24)$$

Now we can define for $t > s \ge 0$ the time-space *duality function* for that path marked case for the situation where we are starting in s in constant path and distance matrix r.

For the duality function we chose again a *number of individuals m* in the dual process, the vector $\underline{\xi}$ of their initial path determined by an m-tuple of locations $\underline{\xi}$, a function $\Psi \in C_b(\mathbb{R}^2, \mathbb{R})$ with $\Psi(t,s) = \widetilde{\Psi}(t - s)$ with $\widetilde{\Psi} \in C_b(\mathbb{R}, \mathbb{R}), \widehat{\chi} \in \mathcal{B}(\mathbb{R} \times D(\mathbb{R}, E^m), \mathbb{R})$ and then *a function $\phi \in C_b(\mathbb{D}_n, \mathbb{R})$*.

Then set for $\mathfrak{u} \in \mathbb{U}$ and $s,t \in \mathbb{R}$ with $s < t$ and $\widetilde{\mathfrak{C}}_t = (p,\underline{r}',\underline{\xi}',t)$:

$$H^{\phi,\psi,\underline{\xi},\widehat{\chi}}\left((\mathfrak{u},s),(p,\underline{r}',\underline{\xi},t)\right) = \qquad\qquad (7.25)$$

$$\psi(s,t) \int\limits_{(U\times V)^n} d\left(\left(\bigotimes_{i=1}^n \mu_{\xi_i(s)}\right)(d\underline{x})\right) \phi\left[(r^p(x_i,x_j))_{1<i<j\leq n} + (r'(i,j))_{1\leq i<j\leq n}\right]$$

$$\widehat{\chi}\left(s,(\xi(u))_{s\leq u\leq t}\right).$$

This amount to having a χ-function with *two* factors one as in (7.18) generating the locations where we sample with the sampling measures the individuals from the population and a second factor to explore the corresponding path at different time points.

Then the *Feynman-Kac duality relation* reads (with $s = 0$):

Lemma 7.4 (FK-duality:path process)**.** *We start the process $\mathfrak{U}$ in a state with constant path. Then:*

$$E[H^{\phi,\psi,\underline{\xi},\widehat{\chi}}((\mathfrak{U}_t,t),(p_0,r_0',\underline{\xi}_0,0))] = E[H^{\phi,\psi,\underline{\xi},\widehat{\chi}}((\mathfrak{U}_0,0),(p_t,r_t',\underline{\xi}_t,t))\exp(\beta_{0,t})](7.26)$$

However in order to obtain above duality later via the generator relation some additional concepts are needed, due to the time *in*homogeneity. These operators corresponding to the processes above involve the explicit time coordinate and therefore we will need a generator relation for all times s between 0 and the time horizon t and hence we have as state of the forward process path which already evolved for some time u and are not in the constant state, which we assumed writing down $H(\cdot,\cdot)$ in a specific way. In fact we have to be more careful writing down the duality function, so that we can use it for the intermediate times s between 0 and the time horizon t. In fact we just saw that we have to generalize this a bit now.

Consider the following objects. For two path $\xi^\uparrow$ and $\xi^\downarrow$ one from $D_{s,u}$ one from $D_{u,t}$ we introduce, for those ones with $\xi^\uparrow(u) = \xi^\downarrow(u)$ the *glued path* $\xi^{*,u} = \xi^\uparrow \vdash \xi^\downarrow$ from $D_{s,t}$ arising by setting

$$\xi^{*,u}(r) = \begin{cases} \xi^\uparrow & \text{for} \quad r \leq u \\ \xi^\downarrow & \text{for} \quad r > u. \end{cases} \qquad (7.27)$$

This is used to generalize the duality as follows. Now the duality function between time s and t given by replacing the path ξ in the formula (7.25) by $\xi^{*,s}$ from above.

We need above relation also for the $t-truncated\ process$, including the smoothed versions. For this purpose the duality function has to be changed by replacing in (7.25) the function ϕ and $\widehat{\chi}$ as follows.

For the truncated case we replace ϕ by the truncated ϕ which is $\phi 1_{[0,t]^n}$ respectively its smooth version. The function $\widehat{\chi}$ is replaced by the function of the truncated path respectively the smoothly truncated one. We have to see to it below that the duality still holds on these truncated test functions by approximation (spatial version of Lemma 5.5 see (7.15)).

Proof of Lemma 7.4: We have established the needed generator relation for the FK-duality for the non-spatial and time-homogeneous case. We note that passing to the time-space process is fine since $\frac{\partial}{\partial t}$ and $-\frac{\partial}{\partial t}$ are *operators in duality* with our condition on ψ.

Consider next the E-marked case. Then the argument for growth and branching operator is easily lifted from U to $U \times V$. Therefore it remains to verify the duality criterion now only for the *migration operator*. Hence we have to show that the generator of the mark evolution is in duality with the one for the dual process, the spatial coalescent enriched with the distance matrix and the mark vector which is a vector of locations. But this is in the literature. For details for the mark part of the operator note that this is the same as in the case of the Fleming-Viot process, see therefore Remark 1.19 in Greven et al. (2016).

