

# A growth-fragmentation-isolation process on random recursive trees

Vincent Bansaye\*, Chenlin Gu†, Linglong Yuan‡

September 14, 2021

## Abstract

We consider a random process on recursive trees, with three types of events. Vertices give birth at a constant rate (growth), each edge may be removed independently (fragmentation of the tree) and clusters are frozen with a rate proportional to their size (isolation of connected component). A phase transition occurs when the isolation is able to stop the growth fragmentation process and cause extinction. When the process survives, we characterize its growth and prove that the empirical measure of clusters a.s. converges to a limit law on recursive trees. We exploit the branching structure associated to the size of clusters, which is inherited from the splitting property of random recursive trees. This issue is motivated by the control of epidemics and contact-tracing where clusters correspond to subtrees of infected individuals that can be identified and isolated.

**Key words:** branching process, random recursive tree, law of large number.

*MSC (2010):* 60J27, 60J85, 60J80

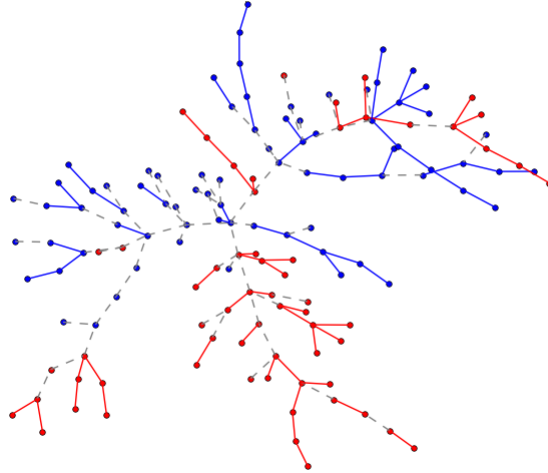


Figure 1: An illustration of the growth-fragmentation-isolation process with 62 active vertices (in red) and 77 inactive vertices (in blue).

\*CMAP, Ecole Polytechnique, IPP, 91128 Palaiseau, France

†DMA/ENS and NYU-ECNU Institute of Mathematical Sciences, NYU Shanghai

‡Department of Mathematical Sciences, University of Liverpool; Xi'an Jiaotong-Liverpool University

# 1 Introduction

The evolution of random trees is motivated by various fields : algorithmic, queuing systems, population modeling... The random deletion of edges of a tree has been studied in particular by [20, 8]. Initially, Meir and Moon [20] were interested in the number of steps needed to isolate a distinguished vertex in a random Cayley tree, when the deleted edge is chosen uniformly. Bertoin [9] and Marzouk [19] have then studied processes where sets of vertices can be fired. More precisely, a connected component of the graph is removed (i.e. isolated) at each step. This component is determined by a uniform choice among the vertices. Such dynamics combine the fragmentation of the tree (when an edge is deleted) and the isolation of connected components of the tree (when a vertex provokes a fire).

In this work, we are interested in the long time behavior of such dynamics when the random recursive tree grows, following a binary branching process. Our original motivation is the description of control policies of an epidemic. The growth of the infected population is modeled by a Yule process in this work, i.e. a binary Markov branching process. The discrete structure of the Yule tree is a random recursive tree. The newly connected vertices are the new contamination that can be detected. The fragmentation occurs when one edge is removed, interpreted as the loss of infector-infectee information, in the tracing of infected individuals along time. Growth and fragmentation will generate connected components, called here clusters. Each cluster is a set of connected infected individuals that can be identified, and isolated all together as soon as one individual among the cluster is detected. An isolated cluster is frozen, in the sense that no more event happen to it. Indeed, the individuals found in the tracing system are supposed to stop contamination's. The second author has studied a similar model with numerical simulations in [13] to estimate the propagation of Covid-19. For similar motivations, we mention [17, 7], which also exploit a branching structure. But the models and the approaches differ. In particular the family of connected components is not described in these papers and the approach in [17] does not allow for backward tracing (i.e. tracing of an ancestor when the descendant is tested), while more general regarding the characteristics of epidemics.

