AGE EVOLUTION IN THE MEAN FIELD FOREST FIRE MODEL VIA MULTITYPE BRANCHING PROCESSES

BY EDWARD CRANE[1], BALÁZS RÁTH[2] AND DOMINIC YEO[3]

[1]*School of Mathematics, University of Bristol, Edward.Crane@bristol.ac.uk*

[2]*MTA-BME Stochastics Research Group, Budapest University of Technology and Economics, rathb@math.bme.hu*

[3]*Department of Statistics, University of Oxford, dominicjyeo@gmail.com*

We study the distribution of ages in the mean field forest fire model introduced by Ráth and Tóth. This model is an evolving random graph whose dynamics combine Erdős–Rényi edge-addition with a Poisson rain of *lightning strikes*. All edges in a connected component are deleted when any of its vertices is struck by lightning. We consider the asymptotic regime of lightning rates for which the model displays self-organized criticality. The *age* of a vertex increases at unit rate, but it is reset to zero at each burning time. We show that the empirical age distribution converges as a process to a deterministic solution of an autonomous measure-valued differential equation. The main technique is to observe that, conditioned on the vertex ages, the graph is an inhomogeneous random graph in the sense of Bollobás, Janson and Riordan. We then study the evolution of the ages via the multitype Galton–Watson trees that arise as the limit in law of the component of an identified vertex at any fixed time. These trees are critical from the gelation time onwards.

1. Introduction. We study a stochastic model where a network grows steadily, but is subject to occasional destructive events. These can spread widely, but by damaging the connectivity, each destructive event makes it harder for future destruction to propagate through the network. As motivation, consider the effect of fires on a dense forest. Given the right conditions, a fire can destroy all the trees in a region, but afterwards, future fires cannot pass through this area until some trees have regrown.

A probabilistic model of a complex interacting system is considered to be particularly interesting if the effect sizes or the correlations between regions follow a power-law decay. These so-called *critical* phenomena are observed in many complex real-world networks. Seminal work of Bak, Tang and Wiesenfeld [3] considers models where from a broad range of initial conditions, the dynamics move the system into a class of states where critical phenomena are observed, and then maintain it there. These authors describe this property as *self-organized criticality (SOC)*, and in recent years a wide range of mathematical models across many contexts have been shown to exhibit such behaviour; see, for example, [20], Section 3, on the SOC of the Abelian sandpile model and [8] for the Curie–Weiss model of SOC.

The *forest fire model* that we study in this paper is a random process taking values in subgraphs of the complete graph K_n. The lattice setting, introduced by Drossel and Schwabl [12], where geometry plays a more central role has also been studied: the subcritical forest fire model on $\mathbb{Z}^d$ is constructed by Dürre [13–15] and the critical model on the half-plane is constructed by Graf [19]. However, the rigorous construction of a self-organized critical forest fire model on $\mathbb{Z}^d$ poses a real mathematical challenge: in fact Kiss, Manolescu and Sidoravicius show in [21] (using the earlier results of van den Berg and Brouwer [31]) that it

is impossible to construct such a model on $\mathbb{Z}^2$ by starting with standard dynamical percolation and requiring edges to be burned as soon as they belong to an infinite cluster.

Our focus is the *mean field* setting where edges may appear between any pair of vertices. Without the destructive dynamics of fires, adding edges uniformly at random defines the famous *random graph process* of Erdős and Rényi [16]. This process experiences a phase transition between a subcritical regime, where the largest components have logarithmic size relative to the size of the graph, and a supercritical regime, where for large graphs an asymptotically strictly positive proportion of the vertices form a unique *giant component*. During the asymptotically instantaneous transition, the graphs display various *critical* properties, notably a power-law decay in component sizes, which matches behaviour that is observed in many real-world networks not just at phase transition times but at all times.

The *mean field forest fire (MFFF)* process was introduced in [27]. It can be viewed as an adjustment to the Erdős–Rényi dynamics, which destroys the edges of potential giant components as they are forming, and thus maintains the system in a critical state forever. The following informal definition will be made precise in Section 2.1:

- The model has n vertices, with some (possibly random) initial set of undirected edges at time 0.
- Each possible edge joining two vertices appears at rate $1/n$, independently.
- At rate $\lambda = \lambda(n)$, each vertex is struck by lightning, independently. When a vertex is struck by lightning, the vertex survives but all of the edges in its connected component (or cluster) are instantaneously deleted. Those edges may subsequently reappear.

If a lightning hits a vertex v at time t, we say that t is a burning time of all the vertices in the connected cluster of v.

The most interesting asymptotic regime for the lightning rate is $1/n \ll \lambda(n) \ll 1$. In this regime, clusters of any fixed finite size are destroyed at a negligible rate when n is large. However, the total rate of lightning strikes in the model diverges, so if a cluster of size comparable to n were able to form then it would only survive for time $o(1)$. In [27], it is shown that in this regime the model displays SOC. Subject to some assumptions on the initial conditions, the limiting cluster size distribution in the $n \to \infty$ limit is deterministic and satisfies a coupled system of differential equations called the *critical forest fire equations*; see Propositions 2.16 and 2.18 below. The limiting cluster size distribution stays subcritical until a certain *gelation time*. At the gelation time and afterwards, the limiting cluster size distribution is critical in the sense that it has a polynomially decaying tail. No giant component forms; in fact, the model is *conservative*, meaning very roughly that at any time t, nearly all of the vertices are contained in small connected components.

The goal of this paper is to describe the local graph structure of the MFFF at time t as $n \to \infty$ in terms of a multitype branching process, and also to give a simple description of the time evolution of the parameters that govern this multitype branching processes.

The simplest initial condition for the MFFF is the *monodisperse* state, where at time 0 there are no edges, so the graph consists of n isolated vertices. In this paper, we will consider some more general initial conditions, but to state our results informally in this Introduction *we will discuss only the monodisperse initial condition*. At any time $t \geq 0$, each vertex v has an *age* $a_t^n(v)$, which is defined to be the time since it was last burned, or t if it has not yet been burned. Let $\pi_t^n = \frac{1}{n} \sum_v \delta_{a_t^n(v)}$ be the empirical measure of these ages.

Our central observation, stated formally as Theorem 2.6, is that conditional on the ages $a_t^n(v)$ and $a_t^n(w)$ of two vertices v and w, the probability that they are joined by an edge at time t is exactly $1 - \exp(-a_t^n(v) \wedge a_t^n(w)/n)$. Furthermore, these events are independent for distinct pairs of vertices. So conditional on π_t^n, the graph seen at time t is an inhomogeneous random graph *(IRG)* in the sense of Bollobás, Janson and Riordan [5].

Our first main result, Theorem 2.19, shows that the empirical age distributions π_t^n converge as $n \to \infty$ to a deterministic limit distribution π_t. In the course of the proof, we identify π_t explicitly as the marginal distribution of a_t in a Markov process $(C_t, a_t)_{t \geq 0}$ that takes values in $E \times [0, \infty)$, where E is the set $\{1, 2, 3, \ldots\}$ equipped with the one-point compactification topology, with 1 identified with the point at ∞. This topology makes the process almost surely continuous from the left at the random times when C_t explodes. The age a_t is simply the time that has elapsed since the last explosion at time t, or $a_t = t$ if there is no explosion in $[0, t]$. (See Section 6 for the definitions in the case of general initial conditions.)

(C_t, a_t) is a *McKean–Vlasov process* with jumps, meaning that it is a Markovian Feller process whose infinitesimal generator at time t is defined in terms of the distribution of (C_t, a_t). We show that (C_t, a_t) is the distributional limit of the process $(C_t^n(\rho^n), a_t^n(\rho^n))$, where ρ^n is a uniformly chosen vertex of the graph and $C_t^n(\rho^n)$ is the size of the cluster of ρ^n at time t. The marginal $(C_t)_{t \geq 0}$ is the *cluster growth process* $(C_t)_{t \geq 0}$ introduced in [11], which is a McKean–Vlasov process on its own; it is an explosive pure jump process with explosion times $\tau_\ell, \ell \in \mathbb{N}$ which returns to $C_{\tau_\ell} = 1$ at each explosion time τ_ℓ. The jump rate of C_t is k when $C_t = k$ and the jump distribution for jumps at time t is the law of C_t itself.

Remarkably, $(a_t)_{t \geq 0}$ on its own is also a McKean–Vlasov process. It increases at rate 1 except at a random discrete set of jump times when it jumps down to 0. We show that the jump rate of a_t at each time t is a function of a_t and the distribution π_t of a_t. This leads to our second main result, Theorem 2.21, which is that $(\pi_t, t \geq 0)$ satisfies an autonomous differential equation that we call the *age differential equation*; see (2), (3) and (4) below. The well-posedness of this differential equation is proved in the companion paper [10], so in fact π_t is uniquely determined from π_0 by the differential equation. However, we stress that in this paper we determine π_t from π_0 using the cluster growth process.

Recall from [27] that there exists a so-called *gelation time* $t_{\mathrm{gel}} \geq 0$, at which the model makes a phase transition from subcritical to critical behaviour. For $0 \leq t < t_{\mathrm{gel}}$, the age of each vertex simply increases at rate 1 unless it burns before t_{gel}. However, only an asymptotically negligible proportion of the vertices burn before t_{gel}, so the limiting age distribution satisfies the simple transport equation

$$(1) \qquad \frac{\mathrm{d}\pi_t}{\mathrm{d}t} = -\delta_0' * \pi_t.$$

Here, δ_0' is the derivative of the Dirac delta at 0, so this statement is an equality of Schwarz distributions. In other words, for $0 \leq t \leq t_{\mathrm{gel}}$ and for any Borel set $A \subseteq [0, \infty)$,

$$\pi_t(A) = \pi_0(\{x - t : x \in A\}).$$

The situation is more interesting for $t \geq t_{\mathrm{gel}}$, when the model is critical. Then, for each such t, there exists a unique nonnegative, continuous and nondecreasing function $s \mapsto \theta_t(s)$ satisfying $\int \theta_t(s) \, d\pi_t(s) = 1$ and

$$(2) \qquad \theta_t(s) = \int_0^\infty \theta_t(u)(u \wedge s) \, \mathrm{d}\pi_t(u), \quad s \in [0, \infty].$$

For a fixed $t \geq t_{\mathrm{gel}}$, we denote by $\theta_t \pi_t$ the probability measure absolutely continuous with respect to π_t with Radon–Nikodym derivative θ_t. We denote by $\varphi(t)$ the limiting total rate of mass of burnt vertices at time t (see Proposition 2.16 for details). In fact, $\varphi(t)$ is also equal to the explosion rate of C_t. Then for $t > t_{\mathrm{gel}}$, π_t satisfies the following distribution-valued differential equation:

$$(3) \qquad \frac{\mathrm{d}\pi_t}{\mathrm{d}t} = -\delta_0' * \pi_t - \varphi(t)\theta_t \pi_t + \varphi(t)\delta_0.$$

We prove (3) by identifying it with the Kolmogorov forward equation of the McKean–Vlasov process $(a_t)_{t\geq 0}$. Let us give interpretations of the individual terms. As in (1), the transport term $-\delta_0' * \pi_t$ describes the constant, deterministic growth of the ages of all vertices not instantaneously involved in fires. The term $-\varphi(t)\theta_t\pi_t$ describes the change due to the removal of burning vertices. The final term $\varphi(t)\delta_0$ corresponds to the fact that all vertices burned at time t reappear with age zero. We will show that criticality of the forest fire equations corresponds to criticality of the IRG which describes the system conditional on the ages. The local structure of the IRG seen at time t is well approximated by a multitype Poisson branching process (see Definition 2.8 and Theorem 2.20(i)), and θ_t is the right eigenfunction corresponding to the principal eigenvalue $\lambda_t = 1$ of the branching operator. The heuristic idea is that $\theta_t\pi_t$ approximates the distribution of ages in very large components of the IRG, which account for nearly all of the burning vertices (see Section 2.6.3 for details). In fact, we do not prove this global statement but instead obtain (3) by considering the local limit, showing that the rate at which a_t jumps down to 0 when $a_t = s$ is $\varphi(t)\theta_t(s)$. By careful analysis of the multitype branching process, we show

$$(4) \qquad \varphi(t) = \left(\int \theta_t(s)^3 \, d\pi_t(s) \right)^{-1}.$$

Combining (2), (3) and (4), we have an autonomous differential equation, describing the evolution of π_t in terms of π_t, without reference to t.

2. Statements of results.

We precisely introduce the mean field forest fire model in Section 2.1, age-driven inhomogeneous random graphs in Section 2.2, age-driven multitype branching processes in Section 2.3, and the critical forest fire equations in Section 2.4. These provide the necessary background to understand the statements of our main results, stated in Section 2.5. In Section 2.6, we set our results in the context of related literature. In Section 2.7, we pose some open questions. In Section 2.8, we give an overview of the contents of the rest of the paper.

2.1. The mean field forest fire.

We will always use the following definition of a mean field forest fire process on vertex set $[n] := \{1, \ldots, n\}$, with lightning rate λ, following [27]. We refer to this model as MFFF(n, λ).

DEFINITION 2.1 (Graphical construction of MFFF(n, λ)). Let $\mathcal{E}$ be a Poisson point process (PPP) of rate $1/n$ on $\binom{[n]}{2} \times [0, \infty)$, and Λ be an independent PPP of rate λ on $[n] \times [0, \infty)$. These PPPs will determine edge arrivals and lightning strikes, respectively, at times given by their second coordinates. Given some (possibly random) initial graph $\mathcal{G}_0^n$ with vertex set $[n]$, we construct the random graph-valued process $(\mathcal{G}_t^n)_{t=0}^\infty$ started from $\mathcal{G}_0^n$ as follows, working through the points of $\mathcal{E} \cup \Lambda$ in increasing order of their time coordinates:

(i) add the edge $\{i, j\}$ to obtain $\mathcal{G}_t^n$ from $\mathcal{G}_{t-}^n$ if $\{i, j\} \times t$ is a point of $\mathcal{E}$, and edge $\{i, j\}$ is not already present in $\mathcal{G}_{t-}^n$ (and otherwise do nothing),
(ii) erase the edges of the connected component $\mathcal{C}_{t-}^n(i)$ of vertex i in $\mathcal{G}_{t-}^n$ to obtain $\mathcal{G}_t^n$ from $\mathcal{G}_{t-}^n$ if $\{i\} \times t$ is a point of Λ.

In the latter case, we say that the vertices of $\mathcal{C}_{t-}^n(i)$ are burned at time t.

REMARK 2.2. We will always view MFFF(n, λ) as a graph-valued process coupled with its lightning process Λ. This will simplify the choice of probability space in subsequent arguments.

In [27], the asymptotic behaviour of MFFF processes is studied as $n \to \infty$, where the lightning rate $\lambda(n)$ satisfies the *critical relation*

$$(5) \qquad\qquad\qquad 1/n \ll \lambda(n) \ll 1.$$

We will assume throughout that this critical relation holds. Informally, this has the effect that small components are negligibly affected by lightning, whereas components of size $\Theta(n)$ are burned into singletons instantly. Roughly speaking, this means that such giant components never appear.

Let us stress that the earlier results of [27] and [11] concerning the MFFF (which will be recalled in Section 2.4 and Section 6.1, resp.) are about the *sizes* of connected clusters in the process $(\mathcal{G}_t^n)$, but the main results of this paper are about the *graph structure* itself for $(\mathcal{G}_t^n)$.

2.2. *Age-driven inhomogeneous random graphs.* We will make a connection between the MFFF(n, λ) and the theory of inhomogeneous random graphs. In order to do so, we need to study the MFFF augmented with extra information about the *ages* of vertices.

DEFINITION 2.3 (Ages in the mean field forest fire model). For $a_0^n(i) \in [0, \infty)$, $i \in [n]$, we define a *mean field forest fire with ages* MFFFA$(n, \underline{a}_0^n, \lambda)$ as follows. We take MFFF(n, λ) as in Definition 2.1, with some initial graph $\mathcal{G}_0^n$, along with its lightning process Λ, and we call $a_0^n(i)$ the *initial age* of vertex i.

For $t > 0$, denote by $a_t^n(i)$ the *age* of vertex i at time t, that is, $a_t^n(i)$ is $a_0^n(i) + t$ if i did not burn during the time interval $[0, t]$, otherwise $a_t^n(i)$ is $t - s$, where s is the last burning time of i on the time interval $[0, t]$. We write $\underline{a}_t^n := (a_t^n(i))_{i \in [n]}$ for the vector of ages, and

$$(6) \qquad\qquad\qquad \pi_t^n := \frac{1}{n} \sum_{i=1}^n \delta_{a_t^n(i)}$$

for the empirical measure of ages.

So the age of a vertex in MFFFA$(n, \underline{a}_0^n, \lambda)$ increases deterministically at unit rate, but is reset to zero at each burning time.

We introduce a special case of the class of inhomogeneous random graphs introduced by Bollobás, Janson and Riordan [5].

DEFINITION 2.4 (Age-driven IRG). Given a (possibly random) sequence

$$\underline{a}^n = (a^n(1), \ldots, a^n(n)) \in [0, \infty)^n,$$

we define the *age-driven inhomogeneous random graph (IRG)* $G^{\text{age}}(n, \underline{a}^n)$ on the vertex set $[n]$ as follows. Conditional on $\underline{a}^n$:

- $a^n(i)$ is said to be the *age* of vertex $i \in [n]$;
- independently for different edges, the edge $\{i, j\} \in \binom{[n]}{2}$ is present with probability

$$1 - \exp\left(-\frac{a^n(i) \wedge a^n(j)}{n}\right).$$

Our main results concern the evolution of the process of ages in the MFFF model started from an age-driven IRG. So, although Definition 2.3 makes sense in the generality stated, from now on, we will usually make the following assumption.

ASSUMPTION 2.5 (MFFFA started from age-driven IRG). The MFFFA$(n, \underline{a}_0^n, \lambda)$ is started from the initial graph $\mathcal{G}_0^n \stackrel{d}{=} G^{\text{age}}(n, \underline{a}_0^n)$.

The central idea underpinning our results is that the dynamics of the MFFFA preserve the class of age-driven IRGs.

THEOREM 2.6 (($\mathcal{G}_t^n$) and ($\underline{a}_t^n$) are intertwined). *If* ($\mathcal{G}_t^n$) *is a MFFFA process with initial condition distributed as* $G^{\text{age}}(n, \underline{a}_0^n)$ *then, for any* $t \geq 0$,

(7) $$\text{conditional on } (\underline{a}_s^n, s \in [0, t]), \text{ we have } \mathcal{G}_t^n \overset{d}{=} G^{\text{age}}(n, \underline{a}_t^n).$$

A version of this result with weaker conditioning and with monodisperse initial condition appears in [33] as Lemma 5.8. We will give a shorter proof here in Section 3.1.

REMARK 2.7. It follows from (7) that $(\underline{a}_t^n)_{t \geq 0}$ is a time-homogeneous Markov process, and by the vertex exchangeability of the MFFF dynamics $(\pi_t^n)_{t \geq 0}$ is also a time-homogeneous Markov process. Using the notation of [29], Section 3.2, the Markov process $(\mathcal{G}_t^n)$ is *intertwined* on top of $(\underline{a}_t^n)$ (cf. [29], Proposition 3.4).

In order to state some of our main results, we need to define the notion of local weak limits of age-driven IRGs: this is what we will do in the next section.

2.3. *Age-driven multitype branching processes.* We introduce a family of branching process trees, also augmented with ages, which appear as local weak limits of age-driven IRGs.

DEFINITION 2.8 (Age-driven multitype branching process (MBP)).
Given a Borel probability measure π on $[0, \infty)$, we define T^π, a multitype Galton–Watson tree with vertex set $V(T^\pi)$ and *with ages* $a : V(T^\pi) \to [0, \infty)$, as follows. The root ρ has age $a(\rho) \overset{d}{=} \pi$, and then any vertex of age s has an independent set of offspring vertices with ages given by a Poisson random measure with intensity $(s \wedge u) \, d\pi(u)$.
We denote by $|T^\pi|$ the cardinality of $V(T^\pi)$.
We write T_s^π for the random tree T^π constructed in the same way but starting with a root vertex of deterministic age s. We also define T_∞^π to be the random tree whose root has infinite age, meaning that its offspring have ages described by a Poisson random measure with intensity $u \, d\pi(u)$. If $\int u \, d\pi(u) < \infty$, then this is a finite intensity measure so the root almost surely has only finitely many offspring.
We write $\mathcal{L}_\pi$ for the branching operator, that is the Perron–Frobenius operator defined for $f \in L^1(\pi)$ by

(8) $$\mathcal{L}_\pi f(s) := \int f(u)(s \wedge u) \, d\pi(u).$$

When $\mathcal{L}_\pi$ maps $L^2(\pi)$ into itself, denote by $\|\mathcal{L}_\pi\|$ the $L^2(\pi)$-operator norm of $\mathcal{L}_\pi$.

$\mathcal{L}_\pi$ has a probabilistic interpretation. The root of T_s^π has a random number K of offspring with ages $a_1 \leq \cdots \leq a_K$, and $\mathcal{L}_\pi f(s) = \mathbb{E}(f(a_1) + \cdots + f(a_K))$.

LEMMA 2.9 (Normalized principal eigenfunction). *Suppose* π *is a Borel probability measure on* $[0, \infty)$ *such that* $0 < \int x \, d\pi(x) < \infty$. *Then the Perron–Frobenius operator* $\mathcal{L}_\pi$ *is a compact self-adjoint operator on* $L^2(\pi)$ *whose principal eigenvalue* λ *satisfies* $0 < \lambda = \|\mathcal{L}_\pi\|$. *Moreover there exists a unique nonnegative eigenfunction* $\theta \in L^2(\pi)$ *for which* $\mathcal{L}_\pi \theta = \lambda \theta$ *and* $\int \theta(x) \, d\pi(x) = 1$.

We will prove a more detailed result in Section 4.1 that includes the statements of Lemma 2.9.

In Section 4.2, we combine [5], Lemma 6.1, (characterizing supercriticality) and some further arguments to prove the following result.

PROPOSITION 2.10 (Trichotomy of $|T^\pi|$). *Let π be a Borel probability measure on $(0, \infty)$ with $\int x \, d\pi(x) < \infty$. Then the following trichotomy holds*:

- $\|\mathcal{L}_\pi\| < 1$ *if and only if* $\mathbb{E}[|T^\pi|] < \infty$;
- $\|\mathcal{L}_\pi\| = 1$ *if and only if* $\mathbb{P}(|T^\pi| < \infty) = 1$ *and* $\mathbb{E}[|T^\pi|] = \infty$;
- $\|\mathcal{L}_\pi\| > 1$ *if and only if* $\mathbb{P}(|T^\pi| = \infty) > 0$.

We will say that π is age-subcritical, age-critical *and* age-supercritical *when* $\|\mathcal{L}_\pi\| < 1$, $\|\mathcal{L}_\pi\| = 1$ *and* $\|\mathcal{L}_\pi\| > 1$, *respectively.*

In fact, the results of Proposition 2.10 hold under slightly weaker conditions on π. Further details are deferred to Section 4.2.

To make sense of approximating age-driven IRGs by age-driven multitype Galton–Watson trees, we briefly introduce one version of *local weak convergence* of random graphs, as studied by Aldous, Benjamini and Schramm. See [32] for a comprehensive account of this.