Similarly in order to now establish the Feynman-Kac duality for the *path marked* case we have to show that the mass flow on path is dual to the migration in the path in the $a(\cdot, \cdot)$-spatial coalescent. Namely in order to check the duality in that case one uses the generator criterion and as test functions polynomials of the form as given in (7.25) with the mentioned restriction.

Here one might wonder whether this is not just the duality of the time-space processes from above. That is unfortunately a bit more tricky as we explain next.

Remark 7.5 (Time-space process for E-marked versus D-marked process). If we observe the state at time t of the *path valued* process $\mathfrak{U}$ and evaluate the specified polynomial on the one hand and compare it with the functional of the path of the E-*marked* process with the specified test-functions are similar if in the latter we consider the time-space process but there is the following difference.

We must observe that in the E-*marked time-space process* we sample from the population at times $t_1, \ldots, t_m, t$ and observe the position at this time say t_k only, while for the path marked case such positions appear also for individuals sampled at later time t_i. In particular in the path marked case we sample from the population at time t_i, so that we prune individuals at those time t_k which do not have descendants at time t_i. Note however that the pruning is independent of path and genealogy *up to time t_i*. This means that the ones sampled at a time t_i but whose path is evaluated at positions $t_k < t_i$ must in the time-space coalescent in the backward picture be activated at time t_i and not t_k that is in the backward time at $t - t_i$ and not $t - t_k$. Similarly the Feynman-Kac term changes.

$\square$

Next we have to argue how to get the duality relation respectively the analogue of (7.23) for the *path-marked* process, first in the time-inhomogeneous setting in (i) then for D^* in (ii).

(i). Denote coalescent and functional by $\widetilde{\mathfrak{K}}$ respectively $\widetilde{\beta}_{0,t}$. We have to address two points (1) How is duality transfered from the mass flow of locations to the mass flow of the *path process* and (2) how this then is transfered to the *measure-valued historical process* i.e. the process of measures on these path and even further from here to $\mathbb{U}^V$-valued processes. How to prove the claim?

To get the duality we have to prove essentially first ((1) and the first part of (2)) that the mass flow induced by the historical process and a system of independent random walk *path* processes are dual w.r.t. $H(\cdot, \cdot)$ from (7.25) using the generator criterion. Some care is needed here due to the time inhomogeneity as we will see below.

To verify the generator criterion for duality, we note that the generator action of $\Omega_s^{\uparrow,\mathrm{anc}}$ (recall this is exclusive the action in the explicit time coordinate) is on the polynomial via the function χ. Here we have to recall the operators defined in

(3.42) and (3.44). Namely for a polynomial of degree n, $\Phi^{n,\phi,\widehat{\chi}}$ we have:

$$\Omega_s^{\uparrow,\mathrm{anc}}\Phi^{n,\phi,\widehat{\chi}} = \sum_{k=1}^{n} \Phi^{n,\phi,\widehat{\chi}^{*,k}}, \tag{7.28}$$

where (recall here (3.44))

$$\widehat{\chi}^{*,k} := A_k^* \, \widehat{\chi}. \tag{7.29}$$

Note here that the rhs. involves s. This expression we have to compare with the action of the *dual process generator* on a polynomial.

We now need the action of the mark evolution of the dual process on the duality function for fixed first argument and how this acts on the test function. To see this we have to calculate the action of the generator of the random walk path process of one moving individual on $H^{n,\phi,\chi}(\mathfrak{u};\cdot)$. This latter process is a pure jump process and the generator action we look for is the *sum* of the n one–individual generators which is given in equation (3.42). We have to argue now what this formula implies if we apply it to the function given via the duality function above. For that purpose we have to view this function as a functional of the function $\widehat{\chi}$ of the path on which we can act directly with the generator. We claim this object we can write again as a polynomial with a *new* $\widehat{\chi}$-function.

To see this we need to analyze what type of function $H(\mathfrak{u},\cdot)$ actually is. This function is an integral over parameters, namely $u \in U$ which is derived from $\mathfrak{u}$, of polynomials of the form (recall (3.24)):

$$\int_{V^n} \widehat{\chi} \, d\nu_*^{\otimes n}. \tag{7.30}$$

In particular does it suffice to have the self-duality in terms of the operation acting on $\widehat{\chi}$, on the function of one path. Here the dual random walk path goes in the revers direction as the underlying motion hence we have the same $a(\cdot,\cdot)$ as in the mass flow term in the operator and we have here therefore the same coefficients for the action on $\widehat{\chi}$.

Hence the dual migration operator acts only on the function $\widehat{\chi}$ and through it on the polynomial $\Phi^{n,\phi,\widehat{\chi}}$ and the operator $\Omega^{\downarrow,\mathrm{anc}}$ does this acting on the polynomial via

$$\Omega^{\downarrow,\mathrm{anc}}\Phi^{n,\phi,\widehat{\chi}} = \sum_{k=1}^{n} \Phi^{n,\phi,\widehat{\chi}_k^{*,\downarrow}}, \tag{7.31}$$

where

$$\widehat{\chi}_k^{*,\downarrow} := A_k^* \, \widehat{\chi}. \tag{7.32}$$

We see that $\widehat{\chi}_k^* = \widehat{\chi}_k^{*,\downarrow}$ which proves the duality of the mass flow.