Our growth-fragmentation-isolation model (GFI) is thus interpreted as a simple branching process to study the effect of identification-tracing-isolation strategy in the context of pandemic, with loss of contact information along time. The approximation of the outbreak by a branching process is classical in the first stages of an epidemic, when the whole population is large and the infected individuals can be neglected, see for instance [3]. We also control in this work the speed of convergences of the number of infectees and the empirical measure and thus check that our asymptotic profile can indeed describe the outbreak at the "beginning" of the epidemics. Besides, more epidemiological and control features could be incorporated to reflect the current reality in subsequent works. No recovery happens in our setting and we consider large homogeneously mixing population. We expect extensions of our results on the long time behavior to take into account these features, even if the probabilistic structure at a fixed time seem to be less tractable.

Random recursive trees (RRT) have a nice splitting property that allows us to characterize a cluster by its size in this model. More precisely, considering the collection of the active (non-isolated) clusters, the size process is a branching process with a countable set of types. At fixed time, conditionally on the size of the clusters, the collection of clusters are independent RRTs. We can then study the ergodic properties of first moment semigroup of this branching process and obtain a phase transition depending on the value of the maximal eigenvalue (Malthusian growth rate). We describe the a.s. behavior of the process when the active clusters survive, proving a strong law of large numbers for distribution of types and

Kesten-Stigum type result for the growth of the population. We can also characterize the a.s. behavior of the process of isolated clusters when the active clusters survive, which is by itself non-Markovian. The fact that the number of types is infinite and the loss of Markov property for the isolated clusters poses mathematical difficulties, in particular to get strong convergence. We refer to [2, 1, 12] for classical references on strong law of large numbers of some classes of multitype branching processes. For the asymptotic analysis of the mean behavior and weak convergences and estimation of the speed of convergence, we follow a now well developed for branching processes with infinite number of types. Roughly, it relies on the ergodic properties of the size of a typical cluster and the fact that the common ancestor of two samples at large time is found at small times. We refer e.g. to [10, 11, 5, 18, 23, 14] and references therein for related works on the asymptotic analysis of growth fragmentation processes. We exploit here the fact that large clusters fragment fast and give one small cluster and one large cluster with high probability. Together with isolation, it allows to control the size of a typical cluster and eigenelements. In particular, we prove that the harmonic function is bounded and large clusters have no major impact on the growth of epidemics. Indeed, large clusters are isolated before creating too many small clusters, since isolation occurs here at the same scale as fragmentation. Once active clusters are well described, we can treat the isolated clusters using an additive functional.

Let us give an informal description of our model. Starting from the patient zero, the virus infects individuals one after another and forms a tree in the course of infection, which is called the cluster. We suppose that one-to-one spread is the only way of infection. At this step, there is a unique cluster which connects everyone in the chain of infection. But for various reasons, including memory and storage of information, links of infection can be lost along time and become inaccessible when needed. Therefore, the identifiable infected tree is actually fragmented in clusters. In each cluster, there is a root and everyone else was infected by someone in the cluster. Each infected individual is detected at a fixed rate. Then all people in the same cluster as this detected individual are put into isolation instantaneously to slow down the epidemic.

We describe now the model more formally. We introduce a stochastic process on a dynamic tree  $G_t = (V_t, E_t)$  with two functions  $\Psi_t : V_t \rightarrow \{0, 1\}$ ,  $\eta_t : E_t \rightarrow \{0, 1\}$ .

- We identify the vertex set  $V_t$  as the set of patients (individuals infected so far), and label them with the infection time  $v \in \mathbb{R}_+$ . The function  $\Psi_t$  represents the state of a vertex, where vertex  $v$  is *active* if  $\Psi_t(v) = 1$  and  $v$  is *inactive* if  $\Psi_t(v) = 0$ . Only active vertices can infect new ones.
- We identify the edge set  $E_t$  as the set of infection links between patients, and the function  $\eta_t$  indicates the information on this edge. For an edge  $e$ , we say  $e$  is open if  $\eta_t(e) = 1$  which means the infection link can still be retrieved (i.e. it can be found who infected the infectee), otherwise  $\eta_t(e) = 0$  and it is *closed* (i.e. no one knows who infected the infectee). A set of vertices connected by open edges is called a *cluster*.