DEFINITION 2.11 (Local weak convergence). Let $(G^n, n \geq 1)$ be a sequence of random graphs where G^n has vertex set $[n]$, and let $\mathcal{B}_k^n(w)$ be the k-neighborhood in G^n of a vertex $w \in [n]$, viewed as a graph rooted at w. Then we say G^n converges in probability in the local weak sense to the random rooted graph (G, ρ) if for every rooted graph (H, v), and every $k \geq 1$,

$$(9) \qquad \frac{1}{n} \sum_{w \in [n]} \mathbb{1}\big[\mathcal{B}_k^n(w) \simeq (H, v)\big] \overset{\mathbb{P}}{\longrightarrow} \mathbb{P}(\mathcal{B}_k(\rho) \simeq (H, v)), \quad n \to \infty,$$

where $\mathcal{B}_k(\rho)$ is the k-neighborhood of ρ in G and the relation $\simeq$ denotes the root-preserving isomorphism of rooted graphs. Let us denote by ρ^n a vertex which is independent of G^n and uniformly distributed in $[n]$. As a consequence of (9), we obtain the weaker condition

$$(10) \qquad \mathbb{P}\big(\mathcal{B}_k^n(\rho^n) \simeq (H, v)\big) \longrightarrow \mathbb{P}\big(\mathcal{B}_k(\rho) \simeq (H, v)\big), \quad n \to \infty,$$

for every rooted graph (H, v), which is called convergence in distribution in the local weak sense. (See [32], Definition 2.7.)

DEFINITION 2.12 (Convergence in probability of age distributions). If π is a probability measure on $\mathbb{R}$ and $\pi^1, \pi^2, \dots$ is a sequence of (possibly random) probability measures on $\mathbb{R}$, then we say that $\pi^n \overset{\mathbb{P}}{\Rightarrow} \pi$ as $n \to \infty$ if $\int f(s) \, d\pi^n(s) \overset{\mathbb{P}}{\to} \int f(s) \, d\pi(s)$ as $n \to \infty$ for any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$.

PROPOSITION 2.13 (IRG locally converges to MBP). *Let $G^n \overset{d}{=} G^{\mathrm{age}}(n, \underline{a}^n)$ be a sequence of age-driven IRGs, for which the empirical age distributions satisfy $\pi^n \overset{\mathbb{P}}{\Rightarrow} \pi$, where $\int x \, d\pi(x) < \infty$. Then G^n converges in probability in the local weak sense to the age-driven multitype branching process tree T^π.*

We deduce the proof of Proposition 2.13 from [32] by a truncation argument in Section 3.2, Theorem 3.11.

2.4. *Critical forest fire equations.* In this section, we recall the main results of [27], which concern the hydrodynamic limit of the component size vector of the MFFF. The assumptions of this paper on the initial state of the MFFF are different from those of [27], so the adaptation of the results of [27] to our setting will require extra work.

DEFINITION 2.14 (Component size vector in MFFF(n, λ)).   For any $t \geq 0$ and for each $k = 1, 2, \ldots$, let us define

$$(11) \qquad v_k^n(t) := \frac{1}{n} \#\{\text{vertices in size } k \text{ components at time } t\}.$$

We write $\underline{v}^n(t)$ for the vector $(v_k^n(t))_{k=1}^\infty$.

The central observation of [27] is that limits of $\underline{v}^n(t)$ in a sequence of forest fires in the self-organized critical regime (5) satisfy a family of coupled differential equations. Our assumptions on the initial state will be different from those of [27], Theorem 1, but the conclusions will be the same.

ASSUMPTION 2.15 ($\underline{v}(0)$ is the law of $|T^{\pi_0}|$).   Suppose that π_0 is a Borel probability measure on $[0, \infty)$ satisfying $\int u \, d\pi_0(u) < \infty$, moreover π_0 is either age-critical or age-subcritical. Let us define $v_k(0) = \mathbb{P}(|T^{\pi_0}| = k)$ for $k = 1, 2, \ldots$ and let $\underline{v}(0) = (v_k(0))_{k=1}^\infty$.

PROPOSITION 2.16 (Critical forest fire equations).   *If the initial condition $\underline{v}(0)$ satisfies Assumption* 2.15, *then the critical forest fire equations*

$$(12) \qquad \frac{\mathrm{d}}{\mathrm{d}t} v_k(t) = \frac{k}{2} \sum_{\ell=1}^{k-1} v_\ell(t) v_{k-\ell}(t) - k v_k(t), \quad k \geq 2,$$

$$(13) \qquad \sum_{k=1}^\infty v_k(t) = 1,$$

have a unique solution $\underline{v}(\cdot)$, which also has the following properties:

    1.

$$(14) \qquad \frac{\mathrm{d}}{\mathrm{d}t} v_1(t) = \begin{cases} -v_1(t) & \text{if } 0 \leq t < t_{\text{gel}}, \\ -v_1(t) + \varphi(t) & \text{if } t > t_{\text{gel}}, \end{cases}$$

where the gelation time t_{gel} *is defined by*

$$(15) \qquad t_{\text{gel}} = \left(\sum_{\ell=1}^\infty \ell v_\ell(0) \right)^{-1} \quad (\text{using the convention } \infty^{-1} = 0),$$

and

$$(16) \qquad \varphi : [t_{\text{gel}}, +\infty) \to (0, +\infty) \text{ is locally Lipschitz-continuous.}$$

    2. $v_k(t)$ *decays exponentially as $k \to \infty$ if $0 \leq t < t_{\text{gel}}$, but*

$$(17) \qquad \sum_{\ell=k}^\infty v_\ell(t) \approx \sqrt{\frac{2\varphi(t)}{\pi}} k^{-1/2} \quad \text{as } k \to \infty \text{ for any } t \geq t_{\text{gel}}$$

(where $a_k \approx b_k$ is a shorthand for $\lim_{k \to \infty} \frac{a_k}{b_k} = 1$).

We will prove Proposition 2.16 in Section 5.4. In [27], Theorem 1, the same conclusions are made under the assumption that $\sum_{\ell=1}^{\infty} \ell^3 v_\ell(0) < +\infty$ (which neither implies nor is implied by Assumption 2.15).

REMARK 2.17.

(i) As soon as one picks any *control function* $\varphi(\cdot)$, one can construct $v_1(\cdot)$ by solving (14), and then $v_k(\cdot)$ for $k = 2, 3, \ldots$ inductively using (12). Proposition 2.16 states that there is a unique $\varphi(\cdot)$ such that the resulting family $(v_k(t))_{k=1}^{\infty}$ of functions satisfies (13) for all $t \geq 0$.

(ii) Currently there is no known explicit solution of $((12) + (13))$ except for the unique stationary solution $v_k(t) = v_k(0) = \frac{2}{k}\binom{2k-2}{k-1}4^{-k}$, $k \geq 1$. Note that this $\underline{v}(0)$ satisfies Assumption 2.15; see Remark 2.23(ii).

The connection of the critical forest fire equations to the MFFF is given by the following result.

PROPOSITION 2.18 (Convergence of component size vector). *Suppose that $\underline{v}(0)$ satisfies Assumption 2.15. Let $\underline{v}(\cdot)$ denote the unique solution to the critical forest fire equations $((12) + (13))$ with initial state $\underline{v}(0)$. Let $(\mathcal{G}^n, n \geq 1)$ be a sequence of $\mathrm{MFFF}(n, \lambda(n))$ processes as defined in Definition 2.1, for which (5) holds, and $|v_k^n(0) - v_k(0)| \xrightarrow{\mathbb{P}} 0$ for all $k \in \mathbb{N}$ as $n \to \infty$. Let $\underline{v}^n(\cdot)$ be defined by (11). Then for any $t_{\max} \in [0, \infty)$, we have*

$$(18) \qquad \sup_{k \geq 1} \sup_{0 \leq t \leq t_{\max}} |v_k^n(t) - v_k(t)| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

The above result for fixed t and fixed k is proved in [27], Theorem 2, while the stronger (18) is proved in [11], Theorem 1.5. Note that the assumptions under which we prove the uniqueness of the solution of the critical forest fire equations (cf. Proposition 2.16) are different from those of the analogous uniqueness result in [27] (cf. Theorem 1 therein), but the proofs of [27], Theorem 2, and [11], Theorem 1.5, only take the uniqueness result as an input, so the same proof also gives our Proposition 2.18.

Proposition 2.18 describes the hydrodynamic limit of the component size vector $\underline{v}^n(t)$ of the MFFF in terms of the critical forest fire equations. We want to give a similar description of the hydrodynamic limit of the empirical age measure π_t^n, which will also allow us to identify the local weak limit of the graph $\mathcal{G}_t^n$.

2.5. *Main results.* Throughout Section 2.5, we enforce Assumption 2.5, that the MFFFA is started from an age-driven IRG.

Our first main result is a limit theorem for the empirical measure of ages in a family of MFFFA processes whose initial empirical age distributions converge in probability.

THEOREM 2.19 (Convergence in probability of empirical age distribution). *Let $(\mathcal{G}_t^n, t \geq 0)$ be a family of $\mathrm{MFFFA}(n, \underline{a}_0^n, \lambda(n))$ processes that satisfy Assumption 2.5. Suppose the initial empirical age measures satisfy $\pi_0^n \xrightarrow{\mathbb{P}} \pi_0$, where $\int x \, d\pi_0(x) < \infty$ and π_0 is either age-critical or age-subcritical. Assume that $\lambda(n)$ satisfies (5). Then for all $t \geq 0$ we have $\pi_t^n \xrightarrow{\mathbb{P}} \pi_t$ as $n \to \infty$, where $(\pi_t)_{t \geq 0}$ is a family of probability measures determined by π_0 that satisfies $\int x \, d\pi_t(x) < \infty$ for all $t \geq 0$.*

The proof of Theorem 2.19 is carried out in Section 6 and is completed in Section 6.5. As we have already mentioned in Section 1, we will identify π_t as the distribution of the age a_t of the watched vertex at time t of the *cluster growth process with ages* $(C_t, a_t)_{t \geq 0}$, the law of which is determined uniquely by π_0. We refer the reader interested in the rigorous definition of $(C_t, a_t)_{t \geq 0}$ to the introduction of Section 6 for details.

Our next main result concerns the local weak limit of the graph of the MFFFA process at time t.

THEOREM 2.20 (The local weak limit of MFFFA at time t).   *Let* $(\mathcal{G}_t^n, t \geq 0)$ *and* π_0 *satisfy the conditions of Theorem* 2.19. *Then the following hold*:

(i) *For any* $t \geq 0$, $\mathcal{G}_t^n$ *converges in probability in the local weak sense to* T^{π_t} *as* $n \to \infty$, *where* π_t *is defined in Theorem* 2.19.

(ii) *The family of functions* $t \mapsto v_k(t)$, $k = 1, 2, \ldots$ *defined by*

$$v_k(t) = \mathbb{P}(|T^{\pi_t}| = k), \quad t \geq 0 \tag{19}$$

coincides with the unique solution of the critical forest fire equations ((12) + (13)) *guaranteed by Proposition* 2.16.

(iii) π_t *is age-subcritical for* $t < t_{\text{gel}}$ *and* π_t *is age-critical for* $t \geq t_{\text{gel}}$ *(where* t_{gel} *was introduced in Proposition* 2.16).

We will prove Theorem 2.20 in Section 6.6.

Our final main result is the precise statement of equations (3) and (4), which describe the driving forces behind the time evolution of π_t.

THEOREM 2.21 (Age differential equations).   *Assume* $\int u \, d\pi_0(u) < \infty$. *Consider the family* (π_t) *from Theorem* 2.19, *and for any* $t \geq t_{\text{gel}}$ *denote by* θ_t *the eigenfunction of* $\mathcal{L}_{\pi_t}$ *corresponding to the eigenvalue* $\lambda = 1$, *as in Theorem* 2.20(iii) *and Lemma* 2.9. *For every compactly supported and continuously differentiable test function* $f : [0, \infty) \to \mathbb{R}$,

$$
\frac{\partial}{\partial t} \int f(s) \, d\pi_t(s)
$$

$$
= \begin{cases} \displaystyle\int f'(s) \, d\pi_t(s) & t < t_{\text{gel}}, \\ \displaystyle\int f'(s) \, d\pi_t(s) - \int f(s)\varphi(t)\theta_t(s) \, d\pi_t(s) + \varphi(t) f(0) & t > t_{\text{gel}}, \end{cases} \tag{20}
$$

where $\varphi(\cdot)$ *denotes the control function appearing in equation* (14) *that corresponds to the solution of* ((12) + (13)) *arising from Theorem* 2.20(ii). *For* $t \geq t_{\text{gel}}$, *we have*

$$\varphi(t) = \left(\int_0^\infty \theta_t(s)^3 \, d\pi_t(s) \right)^{-1}. \tag{21}$$

We prove Theorem 2.21 in Section 6.9. Our proof crucially relies on the precise asymptotics of the generating function of the total number of vertices in the multitype branching process tree $T_s^{\pi_t}$ that we derive in Section 5.

We emphasise that the age differential equation ((20) + (21)) describing the dynamics of π_t is autonomous and time-homogeneous, which is why we have chosen to formulate it in terms of ages rather than "birth-times" of vertices.

In the companion paper [10] it is shown that the system ((20) + (21)) has a unique solution over any time interval $[0, T]$, for each age-critical Borel probability measure π_0 satisfying $\int x \, d\pi_0(x) < \infty$. It is also shown there that π_t is locally Lipschitz in π_0 with respect to the

Wasserstein 1-metric W_1, uniformly over $[0, T]$. W_1 is the L^1 distance between cumulative distribution functions, and it metrizes simultaneous weak convergence and convergence of first moment. It is also shown in [10] that for any solution of $((20) + (21))$, θ_t is a continuous function of $t \in [t_{\text{gel}}, \infty)$, with respect to $L^\infty([0, \infty))$, therefore in (20) we see that $\int f(s)\,d\pi_t(s)$ is continuously differentiable for $t > t_{\text{gel}}$.

COROLLARY 2.22. *Under the assumptions of Theorem* 2.21, *the total burning rate is bounded by one*: *for all* $t \geq t_{\text{gel}}$, $\varphi(t) \leq 1$.

PROOF. Apply Jensen's inequality to the integral in (21), using the convexity of $x \mapsto x^3$ on $[0, \infty)$, noting that $\int \theta_t(s)\,d\pi_t(s) = 1$ and $\theta_t(s) \geq 0$. $\square$

Solving the critical forest fire equations starting from the m-disperse state, where $v_k(0) = \mathbb{1}[k = m]$ (for some $m \geq 2$), at the gelation time $t_{\text{gel}} = 1/m$ the value of φ is m. So Corollary 2.22 genuinely relies on Assumption 2.15.

REMARK 2.23.

(i) If $\underline{v}(0)$ satisfies Assumption 2.15 and $(\underline{v}(t))_{t \geq 0}$ denotes the unique solution of $((12) + (13))$ guaranteed by Proposition 2.16, then $\underline{\widetilde{v}}(0) := \underline{v}(t)$ also satisfies Assumption 2.15 for any $t \geq 0$ by Theorems 2.19, 2.20.

(ii) One can check that the unique fixed point of the age differential equation described in Theorem 2.21 is the age-critical measure π which has density $d\pi(x) = \frac{1}{2}\text{sech}^2(\frac{x}{2})\,dx$.

The principal eigenfunction of the operator $\mathcal{L}_\pi$ is $\theta(x) = 2\tanh(\frac{x}{2})$, $\varphi = 1/2$ and $\mathbb{P}(|T^\pi| = k) = \frac{2}{k}\binom{2k-2}{k-1}4^{-k}$ for any $k \geq 1$; cf. Remark 2.17(ii). Many questions related to the stationary MFFF process are discussed in [9].

REMARK 2.24. Our proof of (21) involves the asymptotics of the generating function $\mathbb{E}(z^{|T^{\pi_t}|})$ as $z \to 1$ (see Section 5), and thus it is analytic in nature. Let us therefore provide a nonrigorous probabilistic explanation of (21). Given an age-critical distribution π_t and $h \in \mathbb{R}_+$, let us define the distribution $\pi_{t,h}$ by $\int f(x)\,d\pi_{t,h}(x) = \int f(t+h)\,d\pi_t(x)$. One can show that

$$(22) \qquad \mathbb{P}(|T^{\pi_{t,h}}| = +\infty) = 2\left(\int_0^\infty \theta_t(s)^3\,d\pi_t(s)\right)^{-1} h + o(h), \quad h \to 0.$$

Note that a similar expression characterising the infinitesimal rate of emergence of the giant component in a family of IRGs near criticality appears in equation (3.12) in Theorem 3.17 of [5]. The fires in the MFFF burn the incipient giant clusters as they try to appear, so it is natural to expect that $\varphi(t)$, the limiting total rate of mass of burnt vertices at time t, is proportional to the speed at which the giant cluster wants to grow, that is, $2(\int_0^\infty \theta_t(s)^3\,d\pi_t(s))^{-1}$. In Section 2.6.2, we further comment on the reason why we have to multiply this speed by $1/2$ to obtain the value of $\varphi(t)$ given in (21).

2.6. *Relation to other work.*

2.6.1. *Erdős–Rényi graphs and the Flory equations.* If we consider the MFFF(n, λ) model of Definition 2.1 without lightning (i.e., $\lambda = 0$) and start it from the empty graph on n vertices, we get back the classical Erdős–Rényi graph process. It is well known [6, 7] that if

we consider the component sizes $(v_k^n(t))$ (cf. Definition 2.14) then the limits $v_k^n(t) \xrightarrow{\mathbb{P}} v_k(t)$ exist as $n \to \infty$ and the limiting $(v_k(t))$ satisfy the Flory equations:

$$(23) \qquad \frac{\mathrm{d}}{\mathrm{d}t} v_k(t) = \frac{k}{2} \sum_{\ell=1}^{k-1} v_\ell(t) v_{k-\ell}(t) - k v_k(t) \sum_{\ell=1}^{\infty} v_\ell(0), \quad k \geq 1,$$

where $\sum_{\ell=1}^{\infty} v_\ell(0) = 1$ follows from our normalization (11). Note that (12) agrees with (23) for any $k \geq 2$, but we obtain the r.h.s. of (14) by adding $\varphi(t)$ to the right-hand side of (23) when $k = 1$ and $t > t_{\mathrm{gel}}$. Also note that (13) fails for the solution of (23) when $t > t_{\mathrm{gel}}$.

The local weak limit of the Erdős–Rényi graph at time t is the Galton–Watson tree with Poi(t) offspring distribution; see, for example, [32], Theorem 2.11. In the notation of Definition 2.8, this tree is T^{δ_t}, where δ_t is the Dirac measure concentrated on t. The explicit solution of (23) is given by

$$v_k(t) = \mathbb{P}(|T^{\delta_t}| = k) = \frac{k^{k-1}}{k!} e^{-kt} t^{k-1}, \quad t \in [0, +\infty),$$

that is, the Borel distribution with parameter t. Note that in the supercritical regime $t > 1$ we have $\sum_{k=1}^{\infty} v_k(t) = 1 - \mathbb{P}(|T^{\delta_t}| = \infty) < 1$, so the Borel distribution is a defective distribution when $t > 1$.

2.6.2. *Stochastic models of Smoluchowski's coagulation equations.* The family of equations (23) is related to Smoluchowski's coagulation equations, for which the final sum in (23) is replaced by $\sum_{\ell=1}^{\infty} v_\ell(t)$:

$$(24) \qquad \frac{\mathrm{d}}{\mathrm{d}t} v_k(t) = \frac{k}{2} \sum_{\ell=1}^{k-1} v_\ell(t) v_{k-\ell}(t) - k v_k(t) \sum_{\ell=1}^{\infty} v_\ell(t), \quad k \geq 1.$$

The solutions to the Flory and Smoluchowski equations coincide until the *gelation time* $t_{\mathrm{gel}} := (\sum_{k=1}^{\infty} k v_k(0))^{-1}$, moreover $\sum_{\ell=1}^{\infty} v_\ell(t) = \sum_{\ell=1}^{\infty} v_\ell(0) = 1$ holds for all $t \in [0, t_{\mathrm{gel}}]$. Beyond t_{gel} the evolution depends, informally, on whether small blocks are allowed to interact with the so-called *gel*, which has mass $\sum_{\ell=1}^{\infty} v_\ell(0) - \sum_{\ell=1}^{\infty} v_\ell(t)$ at time t. In both cases, however, the total mass $\sum_{k=1}^{\infty} v_k(t)$ of small components is strictly decreasing for $t \in [t_{\mathrm{gel}}, \infty)$.

The mean field frozen percolation model [26] is defined exactly like the MFFF (cf. Definition 2.1), only differing in that when a connected component is struck by lightning, the *vertices* of the component get deleted forever. Vertices that have not yet been deleted are called *alive vertices*. The number of alive vertices in the frozen percolation model decreases over time. The analogue of Proposition 2.18 holds in the frozen percolation model by [26], Theorem 1.2, (see also [33], Chapter 4, for a shorter proof with weaker assumptions on the initial state), where the limiting component size densities $(v_k(t))_{k=1}^{\infty}$ solve Smoluchowski's coagulation equations (24).

The solution of (24) is known to be unique and explicit for general initial conditions; see [25] and also [26], Section 2. The frozen percolation model shares the feature of SOC with the MFFF, that is, (17) holds for the solution of (24) with $\varphi(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \sum_{\ell=1}^{\infty} v_\ell(t)$; see [26], Theorem 1.5. One way to approximate SOC is to allow tiny giant clusters to grow and then destroy them successively (see [26], Theorems 1.4, 1.9), and the multiplying factor 1/2 mentioned at the end of Remark 2.24 is explained in [26], Section 7, for an equivalent approximation of SOC in the context of mean field frozen percolation. Roughly speaking, the model spends half of its time in a slightly subcritical state (recovering from the destructions), while the other half is spent in a slightly supercritical state (growing tiny giants).

Another closely related random graph model where (24) describes the limiting component size densities was proposed by Aldous in [1], Section 5.5, and studied by Merle and Normand in [22]. Two connected components of size k and l merge at rate $\frac{kl}{n}$ (just as in Definition 2.1) and connected components disappear if their size exceeds a threshold $\omega(n)$ satisfying $1 \ll \omega(n) \ll n$. [22], Theorem 1.1, states that $\underline{v}^n(t)$ converge to the solution $\underline{v}(t)$ of (24). Note that this result is a special case of the main result of [18], where discrete models of Smoluchowski's coagulation equations with more general coagulation kernels are studied.

Similar to our Theorem 2.20(i), it is shown in [22], Theorem 1.3, that the local weak limit of the "$\omega(n)$-threshold deletion" graph model, started from the empty graph on n vertices, is a Galton–Watson branching process tree with Poisson$(1 \wedge t)$ offspring distribution at time t. Also note that in this case the fraction of alive vertices at time t converges in probability to $1 \wedge 1/t$ as $n \to \infty$ by [22], Theorem 1.2.

2.6.3. *Frozen percolation on inhomogeneous random graphs.* The mean field frozen percolation model of [26] started from an IRG is studied by the third author in [34]. Initially, each of the n vertices has one of k types, and the fraction of vertices with type i is $\pi_i^n(0)$. At time zero, conditional on these types, a pair of vertices with types i and j is connected with probability $1 - \exp(-\kappa_{i,j}/n)$ independently of other pairs, where $(\kappa_{i,j})_{i,j=1}^k$ (the kernel of the IRG) is some symmetric matrix with strictly positive entries. Analogously to our Theorem 2.6, if we condition on the set of alive vertices at any time t and their types, the graph is also an IRG with kernel $(\kappa_{i,j} + t)_{i,j=1}^k$; see [34], Proposition 7.