This proves the Feynman-Kac duality relation for the path valued process. In particular we get also a formula for the finite dimensional distributions i.e. using instead of the mechanism and the functional $\beta_{0,t}(\cdot)$ the one of the dual process $\widetilde{\mathfrak{K}}$ and $\widetilde{\beta}_{0,t}$ as specified above then the analog of (7.23) holds replacing $\mathfrak{K}, \beta$ by $\widetilde{\mathfrak{K}}, \widetilde{\beta}$ in the formulas .q.e.d.

Therefore the uniqueness result for the martingale problem carries over to the D^+-marked case.

(ii). In the final step we have to *lift* this FK-duality now to the marks given by the *adjusted* path, shifted by t to the left for the time t duality i.e. we have to pass to D^*-marks.

Here we apply the shift by t to the path of the dual process (which runs backward and is shifted by t further) i.e. in the duality relation we have to take for a dual path ζ running backwards from time t the adjusted path (with a fixed choice of t):

$$\zeta(\cdot) \longrightarrow \zeta(\cdot + t) := (\zeta)^{*,\downarrow}(\cdot) \tag{7.33}$$

in the duality relation. We need that this does not change the two sides of the duality relation for the *chosen* function. This is true since we shift by t the path on both sides.

Altogether we have now verified the assumption needed to get the uniqueness of the martingale problem, which is needed in the verification of our criterion for the branching property.

Step 2: Existence Next we need *existence* of a solution. In order to prove the existence of a solution to the martingale problems we proceed in principle in two parts. One considers in the first part a solution to the martingale problem on a *finite* geographic space with the "induced" dynamic from the larger space and obtains then in the second part a solution for the possibly infinite space by approximation via a sequence of processes constructed for embedded *finite* geographic spaces namely for $E_n \uparrow E$, with $|E_n| < \infty$. Here we use a suitably chosen *modified migration part* of the dynamic. This approximation argument is standard for measure-valued processes and carries over to the $\mathbb{U}^V$-valued setup making use of the duality for the limit evolution for the latter, see Depperschmidt and Greven (2019), where all details are spelled out. Hence it remains to show here the existence of a solution on *finite* geographic space.

Here, on the finite geographic space (and *finite total population size*), one uses the *approximation* with an *individual based model*. One takes per site a finite number of individuals and then lets this number tend to infinity, see Greven et al. (2009), Greven et al. (2016) where this is carried out for the Fleming-Viot model and Glöde (2012) where the branching model is treated in the non-spatial case, the spatial case in Depperschmidt and Greven (2019). We refrain from giving details in this paper *except for the path-marked situation* where a *new* point arises.

Namely we want to be able to approximate our process by a sequence of individual based branching models. In order to then show that limits of the approximating models solve the martingale problem it is most convenient if we have *continuous* test functions. This is here a problem with *path-marked* genealogies, the spatial component in the polynomials is in that model *not continuous* for these polynomials we use as test functions in the martingale problem, namely in the path-valued case our polynomials are based on evaluations in *fixed time points* and hence are *not continuous* in the mark variable in the Skorohod space of path.

We can obtain continuous test functions by considering moving averages of the functions χ which we are using, more precisely choosing the time points there at random on $\mathbb{R}$, recall 3.8. (This we also use for the topology to define convergence.) Because of the linearity of the operator $\widehat{A}$ we can then obtain $\Omega^{\uparrow,\mathrm{anc}}$ for these test functions as well, but now the test functions are continuous. Furthermore solutions to the martingale problem are also solutions to the one with these test functions,

but *not immediately the other way around.* We must therefore directly work with the given martingale problem.

We can use two known properties here. Namely for the convergence of the projection of the sampling measure on the marks, which yields the historical process, we can use the existence and uniqueness results of Section 7 in Dawson and Perkins (1991). On the other hand projecting on $\mathbb{U}$ by projecting $U \times V$ on U gives for finite space the $\mathbb{U}$-valued Feller diffusion, where we can use Depperschmidt and Greven (2019), Glöde (2012). These two facts are given in detail below and can be used below to establish the convergence of the genealogies of the subpopulations descending from ancestors in some finite subset of geographic space and then it is well known to lift this to the full population as explained above.

The *tightness* of the approximating individual based processes follows, since tightness of genealogies and marks separately implies the *tightness* of the *joint law*, Theorem 3 in Depperschmidt et al. (2011). However we gave the reference for these two points above. Hence we need to prove only *uniqueness* of the *limit points* to prove *convergence.* We know our martingale problem has a unique solution by duality. Hence we get convergence if we show that the limit points of our tight sequence *must solve the martingale problem.*

The convergence to solutions of the martingale problems which are well-posed are known both for the projection on $(\nu_t^*)_{t \geq 0}$ and for the projection on the genealogy. Hence it only remains to deal with the *coupling* of the two components i.e. the *joint distribution* of genealogy *and* marks to conclude convergence as we saw above.

The process has test functions which are polynomials induced by a product of a function ϕ of the distances and $\widehat{\chi}$ a function of the marks. We have by results on the non-spatial case respectively the historical process (Glöde, 2012, Dawson and Perkins, 1991) the convergence of the compensators to the limit compensator if we put either ϕ or $\widehat{\chi}$ equal to a constant. We also know that the operators associated with the genealogy and the mark evolution sum to the complete operator. The branching operator acts on the function ϕ as well as the growth operator, the mark evolution operator on χ but the branching operator in the limit dynamic acts also on χ through the action of duplicating a path i.e. its action on the measure. The latter is of course what connects the evolution of marks and of genealogies here in this model which is for the path marked case not just affecting mass at the site. Still this gives rise to technical problems.