The GFI process  $(G_t, \Psi_t, \eta_t)_{t \geq 0}$  is a Markov jump process, starting from an active vertex as patient zero  $V_0 = \{0\}$ ,  $\Psi_0(0) = 1$ , governed by three positive parameters  $(\beta, \gamma, \theta) \in \mathbb{R}_+^3$  which represent three types of events:

- Infection (growth): every active vertex  $v$  independently attaches a new vertex in an exponential time with parameter  $\beta$ . When a new vertex  $u$  is created and attached, it is active (i.e.  $\Psi_t(u) = 1$ ) and the edge  $\{u, v\}$  is open (i.e.  $\eta_t(\{u, v\}) = 1$ ).
- Information decay (fragmentation): every open edge  $e$  independently becomes closed in an exponential time with parameter  $\gamma$ .

- Confirmation and contact-tracing (isolation): every active vertex independently gets “confirmed” in an exponential time with parameter  $\theta$ , then its associated cluster is isolated and every vertex on this cluster becomes inactive.

See Figure 2 for an illustration of this model. If  $\gamma = \theta = 0$ , this is the well-known Yule tree process; for a static model without isolation, it is the percolation model on the tree. As every vertex is indexed by its infection time, every cluster is a labeled *recursive tree* (see Section 3.2 for rigorous definition).

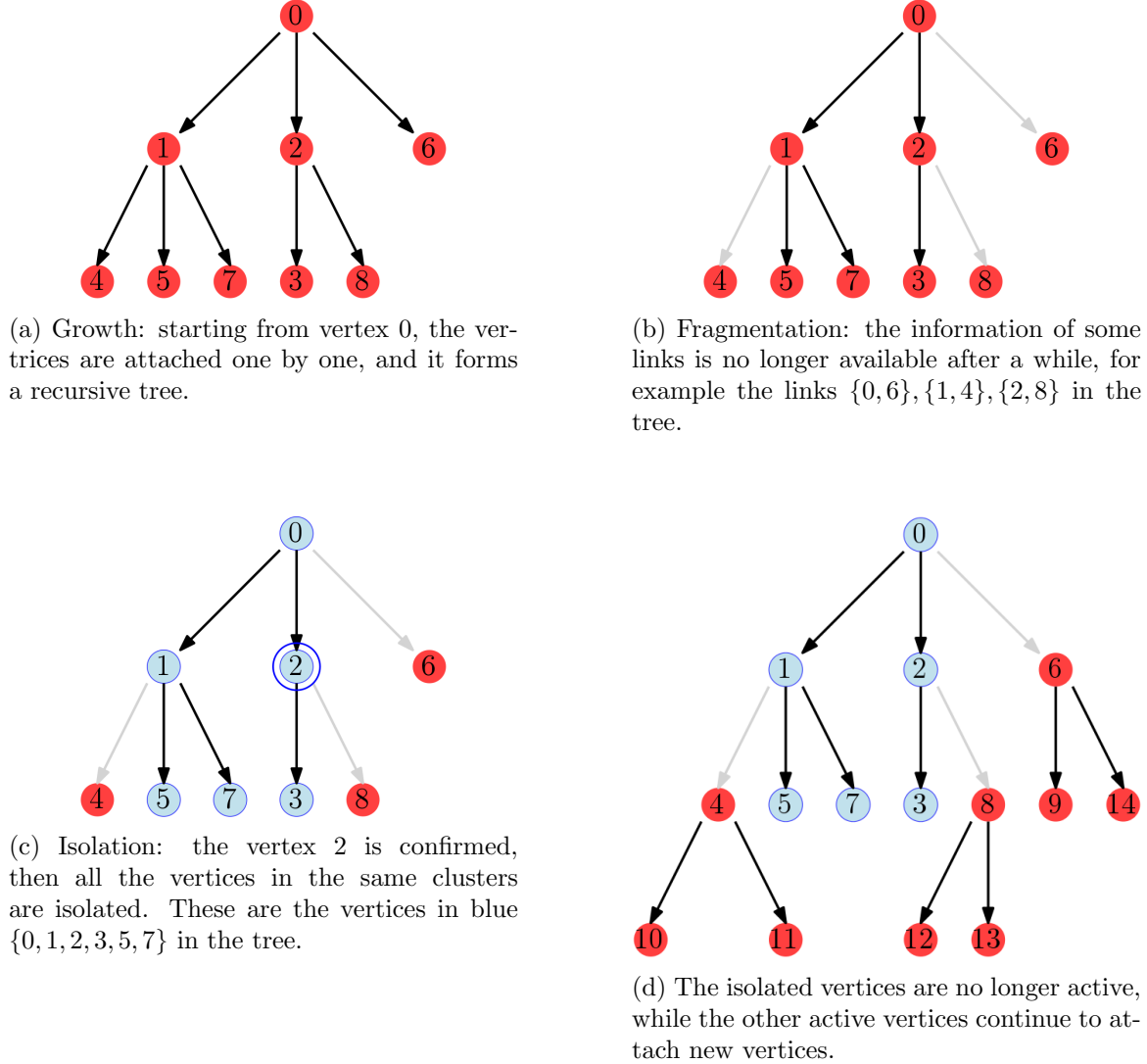


Figure 2: An illustration of GFI process.

