Electron. J. Probab. **21** (2016), no. 58, 1–64.
ISSN: 1083-6489 DOI: 10.1214/16-EJP4514

Continuum space limit of the genealogies of interacting Fleming-Viot processes on $\mathbb{Z}$

Andreas Greven[*] Rongfeng Sun[†] Anita Winter[‡]

Abstract

We study the evolution of genealogies of a population of individuals, whose type frequencies result in an interacting Fleming-Viot process on $\mathbb{Z}$. We construct and analyze the genealogical structure of the population in this genealogy-valued Fleming-Viot process as a marked metric measure space, with each individual carrying its spatial location as a mark. We then show that its time evolution converges to that of the genealogy of a continuum-sites stepping stone model on $\mathbb{R}$, if space and time are scaled diffusively. We construct the genealogies of the continuum-sites stepping stone model as functionals of the Brownian web, and furthermore, we show that its evolution solves a martingale problem. The generator for the continuum-sites stepping stone model has a singular feature: at each time, the resampling of genealogies only affects a set of individuals of measure 0. Along the way, we prove some negative correlation inequalities for coalescing Brownian motions, as well as extend the theory of marked metric measure spaces (developed recently by Depperschmidt, Greven and Pfaffelhuber [DGP11]) from the case of probability measures to measures that are finite on bounded sets.

Keywords: Brownian web; continuum-sites stepping stone model; evolving genealogies; interacting Fleming-Viot process; marked metric measure space; martingale problems; negative correlation inequalities; spatial continuum limit.
AMS MSC 2010: 60K35; 60J65; 60J70; 92D25.
Submitted to EJP on August 28, 2015, final version accepted on September 12, 2016.
Supersedes arXiv:1508.07169.

Contents

[*]Universität Erlangen-Nürnberg, Germany. E-mail: greven@mi.uni-erlangen.de
[†]National University of Singapore, Singapore. E-mail: matsr@nus.edu.sg
[‡]Universität Duisburg-Essen, Germany. E-mail: anita.winter@uni-due.de

1 Introduction and main results

In the study of spatial population models on discrete geographic spaces (for example $\mathbb{Z}^d$), such as branching processes, voter models, or interacting Fisher-Wright diffusions (Fleming-Viot models), the technique of passing to the *spatial continuum limit* has proven to be useful in gaining insight into the qualitative behaviour of these processes. A key example is branching random walks on $\mathbb{Z}^d$, leading to the Dawson-Watanabe process [Daw77] on $\mathbb{R}^d$ and Fisher-Wright diffusions; catalytic branching and mutually catalytic branching on $\mathbb{Z}$, leading to SPDE on $\mathbb{R}$ [KS88, EF96, DP98, DEF$^+$02b, DEF$^+$02a]. The goal of this paper is to carry out this program at the level of *genealogies*, rather than just type or mass configurations. We focus here on *interacting Fleming-Viot models* on $\mathbb{Z}$.

1.1 Background and overview

We summarize below the main results of this paper, recall some historical background, as well as state some open problems.

Summary of results. The purpose of this paper is twofold. On the one hand, we want to understand the formation of large local one-family clusters in Fleming-Viot populations on the geographic space $\mathbb{Z}^1$, by taking a space-time continuum limit of the genealogical configurations equipped with types. On the other hand we use this example to develop the theory of *tree-valued dynamics* via *martingale problems* in some new directions. In particular, this is the first study of a *tree-valued dynamics* on an unbounded geographical space with infinite sampling measure, which requires us to extend both the notion of *marked metric measure spaces* in [GPW09, DGP11] and the martingale

problem formulations in [GPW13, DGP12] to *marked* metric measure spaces with infinite sampling measures that are boundedly finite (i.e., finite on bounded sets).

Here is a summary of our main results:

(1) We extend the theory of *marked metric measure spaces* [GPW09, DGP12] from probability sampling measures to infinite sampling measures that are *boundedly finite*, which serve as the state space of marked genealogies of spatial population models. See Section 1.2.

(2) We characterize the evolution of the genealogies of interacting Fleming-Viot (IFV) models by well-posed martingale problems on spaces of *marked ultrametric measure spaces*. See Section 1.3.

(3) We give a *graphical construction* of the *spatial continuum limit* of the IFV genealogy process, which is the genealogy process of the so-called *Continuum Sites Stepping-stone Model* (CSSM), taking values in the space of ultrametric measure spaces with *spatial marks* and an *infinite total population*. The graphical construction is based on the (dual) Brownian web [FINR04]. The CSSM genealogy process has the feature that, as soon as $t > 0$, the process enters a regular subset of the state space that is not closed under the topology. Only on this subset we can evaluate the action of the operator of the martingale problem in its action on test functions. The nice aspect is that the set of these states are preserved under the dynamic. This leads to a singular structure with complications for the associated martingale problem and for the study of continuity of the process at time 0. See Section 1.4.

(4) We prove that under suitable scaling, the IFV genealogy processes converge to the CSSM genealogy process. The proof is based on duality with spatial coalescents, together with a novel approach of controlling the genealogy structure using a weaker convergence result on the corresponding measure-valued processes, with measures on the geographic and type space (with no genealogies). See Section 1.5.

(5) We show that the CSSM genealogy process solves a martingale problem with a singular generator. More precisely, the generator action involves individuals, which are *not typical* under the sampling measure, so that the dynamic is driven by atypical individuals at atypical locations. In particular, the generator is only defined on a regular subset of the state space. See Section 1.6.

(6) We prove some negative correlation inequalities for coalescing Brownian motions, which are of independent interest. See Appendix C.

Besides the description of the genealogies of the current population, we also prepare the ground for the treatment of all individuals ever alive, i.e. *fossils*, moving from the state space of marked ultrametric measure spaces to the state space of marked measure $\mathbb{R}$-trees, which will be carried out elsewhere.

History of the problem: Why are we particularly interested in *one*-dimensional geographic spaces for our scaling results? Many interacting spatial systems that model evolving populations, i.e., Markov processes with state spaces $I^G (I = \mathbb{R}, \mathbb{N}, [0,1]$, etc., and $G = \mathbb{Z}^d$ or the hierarchical group Ω_N) that evolve by a migration mechanism between sites and a stochastic mechanism acting locally at each site, exhibit a *dichotomy* in their *longtime behavior*. For example, when $G = \mathbb{Z}^d$ and the migration is induced by the simple symmetric random walk: in dimension $d \leq 2$, one observes convergence to laws concentrated on the traps of the dynamic; while in $d \geq 3$, nontrivial equilibrium states are approached and the extremal invariant measures are spatially homogeneous ergodic measures characterized by the intensity of the configuration. Typical examples include the voter model, branching random walks, spatial Moran models, or systems of

interacting diffusions (e.g., Feller, Fisher-Wright or Anderson diffusions). One obtains universal dimension-dependent scaling limits for these models if an additional continuum spatial limit is taken, resulting in, for example, super Brownian motion (see Liggett [Lig85] or Dawson [Daw93]).

In the low-dimension regime, the cases $d = 1$ and $d = 2$ are very different. In $d = 2$, one observes for example in the voter model the formation of mono-type clusters on spatial scales $t^{\alpha/2}$ with a random $\alpha \in [0, 1]$, a phenomenon called diffusive clustering (see Cox and Griffeath [CG86]). In the one-dimensional voter model, the clusters have an extension of a fixed order of magnitude but exhibit random factors in that scale. More precisely, in space-time scales $(\sqrt{t}, t)$ for $t \to \infty$, we get annihilating Brownian motions. Similar results have been obtained for low-dimensional branching systems (Klenke [Kle00], Winter [Win02]), systems of interacting Fisher-Wright diffusions (Fleischmann and Greven [FG94], [FG96] and subsequently [Zho03], [DEF$^+$00]) and for the Moran model in $d = 2$ (Greven, Limic, Winter [GLW05]).

In all these models, one can go further and study the complete space-time genealogy structure of the cluster formation and describe this phenomenon asymptotically by the *spatial continuum limit*. In particular, the description for the one-dimensional voter model can be extended to the complete space-time genealogy structure, obtaining as scaling limit the *Brownian web* [Arr79, Arr81, TW98, FINR04] (see Appendix B, and the recent survey [SSS15]). More precisely, the Brownian web is defined by considering instantaneously coalescing one-dimensional Brownian motions starting from every space-time point in $\mathbb{R} \times \mathbb{R}$. It arises as the diffusive scaling limit of continuous time coalescing simple symmetric random walks starting from every space-time point in $\mathbb{Z} \times \mathbb{R}$, which represent the space-time genealogies of the voter model. This is analogous to the study of historical process for branching processes, which approximates the ancestral paths of branching random walks by that of super Brownian motion (see e.g. [DP91, FG96, GLW05]).

The basic idea behind all this is that, we can identify the *genealogical relationship* between the individuals of the population living at different times and different locations. This raises the question of whether one can obtain a description of the asymptotic behavior of the complete genealogical structure of the process on large space-time scales, which will in turn allow for asymptotic descriptions of interesting genealogical statistics that are not expressible in a natural way in terms of the configuration process.

These observations on the genealogical structure goes back to the graphical construction of the voter model due to Harris, and continues up to the historical process of Dawson and Perkins for branching models [DP91], or representation by contour processes [GJ98, Ald93]. To better describe genealogies, the notion of $\mathbb{R}$-*trees, marked* $\mathbb{R}$-*trees or marked measure* $\mathbb{R}$-*trees* were developed as a framework [Eva97, EPW06]. These objects contain the relevant information *abstracted* from the *labeled* genealogy tree, where every individual is coded with its lifespan and its locations at each time. Such a coding means in particular that all members of the population are distinguished, which information is mostly not needed. In the large population limit, it suffices to consider the statistics of the population via sampling.

For this purpose, one equips the population with a *metric* (genealogical distance), a probability measure (the so-called *sampling measure*, which allows to draw typical finite samples from the population) together with a *mark* (specifying types and locations). This description in terms of random marked metric measure spaces (in fact, the metric defines a tree) is the most natural framework to discuss the asymptotic analysis of population models, since it comprises exactly the information one wants to keep for the analysis in the limits of populations with even locally infinitely many individuals. The evolution of a process with such a state space is described by *martingale problems* [GPW13, DGP12].

Open problems and conjectures. We show in this paper that the spatial continuum limit of the one-dimensional interacting Fleming-Viot genealogy process solves a martingale problem. However, due to the singular nature of the generator, the uniqueness of this martingale problem is non-standard, and we leave it for a future paper. In particular, it is difficult to establish the duality relation at the level of generators, because the generators for the process and its dual are only defined for special test functions *at special points*.

Instead of $\mathbb{Z}$, resp. $\mathbb{R}$, as geographic spaces, one could consider the hierarchical group $\Omega_N = \oplus_{\mathbb{N}} Z_N$, with Z_N being the cyclical group of order N, respectively the continuum hierarchical group $\Omega_N^\infty = \bigoplus_{\mathbb{Z}} Z_N$. Brownian motion on $\mathbb{R}$ can be replaced by suitable Lévy processes on Ω_N^∞ and the program of this paper can then be carried out. The Brownian web would have to be replaced by an object based on Lévy processes as studied in [EF96], [DEF$^+$00]. We conjecture that the analogues of our theorems would hold in this context. A further challenge would be to give a unified treatment of these results on $\mathbb{R}, \Omega_N^\infty$.

Another direction is to consider the genealogy processes of interacting Feller diffusions, catalytic or mutually catalytic diffusions, interacting logistic Feller diffusions, and derive their genealogical continuum limits. A more difficult extension would be to include *ancestral paths* as marks, which raises new challenges related to topological properties of the state space. Namely, as the space of paths is a Polish space, it is not a Heine-Borel space as closed balls around a path are not compact.

Outline of Section 1. The remainder of the introduction is organized as follows. In Subsection 1.2, we recall and extend the notion of marked metric measure spaces needed to describe the genealogies. In Subsection 1.3, we define the interacting Fleming-Viot (IFV) genealogy process via a martingale problem and give a dual representation in terms of a spatial coalescent. In Subsection 1.4, we give, based on the Brownian web, a graphical construction for the continuum-sites stepping stone model (CSSM) on $\mathbb{R}$ and its marked genealogy process, which is the continuum analogue and scaling limit of the interacting Fleming-Viot genealogy process on $\mathbb{Z}$, under diffusive scaling of space and time and normalizing of measure, a fact which we state in Subsection 1.5. In Subsection 1.6, we formulate a martingale problem for the CSSM genealogy process. In Subsection 1.7 we outline the rest of the paper.

1.2 Marked metric measure spaces

In this subsection we introduce the state space of the genealogies of interacting Fleming-Viot processes. We want to describe evolving genealogies of the whole population of all individuals currently alive allowing to sample from this population. We also want to include the individuals' positions in geographic space and possible genetic types. We therefore regard genealogies as (equivalence classes) of marked metric measure spaces. As our geographic space is infinite (and not compact), it won't be possible to sample individuals by means of a finite (or after renormalizing of a probability) measure. We rather require the sampling measure to be finite on all subpopulations which can be obtained by restricting to finite geographic subspaces.

The following definition of *marked metric measure spaces* extends the one introduced in [DGP11] which considered probability measures only.

Definition 1.1 (V-mmm-spaces)**.** *Let (V, r_V) be a complete separable metric space with metric r_V, and let o be a distinguished point in V.*

1. *We call (X, r, μ) a V-marked metric measure space (V-mmm space for short) if:*

(i) (X, r) *is a complete separable metric space,*
(ii) μ *is a measure on the Borel* σ*-algebra of* $X \times V$*, with* $\mu(X \times B_o(R)) < \infty$ *for each ball* $B_o(R) \subset V$ *of finite radius* R *centered at* o*.*

2. *We say two* V*-mmm spaces* (X, r_X, μ_X) *and* (Y, r_Y, μ_Y) *are equivalent if there exists a measureable map* $\varphi : X \to Y$*, such that*

$$r_X(x_1, x_2) = r_Y(\varphi(x_1), \varphi(x_2)) \qquad \text{for all } \; x_1, x_2 \in \operatorname{supp}(\mu_X(\cdot \times V)), \tag{1.1}$$

and if $\widetilde{\varphi} : X \times V \to Y \times V$ *is defined by* $\widetilde{\varphi}(x, v) := (\varphi(x), v)$*, then*

$$\mu_Y = \mu_X \circ \widetilde{\varphi}^{-1}. \tag{1.2}$$

In other words, φ *is an isometry between* $supp(\mu_X(\cdot \times V))$ *and* $supp(\mu_Y(\cdot \times V))$*, and the induced map* $\widetilde{\varphi}$ *is mark and measure preserving. We denote the equivalence class of* (X, r, μ) *by*

$$\overline{(X, r, \mu)}. \tag{1.3}$$

3. *The space of (equivalence classes of)* V*-mmm spaces is denoted by*

$$\mathbb{M}^V := \big\{\overline{(X, r, \mu)} : (X, r, \mu) \text{ is a } V\text{-mmm space}\big\}. \tag{1.4}$$

4. *The subspace of (equivalence classes of)* V*-mmm spaces which admit a mark function is denoted by*

$$\begin{aligned}&\mathbb{M}^V_{\mathrm{fct}}\\ &:= \{\overline{(X, r, \mu)} \in \mathbb{M}^V : \exists \text{ measurable } \kappa : X \to V \text{ s.t. } \mu(\mathrm{d}(x, \cdot)) = \mu(\mathrm{d}x \times V) \otimes \delta_{\kappa(x)}\}.\end{aligned} \tag{1.5}$$

Note that $\mathbb{M}^V$ depends both on the set V and the metric r_V since the latter defines the sets on which the measure must be finite.

Marked metric measure spaces were introduced in [DGP11], which extends the notion of metric measure spaces studied earlier in [GPW09]. Definition 1.1 is exactly the analogue of [DGP11, Def. 2.1], where μ is a probability measure. The basic interpretation in our context is that: X is the space of individuals; $r(x_1, x_2)$ measures the genealogical distance between two individuals x_1 and x_2 in X; the measure μ is a measure on the individuals and the marks they carry (which can be spatial location as well as type, or even ancestral paths up to now), allowing us to draw samples from individuals with marks in a bounded set.

To define a topology on $\mathbb{M}^V$ that makes it a Polish space, we will make use of the *marked Gromov-weak topology* introduced in [DGP11, Def. 2.4] for V-mmm spaces equipped with probability measures. In this topology a sequence converges iff we can embedd the involved metric spaces isometrically into one metric space such that the images of the "sampling" measures converges weakly. The basic idea to extend this to $\mathbb{M}^V$ is that, given our assumption on μ in Definition 1.1.1.(ii), we can localize μ to finite balls in V to reduce μ to a finite measure. We can then apply the *marked Gromov-weak topology* (which also applies to finite measures instead of probability measures) to require convergence for each such localized version of the V-mmm spaces. We will call such a topology V*-marked Gromov-weak*$^\#$ *topology*, replacing *weak* by *weak*$^\#$, following the terminology in [DVJ03, Section A2.6] for the convergence of measures that are bounded on bounded subsets of a complete separable metric space. Note that vague convergence is for measures that are finite on compact rather than bounded subsets. Both notions agree on Heine-Borel spaces (compare, [ALW16]).

Definition 1.2 (V-marked Gromov-weak$^\#$ Topology)**.** *Fix a sequence of continuous functions $\psi_k : V \to [0,1]$, $k \in \mathbb{N}$, such that $\psi_k = 1$ on the ball $B_o(k)$ centered at $o \in V$ of radius k, and $\psi_k = 0$ on $B_o^c(k+1)$. Let $\chi := \overline{(X, r, \mu)}$ and $\chi_n := \overline{(X_n, r_n, \mu_n)}$, $n \in \mathbb{N}$, be elements of $\mathbb{M}^V$. Let $\psi_k \cdot \mu$ be the measure on $X \times V$ defined by $(\psi_k \cdot \mu)(\mathrm{d}(x,v)) := \psi_k(v)\mu(\mathrm{d}(x,v))$, and let $\psi_k \cdot \mu_n$ be defined similarly. We say that $\chi_n \underset{n\to\infty}{\longrightarrow} \chi$ in the V-marked Gromov-weak$^\#$ topology if and only if:*

$$\overline{(X_n, r_n, \psi_k \cdot \mu_n)} \underset{n\to\infty}{\Longrightarrow} \overline{(X, r, \psi_k \cdot \mu)} \quad \text{in the Gromov-weak topology for each } k \in \mathbb{N}. \tag{1.6}$$

When $V = \mathbb{R}^d$, we may choose ψ_k to be infinitely differentiable.

Remark 1.3 (Dependence on o and $(\psi_k)_{k\in\mathbb{N}}$)**.** Note that the V-marked Gromov-weak$^\#$ topology does not depend on the choice $o \in V$ and the sequence $(\psi_k)_{k\in\mathbb{N}}$, as long as ψ_k has bounded support and $A_k := \{v : \psi_k(v) = 1\}$ increases to V as $k \to \infty$.

Remark 1.4 ($\mathbb{M}^V$ as a subspace of $(\mathbb{M}_f^V)^{\mathbb{N}}$)**.** Let $\mathbb{M}_f^V$ denote the space of (equivalent classes of) V-mmm spaces with finite measures, equipped with the V-marked Gromov-weak topology as introduced in [DGP11, Def. 2.4]. Then it is a well-known fact that each element $\overline{(X, r, \mu)} \in \mathbb{M}^V$ can be identified with a sequence $(\overline{(X, r, \psi_1 \cdot \mu)}, \overline{(X, r, \psi_2 \cdot \mu)}, \ldots)$ in the product space $(\mathbb{M}_f^V)^{\mathbb{N}}$, equipped with the product topology. This identification allows us to easily deduce many properties of $\mathbb{M}^V$ from properties of $\mathbb{M}_f^V$ that were established in [DGP11]. In particular, we can metrize the V-marked Gromov-weak$^\#$ topology on $\mathbb{M}^V$ by introducing a metric (which can be called *V-marked Gromov-Prohorov$^\#$ metric*)

$$d_{MGP^\#}(\overline{(X_1, r_1, \mu_1)}, \overline{(X_2, r_2, \mu_2)}) := \sum_{k=1}^{\infty} \frac{d_{MGP}(\overline{(X_1, r_1, \psi_k \cdot \mu_1)}, \overline{(X_2, r_2, \psi_k \cdot \mu_2)}) \wedge 1}{2^k}, \tag{1.7}$$

where d_{MGP} is the marked Gromov-Prohorov metric on $\mathbb{M}_f^V$, which was introduced in [DGP11, Def. 3.1] and metrizes the marked Gromov-weak topology.

The proof of the following result is in Appendix A.

Theorem 1.5 (Polish space)**.** *The space $\mathbb{M}^V$, equipped with the V-marked Gromov-weak$^\#$ topology, is a Polish space.*

Points in $\mathbb{M}^V$, as well as weak convergence of $\mathbb{M}^V$-valued random variables, can be determined by the so-called *polynomials* on $\mathbb{M}^V$, which are defined via sampling a finite subset on the V-mmm space.

Definition 1.6 (Polynomials)**.** *Let $n \in \mathbb{N}$. Let $g \in C_{bb}(V^n, \mathbb{R})$, the space of real-valued bounded continuous function on V^n with bounded support. For $k \in \mathbb{N} \cup \{0, \infty\}$, let $\phi \in C_b^k(\mathbb{R}_+^{\binom{n}{2}}, \mathbb{R})$, the space of k-times continuously differentiable real-valued functions on $\mathbb{R}_+^{\binom{n}{2}}$ with bounded derivatives up to order k. We call the function $\Phi^{n,\phi,g} : \mathbb{M}^V \to \mathbb{R}$ defined by*

$$\Phi^{n,\phi,g}(\overline{(X, r, \mu)}) := \int \cdots \int \phi(\underline{r}) g(\underline{v}) \mu^{\otimes n}(\mathrm{d}(\underline{x}, \underline{v})), \tag{1.8}$$

a monomial of order n, where $\underline{x} := (x_1, \ldots, x_n)$, $\underline{r} = \underline{r}(\underline{x}) := (r(x_i, x_j))_{1\le i<j\le n}$, $\underline{v} := (v_1, \ldots, v_n)$, and $\mu^{\otimes n}(\mathrm{d}(\underline{x}, \underline{v}))$ denotes the n-fold product measure of μ on $(X \times V)^n$.

(a) *Let $\Pi_n^k := \{\Phi^{n,\phi,g} : \phi \in C_b^k(\mathbb{R}_+^{\binom{n}{2}}, \mathbb{R}), g \in C_{bb}(V^n, \mathbb{R})\}$, which we call the space of monomials of order n (with differentiability of order k). Let Π_0^k be the set of constant functions. We then denote by*

$$\Pi^k := \cup_{n\in\mathbb{N}_0} \Pi_n^k$$

the set of all monomials (with differentiability of order k).

(b) *For $V = \mathbb{R}^d$ and $k, l \in \mathbb{N} \cup \{0, \infty\}$, we define*

$$\Pi_n^{k,l} := \{\Phi^{n,\phi,g} : \phi \in C_b^k(\mathbb{R}_+^{\binom{n}{2}}, \mathbb{R}), g \in C_{bb}^l(V^n, \mathbb{R})\},$$

and $\Pi^{k,l} := \cup_{n\in\mathbb{N}_0} \Pi_n^{k,l}$.

(c) *We call the linear spaces*

$$\widetilde{\Pi}^{k,\ell} \text{ generated by } \Pi^{k,\ell}, \tag{1.9}$$

the polynomials (with differentiability of ϕ, resp. g, of order k, resp. ℓ).

Remark 1.7. Note that the polynomials form an algebra of bounded continuous functions since the product of two monomials can be seen as a new monomial as defined in (1.8). However, the sum of two monomials in general is not a monomial.

Throughout the paper we are often interested in the following sub-space of $\mathbb{M}^V$ (compare, e.g., Definition 1.34 of so-called regular space of states). Let b be a measurable function on $\mathbb{R}_+$, and write

$$\mathbb{M}^{(V,\le b)} := \left\{\overline{(X, r, \mu)} \in \mathbb{M}^V : \mu(X \times B_r(o)) \le b(r)\right\}. \tag{1.10}$$

If β is a measure on $\mathcal{B}(V)$ which is finite on bounded subsets, and

$$\mathbb{M}^{(V,\beta)} := \left\{\overline{(X, r, \mu)} \in \mathbb{M}^V : \mu(X \times \cdot) = \beta\right\}, \tag{1.11}$$

then obviously $\mathbb{M}^{(V,\beta)}$ is a closed subspace of $\mathbb{M}^{(V,\le b)}$ with $b(r) := \beta(B_r(o))$.

Theorem 1.8 (Convergence determining class)**.** *Fix a measurable function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$. We have the following properties for Π^k, for each $k \in \mathbb{N} \cup \{0, \infty\}$:*

(i) *Π^k is convergence determining in $\mathbb{M}^{(V,\le b)}$. Namely, $\overline{(X_n, r_n, \mu_n)} \to \overline{(X, r, \mu)}$ in $\mathbb{M}^{(V,\le b)}$ if and only if $\Phi(\overline{(X_n, r_n, \mu_n)}) \to \Phi(\overline{(X, r, \mu)})$ as $n \to \infty$ for all $\Phi \in \Pi^k$.*

(ii) *Π^k is also convergence determining in $\mathcal{M}_1(\mathbb{M}^{(V,\le b)})$, the space of probability measures on $\mathbb{M}^{(V,\le b)}$. Namely, a sequence of $\mathbb{M}^{(V,\le b)}$-valued random variables $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges weakly to a $\mathbb{M}^{(V,\le b)}$-valued random variable $\mathcal{X}$ if and only if $\mathbb{E}[\Phi(\mathcal{X}_n)] \to \mathbb{E}[\Phi(\mathcal{X})]$ as $n \to \infty$ for all $\Phi \in \Pi^k$.*

(iii) *For $V = \mathbb{R}^d$ and each $k, l \in \mathbb{N} \cup \{0, \infty\}$, $\Pi^{k,l}$ is also convergence determining in $\mathbb{M}^{(V,\le b)}$ and $\mathcal{M}_1(\mathbb{M}^{(V,\le b)})$.*

We defer the proof of Theorems 1.5 and 1.8, as well as some additional properties of V-mmm spaces, to Appendix A.

Remark 1.9. For the models we consider, the genealogies lie in certain particular Polish spaces which arise as *closed* subspaces of $\mathbb{M}^V$. Note that the current population alive corresponds to the *leaves* of a genealogical tree, and the associated V-mmm space is *ultrametric*. We will denote the space of V-marked ultrametric measure spaces by $\mathbb{U}^V$. They form a closed subspace of $\mathbb{M}^V$ and hence $\mathbb{U}^V$ is Polish. The same holds for $\mathbb{M}^{(V,\beta)}$, for some Borel measure β on V which is finite on bounded subsets.

Remark 1.10. Recall $\mathbb{M}_{\text{fct}}^V$ from Definition 1.1, and notice that $\mathbb{M}_{\text{fct}}^V$ is not closed, and that we therefore choose the bigger space $\mathbb{M}^V$ as the state space. The space $\mathbb{M}^V$ allows an individual $x \in X$ to carry a set of marks, equipped with the conditional measure of μ on V given $x \in X$. If each individual carries only a single mark which we can identify via a *mark function* $\kappa : X \to V$, the corresponding marked metric measure space is an element of $\mathbb{M}_{\text{fct}}^V$. This will be the case for the interacting Fleming-Viot process that we will study.

It can be shown that every element in $\mathbb{M}_{\mathrm{fct}}^V$ is an element of the closed space

$$\begin{aligned}\mathbb{M}_h^V := \bigcup_{A \text{ bdd}} \bigcap_{\delta>0} \{(X,r,\mu) \in \mathbb{M}^V : \\ \exists \mu'_A \in \mathcal{M}_f(X\times A) : \mu' \le \mu(\cdot \times A), \|\mu' - \mu(\cdot\times A)\|_{\mathrm{TV}} \le h_A(\delta), \\ (\mu'_A)^{\otimes 2}\{r(x_1,x_2) < \delta, d(v_1,v_2) > h_A(\delta)\} = 0\}, \end{aligned} \tag{1.12}$$

for some $h_A \in \mathcal{H}$, where

$$\mathcal{H} := \{h : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\} : \ h \text{ is continuous and increasing, and } h(0) = 0\} \tag{1.13}$$

([KL15, Lemma 2.8]).

1.3 Interacting Fleming-Viot (IFV) genealogy processes

We now study the genealogies of the measure-valued interacting Fleming-Viot (IFV) processes on a countable geographic space and with allelic types, typically taken from the type space $[0,1]$ (see [DGV95] for details on IFV), which is motivated by the following individual-based model, the so-called Moran model.

Consider a population of individuals, X, with locations indexed by a *countable additive group* V (for us this later will be $\mathbb{Z}$). The individuals *migrate* independently according to rate one continuous time random walks with transition probability kernel $a(\cdot,\cdot)$,

$$a(v_1,v_2) = a(0,v_2-v_1) \qquad \text{for all } \ v_1,v_2 \in V. \tag{1.14}$$

We denote the transition kernel of the time reversed random walks by

$$\bar{a}(v_1,v_2) := a(v_2,v_1) = a(0,v_1-v_2). \tag{1.15}$$

Individuals furthermore reproduce by *resampling*, where every pair of individuals at the same site dies together at exponential rate $\gamma > 0$, and with equal probability, one of the two individuals is chosen to give birth to two new individuals at the same site with the same type as the parent. This naturally induces a genealogical structure. The genealogical distance, r, of two individuals at time t is $2\min\{t,T\}$ plus the distance of the ancestors at time 0, where T is the time it takes to go back to the most recent common ancestor. Imposing the Haar measure (here the counting measure), μ, on the population with each individual carrying its location as a *mark*, we obtain a V-mmm space and its equivalence class is the state of the genealogy process.

Letting now the number of individuals per site tend to infinity and normalizing the measure such that each site carries population mass of order one, we obtain a diffusion model, the *interacting Fleming-Viot (IFV) genealogy process* with state space

$$\mathbb{U}_1^V := \mathbb{U}^{(V,\mathfrak{n})}, \tag{1.16}$$

where $\mathfrak{n}$ denotes the *Haar measure* on the countable geographic space V, the 1 indicating that the measure restricted to each colony is a probability measure. This (see Remark 1.9) is again a Polish space. (For the diffusion limit in the case of a finite geographic space V, see [GPW13], [DGP12]).

Remark 1.11. If we introduce as marks (besides *locations* from a countable geographic space $\mathbb{G}$) also allelic *types* from some set $\mathbb{K}$ (typically taken as a closed subset of $[0,1]$), then the type is inherited at reproduction and $V = \mathbb{K}\times\mathbb{G}$ is the product of type space and geographic space. In this case the localization procedure in Definition 1.2 applies to the geographic space, since $\mathbb{K}$ is compact.

1.3.1 The genealogical IFV martingale problem

We now define the interacting Fleming-Viot (IFV) genealogy process via a martingale problem for a linear operator L^{FV} on $C_b(\mathbb{U}_1^V, \mathbb{R})$, acting on polynomials. For simplicity we first leave out allelic types, which we introduce later in Remark 1.13.

The linear operator L^{FV} on $C_b(\mathbb{U}_1^V, \mathbb{R})$, with domain $\Pi^{1,0}$ as introduced in Definition 1.6 (b) with $d = 1$, consists of three terms, corresponding respectively to aging, migration, and reproduction by resampling. With $\mathcal{X} = \overline{(X, r, \mu)}$ and $\Phi = \Phi^{n,\phi,g} \in \Pi^{1,0}$,

$$\begin{aligned} L^{\mathrm{FV}}\Phi(\mathcal{X}) \;=\; & 2 \int_{(X\times V)^n} \mu^{\otimes n}(\mathrm{d}(\underline{x},\underline{v}))\, g(\underline{v}) \sum_{1\le k<\ell\le n} \frac{\partial}{\partial r_{k,\ell}}\, \phi(\underline{r}) \\ & + \int_{(X\times V)^n} \mu^{\otimes n}(\mathrm{d}(\underline{x},\underline{v}))\, \phi(\underline{r}) \sum_{j=1}^{n} \sum_{v'\in V} \bar{a}(v_j, v')(M_{v_j,v'}g - g)(\underline{v}) \qquad (1.17) \\ & + 2\,\gamma \int_{(X\times V)^n} \mu^{\otimes n}(\mathrm{d}(\underline{x},\underline{v}))\; g(\underline{v}) \sum_{1\le k<\ell\le n} 1_{\{v_k=v_\ell\}}(\theta_{k,\ell}\phi - \phi)(\underline{r}), \end{aligned}$$

where $\underline{x} = (x_1, \cdots, x_n)$, $\underline{v} = (v_1, \cdots, v_n)$, $\underline{r} := (r_{k,\ell})_{1\le k<\ell\le n} := (r(x_k, x_\ell)_{1\le k<\ell\le n})$, and

$$(M_{v_j,v'}g)(v_1, \cdots, v_n) := g(v_1, \cdots, v_{j-1}, v', v_{j+1}, \cdots, v_n) \qquad (1.18)$$

encodes the replacement by migration of the j-th sampled individual corresponding to a jump from location v_j to v', while $\theta_{k,\ell}$ encodes the replacement of the ℓ-th individual by the k-th individual (both at the same site). More precisely,

$$(\theta_{k,\ell}\phi)(\underline{r}) := \phi(\theta_{k,\ell}\underline{r}), \quad \text{with} \quad (\theta_{k,\ell}\underline{r})_{i,j} := \begin{cases} r(x_i, x_j) & \text{if } i, j \neq \ell, \\ r(x_i, x_k) & \text{if } j = \ell, \\ r(x_k, x_j) & \text{if } i = \ell. \end{cases} \qquad (1.19)$$

The first result states that there is a unique $\mathbb{U}_1^V$-valued diffusion process associated with this operator.

Theorem 1.12 (Martingale problem characterization of IVF Genealogy processes)**.** *For any $\mathcal{X}_0 = \overline{(X_0, r_0, \mu_0)} \in \mathbb{U}_1^V$, we have:*

(i) *The $(L^{\mathrm{FV}}, \Pi^{1,0}, \delta_{\mathcal{X}_0})$-martingale problem is well-posed, i.e. there exists a $\mathbb{U}_1^V$-valued process $\mathcal{X}^{\mathrm{FV}} := (\mathcal{X}_t^{\mathrm{FV}})_{t\ge 0}$, unique in its distribution, which has initial condition $\mathcal{X}_0$ and càdlàg path, such that for all $\Phi \in \Pi^{1,0}$ and w.r.t. the natural filtration generated by $(\mathcal{X}_t^{\mathrm{FV}})_{t\ge 0}$,*

$$\Big(\Phi(\mathcal{X}_t^{\mathrm{FV}}) - \Phi(\mathcal{X}_0^{\mathrm{FV}}) - \int_0^t (L^{\mathrm{FV}}\Phi)(\mathcal{X}_s^{\mathrm{FV}})\mathrm{d}s\Big)_{t\ge 0} \text{ is a martingale.} \qquad (1.20)$$

(ii) *The solutions (for varying initial conditions) define a strong Markov with continuous path. This Markov process has the Feller property, i.e., the one-dimensional distributions depend continuously on the initial distribution.*

(iii) *If the initial state admits a mark function, then so does the path almost surely.*

Remark 1.13. If we add the type of an individual as an additional mark, i.e., $V := \mathbb{G}\times\mathbb{K}$ with geographic and type space respectively, then the same result holds if we modify as follows. We require the states to satisfy the constraint that the projection of μ on

the geographic space V is still the counting measure. The test functions Φ should be modified so that we multiply $g : \mathbb{G}^n \to \mathbb{R}$, acting on the locations of the n sampled individuals, by another bounded, continuous factor $g_{\text{typ}} : \mathbb{K}^n \to \mathbb{R}$, acting on the types of the individuals. The generator L^{FV} should be modified accordingly, so that g_{typ} changes at resampling from g_{typ} to $g_{\text{typ}} \circ \widetilde{\theta}_{k,\ell}$, with $\widetilde{\theta}_{k,\ell}$ replacing the ℓ-th sampled individual by the k-th one, see [DGP12]).

Remark 1.14. The process $\mathcal{X}_t^{\text{FV}} = \overline{(X_t, r_t, \mu_t)}$ has the property that the measure-valued process $\widehat{\mathcal{X}}_t$ given by the collection $\{\mu_t(X_t \times \{i\} \times \cdot), i \in \mathbb{G}\}$, is the IFV process on $(\mathcal{M}_1(\mathbb{K}))^{\mathbb{G}}$.

Remark 1.15. From Section 1.4 onward, we will choose $V = \mathbb{Z}$. However in the subsequent analysis, it is important to observe that we can embed $\mathbb{Z}$ into $\mathbb{R}$ and view $\mathbb{U}^{\mathbb{Z}}$ as a closed subspace of $\mathbb{U}^{\mathbb{R}}$, and view the IFV genealogy process as $\mathbb{U}^{\mathbb{R}}$-valued process.

1.3.2 Duality

The IFV genealogy process $\mathcal{X}^{\text{FV}}$ can be characterized by a dual process, the *spatial coalescent*. The formulation given here can also incorporate mutation and selection (see [DGP12]).

For each $n \in \mathbb{N}$, let $\tilde{S}_n$ denote the space of *partitions* of the set $\{1, ..., n\}$, i.e., $\pi = \{\mathfrak{p}; \mathfrak{p} \in \pi\} \in \tilde{\mathbb{S}}_n$ is a collection of disjoint subsets $\mathfrak{p} \subseteq \{1, ..., n\}$, referred to as partition elements (or blocks), such that $\uplus_{\mathfrak{p}\in\pi}\mathfrak{p} = \{1, ..., n\}$. Moreover, let $\tilde{\mathbb{S}}_n^V$ denote the space of *marked partitions* of the set $\{1, ..., n\}$ with mark space V, i.e., we regard $\{(\mathfrak{p}, v_{\mathfrak{p}}); \pi = \{\mathfrak{p}; \mathfrak{p} \in \pi\} \in \tilde{\mathbb{S}}_n, v_{\mathfrak{p}} \in V\} \in \tilde{\mathbb{S}}_n^V$ as a partition of $\{1, ..., n\}$ where each partition element is assigned a mark in V. Finally, let $\mathbb{S}_n := \tilde{\mathbb{S}}_n^V \times \mathbb{R}_+^{\binom{n}{2}}$ denote the space of *historical marked partitions* of the set $\{1, ..., n\}$ with mark space V. That is, we regard $(\{(\mathfrak{p}, v_{\mathfrak{p}}); \mathfrak{p} \in \pi\}, (r_{ij})_{1\le i<j\le n}) \in \mathbb{S}_n$ as consisting of a marked partition and a matrix of mutual distances.

The dual process $(\mathcal{K}_t)_{t\ge 0}$ will take values in $\mathbb{S} := \bigcup_{n\in\mathbb{N}} \mathbb{S}_n$ with the following dynamics: given a finite historical marked partition, independently every pair of partition elements with the same mark in V merges at rate γ. Until a pair of partition elements merges, the marks migrate independently of each other on V according to a continuous time random walk with transition kernel $\bar{a}$. After merging the marks of the two involved partition elements will move together forever. At time t, we define the genealogical distance $r_t(i, j)$ of i and j in $\{1, 2, ..., n\}$ as $2\min\{t, T_{i,j}\}$, where $T_{i,j}$ is the first time that i and j belong to the same partition element.

For each n, $\phi \in C_b^\infty(\mathbb{R}_+^{\binom{n}{2}}, \mathbb{R})$, and $g \in C_{bb}(V^n, \mathbb{R})$, we define a *duality function* $H : \mathbb{U}_1^V \times \mathbb{S}_n \to \mathbb{R}$ with

$$H\big(\mathcal{X}, \mathcal{K}\big) := \int_{(X\times V)^{|\pi|}} \mu^{\otimes|\pi|}(\mathrm{d}(\underline{x}, \underline{v}))\, \mathbf{1}_{\{v_{\mathfrak{p}}=\xi'_{\mathfrak{p}};\, \forall \mathfrak{p}\in\pi\}} \cdot g(\underline{v}^\pi)\phi\big(\underline{\underline{r}}^\pi(\underline{x}) + \underline{\underline{r}}'\big), \tag{1.21}$$

where $\mathcal{X} = \overline{(X, r, \mu)} \in \mathbb{U}_1^V$, and $\mathcal{K} = (\{(\mathfrak{p}, \xi'_{\mathfrak{p}}); \mathfrak{p} \in \pi\}, \underline{\underline{r}}') \in \mathbb{S}_n$, and $\underline{v}^\pi = (v_{\mathfrak{p}(i)})_{i=1,...,n}$, and $\underline{\underline{r}}^\pi(\underline{x}) := (r(x_{\mathfrak{p}(i)}, x_{\mathfrak{p}(j)}))_{1\le i<j\le n}$, with $\mathfrak{p}(i)$ being the partition element of π containing $i \in \{1, ..., n\}$.

Remark 1.16. Note that $\{H(\cdot, \mathcal{K}) : \mathcal{K} \in \mathbb{S}_n, \phi \in C_b^\infty(\mathbb{R}_+^{\binom{n}{2}}, \mathbb{R}), g \in C_{bb}(V^n, \mathbb{R}), n \in \mathbb{N}\}$ is law-determining and convergence-determining on $\mathbb{U}_1^V$.

The IFV genealogy process $(\mathcal{X}_t^{\text{FV}})_{t\ge 0}$ is dual to the coalescent $(\mathcal{K}_t)_{t\ge 0}$, and its law and behavior as $t \to \infty$ can be determined as follows.