By [34], Theorem 3, the vector $(\pi_i^n(t))_{i=1}^k$ of proportions of vertices of each type that remain alive at time t has a limit $(\pi_i(t))_{i=1}^k$ as $n \to \infty$. Moreover, if $t \geq t_{\text{gel}}$ then the vector $(\pi_i(t))_{i=1}^k$ obeys the differential equation

$$(25) \qquad \frac{\mathrm{d}}{\mathrm{d}t}\pi_i(t) = -\varphi(t)\theta_i(t)\pi_i(t),$$

where $\varphi(t)$ is the total destruction rate at time t and $(\theta_i(t))_{i=1}^k$ is the right eigenvector corresponding to the principal eigenvalue $\lambda = 1$ of the critical Perron–Frobenius matrix $((\kappa_{i,j} + t)\pi_j(t))_{i,j=1}^k$, normalized so that $\sum_{i=1}^k \theta_i(t)\pi_i(t) = 1$. (Note that the results in [34] are formulated using the left eigenvector $(\mu_i(t))_{i=1}^k$, where $\mu_i(t) = \theta_i(t)\pi_i(t)$.) This result is similar to our Theorem 2.21, but the method of proof is quite different, as we now explain. A key ingredient of the approach of [34] is that the empirical distribution of the vertex types in a very large component of the IRG seen at time t is close to $(\theta_i(t)\pi_i(t))_{i=1}^k$ with high probability (see [34], Section 4, for the precise statement and proof). Instead of proving a similar result for age-driven IRGs, for which the continuous type space presents considerable technical challenges, we derive Theorem 2.21 by directly studying the fine properties of the local limit objects (such as the cluster growth process of [11] and age-driven multitype branching process trees) that arise from the MFFF as $n \to \infty$ (see the introduction of Section 6 for further details about our methods).

2.6.4. *A model with limited aggregations.* In [23], a dynamical variant of the configuration model is discussed where edges matching remaining stubs are added one by one in a uniform fashion. Connected clusters get deleted as soon as their size exceeds a threshold $\alpha(n)$, where n denotes the initial number of vertices and $1 \ll \alpha(n) \ll n$. Similar to our Theorem 2.20, the local weak limit of this random graph model at any time t after gelation turns out to be a critical Galton–Watson tree.

2.7. *Open questions*.

1. For each n, the evolution of the cluster size densities $\underline{v}^n(t)$ of the MFFF can be viewed as a coagulation-fragmentation process (see [27], Section 2.2), which converges in distribution as $t \to \infty$ to $\underline{v}^n(\infty)$, which is distributed according to the unique stationary law of the MFFF. The critical forest fire equations ((12) + (13)) themselves have a unique fixed-point $\underline{v}(\infty)$; see Remark 2.17(ii). However, it has not yet been proved that the limit as $n \to \infty$ of $\underline{v}^n(\infty)$ is deterministic and equal to $\underline{v}(\infty)$; see [9], Conjecture 1.

2. A related question is whether there exist nonconstant periodic or chaotic solutions to the critical forest fire equations ((12) + (13)). Indeed, it is not yet known whether we have $\underline{v}(t) \to \underline{v}(\infty)$ as $t \to \infty$, even in the *monodisperse* case where $\underline{v}(0) = (1, 0, 0, \ldots)$. The analogous question for solutions to the age differential equations ((20) + (21)), for which the expected limit π is discussed in Remark 2.23(ii), is also unknown. In both settings, the conjecture that $\varphi(t) \to \frac{1}{2}$ as $t \to \infty$ is of central importance.

3. Is it possible to write down any explicit nonconstant age-critical solutions of the age differential equations ((20) + (21))? No explicit solutions of the critical forest fire equations ((12) + (13)) are currently known beyond t_{gel}, apart from the stationary solution $\underline{v}(\infty)$.

4. An age-critical or age-subcritical distribution π determines a cluster size distribution $v_k = \mathbb{P}(|T^\pi| = k)$, $k = 1, 2, \ldots$. Not every such probability measure arises from an age distribution π. For example, the support of $(v_k)_{k=1}^\infty$ must be all of $\mathbb{N}$ if $\pi \neq \delta_0$. In the other direction, does $(v_k)_{k=1}^\infty$ uniquely determine π?

2.8. *Overview of the contents of the rest of this paper.* In Section 3, we connect the MFFFA and the notion of age-driven IRG to the literature of IRGs and their local weak limits. In Section 4, we discuss fundamental properties of the branching operator and age-driven MBPs. In Section 5, we state and prove some results about the fine asymptotics of the probability generating functions of the size of age-driven MBP trees. In Section 6, we prove the main results of this paper. We strongly encourage the reader interested in our methods to read the introduction of Section 6 now: in Section 6.1, we recall and introduce some auxiliary Markov processes (e.g., the *cluster growth process with age*) which allow us to describe the asymptotics of the time evolution of the empirical age distribution in the MFFFA. Some properties of these auxiliary Markov processes stated in Section 6.1 (e.g., the *intertwining relation* between them) are interesting in their own right.

3. Age-driven IRGs and their limits. In Section 3.1, we prove Theorem 2.6, that the dynamics of the MFFFA preserve the class of age-driven IRGs.

In Section 3.2, we recall some elements of the theory of IRGs from [5] and [32] and prove Proposition 2.13.

In Section 3.3, we extend this to understand the limiting joint distribution of the ages and component sizes of two i.i.d. uniformly chosen vertices in a large age-driven IRG, for the purpose of running a second moment argument.

3.1. *Forest fire dynamics preserve age-driven IRGs.* PROOF OF THEOREM 2.6. Without loss of generality, we assume that $\underline{a}_0^n$ is deterministic, by conditioning on its value.

In order to prove (7), we will prove an even stronger statement: we will prove that (7) holds even if we further condition on the lightning PPP Λ. In practice, this means that we assume that Λ consists of the deterministic points (i_k, t_k), $k = 1, 2, \ldots$, where $0 = t_0 < t_1 < t_2 < \cdots$, that is, vertex i_k's component is burned at time t_k.

We will prove that (7) holds for all $t \leq t_k$, $k = 0, 1, 2, \ldots$ by induction on k. By construction of $(\mathcal{G}_0^n, \underline{a}_0^n)$ (as an age-driven IRG), (7) holds for $k = 0$.

Now let us assume that (7) holds for all $t \le t_{k-1}$. In particular, the conditional distribution of $(\mathcal{G}_{t_{k-1}}^n, \underline{a}_{t_{k-1}}^n)$ given $\underline{a}_s^n, 0 \le s \le t_{k-1}$ is an age-driven IRG. Since no point of Λ lies in $[n] \times (t_{k-1}, t_k)$, we have

$$(26) \qquad a_t^n(i) = a_{t_{k-1}}^n(i) + (t - t_{k-1}), \quad t_{k-1} \le t < t_k, i \in [n].$$

Edge $\{i, j\}$ is present in $\mathcal{G}_t^n$ either if it is present in $\mathcal{G}_{t_{k-1}}^n$, or if it appears during time interval $(t_{k-1}, t]$. Using the induction hypothesis and the fact that edges appear independently at rate $1/n$ on this interval, we conclude that conditional on $(\underline{a}_s^n, 0 \le s \le t)$ and Λ, each edge $\{i, j\}$ is present in $\mathcal{G}_t^n$ independently with probability

$$(27) \qquad \begin{aligned} & 1 - \exp\left(-\frac{1}{n}\big(a_{t_{k-1}}^n(i) \wedge a_{t_{k-1}}^n(j)\big)\right) \exp\left(-\frac{1}{n}(t - t_{k-1})\right) \\ & \overset{(26)}{=} 1 - \exp\left(-\frac{1}{n}\big(a_t^n(i) \wedge a_t^n(j)\big)\right). \end{aligned}$$

So (7) holds for any $t_{k-1} \le t < t_k$, and if we denote by $\mathcal{G}_{t_k-}^n$ and $\underline{a}_{t_k-}^n$ the graph and the age configuration that we see right before the lightning strike at time t_k, then

$$(28) \qquad \text{conditional on } (\underline{a}_s^n, s \in [0, t_k)), \text{ we have } \mathcal{G}_{t_k-}^n \overset{d}{=} G^{\text{age}}(n, \underline{a}_{t_k-}^n).$$

At time t_k, a lightning strikes vertex i_k and the vertices of the connected component $\mathcal{C}$ of vertex i_k in the graph $\mathcal{G}_{t_k-}^n$ burn, thus we have

$$(29) \qquad a_{t_k}^n(i) = \begin{cases} 0 & \text{if } i \in \mathcal{C}, \\ a_{t_k-}^n(i) = a_{t_{k-1}}^n(i) + (t_k - t_{k-1}) & \text{if } i \in [n] \setminus \mathcal{C}. \end{cases}$$

Note that $\mathcal{C}$ is determined by $\underline{a}_s^n, 0 \le s \le t_k$ because $a_{t_k}^n(i) = 0$ if and only if $i \in \mathcal{C}$. Also note that there are no edges between $\mathcal{C}$ and $[n] \setminus \mathcal{C}$ in $\mathcal{G}_{t_k-}^n$, since $\mathcal{C}$ is a connected component of $\mathcal{G}_{t_k-}^n$. Also note that it follows from (28) and (29) that if we condition on $(\underline{a}_s^n, s \in [0, t_k])$ then the subgraph of $\mathcal{G}_{t_k-}^n$ spanned by the vertices $[n] \setminus \mathcal{C}$ is an age-driven inhomogeneous random graph with ages $(a_{t_k}^n(i), i \in [n] \setminus \mathcal{C})$, that is, independently for different edges, the edge $\{i, j\} \in \binom{[n] \setminus \mathcal{C}}{2}$ is present with probability $1 - \exp(-\frac{1}{n}(a_{t_k}^n(i) \wedge a_{t_k}^n(j)))$. Also note that the subgraphs of $\mathcal{G}_{t_k-}^n$ and $\mathcal{G}_{t_k}^n$ spanned by the vertices $[n] \setminus \mathcal{C}$ are the same, but if either i or j belongs to $\mathcal{C}$, the edge $\{i, j\}$ is not present in $\mathcal{G}_{t_k}^n$. Putting these observations together with Definition 2.4 and (29), we obtain that (7) holds for $t = t_k$. This completes the proof of Theorem 2.6. $\square$

3.2. *Local weak convergence of age-driven IRGs.* The goal of Section 3.2 is to prove Proposition 2.13.

In order to apply the theory of IRGs established in [5], we recall the notion of graphical and irreducible kernels from [5], Definitions 2.7, 2.10, (or, alternatively, from [32], Definition 3.3). In our case the *ground space* $(\mathcal{S}, \mu)$ is $([0, \infty), \pi)$, where π is a Borel probability measure on $[0, \infty)$, and the kernel is $\kappa(x, y) = x \wedge y$. Our *vertex space* (cf. [5], Section 2) is $([0, \infty), \pi, (\underline{a}^n)_{n \ge 1})$, where $\underline{a}^n = (a^n(1), \ldots, a^n(n)) \in [0, \infty)^n$ is a sequence of (possibly random) ages in $G^n = G^{\text{age}}(n, \underline{a}^n)$, which determines the (possibly random) empirical measure π^n. Our definition of $\pi^n \overset{\mathbb{P}}{\Rightarrow} \pi$ (cf. Definition 2.12) is one of the equivalent characterizations given in [5], Lemma A.2. Our particular definition of the conditional edge probabilities given the types (i.e., ages) of the vertices in an age-driven IRG (cf. Definition 2.4) is in line with [5], (2.6).

Consider $G^n \overset{(d)}{=} G^{\text{age}}(n, \underline{a}^n)$ and for any $B \ge 0$ define $a_B^n(i) := a^n(i) \wedge B$ for $i = 1, \ldots, n$. We may use these truncated ages to define an age-driven IRG $G_{(B)}^n \overset{(d)}{=} G^{\text{age}}(n, \underline{a}_B^n)$, coupled

such that $G^n_{(B)}$ is a subgraph of G^n. We also define the bounded continuous kernel $\kappa_B(x, y) = x \wedge y \wedge B$ and note that $G^n_{(B)}$ can be viewed as an IRG with types $\underline{a}^n$ and kernel κ_B. Let us also introduce the multitype Galton–Watson tree $T^\pi_{(B)}$, which is defined in the same way as T^π (cf. Definition 2.8) but using the kernel κ_B, so that the offspring of a vertex of type s are given by a PPP with intensity $(s \wedge u \wedge B) \, d\pi(u)$.

We first show that the "truncated" analogue of Proposition 2.13 holds.

LEMMA 3.1 (Truncated IRG locally converges to MBP). $G^n_{(B)}$ *converges in probability in the local weak sense as* $n \to \infty$ *to* $T^\pi_{(B)}$.

PROOF.   First, we check that κ_B is a *graphical kernel* in our vertex space: see conditions (i), (ii) and (iii) of [5], Definition 2.7.

   (i) κ_B is continuous.
   (ii) κ_B is bounded, which implies $\int \int \kappa_B(x, y) \, d\pi(x) \, d\pi(y) < \infty$.
   (iii) The expected number of edges in $G^{\mathrm{age}}(n, \underline{a}^n)$ divided by n tends to $\frac{1}{2} \int \int \kappa_B(x, y) \, d\pi(x) \, d\pi(y)$ as $n \to \infty$, and this holds for bounded continuous kernels by [5], Lemma 8.1 and Remark 8.4.

Thus our kernel κ_B is graphical. κ_B is also irreducible (cf. [5], Definition 2.10), therefore the proof of Lemma 3.1 follows from [32], Theorem 3.11.   □

REMARK 3.2.   To ensure that κ itself is a graphical kernel, we would have to assume some extra condition on π^n in addition to its convergence in probability to π, for example, $\mathbb{E}(\int x \, d\pi^n(x)) \to \int x \, d\pi(x)$ would suffice. Our truncation argument allows us to avoid making such an extra assumption.

Let us denote by $\mathcal{S}^{n,k}_{(B)}$ the set of vertices v of G^n such that there is a vertex of age B within distance k of v in $G^n_{(B)}$. Note that for any $v \in [n] \setminus \mathcal{S}^{n,k}_{(B)}$ the ball of radius k about v in G^n agrees with that in $G^n_{(B)}$.

LEMMA 3.3 ($G^n_{(B)}$ and G^n locally look similar).   *Let* $\epsilon > 0$ *and* $k \in \mathbb{N}$. *Then there exists a constant* $B = B(\epsilon, k)$ *such that for any sufficiently large* n *we have*

$$(30) \qquad\qquad \mathbb{P}(|\mathcal{S}^{n,k}_{(B)}| \geq n\epsilon) \leq \epsilon.$$

PROOF.   Consider the set $\mathrm{Path}(\ell, B)$ that consists of directed simple paths in $G^n_{(B)}$ of length ℓ, beginning in the set $S = \{v \in [n] : a^n_B(v) = B\}$. Conditional on the ages, the probability that any particular directed path $(v_0, v_1, \ldots, v_\ell) \in \mathrm{Path}(\ell, B)$ is present in $G^n_{(B)}$ is bounded as follows:

$$\mathbb{P}[\{v_{i-1}, v_i\} \in E(G^n_{(B)}) : 1 \leq i \leq \ell | \underline{a}^n_B]$$
$$= \prod_{i=1}^\ell (1 - e^{-(a^n_B(v_{i-1}) \wedge a^n_B(v_i))/n}) \leq \prod_{i=1}^\ell \frac{a^n_B(v_{i-1}) \wedge a^n_B(v_i)}{n} \leq \prod_{i=1}^\ell \frac{a^n_B(v_i)}{n}.$$

Summing over all choices of $v_0 \in S$ and $(v_1, \ldots, v_\ell) \in [n]^\ell$, we obtain

$$\mathbb{E}[|\mathrm{Path}(\ell, B)| \, | \underline{a}^n_B] \leq |S| \left(\sum_{v \in [n]} \frac{a^n_B(v)}{n} \right)^\ell = |S| \left(\int (x \wedge B) \, d\pi^n(x) \right)^\ell.$$

Note that $\mathcal{S}_{(B)}^{n,k}$ is the set of vertices whose k-ball in $G_{(B)}^n$ meets S, thus

$$(31) \qquad \mathbb{E}[|\mathcal{S}_{(B)}^{n,k}|/n|\underline{a}_B^n] \leq \frac{|S|}{n} \sum_{\ell=0}^k \left(\int (x \wedge B) \, d\pi^n(x) \right)^\ell =: Z_n.$$

If $\pi(\{B\}) = 0$, then Z_n converges to $\pi([B, \infty)) \sum_{\ell=0}^k (\int (x \wedge B) \, d\pi(x))^\ell$ in probability, and the latter quantity is at most $\epsilon^2/4$ if we choose B big enough, since $\int x \, d\pi(x) < \infty$. Then for large enough n we have $\mathbb{P}(Z_n > \frac{\epsilon^2}{2}) \leq \frac{\epsilon}{2}$, moreover $\mathbb{P}(|\mathcal{S}_{(B)}^{n,k}|/n \geq \epsilon | Z_n \leq \frac{\epsilon^2}{2}) \leq \frac{\epsilon}{2}$ holds by (31), thus (30) follows. $\square$

Next, we observe that $T_{(B)}^\pi$ converges to T^π in the local weak sense as $B \to \infty$: this follows from the fact that $\kappa_B \nearrow \kappa$, therefore one can couple $T_{(B)}^\pi$ simultaneously for all $B \geq 0$ to T^π so that $T_{(B)}^\pi$ converges monotonically (with respect to the partial ordering of graph inclusion) to T^π almost surely (for the details of this coupling; see the proof of [32], Theorem 3.10).

Now we can conclude the proof of Proposition 2.13. Let us fix some finite rooted graph (H, v) and $k \in \mathbb{N}$. Recalling Definition 2.11, let $\mathcal{B}_k^n(w)$ be the k-neighborhood in G^n of a vertex $w \in [n]$, and similarly let $\mathcal{B}_{k,(B)}^n(w)$ be the k-neighborhood in $G_{(B)}^n$ of w. Let $X_n := \frac{1}{n} \sum_{w \in [n]} \mathbb{1}[\mathcal{B}_k^n(w) \simeq (H, v)]$ and $X_{n,(B)} := \frac{1}{n} \sum_{w \in [n]} \mathbb{1}[\mathcal{B}_{k,(B)}^n(w) \simeq (H, v)]$. Let $q := \mathbb{P}(\mathcal{B}_k(\rho) \simeq (H, v))$ and $q_{(B)} := \mathbb{P}(\mathcal{B}_{k,(B)}(\rho) \simeq (H, v))$, where $\mathcal{B}_k(\rho)$ and $\mathcal{B}_{k,(B)}(\rho)$ denote the k-neighborhoods of the root in T^π and $T_{(B)}^\pi$, respectively.

Given some $\epsilon > 0$, we want to show that for all large enough n we have $\mathbb{P}(|X_n - q| \geq \epsilon) \leq \epsilon$. Recall that for any $v \in [n] \setminus \mathcal{S}_{(B)}^{n,k}$ we have $\mathcal{B}_k^n(v) = \mathcal{B}_{k,(B)}^n(v)$, thus by Lemma 3.3 we can choose $B^* = B^*(\epsilon, k)$ such that if $B \geq B^*$ then we have $\mathbb{P}(|X_n - X_{n,(B)}| \geq \frac{\epsilon}{3}) \leq \frac{\epsilon}{2}$ for large enough values of n. There exists B^{**} such that $|q - q_{(B)}| \leq \frac{\epsilon}{3}$ for any $B \geq B^{**}$, since $T_{(B)}^\pi$ converges to T^π in the local weak sense as $B \to \infty$. Let $B = B^* \vee B^{**}$. By Lemma 3.1, we have $\mathbb{P}(|X_{n,(B)} - q_{(B)}| \geq \frac{\epsilon}{3}) \leq \frac{\epsilon}{2}$ for large enough n. We obtain the desired $\mathbb{P}(|X_n - q| \geq \epsilon) \leq \epsilon$ for large enough n by the triangle inequality and the union bound. The proof of Proposition 2.13 is complete.

3.3. *Joint weak limit of component sizes and ages.* In order to prove that π_t^n is concentrated around π_t in Theorem 2.19 using a second moment argument, we will show in Proposition 6.7 that the age processes $a_t^n(\rho_1^n)$ and $a_t^n(\rho_2^n)$ of two i.i.d. uniformly distributed vertices ρ_1^n, ρ_2^n in an MFFFA evolve asymptotically independently as $n \to \infty$. The following result implies that the ages of ρ_1^n and ρ_2^n in the initial graph $\mathcal{G}_0^n$ are asymptotically independent.

DEFINITION 3.4 (Components and cardinalities). Given an age-driven IRG $G^{\mathrm{age}}(n, \underline{a}^n)$ (cf. Definition 2.4) and a vertex $i \in [n]$, let $\mathcal{C}^n(i)$ denote the connected component of vertex i in the graph $G^{\mathrm{age}}(n, \underline{a}^n)$, moreover let $C^n(i)$ denote the number of vertices in $\mathcal{C}^n(i)$, that is, $C^n(i) = |\mathcal{C}^n(i)|$.

PROPOSITION 3.5 (Weak limits for two i.i.d. vertices). *Let $G^{\mathrm{age}}(n, \underline{a}^n)$ be a sequence of age-driven IRGs, for which the empirical age distributions satisfy $\pi^n \overset{\mathbb{P}}{\Rightarrow} \pi$, where $\int x \, d\pi_0(x) < \infty$ and π is an age-subcritical or age-critical probability measure. Let ρ_1^n, ρ_2^n be i.i.d. with uniform distribution on $[n]$. Then as $n \to \infty$,*

$$(32) \qquad ((a^n(\rho_i^n), C^n(\rho_i^n)), i \in \{1, 2\}) \Rightarrow ((a(\rho^{(i)}), |T^{(i)}|), i \in \{1, 2\}),$$

where $(a(\rho^{(i)}), |T^{(i)}|), i = 1, 2$ are i.i.d. copies of $(a(\rho), |T^\pi|)$, where ρ is the root of T^π and $a(\rho)$ is the age of the root.

Note that when π is age-supercritical, then $\mathbb{P}(|T^\pi| = \infty) > 0$ (cf. Proposition 2.10) and, therefore, $(C^n(\rho_i^n))_{n \in \mathbb{N}}$ is not tight.

Let us now explain why we omit the proof of Proposition 3.5. The statement of Proposition 2.13 concerns only distributions of rooted graphs, not graphs augmented with ages. However, the proof of [32], Theorem 3.11, (which we used in the proof of Lemma 3.1) at its core uses a coupling argument in which an exploration of G^n from a uniformly chosen vertex is coupled to an exploration of T^π from its root. In this coupling, the ages of corresponding vertices are in fact close with high probability. In the discussion in [32], Section 3.5.1, immediately following the statement of the relevant Theorem 3.11, van der Hofstad remarks that the convergence in probability in the local weak sense can be extended to apply to the topology where two local subgraphs are close only if they are isomorphic as rooted graphs by an isomorphism under which the corresponding vertices' types are uniformly close. This is similar to Benjamini et al. [4], who study a setting with marked edges. We actually do not require the full power of this extension, only the fact that the initial step of the coupling argument is to couple a uniformly chosen random vertex ρ^n of G^n with the root ρ of T^π in such a way that with high probability the age of ρ^n is close to the age of ρ. This observation can be used to show that in fact for any continuity set A of the measure π and any finite rooted graph (H, v) and any $k \geq 1$ we have as $n \to \infty$ that

$$(33) \qquad \frac{1}{n} \sum_{w \in [n]} \mathbb{1}\big[\mathcal{B}_k^n(w) \simeq (H, v), a^n(w) \in A\big] \xrightarrow{\mathbb{P}} \mathbb{P}(\mathcal{B}_k(\rho) \simeq (H, v), a(\rho) \in A).$$

In order to prove (32), we only need to show that for any pair of continuity sets A_1, A_2 of π and finite rooted graphs (H_1, v_1), (H_2, v_2), we have

$$(34) \quad \mathbb{P}(\mathcal{B}_k^n(\rho_i^n) \simeq (H_i, v_i), a^n(\rho_i^n) \in A_i, i \in \{1, 2\}) \to \prod_{i=1}^{2} \mathbb{P}(\mathcal{B}_k(\rho) \simeq (H_i, v_i), a(\rho) \in A_i)$$

as n goes to infinity. Taking the product of two instances of (33) with A_i and (H_i, v_i) in place of A and (H, v) (with $i = 1, 2$, resp.), we obtain that

$$(35) \qquad \frac{1}{n^2} \sum_{w_1, w_2 = 1}^{n} \mathbb{1}\big[\mathcal{B}_k^n(w_i) \simeq (H_i, v_i), a^n(w_i) \in A_i, i \in \{1, 2\}\big]$$

converges in probability to the r.h.s. of (34). Taking the expectation, we obtain (34).