Because the vertices on a cluster have the same state, it is very natural to decompose the dynamic tree  $G_t$  into clusters of individuals connected by open edges: for an isolated cluster, we call it *inactive cluster*; otherwise, it is an *active cluster*. In this paper, we use *isolated* and *inactive* interchangeably for clusters and also for vertices/patients/infected individuals. We denote by  $(\mathcal{X}_t, \mathcal{Y}_t)_{t \geq 0}$  the associated *cluster process*, where  $\mathcal{X}_t$  is the set of active clusters and  $\mathcal{Y}_t$  is the set of inactive clusters

$$\begin{aligned}\mathcal{X}_t &= \{\mathcal{C} \mid \mathcal{C} \text{ is a cluster in } G_t; \forall v \in \mathcal{C}, \Psi_t(v) = 1\}, \\ \mathcal{Y}_t &= \{\mathcal{C} \mid \mathcal{C} \text{ is a cluster in } G_t; \forall v \in \mathcal{C}, \Psi_t(v) = 0\},\end{aligned}$$

The process  $(G_t, \Psi_t, \eta_t)_{t \geq 0}$  stops when  $\mathcal{X}_t$  is empty. We denote by  $\tau$  the corresponding stopping time :

$$\tau := \inf\{t \mid \mathcal{X}_t = \emptyset\}, \quad (1.1)$$

and the event  $\{\tau < \infty\}$  is called *extinction*, while  $\{\tau = \infty\}$  is called *survival*.

In this paper, we are motivated by the following questions. For which values of the parameters  $(\beta, \gamma, \theta)$  does the propagation of the epidemic stop (extinction of active clusters) ? If the epidemic propagates (outbreak), what is the long time behavior of the population of active and inactive clusters ? We give some answers to these questions by first determining the asymptotic behavior of the first moment semigroup and show exponential convergence of the renormalized semigroup. The asymptotic behavior is driven by the maximal eigenvalue of the first moment semigroup associated to the active clusters, which yields the Malthusian exponent. When this value is negative (subcritical case) or zero (critical case), the population of active clusters becomes extinct in finite time. It corresponds to the fact that the isolation process is strong enough to stop the epidemic. When this maximal value is positive, the population of active clusters tends to infinity a.s. on the survival event, with an exponential speed given by the Malthusian exponent. We shed some light on the genealogical structure of clusters and describe the asymptotic behavior of the empirical distribution. We prove that a.s. we get a collection of recursive trees whose sizes are distributed following the left eigenvector associated to the maximal eigenvalue of the semigroup. Besides, at conditionally on their size, these clusters are independent RRT.

## 2 Main results

We obtain first the classification for branching structures, where extinction occurs in the subcritical and critical cases. As mentioned above and classically for branching processes, this is related to the *Malthusian* exponent which describes the (mean) exponential growth (or decrease). This growth rate coincides for active and inactive clusters.

**Theorem 1** (Malthusian exponent). *The following limits exist and coincide and are finite*

$$\lambda := \lim_{t \rightarrow \infty} \frac{1}{t} \log(\mathbb{E}[|\mathcal{X}_t|]) = \lim_{t \rightarrow \infty} \frac{1}{t} \log(\mathbb{E}[|\mathcal{Y}_t|]) \in (-\infty, \infty).$$

Here  $|\mathcal{X}_t|$  (resp.  $|\mathcal{Y}_t|$ ) is the number of active (resp. inactive) clusters at time  $t$ . If  $\lambda \leq 0$ , then extinction occurs a.s. :  $\mathbb{P}[\tau < \infty] = 1$ . Otherwise, survival occurs with positive probability  $\mathbb{P}[\tau = \infty] > 0$ .