Theorem 1.17 (Duality and longtime behaviour)**.** *The following properties hold for the IFV genealogy process* $(\mathcal{X}_t^{\text{FV}})_{t\ge 0}$*:*

(a) *For every* $\mathcal{X}_0^{\mathrm{FV}} \in \mathbb{U}_1^V$ *and* $\mathcal{K}_0 \in \mathbb{S}$*, we have*

$$\mathbb{E}[H(\mathcal{X}_t^{\mathrm{FV}}, \mathcal{K}_0)] = \mathbb{E}[H(\mathcal{X}_0^{\mathrm{FV}}, \mathcal{K}_t)], \quad t \geq 0. \tag{1.22}$$

(b) *If* $\widehat{a}(\cdot,\cdot) = \frac{1}{2}(a(\cdot,\cdot) + \bar{a}(\cdot,\cdot))$ *is recurrent, then*

$$\mathcal{L}[\mathcal{X}_t^{\mathrm{FV}}] \underset{t\to\infty}{\Longrightarrow} \Gamma \in \mathcal{M}_1(\mathbb{U}_1^V), \tag{1.23}$$

where Γ *is the unique invariant measure of the process* $\mathcal{X}^{\mathrm{FV}}$ *on* $\mathbb{U}_1^V$*.*

Remark 1.18. If $\widehat{a}$ is transient, then we can decompose $X_t^{\mathrm{FV}} = \overline{(X_t, r_t, \mu_t)}$ in such a way that X_t is the countable union of disjoint sets X_t^i, μ_t is the sum of measures μ_t^i, and $\{(X_t^i, r_t|_{X_t^i \times X_t^i}, \mu_t^i)\}_{i\in\mathbb{N}}$ are V-mmm spaces such that for $x \in X_t^i, x' \in X_t^j$ with $i \neq j$, $r_t(x,x')$ diverges in probability as $t \to \infty$, and each $(X_t^i, r_t|_{X_t^i \times X_t^i}, \mu_t^i)$ converges in law to a limiting V-mmm space. Alternatively we can transform distances $r : r \to 1 - e^{-r}$ and obtain a unique equilibrium in that case. See also [GM].

Remark 1.19 (Strong duality)**.** As the interacting Fleming-Viot process is a population model, its evolving genealogy $(U_t, r_t, \mu_t)_{t\geq 0}$ can be represented by a V-marked $\mathbb{R}$-tree (X, r). That is, we can find a 0-hyperbolic metric space (with distances possibly equal to ∞) (X, r) such that any two $x, y \in X$ of finite distance are connected by a path, and isometries $(\varphi_t)_{t\geq 0}$ with $\varphi_t : U_t \to X$ such that for all $x \in X$ there is a $t \geq 0$ with $\varphi_t^{-1}(x) \in \mathrm{supp}(\mu_t(\cdot \times V))$. We won't write down this $\mathbb{R}$-tree representation explicitly here but it can be derived from the look-down construction presented in [GLW05] in a straightforward way.

We would like to point out that this representation by an $\mathbb{R}$-tree allows to define the ancestor A_s^T of x at time T back at time $T - s$, and the above duality relation can actually be stated in a strong sense. That is, we can construct for each $T > 0$ our model together with a dual process $\tilde{K}^T$ on the same probability space such that for all $X' \subseteq U_T$ with $\#X' < \infty$, $\phi \in C_b^\infty(\mathbb{R}_+^{\binom{\#X'}{2}}, \mathbb{R})$, and $g \in C_{bb}(V^{\#X'}, \mathbb{R})$, and $H = H^{\#X',\varphi,g}$ from (1.21), for all $s \in [0,T]$,

$$H\big(\mathcal{U}_{T-s}, \tilde{\mathcal{K}}_s^T\big) \equiv H\big(\mathcal{U}_T, \tilde{\mathcal{K}}_0^T\big), \text{ almost surely.} \tag{1.24}$$

Notice that whenever a duality relation holds almost surely rather than in expectation, one refers to it as a *strong form of duality*.

Indeed, define for fixed $T \geq 0$ and any finite subset $X' := \{x_1, ..., x_{\#X'}\} \subset U_T$ the map $\tilde{K}^T := (\tilde{K}_s^T := ((\{(\mathfrak{p}, \xi'_{\mathfrak{p}}); \mathfrak{p} \in \pi_s^T\}, r_s^{'T}))_{s\in[0,T]}$ which sends $\mathcal{U} := (\mathcal{U}_t := (U_t, r_t, \mu_t))_{t\geq 0}$ to a path with values in the space of historical marked partitions $\mathbb{S}_{\#X'}$ defined by

$$r_s^{'T}(i,j) = r\big(\varphi_T(x_i), \varphi_T(x_j)\big) \wedge (T - s), \tag{1.25}$$

for all $1 \leq i < j \leq \#X'$, a marked partition of $\{1, 2, ..., n\}$ defined through the equivalence relation

$$i \equiv_s^T j \quad \text{iff} \quad r_s^{'T}(i,j) < (T - s), \tag{1.26}$$

and a family of position functions on $\varphi_T(\mathrm{supp}(\mu_T(\cdot \times V)))$ such that

$$\xi_s^{'T}(\mathfrak{p}) := \kappa_{T-s}\big(\varphi_{T-s}^{-1}(\mathfrak{p})\big). \tag{1.27}$$

By construction, $\tilde{K}^T$ is the dual spatial coalescent. Moreover, for all $s \in [0,T]$, and $(\underline{x}, \underline{v}) \in (X_T \times V)^{|\pi_0|}$,

$$\underline{r}_{\underline{=}T-s}^{\pi_s}(A_s^T(\underline{x})) + \underline{r}'^T_{\underline{=}\,s} = \underline{r}_{\underline{=}T}^{\pi_0}(\underline{x}) + \underline{r}'^T_{\underline{=}\,0}, \tag{1.28}$$

and $\mathrm{s}\kappa_{T-s}(A_s^T(x))$ has under μ_T the same distribution as types under μ_{T-s}. This together implies the strong form of duality as stated in (1.24). ∎

As a further consequence of relation (1.22) we can specify the *finite dimensional distributions* of $(\mathcal{X}_t^{\mathrm{FV}})_{t\geq 0}$ completely in *terms of the spatial coalescent* as follows. Fix a time horizon $T > 0$. The finite dimensional distributions are determined by the expectation of all test functions $\widetilde{\Phi} : C([0,\infty), \mathbb{M}^V)$ of the following form:

$$\widetilde{\Phi}\big((\mathcal{X}_t^{\mathrm{FV}})_{t\geq 0}\big) := \Pi_{k=1}^{\ell} \Phi_k\big(\mathcal{X}_{t_k}^{\mathrm{FV}}\big), \tag{1.29}$$

for some $\ell \in \mathbb{N}$, $0 \leq t_1 < t_2 < ... < t_\ell = t$ and $\Phi_k = \Phi^{n_k, g_k, \phi_k} \in \Pi^{1,0}$ of order $n_k \in \mathbb{N}$, $\phi_k \in C_b^\infty(\mathbb{R}_+^{\binom{n_k}{2}}, \mathbb{R})$, and $g_k \in C_{bb}(V^{n_k}, \mathbb{R})$, for each $k = 1, ..., \ell$.

The dual is the *spatial coalescent with frozen particles* $(\widetilde{\mathcal{K}}_s)_{s\in[0,T]}$, for which the time index $s \in [0,T]$ runs in the opposite direction from the time index of $\mathcal{X}^{\mathrm{FV}}$. Namely, looking backward from time T, for each $1 \leq k \leq \ell$, we start n_k particles at time $T - t_k$ at locations $\xi_{t_k}^1 := (\xi_{t_k,1}^1, ..., \xi_{t_k,n_k}^1) \in V^{n_k}$, each forming its own partition element in the partition $\pi \in \mathbb{S}_n$, $n := n_1 + ... + n_\ell$. The particles perform the usual dynamics of the spatial coalescent with the restriction that all particles starting at time $T - t_k$ were kept frozen before time $T - t_k$. At time s, the genealogical distance $r_s(i,j)$ between two individuals i and j, started respectively at times $T - t_i$ and $T - t_j$, is defined to be

$$2\min\{s, T_{i,j}\} - \min\{s, (T - t_i)\} - \min\{s, (T - t_j)\}, \tag{1.30}$$

where $T_{i,j}$ is the first time the two individuals coalesce.

Denote this new spatial coalescent process with frozen particles by $(\widetilde{\mathcal{K}}_t)_{t\geq 0}$. The state space $\mathbb{S}$ is once more the space of historical marked partitions. We then define the duality function $H : \mathbb{U}_1^V \times \mathbb{S} \to \mathbb{R}$ which determines the finite dimensional distributions of $\mathcal{X}$ for varying $\widetilde{\mathcal{K}}$ as in (1.21) but now with

$$\phi(\underline{r}) := \prod_{k=1}^{\ell} \phi_k\big((r_{i,j})_{n_1+\cdots+n_{k-1}<i<j\leq n_1+\cdots+n_k}\big) \text{ and } g\big(\underline{v}^\pi\big) := \prod_{k=1}^{\ell} g_k\big(t_k, \underline{v}^{\pi_k}\big), \tag{1.31}$$

where π_k are the partition elements started at time t_k.

The following space-time duality is an immediate consequence of the Markov property and the duality applied successively to the time intervals $[t_{\ell-1}, t_\ell]$, $[t_{\ell-2}, t_{\ell-1}]$, ..., $[0, t_1]$.

Corollary 1.20 (Space-time duality)**.** *Let $\mathcal{X}_0^{\mathrm{FV}} \in \mathbb{U}_1^V$, and $\widetilde{\mathcal{K}}_0 \in \mathbb{S}$ with $t_k \leq T$ for all $1 \leq k \leq \ell$. Let $\widetilde{\Phi}$ be defined as above (1.30). Then*

$$\mathbb{E}\big[\widetilde{\Phi}\big((\mathcal{X}_s^{\mathrm{FV}})_{s\in[0,T]}\big)\big] = \mathbb{E}\big[H\big(\mathcal{X}_0^{\mathrm{FV}}, \widetilde{\mathcal{K}}_T\big)\big]. \tag{1.32}$$

In words, the genealogical distances among the $n = n_1 + \cdots + n_\ell$ individuals sampled from $(\mathcal{X}_s^{\mathrm{FV}})_{s\in[0,T]}$, with n_k individuals sampled at time t_k at specified locations, can be recovered by letting these individuals evolve backward in time as a spatial coalescent until time 0, at which point we sample from $\mathcal{X}_0^{\mathrm{FV}}$ according to the location of each partition element in the spatial coalescent.

1.4 Genealogies of continuum-sites stepping stone model (CSSM) on $\mathbb{R}$

If we rescale space and time diffusively, the measure-valued interacting Fleming-Viot process on $\mathbb{Z}$ converges to a continuum space limit, the so-called *continuum-sites stepping stone model* (CSSM). Formally, CSSM is a measure-valued process $\nu := (\nu_t)_{t\geq 0}$ on $\mathbb{R} \times [0,1]$, where $\mathbb{R}$ is the geographical space and $[0,1]$ is the type space. We might think of individuals in the population which undergo independent Brownian motions, and whenever two individuals meet, one of the two individuals is chosen with equal probability and switches its type to that of the second individual. Provided that $\nu_0(\cdot \times [0,1])$ is the

Lebesgue measure on $\mathbb{R}$, CSSM was rigorously constructed in [EF96, Eva97, DEF$^+$00, Zho03] via a moment duality with coalescing Brownian motions. In particular, $\nu_t(\cdot \times [0,1])$ is the Lebesgue measure on $\mathbb{R}$, for all $t > 0$.

In this subsection we will construct the evolving genealogies of the CSSM based on duality to the *(dual) Brownian web*, and establish properties (Proposition 1.25, Theorem 1.27).

For the voter model on $\mathbb{Z}$, the joint genealogy lines of individuals at all space-time points in $\mathbb{Z} \times [0,\infty)$ are distributed as a collection of coalescing random walks evolving backward in time (see [Lig85]). Analogously, for the CSSM on $\mathbb{R}$, when n individuals are sampled from the population at possibly different times, their joint genealogy lines evolve backward in time as coalescing Brownian motions. Upon reaching time 0, each surviving genealogy line then independently selects an ancestral type by sampling according to the conditional distribution of ν_0 on the type space $[0,1]$, conditioned on the spatial location of the genealogy line. Furthermore the joint genealogy lines of individuals at all space-time points in $\mathbb{R} \times [0,\infty)$ are distributed as a collection of coalescing Brownian motions evolving backward in time. Therefore the CSSM on $\mathbb{R}$ is exactly the continuum analogue of the interacting Fleming-Viot process (as well as the voter model) on $\mathbb{Z}$.

Although having an uncountable number (starting from every space-time point in $\mathbb{R} \times [0,\infty)$) of coalescing Brownian motions seems problematic, this object has been constructed rigorously and is now known as the *(dual) Brownian web* $\widehat{\mathcal{W}}$ [FINR04, FINR06], dual to a *forward Brownian web* $\mathcal{W}$ constructed on the *same* probability space. The (dual) Brownian web is essentially a collection of coalescing Brownian motions on $\mathbb{R}$, starting from every point in the space-time plane $\mathbb{R} \times \mathbb{R}$.

In [FINR04], the Brownian web $\mathcal{W}$ is constructed as a random variable where each realization of

$$\mathcal{W} \text{ is a closed subset of } \Pi := \cup_{s\in\mathbb{R}} C([s,\infty),\mathbb{R}), \tag{1.33}$$

the space of continuous paths in $\mathbb{R}$ with any starting time $s \in \mathbb{R}$. In [FINR04], the topology is defined by first compactifying the space $\mathbb{R}^2$ suitably, and then by choosing for Π the *topology of local uniform convergence* and requiring the initial times to be close if paths are close. For each $z := (x,t) \in \mathbb{R}$, we will let $\mathcal{W}(z) := \mathcal{W}(x,t)$ denote the subset of paths in $\mathcal{W}$ with starting position x and starting time t. We can construct $\mathcal{W}$ by first fixing a countable dense subset $\mathcal{D} \subset \mathbb{R}^2$, and then construct a collection of coalescing Brownian motions $\{\mathcal{W}(z) : z \in \mathcal{D}\}$, with one Brownian motion starting from each $z \in \mathcal{D}$. The Brownian web $\mathcal{W}$ is then obtained by taking a suitable closure of $\{\mathcal{W}(z) : z \in \mathcal{D}\}$ in Π, which gives rise to a set of paths starting from every point in the space-time plane $\mathbb{R}^2$. It can be shown that the law of $\mathcal{W}$ does *not* depend on the choice of $\mathcal{D}$ (see Theorem B.1 in Appendix B).

The Brownian web $\mathcal{W}$ has a *graphical* dual called the *dual Brownian web*, which we denote by $\widehat{\mathcal{W}}$. Formally,

$$\widehat{\mathcal{W}} \text{ is a random closed subset of } \widehat{\Pi} := \cup_{s\in\mathbb{R}} C((-\infty,s],\mathbb{R}), \tag{1.34}$$

the space of continuous paths in $\mathbb{R}$ starting at any time $s \in \mathbb{R}$ and running backward in time as coalescing Brownian motions.

The *joint* distribution of $\widehat{\mathcal{W}}$ and $\mathcal{W}$ is uniquely determined by the requirement that the paths of $\widehat{\mathcal{W}}$ never cross paths of $\mathcal{W}$ (see, Theorem B.3 in Appendix B). Thus, jointly, the Brownian web and its dual is a random variable taking values in a Polish space, with

$$(\mathcal{W}, \widehat{\mathcal{W}}) \in \Pi \times \widehat{\Pi}. \tag{1.35}$$

Interpreting coalescing Brownian motions in the (dual) Brownian web as *ancestral lines* specifying the genealogies, we can then give an *almost sure graphical construction*

of the CSSM, instead of relying on moment duality relations as in [EF96, Eva97, DEF+00, Zho03], which nevertheless we get as corollary of the graphical construction. The classical measure-valued CSSM process can be recovered from $(\mathcal{X}_t^{\mathrm{CS}})_{t\geq 0}$ by projecting the sampling measure $(\mu_t^{\mathrm{CS}})_{t\geq 0}$ to the mark space V, if V is chosen to be the product of geographical space $\mathbb{R}$ and type space $[0,1]$. In what follows, we will take V to be only the geographical space $\mathbb{R}$, since types have no influence on the evolution of genealogies.

We next explicitly construct a version of the CSSM genealogy process

$$\mathcal{X}^{\mathrm{CS}} := (\mathcal{X}_t^{\mathrm{CS}})_{t\geq 0}, \quad \mathcal{X}_t^{\mathrm{CS}} = \overline{(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}}, \mu_t^{\mathrm{CS}})}, \qquad t \geq 0, \tag{1.36}$$

as a functional of the (dual) Brownian web $(\mathcal{W}, \widehat{\mathcal{W}})$.

To avoid a disruption of the flow of presentation, background details on the (dual) Brownian web we will need are collected in Appendix B.

We proceed in three steps:

Step 1 (Initial states). Assume that $\mathcal{X}_0^{\mathrm{CS}}$ belongs to the following closed subspace of $\mathbb{U}^{\mathbb{R}}$:

$$\mathbb{U}_1^{\mathbb{R}} := \{\overline{(X, r, \mu)} \in \mathbb{U}^{\mathbb{R}} : \mu(X \times \cdot) \text{ is the Lebesgue measure on } \mathbb{R}\}. \tag{1.37}$$

In other words, $\mathbb{U}_1^{\mathbb{R}}$ is the set of $\mathbb{R}$-marked ultrametric measure spaces where the projection of the measure on the mark space (geographic space) $\mathbb{R}$ is the Lebesgue measure. This is necessary for the duality between CSSM and coalescing Brownian motions. We will see that almost surely $\mathcal{X}_t^{\mathrm{CS}} \in \mathbb{U}_1^{\mathbb{R}}$ for all $t \geq 0$.

Step 2 (The time-t genealogy as a metric space). To define $(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}})$ for every $t > 0$, let us fix a realization of $(\mathcal{W}, \widehat{\mathcal{W}})$, (see the Appendix B for more details). For each $t > 0$, let

$$A_t := \big\{v \in \mathbb{R} : \widehat{\mathcal{W}}(v, t) \text{ contains a single path } \hat{f}_{(v,t)}\big\}, \quad E_t := \mathbb{R}\backslash A_t. \tag{1.38}$$

By Lemma B.4 on the classification of points in $\mathbb{R}^2$ w.r.t. $\mathcal{W}$ and $\widehat{\mathcal{W}}$, almost surely, E_t is a countable set for each $t > 0$. For each $v \in A_t$, we interpret $\hat{f}_{(v,t)}$ as the genealogy line of the individual at the space-time coordinate (v, t). Genealogy lines of individuals at different space-time coordinates evolve backward in time and coalesce with each other. At time 0, each genealogy line traces back to exactly one spatial location in the set

$$\widehat{E} := \{\hat{f}_{(v,s)}(0) : v \in \mathbb{R}, s > 0, \hat{f} \in \widehat{\mathcal{W}}\}, \tag{1.39}$$

where we note that $\widehat{E}$ is almost surely a countable set, because by Theorem B.1 and Lemma B.2, paths in $\widehat{\mathcal{W}}$ can be approximated in a strong sense by a countable subset of paths in $\widehat{\mathcal{W}}$. For each $v \in \widehat{E}$, we then identify a common ancestor $\xi(v)$ for all the individuals whose genealogy lines trace back to spatial location v at time 0, by sampling an individual $\xi(v) \in X_0^{\mathrm{CS}}$ according to the conditional distribution of μ_0^{CS} on X_0^{CS}, conditioned on the spatial mark in the product space $X_0^{\mathrm{CS}} \times \mathbb{R}$ being equal to v.

We next characterize individuals by points in space. Note that there is a natural genealogical distance between points in A_t. For individuals $x, y \in A_t$, if $\hat{f}_{(x,t)}$ and $\hat{f}_{(y,t)}$ coalesce at time $\hat{\tau} < t$, then denoting $u := \hat{f}_{(x,t)}(0)$ and $v := \hat{f}_{(y,t)}(0)$, we define the distance between x and y by

$$r_t(x, y) := \begin{cases} 2(t - \hat{\tau}) & \text{if } \hat{\tau} \geq 0, \\ 2t + r_0^{\mathrm{CS}}(\xi(u), \xi(v)) & \text{if } \hat{\tau} < 0. \end{cases} \tag{1.40}$$

First define $(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}})$ as the *closure* of A_t w.r.t. the metric r_t defined in (1.40). Note that $(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}})$ is ultrametric, and by construction Polish.

Remark 1.21. We may even extend the distance r_t to a distance r between (x,t) with $x \in A_t$ and $t > 0$, and (y,s) with $y \in A_s$ and $s > 0$. More precisely, let

$$r((x,t),(y,s)) := \begin{cases} s+t-2\hat{\tau} & \text{if } \hat{\tau} \geq 0, \\ s+t+r_0^{\mathrm{CS}}(\xi(u),\xi(v)) & \text{if } \hat{\tau} < 0. \end{cases} \tag{1.41}$$

Step 3 (Adding the sampling measure). We now define $\mathcal{X}_t^{\mathrm{CS}} := \overline{(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}}, \mu_t^{\mathrm{CS}})}$. For that purpose, we will represent next X_t^{CS} as an *enriched* copy of $\mathbb{R}$ (see (1.42) below).

By identifying each $x \in A_t$ with the path $\hat{f}_{(x,t)} \in \widehat{\mathcal{W}}$, we can also identify X_t^{CS} with the closure of $\{\hat{f}_{(x,t)} \in \widehat{\mathcal{W}} : x \in A_t\}$ in $\widehat{\Pi}$, because a sequence $x_n \in A_t$ is a Cauchy sequence w.r.t. the metric r_t if and only if the sequence of paths $\hat{f}_{(x_n,t)}$ is a Cauchy sequence when the distance between two paths is measured by the time to coalescence, which by Lemma B.2, is also equivalent to $(\hat{f}_{(x_n,t)})_{n\in\mathbb{N}}$ being a Cauchy sequence in $\widehat{\Pi}$.

When we take the closure of $\{\hat{f}_{(x,t)} \in \widehat{\mathcal{W}} : x \in A_t\}$ in $\widehat{\Pi}$, only a countable number of paths in $\widehat{\mathcal{W}}$ are added, which are precisely the leftmost and rightmost paths in $\widehat{\mathcal{W}}(x,t)$, when $\widehat{\mathcal{W}}(x,t)$ contains more than one path.

Namely for each $x \in E_t$ (recall from (1.38)), let x^+, x^- denote the two copies of x obtained by taking limits of $x_n \in A_t$ with either $x_n \downarrow x$ or $x_n \uparrow x$ in $\mathbb{R}$, and let $E_t^{\pm} := \{x^{\pm} : x \in E_t\}$. We can then take

$$X_t^{\mathrm{CS}} := A_t \cup E_t^+ \cup E_t^-, \tag{1.42}$$

equipped with a metric r_t^{CS}, which is the extension of r_t from A_t to its closure X_t^{CS}, giving a Polish space $(\mathcal{X}_t^{\mathrm{CS}}, r_t^{\mathrm{CS}})$.

Next to get a *sampling measure*, note that each finite ball in $(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}})$ with radius less than t can be identified with an interval in $\mathbb{R}$ (modulo a subset of $E_t \cup E_t^+ \cup E_t^-$), and hence can be assigned the Lebesgue measure of this interval. That is, we define the Borel measure $\widetilde{\mu}_t^{CS}$ on (X_t^{CS}, r_t^{CS}) by

$$\widetilde{\mu}_t^{CS}\big(\{x' \in X_t^{\mathrm{CS}} : r_t^{\mathrm{CS}}(x',x) < \delta\}\big) := R(x,\delta) - L(x,\delta), \quad \delta < t, x \in X_t^{\mathrm{CS}} \tag{1.43}$$

if $\{x' \in X_t^{\mathrm{CS}} : r_t^{\mathrm{CS}}(x',x) < \delta\} = (L(x,\delta), R(x,\delta)) \subseteq \mathbb{R}$ (Note that as we have identified balls with intervals on $\mathbb{R}$ modulo a subset of $E_t \cup E_t^+ \cup E_t^-$, we are free to think of these intervals as open). We then define the sampling measure μ_t^{CS} on $\mathcal{B}((X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}}) \times (\mathbb{R}, d_{\mathrm{eucl}}))$ by

$$\mu_t^{CS}(\mathrm{d}x\mathrm{d}v) := \widetilde{\mu}_t^{CS}(\mathrm{d}x)\delta_x(dv). \tag{1.44}$$

This completes our construction of the CSSM genealogy process $(\mathcal{X}_t^{\mathrm{CS}})_{t\geq 0}$.

Remark 1.22 (Notational simplification)**.** In the sequel we will apply the existence of a mark function and the embedding of the basic set in the enriched reals to simplify notation. Namely, we will write integrals w.r.t. the sampling measure over $X \times V$ as an integral w.r.t. the Lebesgue measure over $\mathbb{R}$. For example, we will write $\int_{\mathbb{R}^2} g(x_1,x_2)\phi(|x_1-x_2|)\mathrm{d}x_1x_2$ rather than $\int_{\mathbb{X}\times\mathbb{V}^2} g(v_1,v_2)\phi(r(x_1,x_2))\mu^{\otimes 2}(\mathrm{d}(\underline{(x,v)})$.

A key feature is again *duality*. We can replace in the Definition of the process $(\mathcal{K}_t)_{t\geq 0}$ from Subsubsection 1.3.2 the random walks by Brownian motions to obtain $(\mathcal{K}_t^{\mathrm{Br}})_{t\geq 0}$, which is dual to the process $(\mathcal{X}_t^{\mathrm{CS}})_{t\geq 0}$ by construction:

Corollary 1.23. *(H-duality) The tree-valued CS-process and marked tree-valued coalescing Brownian motions are in H-duality, i.e., for each H of the form (1.21),*

$$\mathbb{E}[H(\mathcal{X}_t^{\mathrm{CS}}, \mathcal{K}_0^{\mathrm{Br}})] = \mathbb{E}[H(\mathcal{X}_0^{\mathrm{CS}}, \mathcal{K}_t^{\mathrm{Br}})]. \tag{1.45}$$

Furthermore by the above construction strong duality holds. □

Remark 1.24. Note in the continuum case the function $\Phi^{n,g,\varphi}(\cdot, \kappa^{\mathrm{Br}})$ is not a polynomial since we fix the locations we consider. In order to get a polynomial we need to consider a function g on mark space with $g \in C^2_{bb}(\mathbb{R}, \mathbb{R})$ over which we integrate w.r.t. the sampling measure.

We collect below some basic properties for the CSSM genealogy processes that we just constructed.

Proposition 1.25 (Regularity of states)**.** *Let $\mathcal{X}^{\mathrm{CS}} := (\mathcal{X}^{\mathrm{CS}}_t)_{t\geq 0}$ be the CSSM genealogy process constructed from the dual Brownian web $\widehat{\mathcal{W}}$, with $\mathcal{X}^{\mathrm{CS}}_0 \in \mathbb{U}^{\mathbb{R}}_1$. Then almost surely, for every $t > 0$:*

(a) *There exists a continuous (mark) function $\kappa_t : X^{\mathrm{CS}}_t \to \mathbb{R}$, i.e., $\mu^{\mathrm{CS}}_t(\mathrm{d}x\mathrm{d}v) = \mu^{\mathrm{CS}}_t(\mathrm{d}x \times \mathbb{R})\delta_{\kappa_t(x)}(\mathrm{d}v)$;*

(b) *For each $\ell \in (0, t)$, X^{CS}_t is the disjoint union of balls $(B^\ell_i)_{i\in\mathbb{Z}}$ of radius ℓ. Furthermore, there exists a locally finite set $E^\ell_t := \{v_i\}_{i\in\mathbb{Z}} \subset \mathbb{R}$ with $v_{i-1} < v_i$ for all $i \in \mathbb{Z}$, such that $\{\kappa(x) : x \in B^\ell_i\} = [v_{i-1}, v_i]$ and $\mu^{\mathrm{CS}}_t(B^\ell_i) = v_i - v_{i-1}$. We can identify E^ℓ_t from $\widehat{\mathcal{W}}$ by*

$$E^\ell_t := \{x \in E_t : \hat{f}^+_{(x,t)} \text{ and } \hat{f}^-_{(x,t)} \text{ coalesce at some time } s \leq t - \ell\}, \tag{1.46}$$

where $f^+_{(x,t)}$ and $f^-_{(x,t)}$ are respectively the rightmost and leftmost path in $\widehat{\mathcal{W}}(x, t)$;

(c) *$\mathcal{X}^{\mathrm{CS}}_t = \overline{(X^{\mathrm{CS}}_t, r^{\mathrm{CS}}_t, \mu^{\mathrm{CS}}_t)} \in \mathbb{U}^{\mathbb{R}}_1$, i.e., $\mu^{\mathrm{CS}}_t(X^{\mathrm{CS}}_t \times \cdot)$ is the Lebesgue measure on $\mathbb{R}$.*

Remark 1.26. Using the duality between the Brownian web $\mathcal{W}$ and $\widehat{\mathcal{W}}$, as characterized in Appendix B, it is easily seen that we can also write

$$E^l_t = \{f(t) : f \in \mathcal{W}(x, t - s) \text{ for some } x \in \mathbb{R}, s \geq l\}. \tag{1.47}$$

Theorem 1.27 (Markov property, path continuity, Feller property)**.** *Let $\mathcal{X}^{\mathrm{CS}}$ be as in Proposition 1.25. Then*

(a) *$(\mathcal{X}^{\mathrm{CS}}_t)_{t\geq 0}$ is a $\mathbb{U}^{\mathbb{R}}_1$-valued Markov process;*

(b) *Almost surely, $\mathcal{X}^{\mathrm{CS}}_t$ is continuous in $t \geq 0$;*

(c) *For each $m \in \mathbb{N}$, let $\mathcal{X}^{\mathrm{CS},(m)}$ be a CSSM genealogy process with $\mathcal{X}^{\mathrm{CS},(m)}_0 \in \mathbb{U}^{\mathbb{R}}_1$. If $\mathcal{X}^{\mathrm{CS},(m)}_0 \to \mathcal{X}^{\mathrm{CS}}_0$ in $\mathbb{U}^{\mathbb{R}}_1$, then for any sequence $t_m \to t \geq 0$, we have weak convergence in law, i.e. $\mathcal{X}^{\mathrm{CS},(m)}_{t_m} \Rightarrow \mathcal{X}^{\mathrm{CS}}_t$, as $m \to \infty$.*

(d) *For each initial state in $\mathbb{U}^{\mathbb{R}}_1$, $\mathcal{L}[\mathcal{X}^{\mathrm{CS}}_t]$ converges weakly to the unique equilibrium distribution on $\mathbb{U}^{\mathbb{R}}_1$ as $t \to \infty$.*

Remark 1.28. Notice that this equilibrium state can be represented in terms of a functional of the Brownian web.

Remark 1.29. We note that if we allow types as well, we enlarge the mark space from $\mathbb{R}$ to $\mathbb{R} \times [0, 1]$, where each individual carries a type in $[0, 1]$ that is inherited upon resampling. Theorem 1.27 still holds in this case. We will consider such an extended mark space in Theorem 1.32 below.

Remark 1.30. Proposition 1.25 shows that even though $\mathcal{X}^{\mathrm{CS}}_0$ can be any state in $\mathbb{U}^{\mathbb{R}}_1$, for $t > 0$, $\mathcal{X}^{\mathrm{CS}}_t$ can only take on a small subset of the state space $\mathbb{U}^{\mathbb{R}}_1$. This introduces complications in establishing the continuity of the process at $t = 0$, and it will also be an important point when we discuss the generator of the associated martingale problem.

1.5 Convergence of rescaled IFV genealogies

In this subsection, we establish the convergence of the interacting Fleming-Viot genealogy processes on $\mathbb{Z}$ to that of the CSSM, where we view the states as elements of $\mathbb{U}^{\mathbb{R}}$ (see Remark 1.15). We assume that the transition probability kernel $a(\cdot,\cdot)$ in the definition of the IFV process satisfies

$$\sum_{x\in\mathbb{Z}} a(0,x)x = 0 \qquad \text{and} \qquad \sigma^2 := \sum_{x\in\mathbb{Z}} a(0,x)x^2 = \sum_{x\in\mathbb{Z}} \bar{a}(0,x)x^2 \in (0,\infty). \tag{1.48}$$

For each $\epsilon > 0$, we then define a scaling map $S_\epsilon = S_\epsilon^\sigma : \mathbb{U}^{\mathbb{R}} \to \mathbb{U}^{\mathbb{R}}$ (depending on σ) as follows. Let $\mathcal{X} = \overline{(X,r,\mu)} \in \mathbb{U}^{\mathbb{R}}$. Then

$$S_\epsilon \mathcal{X} := \overline{(X, S_\epsilon r, S_\epsilon \mu)}, \tag{1.49}$$

where $(S_\epsilon r)(x,y) := \epsilon^2 r(x,y)$ for all $x,y \in X$, and $S_\epsilon\mu$ is the measure on $X\times\mathbb{R}$ induced by μ and the map $(x,v) \in X\times\mathbb{R} \to (x,\epsilon\sigma^{-1}v)$, and then the mass rescaled by a factor of $\epsilon\sigma^{-1}$. More precisely,

$$(S_\epsilon\mu)(F) = \epsilon\sigma^{-1}\mu\{(x,\epsilon^{-1}\sigma v) : (x,v) \in F\} \qquad \text{for all measurable } F \subset X\times\mathbb{R}. \tag{1.50}$$

We have the following convergence result for rescaled IFV genealogy processes.

Theorem 1.31 (Convergence of Rescaled IFV Genealogies)**.** *Let $\mathcal{X}^{\mathrm{FV},\epsilon} := (\mathcal{X}_t^{\mathrm{FV},\epsilon})_{t\ge0}$ be an IFV genealogy process on $\mathbb{Z}$ with initial condition $\mathcal{X}_0^{\mathrm{FV},\epsilon} \in \mathbb{U}_1^{\mathbb{Z}}$, indexed by $\epsilon > 0$. Assume that $a(\cdot,\cdot)$ satisfies (1.48), and $S_\epsilon\mathcal{X}_0^{\mathrm{FV},\epsilon} \to \mathcal{X}_0^{\mathrm{CS}}$ for some $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_1^{\mathbb{R}}$ as $\epsilon\to0$.*

Then $(S_\epsilon\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t})_{t\ge0}$ converges as $C([0,\infty),\mathbb{U}^{\mathbb{R}})$-valued random variable weakly to the CSSM genealogy process $\mathcal{X}^{\mathrm{CS}} := (\mathcal{X}_t^{\mathrm{CS}})_{t\ge0}$ as $\epsilon\to0$.

To prove Theorem 1.31, we will need an auxiliary result of interest in its own on the convergence of rescaled measure-valued IFV processes to the measure-valued CSSM. The IFV process with mark space $\mathbb{R}\times[0,1]$ is a measure-valued process $(\widehat{\mathcal{X}}_t^{\mathrm{FV}})_{t\ge0}$, where $\widehat{\mathcal{X}}_t^{\mathrm{FV}}$ is a measure on $\mathbb{R}\times[0,1]$, and its projection on $\mathbb{R}$ is the counting measure on $\mathbb{Z}$. Similarly, the CSSM with mark space $\mathbb{R}\times[0,1]$ is a measure-valued process $(\widehat{\mathcal{X}}_t^{\mathrm{CS}})_{t\ge0}$, where $\widehat{\mathcal{X}}_t^{\mathrm{CS}}$ is a measure on $\mathbb{R}\times[0,1]$, and its projection on $\mathbb{R}$ is the Lebesgue measure on $\mathbb{R}$. Define $\widehat{\mathcal{X}}^{\mathrm{FV}}$ and $\widehat{\mathcal{X}}^{\mathrm{CS}}$ respectively as the projection of the measure component of $\mathcal{X}^{\mathrm{FV}}$ and $\mathcal{X}^{\mathrm{CS}}$, projected onto the mark space $\mathbb{R}\times[0,1]$. We then have

Theorem 1.32 (Convergence of Rescaled IFV Processes)**.** *Let $\widehat{\mathcal{X}}^{\mathrm{FV},\epsilon} := (\widehat{\mathcal{X}}_t^{\mathrm{FV},\epsilon})_{t\ge0}$ be a measure-valued IFV process on $\mathbb{Z}$, indexed by $\epsilon > 0$. Assume that $a(\cdot,\cdot)$ satisfies (1.48), and $S_\epsilon\widehat{\mathcal{X}}_0^{\mathrm{FV},\epsilon}$ converges to $\widehat{\mathcal{X}}_0^{\mathrm{CS}}$ w.r.t. the vague topology for some $\widehat{\mathcal{X}}_0^{\mathrm{CS}}$ as $\epsilon\to0$. Then $(S_\epsilon\widehat{\mathcal{X}}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t})_{t\ge0}$ converges as $C([0,\infty),\mathcal{M}(\mathbb{R}\times[0,1]))$-valued random variable weakly to the CSSM process $\widehat{\mathcal{X}}^{\mathrm{CS}} := (\widehat{\mathcal{X}}_t^{\mathrm{CS}})_{t\ge0}$ as $\epsilon\to0$.*

A similar convergence result has previously been established for the voter model in [AS11].

Remark 1.33. As can be seen from the above convergence results and the regularity properties of the limit process in Proposition 1.25, on a macroscopic scale, there are only locally finitely many individuals with descendants surviving for a macroscopic time of δ or more. This phenomenon leads in the continuum limit to a dynamic driven by a thin subset of hotspots only. For similar effects in other population models, see for example [DEF$^+$02b, DEF$^+$02a].

1.6 Martingale problem for CSSM genealogy processes

In this section, we show that the CSSM genealogy processes solves a martingale problem with a singular generator. To identify the generator L^{CS} for the CSSM genealogy

process $(\mathcal{X}_t^{\mathrm{CS}})_{t\geq 0}$, we note that for all $t>0$, $\mathcal{X}_t^{\mathrm{CS}}$ satisfies the regularity properties established in Proposition 1.25. We will see that L^{CS} is only well-defined on $\Phi\in\Pi^{1,2}$ evaluated at points $\mathcal{X}\in\mathbb{U}^{\mathbb{R}}$ with suitable regularity properties for Φ and $\mathcal{X}$.

We now formalize the subset of regular points in $\mathbb{U}_1^{\mathbb{R}}$ as follows, which satisfies exactly the properties in Proposition 1.25.

Definition 1.34 (Regular class of states $\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$)**.** *Let $\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$ denote the set of $\mathcal{X}=\overline{(X,r,\mu)}\in\mathbb{U}^{\mathbb{R}}$ which satisfies the following regularity properties:*

(a) *$\mathcal{X}\in\mathbb{U}_1^{\mathbb{R}}$, i.e., $\mu(X\times\cdot)$ is the Lebesgue measure on $\mathbb{R}$;*

(b) *there exists a mark function $\kappa:X\to\mathbb{R}$, with $\mu(\mathrm{d}x\mathrm{d}v)=\mu(\mathrm{d}x\times\mathbb{R})\delta_{\kappa(x)}(\mathrm{d}v)$;*

(c) *there exists $\delta>0$ such that for each $l\in(0,\delta)$, X is the disjoint union of balls $(B_i^l)_{i\in\mathbb{Z}}$ of radius l. Furthermore, there exists a locally finite set $E^l:=\{v_i\}_{i\in\mathbb{Z}}\subset\mathbb{R}$ with $v_{i-1}<v_i$ for all $i\in\mathbb{Z}$, such that $\{\kappa(x):x\in B_i^l\}=[v_{i-1},v_i]$ and $\mu(B_i^l)=v_i-v_{i-1}$.*

By Proposition 1.25, $\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$ is closed under the dynamics of $\mathcal{X}^{\mathrm{CS}}$ (i.e., $\mathcal{X}_0^{\mathrm{CS}}\in\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$ implies that $\mathcal{X}_t^{\mathrm{CS}}\in\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$ for all $t>0$), separable, metric measurable subset of the Polish space $\mathbb{U}_1^{\mathbb{R}}$. However, it is *not complete*. Note that the first requirement gives rise to a closed set, the second requirement is known to generate a measurable set [KL15], and it is not hard to see that the third requirement also generates a measurable set.

Remark 1.35. Similar to the discussion leading to (1.42), for $\mathcal{X}\in\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$, we can give a representation on an enriched copy of $\mathbb{R}$ as follows. Property (c) in Definition 1.34 implies that any two disjoint balls in X are mapped by κ to two intervals, which overlap at at most a single point in E^l for some $l>0$. Therefore $\kappa^{-1}(x)$ must contain a single point for all $x\in\mathbb{R}$ with x not in

$$E:=\cup_{l>0}E^l, \tag{1.51}$$

and $\kappa^{-1}(x)$ containing two or more points implies that x is in $\kappa(B_1)\cap\kappa(B_2)$ for two disjoint balls in X. By the same reasoning, for each $x\in E$, $\kappa^{-1}(x)$ must contain exactly two points, which we denote by $x^{\pm}$, where x^+ is a limit point of $\{\kappa^{-1}(w):w>x\}$ and x^- is a limit point of $\{\kappa^{-1}(w):w<x\}$. Similar to (1.42), we can then identify X with $(\mathbb{R}\backslash E)\cup E^+\cup E^-$, where $E^{\pm}:=\{x^{\pm}:x\in E\}$. With this identification and with (1.44), we can simplify our notation (with a slight abuse) and let μ be the measure on $(\mathbb{R}\backslash E)\cup E^+\cup E^-$, which assigns no mass to $E^{\pm}$ and is equal to the Lebesgue measure on $\mathbb{R}\backslash E$.