We will circumvent this problem and first prove *convergence* to some limit *without* using uniqueness of the martingale problem. We will see from the proof that this limit point process is a Markov process. We also use that we have tightness in the space $D([0, \infty], \mathbb{U}^V)$ so that it suffices to show f.d.d.-convergence, to have convergence in the path-space for the process. Then we will have to show that this process solves the given martingale problem using the way it is *constructed.* Since we have the distribution of the genealogy part and the mark part converging we obtain the joint distribution by giving a construction of the *conditional distribution* of the *marks given the genealogy.* This we can do by using the so called trunks of the state (recall Figure1) which we introduce below.

For a given $\mathfrak{U}$ we denote by $\widetilde{\mathfrak{U}}$ the projection from the marked genealogy on the genealogy, i.e. we consider the map $\tau : \mathbb{U}^V \to \mathbb{U}$ given by $[U \times V, r \otimes r_V, \nu] \to [U, r, \mu]$ with $\mu(\cdot) = \nu(\{\cdot\} \times V)$ and set

$$\widetilde{\mathfrak{U}} = \tau(\mathfrak{U}). \tag{7.34}$$

Then $\nu = \mu \otimes m$ with m a kernel from $\mathbb{U}$ to V describing the mark configuration given the genealogy, the latter means, we condition on the value of this functional τ to condition on the process of genealogies. Hence the object we have to focus on is the $\mathbb{U} \times \mathcal{M}_{\text{fin}}(D(\mathbb{R}, E))$-*transition kernel* m.

Fix for the state at time t an $h \in [0, t)$. We consider here the *marked* $(t - h)$-*trunks* $\lceil \mathfrak{u} \rceil (t - h)$ of the states $\mathfrak{u}$ at time t to show that the marginal distribution at times t converge, by showing that the law of the *marked* $(t - h)$-trunks *conditioned* on the *genealogy* which is a law on the mark distribution converges for all $h \in [0, t)$. We recall next (resp. define here) $(t - h)$-trunks and *marked* $(t - h)$-trunks.

The $(t - h)$-trunk of an element $\mathfrak{u}$ of $\mathbb{U}$ is an element of $\mathbb{U}(t - h)^{\sqcup}$ (recall (3.3)) which complements the $\lfloor \mathfrak{u} \rfloor (t - h)$-top of the $\mathfrak{u}$, see (3.8) and as $h \uparrow \mathfrak{u}$ the trunks converge to $\mathfrak{u}$. Here is the definition.

For $\mathfrak{u} \in \mathbb{U}(t)^{\sqcup}$ *consider* the $(t - h)$-top denoted $\lfloor \mathfrak{u} \rfloor (t - h)$ and write $\lfloor \mathfrak{u} \rfloor (t - h) = \bigsqcup_{i \in I}^{t-h} \mathfrak{u}_i$, the decomposition into open $(t - h)$-balls $\mathfrak{u}_i = [U_i, r_i, \mu_i]$ for $i \in I$. Then the $(t - h)$-*trunk* of $\mathfrak{u}$ is defined:

$$\lceil \mathfrak{u} \rceil (t - h) = [I, r^*, \mu^*], \text{ with } r^*(i, i') = \inf\left(r(u, v) - 2(t - h) \mid u \in U_i, v \in U_{i'}\right) \tag{7.35}$$

and weights

$$\mu^*(\{i\}) = \mu_i(U_i). \tag{7.36}$$

Next the *marked* $(t - h)$-trunk. We decompose the $(t - h)$-top of the marked state $\mathfrak{u}$ in the elements of $\mathbb{U}^D(t - h)$. Now we proceed as above but the measures ν_i are now projected onto the set of path which are only non-constant in $[0, h]$, i.e. $v^h(s) = v(s)1_{[0,h]}(s) + v(h)1_{[h,\infty)}(s)$. This gives us an element in $\mathbb{U}^D$. Furthermore we have again that the marked $(t - h)$-trunk of $\mathfrak{u}$ namely $\lceil \mathfrak{u} \rceil (t - h)$ satisfies $\lceil \mathfrak{u} \rceil (t - h) \to \mathfrak{u}$ as $h \uparrow t$.

We begin by constructing the announced candidate for the limit dynamic.