4. The operator $\mathcal{L}_\pi$ and criticality of T^π. In Section 4.1, we prove (a strengthening of) Lemma 2.9. In Section 4.2, we prove Proposition 2.10.

4.1. *Basic properties of $\mathcal{L}_\pi$.* The main goal of Section 4.1 is to prove Lemma 2.9. In fact, we will prove a more detailed result.

Let (X, μ) be a measure space, let κ be a measurable real kernel on X and define an integral operator T by $(Tf)(x) = \int \kappa(x, y) f(y) \, d\mu(y)$. Then the Hilbert–Schmidt norm of T is

$$(36) \qquad \|T\|_{\mathrm{HS}} = \left(\int \int \kappa(x, y)^2 \, d\mu(x) \, d\mu(y) \right)^{1/2}.$$

When $\|T\|_{\mathrm{HS}}$ is finite, T is a *Hilbert–Schmidt operator* on $L^2(\mu)$. It is compact (see [28], Theorem VI.22) with operator norm bounded by

$$(37) \qquad \|T\| \leq \|T\|_{\mathrm{HS}}.$$

LEMMA 4.1. *If $\int s \, d\pi(s) < +\infty$ then $\mathcal{L}_\pi$ is Hilbert–Schmidt, and*

(38)
$$\|\mathcal{L}_\pi\| \leq \|\mathcal{L}_\pi\|_{\text{HS}} \leq \int x \, d\pi(x) < \infty.$$

PROOF. If we let $m := \int x \, d\pi(x) < \infty$, then we have

(39)
$$\|\mathcal{L}_\pi\|_{\text{HS}}^2 \stackrel{(36)}{=} \int \int (x \wedge y)^2 \, d\pi(x) \, d\pi(y) \leq \int \int xy \, d\pi(x) \, d\pi(y) = m^2.$$

Thus (38) holds by (37) and (39). □

REMARK 4.2. In fact, [2], Theorem 4.6, shows (after a change of variables) that $\mathcal{L}_\pi$ is a bounded operator on $L^2(\pi)$ if and only if $\pi([x, \infty)) = O(1/x)$ as $x \to \infty$, it is compact if and only if $\pi([x, \infty)) = o(1/x)$ as $x \to \infty$, and it belongs to the trace class if and only if $\int x \, d\pi(x) < \infty$. In the latter case, one can show that the sum of the eigenvalues converges (absolutely) to $\int x \, d\pi(x)$. See also [5], Example 17.6.

LEMMA 4.3 (Properties of $\mathcal{L}_\pi$). *If $\|\mathcal{L}_\pi\|_{\text{HS}} < \infty$ and $\pi \neq \delta_0$, then:*

(i) *$\mathcal{L}_\pi$ is a positive semidefinite and compact self-adjoint operator on $L^2(\pi)$.*

(ii) *Elements of the image of $\mathcal{L}_\pi$ are represented by Lipschitz functions and $\mathcal{L}_\pi$ maps nonnegative functions to nondecreasing nonnegative functions.*

(iii) *$\mathcal{L}_\pi$ has a simple principal eigenvalue λ satisfying $0 < \lambda = \|\mathcal{L}_\pi\|$.*

(iv) *There is a unique eigenfunction $\theta \in L^2(\pi)$ for which $\mathcal{L}_\pi \theta = \lambda \theta$ and $\int \theta(x) \, d\pi(x) = 1$. We identify θ with its Lipschitz-continuous representative, which is a nondecreasing function on $[0, \infty)$ with $\theta(0) = 0$.*

PROOF. We begin with the proof of (i). $\mathcal{L}_\pi$ is self-adjoint because the kernel $x \wedge y$ is real and symmetric. To see that $\mathcal{L}_\pi$ is a positive semidefinite operator, note that for $f, g \in L^2(\pi)$ we have

$$\langle \mathcal{L}_\pi f, g \rangle_\pi = \int \int f(y) g(x) (x \wedge y) \, d\pi(x) \, d\pi(y)$$

$$= \int \int \int_0^\infty f(y) g(x) \mathbb{1}[x \geq u, y \geq u] \, du \, d\pi(x) \, d\pi(y)$$

$$\stackrel{(*)}{=} \int_0^\infty \left(\int_u^\infty f(y) \, d\pi(y) \right) \left(\int_u^\infty g(x) \, d\pi(x) \right) du.$$

The application of Fubini in the equation marked by $(*)$ is justified by the absolute integrability which follows from $\langle \mathcal{L}_\pi |f|, |g| \rangle_\pi \leq \|\mathcal{L}_\pi\| \|f\| \|g\| < \infty$, see (37). In particular, we have $\langle \mathcal{L}_\pi f, f \rangle_\pi = \int_0^\infty (\int_u^\infty f(x) \, d\pi(x))^2 \, du \geq 0$.

Next, we prove (ii). For any $f \in L^2(\pi)$, $\mathcal{L}_\pi f$ is represented by a Lipschitz function. Indeed, since π is a finite measure, the constant function $\mathbf{1}$ belongs to $L^2(\pi)$, and because $|(u \wedge s) - (u \wedge s')| \leq |s - s'|$, we have

$$|\mathcal{L}_\pi f(s) - \mathcal{L}_\pi f(s')| \leq \int |f(u)| |(u \wedge s) - (u \wedge s')| \, d\pi(u) \leq |s - s'| \langle |f|, \mathbf{1} \rangle_\pi.$$

Moreover, if f is a nonnegative function and $0 \leq s \leq s'$ then

$$0 \leq \mathcal{L}_\pi f(s) = \int f(u)(u \wedge s) \, d\pi(u) \leq \int f(u)(u \wedge s') \, d\pi(u) = \mathcal{L}_\pi f(s').$$

Next, we prove (iii). Because $\mathcal{L}_\pi$ is positive semidefinite and self-adjoint, and because Hilbert–Schmidt operators are compact, the spectrum of $\mathcal{L}_\pi$ is contained in the real interval

$[0, \|\mathcal{L}_\pi\|]$, with 0 being the only possible accumulation point and every positive eigenvalue having finite multiplicity. Except in the trivial case $\mathcal{L}_\pi = 0$, which only occurs when $\pi = \delta_0$, the Perron–Frobenius theorem for Hilbert–Schmidt integral operators (see [24], Section 6.5, or [5], Lemma 5.15) guarantees that the leading eigenvalue is $\lambda = \|\mathcal{L}_\pi\|$, and it is simple, that is, has a one-dimensional eigenspace.

Next we prove (iv). [5], Lemma 5.15, shows that there is an eigenfunction $\theta \in L^2(\theta)$ with eigenvalue λ such that θ is positive π-a.e. and since $\int \theta \, d\pi = \langle \theta, \mathbf{1} \rangle_\pi < \infty$ it is valid to normalize θ by multiplying it by a nonzero scalar so that $\int \theta \, d\pi = 1$. We have

$$(40) \qquad \theta(s) = \lambda^{-1} \mathcal{L}_\pi \theta(s) = \lambda^{-1} \int_0^\infty \theta(u)(u \wedge s) \, d\pi(u), \quad s \in [0, +\infty)$$

and we now choose the representative for θ defined by the right-hand side of (40), which is nondecreasing and Lipschitz in s by (ii), and satisfies $\theta(0) = 0$ and $\theta(s) > 0$ for $s > 0$. $\quad\square$

The proof of Lemma 2.9 immediately follows from Lemmas 4.1 and 4.3.

4.2. *Subcriticality, criticality and supercriticality of T^π.* Section 4.2 is devoted to the proof of Proposition 2.10. Note that we will only use the properties of $\mathcal{L}_\pi$ derived in Lemma 4.3, so our proof works if we only assume $\|\mathcal{L}_\pi\|_{\mathrm{HS}} < \infty$ and $\pi \neq \delta_0$ instead of $\int s \, d\pi(s) < +\infty$.

Denote by $\lambda = \|\mathcal{L}_\pi\|$ the principal eigenvalue of $\mathcal{L}_\pi$, see Lemma 4.3(iii). It follows from [5], Theorem 6.1, that $\|\mathcal{L}_\pi\|_{\mathrm{HS}} < \infty$ implies that $\lambda > 1$ if and only if $\mathbb{P}(|T^\pi| = \infty) > 0$. Therefore, it suffices to prove

$$(41) \qquad \begin{cases} \lambda < 1 & \Rightarrow \quad \mathbb{E}[|T^\pi|] < \infty, \\ \lambda = 1 & \Rightarrow \quad \mathbb{E}[|T^\pi|] = \infty. \end{cases}$$

Note that the expected number of vertices at distance k from the root of the tree T^π is given by $\langle \mathbf{1}, \mathcal{L}_\pi^k \mathbf{1} \rangle_\pi$, thus

$$(42) \qquad \mathbb{E}[|T^\pi|] = \sum_{k=0}^\infty \langle \mathbf{1}, \mathcal{L}_\pi^k \mathbf{1} \rangle_\pi.$$

Using this, we can prove that $\lambda < 1$ implies $\mathbb{E}[|T^\pi|] < \infty$:

$$(43) \qquad \mathbb{E}[|T^\pi|] \overset{(42)}{\leq} \sum_{k=0}^\infty \|\mathbf{1}\| \cdot \|\mathcal{L}_\pi^k \mathbf{1}\| \leq \sum_{k=0}^\infty \|\mathbf{1}\|^2 \|\mathcal{L}_\pi^k\| \leq \sum_{k=0}^\infty \langle \mathbf{1}, \mathbf{1} \rangle_\pi \lambda^k = \frac{1}{1 - \lambda}.$$

We now treat the critical case $\lambda = 1$. Recall from Lemma 4.3(iv) that in the $\lambda = 1$ case θ satisfies $\langle \mathbf{1}, \theta \rangle_\pi = 1$ and $\theta = \mathcal{L}_\pi \theta$. Let us write $\mathbf{1} = a\theta + f$, where $a = \langle \mathbf{1}, \theta \rangle_\pi / \langle \theta, \theta \rangle_\pi > 0$, so that $f \in \theta^\perp$. Using that $\mathcal{L}_\pi^k$ is self-adjoint for any $k \geq 0$ (cf. Lemma 4.3(i)), we obtain

$$(44) \qquad \mathbb{E}[|T^\pi|] \overset{(42)}{=} a^2 \sum_{k=0}^\infty \langle \theta, \mathcal{L}_\pi^k \theta \rangle_\pi + \sum_{k=0}^\infty \langle f, \mathcal{L}_\pi^k f \rangle_\pi,$$

where the first sum on the right-hand side is infinite since $\mathcal{L}_\pi^k \theta = \theta$, while the second sum is nonnegative, since $\mathcal{L}_\pi^k$ is positive semidefinite; cf. Lemma 4.3(i). It then follows that $\mathbb{E}[|T^\pi|] = \infty$. The proof of Proposition 2.10 is complete.

5. Multitype Galton–Watson trees and generating functions. In this section, we derive further properties of the multitype Galton–Watson trees T^π and T_s^π defined using a probability measure π on $[0, \infty)$ that satisfies $\int u \, d\pi(u) < \infty$, see Definition 2.8. Recall from (8) that we define the Perron–Frobenius operator $\mathcal{L}_\pi$ acting on $L^2(\pi)$ by

$$(45) \qquad \mathcal{L}_\pi f(x) = \int_0^\infty f(s)(x \wedge s) \, d\pi(s).$$

Recall from Definition 4.3(iv) the notion of the eigenfunction θ corresponding to the principal eigenvalue λ of $\mathcal{L}_\pi$. In the age-critical case (see Proposition 2.10), let us define

$$(46) \qquad \phi := \left(\int \theta(u)^3 \, d\pi(u) \right)^{-1}.$$

One of the main results of Section 5 is the next lemma.

LEMMA 5.1 (Critical generating function asymptotics). *Let π be an age-critical measure satisfying $\int u \, d\pi(u) < \infty$. Then on the intersection of the complex open unit disc $\mathbb{D}$ with a neighborhood of 1 the generating function $\mathbb{E}(z^{|T_s^\pi|})$ agrees with an analytic function of a branch of $\sqrt{1-z}$ that depends continuously on $s \in [0, \infty]$. In particular, as $z \nearrow 1$ through the reals, we have*

$$(47) \qquad \mathbb{E}(z^{|T_s^\pi|}) = 1 - \sqrt{2\phi}\theta(s)\sqrt{1-z} + O(1-z),$$

where the implied constant in the error term is uniform as a function of s, and also

$$(48) \qquad \mathbb{E}(z^{|T^\pi|}) = 1 - \sqrt{2\phi}\sqrt{1-z} + O(1-z).$$

Lemma 5.1 is an ingredient of the proof of Theorem 2.21; see Section 6.9.

In Section 5.1, we set up the notation required for the study of age-driven multitype branching processes using generating functions and write down a system of nonlinear equations satisfied by these generating functions. This leads to the study of a nonlinear Volterra equation of the second kind.

In Section 5.2, we prove that this nonlinear Volterra equation is well-posed.

In Section 5.3, we prove Lemma 5.1.

In Section 5.4, we prove Proposition 2.16.

The details of the proofs presented in Section 5 are independent of the rest of this paper, so the rest of Section 5 can be skipped at first reading.

5.1. *The generating function of the total progeny.* Recall the notion of T^π and T_s^π from Definition 2.8. In Section 5.1, we assume $\int u \, d\pi(u) < \infty$ and that π is either age-subcritical or age-critical (cf. Proposition 2.10).

Let us denote by $\overline{\mathbb{D}}$ the closed complex unit disc.

DEFINITION 5.2 (Generating function of $|T_s^\pi|$ and $|T^\pi|$). We write $[0, \infty]$ for the one-point compactification $[0, \infty) \cup \{\infty\}$. For $s \in [0, \infty]$ and $z \in \overline{\mathbb{D}}$, we define the probability generating functions

$$(49) \qquad f(s, z) := \mathbb{E}(z^{|T_s^\pi|}), \qquad f(z) := \mathbb{E}(z^{|T^\pi|}) = \int f(s, z) \, d\pi(s).$$

LEMMA 5.3 (Basic properties of $f(s, z)$).

(i) $f(s, z)$ *satisfies* $f(s, 1) = 1$ *for all* $s \in [0, \infty]$,
(ii) $f(s, z) \in [0, 1]$ *if* $z \in [0, 1]$ *and* $f(s, z) \in \overline{\mathbb{D}}$ *if* $z \in \overline{\mathbb{D}}$,

(iii) $f(s, z)$ *is strictly increasing in* $z \in [0, 1]$ *for all* $s \in [0, \infty]$,

(iv) *for each fixed* $s \in [0, \infty]$, $f(s, z)$ *is analytic over* $z \in \mathbb{D}$,

(v) $f(s, z)$ *is nonincreasing in s for all* $z \in [0, 1]$,

(vi) $f(s, z)$ *is uniformly continuous in* $(s, z) \in [0, \infty] \times \overline{\mathbb{D}}$.

PROOF. In order to prove (i), we only need to show $\mathbb{P}(|T_s^\pi| = +\infty) = 0$ for all $s \in [0, \infty]$. It is enough to show this just for $s = \infty$, since $|T_\infty^\pi|$ stochastically dominates $|T_s^\pi|$ for each $s \in [0, \infty]$. Note that it follows from $\int u \, d\pi(u) < \infty$ that the function $s \mapsto \mathbb{P}(|T_s^\pi| = +\infty)$ is nondecreasing and continuous on $[0, \infty]$. Also note that $\mathbb{P}(|T^\pi| = +\infty) = 0$ follows from Proposition 2.10 and our assumption that π is either age-subcritical or age-critical. It remains to observe that $\int \mathbb{P}(|T_s^\pi| = +\infty) \, d\pi(s) = \mathbb{P}(|T^\pi| = +\infty) = 0$ and (i) follows

(ii), (iii) and (iv) are standard facts about generating functions.

(v) follows from the fact that $|T_s^\pi|$ is stochastically increasing in s.

(vi) follows as soon as we observe that $\int u \, d\pi(u) < \infty$ implies that the law of T_s^π depends continuously on $s \in [0, \infty]$ (uniform continuity of f follows from continuity and the Heine-Cantor theorem, since $[0, \infty] \times \overline{\mathbb{D}}$ is compact). $\square$

LEMMA 5.4 (A system of nonlinear equations for $f(s, z)$). *For each* $s \in [0, \infty]$, $f(s, z)$ *satisfies the recursive equation*

$$(50) \qquad f(s, z) = z \exp \int (f(u, z) - 1)(u \wedge s) \, d\pi(u), \quad z \in \overline{\mathbb{D}}.$$

PROOF. Denote by λ_s the measure $d\lambda_s(a) = (s \wedge a) \, d\pi(a)$. The root of T_s has offspring of ages $a_1, \ldots, a_K$ which are the points of a Poisson point process on $[0, \infty)$ of intensity λ_s. Thus K is a Poisson random variable of mean $|\lambda_s| := \int d\lambda_s(a) < \infty$ and conditional on K, the ages $a_1, \ldots, a_K$ are independent with law $\lambda_s/|\lambda_s|$,

$$f(s, z) = \mathbb{E}(z^{|T_s^\pi|}) = \mathbb{E}\left(z \prod_{i=1}^K z^{|T_{a_i}^\pi|}\right)$$

$$= z \sum_{k=0}^\infty \frac{e^{-|\lambda_s|}|\lambda_s|^k}{k!} \prod_{i=1}^k \mathbb{E}_{a \sim \lambda_s/|\lambda_s|}(z^{|T_a^\pi|}) = z \sum_{k=0}^\infty \frac{e^{-|\lambda_s|}e_s^k}{k!}$$

$$= z \exp(e_s - |\lambda_s|), \quad \text{where } e_s = \int f(a, z) \, d\lambda_s(a). \qquad \square$$

In the case when π is age-subcritical, we will show that for each s, the generating function $f(s, z)$ has an analytic continuation on a small neighborhood of $z = 1$. On the other hand, in the case when π is age-critical, we will show that $f(s, z)$ agrees with an analytic function of $\sqrt{1 - z}$ in the intersection of $\mathbb{D}$ and a neighborhood of 1.

Let us denote by $B_\delta(z_0) = \{z \in \mathbb{C} : |z - z_0| \leq \delta\}$ the disc of radius δ centered at z_0 in the complex plane.

LEMMA 5.5 (A system of nonlinear equations for $\log f(s, z)$). *There exists* $\delta > 0$ *such that for any* $s \in [0, \infty]$ *and* $z \in \mathbb{D} \cap B_\delta(1)$ *we have* $|\log f(s, z)| \leq 1$ *and*

$$(51) \qquad \log f(s, z) = \log f(\infty, z) - \int_s^\infty (e^{\log f(u, z)} - 1)(u - s) \, d\pi(u).$$

PROOF. In order to achieve $|\log f(s, z)| \leq 1$, we want to choose $\delta > 0$ such that $f(s, z)$ falls in an appropriate small neighborhood of 1 for all $s \in [0, \infty]$ and $z \in \mathbb{D} \cap B_\delta(1)$. Such a choice is possible by statements (i) and (vi) of Lemma 5.3.

We can thus rewrite (50) in terms of the function $\log f(s, z)$ as

$$\log f(s, z) = \log z + \int (e^{\log f(u,z)} - 1)(u \wedge s)\, d\pi(u).$$

In particular, substituting $s = \infty$ we obtain

$$\log f(\infty, z) = \log z + \int (e^{\log f(u,z)} - 1)u\, d\pi(u).$$

Taking the difference of the last two equations and using $u - (u \wedge s) = (u - s)_+$ we obtain (51). $\square$

We will use (51) to show that $f(s, z)$ is an analytic function of $f(\infty, z)$. This will help us in showing that $f(s, z)$ is analytic in z in a small neighborhood of 1 in the age-subcritical case and, more importantly, we will use $w = \log f(\infty, z)$ as a local coordinate to resolve the algebraic singularity of $f(s, \cdot)$ at $z = 1$ in the age-critical case. The next lemma explains how to find $\log f(s, z)$ in terms of w. The continuity statements are given w.r.t. the one-point compactification $[0, \infty]$: a function f defined on $[0, \infty]$ is continuous when its restriction to $[0, \infty)$ is continuous and $\lim_{x \to \infty} f(x) = f(\infty)$.

LEMMA 5.6 (A nonlinear Volterra equation). *There exists $\delta > 0$ such that for each $w \in B_\delta(0) \subset \mathbb{C}$, there exists a unique bounded solution $F(\cdot, w) : [0, \infty] \to \mathbb{C}$ of the equation*

$$(52) \qquad F(s, w) = w - \int_s^\infty (e^{F(u,w)} - 1)(u - s)\, d\pi(u).$$

We have $F(s, 0) = 0$ for all s. The solution depends continuously on $(s, w) \in [0, \infty] \times B_\delta(0)$. For each fixed s, the solution depends analytically on w. When we consider the Taylor series of F as a function of w, each Taylor coefficient is a bounded and continuous function of $s \in [0, \infty]$.

Equation (52) is a nonlinear Volterra equation of the second kind, posed as a final value problem. It is somewhat nonstandard because the integral is with respect to π rather than Lebesgue measure, and the assumption that π has finite mean will be crucial. We give a detailed proof to keep our argument self-contained, and because it would take just as much space to verify that the equation can be transformed into a standard form satisfying suitable hypotheses to guarantee the conclusions of the lemma.

5.2. *The nonlinear Volterra equation is well-posed.* The goal of Section 5.2 is to prove Lemma 5.6.

5.2.1. *Proof of existence.* We begin by proving the existence of a solution of (52) with the required properties. Let us now briefly outline our construction. We break $[0, \infty)$ into finitely many intervals of form $[\alpha_i, \alpha_{i-1}]$, where

$$(53) \qquad \alpha_k = 0 < \alpha_{k-1} < \cdots < \alpha_1 < \alpha_0 = \infty$$

and solve equation (52) piecewise for $i = 0, 1, 2, \ldots$, starting at ∞. On each interval, we use a Picard iteration to solve a final value problem with respect to s. We will show that on each interval $[\alpha_i, \alpha_{i-1}]$ all the iterates depend analytically on w, and are continuous on $[\alpha_i, \alpha_{i-1}] \times B_{\delta_i}(0)$ for some $\delta_i > 0$. We will also show that the iterates converge uniformly. It follows that there exists $\delta > 0$ such that the limit F of the iteration is analytic in $w \in B_\delta(0)$ for each $s \in [0, \infty]$ and continuous in $(s, w) \in [0, \infty] \times B_\delta(0)$, and that for each fixed value of $w \in B_\delta(0)$, $F(\cdot, w)$ solves (52).

Let us now start to carry this plan out in detail. Also note that $F(\infty, w) = w$, since π does not have an atom at ∞.

To define the pieces of the iteration, recall that $\int u \, d\pi(u) < \infty$, so we can choose $k \in \mathbb{N}$ and a finite sequence $\alpha_0, \ldots, \alpha_k$ satisfying (53) such that for each $i = 1, \ldots, k$, we have

$$(54) \qquad \int_{\alpha_i}^{\alpha_{i-1}} (u - \alpha_i) \, d\pi(u) < \frac{1}{2}.$$

We will prove by induction over i that for each i there exists $\delta_i > 0$ such that there is a solution $F(\cdot, w) : [\alpha_i, \infty] \to \mathbb{C}$ with the continuity (on $[\alpha_i, \infty]$) and analyticity (on $B_{\delta_i}(0)$) required by Lemma 5.6.