We now introduce a regular subset of $\Pi^{1,2}$ needed to define the generator L^{CS}.

Definition 1.36 (Regular class of functions $\Pi_{\mathrm{r}}^{1,2}$)**.** *Let $\Pi_{\mathrm{r}}^{1,2}$ denote the set of regular test functions $\Phi^{n,\phi,g}\in\Pi^{1,2}$, defined as in (1.8), with the property that:*

$$\exists\,\delta>0\quad \textit{s.t.}\ \forall\, i\neq j\in\{1,2,\cdots,n\},\quad \frac{\partial\phi}{\partial r_{i,j}}((r_{k,l})_{1\leq k<l\leq n})=0\quad \forall\, r_{i,j}\in[0,\delta]. \tag{1.52}$$

We can now specify the action of the generator L^{CS} on regular Φ evaluated at regular points, namely $L\Phi(\mathcal{X})$ exists for $\Phi=\Phi^{n,\phi,g}\in\Pi_{\mathrm{r}}^{1,2}$ and $\mathcal{X}=\overline{(X,r,\mu)}\in\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$. By Remark 1.35, we can identify X with $(\mathbb{R}\backslash E)\cup E^+\cup E^-$, while μ is identified with the Lebesgue measure on $\mathbb{R}\backslash E$. The generator L^{CS} is given by

$$L^{\mathrm{CS}}\Phi(\mathcal{X}):=L_{\mathrm{d}}^{\mathrm{CS}}\Phi(\mathcal{X})+L_{\mathrm{a}}^{\mathrm{CS}}\Phi(\mathcal{X})+L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}), \tag{1.53}$$

with the component for the massflow (migration) of the population on $\mathbb{R}$ given by

$$L_{\mathrm{d}}^{\mathrm{CS}}\Phi(\mathcal{X}):=\frac{1}{2}\int_{X^n}\phi(\underline{r})\Delta g(\underline{x})\,\mathrm{d}\underline{x}, \tag{1.54}$$

and the component for aging of individuals given by

$$L_{\mathrm{a}}^{\mathrm{CS}}\Phi(\mathcal{X}) := 2\int_{X^n} g(\underline{x}) \sum_{1\le i<j\le n} \frac{\partial\phi}{\partial r_{i,j}}(\underline{r})\, \mathrm{d}\underline{x}. \tag{1.55}$$

These operators are linear operators on the space of bounded continuous functions, $C_b(\mathbb{U}_1^{\mathbb{R}}, \mathbb{R})$, with domain $\Pi^{1,2}$, and maps polynomials to polynomials of the same order.

The component for resampling is, with $\theta_{k,l}\phi$ defined as in (1.19), given by

$$L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}) := \sum_{1\le k\ne l\le n} \int_{X^n} 1_{\{\kappa(x_k)=\kappa(x_l)\}}\, g(\underline{x})\, (\theta_{k,l}\phi - \phi)(\underline{r})\, \mu^*(\mathrm{d}x_k)\mu^*(\mathrm{d}x_l) \prod_{\substack{1\le i\le n\\ i\ne k,l}} \mathrm{d}x_i, \tag{1.56}$$

with effective resampling measure and mark functions

$$\mu^* := \sum_{x\in E} \delta_{x^+} + \sum_{x\in E} \delta_{x^-}, \tag{1.57}$$

$$\kappa(x^{\pm}) = x \text{ for } x \in E, \tag{1.58}$$

where E, and $x^{\pm}$ for $x \in E$, are defined as in Remark 1.35.

Remark 1.37. Note that $L_{\mathrm{r}}^{\mathrm{CS}}$ is *singular*. First because the effective resampling measure μ^* is supported on a countable subset of X and is singular w.r.t. the sampling measure μ on X. Secondly, μ^* is locally infinite because $E \cap (a, b)$ contains infinitely many points for any $a < b$. Therefore the r.h.s. of (1.56) is now in principle a sum of countably many monomials of order $n - 2$.

Indeed, as we partition $X = (\mathbb{R}\backslash E) \cup E^+ \cup E^-$ into balls of radius l, with $l \downarrow 0$, the balls must correspond to smaller and smaller intervals on $\mathbb{R}$ so as not to contradict the fact that each point in X is assigned one value in $\mathbb{R}$. Nevertheless, $L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X})$ in (1.56) is well-defined at least on $\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$ because by our assumption that $\Phi \in \Pi_{\mathrm{r}}^{1,2}$ and condition (1.52) on ϕ, we have $\theta_{k,l}\phi = \phi$ if the resampling is carried out between two individuals x^+ and x^- for some $x \in E$, with $r(x^+, x^-) \le \delta$. Thus only resampling involving $x \in E$ with $r(x^+, x^-) > \delta$ remains in the integration w.r.t. μ^*, and such x are contained in the locally finite set E^δ introduced in Definition 1.34 (c). Together with the assumption that g has bounded support, this implies that the integral in (1.56) is finite.

Remark 1.38. The operator $L_{\mathrm{r}}^{\mathrm{CS}}$, defined on functions in $\Pi_{\mathrm{r}}^{1,2}$ evaluated at regular states in $\mathbb{U}_{\mathrm{r}}^{\mathbb{R}} \subseteq \mathbb{U}_1^{\mathbb{R}}$, is still a linear operator, mapping polynomials to generalized polynomials of degree reduced by two and with domain $\Pi_{\mathrm{r}}^{1,2}$. Here, generalized polynomial means that they are no longer bounded and continuous, and are only measurable functions defined on the subset of points $\mathbb{U}_{\mathrm{r}}^{\mathbb{R}} \subseteq \mathbb{U}_1^{\mathbb{R}}$. Hence L_r^{CS} differs significantly from generators associated with Feller semigroups on Polish spaces.

For $L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X})$ to be well-defined for any $\Phi \in \Pi^{1,2}$, instead of $\Phi \in \Pi_{\mathrm{r}}^{1,2}$, we need to place further regularity assumption on the point $\mathcal{X}$ at which we evaluate Φ. These assumptions are satisfied by typical realizations of the CSSM at a fixed time, as we shall see below.

Definition 1.39 (Regular subclass of states $\mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$)**.** *Let $\mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$ be the set of $\mathcal{X} = \overline{(X, r, \mu)} \in \mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$ with the further property that*

$$\sum_{x\in E\cap[-n,n]} r(x^+, x^-) < \infty \qquad \textit{for all } n \in \mathbb{N}, \tag{1.59}$$

where we have identified X with $(\mathbb{R}\backslash E) \cup E^+ \cup E^-$ as in Remark 1.35.

Theorem 1.40 (Martingale problem for CSSM genealogy processes)**.**

(i) *If $\mathcal{X}_0 \in \mathbb{U}_{\mathrm{I}}^{\mathbb{R}}$, then the $(L^{\mathrm{CS}}, \Pi^{1,2}, \delta_{\mathcal{X}_0})$-martingale problem has a solution, i.e., there exists a process $\mathcal{X} := (\mathcal{X}_t)_{t\geq 0}$ with initial state $\mathcal{X}_0$, with almost surely continuous sample path and with $\mathcal{X}_t \in \mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$ for every $t > 0$, such that for all $\Phi \in \Pi^{1,2}$, and w.r.t. the natural filtration,*

$$\Big(\Phi(\mathcal{X}_t) - \Phi(\mathcal{X}_0) - \int_0^t (L^{\mathrm{CS}}\Phi)(\mathcal{X}_s)ds\Big)_{t\geq 0} \text{ is a martingale.} \tag{1.60}$$

(ii) *The CSSM genealogy process $\mathcal{X}^{\mathrm{CS}}$ constructed in Sec. 1.4 is a solution to the above martingale problem. Apart from the properties established in Prop. 1.25, for each $t > 0$, almost surely $\mathcal{X}_t^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$.*

Remark 1.41. Whether almost surely, $\mathcal{X}_t^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$ for all $t > 0$, remains open.

Remark 1.42. Note that different from the usual martingale or local martingale problems, as for example in [EK86], the test functions here are only defined on a (topologically not closed, only dynamically closed) subset of the state space.

We conjecture that the martingale problems above are in fact well-posed. A proof could be attempted by using the duality between the CSSM genealogy process and the Brownian web. There are however subtle technical complications due to the fact that the generator of the martingale problem is highly singular. We leave this for a future paper.

1.7 Outline

We provide here an outline of the rest of the paper. In Section 2 we prove the results on the IFV genealogy processes. In Section 3 we construct the CSSM genealogy process, and establish in Section 4 the convergence of the IFV genealogies to those of the CSSM, and in Section 5 results on the martingale problem for the CSSM genealogy processes. In Appendix A, we collect further facts and proofs concerning marked metric measure spaces. In Appendix B, we recall the construction of the Brownian web and its dual, and collect some basic properties of the Brownian web and coalescing Brownian motions. In Appendix C we prove some results on coalescing Brownian motions needed in our estimates to derive the martingale problem for the CSSM.

2 Proof of Theorems 1.12 and 1.17

In this section we present the proof of the results on the evolving genealogies for the interacting Fleming-Viot diffusions. This model is a special case of evolving genealogies for the interacting Λ-Fleming-Viot diffusions which are studied in [GKW].

Proof of Theorem 1.12. We will proceed in five steps: (1) We show the result on the martingale problem and the duality for finite geographic spaces V. (2) To prepare for the general case where V is countable, we define an approximation procedure with specific finite geographic space dynamics. (3) We then show the convergence in path space, as the finite spaces approach V. (4) We verify the claimed properties of the solution for general V by a direct argument based on the duality and an explicit look-down construction. (5) Finally, we show that the process admits a mark function.

We will use several known facts on measure-valued Fleming-Viot diffusions. For that we refer to [Daw93, Chapter 5] for the non-spatial case and to [DGV95] for the spatial case.

Step 1 (V finite) The case where V is finite is very similar to the non-spatial case. We therefore just have to modify the arguments of the proof of Theorem 1 in [GPW13] or Theorem 1 in [DGP12].

As usual we will conclude *uniqueness* of the solution of the martingale problem from a duality relation. Note that in contrast to the (non-spatial) interacting Fleming-Viot model with mutation considered in [DGP12], in our spatial interacting Fleming-Viot model, resampling takes place only locally, that is, for individuals at the *same* site. The tree-valued dual will therefore be based on the spatial coalescent considered in [GLW05]. As the calculations to verify the duality relation are the same as in [GPW13] and in [DGP12], we omit them here.

As for *existence* of a solution of the martingale problem we consider the martingale problems for the evolving genealogies of the approximating spatial Moran models. By consistency of the spatial coalescent, we get the uniform convergence of generators for free. Thus we only have to show the compact containment condition. Here we can rely on the general criterion for population dynamics given in Proposition 2.22 in [GPW13]. As V is finite, all arguments given in [GPW13] to verify this criterion simply go through here as well.

Step 2 (A coupled family of approximating finite systems) Let now V be countable, and consider a sequence $(V_n)_{n\in\mathbb{N}}$ of finite sets with $V_n \subseteq V$, and $V_n \uparrow V$. Put for each $n \in \mathbb{N}$, and for all $v_1, v_2 \in V$,

$$a_n\big(v_1, v_2\big) := \begin{cases} \frac{a(v_1,v_2)1_{V_n\times V_n}(v_1,v_2)}{\sum_{v_3\in V_n} a(v_1,v_3)}, & \text{if } v_1 \in V_n, \\ \delta(v_1, v_2), & \text{if } v_1 \notin V_n. \end{cases} \tag{2.1}$$

Denote then by $\mathcal{X}^{\mathrm{FV},V_n}$ a solution of the martingale problem associated with the operator (restricted to V_n)

$$\begin{aligned} L^{\mathrm{FV},V_n}\Phi\big(\mathcal{X}\big) = 2\int_{(X\times V)^n} \mu^{\otimes n}(\mathrm{d}(\underline{x},\underline{v}))\, g(\underline{v}) \sum_{1\le k<\ell\le n} \frac{\partial}{\partial r_{k,\ell}}\, \phi(\underline{r}) \\ +\ \int_{(X\times V)^n} \mu^{\otimes n}(\mathrm{d}(\underline{x},\underline{v}))\, \phi(\underline{r}) \sum_{j=1}^{n} \sum_{v'\in V} \bar{a}_n(v_j, v')(M_{v_j,v'}g - g)(\underline{v}) \\ +\ 2\,\gamma \int_{(X\times V)^n} \mu^{\otimes n}(\mathrm{d}(\underline{x},\underline{v}))\ g(\underline{v}) \sum_{1\le k<\ell\le n} 1_{\{v_k=v_\ell\}}\big(\theta_{k,\ell}\phi - \phi\big)(\underline{r}), \end{aligned} \tag{2.2}$$

and $\mathcal{K}^{V_n}$ the spatial coalescent on V_n with migration rate $\bar{a}_n(\cdot,\cdot)$ rather than $\bar{a}(\cdot,\cdot)$.

Notice that a_n is not necessarily double stochastic anymore, which turns the duality with the spatial coalescent into a *Feynman-Kac duality* where the Feynman-Kac term converges to 1, as $n \to \infty$, on every finite time horizon. The Feynman-Kac duality reads as follows (compare, for example, [Sei14, Proposition 3.11]): for all $\mathcal{X}_0 \in \mathbb{U}_1^V$,

$$\mathbb{E}\big[H(\mathcal{X}_t^{\mathrm{FV},V_n}, \mathcal{K}_0^{V_n})\big] = \mathbb{E}\big[H(\mathcal{X}_0, \mathcal{K}_t^{V_n}) \exp\big(\int_0^t A(\mathcal{K}_s^{V_n})\mathrm{d}s\big)\big], \tag{2.3}$$

where

$$A\big((\pi, \xi', \underline{r})\big) := \sum_{i\in\pi}\big(\sum_{v'\in V} a_n(v', \xi_i') - 1\big). \tag{2.4}$$

which is bounded along the path by $|\pi|\cdot$ Const, for all $t \ge 0$.

Establishing the Feynman-Kac duality requires to check that (compare, Section 4.4 in [EK86, Theorem 4.4.11]):

$$L^{\mathrm{FV}}H\big(\cdot,\mathcal{K}\big)\big)(\mathcal{X}) = L^{\mathrm{dual}}H\big(\mathcal{X},\cdot\big)(\mathcal{K}) + A(\mathcal{K})\cdot H\big(\mathcal{X},\mathcal{K}\big). \tag{2.5}$$

This can be immediately verified by explicit calculation (compare, [GPW13, Section 4] for the generator calculation for the resampling part, and [Sei14, Proposition 3.11] for

the generator calculation for Markov chains - here migration - whose transition matrix is not double stochastic).

As for given $n \in \mathbb{N}$, our dynamics consists of independent components outside V_n, we can apply Step 1, here with the Feynman-Kac duality, to conclude the well-posedness of the martingale problem with respect to L^{FV,V_n}.

Step 3 (V countable) Fix $\mathcal{X}_0 \in \mathbb{U}_1^V$. In this step we want to show that the family $\{\mathcal{X}^{\mathrm{FV},V_n};\, n \in \mathbb{N}\}$ is tight, and that every limit satisfies the $(L^{\mathrm{FV}}, \Pi^{1,2}, \delta_{\mathcal{X}_0})$-martingale problem.

Observe first that the family of laws of the projection of the measures on mark space are tight since the localized state w.r.t. a fixed finite subset A of V has only finitely many marks and weight $|A|$ (uniformly in n). We therefore will here ignore the marks and show tightness in Gromov-weak$^{\#}$-topology. For that we want to apply [EK86, Corollary 4.5.2]. Since $a_n(v_1, v_2) \to a(v_1, v_2)$ for all $v_1, v_2 \in V$, we clearly have that $L^{\mathrm{FV},V_n}\Phi$ converges uniformly to $L^{\mathrm{FV}}\Phi$, as $V_n \uparrow V$, i.e., $\sup_{\mathcal{X}\in\mathbb{U}^V} |L^{\mathrm{FV},V_n}\Phi(\mathcal{X}) - L^{\mathrm{FV}}\Phi(\mathcal{X})| \underset{n\to\infty}{\longrightarrow} 0$, for Φ depending only on finitely many sites, it remains to verify the *compact containment condition*, i.e., to show that for every $T > 0$, and $\varepsilon > 0$ we can find a compact set $K_{T,\varepsilon} \subseteq \mathbb{U}^V$ such that for all $n \in \mathbb{N}$,

$$P\big(\mathcal{X}_t^{\mathrm{FV},V_n} \in K_{T,\varepsilon};\ \text{for all } t \in [0,T]\big) \geq 1 - \varepsilon. \tag{2.6}$$

For that purpose we will once more rely on the criterion for the compact containment condition which was developed in [GPW13, Proposition 2.22] for population dynamics. To see first that the criterion applies, notice that the evolving genealogies of interacting Fleming-Viot diffusions can be read off as a functional of the *look-down construction* given in [GLW05]. Thus the countable representation of the look-down defines a population dynamics. In particular, for each $t \geq 0$, we can read off a representative (X_t, r_t, μ_t) of $\mathcal{X}_t^{\mathrm{FV}}$ in the look-down graph such that *ancestor-descendant* relationship is well-defined. Denote for all $t \geq 0$, $s \in [0,t]$ and $x \in X_t$ by $A_t(x,s) \in X_{t-s}$ the ancestor of $x \in X_t$ back at time s, and for $\mathcal{J} \subseteq X_s$ by $D_t(\mathcal{J}, s) \subseteq X_t$ the set of descendants of a point in $\mathcal{J}$ at time t.

In the following we refer for each finite $A \subseteq V$ to

$$\mathcal{X}^{\mathrm{FV},\cdot,A} \tag{2.7}$$

as the restriction of $\mathcal{X}^{\mathrm{FV},\cdot}$ to marks in A, i.e., obtained by considering the sampling measure $\mu^A(dxdv) := 1_A\mu(dxdv)$. Fix $T > 0$. We then have two show that the following properties are true for all $A \subseteq V$,

- **Tightness of number of ancestors.** For all $t \in [0,T]$ and $\varepsilon \in (0,t)$, the family $\{S_{2\varepsilon}^{V_n}(X_t, r_t, \mu_t);\, n \in \mathbb{N}\}$ is tight, where $S_{2\varepsilon}^{V_n}(X_t, r_t, \mu_t)$ denotes the minimal number of balls of radius 2ε needed to cover X_t up to a set of μ_t-measure ε.
- **Bad sets can be controlled.** For all $\varepsilon \in (0,T)$, there exists a $\delta = \delta(\varepsilon) > 0$ such that for all $s \in [0,T)$, $n \in \mathbb{N}$ and $\sigma(\mathcal{X}_u^{\mathrm{FV},V_n};\, u \in [0,s])$-measurable random subsets $\mathcal{J}^{V_n} \subseteq X_s \times A$ with $\mu_s(\mathcal{J}^{V_n}) \leq \delta$,

$$\limsup_{n\in\mathbb{N}} P\Big(\sup_{t\in[s,T]} \mu_t\big(D_t^{V_n}(\mathcal{J}^{V_n}, s) \times A\big) \geq \epsilon\Big) \leq \epsilon. \tag{2.8}$$

(i) Fix $t \in [0,T]$. W.l.o.g. we assume that V_n is a subgroup of V with addition $+_n$ for each $n \in \mathbb{N}$. We consider for each $n \in \mathbb{N}$ another spatial coalescent $\tilde{K}^{V_n}$ on V_n with migration kernel $\bar{\tilde{a}}_n(\cdot,\cdot)$ rather than $\bar{a}_n(\cdot,\cdot)$ where

$$\tilde{a}_n(v, v') := \sum_{y\in V; y\sim_n v'} a(v, y), \tag{2.9}$$

where $\sim_n$ denotes equivalence modulo $+_n$. Further, denote by $\tilde{\mathcal{X}}^{\mathrm{FV},V_n}$ the evolving genealogies of the interacting Fleming-Viot diffusions whose migration kernel is $\tilde{a}_n(\cdot,\cdot)$ rather than $a_n(\cdot,\cdot)$. We prefer to work with $\tilde{K}^{V_n}$. For this spatial coalescent it was verified in the proof of Proposition 3.4 in [GLW05] that for any time $t > 0$ the total number of partition elements of $\tilde{K}_t^{V_n}$ which are located in A is stochastically bounded uniformly in $n \in \mathbb{N}$. As the kernel $\tilde{a}_n(\cdot,\cdot)$ is double stochastic, $\tilde{\mathcal{X}}^{\mathrm{FV},V_n}$ and $\tilde{K}^{V_n}$ are dual (without a Feynman-Kac potential), and as $d_{\mathrm{GP}^\#}(\tilde{\mathcal{X}}^{\mathrm{FV},V_n}, \mathcal{X}^{\mathrm{FV},V_n}) \to 0$, as $n \to \infty$, the claim follows.

(ii) Fix $T > 0$, $A \subseteq V$ finite, $\varepsilon \in (0,T)$, $s \in [0,T)$, $n \in \mathbb{N}$ and a $\sigma(\mathcal{X}_u^{\mathrm{FV},V_n};\, u \in [0,s])$-measurable random subset $\mathcal{J}^{V_n} \subseteq X_s \times A$. From the generator characterization of $\mathcal{X}^{\mathrm{FV},V_n}$, we can conclude that the process $\{\mu_t(D_t(\mathcal{J}^{V_n}, s));\, t \geq s\}$ is a V-indexed system of interacting (measure-valued) Fisher-Wright diffusions. We have to find $\delta = \delta(\varepsilon)$ such that (2.8) holds if $\mu_s(\mathcal{J}^{V_n}) \leq \delta$.

Notice that $\{\mu_t(D_t^{V_n}(\mathcal{J}^{V_n}, s) \times A);\, t \geq s\}$ is a *semi-martingale* and given by a martingale with continuous paths due to resampling plus a *deterministic flow* in and out of the set A due to migration. Therefore we have to control the fluctuation of the martingale part and the maximal flow out of the set A over a time interval of length $t - s$.

The martingale part is estimated from below with Doob's maximum inequality (the quadratic variation is bounded uniformly in the state and in n by a constant $\cdot|A|$, details are left to the reader). The deterministic out flow occurs at most at a finite rate $c \times |A|$ since the total mass of every site is one. This estimate is uniform in the parameter n (recall the random walk kernel is perturbed by restricting it to V_n). Similarly the flow into A can be bounded by $c \cdot |A|$ independently of n but the flow out of the set A occurs with a maximal rate c independently of n with $n \geq n_0(A)$, and

$$c = 2 \cdot \max\Big\{ \max_{v' \in A^c, v \in A} a(v', v),\, \max_{v' \in A^c, v \in A} a(v', v) \Big\}. \tag{2.10}$$

Step 4 (Feller property). We next prove that $\mathcal{X}^{\mathrm{FV}}$ has the Feller property. From here it is standard to conclude that $\mathcal{X}^{\mathrm{FV}}$ satisfies the strong Markov property (see, for example, [EK86, Theorem 4.2.7]). Consider a sequence $(\mathcal{X}_0^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{U}_1^V$ such that $\mathcal{X}^{(n)} \underset{n\to\infty}{\longrightarrow} \mathcal{X}_0$, Gromov-weak$^\#$ly, for some $\mathcal{X}_0 \in \mathbb{U}_1^V$. Denote by $\mathcal{X}^{\mathrm{FV},\mathcal{X}_0^{(n)}}$ and $\mathcal{X}^{\mathrm{FV},\mathcal{X}_0}$ the evolving genealogies of the interacting Fleming-Viot diffusions started in $\mathcal{X}_0^{(n)}$ and $\mathcal{X}_0$, respectively, and let $\mathcal{K}$ be our tree-valued dual spatial coalescent, and H as in (1.21). Then for each given $t \geq 0$,

$$\mathbb{E}\big[H\big(\mathcal{X}_0^{(n)}, \mathcal{K}_t\big)\big|\mathcal{K}_t\big] \underset{n\to\infty}{\longrightarrow} \mathbb{E}\big[H\big(\mathcal{X}_0, \mathcal{K}_t\big)\big|\mathcal{K}_t\big], \qquad a.s. \tag{2.11}$$

Thus, by our duality relation,

$$\begin{aligned} \mathbb{E}\big[H(\mathcal{X}_t^{\mathrm{FV},\mathcal{X}_0^{(n)}}, \mathcal{K}_0)\big] &= \mathbb{E}\big[H(\mathcal{X}_0^{(n)}, \mathcal{K}_t)\big] \\ &\underset{n\to\infty}{\longrightarrow} \mathbb{E}\big[H\big(\mathcal{X}_0, \mathcal{K}_t\big)\big] = \mathbb{E}\big[H\big(\mathcal{X}_t^{\mathrm{FV},\mathcal{X}_0}, \mathcal{K}_0\big)\big]. \end{aligned} \tag{2.12}$$

Recall from Remark R:gs6 that the family $\{H^{n,\phi}(\cdot,\mathcal{K});\, n \in \mathbb{N}, \phi \in C_b(\mathbb{R}_+^{\binom{n}{2}}); \mathcal{K} \in \mathbb{S}_n\}$ is convergence determining. Thus it follows that $\mathcal{X}_t^{\mathrm{FV},\mathcal{X}_0^{(n)}} \underset{n\to\infty}{\Longrightarrow} \mathcal{X}_t^{\mathrm{FV},\mathcal{X}_0}$, for all $t \geq 0$.

Step 5 (Mark function) Fix $T \geq 0$, and $\mathcal{X}_0 \in \mathbb{U}_{\mathrm{fct}}^V$.

For the proof we will rely once more on the approximation of the solution of the $(L^{\mathrm{FV}}, \Pi^{1,0}, \delta_{\mathcal{X}_0})$-martingale problem by $\mathbb{U}_{\mathrm{fct}}^V$-valued evolving genealogies of Moran models, $\mathcal{X}^{\mathrm{M},\rho}$, where $\rho > 0$ is the local intensity of individuals. By the look-down construction given in [GLW05], we can define the family $\{\mathcal{X}^{\mathrm{M},\rho};\, \rho > 0\}$ on one and the same probability

space. Moreover, as the solution of the $(L^{\mathrm{FV}}, \Pi^{1,0}, \delta_{\mathcal{X}_0})$-martingale problem has continuous paths, due to Skorohod representation theorem we may assume that $\mathcal{X}_t^{\mathrm{M},\rho} \to \mathcal{X}_t^{\mathrm{FV}}$, as $\rho \to \infty$, uniformly for all $t \in [0,T]$, almost surely.

We will rely on Theorem 3.9 in [KL15]. That is, as the solutions of the $(L^{\mathrm{FV}}, \Pi^{1,0}, \delta_{\mathcal{X}_0})$–martingale problem have continuous paths almost surely, we have to construct for each finite $A \subseteq V$, for each $t \in [0,T]$, and $\epsilon, \delta, \rho > 0$ a function $h_{t,\epsilon,A} \in \mathcal{H}$ (recall from (1.13)) and a random measurable set $Y^{\rho}_{t,\epsilon,\delta,A} \subseteq X^{\rho}_t$ such that

$$\limsup_{\delta \downarrow 0} \limsup_{\rho \to \infty} \mathbb{P}\{\mu_t(X^{\rho} \setminus Y^{\rho}_{t,\epsilon,\delta,A}) \leq h_{t,\epsilon,A}(\delta),\ \forall t \in [0,T]\} \geq 1 - \epsilon. \tag{2.13}$$

Assume first that the geographic space V is finite. For the construction of such a function $h_{\epsilon,V} = h_\epsilon \in \mathcal{H}$ and a random measurable set $Y^{\epsilon}_{\rho,\delta,V} = Y^{\epsilon}_{\rho,\delta} \subseteq X^{\rho}_t$, we can proceed exactly as in the proof of Theorem 4.3 of [KL15] where the statement is shown with mutation rather than migration in the non-spatial rather than the finite geographic space.

Let now V be countable, and consider a sequence $(V_n)_{n\in\mathbb{N}}$ of finite sets with $V_n \subseteq V$ and $V_n \uparrow V$. Consider for each $n \in \mathbb{N}$ a solution, $\mathcal{X}^{\mathrm{FV},V_n}$, of the $(L^{\mathrm{FV},V_n}, \Pi^{1,0}, \delta_{\mathcal{X}_0})$-martingale problem with L^{FV,V_n} as defined in (2.2). As we have seen above, $\mathcal{X}_t^{\mathrm{FV},V_n} \in \mathbb{U}_{\mathrm{fct}}^{V_n}$ for all $t \geq 0$, almost surely. Moreover, we have shown in Step 3 that each solution $\mathcal{X}^{\mathrm{FV}}$ of the $(L^{\mathrm{FV}}, \Pi^{1,0}, \delta_{\mathcal{X}_0})$-martingale problem on V can be obtained as the limit of $\mathcal{X}^{\mathrm{FV},V_n}$ as $n \to \infty$. To conclude from here that also $\mathcal{X}_t^{\mathrm{FV}} \in \mathbb{U}_{\mathrm{fct}}^{V}$ for all $t \geq 0$, almost surely, fix a finite set $A \subseteq V$. As done before we denote by $\mathcal{X}^{\mathrm{FV},V_n,A} = (X_t, r_t^{V_n}, \mu_t^{V_n}(\cdot\times(\cdot\cap A)))_{t\geq 0}$ and $\mathcal{X}^{\mathrm{FV},A} = (X_t, r_t, \mu_t(\cdot \times (\cdot \times A)))_{t\geq 0}$ the restrictions of $\mathcal{X}^{\mathrm{FV},V_n}$ and $\mathcal{X}^{\mathrm{FV}}$, respectively, to marks in A (compare (2.7)).

For each $m > n$ we couple $\mathcal{X}^{\mathrm{FV},V_n}$ and $\mathcal{X}^{\mathrm{FV},V_m}$ through the graphical lookdown construction by using the same Poisson point processes and marking every path which leaves V_n in the V_m dynamics by a 1. Moreover, we impose the rule that the 1 is inherited upon lookdown in the sense that *both* new particles carry type 1. The sampling measure of types then follows an interacting Fleming-Viot (in fact two-type Fisher-Wright) diffusion with selection. The corresponding Moran models are coupled and converge in the many particle per site limit to a limit evolution, which is the coupling on the finite geographic spaces and the additional types act upon resampling as under selection.

By construction, if $x, x' \in \mathrm{supp}(\mu_t(\cdot \times A))$, their distance is the same in $\mathcal{X}^{\mathrm{FV},V_n}$ and $\mathcal{X}^{\mathrm{FV},V_m}$ if both carry type 0. Thus for suitably large n (depending on $\epsilon > 0$) such that $A \subseteq V_n$ at any location in A the relative frequencies of types 1 at time t can be made less than any given $\varepsilon > 0$ with probability $\geq 1 - \varepsilon$ by simple random walk estimates. Namely, if $(Z_t^b)_{t\geq 0}$ is a $a(\cdot,\cdot)$-random walk starting in $b \in V_n \subseteq V_m$ and $b' \in \mathcal{C}V_n$,

$$\begin{aligned} &\mathbb{P}(Z_t^b \notin V_n \text{ for some } t \in [0,T], Z_T^b \in A) \\ &\quad + \mathbb{P}(Z_t^{b'} \in V_n \text{ for some } t \in [0,T], Z_T^{b'} \in A) \leq \delta_n \to 0 \text{ as } n \to \infty\ \forall\ m \geq n. \end{aligned} \tag{2.14}$$

Then the expected frequency of type 1 in locations in A is bounded by $F(\delta_m)$ with $F(\delta) \to 0$ as $\delta \to 0$, which follows from the properties of the Fisher-Wright diffusion with selection easily via duality.

As a consequence the supremum along the path of the difference in variational norm of the distance-mark distributions for the V_n and the V_m-evolution for types in the set A can be bounded by a sequence converging to 0 as $n, m \to \infty$.

Therefore also the limit dynamics on countable V has a mark function. ∎

Proof of Theorem 1.17. Let $\mathcal{X}^{\mathrm{FV}}$ be the evolving genealogies of interacting Fleming-Viot diffusions where we have assumed that the symmetrized migration is recurrent. In order to prove ergodicity we proceed in two steps: (1) We start with constructing the limiting

object which is *tree-valued spatial coalescent*. (2) We then prove convergence of $\mathcal{X}_t^{\mathrm{FV}}$ to this tree-valued spatial coalescent, at $t \to \infty$, for any initial state $\mathcal{X}_0 \in \mathbb{U}_1^V$. This immediately implies uniqueness of the invariant distribution.

Step 1 (Tree-valued spatial coalescent) Recall from [GLW05] the spatial coalescent started with infinitely many partition elements per site and with migration mechanism $\bar{a}(v, v')$. If the symmetrized migration is recurrent, we can assign to each realization a marked ultrametric space, $\mathcal{K} = (K, r)$, which admits a mark function. In order to equip it with a locally finite measure on the leaves, we consider a coupled family of sub-coalescents $\{\mathcal{K}^\varrho,\, \rho > 0\}$ such that the number of points of a given mark is Poisson with intensity ρ. If we now assign each point in $\mathcal{K}^\rho$ mass ρ^{-1}, then it follows from [GLW05, Theorem 3] that there exist a measure μ on $K \times V$ such that for each $v \in V$, $\mu(K \times \{v\}) = 1$. This reflects the spatial dust-free property. Thus, we can use the same arguments used in [GPW09, Theorem 4] to show that the family $\{(\mathcal{K}^\rho, \rho^{-1} \sum_{(x,v)\in K^\rho \times V} \delta_{(x,v)}\}$ is tight, and in fact has exactly one limit point,

$$\mathcal{K}^\downarrow := \overline{(K, r^\downarrow, \mu^\downarrow)} \tag{2.15}$$

Step 2 (Convergence into the tree-valued spatial coalescent) For all $\mathcal{X}_0 \in \mathbb{U}_1^V$ and $\mathcal{K}_0 \in \mathbb{S}$, by our duality relation, using the functions $H = H^{n,\phi}$ from (1.21),

$$\begin{aligned} \mathbb{E}\big[H\big(\mathcal{X}_t^{\mathrm{FV}}, \mathcal{K}_0\big)\big] &= \mathbb{E}\big[H\big(\mathcal{X}_0, \mathcal{K}_t\big)\big] \\ &\xrightarrow[t\to\infty]{} \mathbb{E}^{\mathcal{K}_0}\big[\phi\big((r^\downarrow(i,j))_{1\le i<j\le n}\big)\big] \\ &= \mathbb{E}\big[H\big(\mathcal{K}^\downarrow, \mathcal{K}_0\big)\big]. \end{aligned} \tag{2.16}$$

Once more, as the family $\{H^{n,\phi}(\cdot, \mathcal{K});\, n \in \mathbb{N}, \phi \in C_{bb}(\mathbb{R}_+^{\binom{n}{2}}), \mathcal{K} \in \mathbb{S}\}$ is convergence determining by Theorem 1.8(i), we can conclude that for all initial conditions, $\mathcal{X}_t^{\mathrm{FV}}$ converges Gromov-#-weakly to the tree-valued spatial coalescent. ∎

3 Proofs of the properties of CSSM genealogy processes

In this section, we prove Proposition 1.25 and Theorem 1.27 by using properties of the double Brownian web $(\mathcal{W}, \widehat{\mathcal{W}})$, which was used to construct the CSSM genealogy process $\mathcal{X}^{\mathrm{CS}}$ in Section 1.4.

Proof of Prop. 1.25. (a): The existence of a mark function $\kappa : X_t^{\mathrm{CS}} \to \mathbb{R}$ follows by construction (recall (1.44) where the mark is explicitly given). The continuity of κ follows from the property of the dual Brownian web $\widehat{\mathcal{W}}$. More precisely, if $x_n \to x$ in $(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}})$, then identifying x_n and x with points on $\mathbb{R}$, it follows that $\hat{f}_{(x_n,t)} \to \hat{f}_{(x,t)}$ in $\widehat{\Pi}$ for some path $\hat{f}_{(x_n,t)} \in \widehat{\mathcal{W}}(x_n, t)$ and $\hat{f}_{(x,t)} \in \widehat{\mathcal{W}}(x, t)$, which implies that $x_n \to x$ in $\mathbb{R}$.

(b): By (1.42), we identify X_t^{CS} with $\mathbb{R}$, where a countable subset E_t is duplicated. The distance between $x, y \in X_t^{\mathrm{CS}}$ is defined to be twice of the time to coalescence between the dual Brownian web paths $\hat{f}_{(x,t)} \in \widehat{\mathcal{W}}(x, t)$ and $\hat{f}_{(y,t)} \in \widehat{\mathcal{W}}(y, t)$, if the two paths coalesce above time 0. Therefore for $l \in (0, t)$, each ball B_i^l of radius l correspond to a maximal interval $[v_{i-1}, v_i] \subset \mathbb{R}$, where all paths in $\{\hat{f}_{(x,t)} \in \widehat{\mathcal{W}}(x, t) : x \in (v_{i-1}, v_i)\}$ coalesce into a single path by time $t - l$. The collections of such maximal intervals (v_{i-1}, v_i) form a partition of $\mathbb{R}\backslash E_t^l$, where $E_t^l = \{v_i : i \in \mathbb{Z}\}$ is exactly the set defined in (1.46).

(c): Fix an $l \in (0, t)$. By construction, for each ball B_i^l of radius l, $\kappa(B_i^l) = [v_{i-1}, v_i]$ is assigned the Lebesgue measure on $\mathbb{R}$. Together with (b), it implies that $\mu_t^{\mathrm{CS}}(X_t^{\mathrm{CS}} \times \cdot)$ is the Lebesgue measure on $\mathbb{R}$. ∎

Proof of Theorem 1.27. Let us fix a realization of the Brownian web $\mathcal{W}$ and its dual $\widehat{\mathcal{W}}$, and let $\mathcal{X}^{\mathrm{CS}}$ be constructed from $\widehat{\mathcal{W}}$ as just before Proposition 1.25. By (1.42), for each $t > 0$, we can identify X_t^{CS} with $A_t \cup E_t^+ \cup E_t^-$, where A_t and E_t are defined as in (1.38). To simplify notation, we will drop the superscript CS in the remainder of the proof.

(a): The Markov property of $(\mathcal{X}_t)_{t\geq 0}$ follows from the Markov property of $\mathcal{W}$ and $\widehat{\mathcal{W}}$. More precisely, if we denote by $f|_s^t$ the restriction of a path $f \in \Pi$ to the time interval $[s,t]$, and $K|_s^t := \{f|_s^t : f \in K\}$ for a set of paths $K \subset \Pi$, then $\mathcal{W}|_0^s$ is independent of $\mathcal{W}|_s^\infty$ for each $s \geq 0$. The same is true for $\widehat{\mathcal{W}}$ since $\mathcal{W}|_0^s$ and $\widehat{\mathcal{W}}|_0^s$ a.s. uniquely determine each other by Theorem B.3. The Markov process $(\mathcal{X}_t)_{t\geq 0}$ is time homogeneous because $\mathcal{W}|_0^t$ is equally distributed with $\mathcal{W}|_s^{s+t}$, apart from a time shift.

(b): Let $\mathcal{X}_0 \in \mathbb{U}_1^{\mathbb{R}}$. We first prove that $(\mathcal{X}_t)_{t\geq 0}$ is a.s. continuous in $t > 0$. To accomplish this, since $\Pi^{1,2}$ is convergence determining in $\mathbb{U}^{\mathbb{R}}$ as shown in Theorem 1.8, it suffices to show that for any $\Phi := \Phi^{n,\phi,g} \in \Pi^{1,2}$, the evaluated polynomial

$$\Phi(\mathcal{X}_t) = \int_{\mathbb{R}^n} \phi(\underline{\underline{r_t}}) g(\underline{x}) \mathrm{d}\underline{x} \tag{3.1}$$

is continuous in t, where $\underline{x} := (x_1, \ldots, x_n)$, $\underline{\underline{r_t}} := (r_t(x_i, x_j))_{1\leq i<j\leq n}$, and given the identification between X_t and $A_t \cup E_t^+ \cup E_t^-$ and the definition of κ, we have replaced integration w.r.t. μ_t on $X_t \times \mathbb{R}$ by integration w.r.t. the Lebesgure measure on $\mathbb{R}$. By (1.40), for Lebesgue a.e. $x_i, x_j \in \mathbb{R}$, we have

$$r_t(x_i, x_j) = \begin{cases} 2(t - \hat{\tau}) & \text{if } \hat{\tau} \geq 0, \\ 2t + r_0(\xi(u), \xi(v)) & \text{if } \hat{\tau} < 0, \end{cases} \tag{3.2}$$

where $\hat{\tau}$ is the time of coalescence between $\hat{f}_{(x,t)}$ and $\hat{f}_{(y,t)}$, $\hat{f}_{(x,t)}(0) = u$, and $\hat{f}_{(x,t)}(0) = v$. By Lemma B.4, for each $t > 0$, $\widehat{\mathcal{W}}(x,t)$ contains a single path for all but a countable number of $x \in \mathbb{R}$. For such x, by Lemma B.2, the time of coalescence between $\hat{f}_{(x,s)}$ and $\hat{f}_{(x,t)}$ tends to t as $s \to t$, and hence $\lim_{s\to t} r((x,s),(x,t)) = 0$, where $r((x,s),(x,t))$ is defined in (1.41) and extends the definition of $r_t(x,y)$ to individuals at different times. Since

$$|r_s(x_i, x_j) - r_t(x_i, x_j)| \leq r((x_i, s), (x_i, t)) + r((x_j, s), (x_j, t)), \tag{3.3}$$

it follows that when $t > 0$, for Lebesgue a.e. $x_i, x_j \in \mathbb{R}$, $1 \leq i < j \leq n$, we have

$$\lim_{s\to t} r_s(x_i, x_j) = r_t(x_i, x_j). \tag{3.4}$$

We can then apply the dominated convergence theorem in (3.1) to deduce that, for each $t > 0$, a.s.

$$\lim_{s\to t} \Phi(\mathcal{X}_s) = \Phi(\mathcal{X}_t). \tag{3.5}$$

This verifies that $(\mathcal{X}_t)_{t\geq 0}$ is a.s. continuous in $t > 0$.