Construction. We know the process for the genealogy exists (i.e. the state projected on $\mathbb{U}$ from $\mathbb{U}^V$), so we have to *construct* the *conditional law of* $\mathfrak{U}$ *given the genealogy*. We identify first the conditional law of the marks at time t of the *marked* $(t - h)$-trunk *given* the $(t - h)$-trunk. Note here that conditioning on the $(t - h)$-trunk is here equal to conditioning on the genealogy up to time t and then passing to the $(t - h)$-trunks, because of the Markov property and the independent increments of the path process. For $h < t$ the h-trunk can be represented by an $\mathbb{R}$-tree with *finitely* many leaves, so that the (U_t^h, r_t^h) is a finite ultrametric space, where the leaves carry also $\mathbb{R}^+$-valued weights namely the mass of the time t descendants (see Glöde et al., 2019) so that we get an element of $\mathbb{U}$, denoted $[U_t^h, r_t^h, \mu_t^h]$. The corresponding $\mathbb{R}$-tree is represented as a tree with finitely many vertices, binary split points and edges and leaves carrying a weight. The $\mathbb{R}$-tree has a finite number of founding fathers, $F_1, F_2, \cdots, F_K$. The set of leaves with the distances and weights determine an equivalence class of ultrametric measure spaces, $\lceil \widetilde{\mathfrak{U}} \rceil (t - h)$. We represent conditional distribution of the marks of the trunk, by constructing a version explicitly for every realization of $\lceil \widetilde{\mathfrak{U}} \rceil (t - h)$ more generally for every finite element in $\mathbb{U}(t - h)^{\sqcup}$. For the $t - h$-trunk of a $\mathbb{U}$-valued Feller diffusion and via the backbone construction carried out in our framework, see Depperschmidt and Greven (2019) for more details on the diffusion case, while in the particle case one can use results of Chauvin, Rouault and Wakolbinger (Chauvin et al., 1991).

We now generate the *marked* object as follows. We have a collection of independent $\bar{a}$-random walks starting in $e \in E$:

$$\left\{ \left(Y_t^{\{e\},k} \right)_{t \geq 0}, e \in E, k \in \mathbb{N} \right\}. \tag{7.37}$$

We consider the *length of the edges* from the founding fathers to the first split point, say $T_1(F_1), \cdots, T_K(F_K)$. Then we take the random walks

$$\left(Y_t^{e_1,1} \right)_{t \in [0,T_1(F_1)]}, \left(Y_t^{e_2,2} \right)_{t \in [0,T_2(F_2)]}, \cdots, \tag{7.38}$$

where e_1, e_2 are the positions of the founding fathers in E. How to choose these points. They are the value of the mark attained for $t \leq 0$.

Then we continue with the split points replacing the founding fathers, where their position is given as the end-point of the corresponding random walks above. We continue until we reach the level h. Then we can associate with each path in the $\mathbb{R}$-tree from a founding father to a leave, parametrized via the length, a *path* with values in E generated by the random walk pieces. Continuing beyond h and before level 0 as constant path, we have marked $\lceil \widetilde{\mathfrak{U}} \rceil_t(h)$ with a path in $D(\mathbb{R}, E)$ and we can assign the weight which we get from the weight of the leaf. This way we have constructed for a fixed trunk of a given $\mathfrak{u} \in \mathbb{U}$ a random marked trunk, i.e. a random element in $\mathbb{U}(t-h)^{\sqcup}$. This way we have constructed $m_h(\mathfrak{u}, \cdot)$ a kernel from $\mathbb{U} \to \mathcal{M}_{\text{fin}}(D(\mathbb{R}, E))$ based on the $(t-h)$-trunk of $\mathfrak{u}$, $\mathfrak{u}$ an element of $\mathbb{U}$. We have now constructed a realization of the marked $(t-h)$-trunk of the element of $\mathbb{U}^D$ we look for.

From the $(t-h)$-trunks we obtain in the limit $h \uparrow t$, the $\mathbb{U}$-valued state at time t and we want to define the marks, i.e. we need $m(\mathfrak{u}, \cdot)$ as the limit $h \uparrow t$ of $m_h(\mathfrak{u}; \cdot)$, which exists because of the concrete construction.

Suppose next we have a *random variable* with values in $\mathbb{U}^D$. Observe that above construction defines a random measure on path and its *law* defines a transition kernel

$$M_h(\mathfrak{u}, dv) \text{ from } \mathbb{U}(h)^{\sqcup} \text{ into } \mathcal{P}\left(\mathcal{M}_{\text{fin}}(D) \right). \tag{7.39}$$

Hence we specified the law of the marked $(t-h)$-trunk of the stochastic process we look for and this works for every $h < t$ and the arising laws are by construction consistent, i.e. for $h' < h$ we have the law of the h'-truncation of the law for h. As the limit $h \uparrow t$ is taken we obtain $M(\mathfrak{u}, dv)$ from $\mathbb{U}$ into $\mathcal{P}(\mathcal{M}_{\text{fin}}(D))$ the searched for conditional law of the t-marginal.

Precisely we have now the information on the marks which we have to insert in our formalism, namely suppose that we have an $\mathbb{U}$-valued process $(\mathfrak{U}_t)_{t \geq 0}$, then we focus on the measure ν in $\mathfrak{U}_t = [U_t, r_t \otimes r_V, \nu]$, we write as $\mu \otimes m$, where m is a kernel from $\mathbb{U}$ to $(V, \mathcal{B}(V))$ which is read of from the state (U, r, μ) and our construction precisely as follows.