For $i = 0$, equation (52) simply tells us that $F(\infty, w) = w$.

For $i \geq 1$, suppose that we have already a solution on $[\alpha_{i-1}, \infty]$, continuous in (s, w) and analytic in w. Consider $s \in [\alpha_i, \alpha_{i-1}]$. Applying (52) to both $F(s, w)$ and $F(\alpha_{i-1}, w)$ and subtracting, we obtain

$$(55) \qquad \begin{aligned} F(s, w) = {} & F(\alpha_{i-1}, w) - (\alpha_{i-1} - s) \int_{\alpha_{i-1}}^{\infty} \left(e^{F(u,w)} - 1 \right) d\pi(u) \\ & - \int_{s}^{\alpha_{i-1}} \left(e^{F(u,w)} - 1 \right)(u - s) \, d\pi(u). \end{aligned}$$

Note that if $i = 1$ then $\alpha_{i-1} = \infty$ by (53), so the second term on the right-hand side of (55) is zero, and thus for $i = 1$ (55) just becomes (52).

We will solve (55) for $s \in [\alpha_i, \alpha_{i-1}]$ and w in some ball $B_{\delta_i}(0)$, where we may choose $\delta_i \in (0, \delta_{i-1}]$. The first two terms on the right-hand side of (55) are analytic in w and are jointly continuous in s and w by the induction hypothesis. Only the left-hand side and the final term on the right-hand side depend on the restriction of F to $[\alpha_i, \alpha_{i-1}]$. We solve equation (55) by the Picard iteration. Define F_0 by

$$(56) \qquad F_0(s, w) = F(\alpha_{i-1}, w), \quad (s, w) \in [\alpha_i, \alpha_{i-1}] \times B_{\delta_i}(0).$$

Then for $j \geq 1$ define inductively

$$(57) \qquad \begin{aligned} F_j(s, w) = {} & F(\alpha_{i-1}, w) - (\alpha_{i-1} - s) \int_{\alpha_{i-1}}^{\infty} \left(e^{F(u,w)} - 1 \right) d\pi(u) \\ & - \int_{s}^{\alpha_{i-1}} \left(e^{F_{j-1}(u,w)} - 1 \right)(u - s) \, d\pi(u), \quad s \in [\alpha_i, \alpha_{i-1}]. \end{aligned}$$

Analogously to (55), if $i = 1$ then we define the second term on the right-hand side of (57) to be equal to zero.

Denote by $\| \cdot \|_{\infty, i}$ the infinity norm on the domain $[\alpha_i, \alpha_{i-1}] \times B_{\delta_i}(0)$, where δ_i will be specified later. Our next goal is to prove the following lemma.

LEMMA 5.7. *With the above notation, we have*

$$(58) \qquad \|F_{j+1} - F_j\|_{\infty, i} \leq \frac{1}{2} e^{\max(\|F_j\|_{\infty, i}, \|F_{j-1}\|_{\infty, i})} \|F_j - F_{j-1}\|_{\infty, i}, \quad j \geq 1.$$

PROOF. For any $s \in [\alpha_i, \alpha_{i-1}]$ and $w \in B_{\delta_i}(0)$, we have

$$(59) \qquad \begin{aligned} & \left| F_{j+1}(s, w) - F_j(s, w) \right| \\ & \overset{(57)}{\leq} \int_{s}^{\alpha_{i-1}} \left| e^{F_j(u,w)} - e^{F_{j-1}(u,w)} \right|(u - s) \, d\pi(u) \\ & \overset{(*)}{\leq} \int_{s}^{\alpha_{i-1}} e^{\max(|F_j(u,w)|, |F_{j-1}(u,w)|)} \left| F_j(u, w) - F_{j-1}(u, w) \right|(u - s) \, d\pi(u) \\ & \leq e^{\max(\|F_j\|_{\infty, i}, \|F_{j-1}\|_{\infty, i})} \|F_j - F_{j-1}\|_{\infty, i} \int_{s}^{\alpha_{i-1}} (u - \alpha_i) \, d\pi(u), \end{aligned}$$

where in (∗) we used that if $x_1, x_2 \in \mathbb{C}$ and $|x_1| \leq a$, $|x_2| \leq a$ both hold then $|e^{x_1} - e^{x_2}| \leq e^a |x_1 - x_2|$ also holds (this inequality follows from the fact that $|\frac{d}{dz} e^z| \leq e^{|z|}$). By (59) and our assumption (54), we obtain (58). □

Lemma 5.7 implies that we will have a strict contraction in $\| \cdot \|_{\infty, i}$ as long as we can keep control of the norms $\|F_j\|_{\infty, i}$. This is what we show in the next lemma.

LEMMA 5.8. *One can choose $\delta_i > 0$ sufficiently small so that*

$$\|F_0\|_{\infty, i} + 3\|F_1 - F_0\|_{\infty, i} \leq 1/4, \tag{60}$$

$$\|F_j\|_{\infty, i} \leq 1/4, \quad j = 0, 1, \ldots, \tag{61}$$

$$\|F_j - F_{j-1}\|_{\infty, i} \leq \left(\frac{2}{3}\right)^{j-1} \|F_1 - F_0\|_{\infty, i}, \quad j = 1, 2, \ldots. \tag{62}$$

PROOF. First, we show (60). We have

$$\begin{aligned}
&F_1(s, w) - F_0(s, w) \\
&\stackrel{(56),(57)}{=} (\alpha_{i-1} - s) \int_{\alpha_{i-1}}^{\infty} \left(e^{F(u, w)} - 1\right) d\pi(u) \\
&\quad - \left(e^{F(\alpha_{i-1}, w)} - 1\right) \int_s^{\alpha_{i-1}} (u - s) d\pi(u), \quad s \in [\alpha_i, \alpha_{i-1}].
\end{aligned} \tag{63}$$

Recall from below (57) that if $i = 1$ then $\alpha_{i-1} = \infty$ and we define the first term on the right-hand side of (63) to be equal to zero. Our induction hypothesis implies that F is jointly continuous in $(s, w) \in [\alpha_{i-1}, \infty] \times B_{\delta_{i-1}}(0)$; moreover, we have $F_0(s, 0) = 0$ by the definition of F_0 in (56) and our induction hypothesis (which implies $F(\alpha_{i-1}, 0) = 0$). Putting these observations together with (54) and (63), we see that we can choose (and fix) $\delta_i > 0$ sufficiently small to ensure that (60) holds.

We will prove (61) and (62) simultaneously by induction on j. First, observe that (60) implies (61) for $j = 0$ and $j = 1$. Also, (62) obviously holds for $j = 1$. Now let $j \geq 1$ and suppose that (61) and (62) hold for all indices up to j. We have

$$\|F_{j+1} - F_j\|_{\infty, i} \stackrel{(58),(61)}{\leq} \frac{1}{2} e^{1/4} \|F_j - F_{j-1}\|_{\infty, i} \leq \frac{2}{3} \|F_j - F_{j-1}\|_{\infty, i},$$

thus (62) holds for $j + 1$. Next, we observe that

$$\|F_{j+1} - F_0\|_{\infty, i} \leq \sum_{k=1}^{j+1} \|F_k - F_{k-1}\|_{\infty, i} \stackrel{(62)}{\leq} 3\|F_1 - F_0\|_{\infty, i}.$$

This, together with (60) implies that (61) holds for $j + 1$, completing the induction step. □

Now we can finish our construction: since (62) implies that F_j is a Cauchy sequence with respect to $\| \cdot \|_{\infty, i}$, so it converges. By considering the limit as $j \to \infty$ of both sides of (57), we see that the limit function F solves (55) on $[\alpha_i, \alpha_{i-1}] \times B_{\delta_i}(0)$. Since each F_j is jointly continuous in (s, w), the uniform limit F also has this property. Since each F_j is analytic with respect to w on $B_{\delta_i}(0)$, the uniform limit F is also analytic by Morera's theorem. It is easy to see by induction on j that $F_j(s, 0) = 0$ for all $s \in [\alpha_i, \alpha_{i-1}]$, so F also has this property.

By piecing together the solutions for $i = 1, \ldots, k$, and taking $\delta = \delta_k$ we obtain a solution of (55) defined on $[0, \infty] \times B_\delta(0)$ that is continuous in (s, w) and analytic in w for each s. Note that $F(\cdot, w)$ is also bounded, since it is continuous on the compact space $[0, \infty]$. Since the Taylor coefficients of F about $w = 0$ can be expressed by the Cauchy integral formula, the joint continuity of F implies the continuity of the Taylor coefficients as functions of $s \in [0, \infty]$. The Taylor coefficients of F are also bounded, since the space $[0, \infty]$ is compact.

5.2.2. *Proof of uniqueness.* It remains to show the uniqueness of the solution of (52) with the required properties. Suppose that a value $w \in B_\delta(0)$ is fixed, and that $\tilde{F}(\cdot, w)$ is another bounded solution of (52) defined on $[0, \infty]$. We may assume that both $\tilde{F}(\cdot, w)$ and $F(\cdot, w)$ are bounded in absolute value by $M < \infty$. Then for every s we have

$$\left| F(s, w) - \tilde{F}(s, w) \right| \overset{(52)}{\leq} \int_s^\infty \left| e^{F(u,w)} - e^{\tilde{F}(u,w)} \right| (u - s) \, d\pi(u)$$

$$\overset{(*)}{\leq} \int_s^\infty \left| F(u, w) - \tilde{F}(u, w) \right| e^M u \, d\pi(u),$$

where in $(*)$ we used that if $x_1, x_2 \in \mathbb{C}$ and $|x_1| \leq M$, $|x_2| \leq M$ both hold then $|e^{x_1} - e^{x_2}| \leq |x_1 - x_2| e^M$ also holds. Since $e^M u \, d\pi(u)$ is a finite measure, Grönwall's inequality implies that $\tilde{F}(\cdot, w) \equiv F(\cdot, w)$.

The proof of Lemma 5.6 is complete.

5.3. *Taylor coefficients.* The goal of Section 5.3 is to prove Lemma 5.1, therefore, we assume that π is age-critical in Section 5.3. Recall from Lemma 4.3(iv) that $\mathcal{L}_\pi \theta = \theta$, that is,

$$(64) \qquad \theta(s) = \int \theta(u)(u \wedge s) \, d\pi(u)$$

and $\int \theta \, d\pi = 1$ and $\theta(0) = 0$. Let us define

$$(65) \qquad \theta(\infty) := \lim_{s \to \infty} \theta(s) \overset{(64)}{=} \int \theta(u) u \, d\pi(u).$$

Let us recall the notion of the function F from Lemma 5.6.

LEMMA 5.9 (Taylor coefficients). *Suppose that π is age-critical. Then $0 < \theta(\infty) < \infty$ and the Taylor expansion of F about $w = 0$ is*

$$(66) \qquad F(s, w) = \frac{\theta(s)}{\theta(\infty)} w + A_2(s) w^2 + O(w^3),$$

where

$$(67) \qquad A_2(0) = -\frac{\int \theta(u)^3 \, d\pi(u)}{2\theta(\infty)^2} < 0.$$

PROOF. $F(s, 0) \equiv 0$, so the constant term in the Taylor series vanishes identically. The coefficient of w is $A_1(s) := F_w(s, 0) := \cdot \frac{\partial}{\partial w} F(s, w)|_{w=0}$. By Lemma 5.6, $s \mapsto F_w(s, 0)$ is bounded and continuous on $[0, \infty]$. To compute it, we differentiate both sides of (52) and evaluate at $w = 0$,

$$F_w(s, 0) = 1 - \int_s^\infty F_w(u, 0) e^{F(u,0)} (u - s) \, d\pi(u)$$

$$(68) \qquad = 1 - \int_s^\infty F_w(u, 0)(u - s) \, d\pi(u)$$

$$= 1 - \int_0^\infty F_w(u, 0) u \, d\pi(u) + \int_0^\infty F_w(u, 0)(u \wedge s) \, d\pi(u).$$

Denote by $\mathbf{1} : [0, \infty] \to \mathbb{R}$ the function $\mathbf{1}(s) := 1$. From (69), we obtain

$$(69) \qquad (I - \mathcal{L}_\pi) F_w(\cdot, 0) = \left(1 - \int_0^\infty F_w(u, 0) u \, d\pi(u) \right) \mathbf{1}.$$

(Note that the integrals in (69) and (69) converge since the first moment of π is finite). Take the inner product of both sides of (69) in $L^2(\pi)$ with θ. Since $I - \mathcal{L}_\pi$ is self-adjoint and annihilates θ, the left-hand side vanishes. Therefore,

$$(70) \qquad 0 = \left(1 - \int_0^\infty F_w(u, 0) u \, d\pi(u)\right) \int \theta(u) \, d\pi(u).$$

Now we can use $\int \theta(u) \, d\pi(u) = 1$ to conclude

$$(71) \qquad \int_0^\infty F_w(u, 0) u \, d\pi(u) = 1.$$

Substituting (71) into (69), we obtain $F_w(\cdot, 0) = \mathcal{L}_\pi F_w(\cdot, 0)$, so by Lemma 4.3(iv) we must have $F_w(\cdot, 0) = c\theta(\cdot)$ for some $c \in \mathbb{R}$. We have already noted that $F_w(\cdot, 0)$ is bounded and continuous on $[0, \infty]$, thus $\theta(\cdot)$ is bounded and

$$(72) \qquad 1 \overset{(71)}{=} c \int_0^\infty \theta(u) u \, d\pi(u) \overset{(65)}{=} c\theta(\infty).$$

Hence $0 < \theta(\infty) < \infty$ and $c = 1/\theta(\infty)$. Thus the first-order Taylor coefficient in (66) is indeed $A_1(s) = \theta(s)/\theta(\infty)$.

It remains to prove (67). Differentiating both sides of (52) twice with respect to w and substituting $w = 0$, we find

$$A_2(s) = -\int_s^\infty (A_2(u) + A_1(u)^2)(u - s) \, d\pi(u)$$

$$= -\int_0^\infty \left(A_2(u) + \frac{\theta(u)^2}{2\theta(\infty)^2}\right)(u - u \wedge s) \, d\pi(u).$$

Hence

$$(I - \mathcal{L}_\pi) A_2(s) - \frac{\mathcal{L}_\pi \theta^2(s)}{2\theta(\infty)^2} = -\int_0^\infty \left(A_2(u) + \frac{\theta(u)^2}{2\theta(\infty)^2}\right) u \, d\pi(u).$$

The right-hand side of this equation does not depend on s, so it must be some constant c'. Therefore,

$$(73) \qquad (I - \mathcal{L}_\pi) A_2 = c' \mathbf{1} + \frac{1}{2\theta(\infty)^2} \mathcal{L}_\pi \theta^2.$$

To find c', take the inner product of both sides of (73) with θ:

$$(74) \qquad \begin{aligned} 0 &= \langle (I - \mathcal{L}_\pi)\theta, A_2 \rangle_\pi \\ &= \langle \theta, (I - \mathcal{L}_\pi) A_2 \rangle_\pi \\ &\overset{(73)}{=} \langle \theta, c'\mathbf{1} \rangle_\pi + \langle \theta, \mathcal{L}_\pi \theta^2 \rangle_\pi / (2\theta(\infty)^2) \\ &= c' \langle \theta, \mathbf{1} \rangle_\pi + \langle \mathcal{L}_\pi \theta, \theta^2 \rangle_\pi / (2\theta(\infty)^2) \\ &= c' + \langle \theta, \theta^2 \rangle_\pi / (2\theta(\infty)^2) \end{aligned}$$

Substituting $s = 0$ into both sides of (73) and using $(\mathcal{L}_\pi A_2)(0) = 0$ and $\theta(0) = 0$ we obtain $A_2(0) = c'$ and (67) follows. □

PROOF OF LEMMA 5.1. Recall the plan outlined in the paragraph before the statement of Lemma 5.6. We have the following expression for $\log z$ as an analytic function of $w = \log f(\infty, z)$, valid in some disc around $w = 0$:

$$(75) \qquad \log z \overset{(49)}{=} \log f(0, z) \overset{(*)}{=} F(0, w) \overset{(**)}{=} A_2(0) w^2 + O(w^3),$$

where in $(*)$ we used Lemmas 5.5 and 5.6; moreover, in $(**)$ we used Lemma 5.9 and $\theta(0) = 0$. From (75), we obtain $1 - z = -A_2(0)w^2 + O(w^3)$, hence w has an expansion around $z = 1$ as a convergent power series in $(1 - z)^{1/2}$:

$$(76) \qquad w = \frac{1}{\sqrt{-A_2(0)}}(1 - z)^{1/2} + a_2(1 - z) + a_3(1 - z)^{3/2} + \cdots.$$

Recall from (46) that we defined $\phi := 1/\int \theta(u)^3 \, d\pi(u)$. Note that we have $\log f(\infty, z) < 0$ if $0 \leq z < 1$, which determines the correct choice of branch of square root in the above expansion. Let $\sqrt{1 - z}$ denote the branch of $(1 - z)^{1/2}$ that takes positive values for $0 \leq z < 1$. Then we have

$$(77) \quad \log f(s, z) = F(s, w) \stackrel{(66)}{=} \frac{\theta(s)}{\theta(\infty)}w + O(w^2) \stackrel{(67),(76)}{=} -\sqrt{2\phi}\theta(s)\sqrt{1 - z} + O(1 - z).$$

Taking the exponential of both sides of (77), we obtain the desired (47). The implied constant in the error term is uniform as a function of s, since the Taylor coefficients of F are bounded by Lemma 5.6.

Now the proof of (48) follows from (47) using (49), $\int d\pi(s) = 1$ and $\int \theta(s) \, d\pi(s) = 1$. The proof of Lemma 5.1 is complete. $\square$

5.4. *Critical forest fire equations.* The goal of Section 5.4 is to prove Proposition 2.16. In the proof, we will use generating functions. In fact, we will prove a more general result where Assumption 2.15 is replaced with an assumption on the initial state $\underline{v}(0)$ which is formulated entirely in terms of generating functions.

DEFINITION 5.10 (Analytic inverse of generating function). Let $\underline{v} = (v_k)_{k=1}^{\infty}$ satisfy $v_k \geq 0$ and $\sum_{k=1}^{\infty} v_k = 1$. We denote by $f(z) = \sum_{k=1}^{\infty} v_k z^k$ the generating function of $\underline{v}$. Noting that $f(0) = 0$, $f(1) = 1$ and f is strictly increasing on $[0, 1]$, the inverse function f^{-1} of f is well defined on $[0, 1]$ and satisfies $f^{-1}(1) = 1$ and $(f^{-1})'(1) \geq 0$. We say that f has an analytic inverse if there exists some $\delta > 0$ and a complex analytic function $g : B_\delta(1) \to \mathbb{C}$, satisfying either $g'(1) > 0$ or both $g'(1) = 0$ and $g''(1) < 0$, such that

$$(78) \qquad g(z) = f^{-1}(z), \quad z \in [1 - \delta, 1].$$

In other words, the generating function f of $\underline{v}$ has the analytic inverse property if f^{-1} can be extended analytically to $B_\delta(1)$ for some $\delta > 0$; moreover, the second-order Taylor polynomial of f^{-1} about $z = 1$ is not identically zero.

REMARK 5.11. Assume that $\underline{v}$ satisfies Definition 5.10.

If $g'(1) > 0$, then we have $\sum_{k=1}^{\infty} kv_k = f'(1) = 1/g'(1)$ and f is analytic in a small neighborhood of $z = 1$, and thus v_k decays exponentially as $k \to \infty$ by the Vivanti–Pringsheim theorem.

If $g'(1) = 0$ and $g''(1) < 0$ hold, then f cannot be extended analytically to a small neighborhood of $z = 1$, and by Example (c) of Theorem 4 of Chapter XIII.5 of [17] we have $\sum_{\ell=k}^{\infty} v_\ell \approx \sqrt{\frac{-2}{\pi g''(1)}}k^{-1/2}$ as $k \to \infty$.

The next result is a modification of [27], Theorem 1, (but neither result follows from the other one).

PROPOSITION 5.12 (Critical forest fire equations). *Let us assume that the initial condition $\underline{v}(0) = (v_k(0))_{k=1}^{\infty}$ satisfies Definition 5.10. Then the critical forest fire equations ((12) + (13)) have a unique solution $\underline{v}(\cdot)$, which also satisfies the properties listed in Proposition 2.16.*

We will prove Proposition 5.12 in Section 5.4.1. Proposition 2.16 immediately follows from the result of Proposition 5.12 as soon as we prove the next lemma.

LEMMA 5.13 (The law of $|T^\pi|$ satisfies the analytic inverse property). *Suppose that $\int u\,d\pi(u) < \infty$ holds and π is either age-critical or age-subcritical. Let us define $\underline{v} = (v_k)_{k=1}^\infty$ by $v_k := \mathbb{P}(|T^\pi| = k)$, $k = 1, 2, \dots$. Under these conditions, $\underline{v}$ satisfies Definition 5.10. Moreover, if π is age-subcritical then $g'(1) > 0$ and if π is age-critical then $g'(1) = 0$ and $g''(1) < 0$.*

In other words, if $\underline{v}(0)$ satisfies Assumption 2.15 then it also satisfies Definition 5.10.

PROOF OF LEMMA 5.13. Let us recall the notion of $f(z)$ and $f(s, z)$ from (49) and $F(s, w)$ from Lemma 5.6. We will construct the analytic inverse function g of f required by Definition 5.10 by letting $g = G \circ H^{-1}$, where

$$(79) \qquad G(z) := e^{F(0,\log(z))}, \qquad H(z) := \int e^{F(s,\log(z))}\,d\pi(s).$$

Note that $G(1) = 1$ and $H(1) = 1$; moreover, G and H are complex analytic functions in a neighborhood of $z = 1$. By Lemmas 5.5 and 5.6, we have

$$(80) \qquad f(s, z) = e^{F(s,\log(f(\infty,z)))}, \qquad f(0, z) \overset{(49)}{=} z = e^{F(0,\log(f(\infty,z)))},$$

therefore for any $z \in (1 - \epsilon, 1]$, where $\epsilon > 0$ is small enough, we have

$$(81) \qquad G(f(\infty, z)) = z, \qquad f(z) \overset{(49)}{=} H(f(\infty, z)).$$

We want H to be invertible in a small neighborhood of 1, therefore our next goal is to show that $0 < H'(1) < \infty$ holds. First, note that $H'(1) = \int F_w(s, 0)\,d\pi(s)$ follows from the definition of H. Next, note that $F(s, 0) = 0$ and $F(s, w) < 0$ if $w < 0$ by (80), thus $F_w(s, 0) \geq 0$. Also note that $\sup_{s \geq 0} F_w(s, 0) < \infty$ by Lemma 5.6, thus $H'(1) < \infty$. Finally, we have to rule out the possibility that $H'(1) = 0$. Let us denote $A_1(s) = F_w(s, 0)$ and observe that $A_1(s) = 1 - \int_s^\infty A_1(u)(u - s)\,d\pi(u)$ holds by (69) (even if π is age-subcritical). Differentiating this w.r.t. s, we obtain $A_1'(s) = \int_s^\infty A_1(u)\,d\pi(u)$. Now $A_1(\infty) = 1$ and $A_1(s) \geq 0$, so if we indirectly assume $H'(1) = \int_0^\infty A_1(u)\,d\pi(u) = 0$, then $A_1'(s) \equiv 0$, so $A_1(s) \equiv A_1(\infty)$, which contradicts $\int_0^\infty A_1(u)\,d\pi(u) = 0$. This completes the proof of $0 < H'(1) < \infty$, which implies that H has a complex analytic inverse function H^{-1} in a small neighborhood of $z = 1$, so the desired inverse function $g(\cdot)$ of $f(\cdot)$ can be defined as $g = G \circ H^{-1}$ by (81); moreover, g is complex analytic in a small neighborhood of $z = 1$.