Proving the a.s. continuity of $(\mathcal{X}_t)_{t\geq 0}$ at $t = 0$ poses new difficulties because $\mathcal{X}_0$ can be any state in $\mathbb{U}_1^{\mathbb{R}}$, while for any $t > 0$, $\mathcal{X}_t$ is a regular state as shown in Proposition 1.25. We get around this by showing that $\mathcal{X}$ admits a càdlàg version. More precisely, we invoke a part of the proof of the convergence Theorem 1.31 that is independent of the current proof. Note that for any $\mathcal{X}_0 \in \mathbb{U}_1^{\mathbb{R}}$, we can find a sequence $\mathcal{X}_0^{\mathrm{FV},\epsilon} \in \mathbb{U}_1^{\mathbb{Z}}$, indexed by $\epsilon > 0$, such that $S_\epsilon \mathcal{X}_0^{\mathrm{FV},\epsilon} \to \mathcal{X}_0$. Indeed, we only need to approximate the mark space $\mathbb{R}$ by $\epsilon\mathbb{Z}$ in order to construct $S_\epsilon \mathcal{X}_0^{\mathrm{FV},\epsilon}$ from $\mathcal{X}_0$. In the proof of Theorem 1.31, it is shown that the corresponding sequence of interacting Fleming-Viot genealogy process $(S_\epsilon \mathcal{X}_{\epsilon^{-2}t}^{\mathrm{FV},\epsilon})_{t\geq 0}$ is a tight family of $D([0,\infty), \mathbb{U}^{\mathbb{R}})$-valued random variables, where $D([0,\infty), \mathbb{U}^{\mathbb{R}})$ denotes the space of càdlàg paths on $\mathbb{U}^{\mathbb{R}}$ equipped with the Skorohod topology. Furthermore,

$(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t})_{t\geq 0}$ converges in finite-dimensional distribution to the CSSM genealogy process $(\mathcal{X}_t)_{t\geq 0}$. Therefore, $(\mathcal{X}_t)_{t\geq 0}$ must admit a version which is a.s. càdlàg, with $\mathcal{X}_t \to \mathcal{X}_0$ as $t \downarrow 0$. Since we have just shown that the version of $(\mathcal{X}_t)_{t\geq 0}$ constructed in Sec. 1.4 is a.s. continuous in $t > 0$, it follows that the same version must also be a.s. càdlàg, and hence continuous at $t = 0$, which concludes the proof of part (b).

(c): To prove the Feller property, let $\mathcal{X}^{(m)}_0 \to \mathcal{X}_0$ in $\mathbb{U}^{\mathbb{R}}_1$, and let $t_m \to t \geq 0$. To show $\mathcal{X}^{(m)}_{t_m} \Rightarrow \mathcal{X}_t$, by Theorem 1.8, it suffices to show

$$\lim_{m\to\infty} \mathbb{E}[\Phi(\mathcal{X}^{(m)}_{t_m})] = \mathbb{E}[\Phi(\mathcal{X}_t)] \qquad \forall\ \Phi = \Phi^{n,\phi,g} \in \Pi^{1,2}. \tag{3.6}$$

We claim that the convergence in

$$\lim_{s\to t} \mathbb{E}[\Phi(\mathcal{X}_s)] = \mathbb{E}[\Phi(\mathcal{X}_t)] \quad \text{is uniform w.r.t. the initial condition } \mathcal{X}_0. \tag{3.7}$$

In particular, as $m \to \infty$,

$$\mathbb{E}[\Phi(\mathcal{X}^{(m)}_{t_m})] - \mathbb{E}[\Phi(\mathcal{X}^{(m)}_{t})] \to 0.$$

To prove (3.6), it then suffices to show that

$$\lim_{m\to\infty} \mathbb{E}[\Phi(\mathcal{X}^{(m)}_{t})] = \mathbb{E}[\Phi(\mathcal{X}_t)]. \tag{3.8}$$

We prove (3.7) and (3.8) next.

Proof of (3.7). By the Markov property of $\mathcal{X}$, it suffices to show that as $t \downarrow 0$,

$$\big|\mathbb{E}[\Phi(\mathcal{X}_t)] - \mathbb{E}[\Phi(\mathcal{X}_0)]\big| \to 0 \quad \text{uniformly in } \mathcal{X}_0. \tag{3.9}$$

Note that we can write

$$\mathbb{E}[\Phi(\mathcal{X}_0)] = \int_{\mathbb{R}^n} g(\underline{x})\, \mathbb{E}[\phi(\{r_0(\xi(x_i), \xi(x_j))\}_{1\leq i<j\leq n})]\, \mathrm{d}\underline{x}, \tag{3.10}$$

where for each $x \in \mathbb{R}$, $\xi(x) \in X_0$ is sampled according to the conditional distribution of μ_0 on X_0, conditioned on the spatial coordinate in $X_0 \times \mathbb{R}$ being equal to x. On the other hand,

$$\mathbb{E}[\Phi(\mathcal{X}_t)] = \int_{\mathbb{R}^n} g(\underline{x}) \mathbb{E}[\,\phi(\{r_t(x_i, x_j)\}_{1\leq i<j\leq n})]\, \mathrm{d}\underline{x}. \tag{3.11}$$

Let $F(\underline{x}, t)$ denote the event that the dual Brownian web paths $\hat{f}_{(x_1,t)}, \ldots, \hat{f}_{(x_n,t)}$ do not coalesce during the time interval $[0, t]$. We can then partition $\mathbb{E}[\phi(\{r_t(x_i, x_j)\}_{1\leq i<j\leq n})]$ into expectation restricted to $F(\underline{x}, t)$ and $F^c(\underline{x}, t)$ respectively. On the event $F(\underline{x}, t)$, we can replace $\hat{f}_{(x_i,t)}$, $1 \leq i \leq n$, by independent Brownian motions $(x_i(s))_{s\leq t}$, $1 \leq i \leq n$, starting respectively at x_i at time t and *running backward in time*. Then

$$r_t(x_i, x_j) = 2t + r_0(\xi(x_i(0)), \xi(x_j(0))). \tag{3.12}$$

Let $\widetilde{F}(\underline{x}, t)$ denote the event that $(x_i(s))_{s\leq t}$, $1 \leq i \leq n$, do not intersect during the time interval $[0, t]$. Then using (3.12), we can rewrite (3.11) as

$$\begin{aligned}
&\mathbb{E}[\Phi(\mathcal{X}_t)] \\
&= \int_{\mathbb{R}^n} g(\underline{x}) \mathbb{E}[\,\phi(\underline{\underline{r_t}}) 1_{F^c(\underline{x},t)}]\, \mathrm{d}\underline{x} + \int_{\mathbb{R}^n} g(\underline{x}) \mathbb{E}[\,\phi(\{2t + r_0(\xi(x_i(0)), \xi(x_j(0)))\}_{1\leq i<j\leq n})]\, \mathrm{d}\underline{x} \\
&\quad - \int_{\mathbb{R}^n} g(\underline{x}) \mathbb{E}[\,\phi(\{2t + r_0(\xi(x_i(0)), \xi(x_j(0)))\}_{1\leq i<j\leq n}) 1_{\widetilde{F}^c(\underline{x},t)}]\, \mathrm{d}\underline{x}.
\end{aligned} \tag{3.13}$$

Note that for Lebesgue a.e. $x_1, \dots, x_n$, $\mathbb{P}(F^c(\underline{x}, t)) = \mathbb{P}(\widetilde{F}^c(\underline{x}, t)) \to 0$ as $t \downarrow 0$. Therefore by the bounded convergence theorem, the first and third term on the right hand side of (3.13) converges to 0 as $t \downarrow 0$, uniformly in $\mathcal{X}_0$. For the second term on the right hand side of (3.13), we can make the change of variable $y_i := x_i(0)$, $y_i(s) := x_i(s)$, to rewrite it as

$$\int_{\mathbb{R}^n} \mathbb{E}[g(y_1(t), \dots, y_n(t))\, \phi(\{2t + r_0(\xi(y_i), \xi(y_j))\}_{1 \le i < j \le n})]\, \mathrm{d}\underline{y}. \tag{3.14}$$

For each $\underline{y} \in \mathbb{R}^n$, clearly the quantity inside the expectation converges a.s. to the analogue in (3.10) as $t \downarrow 0$, and the speed of convergence does not depend on $\mathcal{X}_0$. Therefore the expectation in (3.14) also converges, uniformly in $\mathcal{X}_0$. Using the fact that g has bounded support, while g and ϕ are both bounded, we can easily dominate the integrand in (3.14) w.r.t. $\mathrm{d}\underline{y}$ by an integrable function as $t \downarrow 0$; (3.9), and hence (3.7) follows.

Proof of (3.8). For each $x_i \in \mathbb{R}$, $1 \le i \le n$, let us denote $\hat{f}_{(x_i,t)} \in \widehat{\mathcal{W}}(x_i, t)$ by $\hat{f}_i$. Then

$$\mathbb{E}[\Phi(\mathcal{X}_t^{(m)})] = \int_{\mathbb{R}^n} g(\underline{x}) \mathbb{E}[\phi(\underline{\underline{r_t}}^{(m)})] \mathrm{d}\underline{x}, \tag{3.15}$$

where $\underline{\underline{r_t}}^{(m)} = \{r_t^{(m)}(x_i, x_j)\}_{1 \le i < j \le n}$ depends on the realization of $(\hat{f}_i)_{1 \le i \le n}$. Let $\hat{\tau}$ be the smallest time in $[0, t]$ when a coalescence occurs among the paths $(\hat{f}_i)_{1 \le i \le n}$. Let $I_1, \dots, I_k$ denote the partition of $\{1, \dots, n\}$, where all $\hat{f}_i$ with i in the same partition element I_j have coalesced into a single path at time $\hat{\tau}$. Conditioned on $(\hat{f}_i)_{1 \le i \le n}$ on the time interval $[\hat{\tau}, t]$, the distribution of the remaining k coalescing Brownian motions on the time interval $[0, \hat{\tau}]$ is then given by the distribution of k Brownian motions conditioned not to intersect on the time interval $[0, \hat{\tau}]$, and their positions at time 0 has a probability density in $\mathbb{R}^k$, which we denote by $g_{\hat{f},\hat{\tau}}(v_1, \dots, v_k)$. Note that conditioned on $(\hat{f}_i)_{1 \le i \le n}$ on the time interval $[\hat{\tau}, t]$ and their positions $v_1, \dots, v_k$ at time 0, $\phi(\{r_t^{(m)}(x_i, x_j)\}_{1 \le i < j \le n})$ only depends on $\{r_0^{(m)}(\xi(v_i), \xi(v_j))\}_{1 \le i < j \le k}$, cf. (3.2). We can therefore write

$$\begin{aligned} \Phi_{\hat{f},\hat{\tau}}(\mathcal{X}_0^{(m)}) &:= \mathbb{E}[\phi(\underline{\underline{r_t}}^{(m)}) \,|\, (\hat{f}_i(s))_{1 \le i \le n, s \in [\hat{\tau}, t]}] \\ &= \int_{\mathbb{R}^k} g_{\hat{f},\hat{\tau}}(v_1, \dots, v_k) \mathbb{E}[\phi_{\hat{f},\hat{\tau}}(\{r_0^{(m)}(\xi(v_i), \xi(v_j))\}_{1 \le i < j \le k})]\, \mathrm{d}\underline{v}, \end{aligned} \tag{3.16}$$

where given the realization of $(\hat{f}_i(s))_{1 \le i \le n, s \in [\hat{\tau}, t]}$ and $\{\hat{f}_i(0)\}_{1 \le i \le n} = \{v_1, \dots, v_k\}$,

$$\phi_{\hat{f},\hat{\tau}}(\{r_0^{(m)}(\xi(v_i), \xi(v_j))\}_{1 \le i < j \le k}) = \phi(\{r_t^{(m)}(x_i, x_j)\}_{1 \le i < j \le n}). \tag{3.17}$$

Note that $\Phi_{\hat{f},\hat{\tau}}$ is a polynomial of order k on $\mathbb{U}^{\mathbb{R}}$, defined from the bounded continuous functions $g_{\hat{f},\hat{\tau}}$ and $\phi_{\hat{f},\hat{\tau}}$, except that $g_{\hat{f},\hat{\tau}}$ does not have bounded support. Nevertheless, $g_{\hat{f},\hat{\tau}}$ is integrable and can be approximated by continuous functions with bounded support. Therefore from the assumption $\mathcal{X}_0^{(m)} \to \mathcal{X}_0$ in $\mathbb{U}_1^{\mathbb{R}}$, we deduce that $\Phi_{\hat{f},\hat{\tau}}(\mathcal{X}_0^{(m)}) \to \Phi_{\hat{f},\hat{\tau}}(\mathcal{X}_0)$ for Lebesgue a.e. $x_1, \dots, x_n$ and a.e. realization of $(\hat{f}_i(s))_{1 \le i \le n, s \in [\hat{\tau}, t]}$. It then follows from the bounded convergence theorem that

$$\mathbb{E}[\Phi(\mathcal{X}_t^{(m)})] = \int_{\mathbb{R}^n} g(\underline{x}) \mathbb{E}[\Phi_{\hat{f},\hat{\tau}}(\mathcal{X}_0^{(m)})] \mathrm{d}\underline{x} \underset{m \to \infty}{\longrightarrow} \int_{\mathbb{R}^n} g(\underline{x}) \mathbb{E}[\Phi_{\hat{f},\hat{\tau}}(\mathcal{X}_0)] \mathrm{d}\underline{x} = \mathbb{E}[\Phi(\mathcal{X}_t)], \tag{3.18}$$

which concludes the proof of the Feller property.

(d): This follows readily from the construction of the CSSM genealogy process $\mathcal{X}_t^{\mathrm{CS}}$. For any $N > 0$, the genealogical distances among individuals with spatial locations in $[-N, N]$ are determined by coalescing Brownian motions in the dual Brownian web $\widehat{\mathcal{W}}$.

The initial condition $\mathcal{X}_0^{\mathrm{CS}}$ affects the genealogical distances only on the event that the backward coalescing Brownian motions starting from $[-N, N]$ at time t do not coalesce into a single path by time 0. As $t \to \infty$, the probability of this event tends to 0, and therefore $(\mathcal{X}_{t+s}^{\mathrm{CS}})_{s\geq 0}$ converges in distribution to the CSSM genealogy process constructed from $\widehat{\mathcal{W}}$ as in Section 1.4, with initial condition being at time $-\infty$. ∎

4 Proof of convergence of rescaled IFV genealogies

In this section, we prove Theorem 1.31, that under diffusive scaling of space and time as well as rescaling of measure, the genealogies of the interacting Fleming-Viot process converges to those of a CSSM. In Section 4.1 we prove f.d.d.-convergence, and in Section 4.2 tightness in path space. In Section 4.3, we prove Theorem 1.32 on the measure-valued process, which is needed to prove tightness in Section 4.2.

As in Theorem 1.31, let $(\mathcal{X}_t^{\mathrm{FV},\epsilon})_{t\geq 0} = \overline{(X_t^{\mathrm{FV},\epsilon}, r_t^{\mathrm{FV},\epsilon}, \mu_t^{\mathrm{FV},\epsilon})}_{t\geq 0}$ be the family of IFV genealogy processes on $\mathbb{Z}$ indexed by $\epsilon > 0$, such that $S_\epsilon \mathcal{X}_0^{\mathrm{FV},\epsilon} \to \mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_1^{\mathbb{R}}$ as $\epsilon \to 0$, and let $(\mathcal{X}_t^{\mathrm{CS}})_{t\geq 0} = \overline{(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}}, \mu_t^{\mathrm{CS}})}_{t\geq 0}$ be the CSSM genealogy process with initial condition $\mathcal{X}_0^{\mathrm{CS}}$.

4.1 Convergence of finite-dimensional distributions

In this subsection we prove the convergence $(S_\epsilon \mathcal{X}_{\epsilon^{-2}t}^{\mathrm{FV},\epsilon})_{t\geq 0} \Rightarrow (\mathcal{X}_t^{\mathrm{CS}})_{t\geq 0}$ in finite-dimensional distribution, i.e.,

$$(S_\epsilon \mathcal{X}_{\epsilon^{-2}t_1}^{\mathrm{FV},\epsilon}, \dots, S_\epsilon \mathcal{X}_{\epsilon^{-2}t_k}^{\mathrm{FV},\epsilon}) \underset{\epsilon\to 0}{\Longrightarrow} (\mathcal{X}_{t_1}^{\mathrm{CS}}, \dots, \mathcal{X}_{t_k}^{\mathrm{CS}}) \qquad \forall\, 0 \leq t_1 < t_2 < \cdots < t_k, \tag{4.1}$$

where $\Rightarrow$ denotes weak convergence of $(\mathbb{U}^{\mathbb{R}})^k$-valued random variables. By [EK86, Prop. 3.4.6] on convergence determining class for product spaces, it suffices to show that for any $\Phi_i := \Phi^{n_i,\phi_i,g_i} \in \Pi^{1,2}$, $1 \leq i \leq k$, we have (recall (1.8))

$$\mathbb{E}\Big[\prod_{i=1}^k \Phi_i(S_\epsilon \mathcal{X}_{\epsilon^{-2}t_i}^{\mathrm{FV},\epsilon})\Big] \underset{\epsilon\to 0}{\longrightarrow} \mathbb{E}\Big[\prod_{i=1}^k \Phi_i(\mathcal{X}_{t_i}^{\mathrm{CS}})\Big]. \tag{4.2}$$

For notational convenience we assume first that the initial tree is the trivial one (all distances are zero) and we shall see at the end of the argument that this easily generalizes. We will first rewrite both sides of this convergence relation in terms of the dual coalescents and then apply the invariance principle for coalescing random walks.

Step 1 (Claim rephrased in terms of coalescents). We can by the definition of the polynomial in (3.1) rewrite the left hand side of (4.2) as

$$\begin{aligned} &\mathbb{E}\Big[\prod_{i=1}^k \Phi_i(S_\epsilon \mathcal{X}_{\epsilon^{-2}t_i}^{\mathrm{FV},\epsilon})\Big] \\ &= \mathbb{E}\Big[\prod_{i=1}^k (\epsilon\sigma^{-1})^{n_i} \sum_{x_1^i,\dots,x_{n_i}^i \in \mathbb{Z}} g_i\big(\epsilon\sigma^{-1}(x_a^i)_{1\leq a\leq n_i}\big)\phi_i\big(\epsilon^2(r_{\epsilon^{-2}t_i}^{\mathrm{FV},\epsilon}(\xi_{t_i}^\epsilon(x_a^i), \xi_{t_i}^\epsilon(x_b^i)))_{1\leq a<b\leq n_i}\big)\Big], \end{aligned} \tag{4.3}$$

where for each time t_i, we sample n_i individuals in $X_{\epsilon^{-2}t_i}^{\mathrm{FV},\epsilon}$ at respective spatial positions $(x_a^i)_{1\leq a\leq n_i}$, with

$$\xi_{t_i}^\epsilon(x_a) \text{ being sampled from } X_{\epsilon^{-2}t_i}^{\mathrm{FV},\epsilon} \text{ according to } \mu_{\varepsilon^{-2}t_i}^{\mathrm{FV}}(\cdot|x_a), \tag{4.4}$$

the conditional distribution of $\mu_{\epsilon^{-2}t_i}^{\mathrm{FV},\epsilon}$ on $X_{\epsilon^{-2}t_i}^{\mathrm{FV},\epsilon}$ conditioned on the spatial mark being equal to $x_a \in \mathbb{Z}$.

By the space-time duality relation (1.32) for the IFV genealogy processes, every summand of the R.H.S. of (4.3) can be calculated in terms of coalescing random walks. Namely, the joint law of the space-time genealogies of the sampled individuals $\xi^\epsilon_{t_i}(x^i_a) \in X^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t_i}$, $1 \le a \le n_i$ and $1 \le i \le k$, is equal to that of a collection of coalescing random walks $(x^i_a(s))_{s \le \epsilon^{-2}t_i}$ (recall here time s runs *backwards*), starting respectively at x^i_a at time $\epsilon^{-2}t_i$, where each walk evolves *backward* in time as rate 1 continuous time random walk on $\mathbb{Z}$ with transition probability kernel $\bar{a}$, and two walks at the same location coalesce at rate γ. From the duality relation we get the following the stochastic representation:

$$r^{\mathrm{FV},\epsilon}_s(x,y) := \begin{cases} 2(s-\hat{\tau}) & \text{if } \hat{\tau} \ge 0, \\ 2s + r^{\mathrm{FV},\epsilon}_0(\xi^\epsilon_0(u), \xi^\epsilon_0(v)) & \text{if } \hat{\tau} < 0, \end{cases} \tag{4.5}$$

where $\hat{\tau}$ denotes the time of coalescence between the two coalescing walks starting at $x, y \in \mathbb{Z}$ at time s, while u, v are the positions of the two walks at time 0.

We observe next that the continuum population is represented by $A_t \cup E^+_t \cup E^-_t$ which is a version of $\mathbb{R}$ marked on E_t by $+, -$, the geographic marks are the reals and since the sampling measure is Lebesgue measure, we can write polynomials based on integration over $\mathbb{R}$ instead of $X \times V$ as:

$$\int_X \int_V \mu^{\otimes n}(d(\underline{x}, \underline{u}))g(\underline{u})\varphi(\underline{r}(\underline{x})) = \int_{\mathbb{R}} \lambda^{\otimes n}(d\underline{u})g(\underline{u})\varphi(\underline{r}(\underline{u})). \tag{4.6}$$

We can rewrite using the duality of Corollary 1.23, see also (1.41), the R.H.S. of (4.2) in the same form as in (4.3):

$$\mathbb{E}\Big[\prod_{i=1}^k \Phi_i(\mathcal{X}^{\mathrm{CS}}_{t_i})\Big] = \mathbb{E}\Big[\prod_{i=1}^k \int_{\mathbb{R}^{n_i}} g_i\big((y^i_a)_{1\le a\le n_i}\big)\phi_i\big((r^{\mathrm{CS}}_{t_i}(y^i_a, y^i_b))_{1\le a<b\le n_i}\big)\mathrm{d}\underline{y}^i\Big], \tag{4.7}$$

where at each time t_i, we sample n_i individuals from $X^{\mathrm{CS}}_{t_i}$ according to $\mu^{\mathrm{CS}}_{t_i}$ (which is Lebesgue measure on $\mathbb{R}$) at positions $(y^i_a)_{1\le a\le n_i}$, and their joint space-time genealogy lines are by construction distributed as a collection of coalescing Brownian motions $(y^i_a(s))_{s\le t_i}$, evolving *backward* in time.

To link (4.3) with (4.7), we note that in (4.3), we can regard

$$\prod_{i=1}^k (\epsilon\sigma^{-1})^{n_i} \sum_{x^i_1,\dots,x^i_{n_i}\in\mathbb{Z}} g_i\big(\epsilon\sigma^{-1}(x^i_a)_{1\le a\le n_i}\big)\delta_{x^i_1}\cdots\delta_{x^i_{n_i}} \tag{4.8}$$

as a finite signed sampling measure (recall that g has bounded support), which is easily seen to converge weakly to the finite signed sampling measure appearing in (4.7), namely

$$\prod_{i=1}^k g_i\big((y^i_a)_{1\le a\le n_i}\big)\mathrm{d}y^i_1\cdots\mathrm{d}y^i_{n_i}. \tag{4.9}$$

To prove (4.2), it then suffices to show that (having (4.8)-(4.9) in mind): If for each $1 \le a \le n_i$ and $1 \le i \le k$, $x^{i,\epsilon}_a \in \mathbb{Z}$ and $\epsilon\sigma^{-1}x^{i,\epsilon}_a \to y^i_a$ as $\epsilon \to 0$, then

$$\mathbb{E}\Big[\prod_{i=1}^k \phi_i\big((\epsilon^2 r^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t_i}(\xi^\epsilon_{t_i}(x^{i,\epsilon}_a), \xi^\epsilon_{t_i}(x^{i,\epsilon}_b)))_{1\le a<b\le n_i}\big)\Big] \underset{\epsilon\to 0}{\longrightarrow} \mathbb{E}\Big[\prod_{i=1}^k \phi_i\big((r^{\mathrm{CS}}_{t_i}(y^i_a, y^i_b))_{1\le a<b\le n_i}\big)\Big]. \tag{4.10}$$

Step 2 (Invariance principle for coalescents). We next prove (4.10) by means of an invariance principle for coalescing random walks.

This *invariance principle* reads as follows. Given a collection of backward coalescing random walks starting at ϵ-dependent positions $x_a^{i,\epsilon} \in \mathbb{Z}$ at time $\epsilon^{-2}t_i$, $1 \le a \le n_i$, $1 \le i \le k$, such that $\epsilon\sigma^{-1}x_a^{i,\epsilon} \to y_a^i$ as $\epsilon \to 0$, the collection of coalescing random walks $(x_a^{i,\epsilon}(s))_{s \le \epsilon^{-2}t_i}$, rescaled diffusively as $(\epsilon\sigma^{-1}x_a^{i,\epsilon}(\epsilon^{-2}s))_{s \le t_i}$, converges in distribution to the collection of coalescing Brownian motions $(y_a^i(s))_{s \le t_i}$ evolving backward in time. Furthermore, the times of coalescence between the coalescing random walks, scaled by ϵ^2, converge in distribution to the times of coalescence between the corresponding Brownian motions. The proof of such an invariance principle can be easily adapted from [NRS05, Section 5], which considered discrete time random walks with instantaneous coalescence. We will omit the details.

Let $\delta > 0$ be small. Note that the collection of rescaled coalescing random walks $(\epsilon\sigma^{-1}x_a^{i,\epsilon}(\epsilon^{-2}s))$ restricted to the time interval $s \in [\delta, t_k]$, together with their times of coalescence, converge in joint distribution to the collection of coalescing Brownian motions $(y_a^i(s))$ restricted to the time interval $s \in [\delta, t_k]$, together with their times of coalescence. Using Skorohod's representation theorem (see e.g. [Bil89]), we can couple $(\epsilon\sigma^{-1}x_a^{i,\epsilon}(\epsilon^{-2}s))_{s\in[\delta,t_k]}$ and $(y_a^i(s))_{s\in[\delta,t_k]}$ such that the paths and their times of coalescence converge almost surely. *Let us assume such a coupling* from now on.

By the same argument as in the proof of Theorem 1.27 (c), we can rewrite the expectations in (4.10) in terms of the backward coalescing random walks $x_a^{i,\epsilon} \in \mathbb{Z}$ and coalescing Brownian motions y_a^i. Furthermore, we can condition on the coalescing random walks $x_a^{i,\epsilon}(s)$ on the time interval $[\delta\epsilon^{-2}, \epsilon^{-2}t_k]$ and condition on the coalescing Brownian motions $y_a^i(s)$ on the time interval $[\delta, t_k]$, coupled as above.

Given the locations $u_1^\epsilon, \dots, u_l^\epsilon \in \mathbb{Z}$ of the remaining coalescing random walks at time $\delta\epsilon^{-2}$, we now make an approximation and replace them by *independent* random walks on the remaining time interval $[0, \delta\epsilon^{-2}]$, and make a similar replacement for the coalescing Brownian motions. Note that the error we introduce to the two sides of (4.10) is bounded by a constant (determined only by $|\phi_i|_\infty$, $1 \le i \le k$) times the probability that there is a coalescence among the random walks (resp. Brownian motions) in the time interval $[0, \delta\epsilon^{-2}]$ (resp. $[0, \delta]$), which tends to 0 as $\delta \downarrow 0$ uniformly in ϵ by the properties of Brownian motion and the invariance principle. Therefore to prove (4.10), it suffices to prove its analogue where we make such an approximation for a fixed $\delta > 0$, replacing coalescing random walks (resp. Brownian motions) on the time interval $[0, \delta\epsilon^{-2}]$ (resp. $[0, \delta]$) by independent ones. Let us fix such a $\delta > 0$ from now on.

By conditioning on the coalescing random walks and the coalescing Brownian motions on the macroscopic time interval $[\delta, t_k]$ and using the a.s. coupling between them, we note that the analogue of (4.10) discussed above follows readily if we show:

Lemma 4.1. *If $u_1^\epsilon, \dots, u_l^\epsilon \in \mathbb{Z}$ satisfy $\epsilon\sigma^{-1}u_i^\epsilon \to u_i$ as $\epsilon \to 0$, then for any bounded continuous function $\psi : \mathbb{R}^{\binom{l}{2}} \to \mathbb{R}$, we have*

$$\begin{aligned} &\sum_{x_1,\dots,x_l \in \mathbb{Z}} g_\delta^\epsilon(x_1, \dots, x_l)\mathbb{E}\big[\psi\big((\epsilon^2 r_0^{\mathrm{FV},\epsilon}(\xi_0^\epsilon(x_i), \xi_0^\epsilon(x_j)))_{1\le i<j\le l}\big)\big] \\ \xrightarrow[\epsilon\to 0]{} &\int_{\mathbb{R}^l} g_\delta(y_1, \dots, y_l)\mathbb{E}\big[\psi\big((r_0^{\mathrm{CS}}(\xi(y_i), \xi(y_j)))_{1\le i<j\le l}\big)\big]\mathrm{d}\underline{y}, \end{aligned} \tag{4.11}$$

where $g_\delta^\epsilon(\underline{x})$ is the probability mass function of l independent random walks at time $\delta\epsilon^{-2}$, starting at $u_1^\epsilon, \dots, u_l^\epsilon$; while $g_\delta(\underline{y})$ is the probability density function of l independent Brownian motions at time δ, starting at $u_1, \dots, u_l$.

Proof. If we can replace $g_\delta^\epsilon(x_1, \dots, x_l)$ in (4.11) by $(\epsilon\sigma^{-1})^l g_\delta(\epsilon\sigma^{-1}x_1, \dots, \epsilon\sigma^{-1}x_l)$, then (4.11) follows immediately by applying the polynomial Φ^{l,ψ,g_δ} to the states $S_\epsilon\mathcal{X}_0^{\mathrm{FV},\varepsilon}$ and $\mathcal{X}_0^{\mathrm{CS}}$, using the assumption $S_\epsilon\mathcal{X}_0^{\mathrm{FV},\varepsilon} \to \mathcal{X}_0^{\mathrm{CS}}$. The only problem is that g_δ does not have bounded support as we require for a polynomial. However, it is continuous

and integrable, and hence can be approximated by continuous functions with bounded support. Therefore the above reasoning is still valid.

To see why we can replace $g^\epsilon_\delta(\underline{x})$ by $(\epsilon\sigma^{-1})^l g_\delta(\epsilon\sigma^{-1}\underline{x})$, note that by the *local central limit theorem* (see e.g. [Spi76]),

$$(\epsilon\sigma^{-1})^{-l} g^\epsilon_\delta(\lceil \epsilon^{-1}\sigma \underline{y}\rceil) \xrightarrow[\epsilon\to 0]{} g_\delta(\underline{y}) \tag{4.12}$$

uniformly in $\underline{y} \in [-L, L]^l$ for any $L > 0$. Therefore when we restrict the summation in (4.11) to $\underline{x} \in [-\epsilon^{-1}L, \epsilon^{-1}L]^l$, the replacement induces an error that tends to 0 as $\epsilon \to 0$. By the central limit theorem, the contribution to the sum in (4.11) from $\underline{x} \notin [-\epsilon^{-1}L, \epsilon^{-1}L]^l$ can be made arbitrarily small (uniformly in ϵ) by choosing L large, and hence can be safely neglected if we first let $\epsilon \to 0$ and then let $L \to \infty$. ∎

4.2 Tightness

In this subsection we prove the tightness of the family of rescaled IFV genealogy processes, $(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon})_{\epsilon>0}$, regarded as $C([0,\infty), \mathbb{U}^{\mathbb{R}})$-valued random variables.

First we note that it is sufficient to prove the tightness of $(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon})_{\epsilon>0}$ as random variables taking values in the Skorohod space $D([0,\infty), \mathbb{U}^{\mathbb{R}})$. Indeed, the tightness of $(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon})_{\epsilon>0}$ in the Skorohod space, together with the convergence of $S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}$ to $\mathcal{X}^{\mathrm{CS}}$ in finite-dimensional distributions, imply that $S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon} \Rightarrow \mathcal{X}^{\mathrm{CS}}$ as $D([0,\infty), \mathbb{U}^{\mathbb{R}})$-valued random variables. In particular, $(\mathcal{X}^{\mathrm{CS}}_t)_{t\geq 0}$ admits a version which is a.s. càdlàg. Together with the fact that $\mathcal{X}^{\mathrm{CS}}_t$ is a.s. continuous in $t > 0$, which was established in the proof of Theorem 1.27 (b), it follows that $\mathcal{X}^{\mathrm{CS}}_t$ must be a.s. continuous in $t \geq 0$. *Note that this concludes the proof of Theorem 1.27 (b).*

Using Skorohod's representation theorem (see e.g. [Bil89]) to couple $(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon})_{\epsilon>0}$ and $\mathcal{X}^{\mathrm{CS}}$ such that the convergence in $D([0,\infty), \mathbb{U}^{\mathbb{R}})$ is almost sure, and using the a.s. continuity of $(\mathcal{X}^{\mathrm{CS}}_t)_{t\geq 0}$, we can then easily conclude that $S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon} \to \mathcal{X}^{\mathrm{CS}}$ a.s. in $C([0,\infty), \mathbb{U}^{\mathbb{R}})$, which implies the tightness of $(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon})_{\epsilon>0}$ as $C([0,\infty), \mathbb{U}^{\mathbb{R}})$-valued random variables.

By Jakubowski's criterion (see e.g. [Daw93, Theorem 3.6.4]), to show that $(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon})_{\epsilon>0}$ is a tight family of random variables in the Skorohod space $D([0,\infty), \mathbb{U}^{\mathbb{R}})$, it suffices to show that the following two conditions are satisfied:

(J1) (Compact Containment) For each $T > 0$ and $\delta > 0$, there exists a compact set $K_{T,\delta} \subset \mathbb{U}^{\mathbb{R}}$ such that for all $\epsilon > 0$,

$$\mathbb{P}\big(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t} \in K_{T,\delta}\, \forall\, 0 \leq t \leq T\big) \geq 1 - \delta; \tag{4.13}$$

(J2) (Tightness of Evaluations) For each $f \in \widetilde{\Pi}^{1,2}$, $(f(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}))_{t\geq 0}$, indexed by $\epsilon > 0$, is a tight family of $D([0,\infty), \mathbb{R})$-valued random variables.

Note that $\widetilde{\Pi}^{1,2}$ (recall from (1.9)) separates points in $\mathbb{U}^{\mathbb{R}}$ by Theorem 1.8, and is closed under addition.

We first prove **(J2)**, following an approach used in [AS11], where a family of rescaled measure-valued processes induced by the voter model on $\mathbb{Z}$ is shown to be tight and converge weakly to the measure-valued CSSM as $\epsilon \to 0$. We will verify **(J2)** via Aldous' tightness criterion (see e.g. [Daw93, Theorem 3.6.5]), reducing **(J2)** to the following conditions:

(A1) For each rational $t \geq 0$, $\{f(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t})\}_{\epsilon>0}$ is a tight family of $\mathbb{R}$-valued random variables.

(A2) Given $T > 0$ and any stopping time $\tau_\epsilon \le T$, if $\delta_\epsilon \downarrow 0$ as $\epsilon \downarrow 0$, then for each $\eta > 0$,

$$\lim_{\epsilon \downarrow 0} \mathbb{P}\big(\big|f(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}(\tau_\epsilon+\delta_\epsilon)}) - f(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\tau_\epsilon})\big| > \eta\big) = 0. \tag{4.14}$$

Proof of (J2) via (A1)–(A2). Note that each $f \in \widetilde{\Pi}^{1,2}$ can be written as $f = \sum_{i=1}^k c_i \Phi_i$ for finitely many $\Phi_i \in \Pi^{1,2}$ and $c_i \in \mathbb{R}$. It then follows by the triangle inequality that to verify **(A1)–(A2)**, it suffices to consider $f = \Phi \in \Pi^{1,2}$. We abbreviate for $\Phi \in \Pi^{1,2}$, $\Phi^\epsilon_t := \Phi(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t})$.

(A1). Note that for $\Phi = \Phi^{n,\phi,g} \in \Pi^{1,2}$, $(\Phi^\epsilon_t)_{\epsilon>0}$ is uniformly bounded, because ϕ is bounded, g is bounded with bounded support, and the projection of $S_\epsilon \mu^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}$ on the mark space $\mathbb{R}$ converges to the Lebesgue measure as $\epsilon \downarrow 0$, giving **(A1)**.

(A2). We will use the duality between interacting Fleming-Viot processes and coalescing random walks (see Theorem 1.17 and Corollary 1.20). Recall that $\Phi^\epsilon_t = \Phi(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t})$. First, we bound

$$\begin{aligned}
&\mathbb{P}(|\Phi^\epsilon_{\tau_\epsilon+\delta_\epsilon} - \Phi^\epsilon_{\tau_\epsilon}| > \eta) \\
&\le \frac{1}{\eta^2}\mathbb{E}[(\Phi^\epsilon_{\tau_\epsilon+\delta_\epsilon} - \Phi^\epsilon_{\tau_\epsilon})^2] \\
&= \frac{1}{\eta^2}\mathbb{E}\big[\mathbb{E}[(\Phi^\epsilon_{\tau_\epsilon+\delta_\epsilon} - \Phi^\epsilon_{\tau_\epsilon})^2 \,|\, \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\tau_\epsilon}]\big] \\
&\le \frac{2}{\eta^2}\mathbb{E}\big[\mathrm{Var}\big(\Phi^\epsilon_{\tau_\epsilon+\delta_\epsilon} \,|\, \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\tau_\epsilon}\big)\big] + \frac{2}{\eta^2}\mathbb{E}\big[\big(\mathbb{E}[\Phi^\epsilon_{\tau_\epsilon+\delta_\epsilon} \,|\, \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\tau_\epsilon}] - \Phi^\epsilon_{\tau_\epsilon}\big)^2\big],
\end{aligned} \tag{4.15}$$

where in the second inequality, we added and subtracted $\mathbb{E}[\Phi^\epsilon_{\tau_\epsilon+\delta_\epsilon} \,|\, \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\tau_\epsilon}]$ from $\Phi^\epsilon_{\tau_\epsilon+\delta_\epsilon} - \Phi^\epsilon_{\tau_\epsilon}$ and used $(a+b)^2 \le 2a^2 + 2b^2$. We treat the two terms on the r.h.s. separately.

First term in (4.15). We bound this term by bounding $\mathrm{Var}\big(\Phi^\epsilon_{\tau_\epsilon+\delta_\epsilon} \,|\, \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\tau_\epsilon}\big)$ uniformly in $\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\tau_\epsilon}$. First note that $\mathcal{X}^{\mathrm{FV},\epsilon}$ is a strong Markov process by Theorem 1.12. Therefore $\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}(\tau_\epsilon+\delta_\epsilon)}$ can be seen as the IFV genealogy process $\mathcal{X}^{\mathrm{FV},\epsilon}$ at time $\epsilon^{-2}\delta_\epsilon$ with initial condition $\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\tau_\epsilon}$. In particular, it suffices to bound $\mathrm{Var}(\Phi^\epsilon_{\delta_\epsilon})$ uniformly in the initial condition $\mathcal{X}^{\mathrm{FV},\epsilon}_0$, which we can assume to be deterministic.

Let $\Phi = \Phi^{n,\phi,g}$, and denote $\underline{x} := (x_1, \dots, x_n) \in \mathbb{Z}^n$, $\underline{y} := (y_1, \dots, y_n) \in \mathbb{Z}^n$. Then by the definition of the scaling map S_ϵ in (1.49), we have

$$\begin{aligned}
&\mathrm{Var}(\Phi^\epsilon_{\delta_\epsilon}) \\
&= \mathrm{Var}\big(\Phi^{n,\phi,g}(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\delta_\epsilon})\big) = \mathbb{E}[(\Phi^\epsilon_{\delta_\epsilon})^2] - \mathbb{E}[\Phi^\epsilon_{\delta_\epsilon}]^2 \\
&= (\epsilon\sigma^{-1})^{2n} \sum_{\underline{x},\underline{y} \in \mathbb{Z}^n} g(\epsilon\sigma^{-1}\underline{x}) g(\epsilon\sigma^{-1}\underline{y})\, \mathrm{Cov}\big(\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\delta_\epsilon}(\underline{x})),\ \phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\delta_\epsilon}(\underline{y}))\big),
\end{aligned} \tag{4.16}$$

where $r^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\delta_\epsilon}(\underline{x})$ denotes the distance matrix $r^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\delta_\epsilon}(\xi(x_i), \xi(x_j))_{1 \le i < j \le n}$ of n individuals $\xi(x_1), \dots, \xi(x_n)$ sampled independently from $\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\delta_\epsilon}$ at positions $x_1, \dots, x_n$ respectively. In order to evaluate the r.h.s. of (4.16) we represent the quantity using the duality in terms of a collection of coalescing random walks as follows.