We need for given projection on the genealogy for the measure ν the conditional law given the first component, more precisely a regular version of the conditional distribution of ν^*, which we called above $m(\mathfrak{u}, \cdot)$, which is a finite measure on $(V, \mathcal{B}(V))$. This measure is for us only relevant projected on the trunks that is we keep the path constant beyond $t-h$, equal to its value at this time. This projection we read of from our construction by taking the realization of our random walks, leading from the founders to a leaf in the trunk and the weight prescribed there from the condition is their weight and gives $m_h(\mathfrak{u}, \cdot)$, as an atomic measure on

$D(\mathbb{R}, E)$ and as $h \uparrow t$ we obtain $m(\mathfrak{u}, \cdot)$ as we saw above and applying this to $\mathfrak{U}_t$ we get m_t, M_t. Finally we set therefore for our derived object: $M_t(\mathfrak{u}, \cdot) = \mathcal{L}[m_t(\mathfrak{u}, \cdot)]$ which gives us averaging over $\mathfrak{u}$ with the law of the $\mathbb{U}$-valued process now the $\mathbb{U}^D$-valued one namely $\mathcal{L}[\mathfrak{U}_t]$.

We claim and prove further below that this way we have constructed a realization of the conditioned distribution of $\lceil \widetilde{\mathfrak{U}}_t \rceil(h)$ given $\widetilde{\mathfrak{U}}_t$, with $(\mathfrak{U}_t)_{t \geq 0}$ being the limit process arising from the individual based models.

We need next the conditional law given the $\mathbb{U}$-projection of the *finite dimensional distributions*. We have for that purpose to carry out the construction for the states at times $0 \leq t_1 < t_2 < \cdots < t_n \leq t$ jointly. This means we consider the n-states and the trunks formed at heights $h_1, h_2, \cdots, h_n$ of these states. Here we have $h_k < t_k$ and we make the convention to consider the case where $h_k > t_{k-1}$, for all k.

We have now to construct the conditional distribution of the marks given the states in $\mathbb{U}$ in the n-time points. We consider first the n trunks specified by the $(h_i)_{i=1,\cdots,n}$ i.e. $(t - h_1)$-trunks. We begin the construction as before for each time $t_1, \cdots, t_n$ by using random walk increments for the time intervals $[0, t_1], [t_1, t_2]$, etc by taking n *independent copies* of our random walk collection, the point then is to match the pieces of the path by choosing the starting points of the random walks consistently. However we observe that the $(t - h_1)$-trunk of the state at time t_1 is as ultrametric space embedded in the $(t - h_2)$-trunk of the state at time t_2, there is only pruning since some of the leaves of the first will have no descendants at time t_2. But therefore there is no consistency problem, constructing the random walk path.

Similarly we proceed with the rescaled approximating branching random walk population and construct a representation as the one we used to define the *infinite* population per site process.

We are now in the situation that we have constructed the finite dimensional distributions of an $\mathbb{U}^{D^+}$-*valued Markov process*, the Markov property being a consequence of the Markov property of the $\mathbb{U}$-valued process we use and the independent increments property of the random walks used in the construction of the paths which are independent of the chosen genealogy. This *process is the candidate for the limit process* arising from the *sequence of individual based approximations*.

A consequence of the construction above is that in order to establish the convergence of the individual based approximations it will suffice to show convergence of the marginal distributions for fixed time t, if we can establish the *Feller property* for the involved processes.

Now we turn to the issue of the *convergence* of the approximations with individual based processes. Now we claim that our conditional distributions converge weakly, i.e. $\mathcal{L}[m_n(\mathfrak{u}, \cdot)]$ converge to $\mathcal{L}[m(\mathfrak{u}, \cdot)]$. For the approximating individual based models we have however the same $m(\mathfrak{u}, \cdot)$ and $M(\mathfrak{u}, \cdot)$ by construction. The map $\mathfrak{u} \to M(\mathfrak{u}, \cdot)$ is continuous.

We can use for the approximating system now the same random walk system and we pick the same increments and obtain this way a coupling by matching the starting positions cleverly. We also note that we can choose the genealogies all on one probability space such that we have convergence in the sense of ultrametric measure spaces for the trunks on which we condition (by conditioning on the state at time of the genealogy process). Therefore the approximation by individual based models converges to a limit law in the sense of convergence of the finite-dimensional

marginals. This limit is given by our *construction* we have given based on the realization of the $\mathbb{U}$-valued process and the collection of random walks.

We have to show that this process we have constructed solves the martingale problem *based on the construction* not its property as limit. The latter we would know so far only if in our polynomial either ϕ or χ is constant (recall the results we quoted in the beginning).

We have to show that the branching operator $(\Omega^{\mathrm{grow}} + \Omega^{\uparrow,\mathrm{bra}})$ and the mark evolution operator $\Omega^{\uparrow,\mathrm{mig}}$ act on a polynomial $\Phi^{n,\phi,\widehat{\chi}}$ by changing ϕ resp. $\widehat{\chi}$ *separately* by replacing ϕ by a suitable ϕ' and $\widehat{\chi}$ by a suitable $\widehat{\chi}'$ as given in (3.34), (3.36). Since the construction uses the $\mathbb{U}$-valued process as given element the action of the branching operator which is only lifted to $U \times V$ from U and the mark evolution is constructed based on the collection of independent random walks which evolve according to the path process dynamic replacing χ by χ^* as calculated earlier and which merge based on the underlying genealogy giving then the lifted, from V to $U \times V$, term.