Note that $f'(1) = \sum_{k=1}^\infty k v_k = \mathbb{E}(|T^\pi|)$.

If π is age-subcritical, then $f'(1) < \infty$ by Proposition 2.10, thus $g'(1) > 0$.

If π is age-critical, then $f'(1) = \infty$ by Proposition 2.10, thus $g'(1) = 0$. It remains to show that $g''(1) < 0$. One can check that $G'(1) = 0$ and thus $g''(1) = G''(1)/(H'(1))^2$. Noting that by Lemma 5.9 we have $H'(1) = 1/\theta(\infty)$, we can therefore write $g''(1) = G''(1)\theta^2(\infty)$. We can then use $F_w(0, 0) = A_1(0) = \theta(0)/\theta(\infty) = 0$ to deduce $G''(1) = 2A_2(0)$, whence $g''(1) = -1/\phi < 0$ by (46) and (67). $\qquad\square$

5.4.1. *Analytic inverse property implies well-posedness.* PROOF OF PROPOSITION 5.12. Let $f_0(z) = \sum_{k=1}^\infty v_k(0)z^k$. The condition that $\sum_{k=1}^\infty k v_k(0) < \infty$ is equivalent to $f_0'(1) < \infty$, which is in turn equivalent to $g'(1) > 0$, where g is the inverse function of f_0.

If $g'(1) > 0$, then both f_0 and g are invertible complex analytic functions in a small neighborhood of 1. In particular, this implies that $f_0'''(1) < \infty$ and thus $\sum_{k=1}^\infty k^3 v_k(0) < \infty$, hence we can apply [27], Theorem 1, to conclude the proof of Proposition 5.12.

The condition $g'(1) = 0$ is equivalent to $\sum_{k=1}^{\infty} k v_k(0) = \infty$, which is in turn equivalent to $t_{\text{gel}} = 0$ (see (15)). Recall from [27] that we define $V_0(x) = f_0(e^{-x}, (41)) - 1$ for any $x \in [0, \infty)$. Recalling [27], (57), we define $E(x) = -V_0'(x)^3 / V_0''(x)$. In [27], Section 4.2, Section 4.3, the uniqueness of the equations ((12) + (13)) and all of the other conclusions of Proposition 2.16 are proved under the assumptions

$$(82) \qquad \lim_{x \to 0_+} E(x) \in (0, \infty), \qquad E'(x) = O(x^{-1/2}) \quad \text{as } x \to 0_+$$

on the initial condition $(v_k(0))_{k=1}^{\infty}$, thus we only need to show that the assumptions of Proposition 5.12 imply (82). Indeed: denote the inverse function of $-V_0(x)$ by $X(u)$, that is, $X(-V_0(x)) = x$, $x \in [0, \infty)$. By [27], (63), we have

$$(83) \qquad E(x) = 1/X''(-V_0(x)).$$

Using the above definitions, we can write $X(u) = -\log(g(1-u))$, thus we can use $g'(1) = 0$ to derive

$$(84) \qquad \lim_{x \to 0_+} E(x) = 1/X''(0) = -1/g''(1) \in (0, \infty)$$

using the assumption of Proposition 5.12. Thus the first statement of (82) holds. Differentiating (83) with respect to x, we obtain

$$(85) \qquad E'(x) = E(x)^2 X'''(-V_0(x)) V_0'(x).$$

Now $\lim_{x \to 0_+} E(x)^2 X'''(-V_0(x))$ exists and is finite by the assumption that g is analytic near 1, and $|V_0'(x)| = O(x^{-1/2})$ follows from $t_{\text{gel}} = 0$, (84) and [27], Lemma 3, (61). This concludes the proof of (82). $\square$

Putting together the results of Proposition 5.12 and Lemma 5.13, we obtain the proof of Proposition 2.16.

6. Age evolution.

In Section 6, we prove our main results stated in Section 2.5. In order to do so, we need to introduce some auxiliary Markov processes. In Section 6.1, below we will provide the detailed definitions of these processes and state some key lemmas about them. Let us now give a brief outline of the contents of Section 6.1.

We first recall from [11] the cluster growth process (C_t) driven by a solution to the forest fire equations ((12) + (13)) and augment it with its age process (a_t) to obtain the *cluster process with age* (a_t, C_t) (see Definition 6.1). We will identify the probability distribution π_t that appears in the statements of our main results (Theorems 2.19, 2.20 and 2.21) as the marginal distribution of a_t.

The dynamics of the process (C_t) are approximated by the dynamics of the component size $(C_t^n(\rho^n))$ of a uniformly chosen vertex ρ^n in the MFFF when $n \gg 1$. The two processes can be coupled in a way that with high probability they agree exactly except close to their burning or explosion times; see Theorem 6.4.

In Theorem 6.5, we will upgrade the result of Theorem 6.4 and couple the cluster processes $(C_t^n(\rho_1^n), C_t^n(\rho_2^n))$ of i.i.d. uniform vertices ρ_1^n and ρ_2^n to a pair $(C_t^{(1)}, C_t^{(2)})$ of i.i.d. copies of (C_t). In Proposition 6.7, we extend the coupling of Theorem 6.5 to the setting where we keep track of the ages of the tagged vertices as well as the size of their clusters. In Corollary 6.8, we conclude that, for any fixed t, $(a_t^n(\rho_i^n), C_t^n(\rho_i^n))_{i \in \{1,2\}}$ weakly converge to i.i.d. copies $(a_t^{(i)}, C_t^{(i)})_{i \in \{1,2\}}$ of the joint distribution at time t of the age a_t and cluster size C_t in the cluster process with age.

We prove concentration of the empirical age measure π_t^n around π_t using Corollary 6.8 and a second moment argument, which gives the proof of Theorem 2.19. The proof of the

local weak limit statements of Theorem 2.20 will also follow using the fact that given the ages, the graph that we see in the MFFFA at time t is an age-driven IRG.

In Lemma 6.9, we show that conditional on $(a_s)_{0 \le s \le t}$, the law of C_t agrees with the size of the multitype Galton–Watson tree $T_{a_t}^{\pi_t}$ where the root has age a_t. Finally, we show in Lemma 6.12 that the process $(a_t)_{t \ge 0}$ itself has the Markov property, and the proof of Theorem 2.21 will follow as soon as we look at the Kolmogorov forward equations of $(a_t)_{t \ge 0}$ the right way.

6.1. *Cluster growth processes with age.*

We extend the definition of the cluster growth process associated with the forest fire model from [11], Definition 1.6, by augmenting it with the age of the root vertex. Denote by $E = (\mathbb{N}, d_E)$ the metric space with metric $d_E(i, j) = |f(i) - f(j)|$, where $f(i) = 1/i$ for $i \in \{2, 3, 4, \dots\}$ but $f(1) = 0$. Thus E is a compact metric space in which $\lim_{n \to \infty} n = 1$, and 1 is the only accumulation point of E. A function $g : E \to \mathbb{R}$ is continuous if and only if $\lim_{k \to \infty} g(k) = g(1)$.

DEFINITION 6.1 (Cluster growth process with age). Given an initial distribution π_0 satisfying $\int u \, d\pi_0(u) < \infty$ that is age-critical or age-subcritical, denote by (a_0, C_0) the joint realization of the age of the root and the total number of vertices of T^{π_0}. Let $v_k(0) := \mathbb{P}(|T^{\pi_0}| = k)$ and denote by $(\underline{v}(t))$ the corresponding solution of the critical forest fire equations (cf. Proposition 2.16). We will use this solution to construct (in Section 6.2) the *cluster growth process with age* $(a_t, C_t)_{t \ge 0}$, which is an inhomogeneous $([0, \infty) \times E)$-valued continuous-time càdlàg Markov process such that:

- a_t increases at rate 1,
- C_t jumps from state i with rate i, and conditional on jumping from i at time s, it jumps to $i + k$ with probability $v_k(s)$,
- whenever C_t explodes, C_t returns immediately to state 1 and a_t jumps immediately to 0.

For any $t \in [0, \infty)$, we denote by π_t the distribution of a_t.

The marginal process $(C_t : t \ge 0)$ is a Markov process on its own. It is the cluster growth process studied in [11], extended to allow a critical initial cluster size distribution $v_k(0) = \mathbb{P}(|T^{\pi_0}| = k)$ by Propositions 2.16 and 2.18.

Let us stress that the probability distribution π_t from Definition 6.1 will play the role of the π_t in our main results (i.e., Theorems 2.19, 2.20, 2.21).

Note that Definition 6.1 implies

$$\text{(86)} \qquad \int u \, d\pi_t(u) = \mathbb{E}(a_t) \le \mathbb{E}(a_0 + t) = \int u \, d\pi_0(u) + t < \infty.$$

Note that almost surely, $(C_t : t \ge 0)$ has infinitely many explosion times, but the set of explosion times has no accumulation points (cf. [11], Lemma 3.16). In [11], Section 3.4, it is proved that for any $t, h \in \mathbb{R}_+$ we have

$$\text{(87)} \qquad \int_t^{t+h} \varphi(s) \, ds = \mathbb{E}\big(\#\{s \in [t, t+h] : C \text{ explodes at time } s\}\big),$$

where $\varphi(s)$ is defined in Proposition 2.16 for $s \ge t_{\text{gel}}$ and $\varphi(s) = 0$ if $s < t_{\text{gel}}$.

Let us recall [11], Proposition 1.9, about the marginal distributions of C. Again the proof of this extends to include our critical initial condition once we have Propositions 2.16 and 2.18.

PROPOSITION 6.2 (The law of C_t is $\underline{v}(t)$). *For all $t \geq 0$ and all $k \in \mathbb{N}$, the cluster growth process C satisfies*

(88) $$\mathbb{P}[C_t = k] = v_k(t),$$

where $(\underline{v}(t))$ denotes the solution of the critical forest fire equations (cf. Proposition 2.16) with initial condition $\underline{v}(0)$.

The main result of [11] is stated in Theorem 6.4 below. Loosely speaking, Theorem 6.4 states that for large n, the time evolution of the cluster size of a uniformly chosen vertex ρ^n in the self-organized critical MFFF looks like the cluster growth process C. In particular, burning times of ρ^n correspond to explosion times of C. For further intuitive discussion of the cluster growth process C and how it arises from the MFFF, see [11], Section 1.2.

DEFINITION 6.3 (Components and cardinalities at time t). Given a MFFF(n, λ) process (see Definition 2.1) and a vertex $i \in [n]$, let $\mathcal{C}_t^n(i)$ denote the connected component of vertex i in the graph at time t; moreover, let $C_t^n(i)$ denote the number of vertices in $\mathcal{C}_t^n(i)$, that is, $C_t^n(i) = |\mathcal{C}_t^n(i)|$.

The next theorem is [11], Theorem 1.7, extended to include the case where the limiting initial condition $(v_k(0))$ is the law of $|T^{\pi_0}|$ for an age-critical π_0 with finite mean. The proof in [11] extends to this case once we have Propositions 2.16 and 2.18.

THEOREM 6.4 (Coupling of $(C_t^n(\rho^n))$ and (C_t)). *Let $(\mathcal{G}^n)$ be a sequence of MFFF processes satisfying the conditions of Proposition 2.18, and let ρ^n be a uniformly chosen vertex from $[n]$. There exists a coupling of $(C_t^n(\rho^n))$ and (C_t) such that for any $t_{\max} > 0$,*

$$\sup_{t \in [0, t_{\max}]} d_E(C_t^n(\rho^n), C_t) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty.$$

We will show a convergence result analogous to Theorem 6.4 when we study the clusters of *two* uniformly chosen vertices in a sequence of MFFF processes, and show that they are asymptotically independent.

THEOREM 6.5 (Coupling of $(C_t^n(\rho_i^n))_{i=1}^2$ and $(C_t^{(i)})_{i=1}^2$). *Let $(\mathcal{G}^n)$ be a sequence of MFFF processes as in Theorem 6.4 and let ρ_1^n, ρ_2^n be independent uniform choices of vertices from $[n]$. For each n, there exists a coupling of the process $(C_t^n(\rho_1^n), C_t^n(\rho_2^n); t \in [0, t_{\max}])$ to the process $(C_t^{(1)}, C_t^{(2)}; t \in [0, t_{\max}])$ consisting of two independent copies of the limit cluster growth process (C_t), such that*

(89) $$\sup_{t \in [0, t_{\max}]} d_E(C_t^n(\rho_i^n), C_t^{(i)}) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty, \text{ for } i = 1, 2.$$

We will prove Theorem 6.5 in Section 6.3.

REMARK 6.6. Let us fix $k \in \mathbb{N}$. The proof of Theorem 6.5 can be easily adjusted to the case of $(C_t^n(\rho_i^n))_{i=1}^k$, where $\rho_1^n, \dots, \rho_k^n$ are i.i.d. with uniform distribution on $[n]$, and in this case the joint distributional limit $(C_t^{(i)})_{i=1}^k$ consists of k i.i.d. copies of (C_t). Thus, in the terminology of statistical physics, the MFFF model has the *propagation of chaos* property.

Note that (C_t) fits into the framework of McKean–Vlasov equations with jumps. McKean–Vlasov equations are stochastic differential equations whose coefficients at each time depend

on the law of their own solution at that time (similarly to our Proposition 6.2). McKean–Vlasov equations arise as weak limits of the motion of a tagged particle in an exchangeable interacting particle system which, similar to our Theorem 6.5, exhibit propagation of chaos (cf. [30]).

We now lift Theorem 6.5 to the setting with ages. Let us stress that we start the MFFFA from an age-driven IRG with possibly random initial ages (cf. Assumption 2.5).

PROPOSITION 6.7 (Coupling of cluster size and age processes). *Let $(\mathcal{G}^n)$ be a sequence of MFFFAs satisfying the conditions of Theorem* 2.19, *and let ρ_1^n, ρ_2^n be i.i.d. uniformly chosen vertices in $[n]$. There is a coupling of $(\mathcal{G}^n, \rho_1^n, \rho_2^n)$ with $((a_t^{(i)}, C_t^{(i)}), t \in [0, t_{\max}])_{i \in \{1,2\}}$, which are two i.i.d. copies of the process defined in Definition* 6.1, *such that we have both* (89) *and*

$$\sup_{t \in [0, t_{\max}]} \left| \frac{a_t^n(\rho_i^n)}{C_t^n(\rho_i^n)} - \frac{a_t^{(i)}}{C_t^{(i)}} \right| \overset{\mathbb{P}}{\to} 0 \quad \text{as } n \to \infty, \text{ for } i = 1, 2.$$

In the coupling the age of the tagged vertex ρ_i^n does not jump to 0 at exactly the same times as the age $a_t^{(i)}$ jumps to zero, so there are short intervals during which one age is close to zero and the other is not. The odd-looking division of the age by the cluster size takes care of this, since during this interval where the ages disagree substantially, one cluster size is very large and the other age is very small.

We deduce Proposition 6.7 from Theorem 6.5 in Section 6.4.

Proposition 6.7 implies the following statement, which is in fact all we will need for the proof of Theorems 2.19, 2.20 and 2.21.

COROLLARY 6.8 (Convergence of ages and cluster sizes at a fixed time). *In the situation of Proposition* 6.7, *for each fixed $t \geq 0$ we have*

$$(a_t^n(\rho_i^n), C_t^n(\rho_i^n))_{i \in \{1,2\}} \Rightarrow (a_t^{(i)}, C_t^{(i)})_{i \in \{1,2\}} \quad \text{as } n \to \infty,$$

where the convergence in distribution is with respect to $([0, \infty) \times E)^2$.

We will deduce Corollary 6.8 from Proposition 6.7 in Section 6.4.

Recall the notion of the empirical age measure π_t^n from (6). In Section 6.5, we will use Corollary 6.8 and a second moment argument to prove the convergence of $\pi_t^n \overset{\mathbb{P}}{\Rightarrow} \pi_t$ stated in Theorem 2.19.

In Section 6.6, we will use Theorem 2.19 to deduce Theorem 2.20 about the weak local convergence of the MFFF graph $\mathcal{G}_t^n$ to the MBP tree T^{π_t}.

Recall from Definition 2.8 the notion of the age-driven MBP tree T_s^π where the age of the root is specified to be s.

The following statement bears some resemblance to Theorem 2.6.

LEMMA 6.9 ((C_t) and (a_t) are intertwined). *Conditional on $(a_s)_{0 \leq s \leq t}$, $C_t \overset{d}{=} |T_{a_t}^{\pi_t}|$.*

We will prove Lemma 6.9 in Section 6.7.

REMARK 6.10. Similar to Remark 2.7, in the terminology of of [29], Section 3.2, the Markov process (C_t) is *intertwined* on top of (a_t) (cf. [29], Proposition 3.4).

Recall the notion of the control function $\varphi(t)$ associated to the solution of the critical forest fire equations from Proposition 2.16. For $t < t_{\text{gel}}$ we may assume that $\varphi(t) = 0$. Note that $\int s \, d\pi_t(s) < \infty$ for all $t \geq t_{\text{gel}}$ by (86). For $t \geq t_{\text{gel}}$, the probability distribution π_t is age-critical by Theorem 2.20(iii). Moreover, the eigenfunction $\theta_t(\cdot)$ of $\mathcal{L}_{\pi_t}$ corresponding to the eigenvalue $\lambda = 1$ is uniquely defined for all $t \geq 0$ by Lemma 4.3(iv).

Let us define an auxiliary $[0, \infty)$-valued continuous-time inhomogeneous Markov process that we will use to give an autonomous description of the time evolution of the ages in $(a_t, C_t)_{t \geq 0}$.

DEFINITION 6.11 (An auxiliary Markov process). Let us define a Markov process $(A_t)_{t \geq 0}$, with initial distribution $A_0 \sim \pi_0$ and infinitesimal generator

$$(90) \qquad \lim_{h \to 0_+} \frac{1}{h} \mathbb{E}(f(A_{t+h}) - f(s) | A_t = s) = f'(s) + \varphi(t)\theta_t(s)(f(0) - f(s)),$$

for every $f : [0, \infty) \to \mathbb{R}$, a compactly supported smooth test function.

So A increases deterministically at speed 1 and if $A_t = s$ then A jumps to zero at time t at rate $\varphi(t)\theta_t(s)$.

LEMMA 6.12 (The age process is Markov). *The process* $(a_t)_{t \geq 0}$ (*defined in Definition* 6.1) *has the same law as the process* $(A_t)_{t \geq 0}$ (*defined in Definition* 6.11).

We will prove Lemma 6.12 in Section 6.8. As an ingredient of that proof, we will also prove $\varphi(t) = (\int \theta_t(s)^3 \, d\pi_t(s))^{-1}$ when $t \geq t_{\text{gel}}$ (i.e., (21)).

REMARK 6.13. Similar to Remark 6.6, the McKean–Vlasov analogy applies to the age process (a_t), which arises as the weak limit of the motion of a tagged particle in an exchangeable interacting particle system which exhibits propagation of chaos (cf. Proposition 6.7). In addition, (a_t) is a Markov process governed by its own distribution: on the one hand the coefficient $\varphi(t)\theta_t(s)$ in the infinitesimal generator (90) (which governs (a_t) by Lemma 6.12) is determined by π_t (cf. (21) and Lemma 4.3(iv)), on the other hand π_t is the distribution of a_t (cf. Definition 6.1).

Finally, in Section 6.9 we deduce Theorem 2.21, that is, that (π_t) satisfies the age differential equations, which turn out to be closely related to the Kolmogorov forward equations of the Markov process (a_t).

6.2. *Construction of the cluster growth process with age.* In our construction of (a_t, C_t) satisfying Definition 6.1, we rely on [11], Lemma 3.16, about the almost sure explosion in finite time of C_t, where the stated hypotheses require the initial condition to be subcritical. However, the proof of [11], Lemma 3.16, extends to the case of a critical initial cluster size distribution $v_k(0) = \mathbb{P}(|T^{\pi_0}| = k)$ once we have Propositions 2.16 and 2.18.

We give a doubly-inductive construction, using an array $(\epsilon_{\ell,m})_{\ell \in \mathbb{N}_0, m \in \mathbb{N}}$ of independent Exp(1) r.v.s and an independent array of $(U_{\ell,m})_{\ell \in \mathbb{N}_0, m \in \mathbb{N}}$ of independent $U[0, 1]$ r.v.s. The pair (a_0, C_0) is random and independent of these arrays, jointly distributed as the root age and the number of vertices of the age-driven MBP tree T^{π_0}, defined in Definition 2.8.

We construct $(C_t)_{t>0}$ given C_0, by concatenating a countably infinite sequence of time-inhomogenous continuous-time branching processes, each of which almost surely explodes in finite time. The outer induction is indexed by $\ell \geq 0$ and for $\ell \geq 1$ the ℓth explosion time will be denoted τ_ℓ. Given some $\ell \geq 0$, the inner induction constructs the ℓth branching process which jumps at times $(J_{\ell,m})_{m \in \mathbb{N}}$.

We define $J_{0,0} = 0$. If $\ell \geq 1$ then we suppose that τ_ℓ has already been defined and C_t has already been constructed up to time τ_ℓ, and define $J_{\ell,0} := \tau_\ell$ if $\ell \geq 1$. If $\ell = 0$, then define $C_{J_{\ell,0}} := C_0$, and if $\ell \geq 1$ then we define $C_{J_{\ell,0}} = C_{\tau_\ell} := 1$. Given some $\ell \geq 0$, we recursively define for each $m \geq 1$,

$$(91) \qquad J_{\ell,m} := J_{\ell,m-1} + \frac{\epsilon_{\ell,m}}{C_{J_{\ell,m-1}}},$$

$$(92) \qquad C_t := C_{J_{\ell,m-1}}, \quad \text{for } t \in [J_{\ell,m-1}, J_{\ell,m}),$$

$$(93) \qquad C_{J_{\ell,m}} := C_{J_{\ell,m-1}} + \chi(J_{\ell,m}, U_{\ell,m}), \quad \text{where}$$

$$(94) \qquad \chi : [0, \infty) \times [0, 1) \to \mathbb{N}_+, \qquad \chi(t, x) = k, \quad \forall x \in \left[\sum_{\ell=1}^{k-1} v_\ell(t), \sum_{\ell=1}^{k} v_\ell(t) \right).$$

So $\chi(t, \cdot)$ is the inverse of the cumulative distribution function which corresponds to the distribution $\underline{v}(t)$. We define $\tau_{\ell+1} := \lim_{m \to \infty} J_{\ell,m}$. Lemma 3.16 in [11] shows that the properties of $(v_k(s))_{k=1}^\infty$ (see in particular (17)) imply that τ_ℓ is almost surely finite. Since $\tau_{\ell+1} - \tau_\ell \geq J_{\ell,1} - \tau_\ell = \epsilon_{\ell,1}$, we have $\tau_\ell \to \infty$ as $\ell \to \infty$. The inductive construction therefore constructs C_t for all $t \geq 0$ almost surely. We have $\lim_{h \to 0_+} C_{\tau_\ell - h} = \infty$ w.r.t. the usual topology of $\mathbb{N}$. This means that $\lim_{h \to 0_+} C_{\tau_\ell - h} = 1$ w.r.t. the topology of E, so (C_t) is continuous from the left at time τ_ℓ, since $C_{\tau_\ell} = 1$.

We now construct a_t for all $t > 0$. Denote $\tau_0 := -a_0$ and define the age a_t at time t by

$$(95) \qquad a_t := t - \tau_{L_t}, \quad \text{where } L_t := \max\{\ell \geq 0 : \tau_\ell \leq t\}.$$

This completes the construction of (a_t, C_t). The marginal (C_t) is precisely the cluster growth process of [11], therefore the properties of (C_t) discussed in Section 6.1 apply to it.