Let $(X^{x_i}_t)_{1 \le i \le n}$ and $(X^{y_i}_t)_{1 \le i \le n}$ denote a family of rate 1 continuous time random walks on $\mathbb{Z}$ with transition kernel $\bar{a}$ as in (1.15), and every pair of walks at the same site coalesce at rate γ. The coalescence gives a partition of the set of coalescing random walks at time $\epsilon^{-2}\delta_\epsilon$, and independently for each partition element, say at position $z \in \mathbb{Z}$, we sample an individual from $\mathcal{X}^{\mathrm{FV},\epsilon}_0$ at position z. Let $r^{\mathrm{FV},\epsilon}_0(X^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon})$ denote the distance matrix of the collection of sampled individuals associated with the walks $X^{x_1}_{\epsilon^{-2}\delta_\epsilon}, \dots, X^{x_n}_{\epsilon^{-2}\delta_\epsilon}$ at time $\epsilon^{-2}\delta_\epsilon$, and let $r^{\mathrm{FV},\epsilon}_0(X^{\underline{y}}_{\epsilon^{-2}\delta_\epsilon})$ be defined similarly. We can further

construct $(\widetilde{X}_t^{y_i})_{1\le i\le n}$, a copy of $(X_t^{y_i})_{1\le i\le n}$, which coincides with $(X_t^{y_i})_{1\le i\le n}$ up to time $\epsilon^{-2}\delta_\epsilon$ on the event

$$G_{\epsilon^{-2}\delta_\epsilon}(\underline{x},\underline{y}) := \{\text{none of } (X^{x_i})_{1\le i\le n} \text{ coalesces with any } (X^{y_i})_{1\le i\le n} \text{ before time } \epsilon^{-2}\delta_\epsilon\},$$

such that $(\widetilde{X}_t^{y_i})_{1\le i\le n}$ is independent of $(X_t^{x_i})_{1\le i\le n}$. Let $r_0^{\mathrm{FV},\epsilon}(\widetilde{X}^{\underline{y}}_{\epsilon^{-2}\delta_\epsilon})$ be the associated distance matrix, which is independent of $r_0^{\mathrm{FV},\epsilon}(X^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon})$. By the duality relation (see Theorem 1.17), we have

$$\begin{aligned}
&\mathrm{Cov}\big(\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\delta_\epsilon}(\underline{x})),\ \phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\delta_\epsilon}(\underline{y}))\big)\\
&= \mathrm{Cov}\big(\phi(\epsilon^2 r_0^{\mathrm{FV},\epsilon}(X^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon})),\ \phi(\epsilon^2 r_0^{\mathrm{FV},\epsilon}(X^{\underline{y}}_{\epsilon^{-2}\delta_\epsilon}))\big)\\
&= \mathbb{E}\big[\phi(\epsilon^2 r_0^{\mathrm{FV},\epsilon}(X^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon}))\phi(\epsilon^2 r_0^{\mathrm{FV},\epsilon}(X^{\underline{y}}_{\epsilon^{-2}\delta_\epsilon}))\big] - \mathbb{E}\big[\phi(\epsilon^2 r_0^{\mathrm{FV},\epsilon}(X^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon}))\big]\mathbb{E}\big[\phi(\epsilon^2 r_0^{\mathrm{FV},\epsilon}(X^{\underline{y}}_{\epsilon^{-2}\delta_\epsilon}))\big]\\
&= \mathbb{E}\big[\phi(\epsilon^2 r_0^{\mathrm{FV},\epsilon}(X^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon}))\big\{\phi(\epsilon^2 r_0^{\mathrm{FV},\epsilon}(X^{\underline{y}}_{\epsilon^{-2}\delta_\epsilon})) - \phi(\epsilon^2 r_0^{\mathrm{FV},\epsilon}(\widetilde{X}^{\underline{y}}_{\epsilon^{-2}\delta_\epsilon}))\big\}\big]\\
&\le 2|\phi|_\infty^2 \mathbb{P}(G_{\epsilon^{-2}\delta_\epsilon}(\underline{x},\underline{y})^c)\\
&\le 2|\phi|_\infty^2 \sum_{1\le i,j\le n} \mathbb{P}(\tau_{x_i,y_j}\le \epsilon^{-2}\delta_\epsilon),
\end{aligned} \tag{4.17}$$

where τ_{x_i,y_i} denotes the time it takes for the two walks $X^{x_i}_\cdot$ and $X^{y_j}_\cdot$ to meet. Note that this bound is uniform w.r.t. $\mathcal{X}_0^{\mathrm{FV},\epsilon}$. Substituting it into (4.16) then gives

$$\begin{aligned}
&\mathrm{Var}(\Phi^\epsilon_{\delta_\epsilon})\\
&\le |\phi|_\infty^2 \sum_{1\le i,j\le n} (\epsilon\sigma^{-1})^{2n} \sum_{\underline{x},\underline{y}\in\mathbb{Z}^n} g(\epsilon\sigma^{-1}\underline{x})g(\epsilon\sigma^{-1}\underline{y})\mathbb{P}(\tau_{x_i,y_j}\le\epsilon^{-2}\delta_\epsilon)\\
&= |\phi|_\infty^2 \sum_{1\le i,j\le n} \sum_{\tilde{\underline{x}},\tilde{\underline{y}}\in\epsilon\sigma^{-1}\mathbb{Z}^n} (\epsilon\sigma^{-1})^{2n} g(\tilde{\underline{x}})g(\tilde{\underline{y}})\mathbb{P}(\tau_{\epsilon^{-1}\sigma\tilde{x}_i,\epsilon^{-1}\sigma y_j}\le\epsilon^{-2}\delta_\epsilon),
\end{aligned} \tag{4.18}$$

where $\tilde{\underline{x}} := \epsilon\sigma^{-1}\underline{x}$ and $\tilde{\underline{y}} := \epsilon\sigma^{-1}\underline{y}$.

We claim that the r.h.s. of (4.18) tends to 0 as $\epsilon\downarrow 0$. Indeed, the measure

$$(\epsilon\sigma^{-1})^{2n} \sum_{1\le i,j\le n} \sum_{\tilde{\underline{x}},\tilde{\underline{y}}\in\epsilon\sigma^{-1}\mathbb{Z}^n} \delta_{\tilde{x}_i}\delta_{\tilde{y}_j} g(\tilde{\underline{x}})g(\tilde{\underline{y}})$$

converges weakly to the finite measure $g(\tilde{\underline{x}})g(\tilde{\underline{y}})\mathrm{d}\tilde{\underline{x}}\mathrm{d}\tilde{\underline{y}}$ on $\mathbb{R}^{2n}$ as $\epsilon\downarrow 0$. By Donsker's invariance principle and the fact that $\delta_\epsilon\to 0$ as $\epsilon\downarrow 0$, we note that for any $\lambda>0$,

$$\mathbb{P}(\tau_{\epsilon^{-1}\sigma\tilde{x}_i,\epsilon^{-1}\sigma y_j}\le\epsilon^{-2}\delta_\epsilon) \underset{\epsilon\to 0}{\longrightarrow} 0$$

uniformly in $\tilde{x}_i$ and $\tilde{y}_j$ with $|\tilde{x}_i-\tilde{y}_j|>\lambda$. It follows that when restricted to $\tilde{x}_i$ and $\tilde{y}_j$ with $|\tilde{x}_i-\tilde{y}_j|>\lambda$, the inner sum in (4.18) tends to 0 as $\epsilon\downarrow 0$. On the other hand, when restricted to $\tilde{x}_i$ and $\tilde{y}_j$ with $|\tilde{x}_i-\tilde{y}_j|\le\lambda$, the inner sum in (4.18) can be bounded from above by replacing $\mathbb{P}(\cdot)$ with 1, which then converges to the integral of the finite measure $g(\tilde{\underline{x}})g(\tilde{\underline{y}})\mathrm{d}\tilde{\underline{x}}\mathrm{d}\tilde{\underline{y}}$ over the subset of $\mathbb{R}^{2n}$ with $|\tilde{x}_i-\tilde{y}_j|\le\lambda$, and can be made arbitrarily small by choosing $\lambda>0$ small.

This proves that $\mathrm{Var}(\Phi^\epsilon_{\delta_\epsilon})$ tends to 0 uniformly in $\mathcal{X}_0^{\mathrm{FV},\epsilon}$ as $\epsilon\downarrow 0$, and hence the first term in (4.15) tends to 0 as $\epsilon\downarrow 0$.

Second term in (4.15). By the strong Markov property of $\mathcal{X}^{\mathrm{FV},\epsilon}$, it suffices to bound $\big|\mathbb{E}[\Phi^\epsilon_{\delta_\epsilon}]-\Phi^\epsilon_0\big|$ uniformly in the (deterministic) initial condition $\mathcal{X}_0^{\mathrm{FV},\epsilon}$.

Let $(X_t^{x_i})_{1\le i\le n}$, $r_0^{\mathrm{FV},\epsilon}(X^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon})$, and $r_t^{\mathrm{FV},\epsilon}(\underline{x})$ be defined as before (4.17). Let $(\widetilde{X}_t^{x_i})_{1\le i\le n}$ be a collection of independent random walks, such that $(\widetilde{X}_t^{x_i})_{1\le i\le n}$ coincides with $(X_t^{x_i})_{1\le i\le n}$ up to time $\epsilon^{-2}\delta_\epsilon$ on the event

$$G_{\epsilon^{-2}\delta_\epsilon}(\underline{x}) := \{\text{no coalescence has taken place among } (X^{x_i})_{1\le i\le n} \text{ before time } \epsilon^{-2}\delta_\epsilon\}.$$

Then we have

$$
\begin{aligned}
&\big|\mathbb{E}[\Phi^\epsilon_{\delta_\epsilon}] - \Phi^\epsilon_0\big| \\
= \;& (\epsilon\sigma^{-1})^n \Big| \sum_{\underline{x}\in\mathbb{Z}^n} g(\epsilon\sigma^{-1}\underline{x})\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\delta_\epsilon}(\underline{x}))\big] - \sum_{\underline{y}\in\mathbb{Z}^n} g(\epsilon\sigma^{-1}\underline{y})\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(\underline{y}))\big]\Big| \\
= \;& (\epsilon\sigma^{-1})^n \Big| \sum_{\underline{x}\in\mathbb{Z}^n} g(\epsilon\sigma^{-1}\underline{x})\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(X^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon}))\big] - \sum_{\underline{y}\in\mathbb{Z}^n} g(\epsilon\sigma^{-1}\underline{y})\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(\underline{y}))\big]\Big| \\
\le \;& (\epsilon\sigma^{-1})^n \Big| \sum_{\underline{x}\in\mathbb{Z}^n} g(\epsilon\sigma^{-1}\underline{x})\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(\widetilde{X}^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon}))\big] - \sum_{\underline{y}\in\mathbb{Z}^n} g(\epsilon\sigma^{-1}\underline{y})\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(\underline{y}))\big]\Big| \\
&+ (\epsilon\sigma^{-1})^n \Big| \sum_{\underline{x}\in\mathbb{Z}^n} g(\epsilon\sigma^{-1}\underline{x})\mathbb{E}\Big[\Big\{\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(X^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon})) - \phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(\widetilde{X}^{\underline{x}}_{\epsilon^{-2}\delta_\epsilon}))\Big\} 1_{\{G^c_{\epsilon^{-2}\delta_\epsilon}(\underline{x})\}}\Big]\Big|.
\end{aligned}
$$

Note that the second term in the bound above tends to 0 as $\epsilon \downarrow 0$ by the same argument as the one showing that the bound for $\mathrm{Var}(\Phi^\epsilon_{\delta_\epsilon})$ in (4.18) tends to 0 as $\epsilon \downarrow 0$. To bound the first term in the bound above, we decompose according to the positions of the random walks and rewrite it as follows, where $p_t(x)$ denotes the transition probability kernel of $\widetilde{X}^0_t$:

$$
\begin{aligned}
&\big(\tfrac{\epsilon}{\sigma}\big)^n \Big| \sum_{\underline{x}\in\mathbb{Z}^n} g\big(\tfrac{\epsilon}{\sigma}\underline{x}\big)\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(\widetilde{X}^{\underline{x}}_{\delta_\epsilon\epsilon^{-2}}))\big] - \sum_{\underline{y}\in\mathbb{Z}^n} g\big(\tfrac{\epsilon}{\sigma}\underline{y}\big)\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(\underline{y}))\big]\Big| \\
&= \big(\tfrac{\epsilon}{\sigma}\big)^n \Big| \sum_{\underline{y}\in\mathbb{Z}^n}\sum_{\underline{x}\in\mathbb{Z}^n}\prod_{i=1}^n p_{\frac{\delta_\epsilon}{\epsilon^2}}(y_i - x_i) g\big(\tfrac{\epsilon}{\sigma}\underline{x}\big)\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(\underline{y})\big] - \sum_{\underline{y}\in\mathbb{Z}^n} g\big(\tfrac{\epsilon}{\sigma}\underline{y}\big)\mathbb{E}\big[\phi(\epsilon^2 r^{\mathrm{FV},\epsilon}_0(\underline{y}))\big]\Big| \\
&\le |\phi|_\infty \sum_{\underline{y}\in\mathbb{Z}^n} \big(\tfrac{\epsilon}{\sigma}\big)^n \Big| \sum_{\underline{x}\in\mathbb{Z}^n}\prod_{i=1}^n p_{\frac{\delta_\epsilon}{\epsilon^2}}(y_i - x_i)\big(g\big(\tfrac{\epsilon}{\sigma}\underline{x}\big) - g\big(\tfrac{\epsilon}{\sigma}\underline{y}\big)\big)\Big| \\
&= |\phi|_\infty \sum_{\underline{y}\in\mathbb{Z}^n} \big(\tfrac{\epsilon}{\sigma}\big)^n \Big| \sum_{\underline{x}\in\mathbb{Z}^n}\prod_{i=1}^n p_{\frac{\delta_\epsilon}{\epsilon^2}}(y_i - x_i)\Big\{\tfrac{\epsilon}{\sigma}\langle \underline{x}-\underline{y}, \nabla g\big(\tfrac{\epsilon}{\sigma}\underline{y}\big)\rangle \\
&\qquad\qquad + \tfrac{\epsilon^2}{2\sigma^2}\langle \underline{x}-\underline{y}, \nabla^2 g\big(\tfrac{\epsilon}{\sigma}\underline{u}(\underline{x},\underline{y})\big)(\underline{x}-\underline{y})\rangle\Big\}\Big| \\
&\le C|\phi|_\infty|\nabla^2 g|_\infty\big(\tfrac{\epsilon}{\sigma}\big)^{n+2} \sum_{\underline{x},\underline{y}\in\mathbb{Z}^n}\prod_{i=1}^n p_{\frac{\delta_\epsilon}{\epsilon^2}}(y_i - x_i)\big(1_{\{\frac{\epsilon}{\sigma}\underline{x}\in\mathrm{supp}(g)\}} + 1_{\{\frac{\epsilon}{\sigma}\underline{y}\in\mathrm{supp}(g)\}}\big)\sum_{i=1}^n (y_i - x_i)^2 \\
&\le 2nC|\phi|_\infty|\nabla^2 g|_\infty\big(\tfrac{\epsilon}{\sigma}\big)^n \sum_{\underline{y}\in\mathbb{Z}^n} 1_{\{\epsilon\sigma^{-1}\underline{y}\in\mathrm{supp}(g)\}}\ \delta_\epsilon,
\end{aligned}
\tag{4.19}
$$

where C is a constant depending only on n. In the derivation above, we Taylor expanded $g(\epsilon\sigma^{-1}\underline{x})$ around $\epsilon\sigma^{-1}\underline{y}$ when either $\epsilon\sigma^{-1}\underline{x}$ or $\epsilon\sigma^{-1}\underline{y}$ is not in the support of g, ∇g and $\nabla^2 g$ denote the first and second derivatives of g, and $\underline{u}(\underline{x},\underline{y})$ is some point on the line segment connecting $\underline{x}$ and $\underline{y}$. Lastly, we used the fact that $\sum_{z\in\mathbb{Z}} z p_t(z) = 0$ and $\sum_{z\in\mathbb{Z}} z^2 p_t(z) = t\sigma^2$. Since g has bounded support, the bound we obtained above is bounded by $C'\delta_\epsilon$ for some C' depending only on n, ϕ and g, and hence tends to 0 as $\epsilon \downarrow 0$.

This verifies that the second term in (4.15) also tends to 0 as $\epsilon \downarrow 0$, which concludes the proof of **(A2)**. ∎

We have verified **(J2)** above and hence to conclude the proof of tightness of $(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon})_{\epsilon>0}$ as a family of random variables in the Skorohod space $D([0,\infty), \mathbb{U}^{\mathbb{R}})$, it only remains to verify the compact containment condition **(J1)**. Some technical difficulties arise. Because the geographical space is unbounded, truncation in space is needed. We also need to control how the sizes of different families fluctuate over time, as well as how

the population flux across the truncation boundaries affect the family sizes. Our strategy is to enlarge the mark space by assigning types to different families. Using a weaker convergence result, Theorem 1.32, for measure-valued IFV processes with types (but no genealogies), we can control the evolution of family sizes as well as their dispersion in space, which can then be strengthened to control the genealogical structure of the population. As we will point out later, Theorem 1.32 can be proved by adapting what we have done so far, because condition **(J1)** is trivial in that context. Therefore invoking Theorem 1.32 to prove **(J1)** for the genealogy processes is justified.

Proof of (J1). As noted in Remark 1.4, we can regard $\mathbb{U}^{\mathbb{R}}$ as a subset of $(\mathbb{U}_f^{\mathbb{R}})^{\mathbb{N}}$, endowed with the product $\mathbb{R}$-marked Gromov-weak topology. Since the product of compact sets gives a compact set in the product space $(\mathbb{U}_f^{\mathbb{R}})^{\mathbb{N}}$, to prove **(J1)**, it suffices to show that for each $k \in \mathbb{N}$, the restriction of $(S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t})_{t\geq 0,\epsilon>0}$ to the subset of marks $(-k,k) \subset \mathbb{R}$, i.e.,

$$S^{(k)}_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t} := \overline{(X^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}, S_\epsilon r^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}, 1_{\{|v|<k\}} S_\epsilon \mu^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}(\mathrm{d}x\mathrm{d}v))}, \qquad t \geq 0, \epsilon > 0,$$

satisfies the compact containment condition (4.13). More precisely, it suffices to show that for each $T > 0$ (which we will assume to be 1 for simplicity), and for each $\delta > 0$, there exists a compact $K_\delta \subset \mathbb{U}_f^{\mathbb{R}}$, such that for all $\epsilon > 0$ sufficiently small,

$$\mathbb{P}\big(S^{(k)}_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t} \in K_\delta \ \forall\, 0 \leq t \leq 1\big) \geq 1 - \delta, \tag{4.20}$$

where $k \in \mathbb{N}$ will be fixed for the rest of the proof.

We will construct $K^1_\delta, K^2_\delta, K^3_\delta \subset \mathbb{U}_f^{\mathbb{R}}$, which satisfy respectively conditions (i)–(iii) in Theorem A.1 for the relative compactness of subsets of $\mathbb{U}_f^{\mathbb{R}}$. We can then take $K_\delta := \overline{K^1_\delta \cap K^2_\delta \cap K^3_\delta}$, which is a compact subset of $\mathbb{U}_f^{\mathbb{R}}$. To prove (4.20), it then suffices to prove the same inequality but with K_δ replaced by K^i_δ for each $1 \leq i \leq 3$, which we do in **(1)–(3)** below.

Later when we construct K^2_δ and K^3_δ, we will keep track of the mass of different individuals having some specified properties. The way to do this is to introduce additional marks. We will enlarge the mark space from $\mathbb{R}$ to $\mathbb{R} \times [0,1]$, where $[0,1]$ is the space of the additional types, and $\mathcal{X}^{\mathrm{FV},\epsilon}_t$ and $\mathcal{X}^{\mathrm{CS}}_t$ then become random variables taking values in the space $\mathbb{U}^{\mathbb{R}\times[0,1]}$. For each $\overline{(X,r,\mu)} \in \mathbb{U}^{\mathbb{R}\times[0,1]}$, $\mu(\mathrm{d}x\mathrm{d}v\mathrm{d}\tau)$ is then a measure on $X \times \mathbb{R} \times [0,1]$. The types of individuals in $\mathcal{X}^{\mathrm{FV},\epsilon}_0$ and $\mathcal{X}^{\mathrm{CS}}_0$ will be assigned later as we see fit.

(1) First let K^1_δ be the subset of $\mathbb{U}_f^{\mathbb{R}}$, such that for each $\overline{(X,r,\mu)} \in K^1_\delta$, $\mu(X \times \cdot)$ is supported on $[-k,k]$, with total mass bounded by $4k$. Since the family of measures on $[-k,k]$, with total mass bounded by $4k$, is relatively compact w.r.t. the weak topology, K^1_δ satisfies condition (i) in Theorem A.1. We further note that a.s., $S^{(k)}_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t} \in K^1_\delta$ for all $t \geq 0$, and hence (4.20) holds with K_δ replaced by K^1_δ.

(2) For each $n \in \mathbb{N}$, we will find below $L(n)$ such that if $K^{2,n}_\delta$ denotes the subset of $\mathbb{U}_f^{\mathbb{R}}$ with

$$\iint_{(X\times\mathbb{R})^2} 1_{\{r(x,y)>L(n)\}} \mu(\mathrm{d}x\mathrm{d}u)\mu(\mathrm{d}y\mathrm{d}v) < \frac{1}{n} \qquad \text{for each } \overline{(X,r,\mu)} \in K^{2,n}_\delta, \tag{4.21}$$

then uniformly in $\epsilon > 0$ sufficiently small, we will have

$$\mathbb{P}\big(S^{(k)}_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t} \notin K^{2,n}_\delta \text{ for some } 0 \leq t \leq 1\big) \leq \frac{\delta}{2^n} =: \delta_n. \tag{4.22}$$

We can then take $K^2_\delta := \cap_{n\in\mathbb{N}} K^{2,n}_\delta$, which clearly satisfies condition (ii) in Theorem A.1, while (4.22) implies that (4.20) holds with K_δ replaced by K^2_δ.

In order to find $L(n)$, we proceed in two steps. First we find an analogue of $L(n)$ for the limiting CSSM genealogies, and then in a second step, we use the convergence of the measure-valued IFV to obtain $L(n)$.

Fix $n \in \mathbb{N}$. To find $L(n)$ such that (4.22) holds, we first prove an analogue of (4.22) for the continuum limit $\mathcal{X}^{\mathrm{CS}}$ by utilizing the types of the individuals. Given $\eta \geq 0$ and $\gamma \in [0,1]$, let

$$G_{\eta,\gamma} := \big\{\overline{(X,r,\mu)} \in \mathbb{U}_f^{\mathbb{R}\times[0,1]} : \mu(X \times [-k,k] \times [0,\gamma]) \leq \eta\big\}. \tag{4.23}$$

We claim that we can find A sufficiently large, such that if all individuals in $\mathcal{X}_0^{\mathrm{CS}}$ with spatial mark outside $[-A,A]$ are assigned type 0, and all other individuals are assigned type 1, then

$$\mathbb{P}\big(\mathcal{X}_t^{\mathrm{CS}} \in G_{0,0}\ \forall\, 0 \leq t \leq 1\big) \geq 1 - \frac{\delta_n}{4}. \tag{4.24}$$

In other words, with probability at least $1 - \delta_n/4$, the following event occurs: for all $0 \leq t \leq 1$, no individual in $\mathcal{X}_t^{\mathrm{CS}}$ with spatial mark in $[-k,k]$ can trace its genealogy back to some individual at time 0 with spatial mark outside $[-A,A]$. This will allow us to restrict our attention to descendants of the population in $[-A,A]$ at time 0.

Indeed, by the construction of the CSSM (Section 1.4) using the Brownian web, the measure-valued process $\widehat{\mathcal{X}}_t^{\mathrm{CS}}$, which is the measure μ_t^{CS} projected on the geographic and type space, is given by

$$\widehat{\mathcal{X}}_t^{\mathrm{CS}}(\mathrm{d}v\mathrm{d}\tau) = \delta_0(\mathrm{d}\tau)\big(1_{\{v>f_{(A,0)}(t)\}} + 1_{\{v<f_{(-A,0)}(t)\}}\big)\mathrm{d}v + \delta_1(\mathrm{d}\tau)1_{\{f_{(-A,0)}(t)\leq v\leq f_{(A,0)}(t)\}}\mathrm{d}v, \tag{4.25}$$

where $f_{(\pm A,0)}(\cdot)$ are the two coalescing Brownian motions in the Brownian web $\mathcal{W}$, starting respectively at $(\pm A, 0)$. The event in (4.24) occurs if $f_{(\pm A,0)}$ do not enter $[-k,k]$ before time 1, the probability of which can be made arbitrarily close to 1 by choosing A large.

Let $A > 0$ be chosen such that (4.24) holds. Since $\mathcal{X}_0^{\mathrm{CS}} = \overline{(X_0^{\mathrm{CS}}, r_0^{\mathrm{CS}}, \mu_0^{\mathrm{CS}})} \in \mathbb{U}^{\mathbb{R}\times[0,1]}$, the population in X_0^{CS} with spatial marks in $[-A,A]$ can be partitioned into a countable collection of disjoint balls of radius $\frac{1}{n}$. Let us label these balls by $B_{n,1}, B_{n,2}, \ldots \subset X_0^{\mathrm{CS}}$ in decreasing order of their total measure, and all individuals in B_i is given type $\frac{1}{i} \in [0,1]$, while all individuals with spatial mark outside $[-A,A]$ are given type 0. Note that by choosing M large, the total measure of individuals in $\mathcal{X}_0^{\mathrm{CS}}$ with types in $(0, \frac{1}{M}]$ can be made arbitrarily small. It then follows by the Feller continuity of the measure-valued process $\widehat{\mathcal{X}}^{\mathrm{CS}}$ stated in Remark 1.29, that as $M \to \infty$, the total measure of individuals in $\mathcal{X}_t^{\mathrm{CS}}$, $t \in [0,1]$, with spatial marks in $[-k,k]$ and types in $(0, \frac{1}{M}]$, converge in probability to the stochastic process which is identically 0 on the time interval $[0,1]$. Combined with (4.24), this implies that we can choose M large, such that

$$\mathbb{P}\big(\mathcal{X}_t^{\mathrm{CS}} \notin G_{\frac{1}{4kn},\frac{1}{M}} \text{ for some } 0 \leq t \leq 1\big) \leq \frac{\delta_n}{2}. \tag{4.26}$$

Let $\widetilde{L}(n)$ denote the maximal distance between the balls $B_{n,1}, \ldots, B_{n,M}$. Note that on the event that $\{\mathcal{X}_t^{\mathrm{CS}} \in G_{\frac{1}{4kn},\frac{1}{M}} \text{ for all } 0 \leq t \leq 1\}$, for each $t \in [0,1]$, the set of individuals in $\mathcal{X}_t^{\mathrm{CS}}$ with spatial marks in $[-k,k]$ and types in $[0, \frac{1}{M}]$ have total measure at most $\frac{1}{4kn}$; while the rest of the population with spatial marks in $[-k,k]$ trace their genealogies back at time 0 to $B_{n,1}, \ldots, B_{n,M}$, and their mutual distance is bounded by $\widetilde{L}(n) + 2$. Letting $L(n) = \widetilde{L}(n) + 3$ in (4.21) then readily implies that $\overline{(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}}, 1_{\{|v|\leq k\}}\mu_t^{\mathrm{CS}}(\mathrm{d}x\mathrm{d}v))} \in K_\delta^{2,n}$ on the event $\mathcal{X}_t^{\mathrm{CS}} \in G_{\frac{1}{4kn},\frac{1}{M}}$, and (4.26) implies that

$$\mathbb{P}\big(\overline{(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}}, 1_{\{|v|\leq k\}}\mu_t^{\mathrm{CS}}(\mathrm{d}x\mathrm{d}v))} \notin K_\delta^{2,n} \text{ for some } 0 \leq t \leq 1\big) \leq \frac{\delta_n}{2}, \tag{4.27}$$

which is the analogue of (4.22) for $\mathcal{X}^{\mathrm{CS}}$.

Next we turn to $\mathcal{X}^{\mathrm{FV},\varepsilon}$ and we will deduce (4.22) from (4.27) by exploiting Theorem 1.32 on the convergence of the measure-valued processes $S_\epsilon\widehat{\mathcal{X}}^{\mathrm{FV},\epsilon}$ to $\widehat{\mathcal{X}}^{\mathrm{CS}}$. Let A be the same as above. Let individuals in $\mathcal{X}_0^{\mathrm{CS}}$ be assigned types in $\{0\}\cup\{\frac{1}{i}:i\in\mathbb{N}\}$ as before (4.26), where $\widetilde{L}(n)$ denotes the maximal distance between any pair of individuals in $\mathcal{X}_0^{\mathrm{CS}}$ with types in $[\frac{1}{M},1]$. The assumption that $S^\epsilon\mathcal{X}_0^{\mathrm{FV},\epsilon}$ converges to $\mathcal{X}_0^{\mathrm{CS}}$ in $\mathbb{U}^{\mathbb{R}}$ as $\epsilon\downarrow 0$ allows a coupling between $S_\epsilon X_0^{\mathrm{FV},\epsilon}$ and X_0^{CS}, such that for most of the individuals in $S_\epsilon X_0^{\mathrm{FV},\epsilon}$ with spatial marks in $[-A,A]$, their genealogical distances and spatial marks are close to their counterparts in X_0^{CS}. We can then assign types to individuals in $S_\epsilon X_0^{\mathrm{FV},\epsilon}$ in such a way that: individuals with spatial marks outside $[-A,A]$ are assigned type 0 while those with spatial marks in $[-A,A]$ are assigned types in $\{\frac{1}{i}:i\in\mathbb{N}\}$. The distance between any pair of individuals with types in $[\frac{1}{M},1]$ is bounded by $\widetilde{L}(n)+1$, and the measure on geographic and type space, $S_\epsilon\widehat{\mathcal{X}}_0^{\mathrm{FV},\epsilon}$, converges vaguely to $\widehat{\mathcal{X}}_0^{\mathrm{CS}}$ as $\epsilon\downarrow 0$. By Theorem 1.32, $(S_\epsilon\widehat{\mathcal{X}}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t})_{0\le t\le 1}$ converges weakly to $(\widehat{\mathcal{X}}_t^{\mathrm{CS}})_{0\le t\le 1}$ as random variables in $C([0,1],\mathcal{M}(\mathbb{R}\times[0,1]))$. Applying this weak convergence result to (4.26) then implies that for $\epsilon>0$ sufficiently small, we have the following analogue of (4.26):

$$\mathbb{P}\big(S_\epsilon\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}\notin G_{\frac{1}{2kn},\frac{1}{M}}\text{ for some }0\le t\le 1\big)\le\delta_n. \tag{4.28}$$

Note that on the event $S_\epsilon\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}\in G_{\frac{1}{2kn},\frac{1}{M}}$ for all $0\le t\le 1$, for each $0\le t\le 1$, the set of individuals in $S_\epsilon\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}$ with spatial marks in $[-k,k]$ and types in $[0,\frac{1}{M}]$ have total measure at most $\frac{1}{2kn}$; while the rest of the population with spatial marks in $[-k,k]$ trace their genealogies back to an individual at time 0 with type in $[\frac{1}{M},1]$, and hence their pairwise distance is bounded by $L(n)=\widetilde{L}(n)+3$. It follows that $S_\epsilon^{(k)}\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}\in K_\delta^{2,n}$ on the event $S_\epsilon\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}\in G_{\frac{1}{2kn},\frac{1}{M}}$. Together with (4.28), this implies (4.22).

(3) Our procedure for constructing K_δ^3 is similar to that of K_δ^2. For each $n\in\mathbb{N}$, we will find $M=M(n)$ such that if $K_\delta^{3,n}$ denotes the subset of $\mathbb{U}_f^{\mathbb{R}}$ with the property that for each $\overline{(X,r,\mu)}\in K_\delta^{3,n}$, we can find M balls of radius $\frac{1}{n}$ in X, say $B_1,\dots,B_M$ with $B:=\cup_{i=1}^M B_i$, such that $\mu(X\backslash B\times\mathbb{R})<\frac{1}{n}$, then uniformly in $\epsilon>0$ sufficiently small, we have

$$\mathbb{P}\big(S_\epsilon^{(k)}\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}\notin K_\delta^{3,n}\text{ for some }0\le t\le 1\big)\le\frac{\delta}{2^n}=:\delta_n. \tag{4.29}$$

We can then take $K_\delta^3:=\cap_{n\in\mathbb{N}}K_\delta^{2,n}$, which clearly satisfies condition (iii) in Theorem A.1, while (4.29) implies that (4.20) holds with K_δ replaced by K_δ^3.

To find $M(n)$ such that (4.29) holds, we partition the time interval $[0,1]$ into $[\frac{i-1}{2n},\frac{i}{2n}]$ for $1\le i\le 2n$. It then suffices to show that for each $1\le i\le 2n$, we can find $M_i(n)$ such that if $K_\delta^{3,n}$ is defined using M_i, then uniformly in $\epsilon>0$ small,

$$\mathbb{P}\big(S_\epsilon^{(k)}\mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t}\notin K_\delta^{3,n}\text{ for some }\frac{i-1}{2n}\le t\le\frac{i}{2n}\big)\le\frac{\delta_n}{2n}=\frac{\delta}{2n2^n}. \tag{4.30}$$

Again we first determine $M(n)$ for CSSM and then use the convergence of measure-valued IFV to measure-valued CSSM. We now prove an analogue of (4.30) for $\mathcal{X}^{\mathrm{CS}}$. Since $\mathcal{X}^{\mathrm{CS}}_{\frac{i-1}{2n}}\in\mathbb{U}^{\mathbb{R}\times[0,1]}$ almost surely, we can condition on its realization and partition the population in $\mathcal{X}^{\mathrm{CS}}_{\frac{i-1}{2n}}$ with spatial marks in $[-A,A]$ into disjoint balls of radius $\frac{1}{3n}$, $B_1,B_2,\dots$, as we did in the argument leading to (4.26). Repeating the same argument there and assigning type $\frac{1}{j}$ to individuals in B_j, we readily obtain the following analogue of (4.26): we can choose M_i large enough such that

$$\mathbb{P}\big(\mathcal{X}_t^{\mathrm{CS}}\notin G_{\frac{1}{2n},\frac{1}{M_i}}\text{ for some }\frac{i-1}{2n}\le t\le\frac{i}{2n}\big)\le\frac{\delta_n}{4n}. \tag{4.31}$$

Note that on the event $\mathcal{X}_t^{\mathrm{CS}} \in G_{\frac{1}{2n},\frac{1}{M_i}}$ for all $\frac{i-1}{2n} \le t \le \frac{i}{2n}$, for each $t \in [\frac{i-1}{2n}, \frac{i}{2n}]$, the set of individuals in $\mathcal{X}_t^{\mathrm{CS}}$ with spatial marks in $[-k,k]$ and types in $[0, \frac{1}{M_i}]$ have total measure at most $\frac{1}{2n}$; while the rest of the individuals in $\mathcal{X}_t^{\mathrm{CS}}$ with spatial marks in $[-k,k]$ trace their genealogies back to an individual at time $\frac{i-1}{2n}$ in $B_1 \cup B_2 \cdots \cup B_{M_i}$, and hence they are contained in M_i balls of radius $\frac{1}{3n} + t \le \frac{5}{6n}$. Therefore $\overline{(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}}, 1_{\{|v|\le k\}}\mu_t^{\mathrm{CS}}(\mathrm{d}x\mathrm{d}v))} \in K_\delta^{3,n}$, and the analogue of (4.30) holds for $\mathcal{X}^{\mathrm{CS}}$.

To establish (4.30) uniformly in small $\epsilon > 0$, we again apply the convergence result in Theorem 1.32. Note that by the f.d.d. convergence established in Section 4.1, $S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\frac{i-1}{2n}}$ converges in distribution to $\mathcal{X}^{\mathrm{CS}}_{\frac{i-1}{2n}}$ as $\epsilon \downarrow 0$. Following the same argument as those leading to (4.28), we can then assign types to individuals in $S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\frac{i-1}{2n}}$ such that the associated measure-valued process, $(S_\epsilon \widehat{\mathcal{X}}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}t})_{\frac{i-1}{2n}\le t\le \frac{i}{2n}}$ converges weakly to $(\widehat{\mathcal{X}}_t^{\mathrm{CS}})_{\frac{i-1}{2n}\le t\le \frac{i}{2n}}$, and individuals in $S_\epsilon \mathcal{X}^{\mathrm{FV},\epsilon}_{\epsilon^{-2}\frac{i-1}{2n}}$ with spatial marks in $[-A,A]$ and types in $[\frac{1}{M_i}, 1]$ are contained in M_i balls with radius at most $\frac{1}{2n}$. Applying this convergence result to (4.31) then implies that for $\epsilon > 0$ sufficiently small, we have the following analogue of (4.31):

$$\mathbb{P}\big(S_\epsilon \mathcal{X}_t^{\mathrm{FV},\epsilon} \notin G_{\frac{1}{n},\frac{1}{M_i}} \text{ for some } \frac{i-1}{2n} \le t \le \frac{i}{2n}\big) \le \frac{\delta_n}{2n}. \tag{4.32}$$

This is then easily seen to imply (4.30).

Combining parts (1)-(3) concludes the proof of (4.20) and hence establishes the compact containment condition **(J1)**. ∎

4.3 Proof of convergence of rescaled IFV processes (Theorem 1.32)

In [AS11, Theorem 1.1], a convergence result similar to Theorem 1.32 was proved for the voter model on $\mathbb{Z}$, where the type space consists of only $\{0,1\}$, and a special initial condition was considered, where the population to the left of the origin all have type 1, and the rest of the population have type 0. The proof consists of two parts: proof of tightness, and convergence of finite-dimensional distribution.

In [AS11], the proof of tightness for the voter model does not depend on the initial condition, and is based on the verification of Jakubowski's criterion and Aldous' criterion as we have done in Sections 4.1 and 4.2 for the genealogical process. Because the IFV process ignores the genealogical distances, the verification of the compact containment condition **(J1)** in Jakubowski's criterion is trivial, as in the case for the voter model. Using the duality between the IFV process and coalescing random walks with *delayed* coalescence, Aldous' criterion on the tightness of evaluations can be verified by exactly the same calculations as that for the voter model in [AS11], which uses the duality between the voter model and coalescing random walks with *instantaneous* coalescence. Recall here that in the rescaling the difference between instantaneous and delayed coalescence disappears because of recurrence of the difference walk. Lastly, the convergence of the finite-dimensional distribution for rescaled IFV process follows the same calculations as in Section 4.1, where we can simply enlarge the mark space to $\mathbb{R} \times [0,1]$ and suppress the genealogical distances by choosing $\phi_i \equiv 1$. ∎

5 Martingale problem for CSSM genealogy processes

In this section, we show that the CSSM genealogy process constructed in Section 1.4 solves the martingale problem formulated in Theorem 1.40. We will first identify the generator action on regular test functions evaluated at regular states, and then extend it to more general test functions and verify the martingale property. Complications

arise mainly from the singular nature of the resampling component of the generator, which are only well-defined a priori on regular test functions evaluated at regular states. Fortunately by Proposition 1.25, the CSSM genealogy process enters these regular states as soon as $t > 0$, even though the initial state may not be regular.

5.1 Generator action on regular test functions

In this section, we identify the generator of the CSSM genealogy process $\mathcal{X}^{\mathrm{CS}}$, acting on $\Phi \in \Pi_{\mathrm{r}}^{1,2}$ and evaluated at $\mathcal{X}_t^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$, where $\Pi_{\mathrm{r}}^{1,2}$ and $\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$ are introduced in Section 1.6. The advantage of working with such regular Φ and $\mathcal{X}_t^{\mathrm{CS}}$ is that, the relevant resamplings only occur at the boundary points of balls of radius at least δ, which is a locally finite set. In Section 5.2, we will extend it to the case $\Phi \in \Pi^{1,2}$, where we need to consider the boundaries of all balls in $\mathcal{X}_t^{\mathrm{CS}}$, which is a locally infinite set.

Proposition 5.1. [Generator action on regular test functions]
Let $\mathcal{X}^{\mathrm{CS}} := (\mathcal{X}_t^{\mathrm{CS}})_{t\geq 0}$ be the CSSM genealogy process with $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$. Let $\Phi = \Phi^{n,\phi,g} \in \Pi_{\mathrm{r}}^{1,2}$, defined as in Definition 1.36. Let $L^{\mathrm{CS}} = L_{\mathrm{d}}^{\mathrm{CS}} + L_{\mathrm{a}}^{\mathrm{CS}} + L_{\mathrm{r}}^{\mathrm{CS}}$ be defined as in (1.53)–(1.56)*. Then we have*

$$\lim_{t\downarrow 0} \frac{\mathbb{E}[\Phi(\mathcal{X}_t^{\mathrm{CS}})] - \Phi(\mathcal{X}_0^{\mathrm{CS}})}{t} = L^{\mathrm{CS}}\Phi(\mathcal{X}_0^{\mathrm{CS}}). \tag{5.1}$$

Furthermore,

$$\mathbb{E}[\Phi(\mathcal{X}_t^{\mathrm{CS}})] = \Phi(\mathcal{X}_0^{\mathrm{CS}}) + \int_0^t \mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}})]\mathrm{d}s, \tag{5.2}$$

where $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ is continuous in $t \geq 0$.