Precisely we proceed as follows. We define again for $\mathfrak{w} \in \mathbb{U}^V, \mathfrak{w} \sim (\mathfrak{u}, \mathfrak{v})$ with $\mathfrak{u} \in \mathbb{U}$ and $\mathfrak{v} \in \mathcal{M}(V)$ by $\mathfrak{w} = [U \times V, r \otimes r_V, \nu], \mathfrak{u} = [U, r, \mu]$ and $\mathfrak{v} = \nu(U \times \cdot)$. We define for $\Omega^{\uparrow,\mathrm{bran}}$ the lifted version $(\widetilde{\Omega}^{\uparrow,\mathrm{bran}} \Phi^{n,\phi,\chi})(\mathfrak{u}, \mathfrak{v}) = (\Omega^{\uparrow,\mathrm{bran}} \Phi^{n,\phi,\chi}(\cdot, \mathfrak{v}))(\mathfrak{u}, \mathfrak{v})$ and similarly for $\Omega^{\uparrow,\mathrm{grow}}$. On the other hand we have an operator describing the evolution of the measures on path driven by the path process which is defined on $\Phi^{n,\widehat{\chi}}$ to describe the evolution of the historical process and which we define on $\Phi^{n,\phi,\widehat{\chi}}$ now, to describe the mark evolution. To derive the expression we have to use the construction we gave using the collection of random walks, which shows that $\Omega^{\uparrow,\mathrm{anc}}(\Phi^{n,\phi,\widehat{\chi}}(\mathfrak{u},\cdot))(\mathfrak{u}, \mathfrak{v}) = \Omega^{\mathrm{hist}}(\Phi^{n,\phi,\widehat{\chi}}(\mathfrak{u},\cdot))(\mathfrak{u}, \mathfrak{v}) = \sum\limits_{k=1}^{u} \Phi^{n,\phi,\widehat{\chi}^{*,k}}(\mathfrak{u}, \mathfrak{v})$, where $\widehat{\chi}'$ is defined in (7.28).

Altogether we have now constructed a solution to the martingale problem which concludes the Step 2 on the *existence* of the solution.

Both Step 1 and Step 2 together prove the *wellposedness* of the martingale process as claimed in Proposition 3.9 q.e.d.

References

Aldous, D. J. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.*, **3** (4), 450–465 (1990). MR1069105.

Aldous, D. J. The Continuum Random Tree. I. *Ann. Probab.*, **19** (1), 1–28 (1991a). MR1085326.

Aldous, D. J. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pp. 23–70. Cambridge Univ. Press, Cambridge (1991b). MR1166406.

Aldous, D. J. The continuum random tree. III. *Ann. Probab.*, **21** (1), 248–289 (1993). MR1207226.

Athreya, K. B. and Ney, P. E. *Branching processes.* Dover Publications, Inc., Mineola, NY (2004). ISBN 0-486-43474-5. MR2047480.

Bertoin, J. and Le Gall, J.-F. The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Related Fields*, **117** (2), 249–266 (2000). MR1771663.

Bertoin, J. and Le Gall, J.-F. Stochastic flows associated to coalescent processes. *Probab. Theory Related Fields*, **126** (2), 261–288 (2003). MR1990057.

Bertoin, J. and Le Gall, J.-F. Stochastic flows associated to coalescent processes. II. Stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.*, **41** (3), 307–333 (2005). MR2139022.

Caballero, M. E., Lambert, A., and Uribe Bravo, G. Proof(s) of the Lamperti representation of continuous-state branching processes. *Probab. Surv.*, **6**, 62–89 (2009). MR2592395.

Chauvin, B., Rouault, A., and Wakolbinger, A. Growing conditioned trees. *Stochastic Process. Appl.*, **39** (1), 117–130 (1991). MR1135089.

Dawson, D. A. The critical measure diffusion process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **40** (2), 125–145 (1977). MR478374.

Dawson, D. A. Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI—1991*, volume 1541 of *Lecture Notes in Math.*, pp. 1–260. Springer, Berlin (1993). MR1242575.

Dawson, D. A. and Greven, A. Multiple space-time scale analysis for interacting branching models. *Electron. J. Probab.*, **1**, no. 14, approx. 84 pp. (1996). MR1423467.

Dawson, D. A. and Greven, A. State dependent multitype spatial branching processes and their longtime behavior. *Electron. J. Probab.*, **8**, no. 4, 93 (2003). MR1961166.

Dawson, D. A. and Greven, A. *Spatial Fleming-Viot models with selection and mutation*, volume 2092 of *Lecture Notes in Mathematics*. Springer, Cham (2014). ISBN 978-3-319-02152-2; 978-3-319-02153-9. MR3155790.

Dawson, D. A. and Perkins, E. A. Historical processes. *Mem. Amer. Math. Soc.*, **93** (454), iv+179 (1991). MR1079034.

Depperschmidt, A. and Greven, A. Tree-valued Feller diffusion. *ArXiv Mathematics e-prints* (2019). arXiv: 1904.02044.

Depperschmidt, A. and Greven, A. *Stochastic Evolution of Genealogies of Spatial Populations: State Description, Characterization of Dynamics and Properties*, pp. 39–79 (2020). ISBN 978-981-12-0608-5. DOI: 10.1142/9789811206092_0002.

Depperschmidt, A., Greven, A., and Pfaffelhuber, P. Marked metric measure spaces. *Electron. Commun. Probab.*, **16**, 174–188 (2011). MR2783338.