6.3. *Asymptotic independence of cluster processes.* Section 6.3 is devoted to the proof of Theorem 6.5.

Recall from Definition 6.3 the notion of the connected cluster $\mathcal{C}_t^n(i)$ of vertex i at time t as well as the size $C_t^n(i)$ of this cluster.

Our proof of Theorem 6.5 proceeds by showing that the evolutions of $C_t^n(\rho_1^n)$ and $C_t^n(\rho_2^n)$ are well approximated by the limiting cluster process whenever they are not too large, and are approximately independent except when $\mathcal{C}_t^n(\rho_1^n) = \mathcal{C}_t^n(\rho_2^n)$. The following lemma controls the probability of this pathological event.

LEMMA 6.14 (The clusters of ρ_1^n and ρ_2^n do not merge). *Let*

$$(96) \qquad \tau^n := \inf\{t \geq 0 : \mathcal{C}_t^n(\rho_1^n) = \mathcal{C}_t^n(\rho_2^n)\} = \inf\{t \geq 0 : \rho_2^n \in \mathcal{C}_t^n(\rho_1^n)\}.$$

Then for any $t_{\max} > 0$, we have

$$(97) \qquad \lim_{n \to \infty} \mathbb{P}(\tau^n \leq t_{\max}) = 0.$$

PROOF. We begin by observing that the model asymptotically almost surely never includes a giant component, that is, for any $\epsilon > 0$,

$$(98) \qquad \lim_{n \to \infty} \mathbb{P}(\exists t \in [0, t_{\max}], \exists v \in [n] \text{ s.t. } C_t^n(v) > n\epsilon) = 0.$$

The proof of (98) follows from Propositions 2.16 and 2.18, as we now explain. Recall the notion of the solution $(v_k(t))$ of the critical forest fire equations ((12) + (13)). Recall the notion $v_k^n(t)$ of the fraction of vertices contained in components of size k at time t from (11).

It follows from (13) and Dini's theorem that there exists $K \in \mathbb{N}$ such that $\sum_{k=1}^{K} v_k(t) \geq 1 - \frac{\epsilon}{2}$ for all $t \leq t_{\max}$. Then we can use (18) to derive

$$(99) \qquad \lim_{n \to \infty} \mathbb{P}\left(\inf_{t \leq t_{\max}} \sum_{k=1}^{K} v_k^n(t) \geq 1 - \epsilon \right) = 1.$$

Now (98) follows from (99) as soon as we observe that if $n\epsilon > K$ and if $C_t^n(v) > n\epsilon$ for some $t \leq t_{\max}$ and $v \in [n]$ then $\sum_{k=1}^{K} v_k^n(t) < 1 - \epsilon$.

Next, consider the number $\mathcal{N}$ of fires involving ρ_1^n up to time t. Since $C^n(\rho_1^n)$ spends an $\mathrm{Exp}(1 + \lambda(n))$ holding time at size 1 after each fire, $\mathcal{N}$ is stochastically bounded by $1 + \mathrm{Poisson}(t_{\max}(1 + \lambda(n)))$. So we can bound in expectation the total number $\mathcal{M}_{t_{\max}}^n$ of vertices which are ever in the same component as ρ_1^n on time interval $[0, t_{\max}]$ as

$$\mathbb{E}(\mathcal{M}_{t_{\max}}^n) = \mathbb{E}\left[\left| \bigcup_{t \in [0, t_{\max}]} C_t^n(\rho_1^n) \right| \right]$$

$$\leq n\mathbb{P}(\exists t \in [0, t_{\max}], C_t^n(\rho_1^n) \geq n\epsilon) + (1 + t_{\max}(1 + \lambda(n)))n\epsilon.$$

Since this holds for all $\epsilon > 0$, using (98) we obtain $\lim_{n \to \infty} \frac{1}{n}\mathbb{E}(\mathcal{M}_{t_{\max}}^n) = 0$, from which (97) follows as soon as we note that ρ_1^n and ρ_2^n are i.i.d. with uniform distribution on $[n]$. $\square$

PROOF OF THEOREM 6.5. The proof of Theorem 6.4 (cf. [11], Section 4) involves a coupling of the pair $(C^n(\rho_1^n), C)$ such that these processes are close with respect to d_E throughout $[0, t_{\max}]$ with high probability. This means they are equal at all times when neither is very large, and when either is very large then the other is either very large or equal to 1. We will restate this coupling, and explain how it may be extended to a coupling of $(C^n(\rho_1^n), C^n(\rho_2^n), C^{(1)}, C^{(2)})$, where the final two processes are independent copies of the limit cluster growth process C. We will follow the style of [11] and state the coupling in explicit detail, from which it will be easy to derive the rigorous results required.

As in [11], Section 4, we work with a modified version of the MFFF process, which alters the rates of clocks on edges within clusters, and introduces clocks on loop edges (i.e., edges from a vertex to itself). Precisely, we demand:

$$(100) \qquad \begin{array}{l} \text{edge clocks \textit{within a cluster} ring at rate } 2/n \text{ rather than } 1/n \\[4pt] \text{and each loop edge has an independent clock with rate } 1/n. \end{array}$$

These adjustments have no effect on the evolution of the cluster sizes, but simplify the statements of some Poisson process calculations to follow.

For each n, we define the (random) function $\chi^n : [0, \infty) \times [0, 1) \to \mathbb{N}$ similar to (94):

$$(101) \qquad \chi^n(t, x) = k, \quad \forall x \in \left[\sum_{\ell=1}^{k-1} v_\ell^n(t), \sum_{\ell=1}^{k} v_\ell^n(t) \right).$$

We also introduce a truncation parameter $K \in \mathbb{N}$, which will be taken large enough at the end.

We now describe the coupling of $(C^n(\rho_1^n), \tilde{C}, S)$, where $\tilde{C}$ will be a copy of the cluster growth process C; moreover, we augment the coupling with a *failure parameter* $S \in \{0, 1\}$, which, informally, will switch from 0 to 1 at a stopping time when the coupling of $C^n(\rho_1^n)$ and $\tilde{C}$ ceases to be effective.

(i) Sample $U_0 \sim U[0, 1]$, and set $\tilde{C}_0 = \chi(0, U_0)$, and select ρ_1^n uniformly at random from the vertices $v \in [n]$ for which $|\mathcal{C}_0^n(v)| = \chi^n(0, U_0)$, so $C_0^n(\rho_1^n) = \chi^n(0, U_0)$. From now on, we write C_t^n for $C_t^n(\rho_1^n)$. If $\tilde{C}_0 = C_0^n$, set $S = 0$, otherwise, set $S = 1$.

(ii) Enumerate the burning times and growth times of $C_t^n(\rho_1^n)$ as follows. Set $I_{0,0} = 0$ and inductively

(102) $$I_{\ell,0} := \inf\{s > I_{\ell-1,0} : C^n(\rho_1^n) \text{ is burned at time } s\}, \quad \ell \geq 1.$$

So $I_{\ell,0}$ is the time of the ℓth fire involving ρ_1^n. Now, define $I_{\ell,1} \leq I_{\ell,2} \leq \cdots \leq I_{\ell,g_\ell}$ to be the sequence of times between $I_{\ell,0}$ and $I_{\ell+1,0}$ at which an edge clock rings, for which at least one incident vertex is in $C^n(\rho_1^n)$ at that time. This sequence includes times when a clock corresponding to an edge *within* $C_t^n(\rho_1^n)$ rings, which does not lead to a change in C_t^n.

(iii) For every $\ell \geq 0$, $m \in [1, g_\ell]$, as a result of the assumption (100), the random variable $I_{\ell,m} - I_{\ell,m-1}$ has exponential distribution with rate $C_{I_{\ell,m-1}}^n(\rho_1^n)$. Moreover, we may assume that the edge whose clock rings at time $I_{\ell,m}$ is $(i_{\ell,m}, j_{\ell,m})$, where $i_{\ell,m}$ is sampled uniformly from $C_{I_{\ell,m-1}}^n(\rho_1^n)$ and, independently, $j_{\ell,m}$ is sampled uniformly from $[n]$. So if we define $L_{\ell,m}^n := \chi^n(I_{\ell,m}-, U_{\ell,m})$, where $(U_{\ell,m}, \ell \geq 0, m \geq 1)$ are i.i.d. $U[0,1]$ variables, we may assume that $j_{\ell,m}$ is chosen uniformly from those vertices $v \in [n]$ for which $C_{I_{\ell,m}-}^n(v) = L_{\ell,m}^n$.

(iv) For each growth time $I_{\ell,m}$, if $\max(C_{I_{\ell,m}-}^n, \tilde{C}_{I_{\ell,m}-}) \leq K$ and $S = 0$, we do the following. Set $\tilde{L}_{\ell,m} := \chi(I_{\ell,m}-, U_{\ell,m})$, and $\tilde{C}_{I_{\ell,m}} = \tilde{C}_{I_{\ell,m}-} + \tilde{L}_{\ell,m}$. Then:

- if $\tilde{L}_{\ell,m} \neq L_{\ell,m}^n$, set $S = 1$;
- if $j_{\ell,m} \in C_{I_{\ell,m}-}^n(\rho_1^n)$, then $C_{I_{\ell,m}}^n = C_{I_{\ell,m}-}^n$, and set $S = 1$;
- otherwise, we have $C_{I_{\ell,m}}^n = C_{I_{\ell,m}-}^n + L_{\ell,m}^n$ and $\tilde{C}_{I_{\ell,m}} = \tilde{C}_{I_{\ell,m}-} + \tilde{L}_{\ell,m}$ (in this case we have $\tilde{L}_{\ell,m} = L_{\ell,m}^n$) and we retain $S = 0$.

(v) Whenever $\max(C_t^n, \tilde{C}_t) > K$, or when $S = 1$, then let the two processes evolve independently. During this, preserve the value of S, *unless* one process jumps away from 1 while the other process is greater than K. If this happens while $S = 0$, move to the failure state $S = 1$.

(vi) For each burning time $I_{\ell,0}$, if $C_{I_{\ell,0}-}^n \leq K$, set $S = 1$ thereafter.

The effect of these dynamics is that while $S = 0$, we either have $C_t^n = \tilde{C}_t \leq K$, or

$$(C_t^n, \tilde{C}_t) \in \{K+1, \ldots\}^2 \cup (\{1\} \times \{K+1, \ldots\}) \cup (\{K+1, \ldots\} \times \{1\}).$$

Crucially, at all times the evolution of $\tilde{C}$ matches with the construction of the cluster growth process given in Section 6.2, and so we have

(103) $$\tilde{C} \overset{d}{=} C.$$

For the purposes of this article, we require a similar coupling $(C^n(\rho_1^n), C^n(\rho_2^n), \tilde{C}^{(1)}, \tilde{C}^{(2)})$ for two independently chosen watched vertices $\rho_1, \rho_2 \in [n]$, which we now describe. The use of the "phantom clocks," though notationally heavy, ensures that independence of $\tilde{C}^{(1)}, \tilde{C}^{(2)}$ is immediate.

- The evolution of $(C^n(\rho_1^n), \tilde{C}^{(1)})$ is exactly as in the original coupling, with failure parameter $S^{(1)}$, and the additional stipulation that whenever $\tilde{C}^{(1)}$'s evolution was specified to be independent of $C^n(\rho_1^n)$, it is now also independent of $C^n(\rho_2^n)$ and $\tilde{C}^{(2)}$.
- The construction of $\tilde{C}^{(2)}$ will mostly be the same, but we require extra machinery to ensure that it stays independent of $\tilde{C}^{(1)}$. The addition required is a family of independent *phantom clocks*, all with rate $1/n$, attached to edges between $C_t^n(\rho_1^n)$ and $C_t^n(\rho_2^n)$ for $t \leq \tau^n$ (see (96)). We define the rate of phantom clocks to be zero for $t > \tau^n$.
- Now, to determine the evolution of $\tilde{C}^{(2)}$ before τ^n, we use the regular edge clocks from the forest-fire model associated to the edges between $C_t^n(\rho_2^n)$ and $[n] \backslash C_t^n(\rho_1^n)$, but use the phantom clocks between $C_t^n(\rho_1^n)$ and $C_t^n(\rho_2^n)$.

- We define $\bar\tau^n$ to be the first time that a phantom clock rings, which we think of as a *phantom growth time* for $\mathcal{C}^n(\rho_2^n)$. Note that $\bar\tau^n \geq \tau^n$ implies $\bar\tau^n = \infty$. We also define $\bar{I}_{0,0} = 0$ and the burning times $(\bar{I}_{\ell,0}, \ell \geq 1)$ for $\mathcal{C}_t^n(\rho_2^n)$ exactly as in (102), and the growth times $(\bar{I}_{\ell,m}, \ell \geq 0, m = 1, \ldots, \bar{g}_\ell)$ similarly, *up to and including* the phantom growth time $\bar\tau^n$. It remains the case, as in (iii), that whenever a regular (nonphantom) clock attached to the component of ρ_2^n rings, the incident vertices may be given by a uniform choice from $\mathcal{C}_{\bar{I}_{\ell,m}-}^n(\rho_2^n)$, and a uniform choice from $[n]$.

- We define the evolution of $\tilde{C}^{(2)}$, and its failure parameter $S^{(2)}$ using the burning times and growth times $(\bar{I}_{\ell,m})$ exactly as in (iv), (v), (vi) above; moreover, we add the stipulation that $S^{(2)}$ moves to the failure state $S^{(2)} = 1$ at $\tau^n \wedge \bar\tau^n$. So $\tilde{C}^{(2)}$ evolves independently of the forest-fire model and of $\tilde{C}^{(1)}$ whenever $\max(C_t^n(\rho_2^n), \tilde{C}^{(2)}) > K$, or when $S^{(2)} = 1$. It follows, analogously to (103), that $\tilde{C}^{(2)}$ is a copy of C.

- The jumps of $\tilde{C}^{(1)}$ and $\tilde{C}^{(2)}$ are independent, as we now explain. In the case where all $C_t^n(\rho_1^n), C_t^n(\rho_2^n), \tilde{C}^{(1)}, \tilde{C}^{(2)} \leq K$ and $S^{(2)} = 0$, the jumps of $\tilde{C}^{(1)}$ and $\tilde{C}^{(2)}$ are driven by disjoint sets of edge clocks in the underlying forest fire model. This is because $\mathcal{C}_t^n(\rho_1^n)$ and $\mathcal{C}_t^n(\rho_2^n)$ are disjoint if $S^{(2)} = 0$, and the true edges between them drive $\tilde{C}^{(1)}$, while the phantom edges drive $\tilde{C}^{(2)}$. The other cases are independent by definition, since at least one of $\tilde{C}^{(1)}, \tilde{C}^{(2)}$ is evolving independently of everything else. In other words,

$$(104) \qquad \tilde{C}^{(1)} \text{ and } \tilde{C}^{(2)} \text{ are i.i.d. copies of } C.$$

Theorem 1.7 of [11] shows that for any $\epsilon > 0$, for large enough K we have

$$(105) \qquad \lim_{n\to\infty} \mathbb{P}\Big(\sup_{t\in[0,t_{\max}]} d_E\big(C_t^n(\rho_1^n), \tilde{C}_t^{(1)}\big) > \epsilon \Big) = 0.$$

Now note that while $S^{(2)} = 0$, the times $\bar\tau^n$ and τ^n are reached at the same (random, time-dependent) rates, being given by the number of real and phantom edges, respectively, between the two watched clusters. Since the rate of the phantom clocks is set to zero after τ^n, if $\bar\tau^n \geq \tau^n$ then $\bar\tau^n = \infty$. Hence $\bar\tau^n$ stochastically dominates τ^n, and so using Lemma 6.14, we have, as $n \to \infty$,

$$\mathbb{P}(\tau^n \wedge \bar\tau^n \leq t_{\max}) = \mathbb{P}(\{\tau^n \leq t_{\max}\} \cup \{\bar\tau^n \leq t_{\max}\}) \leq 2\mathbb{P}(\tau^n \leq t_{\max}) \to 0.$$

Then we may obtain, similar to (105), that

$$(106) \qquad \lim_{n\to\infty} \mathbb{P}\Big(\sup_{t\in[0,t_{\max}]} d_E\big(C_t^n(\rho_2^n), \tilde{C}_t^{(2)}\big) > \epsilon \Big) \leq 0 + \lim_{n\to\infty} \mathbb{P}(\tau^n \wedge \bar\tau^n \leq t_{\max}) = 0$$

and so the proof of Theorem 6.5 follows from (104), (105) and (106). $\quad\square$

6.4. *Coupling of cluster size and age processes.* Section 6.4 is devoted to the proof of Proposition 6.7 and Corollary 6.8.

We want to deduce the result about the coupling of the age processes from the properties of the coupling of the cluster size processes of the watched vertices (i.e., Theorem 6.5). The $t = 0$ case will follow from Proposition 3.5. As for the proof of the coupling of the whole age processes, the main idea is that the last burning time (cf. Definition 2.1) of the vertex ρ_i^n can almost be read off from the evolution of the cluster size $C_t^n(\rho_i^n)$, since it is the same as the elapsed time since the cluster size last dropped to one, except in the unlikely case that lightning strikes the vertex ρ_i^n while it is a singleton.

We will slightly modify the notion of the age of a vertex j at time t in the MFFF (cf. Definition 2.3).

DEFINITION 6.15 (Modified ages).   For $t > 0$, let $\bar{a}_t^n(j)$ denote the minimum of $a_0^n(t) + t$ and the time that has elapsed since the last burning time s of vertex j for which $C_{s-}^n(j) > 1$ (and we let $s := -\infty$ if $[0, t]$ does not contain such a burning time).

In other words, $\bar{a}_t^n(j)$ ignores instances where vertex j is struck by lightning while a singleton.

We will prove Proposition 6.7 by first showing in (107) that with asymptotically high probability the above modification of ages does not actually change the ages during $[0, t_{max}]$. Then we will show that the analogue of Proposition 6.7 holds if we use the modified ages.

PROOF OF PROPOSITION 6.7.   The rate of lightning striking the vertices ρ_1^n and ρ_2^n while they are singletons is $\lambda(n) \ll 1$ (by assumption (5)), thus we have

$$(107) \qquad \mathbb{P}\big(\bar{a}_t^n(\rho_i^n) = a_t^n(\rho_i^n), i \in \{1, 2\}, t \in [0, t_{max}]\big) \geq \exp(-2\lambda(n)t_{max}) = 1 - o_n(1).$$

For $i \in \{1, 2\}$, we have

$$\bar{a}_t^n(\rho_i^n) = t - (-a_0^n(\rho_i^n)) \vee \sup\{u \leq t : C_{u-}^n(\rho_i^n) > 1, C_u^n(\rho_i^n) = 1\},$$

$$a_t^{(i)} = t - (-a_0^{(i)}) \vee \sup\{u \leq t : C_{u-}^{(i)} > 1, C_u^{(i)} = 1\}.$$

The key point here is that in both cases the modified age process is a function of the initial age and the cluster size process.

We will improve the coupling that we gave in the proof of Theorem 6.5, by augmenting it with initial ages.

We will show that for any $\epsilon, \delta > 0$ there exists n_0 such that for any $n \geq n_0$ the event that $\sup_{t \in [0, t_{max}]} |\bar{a}_t^n(\rho_i^n)/C_t^n(\rho_i^n) - a_t^{(i)}/C_t^{(i)}| \leq \epsilon$ occurs for both $i = 1, 2$ is greater than or equal to $1 - 8\delta$.

We use Proposition 3.5 to control the initial coupling at time 0. We couple the pairs $(a_0^n(\rho_i^n), C_0^n(\rho_i^n))_{i \in \{1,2\}}$ to $(a_0^{(i)}, C_0^{(i)})_{i \in \{1,2\}}$ so that for n large enough we have $|a_0^n(\rho_i^n) - a_0^{(i)}| < \epsilon$ and $C_0^n(\rho_i^n) = C_0^{(i)}$ for both $i = 1, 2$ with probability at least $1 - \delta$.

Choose a_{max} sufficiently large that $\pi_0([a_{max}, \infty)) < \delta$, so with probability at least $1 - 2\delta$ we have $a_0^{(i)} < a_{max}$ for $i = 1, 2$, and for sufficiently large n, with probability at least $1 - 3\delta$ we have $a_0^n(\rho_i^n) < a_{max}$ for $i = 1, 2$. Suppose that these events occur. Then all four ages cannot exceed $a_{max} + t_{max}$ during the time interval $[0, t_{max}]$. Recall the threshold parameter K in the coupling of Theorem 6.5, and set $K > (a_{max} + t_{max})/\epsilon$. Then whenever $C_t^{(i)} \geq K$, we have $a_t^{(i)}/C_t^{(i)} < \epsilon$, and similarly whenever $C_t^n(\rho_i^n) \geq K$, we have $\bar{a}_t^n(\rho_i^n)/C_t^n(\rho_i^n) < \epsilon$. Take n large enough that with this choice of K the probability of coupling failure occurring at a strictly positive time $t \leq t_{max}$ is at most δ.

Finally, we will bound the probability that some pair of corresponding explosion and burning times in the coupling differ by at least ϵ. Because the interexplosion times of $C_t^{(i)}$ stochastically dominate a sequence of independent $\text{Exp}(1)$ random variables, there exists N depending on t_{max} and δ such that with probability at least $1 - \delta$, $C_t^{(i)}$ explodes at most N times during $[0, t_{max}]$. If we take K sufficiently large, then (conditional on coupling success) the probability that the ℓth burning time of ρ_i^n is at least ϵ away from the ℓth explosion time of $C^{(i)}$ is at most $\delta/(2N)$ for n sufficiently large. This follows from [11], Lemmas 4.1 and 4.2, which say that the watched cluster burns quickly with high probability once it is sufficiently large, and the corresponding statement for the cluster growth process. A union bound now shows that with probability at least $1 - \delta$ all of the corresponding explosion and burning times in both $C^{(1)}$ and $C^{(2)}$ differ by less than ϵ.

Now the modified age processes can be read off as above from the initial ages and the cluster size processes. We find that with probability at least $1 - 8\delta$ we have for both $i = 1$

and $i = 2$ that $|a_t^{(i)} - \bar{a}_t^n(\rho_i^n)| < \epsilon$ throughout $[0, t_{\max}]$ except at times when exactly one of $C_t^{(i)}$ and $C_t^n(\rho_i^n)$ is at least K and the other is equal to 1. During those times, we have $0 \le a_t^{(i)}/C_t^{(i)} < \epsilon$ and $0 \le \bar{a}_t^n(\rho_i^n)/C_t^n(\rho_i^n) < \epsilon$.

The proof of Proposition 6.7 follows using (107) to replace $\bar{a}^n$ by a^n, with an asymptotically negligible probability of failure. $\square$

PROOF OF COROLLARY 6.8. Let us fix $0 < \epsilon < t_{\max}^{-2}$. Suppose that for $i = 1, 2$ we have

$$a_0^n(\rho_i^n) < \epsilon^{-1/2}, \qquad a_0^{(i)} < \epsilon^{-1/2},$$

$$d_E(c_t^n(\rho_i^n), C_t^{(i)}) \le \epsilon, \quad \text{and} \quad \left| \frac{a_t^n(\rho_i^n)}{C_t^n(\rho_i^n)} - \frac{a_t^{(i)}}{C_t^{(i)}} \right| \le \epsilon.$$

(This happens with high probability if ϵ is small and then $n = n(\epsilon)$ is sufficiently large.) A calculation using the definition of the metric d_E shows that one of the following three options must hold:

1. $C_t^n(\rho_i^n) = C_t^{(i)} \le 1/\sqrt{\epsilon}$, in which case $|a_t^n(\rho_i^n) - a_t^{(i)}| \le \sqrt{\epsilon}$, or
2. $C_t^{(i)} \ge 1/\sqrt{\epsilon}$, or
3. $C_t^{(i)} = 1$ and $C_t^n(\rho_i^n) > 1/\epsilon$, in which case

$$\frac{a_t^n(\rho_i^n)}{C_t^n(\rho_i^n)} < \epsilon(a_0^n(\rho_i^n) + t) \le 2\epsilon^{1/2},$$

so $a_t^{(i)} \le \epsilon + 2\epsilon^{1/2}$.