The proof is fairly long and technical and will be broken into parts, with (5.1) and (5.2) proved respectively in Sections 5.1.1 and 5.1.2.

5.1.1 Proof of (5.1) in Proposition 5.1

For each $t \geq 0$, denote $\mathcal{X}_t^{\mathrm{CS}} = \overline{(X_t^{\mathrm{CS}}, r_t^{\mathrm{CS}}, \mu_t^{\mathrm{CS}})}$. Let $L > 0$ be chosen such that the support of $g(\underline{x})$ is contained in $(-L, L)^n$. Let $\delta > 0$ be determined by ϕ as in (1.52), so that $\phi((r_{i,j})_{1\leq i<j\leq n})$ is constant when any coordinate $r_{i,j}$ varies on the interval $r_{i,j} \in [0, \delta]$.

We proceed in five steps, first giving a suitable representation of $\Phi(X_0^{\mathrm{CS}})$ and $\Phi(X_t^{\mathrm{CS}})$, and then calculating actions that lead to the different parts of the generator.

Step 1 (Representation of $\Phi(X_0^{\mathrm{CS}})$). We derive here a representation of $\Phi(X_0^{\mathrm{CS}})$ by partitioning X_0^{CS} into disjoint balls of radius at least $\delta/4$ and utilizing the fact that $\Phi \in \Pi_{\mathrm{r}}^{1,2}$.

Since $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$, by Remark 1.35 and (1.42), we can identify X_t^{CS} for each $t \geq 0$ with

$$\mathbb{R}_t^* := \bigcup_{x\in E_t} \{x^+, x^-\} \bigcup (\mathbb{R}\backslash E_t), \tag{5.3}$$

where x^+ and x^- are duplicates of the point x in

$$E_t = \cup_{l>0} E_t^l, \tag{5.4}$$

where E_t^l is the set of points in $\mathbb{R}$ that lie at the boundary of two disjoint balls of radius l in X_t^{CS}, which is consistent with the definition in (1.46).

Denote

$$\{y_1 < y_2 < \cdots < y_m\} := E_0^{\frac{\delta}{4}} \cap [-2L, 2L], \tag{5.5}$$

and let y_0, resp. y_{m+1}, be the point in $E_0^{\frac{\delta}{4}}$ adjacent to y_1, resp. y_m. Note that the intervals $[y_i^+, y_{i+1}^-] := (y_i, y_{i+1}) \cup \{y_i^+, y_{i+1}^-\}$ form disjoint open balls of radius $\frac{\delta}{4}$ in X_0^{CS}. Therefore

for all $x_1 \in [y_i^+, y_{i+1}^-]$ and $x_2 \in [y_j^+, y_{j+1}^-]$ with $i \neq j$, $d_0^{ij} := r_0^{\mathrm{CS}}(x_1, x_2)$ is constant, and $d_0 := (d_0^{ij})_{0 \le i \le j \le m}$ forms a distance matrix, where we set $d_0^{ii} := 0$ for $0 \le i \le m$. We can then write

$$\Phi(\mathcal{X}_0^{\mathrm{CS}}) = \Phi^{n,\phi,g}(\mathcal{X}_0^{\mathrm{CS}}) = \sum_{\underline{k} \in \{0,\cdots,m\}^n} \phi(d_0^{\underline{k}\times\underline{k}})\, G(\underline{y}, \underline{k}), \tag{5.6}$$

where $\underline{y} = (y_0, \cdots, y_{m+1})$, $d_0^{\underline{k}\times\underline{k}}$ is the distance matrix with $d_{0,ij}^{\underline{k}\times\underline{k}} = d_0^{k_i k_j}$, $1 \le i \le j \le n$, and

$$G(\underline{y}, \underline{k}) = \int_{\substack{y_{k_i} < x_i < y_{k_i+1} \\ 1 \le i \le n}} \cdots \int g(\underline{x})\, d\underline{x}. \tag{5.7}$$

Step 2 (Representation of $\Phi(X_t^{\mathrm{CS}})$). We next write $\Phi(\mathcal{X}_t^{\mathrm{CS}})$, for $t > 0$, in terms of coalescing Brownian motions running forward in time (in contrast to the spatial genealogies which run backward in time), which determine the evolution of boundaries between disjoint balls in $\mathcal{X}_t^{\mathrm{CS}}$.

Let $\{y_i + B_i\}_{0 \le i \le m+1}$ be independent Brownian motions starting from $\{y_i\}_{0 \le i \le m+1}$, from which we construct a family of coalescing Brownian motions $\{y_i + \widetilde{B}_i\}_{0 \le i \le m+1}$. Namely, let $y_0 + \widetilde{B}_0 := y_0 + B_0$ for all time, and let $y_1 + \widetilde{B}_1(s) := y_1 + B_1(s)$ until the first time $y_1 + B_1(s)$ hits $y_0 + \widetilde{B}_0$. From that time onward, define $y_1 + \widetilde{B}_1$ to coincide with $y_0 + \widetilde{B}_0$. In the same way, we successively define $\{y_i + \widetilde{B}_i\}_{2 \le i \le m+1}$ from $\{y_i + B_i\}_{2 \le i \le m+1}$ by adding one path at a time. Without loss of generality, we may assume that $y_i + \widetilde{B}_i$ is the a.s. unique path in the Brownian web $\mathcal{W}$ starting from $(y_i, 0)$.

To write $\Phi(\mathcal{X}_t^{\mathrm{CS}})$ in terms of the forward coalescing Brownian motions $\widetilde{B}$, we observe that

$$E_s^\delta \cap [y_0 + \widetilde{B}_0(s), y_{m+1} + \widetilde{B}_{m+1}(s)] \subset \bigcup_{i=0}^{m+1} \{y_i + \widetilde{B}_i(s)\} \quad \forall\, 0 \le s < \min\Big\{\frac{\delta}{4}, t\Big\}, \tag{5.8}$$

since by our construction of $\mathcal{X}_s^{\mathrm{CS}}$ in Section 1.4, $(y_i + \widetilde{B}_i(s))_{0 \le i \le m+1}$ are boundaries of disjoint balls in $\mathcal{X}_s^{\mathrm{CS}}$, and $[(y_i + \widetilde{B}_i(s))^+, (y_{i+1} + \widetilde{B}_{i+1}(s))^-]$, $0 \le i \le m$, consists of either empty sets if $y_i + \widetilde{B}_i(s) = y_{i+1} + \widetilde{B}_{i+1}(s)$, or disjoint open balls of radius $\frac{\delta}{4} + s$ in X_s^{CS}.

Let $\{\mathrm{meet}\}_t$ denote the event that either $y_0 + B_0(s)$ reaches level $-L$ before time t, or $y_{m+1} + B_{m+1}(s)$ reaches level L before time t, or one of the pair $(y_i + B_i(s), y_{i+1} + B_{i+1}(s))$ meet before time t. On the complementary event $\{\mathrm{meet}\}_t^c$, $\widetilde{B}_i = B_i$ for all i. Therefore, for all $0 \le t < \frac{\delta}{4}$, we can write

$$\Phi(\mathcal{X}_t^{\mathrm{CS}}) = 1_{\{\mathrm{meet}\}_t} \Phi(\mathcal{X}_t^{\mathrm{CS}}) + 1_{\{\mathrm{meet}\}_t^c} \widetilde{\Phi}(\mathcal{X}_t^{\mathrm{CS}}) \tag{5.9}$$

with

$$\widetilde{\Phi}(\mathcal{X}_t^{\mathrm{CS}}) = \sum_{\underline{k} \in \{0,\cdots,m\}^n} \phi(d_t^{\underline{k}\times\underline{k}})\, G(\underline{y} + \underline{B}(t), \underline{k}), \tag{5.10}$$

where $d_t^{\underline{k}\times\underline{k}}$ is the distance matrix with $d_{t,ij}^{\underline{k}\times\underline{k}} = d_0^{k_i k_j} + 2t(1 - \delta_{k_i,k_j})$ for $1 \le i \le j \le n$, and $\underline{B}(t) = (B_0(t), \cdots, B_{m+1}(t))$.

Since

$$|\Phi(\mathcal{X}_t^{\mathrm{CS}}) - \widetilde{\Phi}(\mathcal{X}_t^{\mathrm{CS}})| = 1_{\{\mathrm{meet}\}_t} |\Phi(\mathcal{X}_t^{\mathrm{CS}}) - \widetilde{\Phi}(\mathcal{X}_t^{\mathrm{CS}})| \le C_\Phi 1_{\{\mathrm{meet}\}_t} \tag{5.11}$$

for some C_Φ depending on Φ, and the probability of the event $\{\mathrm{meet}\}_t$ decays exponentially in t^{-1} by elementary estimates for Brownian motions, we have $\lim_{t\to 0} \frac{\mathbb{E}[\Phi(\mathcal{X}_t^{\mathrm{CS}})] - \mathbb{E}[\widetilde{\Phi}(\mathcal{X}_t^{\mathrm{CS}})]}{t} = 0$. Thus, we may replace $\Phi(\mathcal{X}_t^{\mathrm{CS}})$ by $\widetilde{\Phi}(\mathcal{X}_t^{\mathrm{CS}})$ up to error $o(t)$ as $t \to 0$.

By (5.6) and (5.10), we can write

$$\begin{aligned}&\widetilde{\Phi}(\mathcal{X}_t^{\mathrm{CS}}) - \Phi(\mathcal{X}_0^{\mathrm{CS}}) \\ &= \sum_{\underline{k}\in\{0,\cdots,m\}^n} \big(\phi(d_t^{\underline{k}\times\underline{k}}) - \phi(d_0^{\underline{k}\times\underline{k}})\big)\, G(\underline{y}+\underline{B}(t),\underline{k}) + \sum_{\underline{k}\in\{0,\cdots,m\}^n} \phi(d_0^{\underline{k}\times\underline{k}})\big(G(\underline{y}+\underline{B}(t),\underline{k}) - G(\underline{y},\underline{k})\big).\end{aligned} \tag{5.12}$$

Step 3 (Aging). We first identify the aging term $L_{\mathrm{a}}^{\mathrm{CS}}$ in the generator, defined in (1.55). For each term in the first sum in (5.12), by Taylor expanding $\phi(d_t^{\underline{k}\times\underline{k}})$ in t, it is easy to see that

$$\lim_{t\to 0} t^{-1}\mathbb{E}[\big(\phi(d_t^{\underline{k}\times\underline{k}}) - \phi(d_0^{\underline{k}\times\underline{k}})\big)\, G(\underline{y}+\underline{B}(t),\underline{k})] = 2\sum_{1\le i<j\le n} \frac{\partial\phi}{\partial r_{ij}}(d_0^{\underline{k}\times\underline{k}})\, G(\underline{y},\underline{k}), \tag{5.13}$$

where we note that $\frac{\partial\phi}{\partial r_{ij}}(d_0^{\underline{k}\times\underline{k}}) = 0$ if $k_i = k_j$, since in this case the ij-th argument of ϕ, $d_0^{k_ik_j}$, is less than δ. Summing the above limit over $\underline{k}\in\{0,\cdots,m\}^n$ and using the definition of G, we find

$$\begin{aligned}&\lim_{t\to 0}\ \frac{1}{t} \sum_{\underline{k}\in\{0,\cdots,m\}^n} \mathbb{E}\left[\big(\phi(d_t^{\underline{k}\times\underline{k}}) - \phi(d_0^{\underline{k}\times\underline{k}})\big)\, G(\underline{y}+\underline{B}(t),\underline{k})\right] \\ &\qquad\qquad = 2\int_{\mathbb{R}^n} g(\underline{x}) \sum_{1\le i<j\le n} \frac{\partial\phi}{\partial r_{ij}}(\underline{r})\,\mathrm{d}\underline{x},\end{aligned} \tag{5.14}$$

where $\underline{\underline{r}} := (r_0^{\mathrm{CS}}(x_i,x_j))_{1\le i,j\le n}$. This gives the aging term $L_{\mathrm{a}}^{\mathrm{CS}}$.

Step 4 (Resampling). We next identify the resampling term $L_{\mathrm{r}}^{\mathrm{CS}}$, defined in (1.56). For the second sum in (5.12), we need to compute

$$\begin{aligned}&\lim_{t\to 0}\ \frac{1}{t} \sum_{\underline{k}\in\{0,\cdots,m\}^n} \phi(d_0^{\underline{k}\times\underline{k}})\mathbb{E}[G(\underline{y}+\underline{B}(t),\underline{k}) - G(\underline{y},\underline{k})] \\ &= \sum_{\underline{k}\in\{0,\cdots,m\}^n} \phi(d_0^{\underline{k}\times\underline{k}})\ \lim_{t\to 0}\frac{1}{t}\ \mathbb{E}[G(\underline{y}+\underline{B}(t),\underline{k}) - G(\underline{y},\underline{k})].\end{aligned} \tag{5.15}$$

For each $\underline{k}\in\{0,\cdots,m\}^n$,

$$G(\underline{y}+\underline{B}(t),\underline{k}) - G(\underline{y},\underline{k}) = \int_{\mathbb{R}^n} g(\underline{x})\Big(\prod_{i=1}^n 1_{[y_{k_i}+B_{k_i}(t),\, y_{k_i+1}+B_{k_i+1}(t)]}(x_i) - \prod_{i=1}^n 1_{[y_{k_i},\, y_{k_i+1}]}(x_i)\Big)\mathrm{d}\underline{x}. \tag{5.16}$$

We can rewrite the difference of the product of indicators as

$$\prod_{i=1}^n \big(1_{[y_{k_i},\, y_{k_i+1}]}(x_i) - 1_{[y_{k_i},\, y_{k_i}+B_{k_i}(t)]}(x_i) + 1_{[y_{k_i+1},\, y_{k_i+1}+B_{k_i+1}(t)]}(x_i)\big) - \prod_{i=1}^n 1_{[y_{k_i},\, y_{k_i+1}]}(x_i). \tag{5.17}$$

We can expand the first product above and sort the result into three groups of terms, (G_1), (G_2) and (G_3), depending on whether each term contains one, two, or more factors of the form $1_{[y_i,\, y_i+B_i(t)]}$ for some $0\le i\le m+1$. If $h(\underline{x})$ denotes a term in (G_3), then necessarily $\int_{\mathbb{R}^n} g(\underline{x})h(\underline{x}) \le C_g|B_{i_1}(t)B_{i_2}(t)B_{i_3}(t)|$ for some C_g depending only on g and some $i_1,i_2,i_3\in\{0,\cdots,m+1\}$. Since $\mathbb{E}|B_{i_1}(t)B_{i_2}(t)B_{i_3}(t)| \le ct^{\frac{3}{2}}$, terms in ($\mathrm{G}_3$) do not contribute to the limit in (5.15).

Each term in (G_2) is of the following form, where $1 \le i \ne j \le n$,

$$\begin{aligned}
&1_{[y_{k_i}, y_{k_i}+B_{k_i}(t)]}(x_i) 1_{[y_{k_j}, y_{k_j}+B_{k_j}(t)]}(x_j) \prod_{\substack{1\le\tau\le n\\ \tau\ne i,j}} 1_{[y_{k_\tau}, y_{k_\tau+1}]}(x_\tau),\\
&1_{[y_{k_i+1}, y_{k_i+1}+B_{k_i+1}(t)]}(x_i) 1_{[y_{k_j+1}, y_{k_j+1}+B_{k_j+1}(t)]}(x_j) \prod_{\substack{1\le\tau\le n\\ \tau\ne i,j}} 1_{[y_{k_\tau}, y_{k_\tau+1}]}(x_\tau),\\
&-1_{[y_{k_i+1}, y_{k_i+1}+B_{k_i+1}(t)]}(x_i) 1_{[y_{k_j}, y_{k_j}+B_{k_j}(t)]}(x_j) \prod_{\substack{1\le\tau\le n\\ \tau\ne i,j}} 1_{[y_{k_\tau}, y_{k_\tau+1}]}(x_\tau),\\
&-1_{[y_{k_i}, y_{k_i}+B_{k_i}(t)]}(x_i) 1_{[y_{k_j+1}, y_{k_j+1}+B_{k_j+1}(t)]}(x_j) \prod_{\substack{1\le\tau\le n\\ \tau\ne i,j}} 1_{[y_{k_\tau}, y_{k_\tau+1}]}(x_\tau).
\end{aligned} \tag{5.18}$$

Denote the four terms in (5.18) respectively by $h_{ij}^{(1)}(\underline{x})$, $h_{ij}^{(2)}(\underline{x})$, $h_{ij}^{(3)}(\underline{x})$ and $h_{ij}^{(4)}(\underline{x})$. Then

$$\int_{\mathbb{R}^n} h_{ij}^{(1)}(\underline{x}) g(\underline{x}) \mathrm{d}\underline{x} = \int_{\substack{y_{k_\tau}<x_\tau<y_{k_\tau+1}\\ \tau\ne i,j}} \cdots \int \int_{x_i=y_{k_i}}^{y_{k_i}+B_{k_i}(t)} \int_{x_j=y_{k_j}}^{y_{k_j}+B_{k_j}(t)} \big(g(x_1,\cdots,y_{k_i},\cdots,y_{k_j},\cdots,x_n)+o(1)\big)\mathrm{d}\underline{x}, \tag{5.19}$$

where we replaced $g(\underline{x})$ by $g(x_1,\cdots,y_{k_i},\cdots,y_{k_j},\cdots,x_n)$, with an error of $o(1)$ as $t\downarrow 0$. Therefore

$$\lim_{t\to 0}\frac{1}{t}\mathbb{E}\left[\int_{\mathbb{R}^n} h_{ij}^{(1)}(\underline{x}) g(\underline{x})\mathrm{d}\underline{x}\right] = \delta_{k_i,k_j} \int_{\substack{y_{k_\tau}<x_\tau<y_{k_\tau+1}\\ \tau\ne i,j}} \cdots \int g(x_1,\cdots,y_{k_i},\cdots,y_{k_j},\cdots x_n) \prod_{\substack{1\le\tau\le n\\ \tau\ne i,j}} \mathrm{d}x_\tau. \tag{5.20}$$

We obtain similar results for $h_{ij}^{(2)}$, $h_{ij}^{(3)}$ and $h_{ij}^{(4)}$.

For a fixed pair $i\ne j$, when we sum over $\underline{k}$ and all contributions from $h_{ij}^{(\cdot)}$ in (5.15), we obtain an integral for $\phi(\underline{r})g(\underline{x})$, where x_τ, $\tau\ne i,j$, are still integrated over $\mathbb{R}^n$, however the integration for x_i and x_j are replaced by summation over $\{y_\sigma\}_{1\le\sigma\le m}$. Contributions only come from $x_i=x_j$, and is positive when $k_i=k_j$, and negative when $k_i=k_j+1$ or $k_j=k_i+1$. Writing everything in terms of $\mathcal{X}_0^{\mathrm{CS}}$, we easily verify that the contribution of terms in (G_2) to the limit in (5.15) is exactly

$$\sum_{1\le i\ne j\le n} \int_{\mathbb{R}^{n-2}} \sum_{x_i,x_j\in\{y_\sigma^+,y_\sigma^-:1\le\sigma\le m\}} 1_{\{x_i=x_j\}} g(\underline{x})(\theta_{ij}\phi-\phi)(\underline{r}) \prod_{\substack{1\le\tau\le n\\ \tau\ne i,j}} \mathrm{d}x_\tau, \tag{5.21}$$

where $\theta_{ij}\phi$ is defined as in (1.19), and $\underline{r} := (r_0^{\mathrm{CS}}(x_i,x_j))_{1\le i,j\le n}$. This gives the resampling term $L_{\mathrm{r}}^{\mathrm{CS}}$ defined in (1.56).

Step 5 (Migration). Lastly we identify the diffusion (migration) term $L_{\mathrm{d}}^{\mathrm{CS}}$, defined in (1.54). We note that each term in group (G_1) is of the following form, where $1\le j\le n$,

$$\begin{aligned}
&-1_{[y_{k_j}, y_{k_j}+B_{k_j}(t)]}(x_j) \prod_{\substack{1\le i\le n\\ i\ne j}} 1_{[y_{k_i}, y_{k_i+1}]}(x_i),\\
&1_{[y_{k_j+1}, y_{k_j+1}+B_{k_j+1}(t)]}(x_j) \prod_{\substack{1\le i\le n\\ i\ne j}} 1_{[y_{k_i}, y_{k_i+1}]}(x_i),
\end{aligned} \tag{5.22}$$

Denote the two terms respectively by $h_j^{(1)}(\underline{x})$ and $h_j^{(2)}(\underline{x})$. Then

$$\begin{aligned}&\int_{\mathbb{R}^n} h_j^{(1)}(\underline{x})g(\underline{x})\mathrm{d}\underline{x}\\&= -\int_{\substack{y_{k_i}<x_i<y_{k_i+1}\\ i\neq j}}\cdots\int\int_{x_j=y_{k_j}}^{y_{k_j}+B_{k_j}(t)}\Big(g(\underline{x}\big|_{x_j=y_{k_j}})+(x_j-y_{k_j})\frac{\partial g}{\partial x_j}(\underline{x}\big|_{x_j=y_{k_j}})+o(|x_j-y_{k_j}|)\Big)\mathrm{d}\underline{x}.\end{aligned} \tag{5.23}$$

Therefore,

$$\lim_{t\to 0}\frac{1}{t}\mathbb{E}\left[\int_{\mathbb{R}^n} h_j^{(1)}(\underline{x})g(\underline{x})\mathrm{d}\underline{x}\right] = -\frac{1}{2}\int_{\substack{y_{k_i}<x_i<y_{k_i+1}\\ i\neq j}}\cdots\int\frac{\partial g}{\partial x_j}(\underline{x}\big|_{x_j=y_{k_j}})\prod_{\substack{1\le i\le n\\ i\neq j}}\mathrm{d}x_i. \tag{5.24}$$

A similar result holds for $h_j^{(2)}$. Combining the two, we see that

$$\lim_{t\to 0}\frac{1}{t}\mathbb{E}\left[\int_{\mathbb{R}^n}\big(h_j^{(1)}(\underline{x})+h_j^{(2)}(\underline{x})\big)g(\underline{x})\mathrm{d}\underline{x}\right] = \frac{1}{2}\int_{y_{k_i}<x_i<y_{k_i+1}}\cdots\int\frac{\partial^2 g}{\partial^2 x_j}(\underline{x})\prod_{1\le i\le n}\mathrm{d}x_i. \tag{5.25}$$

Therefore, the contribution from terms in (G_1) to the limit in (5.15) gives precisely the migration term $L_{\mathrm{d}}^{\mathrm{CS}}$ defined in (1.54). This establishes the generator formula (5.1). ∎

5.1.2 Proof of (5.2) in Proposition 5.1

The complications in proving (5.2) arise from trying to prove uniform integrability for various quantities. We proceed in three steps. First we show that, for each $t>0$,

$$\lim_{h\downarrow 0}\frac{\mathbb{E}[\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}})]-\mathbb{E}[\Phi(\mathcal{X}_t^{\mathrm{CS}})]}{h} = \mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]. \tag{5.26}$$

By the Markov property of $(\mathcal{X}_t^{\mathrm{CS}})_{t\ge 0}$,

$$\lim_{h\downarrow 0}\frac{\mathbb{E}[\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}})]-\mathbb{E}[\Phi(\mathcal{X}_t^{\mathrm{CS}})]}{h} = \lim_{h\downarrow 0}\mathbb{E}\left[\frac{\mathbb{E}[\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}})|\mathcal{X}_t^{\mathrm{CS}}]-\Phi(\mathcal{X}_t^{\mathrm{CS}})}{h}\right]. \tag{5.27}$$

Since $\mathcal{X}_t^{\mathrm{CS}}\in\mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$ a.s. by Proposition 1.25, $\lim_{h\downarrow 0}\frac{\mathbb{E}[\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}})|\mathcal{X}_t^{\mathrm{CS}}]-\Phi(\mathcal{X}_t^{\mathrm{CS}})}{h} = L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})$ almost surely. Therefore, the first step is to interchange limit and expectation in (5.27) and to deduce (5.26). We need to show that

$$\Big(\frac{\mathbb{E}[\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}})|\mathcal{X}_t^{\mathrm{CS}}]-\Phi(\mathcal{X}_t^{\mathrm{CS}})}{h}\Big)_{h>0}\text{ is uniformly integrable for } h>0 \text{ small}, \tag{5.28}$$

say $0<h<\frac{\delta\wedge t}{2}$.

Once the uniform integrability has been verified, in Step 2 we prove that $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ is continuous in t, and then in Step 3 we put things together and prove (5.2).

Step 1 (Uniform integrability). This step constitutes the bulk of the proof of (5.2). Let us fix $\mathcal{X}_t^{\mathrm{CS}}$ and examine the error terms in our earlier calculation of the generator formula (5.1) with $\mathcal{X}_0^{\mathrm{CS}}$ replaced by $\mathcal{X}_t^{\mathrm{CS}}$. Instead of partitioning X_t^{CS} into disjoint open balls of radius $\delta/4$ as done in (5.5), we set

$$\{y_1<\cdots<y_{m_t}\} := E_t^{\frac{\delta\wedge t}{4}}\cap[-2L,2L], \tag{5.29}$$

which is determined by the dual Brownian web $\widehat{\mathcal{W}}$ as seen from the definition in (1.46). This choice of partitioning of X_t^{CS} removes the dependence of $\underline{y} = (y_1, \dots, y_{m_t})$ on the initial condition $\mathcal{X}_0^{\mathrm{CS}}$. Note that m_t depends on $\mathcal{X}_t^{\mathrm{CS}}$.

Let $\mathbb{E}_t[\cdot]$ denote the conditional expectation $\mathbb{E}[\cdot|\mathcal{X}_t^{\mathrm{CS}}]$. Following the arguments leading to (5.11),

$$\begin{aligned} & h^{-1}\big|\mathbb{E}_t[\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}})] - \Phi(\mathcal{X}_t^{\mathrm{CS}})\big| \\ = \;& h^{-1}\Big|\mathbb{E}_t[\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}}) - \widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}})] + h^{-1}\mathbb{E}_t[\widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}}) - \Phi(\mathcal{X}_t^{\mathrm{CS}})]\Big| \\ \le \;& h^{-1}\mathbb{E}_t\big[1_{\{\mathrm{meet}\}_h}|\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}}) - \widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}})|\big] + h^{-1}\big|\mathbb{E}_t[\widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}}) - \Phi(\mathcal{X}_t^{\mathrm{CS}})]\big|, \end{aligned} \tag{5.30}$$

where $\{\mathrm{meet}\}_h$ is the event that either $y_0 + B_0(s)$ hits level $-L$ before time h, or $y_{m_t+1} + B_{m_t+1}(s)$ hits level L before time h, or one of the pair $(y_i + B_i(s), y_{i+1} + B_{i+1}(s))$ coalesces before time h. We now estimate the two terms in (5.30).

(i) We start with the second term in (5.30). Based on the decomposition (5.12), $\widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}}) - \Phi(\mathcal{X}_t^{\mathrm{CS}}) = H(h, \underline{B}(h))$ for some function $H(h, z_0, \cdots, z_{m_t+1})$ which is continuously differentiable in h and three times continuously differentiable in $z_0, \dots, z_{m_t+1}$ with uniformly bounded derivatives. Since the generator formula (5.1) is derived by Taylor expanding $H(h, \underline{B}(h))$, it is not hard to see that uniformly in $h \in (0, \frac{\delta \wedge t}{4})$,

$$\begin{aligned} \Big|\mathbb{E}_t[\widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}}) - \Phi(\mathcal{X}_t^{\mathrm{CS}})] - hL^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})\Big| &= \big|\mathbb{E}_t[H(h, \underline{B}(h))] - hL^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})\big| \\ &\le C_{g,\phi}(m_t + 3)h \end{aligned} \tag{5.31}$$

for some $C_{g,\phi}$ depending on g and ϕ, and where $m_t + 3$ denotes the number of variables in $H(h, z_0, \dots, z_{m_t+1})$. Therefore

$$h^{-1}|\mathbb{E}_t[\widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}}) - \Phi(\mathcal{X}_t^{\mathrm{CS}})]| \le C_{g,\phi}(m_t + 3) + |L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})| \tag{5.32}$$

uniformly for $h \in (0, \frac{\delta \wedge t}{4})$. From the definition of $L^{\mathrm{CS}}\Phi$, we note that

$$|L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})| \le C_{g,\phi}(1 + m_t) < \infty. \tag{5.33}$$

By the definition of m_t in (5.29) and by Lemma B.5, we have $\mathbb{E}[m_t] \le \frac{4L}{\sqrt{\pi \frac{\delta \wedge t}{4}}} < \infty$. This implies the uniform integrability of the second term in (5.30) for $h \in (0, \frac{\delta \wedge t}{4})$.

(ii) We now consider the first term in (5.30). For $0 \le i \le m_t$, let $\tau_{i,i+1}$ be the first hitting time between $y_i + B_i(s)$ and $y_{i+1} + B_{i+1}(s)$. Let τ_0 be the first hitting time of level $-L$ by $y_0 + B_0(s)$, and τ_{m_t+1} the first hitting time of level L by $y_{m_t+1} + B_{m_t+1}(s)$. Then

$$\begin{aligned} & h^{-1}\mathbb{E}_t\big[1_{\{\mathrm{meet}\}_h}|\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}}) - \widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}})|\big] \\ \le \;& 2|\Phi|_\infty h^{-1}\mathbb{E}_t\big[1_{\{\tau_0 \le h\}} + 1_{\{\tau_{m_t+1} \le h\}}\big] + h^{-1}\mathbb{E}_t\Big[\sum_{i=0}^{m_t} 1_{\{\tau_{i,i+1} \le h\}}\big|\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}}) - \widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}})\big|\Big]. \end{aligned} \tag{5.34}$$

Since $y_0 < -2L$ and $y_{m_t+1} > 2L$, the probability of the events $\{\tau_0 \le h\}$ and $\{\tau_{m_t+1} \le h\}$ decay exponentially fast in h^{-1} and the events are independent of $\mathcal{X}_t^{\mathrm{CS}}$. Therefore the first term in (5.34) is uniformly bounded in $h > 0$.

Bounding the second term in (5.34) is more delicate, because it remains of order 1 as $h \downarrow 0$. We will need to use negative correlation inequalities for coalescing Brownian motions established in Appendix C.

Let us recall the definition of $\widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}})$ from (5.10), where we replace the pair $\{0, t\}$ by $\{t, t+h\}$. By integrating over the population at time $t + h$, we note that both $\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}})$ and $\widetilde{\Phi}(\mathcal{X}_{t+h}^{\mathrm{CS}})$ can be written as integrals of $g(\underline{x})\phi(\underline{r}(\underline{x}))$ integrated over $\underline{x} = (x_1, \cdots, x_n) \in$

$(-L,L)^n$, except that: for a given $\underline{x}$, the distance matrix $\underline{\underline{r}}(\underline{x})$ may be different for Φ and $\widetilde{\Phi}$, and for $\widetilde{\Phi}$, the same point $\underline{x}$ may be integrated over several times with different distance matrix $\underline{\underline{r}}(\underline{x})$ due to the fact that $\{y_i + B_i(h)\}_{0\le i\le m_t+1}$ may not have the same order as $\{y_i\}_{0\le i\le m+1}$. However, for $\underline{x}$ not in

$$D := \Big\{\underline{x}\in\mathbb{R}^n : x_i \in \bigcup_{s\in[0,h]} \{y_j + B_j(s)\} \text{ for some } 1\le i\le n \text{ and } 0\le j\le m_t+1\Big\}, \quad (5.35)$$

$\underline{x}$ is integrated over exactly once in $\widetilde{\Phi}$, and the associated distance matrix $\underline{\underline{r}}(\underline{x})$ is the same for both Φ and $\widetilde{\Phi}$. Therefore, contributions from $\underline{x}\notin D$ cancel out in $|\Phi(\mathcal{X}^{\mathrm{CS}}_{t+h}) - \widetilde{\Phi}(\mathcal{X}^{\mathrm{CS}}_{t+h})|$. Since g has compact support, the contribution from $\underline{x}\in D$ to $|\Phi(\mathcal{X}^{\mathrm{CS}}_{t+h}) - \widetilde{\Phi}(\mathcal{X}^{\mathrm{CS}}_{t+h})|$, including multiple integrations over the same $\underline{x}$ by $\widetilde{\Phi}$, is at most $C_{\phi,g}n|g|_\infty|\phi|_\infty\sum_{i=0}^{m_t+1}R_i(h)$, where $R_i(h) := (\sup_{0\le s\le h}B_i(s) - \inf_{0\le s\le h}B_i(s))$. Therefore,

$$\begin{aligned}
&h^{-1}\mathbb{E}_t\Big[\sum_{i=0}^{m_t} 1_{\{\tau_{i,i+1}\le h\}}\big|\Phi(\mathcal{X}^{\mathrm{CS}}_{t+h}) - \widetilde{\Phi}(\mathcal{X}^{\mathrm{CS}}_{t+h})\big|\Big]\\
&\le C_{n,\phi,g}h^{-1}\mathbb{E}_{\underline{B}}\Big[\sum_{i=0}^{m_t} 1_{\{\tau_{i,i+1}\le h\}}\sum_{j=0}^{m_t+1}R_j(h)\Big]\\
&\le C_{n,\phi,g}h^{-1}\sum_{i=0}^{m_t}\sum_{j=0}^{m_t+1}\mathbb{P}_{\underline{B}}(\tau_{i,i+1}\le h)^{\frac12}\mathbb{E}_{\underline{B}}[R_j(h)^2]^{\frac12}\\
&\le C'_{n,\phi,g}h^{-\frac12}(m_t+2)\sum_{i=0}^{m_t}\mathbb{P}_{\underline{B}}(\tau_{i,i+1}\le h)^{\frac12},
\end{aligned} \quad (5.36)$$

where $\mathbb{E}_{\underline{B}}$ denotes expectation w.r.t. the Brownian motions $\underline{B} = (B_0,\dots,B_{m_t+1})$, and we applied Hölder inequality and the fact that $\mathbb{E}[R_j(h)^2]^{\frac12} = c\sqrt{h}$ for some $c>0$. It only remains to prove the uniform integrability of the r.h.s. of (5.36) w.r.t. the law of $\mathcal{X}^{\mathrm{CS}}_t$ for $0<h<\frac{\delta\wedge t}{4}$.

Note that the r.h.s. of (5.36) depends on y_0 and y_{m_t+1}, which lie outside $[-2L,2L]$. To control the dependence on y_0 and y_{m_t+1}, we enlarge the interval and let $\{z_1,\cdots,z_{M+1}\} := E_t^{\frac{\delta\wedge t}{4}}\cap(-2L-1,2L+1)$ as in (5.29), which contains $\{y_1,\cdots,y_{m_t}\}$ as a subset. Denote

$$\psi\big(\frac{y_{i+1}-y_i}{\sqrt{h}}\big) := \mathbb{P}(\tau_{i,i+1}\le h)^{\frac12} = \Big(2\int_{\frac{y_{i+1}-y_i}{\sqrt{2h}}}^{\infty} e^{-\frac{x^2}{2}}dx\Big)^{\frac12},$$

we can then replace the r.h.s. of (5.36) by

$$F_h(\mathcal{X}^{\mathrm{CS}}_t) := \frac{M}{\sqrt{h}}\sum_{i=1}^{M}\psi\big(\frac{z_{i+1}-z_i}{\sqrt{h}}\big), \quad (5.37)$$

because $C'_{n,\phi,g}F_h(\mathcal{X}^{\mathrm{CS}}_t)$ dominates the r.h.s. of (5.36) except for possible missing terms $\frac{1}{\sqrt{h}}\psi\big(\frac{y_1-y_0}{\sqrt{h}}\big)(m_t+2)$, resp. $\frac{1}{\sqrt{h}}\psi\big(\frac{y_{m_t+1}-y_{m_t}}{\sqrt{h}}\big)(m_t+2)$, on the event $y_0\le -2L-1$, resp. $y_{m_t+1}\ge 2L+1$. Since $y_1\ge -2L$ and $y_{m_t}\le 2L$ by definition, both missing terms are bounded by $C_1m_t+C_2$ uniformly for $h>0$, and is thus uniformly integrable.

It only remains to prove the uniform integrability of $F_h(\mathcal{X}^{\mathrm{CS}}_t)$ uniformly in $0<h<\frac{\delta\wedge t}{4}$. We achieve this by bounding its second moment. Note that by Cauchy-Schwarz,

$$\mathbb{E}[F_h(\mathcal{X}^{\mathrm{CS}}_t)^2] = \frac{1}{h}\,\mathbb{E}\left[M^2\Big(\sum_{i=1}^{M}\psi\big(\frac{z_{i+1}-z_i}{\sqrt{h}}\big)\Big)^2\right]$$

$$\le \frac{1}{h}\,\mathbb{E}\Big[M^3\sum_{i=1}^{M}\psi\big(\frac{z_{i+1}-z_i}{\sqrt{h}}\big)^2\Big] \tag{5.38}$$
$$\le \frac{1}{h}\,\mathbb{E}\Big[\sum_{i_1,i_2,i_3=1}^{M+1}\sum_{i_4=1}^{M}\psi\big(\frac{z_{i_4+1}-z_{i_4}}{\sqrt{h}}\big)^2\Big].$$

Note that the above summation can be seen as a summation of $x_1 := z_{i_1}, x_2 := z_{i_2}, x_3 := z_{i_3}$ in $\{z_1,\dots,z_{M+1}\}$ and $x_4 := z_{i_4}$ in $\{z_1,\dots,z_M\}$. By integrating over x_1,x_2,x_3,x_4 over $[-2L-1,2L+1]$ and the position $x_5 := z_{i_4+1} \in (x_4, 2L+1)$, we can rewrite the r.h.s. of (5.38) in terms of the correlation functions of the translation invariant simple point process $\xi := E_t^{\frac{\delta\wedge t}{4}} \subset \mathbb{R}$. More precisely, it can be written as

$$\int\cdots\int_{\substack{x_4<x_5\\ |x_1|,|x_2|,|x_3|,|x_4|,|x_5|<2L+1}} \frac{1}{h}\psi\big(\frac{x_5-x_4}{\sqrt{h}}\big)^2 K^{\mathrm{c}}_{4,5}(x_1,x_2,x_3,x_4,x_5)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}x_5, \tag{5.39}$$

where

$$K^{\mathrm{c}}_{4,5}(x_1,\cdots,x_5) := \lim_{\epsilon\downarrow 0}\epsilon^{-5}\mathbb{P}\big([x_j,x_j+\epsilon]\cap\xi\neq\emptyset\ \forall\ 1\le j\le 5, (x_4,x_5)\cap\xi=\emptyset\big) \tag{5.40}$$

is the density of finding a point at x_i for each $1\le i\le 5$, with no point in (x_4,x_5). By the definition of E_t^l in (1.46) and the duality between the forward and dual Brownian web $(\mathcal{W},\widehat{\mathcal{W}})$, we see that ξ is the point process generated on $\mathbb{R}$ at time t by coalescing Brownian motions in the Brownian web $\mathcal{W}$ starting from every point in $\mathbb{R}$ at time $t-\frac{\delta\wedge t}{4}$. By the negative correlation inequality in Lemma C.3, we can bound

$$\begin{aligned}
&K^{\mathrm{c}}_{4,5}(x_1,\cdots,x_5)\\
&\le \lim_{\epsilon\downarrow 0}\epsilon^{-5}\mathbb{P}\big([x_j,x_j+\epsilon]\cap\xi\neq\emptyset,\ j=1,2,3\big)\mathbb{P}\big([x_j,x_j+\epsilon]\cap\xi\neq\emptyset, j=4,5;\ (x_4,x_5)\cap\xi=\emptyset\big)\\
&= K(0)^3K^{\mathrm{c}}(x_4,x_5),
\end{aligned}$$

where by Lemma B.5,

$$K(0) = \lim_{\epsilon\downarrow 0}\epsilon^{-1}\mathbb{P}\big([x,x+\epsilon]\cap\xi\neq\emptyset\big) = \frac{2}{\sqrt{\pi\delta\wedge t}} \qquad \text{for all } x\in\mathbb{R},$$

and

$$K^{\mathrm{c}}(x,y) = \lim_{\epsilon\downarrow 0}\epsilon^{-2}\mathbb{P}\big([x,x+\epsilon]\cap\xi\neq\emptyset,(x+\epsilon,y)\cap\xi=\emptyset,[y,y+\epsilon]\cap\xi\neq\emptyset\big) \quad \text{for } x<y.$$

By Lemma C.6, for $x<y$,

$$K^{\mathrm{c}}(x,y)\le C_{\delta,t}(y-x)\wedge 1$$

for some $C_{\delta,t}>0$ depending only on δ and t. Substituting the above bounds into (5.39), using the definition of ψ, and separating the integration into two regions depending on whether $0<x_5-x_4<\sqrt{h}$ or $x_5-x_4\ge\sqrt{h}$, it is easily seen that the integral in (5.39) is uniformly bounded for $0<h<\frac{\delta\wedge t}{4}$. This implies the uniform integrability of $F_h(\mathcal{X}_t^{\mathrm{CS}})$ for $0<h<\frac{\delta\wedge t}{4}$, and hence that of $\frac{\mathbb{E}_t[\Phi(\mathcal{X}_{t+h}^{\mathrm{CS}})]-\Phi(\mathcal{X}_t^{\mathrm{CS}})}{h}$.