Depperschmidt, A., Greven, A., and Pfaffelhuber, P. Tree-valued Fleming-Viot dynamics with mutation and selection. *Ann. Appl. Probab.*, **22** (6), 2560–2615 (2012). MR3024977.

Depperschmidt, A., Greven, A., and Pfaffelhuber, P. Path-properties of the tree-valued Fleming-Viot process. *Electron. J. Probab.*, **18**, no. 84, 47 (2013). MR3109623.

Duquesne, T. and Le Gall, J.-F. Random trees, Lévy processes and spatial branching processes. *Astérisque*, (281), vi+147 (2002). MR1954248.

Etheridge, A. M. *An introduction to superprocesses*, volume 20 of *University Lecture Series*. American Mathematical Society, Providence, RI (2000). ISBN 0-8218-2706-5. MR1779100.

Ethier, S. N. and Kurtz, T. G. *Markov processes. Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York (1986). ISBN 0-471-08186-8. MR838085.

Evans, S. N. Kingman's coalescent as a random metric space. In *Stochastic models (Ottawa, ON, 1998)*, volume 26 of *CMS Conf. Proc.*, pp. 105–114. Amer. Math.

Soc., Providence, RI (2000). MR1765005.

Evans, S. N. and Molchanov, I. The semigroup of metric measure spaces and its infinitely divisible probability measures. *Trans. Amer. Math. Soc.*, **369** (3), 1797–1834 (2017). MR3581220.

Evans, S. N., Pitman, J., and Winter, A. Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields*, **134** (1), 81–126 (2006). MR2221786.

Fleischmann, K. and Greven, A. Diffusive clustering in an infinite system of hierarchically interacting diffusions. *Probab. Theory Related Fields*, **98** (4), 517–566 (1994). MR1271108.

Glöde, P. *Dynamics of genealogical trees for autocatalytic branching processes.* Ph.D. thesis, Department Mathematik, Erlangen, Germany (2012). URL http://www.opus.ub.uni-erlangen.de/opus/volltexte/2013/4545/.

Glöde, P., Greven, A., and Rippl, T. Branching trees I: Concatenation and infinite divisibility. *Electron. J. Probab.*, **24**, Paper No. 52, 55 (2019). MR3968714.

Greven, A., Klenke, A., and Wakolbinger, A. Interacting diffusions in a random medium: comparison and longtime behavior. *Stochastic Process. Appl.*, **98** (1), 23–41 (2002). MR1884922.

Greven, A., Limic, V., and Winter, A. Representation theorems for interacting Moran models, interacting Fisher-Wright diffusions and applications. *Electron. J. Probab.*, **10**, no. 39, 1286–1356 (2005). MR2176385.

Greven, A., Pfaffelhuber, P., and Winter, A. Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Related Fields*, **145** (1-2), 285–322 (2009). MR2520129.

Greven, A., Pfaffelhuber, P., and Winter, A. Tree-valued resampling dynamics martingale problems and applications. *Probab. Theory Related Fields*, **155** (3-4), 789–838 (2013). MR3034793.

Greven, A., Sun, R., and Winter, A. Continuum space limit of the genealogies of interacting Fleming-Viot processes on $\mathbb{Z}$. *Electron. J. Probab.*, **21**, Paper No. 58, 64 (2016). MR3563886.

Greven, A., Sun, R., and Winter, A. The evolving genealogy of fossils: Unique characterization by martingale problems and applications (2020). In preparation 2020.

Harris, T. E. *The theory of branching processes.* Dover Phoenix Editions. Dover Publications, Inc., Mineola, NY (2002). ISBN 0-486-49508-6. MR1991122.

Lambert, A. Quasi-stationary distributions and the continuous-state branching process conditioned to be never extinct. *Electron. J. Probab.*, **12**, no. 14, 420–446 (2007). MR2299923.

Le Gall, J.-F. Marches aléatoires, mouvement brownien et processus de branchement. In *Séminaire de Probabilités, XXIII*, volume 1372 of *Lecture Notes in Math.*, pp. 258–274. Springer, Berlin (1989). MR1022916.

Le Gall, J.-F. The uniform random tree in a Brownian excursion. *Probab. Theory Related Fields*, **96** (3), 369–383 (1993). MR1231930.

Le Gall, J.-F. *Spatial branching processes, random snakes and partial differential equations.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (1999). ISBN 3-7643-6126-3. MR1714707.

Le Jan, Y. Superprocesses and projective limits of branching Markov process. *Ann. Inst. H. Poincaré Probab. Statist.*, **27** (1), 91–106 (1991). MR1098565.

Liggett, T. M. and Spitzer, F. Ergodic theorems for coupled random walks and other systems with locally interacting components. *Z. Wahrsch. Verw. Gebiete*, **56** (4), 443–468 (1981). MR621659.

Neveu, J. Erasing a branching tree. *Adv. in Appl. Probab.*, (suppl.), 101–108 (1986). MR868511.

Neveu, J. and Pitman, J. W. The branching process in a Brownian excursion. In *Séminaire de Probabilités, XXIII*, volume 1372 of *Lecture Notes in Math.*, pp. 248–257. Springer, Berlin (1989). MR1022915.