For any fixed time $t > 0$, the probability that $C_t^{(i)} > 1/\sqrt{\epsilon}$ tends to 0 as $\epsilon \to 0$ by Propositions 6.2 and 2.16, thus the probability of option 2 above is small.

The probability that $C^{(i)}$ has an explosion in the time interval $[t - \epsilon - 2\epsilon^{1/2}, t]$ also tends to 0 as $\epsilon \to 0$, because the expected number of explosions of C_t in any interval $[a, b]$ is $\int_a^b \varphi(s) \, ds$ (see [11], Section 3.4) and $\varphi(s)$ is bounded on $[0, t_{\max}]$ (see Proposition 2.16), thus the probability of option 3 above is also small.

It follows that option 1 above occurs with asymptotically high probability, completing the proof of Corollary 6.8. $\square$

6.5. *Convergence of empirical age evolution.* The goal of Section 6.5 is to prove Theorem 2.19. We will prove that for any fixed $t \in [0, \infty)$, we have

$$(108) \qquad \qquad \pi_t^n \overset{\mathbb{P}}{\Rightarrow} \pi_t,$$

where π_t is the distribution of a_t (cf. Definition 6.1). By Definition 2.12, we only need to check that for any bounded continuous function $f : [0, \infty) \to \mathbb{R}$ we have

$$(109) \qquad \qquad \lim_{n \to \infty} \mathbb{E}\left(\int f(s) \, d\pi_t^n(s) \right) = \int f(s) \, d\pi_t(s),$$

$$(110) \qquad \qquad \lim_{n \to \infty} \text{Var}\left(\int f(s) \, d\pi_t^n(s) \right) = 0.$$

First, we show (109). Corollary 6.8 gives $a_t^n(\rho_1^n) \Rightarrow \pi_t$. This immediately gives (109), since $\mathbb{E}(\int f(s) \, d\pi_t^n(s)) = \mathbb{E}(f(a_t^n(\rho_1^n)))$.

Corollary 6.8 also implies

$$\lim_{n \to \infty} \mathbb{E}\left[\left(\int f(s) \, d\pi_t^n(s) \right)^2 \right] = \lim_{n \to \infty} \mathbb{E}[f(a_t^n(\rho_1^n)) f(a_t^n(\rho_2^n))]$$

(111)
$$= \mathbb{E}[f(a_t^{(1)}) f(a_t^{(2)})]$$

$$= \left(\int f(s) \, d\pi_t(s) \right)^2.$$

The required variance statement (110) follows from (109) and (111). This completes the proof of (108). The proof of Theorem 2.19 is complete.

6.6. *The local weak limit of MFFFA at time t.* The goal of Section 6.6 is to prove Theorem 2.20.

We begin by proving (i). We have $\pi_t^n \overset{\mathbb{P}}{\Rightarrow} \pi_t$ by Theorem 2.19. Moreover by Theorem 2.6, the MFFFA$(n, \underline{a}_0^n, \lambda)$ graph $\mathcal{G}_t^n$ is an age-inhomogeneous random graph $G^{\text{age}}(n, \underline{a}_t^n)$. We may thus use Proposition 2.13 to conclude that the graph $\mathcal{G}_t^n$ converges in probability in the local weak sense to the age-driven multitype branching process tree T^{π_t} as $n \to \infty$. This proves (i).

Next, we prove (ii). Recall from (11) that $v_k^n(t)$ is the proportion of vertices of $\mathcal{G}_t^n$ in size k components at time t. On the one hand, (i) implies that $v_k^n(t) \overset{\mathbb{P}}{\to} \mathbb{P}(|T^{\pi_t}| = k)$ as $n \to \infty$ for any $k \geq 1$ and $t \geq 0$. On the other hand, Propositions 2.16 and 2.18 imply that $v_k^n(t) \overset{\mathbb{P}}{\to} v_k(t)$ as $n \to \infty$ where $\underline{v}(\cdot)$ is the unique solution to the critical forest fire equations ((12) + (13)) with initial state $\underline{v}(0)$ given by $v_k(0) = \mathbb{P}(|T^{\pi_0}| = k)$. This proves (ii).

It remains to prove (iii). By Proposition 2.16, we have $\sum_{k=1}^{\infty} v_k(t) = 1$,

$$\sum_{k \geq 1} k v_k(t) < \infty, \quad t < t_{\text{gel}}, \quad \text{and} \quad \sum_{k \geq 1} k v_k(t) = \infty, \quad t \geq t_{\text{gel}}.$$

By (ii), we have $\mathbb{P}(|T^{\pi}| < \infty) = \sum_{k=1}^{\infty} v_k(t)$ and $\mathbb{E}(|T^{\pi}|) = \sum_{k=1}^{\infty} k v_k(t)$, thus (iii) follows by Proposition 2.10. The proof of Theorem 2.20 is complete.

6.7. *(C_t) and (a_t) are intertwined.* In Section 6.7, we prove Lemma 6.9.

Theorem 2.19 gives $\pi_t^n \overset{\mathbb{P}}{\Rightarrow} \pi_t$, and so we may apply Proposition 3.5 to $\mathcal{G}_t^n$, to obtain

$$(a_t^n(\rho^n), C_t^n(\rho^n)) \Rightarrow (a(\rho), |T^{\pi_t}|),$$

where ρ^n is uniformly distributed on $[n]$ and $a(\rho)$ is the age of the root in the age-labeled tree T^{π_t}. However, Corollary 6.8 gives

$$(a_t^n(\rho^n), C_t^n(\rho^n)) \Rightarrow (a_t, C_t),$$

and so we know that conditional on a_t, we have $C_t \overset{d}{=} |T_{a_t}^{\pi_t}|$.

It remains to show that conditional on a_t, C_t is independent of $(a_s)_{0 \leq s < t}$.

Recall that $0 < \tau_1 < \tau_2 < \cdots$ are the consecutive explosion times of C and $\tau_0 = -a_0$. Recall from (95) that we denote by $L_t := \max\{\ell \geq 0 : \tau_\ell \leq t\}$ the index of the last explosion time τ_{L_t} before t.

Assume that we are given the value of a_t. If $a_t \geq t$, then we know that C did not explode on $[0, t]$, thus for all $s \in [0, t]$ we have $a_s = a_t - (t - s)$. So in this case, conditional on the value of a_t, the cluster size C_t is independent of $(a_s)_{0 \leq s \leq t}$, since the latter is fully determined by the value of a_t.

Now assume that we are given the value a_t, and we have $a_t \in [0, t)$. This is equivalent to assuming that $\tau_{L_t} = t - a_t$. Under this assumption, one has $a_s = s - \tau_{L_t}$ for any $s \in [\tau_{L_t}, t]$, therefore we only need to show that conditional on τ_{L_t}, C_t is independent of $(a_s)_{0 \leq s < \tau_{L_t}}$.

Note that $(a_s)_{0 \leq s < \tau_{L_t}}$ is determined by a_0 and $(C_s)_{0 \leq s < \tau_{L_t}}$ via (95), so it is enough to show that that conditional on τ_{L_t}, $(C_s)_{\tau_{L_t} \leq s}$ is independent of $(a_0, (C_s)_{0 \leq s < \tau_{L_t}})$. This statement follows from the construction of the cluster growth process with age given in Section 6.2: $(C_s)_{\tau_{L_t} \leq s}$ is constructed via (91)–(93) using the value of $J_{L_t, 0}$ (i.e., τ_{L_t}) and the arrays $(\epsilon_{\ell, m})_{L_t \leq \ell, 1 \leq m}$ and $(U_{\ell, m})_{L_t \leq \ell, 1 \leq m}$, and the joint distribution of these arrays does not depend on $(a_0, (C_s)_{0 \leq s < \tau_{L_t}})$ (in particular, it does not depend on the number L_t of explosions in $[0, t]$). The proof of Lemma 6.9 is complete.

6.8. *The age process is Markov.* The goal of Section 6.8 is to prove Lemma 6.12 and also the formula (21).

The processes (a_t) and (A_t) have the same initial distribution, and as long as $t \in [0, t_{gel})$, both processes simply increase deterministically at speed 1. Indeed, for (A_t), this follows from Definition 6.11 and our assumption that $\varphi(t) = 0$ if $t \in [0, t_{gel})$. On the other hand, the cluster growth process (C_t) does not explode in $[0, t_{gel})$ (see [11], Lemma 3.11), thus by Definition 6.1 we have $a_t = a_0 + t$ for any $t < t_{gel}$.

Thus w.l.o.g. we may assume $t_{gel} = 0$ for the rest of Section 6.8 (cf. Remark 2.23(i)). Under this assumption π_t is age-critical for any $t \geq 0$, by Theorem 2.20(iii). Moreover, the eigenfunction $\theta_t(\cdot)$ of $\mathcal{L}_{\pi_t}$ corresponding to the eigenvalue $\lambda = 1$ is well defined for all $t \geq 0$ by Lemma 4.3(iv).

Our next step in the proof of Lemma 6.12 is to show (21), that $\varphi(t) = (\int \theta_t(s)^3 \, d\pi_t(s))^{-1}$ when $t \geq t_{gel}$.

PROOF OF (21). Equation (48) of Lemma 5.1 implies

$$(112) \qquad \lim_{\varepsilon \to 0_+} \frac{1}{\sqrt{\varepsilon}} \mathbb{E}(1 - (1 - \varepsilon)^{|T^{\pi_t}|}) = \sqrt{2\phi_t},$$

where, following (46), we take $\phi_t = 1 / \int \theta_t(s)^3 \, d\pi_t(s)$. Let us compare this with (17), which is equivalent to

$$(113) \qquad \lim_{\varepsilon \to 0_+} \frac{1}{\sqrt{\varepsilon}} \left(1 - \sum_{k=1}^{\infty} v_k(t)(1 - \varepsilon)^k\right) = \sqrt{2\varphi(t)}$$

by Example (c) of Theorem 4, Chapter XIII.5 of [17].

By Theorem 2.20(ii), we have

$$(114) \qquad \sum_{k=1}^{\infty} v_k(t)(1 - \varepsilon)^k = \mathbb{E}((1 - \varepsilon)^{|T^{\pi_t}|}), \quad \varepsilon \in (0, 1),$$

whence $\phi_t = \varphi(t)$, which proves (21). $\square$

PROOF OF LEMMA 6.12. We now assume $t_{gel} = 0$. The initial distribution of a_0 agrees with the initial distribution of A_0. Both a and A increase deterministically at speed 1 and occasionally jump back to zero. Recalling the definition of L_t from (95), it follows that $\tau_{L_t + 1}$ is the time of the first explosion after t. We only need to show

$$(115) \qquad \lim_{h \to 0_+} \frac{1}{h} \mathbb{P}(\tau_{L_t + 1} \in [t, t + h] | (a(s))_{0 \leq s \leq t}) = \varphi(t)\theta_t(a(t))$$

in order to conclude the proof of Lemma 6.12. Let us define

$$\psi_{t+h}(t) := \mathbb{P}(\tau_{L_t+1} > t + h | C_t = 1). \tag{116}$$

(We explain the history of this notation in Remark 6.16).

Let us make the following observation, which follows from the branching structure underlying Definition 6.1. The cluster growth process started at time t from the state $C_t = k$ evolves until the first explosion according to the same dynamics as the sum of k i.i.d. copies of the cluster growth process started at time t from the state $C_t = 1$, where the explosion time of the sum is the minimum of the k i.i.d. explosion times for its summands. From this observation, we conclude

$$\mathbb{P}(\tau_{L_t+1} > t + h | C_t = k) = \psi_{t+h}(t)^k, \tag{117}$$

therefore by Lemma 6.9 we have

$$\mathbb{P}(\tau_{L_t+1} \in [t, t+h] | (a(s))_{0 \le s \le t}) = 1 - \mathbb{E}(\psi_{t+h}(t)^{|T_{a_t}^{\pi_t}|} | a_t). \tag{118}$$

Thus in order to prove (115), we only need to show

$$\lim_{h \to 0_+} \frac{1}{h}(1 - \mathbb{E}(\psi_{t+h}(t)^{|T_s^{\pi_t}|})) = \varphi(t)\theta_t(s), \quad s \ge 0. \tag{119}$$

First, we observe that

$$
\begin{aligned}
\varphi(t) &\overset{(87)}{=} \lim_{h \to 0_+} \frac{1}{h}\mathbb{P}(\tau_{L_t+1} \in [t, t+h]) \\
&\overset{(88),(117)}{=} \lim_{h \to 0_+} \frac{1}{h}\left(1 - \sum_{k=1}^{\infty} \psi_{t+h}(t)^k v_k(t)\right) \\
&\overset{(19)}{=} \lim_{h \to 0_+} \frac{1}{h}(1 - \mathbb{E}(\psi_{t+h}(t)^{|T^{\pi_t}|})).
\end{aligned}
\tag{120}
$$

If we denote $\varepsilon(t, h) = 1 - \psi_{t+h}(t)$, then

$$
\begin{aligned}
\varphi(t)h + o(h) &\overset{(120)}{=} 1 - \mathbb{E}((1 - \varepsilon(t, h))^{|T^{\pi_t}|}) \\
&\overset{(21),(112)}{=} \sqrt{2\varphi(t)}\sqrt{\varepsilon(t, h)} + o(\sqrt{\varepsilon(t, h)}), \quad h \to 0_+.
\end{aligned}
\tag{121}
$$

Taking the squares of both sides of (121) and rearranging the terms, we obtain $\frac{\varphi(t)}{2}h^2 + o(h^2) = \varepsilon(t, h) + o(\varepsilon(t, h))$, that is,

$$\psi_{t+h}(t) = 1 - \frac{\varphi(t)}{2}h^2 + o(h^2), \quad h \to 0_+. \tag{122}$$

Equation (47) of Lemma 5.1 together with $\phi_t = \varphi(t)$ (i.e., (21)) imply

$$\mathbb{E}((1 - \varepsilon)^{|T_s^{\pi_t}|}) = 1 - \sqrt{\varepsilon}\sqrt{2\varphi(t)}\theta_t(s) + o(\sqrt{\varepsilon})\varepsilon \to 0_+. \tag{123}$$

Putting the above formulas together we obtain (119):

$$
\begin{aligned}
1 - \mathbb{E}(\psi_{t+h}(t)^{|T_s^{\pi_t}|}) &\overset{(122)}{=} 1 - \mathbb{E}\left(\left(1 - \left(\frac{\varphi(t)}{2} + o(1)\right)h^2\right)^{|T_s^{\pi_t}|}\right) \\
&\overset{(123)}{=} \sqrt{\left(\frac{\varphi(t)}{2} + o(1)\right)h^2}\sqrt{2\varphi(t)}\theta_t(s) + o(h) \\
&= \varphi(t)\theta_t(s)h + o(h), \quad h \to 0_+.
\end{aligned}
$$

This completes the proof of Lemma 6.12. $\square$

REMARK 6.16. For any $y > t_{\text{gel}}$, the characteristic curve $\psi_y(t), t \in [0, y]$ is defined in [11], Lemma 3.5, to be the unique solution of the ODE

$$(124) \qquad \psi_y(y) = 1, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\psi_y(t) = \psi_y(t)\big(1 - X_t(\psi_y(t))\big), \qquad X_t(z) := \sum_{k=1}^{\infty} z^k v_k(t)$$

for which $\psi_y(t) < 1$ for all $t \in [0, y]$.

This definition and our definition (116) are equivalent by [11], (3.16).

6.9. *Age differential equations.* The goal of Section 6.9 is to prove Theorem 2.21.

Let us first note that for any $t \in [0, t_{\text{gel}})$ we have $\frac{\partial}{\partial t} \int f(s)\,\mathrm{d}\pi_t(s) = \int f'(s)\,\mathrm{d}\pi_t(s)$, since π_t is the distribution of the age a_t by Definition 6.1, and we have already seen at the beginning of Section 6.8 that $a_t = a_0 + t$ for any $t \in [0, t_{\text{gel}})$. This proves (20) for $t < t_{\text{gel}}$. We can thus assume $t > t_{\text{gel}}$ w.l.o.g. for the rest of Section 6.9.

Denote by $\widetilde{\pi}_t$ the distribution of A_t; cf. Definition 6.11. Integrating (90) with respect to $\widetilde{\pi}_t$ we obtain the weak formulation of the Kolmogorov forward equation of A:

$$\frac{\partial}{\partial t} \int f(s)\,\mathrm{d}\widetilde{\pi}_t(s) = \int f'(s)\,\mathrm{d}\widetilde{\pi}_t(s) - \varphi(t)\int f(s)\theta_t(s)\,\mathrm{d}\widetilde{\pi}_t(s) + \varphi(t)f(0)\int \theta_t(s)\,\mathrm{d}\widetilde{\pi}_t(s).$$

By Lemma 6.12, the marginal distributions of a_t and A_t agree, that is, we have $\widetilde{\pi}_t = \pi_t$. Plugging this identity in the above equation and using $\int \theta_t(s)\,\mathrm{d}\pi_t(s) = 1$, we obtain the desired (20) for $t > t_{\text{gel}}$. Noting that we have already proved (21) in Section 6.8, the proof of Theorem 2.21 is complete.

REFERENCES

[1] ALDOUS, D. J. (2000). The percolation process on a tree where infinite clusters are frozen. *Math. Proc. Cambridge Philos. Soc.* **128** 465–477. MR1744108 https://doi.org/10.1017/S0305004199004326

[2] ALEKSANDROV, A. B., JANSON, S., PELLER, V. V. and ROCHBERG, R. (2002). An interesting class of operators with unusual Schatten–von Neumann behavior. In *Function Spaces, Interpolation Theory and Related Topics* (*Lund*, 2000) (M. Cwikel et al., eds.) 61–149. de Gruyter, Berlin. MR1943283

[3] BAK, P., TANG, C. and WIESENFELD, K. (1988). Self-organized criticality. *Phys. Rev. A (3)* **38** 364–374. MR0949160 https://doi.org/10.1103/PhysRevA.38.364

[4] BENJAMINI, I., LYONS, R. and SCHRAMM, O. (2015). Unimodular random trees. *Ergodic Theory Dynam. Systems* **35** 359–373. MR3316916 https://doi.org/10.1017/etds.2013.56

[5] BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. MR2337396 https://doi.org/10.1002/rsa.20168

[6] BUFFET, E. and PULÉ, J. V. (1990). On Lushnikov's model of gelation. *J. Stat. Phys.* **58** 1041–1058. MR1049055 https://doi.org/10.1007/BF01026562

[7] BUFFET, E. and PULÉ, J. V. (1991). Polymers and random graphs. *J. Stat. Phys.* **64** 87–110. MR1117648 https://doi.org/10.1007/BF01057869

[8] CERF, R. and GORNY, M. (2014). A Curie–Weiss model of self-organized criticality. *Ann. Probab.* **44** 444–478.

[9] CRANE, E. (2018). Steady state clusters and the Ráth–Tóth mean field forest fire model. Available at arXiv:1809.03462 [math.PR].

[10] CRANE, E. (2020). Well-posedness of the mean field forest fire age evolution equation. Available at arXiv:2007.05807 [math.AP].

[11] CRANE, E., FREEMAN, N. and TÓTH, B. (2015). Cluster growth in the dynamical Erdős–Rényi process with forest fires. *Electron. J. Probab.* **20** 101. MR3407218 https://doi.org/10.1214/EJP.v20-4035

[12] DROSSEL, B. and SCHWABL, F. (1992). Self-organized critical forest fires model. *Phys. Rev. Lett.* **69** 1629–1632.

[13] DÜRRE, M. (2006). Existence of multi-dimensional infinite volume self-organized critical forest-fire models. *Electron. J. Probab.* **11** 513–539. MR2242654 https://doi.org/10.1214/EJP.v11-333

[14] DÜRRE, M. (2006). Uniqueness of multi-dimensional infinite volume self-organized critical forest-fire models. *Electron. Commun. Probab.* **11** 304–315. MR2266720 https://doi.org/10.1214/ECP.v11-1229

[15] DÜRRE, M. (2009). Self-organized Critical Phenomena; Forest Fire and Sandpile Model Ph.D. thesis Ludwig-Maximilians-Universität München.

[16] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. MR0125031

[17] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II. Second Edition.* Wiley, New York. MR0270403

[18] FOURNIER, N. and LAURENÇOT, P. (2009). Marcus–Lushnikov processes, Smoluchowski's and Flory's models. *Stochastic Process. Appl.* **119** 167–189. MR2485023 https://doi.org/10.1016/j.spa.2008.02.003

[19] GRAF, R. (2014). A forest-fire model on the upper half-plane. *Electron. J. Probab.* **19** 8. MR3164761 https://doi.org/10.1214/EJP.v19-2625

[20] JÁRAI, A. A. (2018). Sandpile models. *Probab. Surv.* **15** 243–306. MR3857602 https://doi.org/10.1214/14-PS228

[21] KISS, D., MANOLESCU, I. and SIDORAVICIUS, V. (2015). Planar lattices do not recover from forest fires. *Ann. Probab.* **43** 3216–3238. MR3433580 https://doi.org/10.1214/14-AOP958

[22] MERLE, M. and NORMAND, R. (2014). Self-organized criticality in a discrete model for Smoluchowski's equation. Available at arXiv:1410.8338 [math.PR].

[23] MERLE, M. and NORMAND, R. (2015). Self-organized criticality in a discrete model for Smoluchowski's equation with limited aggregations. Available at arXiv:1509.00934 [math.PR].

[24] MODE, C. J. (1971). *Multitype Branching Processes. Theory and Applications.* Elsevier, New York. MR0279901

[25] NORMAND, R. and ZAMBOTTI, L. (2011). Uniqueness of post-gelation solutions of a class of coagulation equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **28** 189–215. MR2784069 https://doi.org/10.1016/j.anihpc.2010.10.005

[26] RÁTH, B. (2009). Mean field frozen percolation. *J. Stat. Phys.* **137** 459–499. MR2564286 https://doi.org/10.1007/s10955-009-9863-5

[27] RÁTH, B. and TÓTH, B. (2009). Erdős–Rényi random graphs + forest fires = self-organized criticality. *Electron. J. Probab.* **14** 1290–1327. MR2511285 https://doi.org/10.1214/EJP.v14-653

[28] REED, M. and SIMON, B. (1980). *Methods of Modern Mathematical Physics. I: Functional Analysis*, 2nd ed. Academic Press, New York. MR0751959

[29] SWART, J. M. (2013). Duality and Intertwining of Markov Chains. ALEA in Europe School, CIRM, Luminy (Marseille). Available at: http://staff.utia.cas.cz/swart/dualtwine.pdf.

[30] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX—1989. Lecture Notes in Math.* **1464** 165–251. Springer, Berlin. MR1108185 https://doi.org/10.1007/BFb0085169

[31] VAN DEN BERG, J. and BROUWER, R. (2004). Self-destructive percolation. *Random Structures Algorithms* **24** 480–501. MR2060632 https://doi.org/10.1002/rsa.20003

[32] VAN DER HOFSTAD, R. *Random Graphs and Complex Networks. Vol.* 2. *Cambridge Series in Statistical and Probabilistic Mathematics.* To appear. Draft available at http://www.win.tue.nl/~rhofstad/.

[33] YEO, D. (2016). Self-organised criticality in random graph processes DPhil thesis, University of Oxford. Available at https://ora.ox.ac.uk.

[34] YEO, D. (2018). Frozen percolation on inhomogeneous random graphs. Available at arXiv:1810.02750 [math.PR].