Step 2 (Continuity of $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$). Recall from (1.53)–(1.56) that $L^{\mathrm{CS}} = L^{\mathrm{CS}}_{\mathrm{d}}+L^{\mathrm{CS}}_{\mathrm{a}}+L^{\mathrm{CS}}_{\mathrm{r}}$. We will prove the continuity for each component of the generator.

By our assumptions on g and ϕ, $|L^{\mathrm{CS}}_{\mathrm{d}}\Phi|_\infty, |L^{\mathrm{CS}}_{\mathrm{a}}\Phi|_\infty \le C_\Phi<\infty$. It was shown in (3.4) that for any $t>0$, for Lebesgue a.e. $\underline{x}=(x_1,\cdots,x_n)\in\mathbb{R}^n$, the distance matrix $r_s^{\mathrm{CS}}(x_i,x_j)$

converges a.s. to $r_t^{\mathrm{CS}}(x_i, x_j)$ as $s \to t$, and this conclusion is also easily seen to hold for $t = 0$ by the assumption $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$. Therefore, almost surely w.r.t. $(\mathcal{X}_s^{\mathrm{CS}})_{s\geq 0}$,

$$\lim_{s\to t} L_{\mathrm{d}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}}) = L_{\mathrm{d}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}}) \quad \text{and} \quad \lim_{s\to t} L_{\mathrm{a}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}}) = L_{\mathrm{a}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}}). \tag{5.41}$$

Therefore $\mathbb{E}[L_{\mathrm{d}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ and $\mathbb{E}[L_{\mathrm{a}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ are continuous in $t \geq 0$ by the bounded convergence theorem.

We now turn to the continuity of $\mathbb{E}[L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$. We first prove the continuity at $t = 0$ and later point out how it extends to $t > 0$. Let $E_0^{\frac{\delta}{4}} \subset \mathbb{R}$ be defined from $\mathcal{X}_0^{\mathrm{CS}}$ as in (1.46), which are the boundaries of balls of radius $\frac{\delta}{4}$ in $\mathcal{X}_0^{\mathrm{CS}}$. Let us follow paths in the Brownian web $\mathcal{W}$ starting from points in $E_0^{\frac{\delta}{4}} \subset \mathbb{R}$ at time 0, which determine the evolution of these boundaries, and denote the point set on $\mathbb{R}$ generated at time $s > 0$ by $E_0^{\frac{\delta}{4}}(s)$. Then $E_s^{\frac{\delta}{4}+s} \subset E_0^{\frac{\delta}{4}}(s)$. By our regularity assumption on ϕ, only resampling at the boundaries of balls of radius δ or more has an effect on $\Phi(\mathcal{X}_s^{\mathrm{CS}})$. Therefore for $s \in (0, \delta/4)$, we have

$$L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}}) = \sum_{1\leq i\neq j\leq n} \int_{\mathbb{R}^{n-2}} \sum_{x_i,x_j\in\{y^+,y^-:y\in E_0^{\delta/4}(s)\}} 1_{\{x_i=x_j\}} g(\underline{x})(\theta_{ij}\phi - \phi)(\underline{r}) \prod_{\substack{1\leq\tau\leq n\\ \tau\neq i,j}} \mathrm{d}x_\tau, \tag{5.42}$$

where $\theta_{ij}\phi$ is defined as in (1.19), and $\underline{\underline{r}} := (r_s^{\mathrm{CS}}(x_i, x_j))_{1\leq i,j\leq n}$. By our assumptions on g and ϕ, the fact $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$, and our construction of $\mathcal{X}_s^{\mathrm{CS}}$ in terms of the (dual) Brownian web, it is then easily seen by dominated convergence that

$$\lim_{s\downarrow 0} L_{\mathrm{r}}\Phi(\mathcal{X}_s^{\mathrm{CS}}) = L_{\mathrm{r}}\Phi(\mathcal{X}_0^{\mathrm{CS}}) \quad \text{almost surely.} \tag{5.43}$$

To prove the continuity of $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ at $t = 0$, it only remains to verify the uniform integrability of $L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}})$ for s close to 0, say $s \in [0, \delta/4]$. We will achieve this by showing that $L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}})$ has uniformly bounded second moments.

Note that because g is assumed to be supported on $[-L, L]^n$, we have

$$|L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}})| \leq C_{n,\phi,g}\big|E_0^{\delta/4}(s) \cap (-L, L)\big|. \tag{5.44}$$

By Lemma C.2, $E_0^{\delta/4}(s)$ is negatively correlated, and hence by Lemma C.5, we have

$$\mathbb{E}\Big[\big|E_0^{\delta/4}(s) \cap (-L, L)\big|^2\Big] \leq 2\mathbb{E}\Big[\big|E_0^{\delta/4}(s) \cap (-L, L)\big|\Big] + \Big[\mathbb{E}\big|E_0^{\delta/4}(s) \cap (-L, L)\big|\Big]^2. \tag{5.45}$$

Thus it suffices to bound $\mathbb{E}\big[\big|E_0^{\delta/4}(s) \cap (-L, L)\big|\big]$ uniformly in $s \in [0, \delta/4]$.

Since $E_0^{\delta/4}(s)$ is obtained by evolving coalescing Brownian motions starting from $E_0^{\delta/4}$, we can bound

$$\mathbb{E}\big[\big|E_0^{\delta/4}(s) \cap (-L, L)\big|\big] \leq \big|E_0^{\delta/4} \cap [-2L, 2L]\big| + 2\sum_{i\geq 2L} \mathbb{E}\big[\big|\xi_s^{[i,i+1]\times\{0\}} \cap (-L, L)\big|\big], \tag{5.46}$$

where $\xi_s^{[i,i+1]\times\{0\}} \subset \mathbb{R}$ is the point set generated at time s by coalescing Brownian motions in the Brownian web $\mathcal{W}$ starting from everywhere in $[i, i + 1]$ at time 0.

Note that the first term in (5.46) is finite and independent of $s \geq 0$. We now treat the second term. For each $i \geq 2L$, let us denote $\xi_s^{[i,i+1]\times\{0\}} \cap (-L, L)$ by $\xi_{s,L}^i$, which is also a point process on $(-L, L)$ satisfying negative correlation. In particular,

$$\mathbb{E}[|\xi_{s,L}^i|] = \sum_{n=1}^{\infty} \mathbb{P}(|\xi_{s,L}^i| \geq n) \leq \sum_{n=1}^{\infty} \mathbb{P}(|\xi_{s,L}^i \geq 1)^n \leq \sum_{n=1}^{\infty} \Big(\frac{1}{\sqrt{2\pi s}} \int_{i-L}^{\infty} e^{-\frac{x^2}{2s}} dx\Big)^n, \tag{5.47}$$

where we used Lemma C.4 and the observation that, $\xi^i_{s,L} \neq \emptyset$ implies that the Brownian motion starting at $(i,0)$ in the Brownian web must be to the left of L at time s. Since L is fixed and large, $\frac{1}{\sqrt{2\pi s}} \int_{i-L}^{\infty} e^{-\frac{x^2}{2s}} dx \leq \alpha$ for some $\alpha \in (0,1)$ uniformly in $i \geq 2L$ and $s \in [0, \delta/4]$, and hence

$$\mathbb{E}[|\xi^i_{s,L}|] \leq \frac{\int_{i-L}^{\infty} e^{-\frac{x^2}{2s}} dx}{(1-\alpha)\sqrt{2\pi s}} = \frac{\int_{\frac{i-L}{\sqrt{s}}}^{\infty} e^{-\frac{x^2}{2}} dx}{(1-\alpha)\sqrt{2\pi}}. \tag{5.48}$$

It is then clear that $\sum_{i \geq 2L} \mathbb{E}[|\xi^i_{s,L}|]$ tends to 0 as $s \downarrow 0$ and is uniformly bounded for $s \in [0, \delta/4]$, which concludes the proof of the uniform integrability of $L_{\mathrm{r}}\Phi(\mathcal{X}^{\mathrm{CS}}_s)$ for $s \in [0, \delta/4]$. Therefore $\mathbb{E}[L^{\mathrm{CS}}_{\mathrm{r}}\Phi(\mathcal{X}^{\mathrm{CS}}_t)]$ is continuous at $t = 0$.

The continuity of $\mathbb{E}[L^{\mathrm{CS}}_{\mathrm{r}}\Phi(\mathcal{X}^{\mathrm{CS}}_t)]$ at $t > 0$ can be proved similarly. For $s \in [t - \frac{\delta \wedge t}{4}, t + \frac{\delta \wedge t}{4}]$, we can use the same representation as in (5.42) by treating $t - \frac{\delta \wedge t}{4}$ as the starting time 0 in (5.42). The a.s. convergence $\lim_{s \to t} L_{\mathrm{r}}\Phi(\mathcal{X}^{\mathrm{CS}}_s) = L_{\mathrm{r}}\Phi(\mathcal{X}^{\mathrm{CS}}_t)$ then follows as before. The uniform integrability of $L_{\mathrm{r}}\Phi(\mathcal{X}^{\mathrm{CS}}_s)$ follows easily from the density estimate, Lemma B.5, if we restrict to $s \in [t - \frac{\delta \wedge t}{8}, t + \frac{\delta \wedge t}{8}]$, since only resampling at locations occupied at time s by Brownian web paths starting from $\mathbb{R}$ at time $t - \frac{\delta \wedge t}{4}$ has an effect on $\Phi(\mathcal{X}^{\mathrm{CS}}_s)$.

Step 3 (Proof of (5.2)). We have thus far shown that $\mathbb{E}[\Phi(\mathcal{X}^{\mathrm{CS}}_t)]$ has a continuous right derivative $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}^{\mathrm{CS}}_t)]$ on $[0, \infty)$. Note that $\mathbb{E}[\Phi(\mathcal{X}^{\mathrm{CS}}_t)]$ is also continuous on $[0, \infty)$ since $(\mathcal{X}^{\mathrm{CS}}_s)_{s \geq 0}$ has a.s. continuous sample paths by Theorem 1.27. Equation (5.2) then follows from a variant of the fundamental theorem of calculus, formulated as Lemma 5.2 below. This finally completes the proof of (5.2) and that of Prop. 5.1. ∎

Lemma 5.2. [Fundamental Theorem of Calculus]
Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function with a continuous right derivative $D^+f(x) := \lim_{h \downarrow 0} \frac{1}{h}(f(x+h) - f(x))$. Then

$$f(x) = f(0) + \int_0^x D^+f(y)\mathrm{d}y \qquad \textit{for all } x \in \mathbb{R}.$$

Proof. Let $g(x) := f(0) + \int_0^x D^+f(y)\mathrm{d}y$, which is clearly differentiable with continuous right derivative D^+f. Then $h := f - g$ is continuous with $h(0) = 0$ and right derivative $D^+h \equiv 0$. It suffices to show that $h \equiv 0$.

For $\epsilon > 0$, let $h_\epsilon(x) = h(x) + \epsilon x$. Then $D^+h_\epsilon \equiv \epsilon > 0$. This implies that h_ϵ is non-decreasing on $\mathbb{R}$, since otherwise if $h_\epsilon(x) > h_\epsilon(y)$ for some $x < y$, then at the point $x_0 \in [x, y]$ where h_ϵ achieves its maximum on $[x, y]$, we have $D^+h_\epsilon(x_0) \leq 0$, contradicting $D^+h_\epsilon \equiv \epsilon$. Letting $\epsilon \downarrow 0$, the monotonicity of h_ϵ then implies that h is also non-decreasing. Applying the same argument to $-h$ shows that h is also non-increasing. Thus $h \equiv 0$. ∎

5.2 Generator action on general test functions

In this section, we extend the validity of the integral equation (5.2) in Prop. 5.1 to general $\Phi \in \Pi^{1,2}$, assuming $\mathcal{X}^{\mathrm{CS}}_0 \in \mathbb{U}^{\mathbb{R}}_{\mathrm{r}}$. The complication is that the generator identified in (5.1), which acts on regular test functions $\Phi \in \Pi^{1,2}_{\mathrm{r}}$ and evaluated at regular states $\mathcal{X}^{\mathrm{CS}}_t \in \mathbb{U}^{\mathbb{R}}_{\mathrm{r}}$, can only be extended to general test functions $\Phi \in \Pi^{1,2}$ provided that we restrict to the regular subclass of states $\mathcal{X}^{\mathrm{CS}}_t \in \mathbb{U}^{\mathbb{R}}_{\mathrm{rr}}$ introduced in Definition 1.39. Fortunately for each $t > 0$, almost surely $\mathcal{X}^{\mathrm{CS}}_t \in \mathbb{U}^{\mathbb{R}}_{\mathrm{rr}}$, which makes it possible to extend from $\Phi \in \Pi^{1,2}_{\mathrm{r}}$ to $\Phi \in \Pi^{1,2}$.

Proposition 5.3. [Generator action on general test functions]
Let $\mathcal{X}^{\mathrm{CS}}$ be the CSSM genealogy process with $\mathcal{X}^{\mathrm{CS}}_0 \in \mathbb{U}^{\mathbb{R}}_{\mathrm{r}}$, and let $\Phi = \Phi^{n,\phi,g} \in \Pi^{1,2}$. Then

(a) *The integral equation* (5.2) *still holds, and* $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ *is continuous in* $t > 0$.
(b) *If* $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$, *defined as in Definition 1.39, then the generator equation* (5.1) *still holds, and* $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ *is continuous in* $t \geq 0$.

Proof. We will approximate $\Phi = \Phi^{n,\phi,g} \in \Pi^{1,2}$ by $\Phi_l \in \Pi_{\mathrm{r}}^{1,2}$ as follows. Let $\rho : [0,\infty) \to \mathbb{R}$ be continuously differentiable with $\rho(x) = 0$ for $x \in [0,2]$, $\rho'(x) > 0$ for $x \in (2,3]$, and $\rho(x) = x$ for $x \in [3,\infty)$. For each $l > 0$, denote $\rho_l(x) = l\rho(x/l)$. Then ρ_l is a smooth truncation with $\sup_{x\geq 0}|\rho_l(x) - x| \to 0$ and $\rho_l'(x) \to 1$ for each $x > 0$ as $l \downarrow 0$. Given $\phi = \phi((r_{ij})_{1\leq i<j\leq n})$, let $\phi_l := \phi \circ \rho_l$. We then define the truncated version of Φ by $\Phi_l := \Phi^{n,\phi_l,g}$.

It is easily seen that $\Phi_l \in \Pi_{\mathrm{r}}^{1,2}$, and hence Proposition 5.1 can be applied. We will then deduce Proposition 5.3 by taking the limit $l \downarrow 0$.

(a) We will take the limit $l \downarrow 0$ in the integral equation (5.2). In **Step 1**, we will show that (5.2) also holds for Φ. In **Step 2**, we will prove the continuity of $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ in $t > 0$. Note that without assuming $\Phi \in \Pi_{\mathrm{r}}^{1,2}$ or $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$, $L^{\mathrm{CS}}\Phi(\mathcal{X}_0^{\mathrm{CS}})$ may not be well-defined.

Step 1 First note that $|\Phi_l - \Phi|_\infty \to 0$ as $l \downarrow 0$ by our assumption on g and ϕ and the fact that $\sup_{x\geq 0}|\rho_l(x) - x| \to 0$ as $l \downarrow 0$. Therefore,

$$\lim_{l\downarrow 0} \mathbb{E}[\Phi_l(\mathcal{X}_t^{\mathrm{CS}})] = \mathbb{E}[\Phi(\mathcal{X}_t^{\mathrm{CS}})] \qquad \text{and} \qquad \lim_{l\downarrow 0} \Phi_l(\mathcal{X}_0^{\mathrm{CS}}) = \Phi(\mathcal{X}_0^{\mathrm{CS}}). \tag{5.49}$$

Recall the decomposition of $L^{\mathrm{CS}} = L_{\mathrm{d}}^{\mathrm{CS}} + L_{\mathrm{a}}^{\mathrm{CS}} + L_{\mathrm{r}}^{\mathrm{CS}}$ in Prop. 5.1. Note that $|L_{\mathrm{d}}^{\mathrm{CS}}\Phi|_\infty$, $|L_{\mathrm{d}}^{\mathrm{CS}}\Phi_l|_\infty$, $|L_{\mathrm{a}}^{\mathrm{CS}}\Phi|_\infty$, $|L_{\mathrm{a}}^{\mathrm{CS}}\Phi_l|_\infty$ are all uniformly bounded by some $C_\Phi < \infty$ independent of l. Also, by our assumptions on g, ϕ and ρ_l, for each $s > 0$ and a.s. every realization of $\mathcal{X}_s^{\mathrm{CS}}$, we have

$$L_{\mathrm{d}}^{\mathrm{CS}}\Phi_l(\mathcal{X}_s^{\mathrm{CS}}) \to L_{\mathrm{d}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}}) \text{ and } L_{\mathrm{a}}^{\mathrm{CS}}\Phi_l(\mathcal{X}_s^{\mathrm{CS}}) \to L_{\mathrm{a}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}}) \text{ as } l \downarrow 0. \tag{5.50}$$

Therefore, by the bounded convergence theorem,

$$\lim_{l\downarrow 0} \int_0^t \mathbb{E}[(L_{\mathrm{d}}^{\mathrm{CS}}\Phi_l + L_{\mathrm{a}}^{\mathrm{CS}}\Phi_l)(\mathcal{X}_s^{\mathrm{CS}})]\mathrm{d}s = \int_0^t \mathbb{E}[(L_{\mathrm{d}}^{\mathrm{CS}}\Phi + L_{\mathrm{a}}^{\mathrm{CS}}\Phi)(\mathcal{X}_s^{\mathrm{CS}})]\mathrm{d}s. \tag{5.51}$$

For the resampling generator $L_{\mathrm{r}}^{\mathrm{CS}}$, note that for any $s > 0$ and a.s. every realization of $\mathcal{X}_s^{\mathrm{CS}}$,

$$\begin{aligned} |L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}})| &\leq \sum_{1\leq i\neq j\leq n} \int_{\mathbb{R}^{n-2}} \sum_{x_i,x_j\in\{x^+,x^-:x\in E_s\}} 1_{\{x_i=x_j\}} |g(\underline{x})| \left|(\theta_{ij}\phi - \phi)(\underline{r}_{\underline{=}s}^{\mathrm{CS}})\right| \prod_{\substack{1\leq\tau\leq n\\ \tau\neq i,j}} \mathrm{d}x_\tau \\ &\leq C_{n,g,\phi} \sum_{x\in(-M,M)\cap E_s} d_x \wedge 1, \end{aligned} \tag{5.52}$$

where E_s is defined as in (5.4), which is the subset of $\mathbb{R}$ that are points of multiplicity in the dual Brwonian web $\widehat{\mathcal{W}}$, $d_x = r_s^{\mathrm{CS}}(x^+,x^-)$ is the distance between the two points in X_s^{CS} with the same spatial location as $x \in E_s$, $M > 0$ is chosen such that $\mathrm{supp}(g) \subset (-M,M)^n$, and we used the assumption that ϕ has bounded derivative. Such a bound holds for ϕ_l uniformly in $l > 0$. If $\sum_{x\in(-M,M)\cap E_s} d_x \wedge 1 < \infty$, then by dominated convergence, we have

$$\lim_{l\downarrow 0} L_{\mathrm{r}}^{\mathrm{CS}}\Phi_l(\mathcal{X}_s^{\mathrm{CS}}) = L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}}). \tag{5.53}$$

To prove the analogue of (5.51) for $L_{\mathrm{r}}^{\mathrm{CS}}\Phi$, by dominated convergence, it only remains to show

$$\int_0^t \mathbb{E}\Big[\sum_{x\in(-M,M)\cap E_s} d_x \wedge 1\Big]\mathrm{d}s < \infty. \tag{5.54}$$

Note that for each $s \in (0,t)$,

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{x\in(-M,M)\cap E_s} d_x \wedge 1\Big] \\
&= \mathbb{E}\Big[\sum_{x\in(-M,M)\cap E_s} \int_0^1 1_{\{d_x\wedge 1>u\}}\mathrm{d}u\Big] = \int_0^1 \mathbb{E}\Big[\Big|\{x\in(-M,M)\cap E_s : d_x > u\}\Big|\Big]\mathrm{d}u \\
&\le \int_0^{2s\wedge 1} \mathbb{E}\Big[\Big|\{x\in(-M,M)\cap E_s : d_x > u\}\Big|\Big]\mathrm{d}u + \int_{2s\wedge 1}^1 \mathbb{E}\Big[\Big|\{x\in(-M,M)\cap E_s : d_x > 2s\}\Big|\Big]\mathrm{d}u \\
&= \int_0^{2s\wedge 1} \mathbb{E}\Big[\,\Big|\xi_s^{\mathbb{R}\times\{s-u/2\}}\cap(-M,M)\Big|\,\Big]\mathrm{d}u + \int_{2s\wedge 1}^1 \mathbb{E}\Big[\,\Big|\xi_s^{\mathbb{R}\times\{0\}}\cap(-M,M)\Big|\,\Big]\mathrm{d}u \\
&= CM\sqrt{2s\wedge 1} + CM\frac{1-2s\wedge 1}{\sqrt{s}} \le CM\big(1+\tfrac{1}{\sqrt{s}}\big),
\end{aligned} \tag{5.55}$$

where ξ_s^A denotes the point set on $\mathbb{R}$ generated at time s by the collection of paths in the Brownian web $\mathcal{W}$ starting from the space time set $A \subset \mathbb{R}^2$, and we used $\mathbb{E}[\,|\xi_s^{\mathbb{R}\times\{s-u\}}\cap[a,b]|\,] = \frac{b-a}{\sqrt{\pi u}}$ by Lemma B.5. Inequality (5.54) then follows. To summarize, we have thus shown that (5.2) is also valid for a general polynomial $\Phi \in \Pi^{1,2}$.

Step 2 To prove the continuity of $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ in $t>0$, we again decompose L^{CS} into its three summands. First note that the continuity of $\mathbb{E}[L_{\mathrm{d}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ and $\mathbb{E}[L_{\mathrm{a}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ on $[0,\infty)$ follow by the same arguments as that for $\mathbb{E}[L_{\mathrm{d}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ and $\mathbb{E}[L_{\mathrm{a}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ in the proof of Prop. 5.1.

To prove the continuity of $\mathbb{E}[L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ in $t>0$, we fix $t>0$ and a truncation parameter $\epsilon\in(0,t)$. For each $s\in(t-\epsilon,t+\epsilon)$, we decompose $L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}})$ into $L_{\mathrm{r}}^{<\epsilon+s-t}\Phi(\mathcal{X}_s^{\mathrm{CS}})$ and $L_{\mathrm{r}}^{\ge\epsilon+s-t}\Phi(\mathcal{X}_s^{\mathrm{CS}})$, where both $L_{\mathrm{r}}^{<\epsilon+s-t}\Phi(\mathcal{X}_s^{\mathrm{CS}})$ and $L_{\mathrm{r}}^{\ge\epsilon+s-t}\Phi(\mathcal{X}_s^{\mathrm{CS}})$ are defined as in (5.42), except that resampling therein is carried out by summing over

$$\begin{aligned}
&\sum_{x_i,x_j\in\{y^+,y^-:y\in E_s^{\epsilon+s-t}\}} 1_{\{x_i=x_j\}} \quad \text{for} \quad L_{\mathrm{r}}^{\ge\epsilon+s-t}\Phi(\mathcal{X}_s^{\mathrm{CS}}), \\
&\sum_{x_i,x_j\in\{y^+,y^-:y\in E_s\setminus E_s^{\epsilon+s-t}\}} 1_{\{x_i=x_j\}} \quad \text{for} \quad L_{\mathrm{r}}^{<\epsilon+s-t}\Phi(\mathcal{X}_s^{\mathrm{CS}}),
\end{aligned} \tag{5.56}$$

where E_s and E_s^δ are defined as in (5.4).

The same argument as in the proof of the continuity of $\mathbb{E}[L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ in Prop. 5.1 shows that

$$\lim_{s\to t}\mathbb{E}[|L_{\mathrm{r}}^{\ge\epsilon+s-t}\Phi(\mathcal{X}_s^{\mathrm{CS}}) - L_{\mathrm{r}}^{\ge\epsilon+s-t}\Phi(\mathcal{X}_t^{\mathrm{CS}})|] = 0. \tag{5.57}$$

On the other hand, for $s\in[t-\frac{\epsilon}{2},t+\frac{\epsilon}{2}]$, by the same calculations as those leading to (5.55), we have

$$\begin{aligned}
\mathbb{E}[|L_{\mathrm{r}}^{<\epsilon+s-t}\Phi(\mathcal{X}_s^{\mathrm{CS}})|] \;&\le\; C_{n,g,\varphi}\mathbb{E}\Big[\sum_{x\in(-M,M)\cap E_s\setminus E_s^{\epsilon+s-t}} d_x\wedge 1\Big] \\
= \;\; C_{n,g,\varphi}\mathbb{E}\Big[\sum_{\substack{x\in(-M,M)\cap E_s\\ d_x\le 2(\epsilon+s-t)}} d_x\wedge 1\Big] \;&\le\; CM\sqrt{\epsilon+s-t} \;\le\; C'M\sqrt{\epsilon}.
\end{aligned} \tag{5.58}$$

Since $\epsilon>0$ can be chosen arbitrarily small, (5.57) and (5.58) together imply that

$$\lim_{s\to t}\mathbb{E}[L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}})] = \mathbb{E}[L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})] \qquad \text{when } t>0.$$

(b) We now verify the continuity of $\mathbb{E}[L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ (and hence of $\mathbb{E}[L^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$) at $t=0$ under the additional assumption that $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$. Together with (5.2), this also implies that the generator equation (5.1) holds for a general polynomial $\Phi \in \Pi^{1,2}$.

As before, we separate $L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_s^{\mathrm{CS}})$ into $L_{\mathrm{r}}^{\geq \epsilon+s}\Phi(\mathcal{X}_s^{\mathrm{CS}})$ and $L_{\mathrm{r}}^{<\epsilon-s}\Phi(\mathcal{X}_s^{\mathrm{CS}})$, where the truncation parameter $\epsilon > 0$ is fixed and small. Equation (5.57) also holds with $t=0$ by the same argument as that for the continuity of $\mathbb{E}[L_{\mathrm{r}}^{\mathrm{CS}}\Phi(\mathcal{X}_t^{\mathrm{CS}})]$ at $t=0$, when $\Phi \in \Pi_{\mathrm{r}}^{1,2}$. Indeed, in both cases, only resampling between individuals with sufficiently large genealogical distance contribute to the generator action on Φ.

It remains to show that

$$\sup_{0\leq s\leq \epsilon/2} \mathbb{E}[|L_{\mathrm{r}}^{<\epsilon+s}\Phi(\mathcal{X}_s^{\mathrm{CS}})|] \underset{\epsilon\downarrow 0}{\longrightarrow} 0. \tag{5.59}$$

For $0 < s < \epsilon/2$, we can separate the resampling terms according to whether the genealogies of the two resampled individuals merge above or below time 0, and write

$$\begin{aligned}
\mathbb{E}[|L_{\mathrm{r}}^{<\epsilon+s}\Phi(\mathcal{X}_s^{\mathrm{CS}})|] &\leq C_{n,g,\varphi}\mathbb{E}\Big[\sum_{\substack{x\in(-M,M)\cap E_s\\ d_x\leq 2(\epsilon+s)}} d_x\Big] \\
&= C_{n,g,\varphi}\mathbb{E}\Big[\sum_{\substack{x\in(-M,M)\cap E_s\\ d_x\in(0,2s]}} d_x + \sum_{\substack{x\in(-M,M)\cap E_s\\ d_x\in(2s,2(\epsilon+s))}} d_x\Big] \\
&\leq C_{n,g,\varphi}M\sqrt{s} + C_{n,g,\varphi}\,\mathbb{E}\Big[\sum_{\substack{x\in(-M,M)\cap E_s\\ d_x\in(2s,2(\epsilon+s))}} d_x\Big].
\end{aligned} \tag{5.60}$$

It remains to bound the expectation on the r.h.s. above, which originate from resampling between individuals whose genealogical distance depend on the distance of their ancestors in $\mathcal{X}_0^{\mathrm{CS}}$ at time 0. Let $\hat{\xi}_0^{[-M,M]\times\{s\}}$ be the point set on $\mathbb{R}$ generated by the collection of paths in the dual Brownian web $\widehat{\mathcal{W}}$ starting from the space-time set $[-M,M]\times\{s\}$. Let us order the points in $\hat{\xi}_0^{[-M,M]\times\{s\}}$ and denote them by $z_1 < z_2 \cdots < z_{M_s+1}$, with a total of M_s+1 points. Then by our construction of $\mathcal{X}_s^{\mathrm{CS}}$, we have

$$\begin{aligned}
\mathbb{E}\Big[\sum_{\substack{x\in(-M,M)\cap E_s\\ d_x\in(2s,2(\epsilon+s))}} d_x\Big] &= \mathbb{E}\Big[\sum_{i=1}^{M_s} 1_{\{r_0^{\mathrm{CS}}(z_i,z_{i+1})<2\epsilon\}}(2s + r_0^{\mathrm{CS}}(z_i,z_{i+1}))\Big] \\
&\leq CM\sqrt{s} + \mathbb{E}\Big[\sum_{i=1}^{M_s} 1_{\{r_0^{\mathrm{CS}}(z_i,z_{i+1})<2\epsilon\}} r_0^{\mathrm{CS}}(z_i,z_{i+1})\Big],
\end{aligned} \tag{5.61}$$

where we used $\mathbb{E}[M_s] = \frac{2M}{\sqrt{\pi s}}$ because by duality between the Brownian web and its dual, M_s is exactly the number of points in the interval $[-M,M]$ occupied at time s by Brownian web paths starting from $\mathbb{R}$ at time 0, and hence Lemma B.5 can be applied. For each pair (z_i,z_{i+1}) with $r_0^{\mathrm{CS}}(z_i,z_{i+1}) < 2\epsilon$, by the properties of $\mathcal{X}_0^{\mathrm{CS}}$, there must exist a $y_i \in E_0$ with $z_i < y_i < z_{i+1}$ such that $r_0^{\mathrm{CS}}(z_i,z_{i+1}) = d_{y_i}$. Note that the y_i, $1\leq i\leq M_s$, are all distinct with $d_{y_i} < 2\epsilon$. We can now separate the contribution to the second term in (5.61) into two groups.

The first group consists of contributions from pairs (z_i,z_{i+1}) with $-2M \leq z_i < z_{i+1} \leq 2M$. The total contribution from these terms is uniformly dominated by $\sum_{\substack{x\in[-2M,2M]\cap E_0\\ d_x\leq 2\epsilon}} d_x$, which tends to 0 as $\epsilon\downarrow 0$ by the assumption $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$.

The second group consists of contributions from pairs (z_i,z_{i+1}) with either $z_i \leq -2M$ or $z_{i+1} \geq 2M$. We bound these terms by 2ϵ times the expected cardinality of such pairs. By the duality between $\mathcal{W}$ and $\widehat{\mathcal{W}}$, the cardinality of such pairs of (z_i,z_{i+1}) is bounded by the cardinality of $\xi_s^{\{x\in\mathbb{R}:|x|\geq 2M\}\times\{0\}} \cap (-M,M)$, the point set on $(-M,M)$ generated at

time s by paths in $\mathcal{W}$ starting from the space-time set$\{x \in \mathbb{R} : |x| \geq 2M\} \times \{0\}$. As shown in (5.47), the expected cardinality of $\xi_s^{\{x\in\mathbb{R}:|x|\geq 2M\}\times\{0\}} \cap (-M, M)$ is uniformly bounded in $s \in [0, l]$ for any $l > 0$. Therefore, uniformly for $s \in [0, \epsilon/2]$, the second term in (5.61) is also bounded by a function of ϵ, which tends to 0 as $\epsilon \downarrow 0$. This concludes the proof of (5.59) and Proposition 5.3. ∎

5.3 Proof of Theorem 1.40

We now prove Theorem 1.40. Note that the various path properties listed in Theorem 1.40 has been verified for the CSSM genealogy process $\mathcal{X}^{\mathrm{CS}}$ in Prop. 1.25 and Theorem 1.27, except for the claim that for each $t > 0$, $\mathcal{X}_t^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$ almost surely. To verify this last claim, we need to show that a.s.,

$$\sum_{x\in E_t\cap(-M,M)} r_t^{\mathrm{CS}}(x^+, x^-) < \infty \qquad \text{for each } M > 0. \tag{5.62}$$

where E_t is defined as in (5.4). Since for each $\epsilon > 0$, the set $\{x \in E_t \cap (-M, M) : r_t^{\mathrm{CS}}(x^+, x^-) \geq \epsilon\}$ is a finite set because $\mathcal{X}_t^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$, (5.62) is reduced to showing that a.s.,

$$\sum_{x\in E_t\cap(-M,M)} 1_{\{r_t^{\mathrm{CS}}(x^+,x^-)<\epsilon\}} r_t^{\mathrm{CS}}(x^+, x^-) < \infty \qquad \text{for each } M > 0.$$

The l.h.s. above has been shown in (5.55) to have finite expectation, and hence is finite a.s. Therefore $\mathcal{X}_t^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{rr}}^{\mathbb{R}}$ a.s.

To conclude the proof of Theorem 1.40, it only remains to verify the martingale property, namely that given $\Phi \in \Pi^{1,2}$ and $\mathcal{X}_0^{\mathrm{CS}} \in \mathbb{U}_1^{\mathbb{R}}$, for each $0 \leq s < t$, we have a.s.

$$\mathbb{E}\Big[\Phi(\mathcal{X}_t^{\mathrm{CS}}) - \Phi(\mathcal{X}_s^{\mathrm{CS}}) - \int_s^t (L^{\mathrm{CS}}\Phi)(\mathcal{X}_u^{\mathrm{CS}})\mathrm{d}u \,\Big|\, (\mathcal{X}_u^{\mathrm{CS}})_{0\leq u\leq s}\Big] = 0. \tag{5.63}$$

Consider first the case $s > 0$. Then a.s. $\mathcal{X}_s^{\mathrm{CS}} \in \mathbb{U}_{\mathrm{r}}^{\mathbb{R}}$. We can use the Markov property of $\mathcal{X}^{\mathrm{CS}}$ and apply the integral equation (5.2) for $\Phi \in \Pi^{1,2}$ and initial state $\mathcal{X}_s^{\mathrm{CS}}$, as established in Prop. 5.3. Equation (5.63) then follows provided that we can apply Fubini's Theorem to interchange the integral with expectation in

$$\mathbb{E}\Big[\int_s^t (L^{\mathrm{CS}}\Phi)(\mathcal{X}_u^{\mathrm{CS}})\mathrm{d}u\Big] = \int_s^t \mathbb{E}[(L^{\mathrm{CS}}\Phi)(\mathcal{X}_u^{\mathrm{CS}})]\mathrm{d}u. \tag{5.64}$$

As in the proof of Prop. 5.3, we can decompose L^{CS} as $L_{\mathrm{a}}^{\mathrm{CS}} + L_{\mathrm{d}}^{\mathrm{CS}} + L_{\mathrm{r}}^{\mathrm{CS}}$. The aging and diffusion part of the generator action on Φ can be uniformly bounded in time and in $\mathcal{X}_u^{\mathrm{CS}}$, while the resampling part can be bounded as in (5.52), which is shown to be integrable w.r.t. $E[\int_0^t \cdot]$ in (5.54). Therefore Fubini can be applied, and (5.63) holds for $s > 0$ a.s.

To treat the case $s = 0$, we take expectation in (5.63) for $s > 0$, i.e.,

$$\mathbb{E}\Big[\Phi(\mathcal{X}_t^{\mathrm{CS}}) - \Phi(\mathcal{X}_s^{\mathrm{CS}}) - \int_s^t (L^{\mathrm{CS}}\Phi)(\mathcal{X}_u^{\mathrm{CS}})\mathrm{d}u\Big] = 0,$$

and then let $s \downarrow 0$. By the a.s. continuity of $\mathcal{X}_s^{\mathrm{CS}}$ in $s \geq 0$, we have $\mathbb{E}[\Phi(\mathcal{X}_s^{\mathrm{CS}})] \to \Phi(\mathcal{X}_0^{\mathrm{CS}})$. The convergence of
$mathbbE\Big[\int_s^t (L^{\mathrm{CS}}\Phi)(\mathcal{X}_u^{\mathrm{CS}})\mathrm{d}u\Big]$ as $s \downarrow 0$ follows by dominated convergence, using the same argument as that for (5.64). Therefore (5.63) also holds for $s = 0$, which proves the desired martingale property. ∎

A Proofs on marked metric measure spaces

In this section, we prove Theorems 1.5 and 1.8, derive a relative compactness condition for subsets of $\mathbb{M}^V$, and a tightness criterion for laws on $\mathbb{M}^V$. Furthermore we formally treat pasting of trees. Our starting point for the first of these points is Remark 1.4, that $\mathbb{M}^V$ can be identified as a subspace of $(\mathbb{M}_f^V)^{\mathbb{N}}$, endowed with the product V-marked Gromov-weak topology.

Proof of Theorem 1.5. As noted in Remark 1.4, we can identify $\mathbb{M}^V$ as a subspace of $(\mathbb{M}_f^V)^{\mathbb{N}}$, endowed with the product V-marked Gromov-weak topology. Furthermore, under this identification, $\mathbb{M}^V$ is a closed subspace of $(\mathbb{M}_f^V)^{\mathbb{N}}$. It was shown in [DGP11, Theorem 2] that $\mathbb{M}_1^V$, the space of V-mmm spaces with probability measures, equipped with the V-marked Gromov-weak topology, is a Polish space. The same conclusion is easily seen to hold for $\mathbb{M}_f^V$. Therefore $(\mathbb{M}_f^V)^{\mathbb{N}}$ is also Polish, which implies that any closed subspace, including $\mathbb{M}^V$, is also Polish. ∎

Proof of Theorem 1.8. For each $k \in \mathbb{N} \cup \{0, \infty\}$, let

$$\widetilde{\Pi}^k := \cup_{n\in\mathbb{N}_0} \widetilde{\Pi}_n^k := \cup_{n\in\mathbb{N}_0} \big\{\Phi^{n,\phi} : \phi \in C^k(\mathbb{R}_+^{\binom{n}{2}} \times V^n, \mathbb{R})\big\},$$

where $C^k(\mathbb{R}_+^{\binom{n}{2}} \times V^n, \mathbb{R})$ is the space of bounded continuous real-valued functions on $\mathbb{R}_+^{\binom{n}{2}} \times V^n$ that are k times continuously differentiable in the first $\binom{n}{2}$ coordinates, and

$$\Phi^{n,\phi}(\overline{(X,r,\mu)}) := \int \cdots \int \phi(\underline{\underline{r}}, \underline{v})\mu^{\otimes n}(d(\underline{x}, \underline{v})) \qquad \text{for each } \overline{(X,r,\mu)} \in \mathbb{M}_1^V,$$

with $\underline{\underline{r}} := (r(x_i, x_j))_{1\le i<j\le n}$ and $\underline{v} := (v_1, \dots, v_n)$.

We use once more Remark 1.4 to now identify $\mathbb{M}^{(V,\le b)}$ as a subspace of $\prod_{k\in\mathbb{N}} \mathbb{M}_{\le c_k}^V$, endowed with the product V-marked Gromov-weak topology. Here $\mathbb{M}_{\le c}^V$ denotes the space of V-mmm spaces with a measure of total mass at most c, and e.g. $c_k := \|\psi_k\|_\infty b(r)$ for any $r > 0$ with $\mathrm{supp}(\psi_k) \subseteq B_r(o)$.

By [DGP11, Theorem 5], $\widetilde{\Pi}^k := \cup_n \widetilde{\Pi}_n^k$ is convergence determining in $\mathbb{M}_1^V$ and $\mathcal{M}_1(\mathbb{M}_1^V)$, and hence also in $\mathbb{M}_{\le c}^V$ and $\mathcal{M}_1(\mathbb{M}_{\le c}^V)$. We now want to argue that this holds as well for $\bigcup_n \Pi^k \subseteq \bigcup_n \widetilde{\Pi}^k$. This follows immediately from [EK86, Prop. 3.4.6] for measures on product spaces. ∎

The following relative compactness criterion for subsets of $\mathbb{M}^V$ follow easily from the identification of $\mathbb{M}^V$ as a subspace of $(\mathbb{M}_f^V)^{\mathbb{N}}$, and the relative compactness criterion formulated for subsets of $\mathbb{M}_1^{\{o\}}$ in [GPW13, Prop. 6.1], for subsets of $\mathbb{M}_1^V$ in [DGP11, Thm. 3] and for subsets in $\mathbb{M}^{\{o\}}$ [ALW16, Corollary 4.3].

Theorem A.1 (Relative compactness of subsets of $\mathbb{M}^V$)**.** *Let $\Gamma \subset \mathbb{M}^V$, and let o be any point in V. For each $k \in \mathbb{N}$, let $B_k(o)$ denote the open ball of radius k centered at o. Then Γ is relatively compact w.r.t. the V-marked Gromov-weak$^\#$ topology if for each $k \in \mathbb{N}$, $\{\overline{(X,r,1_{\{v\in B_k(o)\}}\mu(\mathrm{d}x\mathrm{d}v))} : \overline{(X,r,\mu)} \in \Gamma\}$ is a relatively compact subset of $\mathbb{M}_f^V$, i.e.,*

(i) *The family of finite measures on V,*

$$\Lambda_k := \{\mu(X \times \mathrm{d}v)1_{\{v\in B_k(o)\}} : \overline{(X,r,\mu)} \in \Gamma\},$$

is relatively compact w.r.t. the weak topology;

(ii) *For each $\epsilon > 0$, there exists $L > 0$ such that uniformly in $\overline{(X,r,\mu)} \in \Gamma$,*

$$\iint_{(X\times V)^2} 1_{\{r(x,y)>L\}} 1_{\{u,v\in B_k(o)\}} \mu(\mathrm{d}x\mathrm{d}u)\mu(\mathrm{d}y\mathrm{d}v) \le \epsilon;$$

(iii) *For each $\epsilon > 0$, there exists $M \in \mathbb{N}$ such that uniformly in $\overline{(X,r,\mu)} \in \Gamma$, we can find M balls of radius ϵ in X, say $B_{\varepsilon,1}, \dots, B_{\varepsilon,M} \subset X$ with $B = \cup_{i=1}^{M} B_{\varepsilon,i}$, such that $\mu((X \backslash B) \times B_k(o)) \leq \epsilon$.*

By the tightness criteria formulated for random variables with values in $\mathbb{M}_1^{\{o\}}$ in [GPW13, Theorem 3], with values in $\mathbb{M}_1^V$ in [DGP11, Theorem 4] and with values in $\mathbb{M}^{\{o\}}$ in [ALW16, Corollary 4.6], we obtain the following tightness criteria for $\mathbb{M}^V$-valued random variables.

Theorem A.2 (Tightness of $\mathbb{M}^V$-valued random variables)**.** *Let o be any point in V, and let $B_k(o)$ denote the open ball of radius k centered at o. A family of $\mathbb{M}^V$-valued random variables $\{\mathcal{X}_i = \overline{(X_i, r_i, \mu_i)}\}_{i\in I}$ is tight if for each $k \in \mathbb{N}$, $\{\overline{X_i, r_i, 1_{\{v\in B_k(o)\}}\mu_i(\mathrm{d}x\mathrm{d}v)}\}_{i\in I}$ is a tight family of $\mathbb{M}_f^V$-valued random variables, i.e.,*

(i) *$\{1_{\{v\in B_k(o)\}}\mu_i(X \times \mathrm{d}v)\}_{i\in I}$ is a tight family of random variables taking values in the space of finite measures on V (equipped with the weak topology);*

(ii) *$\{\overline{X_i, r_i, \mu_i(\mathrm{d}x \times B_k(o))}\}_{i\in I}$ is a tight family of random variables taking values in the space of metric measure spaces (equipped with the Gromov-weak topology).*

Using the characterization of relatively compact sets in $\mathbb{M}^V$ given in Theorem A.1, one can also formulate more concrete conditions for the tightness of a family of $\mathbb{M}^V$-valued random variables, using concrete conditions for the tightness of a family of random metric measure spaces formulated in [GPW09, Thm. 3].

B The Brownian web

In this section, we recall the construction and basic properties of the Brownian web. For a recent comprehensive survey, see [SSS15]. Recall from Subsection 1.4 the random variable $(\mathcal{W}, \widehat{\mathcal{W}})$ as constructed in [FINR04, FINR06], and in particular from (1.33)-(1.35), the state spaces of Π of $\mathcal{W}$ and $\widehat{\Pi}$ of $\widehat{\mathcal{W}}$. It has been shown [FINR04, Theorem 2.1] that the Brownian web $\mathcal{W}$ can be characterized as follows:

Theorem B.1 (Characterization of the Brownian web)**.** *The Brownian web $\mathcal{W}$ is a random closed subset of Π, whose law is uniquely determined by the following properties:*

(i) *For every $z \in \mathbb{R}^2$, almost surely $\mathcal{W}(z)$ contains a unique path.*

(ii) *For every finite n and deterministic points $z_1, \cdots, z_n \in \mathbb{R}^2$, $\{\mathcal{W}(z_i) : i = 1, \cdots, n\}$ are distributed as n coalescing Brownian motions starting from $z_1, \cdots, z_n$.*

(iii) *For every deterministic countable dense subset $\mathcal{D} \subset \mathbb{R}^2$, $\mathcal{W}$ is almost surely the closure of $\{\mathcal{W}(z); z \in \mathcal{D}\}$ in Π.*

The following result shows that every path in $\mathcal{W}$ can be approximated by the countable set of paths $\{\mathcal{W}(z); z \in \mathcal{D}\}$ in a very strong sense (see e.g. [SS08, Lemma 3.4]).

Lemma B.2 (Convergence of Paths in $\mathcal{W}$)**.** *Almost surely, if $(f_n)_{n\in\mathbb{N}}$ and f are paths in $\mathcal{W}$, starting respectively at times $(s_n)_{n\in\mathbb{N}}$ and s, and $f_n \to f$ in Π, then $\sup\{t : f_n(t) \neq f(t)\} \to s$ as $n \to \infty$.*

The dual Brownian web $\widehat{\mathcal{W}}$ can be characterized as follows [FINR06, Theorem 3.7]:

Theorem B.3 (Characterization of the dual Brownian web)**.** *Let $\mathcal{W}$ be a Brownian web. Then there exists an almost surely uniquely determined $\widehat{\Pi}$-valued random variable $\widehat{\mathcal{W}}$ defined on the same probability space as $\mathcal{W}$, called the dual Brownian web, such that:*

(i) *Almost surely, paths in $\mathcal{W}$ and $\widehat{\mathcal{W}}$ do not cross, i.e., there exist no $f \in \mathcal{W}$ and $\hat{f} \in \widehat{\mathcal{W}}$ and $s \neq t$ such that $(f(s) - \hat{f}(s))(f(t) - \hat{f}(t)) < 0$;*

(ii) *$\mathcal{R}\widehat{\mathcal{W}}$ has the same law as $\mathcal{W}$, where $\mathcal{R}$ denotes the reflection map that maps each $\hat{f} \in \widehat{\mathcal{W}}$ to an $f \in \Pi$ such that the graph of f in $\mathbb{R}^2$ is the reflection of the graph of $\hat{f}$ with respect to the origin.*

Below we collect some basic properties of the Brownain web which we will use. For further details, see [FINR04, FINR06, SS08]. The first property concerns the configuration of paths in $\mathcal{W}$ and $\widehat{\mathcal{W}}$ entering and leaving a point $z \in \mathbb{R}^2$. For each $z = (x,t) \in \mathbb{R}^2$, let $m_{\text{out}}(z)$ denote the cardinality of $\mathcal{W}(z)$. We will let $m_{\text{in}}(z)$ denote the number of equivalence classes of paths in $\mathcal{W}$ entering z, where a path $f \in \mathcal{W}$ is said to enter z if it starts before time t and $f(t) = x$, while two paths $f, g \in \mathcal{W}$ entering z are called equivalent if they coalesce before time t. Note that $m_{\text{in}}(z) = 2$ if and only if z is a point of coalescence between two paths in $\mathcal{W}$. Similarly, we can define $\hat{m}_{\text{in}}(z)$ and $\hat{m}_{\text{out}}(z)$, based on the configuration of paths in $\widehat{\mathcal{W}}$. The pair $(m_{\text{in}}, m_{\text{out}})$ is called the type of z in $\mathcal{W}$.

We cite the following result from [TW98, Prop. 2.4] and [FINR06, Thm. 3.11–3.14].

Lemma B.4 (Special points for the Brownian web)**.** *Let $\mathcal{W}$ and $\widehat{\mathcal{W}}$ be a Brownian web and its dual. Almost surely:*

(1) *The set of $z \in \mathbb{R}^2$ with $(m_{\text{in}}(z), m_{\text{out}}(z)) = (\hat{m}_{\text{in}}(z), \hat{m}_{\text{out}}(z)) = (0,1)$ has full Lebesgue measure on $\mathbb{R}^2$.*

(2) *For each $t \in \mathbb{R}$, the set of $z = (x',t') \in \mathbb{R} \times \{t\}$ with $\hat{m}_{\text{out}}(z) \geq 2$ is a countable set, with $m_{\text{in}}(z) \geq 1$ for each such z, i.e., z lies on the graph of some path in $\mathcal{W}$ starting before time t.*

Next we cite a result on the decay of the density of coalescing paths started at time 0.

Lemma B.5 (Density for the Brownian web)**.** *For $t > 0$, let $\xi_t^{\mathbb{R}\times\{0\}} := \{f(t) : f \in \cup_{x\in\mathbb{R}} \mathcal{W}(x,0)\}$ denote the point set on $\mathbb{R}$ generated at time t by the collection of coalescing paths in the Brownian web $\mathcal{W}$ started at time 0. Then for any $a < b$,*

$$\mathbb{E}[|\xi_t^{\mathbb{R}\times\{0\}} \cap [a,b]|] = \frac{b-a}{\sqrt{\pi t}}. \tag{B.1}$$

This result can be easily derived by using the duality between $\mathcal{W}$ and $\widehat{\mathcal{W}}$, namely that $\xi_t^{\mathbb{R}\times\{0\}} \cap (x, x+\epsilon) \neq \emptyset$ if and only if the two paths in $\widehat{\mathcal{W}}$ starting from (x,t) and $(x+\epsilon, t)$ do not collide on the time interval $[0,t]$. See e.g. [SS08, Prop. 1.12], where such a density calculation is carried out for a generalization of the Brownian web known as the *Brownian net*, which in addition allows branching of paths.

C Correlation inequalities for coalescing Brownian motions

In this section, we prove some negative correlation inequalities for a collection of coalescing Brownian motions, which are used in Section 5. Similar inequalities have previously been established in [MRTZ06], see also [NRS05, Remark 7.5]. Here we deduce more general negative correlation inequalities from *Reimer's inequality* applied to coalescing random walks.

In van den Berg and Kesten [vdBK02], Reimer's inequality was applied to continuous time coalescing random walks with a generalized coalescing rule. Since we are interested in coalescing Brownian motions, discrete space-time coalescing random walks with instantaneous coalescing already provide an adequate approximation, and Reimer's inequality can be applied without any complication to the latter.

First we recall Reimer's inequality [Rei00]. For each $i \in I := \{1, \cdots, n\}$, let S_i be a finite set with a probability measure μ_i on S_i. Let $\Omega = S_1 \times S_2 \cdots \times S_n$ and $\mu = \mu_1 \times \cdots \times \mu_n$. For $K \subset I$ and $\omega = (\omega_i)_{i\in I}$, define the cylinder set $C(K,\omega) := \{\omega' \in \Omega : \omega'_i = \omega_i \ \forall\ i \in K\}$.

Given two events $A, B \subset \Omega$, we say *A and B occur disjointly* for a configuration $\omega \in \Omega$ if there exists $K \subset I$ such that $C(K,\omega) \subset A$ and $C(I\backslash K,\omega) \subset B$. The set of $\omega \in \Omega$ for which A and B occur disjointly, which we call the disjoint intersection of A and B, is denoted by

$$A\square B := \{\omega \in \Omega : \exists\, K \subset I \text{ s.t. } C(K,\omega) \subset A \text{ and } C(I\backslash K,\omega) \subset B\}.$$

Then Reimer's inequality asserts that, for any two events $A, B \subset \Omega$,

$$\mu(A\square B) \leq \mu(A)\mu(B). \tag{C.1}$$

Now we apply this inequality to coalescing random walks. We recall first the construction of discrete space-time coalescing random walks. Let $\mathbb{Z}^2_{\text{even}} = \{(x,t) \in \mathbb{Z}^2 : x+t \text{ is even}\}$. Let $\{\omega_z\}_{z\in\mathbb{Z}^2_{\text{even}}}$ be i.i.d. random variables taking values in $\{\pm 1\}$. A directed edge is drawn from each $z = (x,t) \in \mathbb{Z}^2_{\text{even}}$, which ends at $(x+1, t+1)$ if $\omega_z = 1$, and ends at $(x-1, t+1)$ if $\omega_z = -1$. This provides a graphical construction of a collection of coalescing random walks, where the random walk path starting from each $z \in \mathbb{Z}^2_{\text{even}}$ is constructed by following the directed edges in $\mathbb{Z}^2_{\text{even}}$ drawn according to ω.

To illustrate how Reimer's inequality is applied, let $(X_j^{(x_i,t_i)})_{j\geq t_i}$, $1 \leq i \leq n$, be a collection of coalescing random walks constructed as above with starting points $(x_i, t_i) \in \mathbb{Z}^2_{\text{even}}$, and assume for simplicity $t_i \leq 0$ for all $1 \leq i \leq n$. For $t \in \mathbb{N}$, let $\xi_t = \{x \in \mathbb{Z} : x = X_t^{(x_i,t_i)} \text{ for some } 1 \leq i \leq n\}$. Let $O_1, \cdots, O_k$ be disjoint subsets of $\mathbb{Z}$, and let $A_i = \{\omega : \xi_t \cap O_i \neq \emptyset\}$. The crucial observation is that, if the events $(A_i)_{1\leq i\leq n}$ occur simultaneously, then they must occur disjointly w.r.t. $(\omega_z)_{z\in\mathbb{Z}^2_{\text{even}}}$ because of the coalescence. Namely,

$$\bigcap_{i=1}^{k} A_i = A_1\square A_2 \cdots \square A_k. \tag{C.2}$$

Reimer's inequality (C.1) then gives the negative correlation inequality

$$\mathbb{P}\big(\bigcap_{i=1}^{k} A_i\big) \leq \prod_{i=1}^{k} \mathbb{P}(A_i). \tag{C.3}$$

The same reasoning allows us to choose each A_i to be an increasing event of the occupation configuration $\xi_t \cap O_i$, i.e., given ω and ω' with respective occupation configurations $\xi_t \cap O_i \subset \xi'_t \cap O_i$, if $\omega \in A_i$, then also $\omega' \in A_i$. Note that Reimer's inequality may even be applied to the disjoint occurence of the event A_1 with itself, which we use later in the proof of Lemma C.6.

Using Reimer's inequality as illustrated above, together with the invariance principle for coalescing random walks, we will deduce a host of negative correlation inequalities for coalescing Brownian motions, which we formulate next.

Definition C.1 (Negatively correlated point processes)**.** *We say a simple point process ξ on $\mathbb{R}$ is negatively correlated, if for any $n \in \mathbb{N}$ and any disjoint open intervals $O_1, \cdots, O_n$, we have $\mathbb{P}(\cap_{i=1}^n \{\xi \cap O_i \neq \emptyset\}) \leq \prod_{i=1}^n \mathbb{P}(\xi \cap O_i \neq \emptyset)$.*

Lemma C.2 (Negative correlation for colaescing Brownian motions)**.** *Let $A \subset \mathbb{R}\times(-\infty, 0]$, and let ξ_t^A denote the point set on $\mathbb{R}$ generated at time $t > 0$ by the collection of coalescing Brownian motions in the Brownian web $\mathcal{W}$ starting from A. Then ξ_t^A is negatively correlated.*

Proof. By monotone convergence, it suffices to consider the case when A consists of a finite number of points $\{z_1, \cdots, z_k\}$. The fact that ξ_t^A is negatively correlated then follows directly from the negative correlation inequality (C.3) for coalescing random

walks, and the distributional convergence of coalescing random walks to coalescing Brownian motions in the local uniform topology (see e.g. [NRS05, Section 5]). ∎

Lemma C.3 (Decoupling of correlation functions)**.** *Let ξ_t^A be as in Lemma C.2. Let $a_1 < b_1 \le a_2 < b_2 \cdots \le a_n < b_n$. Then for any $1 \le j \le n-1$,*

$$\begin{aligned}&\mathbb{P}\Big(\bigcap_{i=1}^{n}\{\xi_t^A \cap (a_i,b_i) \neq \emptyset\} \cap \{\xi_t^A \cap (b_j, a_{j+1}) = \emptyset\}\Big)\\ &\le \mathbb{P}\Big(\xi_t^A \cap (a_j,b_j) \neq \emptyset, \xi_t^A \cap (b_j,a_{j+1}) = \emptyset, \xi_t^A \cap (a_{j+1},b_{j+1}) \neq \emptyset\Big) \prod_{\substack{1\le i\le n\\ i\neq j,j+1}} \mathbb{P}(\xi_t^A \cap (a_i,b_i) \neq \emptyset).\end{aligned} \tag{C.4}$$

Proof. As before, this follows from approximation by discrete space-time coalescing random walks and Reimer's inequality. Note that for the discrete analogue of the events in the second line of (C.4), if they all occur, then they must occur disjointly. ∎

Lemma C.4 (Negative correlation for occupation number)**.** *Let ξ_t^A be as in Lemma C.2, and let $B \subset \mathbb{R}$ have finite Lebesgue measure. Then for any $k \in \mathbb{N}$,*

$$\mathbb{P}(|\xi_t^A \cap B| \ge k) \le \mathbb{P}(|\xi_t^A \cap B| \ge 1)^k. \tag{C.5}$$

Proof. By monotone convergence and approximation by open sets, it suffices to consider the case when $A \subset \mathbb{R} \times (-\infty, 0]$ consists of a finite number of points, and B is the finite union of disjoint open intervals. In fact, the argument is the same if B is a bounded open interval, say $(0,1)$.

We proceed by discrete approximation. Let $(Z_t)_{t\in\mathbb{N}}$ be the subset of $\mathbb{Z}$ occupied at time $t \in \mathbb{N}$ by a collection of coalescing random walks on $\mathbb{Z}^2_{\text{even}}$ starting from $z_1 = (x_1,t_1), \cdots, z_n = (x_n,t_n) \in \mathbb{Z}^2_{\text{even}}$ with $t_i \le 0$ for all $1 \le i \le n$. Given $O \subset \mathbb{Z}$ and for $k \in \mathbb{N}$, let $A_k = \{\omega \in \Omega : |Z_t \cap O| \ge k\}$, where $\omega = (\omega_z)_{z\in\mathbb{Z}^2_{\text{even}}}$ are the i.i.d. $\{\pm 1\}$-valued random variables underlying the graphical construction of the coalescing random walks. We note that

$$A_k \subset \overbrace{A_1 \square \cdots \square A_1}^{k}. \tag{C.6}$$

Indeed, if A_k occurs, then we can find k disjoint random walk paths, each of which occupies a distinct site in O at time t. Reimer's inequality (C.1) then implies $\mathbb{P}(A_k) \le \mathbb{P}(A_1)^k$. Inequality (C.5) then follows by the distributional convergence of coalescing random walks to coalescing Brownian motions in the local uniform topology. ∎

Lemma C.5 (Moment bounds for occupation number)**.** *Let ξ be a simple point process on $\mathbb{R}$ with a locally finite intensity measure μ, which is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}$. If ξ is negatively correlated, then for any Lebesgue measurable $B \subset \mathbb{R}$ with $\mu(B) < \infty$, and for any $k \in \mathbb{N}$, we have*

$$\mathbb{E}\Big[|\xi \cap B|^k\Big] \le \sum_{m=1}^{k} \binom{k}{m} m^{k-m} \mu(B)^m. \tag{C.7}$$

Proof. Let B be an open interval, say $(0,1)$. For $n \in \mathbb{N}$, let $D_n = \{i2^{-n} : 0 \le i \le 2^n\}$, and let $D = \bigcup_{n\in\mathbb{N}} D_n$. By our assumption that the intensity measure μ is absolutely continuous w.r.t. Lebesgue measure, $\xi \cap D = \emptyset$ almost surely. For $1 \le i \le 2^n$, let $I_i^{(n)}(\xi) = 1_{\{\xi\cap((i-1)2^{-n}, i2^{-n}) \neq \emptyset\}}$. By the assumption that ξ is a simple point process, and by monotone convergence,

$$\mathbb{E}\big[|\xi\cap(0,1)|^k\big] = \mathbb{E}\Big[\lim_{n\to\infty}\big(\sum_{1\le i\le 2^n} I_i^{(n)}\big)^k\Big] = \lim_{n\to\infty}\mathbb{E}\Big[\big(\sum_{1\le i\le 2^n} I_i^{(n)}\big)^k\Big] = \lim_{n\to\infty}\sum_{1\le i_1,\cdots,i_k\le 2^n}\mathbb{E}\Big[\prod_{j=1}^{k} I_{i_j}^{(n)}\Big]. \tag{C.8}$$

Note that $\mu(0,1) = \mathbb{E}[|\xi \cap (0,1)|] = \lim_{n\to\infty} \sum_{j=1}^{2^n} \mathbb{E}[I_j^{(n)}]$. Given $\vec{i} := (i_1, \cdots, i_k) \in \{1, \cdots, 2^n\}^k$, let $m(\vec{i}) = |\{i_1, \cdots, i_k\}|$, and denote the $m(\vec{i})$ distinct indices in $\{i_1, \cdots, i_k\}$ by $\sigma_1(\vec{i}) < \cdots < \sigma_m(\vec{i})$. Then by the negative correlation assumption,

$$\begin{aligned}
&\sum_{1\le i_1,\cdots,i_k\le 2^n} \mathbb{E}\Big[\prod_{j=1}^{k} I_{i_j}^{(n)}\Big] \\
&\le \sum_{1\le i_1,\cdots,i_k\le 2^n} \prod_{j=1}^{m(\vec{i})} \mathbb{E}[I_{\sigma_j(\vec{i})}^{(n)}] = \sum_{m=1}^{k} \sum_{1\le\sigma_1<\cdots<\sigma_m\le 2^n} f(k,m)m! \prod_{j=1}^{m} \mathbb{E}[I_j^{(n)}] \qquad \text{(C.9)} \\
&\le \sum_{m=1}^{k} f(k,m)\Big(\sum_{j=1}^{2^n} \mathbb{E}[I_j^{(n)}]\Big)^m \underset{n\to\infty}{\longrightarrow} \sum_{m=1}^{k} f(k,m)\mu(0,1)^m,
\end{aligned}$$

where $f(k,m)m! = |\{(i_1, \cdots, i_k) \in \{1, \cdots, 2^n\}^k : \{i_1, \cdots, i_k\} = \{\sigma_1, \cdots, \sigma_m\}\}|$, which is easily seen to be independent of the choice of $1 \le \sigma_1 < \cdots < \sigma_m \le 2^n$. By first picking m indices out of $\{i_1, \cdots, i_k\}$ and assign them values $\sigma_1 < \cdots < \sigma_m$ respectively, we easily verify that $f(k,m) \le \binom{k}{m} m^{k-m}$, which proves (C.7) for $B = (0,1)$. By the same argument, (C.7) holds for finite unions of disjoint open intervals, and by monotone convergence, for open sets as well. Since any Lebesgue measurable set can be approximated from outside by open sets, again by monotone convergence and the fact $\xi \cap E = \emptyset$ a.s. for a given $E \subset \mathbb{R}$ with zero Lebesgue measure, (C.7) also holds for any Lebesgue measurable B. ∎

We also need the following estimate on the constrained two-point correlation function for the Brownian web.

Lemma C.6 (Constrained two point function for the Brownian web)**.** *Let $\xi_t^{\mathbb{R}\times\{0\}}$ be as in Lemma B.5. Let $t > 0$. For $a < b$, let $I_{[a,b]}$ denote the event that $\xi^{\mathbb{R}\times\{0\}} \cap [a,b] \neq \emptyset$. Then for any $x_1 < x_2$ with $\Delta := x_2 - x_1$, we have*

$$K_t^c(x_1,x_2) := K_t^c(\Delta) = \lim_{\delta\downarrow 0} \frac{1}{\delta^2} \mathbb{P}\big[I_{[x_1,x_1+\delta]} \cap I^c_{[x_1+\delta,x_2]} \cap I_{[x_2,x_2+\delta]}\big] = \frac{\Delta e^{-\frac{\Delta^2}{4t}}}{2\sqrt{\pi}t^{\frac{3}{2}}}. \qquad \square \text{ (C.10)}$$

Proof. By translation invariance of $\xi^{\mathbb{R}\times\{0\}}$, we may assume $x_1 = 0$, and let $x_2 - x_1 = \Delta > 0$. Let $\hat{w}_{(0,t)}$, $\hat{w}_{(\delta,t)}$, $\hat{w}_{(\Delta,t)}$ and $\hat{w}_{(\Delta+\delta,t)}$ be the dual coalescing Brownian motions in $\widehat{\mathcal{W}}$ starting at respectively $(0,t)$, (δ,t), (Δ,t) and $(\Delta+\delta,t)$. Let $\hat{\tau}_{a,b}$ be the time of coalescence between $\hat{w}_{(a,t)}$ and $\hat{w}_{(b,t)}$. Then by the duality between $\mathcal{W}$ and $\widehat{\mathcal{W}}$, almost surely, the event $I_{[0,\delta]} \cap I_{[\delta,\Delta]}$ occurs if and only if $\hat{\tau}_{0,\delta} < 0$ and $\hat{\tau}_{\delta,\Delta} < 0$, and the event $I_{[0,\delta]} \cap I_{[\delta,\Delta+\delta]}$ occurs if and only if $\hat{\tau}_{0,\delta} < 0$ and $\hat{\tau}_{\delta,\Delta+\delta} < 0$. Since $I_{[\delta,\Delta]} \subset I_{[\delta,\Delta+\delta]}$ and $I_{[\delta,\Delta+\delta]} \backslash I_{[\delta,\Delta]} = I^c_{[\delta,\Delta]} \cap I_{(\Delta,\Delta+\delta]}$, we have

$$\begin{aligned}
&\mathbb{P}\big[I_{[x_1,x_1+\delta]} \cap I^c_{[x_1+\delta,x_2]} \cap I_{[x_2,x_2+\delta]}\big] \\
&= \mathbb{P}^{B_1,B_2,B_3}_{0,\delta,\Delta}(\tau_{0,\delta} > t, \tau_{\delta,\Delta+\delta} > t) - \mathbb{P}^{B_1,B_2,B_3}_{0,\delta,\Delta+\delta}(\tau_{0,\delta} > t, \tau_{\delta,\Delta} > t),
\end{aligned} \qquad \text{(C.11)}$$

where we have reversed time and replaced $\hat{w}_{(a,t)}$ for $a = 0, \delta, \Delta, \Delta+\delta$ by independent standard Brownian motions B_1, B_2 and B_3 starting from 0, δ, and Δ (or $\Delta+\delta$), and $\tau_{a,b}$ is the first hitting time between the two Brownian motions starting from a and b.

By the Karlin-McGregor formula [KM59], the transition density for three Brownian motions B_1, B_2 and B_3 starting from $x_1 < x_2 < x_3$ to end at locations $y_1 < y_2 < y_3$ at time t without ever intersecting along the way is given by the determinant $\mathrm{Det}(p_t(x_i,y_j)_{1\le i,j\le 3})$, where $p_t(x,y) = \frac{e^{-\frac{(y-x)^2}{2t}}}{\sqrt{2\pi t}}$. Therefore we define

$$D := \mathbb{P}^{B_1,B_2,B_3}_{0,\delta,\Delta}(\tau_{0,\delta} > t, \tau_{\delta,\Delta+\delta} > t) - \mathbb{P}^{B_1,B_2,B_3}_{0,\delta,\Delta+\delta}(\tau_{0,\delta} > t, \tau_{\delta,\Delta} > t). \qquad \text{(C.12)}$$

The r.h.s. can be written as

$$\iiint\limits_{y_1<y_2<y_3} \begin{vmatrix} p_t(0,y_1) & p_t(0,y_2) & p_t(0,y_3) \\ p_t(\delta,y_1) & p_t(\delta,y_2) & p_t(\delta,y_3) \\ p_t(\Delta+\delta,y_1) & p_t(\Delta+\delta,y_2) & p_t(\Delta+\delta,y_3) \end{vmatrix} - \begin{vmatrix} p_t(0,y_1) & p_t(0,y_2) & p_t(0,y_3) \\ p_t(\delta,y_1) & p_t(\delta,y_2) & p_t(\delta,y_3) \\ p_t(\Delta,y_1) & p_t(\Delta,y_2) & p_t(\Delta+,y_3) \end{vmatrix} d\vec{y}. \tag{C.13}$$

After some elementary manipulation, we obtain that D is given by the expression:

$$D = \iiint\limits_{y_1<y_2<y_3} \frac{e^{-\frac{y_1^2+y_2^2+y_3^2+\delta^2+\Delta^2}{2t}}}{(2\pi t)^{\frac{3}{2}}} \begin{vmatrix} 1 & 1 & 1 \\ e^{\frac{\delta y_1}{t}}-1 & e^{\frac{\delta y_2}{t}}-1 & e^{\frac{\delta y_3}{t}}-1 \\ e^{\frac{\Delta y_1}{t}} f(t,y_1,\delta,\Delta) & e^{\frac{\Delta y_2}{t}} f(t,y_2,\delta,\Delta) & e^{\frac{\Delta y_3}{t}} f(t,y_3,\delta,\Delta) \end{vmatrix} d\vec{y}, \tag{C.14}$$

where $f(t,y_i,\delta,\Delta) = e^{\frac{2(y_i-\Delta)\delta-\delta^2}{2t}} - 1$. We then Taylor expand in δ, and note that the factor $e^{-\frac{\|y\|^2}{2t}}$ allows us to take $D\delta^{-2}$ and pass the limit $\delta \downarrow 0$ inside the integral to obtain

$$K_t^c(\Delta) = \lim_{\delta\downarrow 0} \frac{D}{\delta^2} = \iiint\limits_{y_1<y_2<y_3} \frac{e^{-\frac{y_1^2+y_2^2+y_3^2+\Delta^2}{2t}}}{(2\pi)^{\frac{3}{2}} t^{\frac{7}{2}}} \begin{vmatrix} 1 & 1 & 1 \\ y_1 & y_2 & y_3 \\ e^{\frac{\Delta y_1}{t}}(y_1-\Delta) & e^{\frac{\Delta y_2}{t}}(y_2-\Delta) & e^{\frac{\Delta y_3}{t}}(y_3-\Delta) \end{vmatrix} d\vec{y}. \tag{C.15}$$

Expanding the determinant and performing the change of variable $y_i = x_i\sqrt{t}$ and $\Delta = \bar{\Delta}\sqrt{t}$ gives $K_t^c = \left((2\pi)^{\frac{3}{2}}t\right)^{-1}(I_1+I_2+I_3)$, where

$$\begin{aligned} I_1 &= \iiint\limits_{x_1<x_2<x_3} e^{-\frac{(x_1-\bar{\Delta})^2+x_2^2+x_3^2}{2}}(x_1-\bar{\Delta})(x_3-x_2)d\vec{x}, \\ I_2 &= \iiint\limits_{x_1<x_2<x_3} e^{-\frac{x_1^2+(x_2-\bar{\Delta})^2+x_3^2}{2}}(x_2-\bar{\Delta})(x_1-x_3)d\vec{x}, \\ I_3 &= \iiint\limits_{x_1<x_2<x_3} e^{-\frac{x_1^2+x_2^2+(x_3-\bar{\Delta})^2}{2}}(x_3-\bar{\Delta})(x_2-x_1)d\vec{x}. \end{aligned} \tag{C.16}$$

We then have

$$\begin{aligned} I_1+I_3 &= -\iint\limits_{x_2<x_3} e^{-\frac{x_2^2+(x_2-\bar{\Delta})^2+x_3^2}{2}}(x_3-x_2)dx_2dx_3 + \iint\limits_{x_1<x_2} e^{-\frac{x_1^2+x_2^2+(x_2-\bar{\Delta})^2}{2}}(x_2-x_1)dx_1dx_2 \\ &= \iint\limits_{\mathbb{R}^2} e^{-\frac{x_1^2+x_2^2+(x_2-\bar{\Delta})^2}{2}}(x_2-x_1)dx_1dx_2 = \frac{\pi}{\sqrt{2}}\bar{\Delta}e^{-\frac{\bar{\Delta}^2}{4}}, \end{aligned} \tag{C.17}$$

and

$$\begin{aligned} I_2 &= \iint\limits_{x_1<x_3} e^{-\frac{x_1^2+x_3^2}{2}}\left(e^{-\frac{(x_1-\bar{\Delta})^2}{2}} - e^{-\frac{(x_3-\bar{\Delta})^2}{2}}\right)(x_1-x_3)dx_1dx_3 \\ &= \iint\limits_{\mathbb{R}^2} e^{-\frac{x_1^2+(x_1-\bar{\Delta})^2+x_3^2}{2}}(x_1-x_3)dx_1dx_3 = \frac{\pi}{\sqrt{2}}\bar{\Delta}e^{-\frac{\bar{\Delta}^2}{4}}. \end{aligned} \tag{C.18}$$

Together, they give $K_t^c = \frac{\Delta e^{-\frac{\Delta^2}{4t}}}{2\sqrt{\pi}t^{\frac{3}{2}}}$. ∎

Acknowledgments. Many thanks to the referees who read the manuscript very thoroughly, and thereby helped to improve the presentation. Rongfeng Sun is supported by AcRF Tier 1 grant R-146-000-185-112. Andreas Greven was supported by the Deutsche Forschungsgemeinschaft DFG, Grant GR 876/15-1 und 15-2. Anita Winter is supported by DFG SPP 1590.

References

[Ald93] D. Aldous. The continuum random tree III. *Ann. Probab.*, 21:248–289, 1993. MR-1207226

[ALW16] S. Athreya, W. Löhr, and A. Winter. The gap between Gromov-vague and Gromov-Hausdorff-vague topology. *Stoch Processes Appl.*, 126(9):2527–2553, 2016. MR-3522292

[Arr79] R. Arratia. *Coalescing Brownian motions on the line*. PhD thesis, University of Wisconsin, 1979.

[Arr81] R. Arratia. Coalescing Brownian motions and the voter model on $\mathbb{Z}$. Unpublished partial manuscript, 1981.

[AS11] S. Athreya and R. Sun. One-dimensional Voter model interface revisited. *Electron. Commun. Probab.*, 16:792–800, 2011. MR-2868600

[Bil89] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 2nd edition, 1989. MR-0233396

[CG86] J.T. Cox and D. Griffeath. Diffusive clustering in the two dimensional voter model. *Ann. Probab.*, 14:347–370, 1986. MR-0832014

[Daw77] D. A. Dawson. The critical measure diffusion process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 40:125–145, 1977. MR-0478374

[Daw93] D.A. Dawson. *Measure-valued Markov processes*, volume 1541. Springer-Verlag, 1993. MR-1242575

[DEF+00] P. Donnelly, S.N. Evans, K. Fleischmann, T.G. Kurtz, and X. Zhou. Continuum-sites stepping-stone models, coalescing exchangeable partitions and random trees. *Ann. Probab.*, 28:1063–1110, 2000. MR-1797304

[DEF+02a] D. A. Dawson, A.M. Etheridge, K. Fleischmann, E.A. Perkins L. Mytnik, and J. Xiong. Mutually catalytic branching in the plane: Infinite measure states. *Electron. Commun. Probab.*, 7(15):1–61, 2002. MR-1921744

[DEF+02b] D. A. Dawson, A.M. Etheridge, K. Fleischmann, L. Mytnik, E.A. Perkins, and J. Xiong. Mutually catalytic branching in the plane: Finite measure states. *Ann. of Probab.*, 30:1681–1762, 2002. MR-1944004

[DGP11] A. Depperschmidt, A. Greven, and P. Pfaffelhuber. Marked metric measure spaces. *Electron. Commun. Probab.*, 16:174–188, 2011. MR-2783338

[DGP12] A. Depperschmidt, A. Greven, and P. Pfaffelhuber. Tree-valued Fleming-Viot dynamics with mutation and selection. *Ann. Appl. Probab.*, 22:2560–2615, 2012. MR-3024977

[DGV95] D.A. Dawson, A. Greven, and J. Vaillancourt. Equilibria and quasi-equilibria for infinite systems of Fleming-Viot processes. *Trans. Amer. Math. Soc.*, 347:2277–2360, 1995. MR-1297523

[DP91] D.A. Dawson and E. Perkins. Historical processes. *Mem. Amer. Math. Soc.*, 93(454), 1991. MR-1079034

[DP98] D. A. Dawson and E. Perkins. Longtime behavior and coexistence in a mutually catalytic branching model. *Ann. Probab.*, 26(3):1088–1138, 1998. MR-1634416

[DVJ03] D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I. Elementary theory and methods*. Springer, New York, 2nd edition, 2003. MR-1950431

[EF96] S. Evans and K. Fleischmann. Cluster formation in a stepping stone model with continuous, hierarchically structured sites. *Ann. Probab.*, 24:1926–1952, 1996. MR-1415234

[EK86] S.N. Ethier and T. Kurtz. *Markov processes: characterization and convergence*. John Wiley, New York, 1986. MR-0838085

[EPW06] S.N. Evans, J. Pitman, and A. Winter. Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Relat. Fields*, 134:81–126, 2006. MR-2221786

[Eva97] S. Evans. Coalescing markov labelled partitions and a continuous sites genetics model with infinitely many types. *Ann. Inst. H. Poincaré Probab. Statist.*, 33:339–358, 1997. MR-1457055

[FG94] K. Fleischmann and A. Greven. Diffusive clustering in an infinite system of hierarchically interacting Fisher-Wright diffusions. *Probab. Theory Related Fields*, 98:517–566, 1994. MR-1271108

[FG96] K. Fleischmann and A. Greven. Time–space analysis of the cluster formation in interacting diffusions. *Electron. J. Probab.*, 1:1–46, 1996. MR-1386298

[FINR04] L.R.G. Fontes, M. Isopi, C.M. Newman, and K. Ravishankar. The Brownian web: characterization and convergence. *Ann. Probab.*, 32:2857–2883, 2004. MR-2094432

[FINR06] L.R.G. Fontes, M. Isopi, C.M. Newman, and K. Ravishankar. Coarsening, nucleation, and the marked Brownian web. *Ann. Inst. H. Poincaré Probab. Stat.*, 42:37–60, 2006. MR-2196970

[GJ98] J.F. Le Gall and Y. Le Jan. Branching processes in Lévy processes: Laplace functionals of snakes and superprocesses. *Ann. Probab.*, 26:1407–1432, 1998. MR-1675019

[GKW] A. Greven, A. Klimovsky, and A. Winter. Evolving genealogies of spatial Λ-Fleming-Viot processes with mutation. in preparation.

[GLW05] A. Greven, V. Limic, and A. Winter. Representation theorems for interacting Moran models, interacting Fisher–Wright diffusions and applications. *Electron. J. Probab.*, 10:1286–1358, 2005. MR-2176385

[GM] A. Greven and C. Mukherjee. Genealogies of spatial logistic branching models. in preparation.

[GPW09] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (the Lambda-coalescent measure tree). *Probab. Theory Related Fields*, 145:285–322, 2009. MR-2520129

[GPW13] A. Greven, P. Pfaffelhuber, and A. Winter. Tree-valued resampling dynamics: martingale problems and applications. *Probab. Theory Related Fields*, 155(3–4):789–838, 2013. MR-3034793

[KL15] S. Kliem and W. Löhr. Existence of mark functions in marked metric measure spaces. *Electronic Journal of Probability*, 20(72):1–24, 2015. MR-3371432

[Kle00] A. Klenke. Diffusive clustering of interacting Brownian motions on $\mathbb{Z}^2$. *Stochastic Process. Appl.*, 89:261–268, 2000. MR-1780290

[KM59] S. Karlin and J. McGregor. Coincidence probabilities. *Pacific J. Math.*, 9:1141–1164, 1959. MR-0114248

[KS88] N. Konno and T. Shiga. Stochastic differential equations for some measure-valued diffusions. *Probab. Theory Related Fields*, 79:201–225, 1988. MR-0958288

[Lig85] T.M. Liggett. *Interacting Particle Systems*. Springer, New York, 1985. Reprint of the 1985 original. Classics in Mathematics. Springer-Verlag, Berlin, 2005. MR-2108619

[MRTZ06] R. Munasinghe, R. Rajesh, R. Tribe, and O. Zaboronski. Multi-scaling of the n-point density function for coalescing Brownian motions. *Commun. Math. Phys.*, 268:717–725, 2006. MR-2259212

[NRS05] C. M. Newman, K. Ravishankar, and R. Sun. Convergence of coalescing nonsimple random walks to the Brownian web. *Electron. J. Probab.*, 10:21–60, 2005. MR-2120239

[Rei00] D. Reimer. Proof of the van den Berg-Kesten conjecture. *Combin. Probab. Comput.*, 9:27–32, 2000. MR-1751301

[Sei14] P. Seidel. *The historical process of the spatial Moran model with selection and mutation*. Universität Erlangen-Nürnberg, Erlangen, 2014. Dissertation.

[Spi76] F. Spitzer. *Principles of Random Walk*. Springer-Verlag, 2nd edition, 1976. MR-0388547

[SS08] R. Sun and J.M. Swart. The Brownian net. *Ann. Probab.*, 36:1153–1208, 2008. MR-2408586

[SSS15] E Schertzer, R. Sun, and J.M. Swart. The Brownian web, the Brownian net, and their universality. arXiv:1506.00724, 2015.

[TW98] B. Toth and W. Werner. The true self-repelling motion. *Probab. Theory Related Fields*, 111:375–452, 1998. MR-1640799

[vdBK02] J. van den Berg and H. Kesten. Randomly coalescing random walk in dimension ≥ 3. *In and out of equilibrium. Progr. Probab.*, 51:1–45, 2002. MR-1901947

[Win02] A. Winter. Multiple scale analysis of spatial branching processes under the palm distribution. *Electron. J. Probab.*, 7:1–74, 2002. MR-1921742

[Zho03] X. Zhou. Clustering behavior of a continuous-sites stepping-stone model with Brownian migration. *Electron. J. Probab.*, 8:1–15, 2003. MR-1986843