The Malthus exponent  $\lambda$  corresponds to the maximal eigenvalue of the first moment semigroup and is also called *Perron's root*. The fact that the cluster size can be any positive integer leads us to using techniques for ergodic behavior in infinite dimension, where the control of large sizes is involved. As usual, irreducibility on the state of sizes ensures that the value  $\lambda$  does not depend on the initial state. The diagram of these different phases is illustrated in Figure 3, for fixed  $\beta > 0$ . Note that  $\theta \geq \min(\beta, \gamma)$  implies a.s. extinction since in the case  $\theta \geq \beta$ , individuals are detected faster than they contaminate and in the case  $\theta \geq \gamma$ , isolation is faster than fragmentation.

To study the asymptotic increment of  $|\mathcal{X}_t|$  and  $|\mathcal{Y}_t|$ , we introduce the following *size process*  $(X_t, Y_t)_{t \geq 0}$ , where two empirical measures count the clusters of different sizes

$$X_t = \sum_{C \in \mathcal{X}_t} \delta_{|C|}, \quad Y_t = \sum_{C \in \mathcal{Y}_t} \delta_{|C|}. \quad (2.2)$$

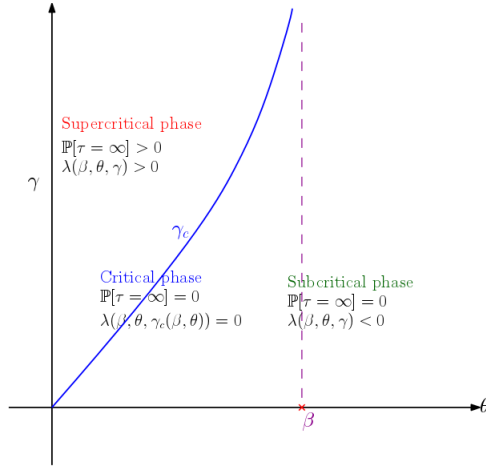


Figure 3: An illustration of different phases.

Here we denote by  $|\mathcal{C}|$  the number of vertices in the cluster  $\mathcal{C}$ , and we call it *the size of cluster*. This reduction of the state space of the empirical measure is simplifying the analysis of the process.

The process  $(X_t)_{t \geq 0}$  is still a branching Markov process with respect to its natural filtration. This comes from the fact that every cluster in  $(\mathcal{X}_t)_{t \geq 0}$  is a random recursive tree and its nice *splitting property* (see Proposition 1 in Section 3.2), which says that the subclusters after fragmentation are still random recursive trees. Up to the times of infection, we can thus replace the study of our GFI process  $(G_t, \Psi_t, \eta_t)_{t \geq 0}$  by the study of the branching process of sizes  $(X_t, Y_t)_{t \geq 0}$ ; see Figure 4.

We prove the following strong law of large numbers in the supercritical case. This provides the asymptotic behavior of  $\langle X_t, f \rangle = \sum_{\mathcal{C} \in \mathcal{X}_t} f(|\mathcal{C}|)$ , where  $f$  has at most polynomial growth, i.e.  $\exists p \in \mathbb{N}_+$  such that  $\limsup_{n \rightarrow \infty} |f(n)|/n^p < \infty$ . In particular,  $f = \mathbf{1}_m$  yields the number of active clusters of size  $m$ , while  $f(n) = n \geq 1$  for all  $n$  provides the number of infected active individuals.

**Theorem 2** (Law of large numbers for  $(X_t)_{t \geq 0}$ ). *Assume that  $\lambda > 0$ . Then there exists a probability distribution  $\pi$  on  $\mathbb{N}_+$  and a random variable  $W \geq 0$ , such that for any function  $f : \mathbb{N}_+ \rightarrow \mathbb{R}$  of at most polynomial growth, we have*

$$e^{-\lambda t} \langle X_t, f \rangle \xrightarrow{t \rightarrow \infty} W \langle \pi, f \rangle, \quad \text{a.s. and in } L^2. \quad (2.3)$$

Besides,  $\{\tau = \infty\} = \{W > 0\}$  a.s. and on this event

$$\frac{\langle X_t, f \rangle}{\langle X_t, 1 \rangle} \xrightarrow{t \rightarrow \infty} \langle \pi, f \rangle \quad \text{a.s.} \quad (2.4)$$

Using the previous theorem about active clusters and the eigenelements of the first moment semigroup, we can derive the asymptotic behavior of isolated clusters. They are created from the active clusters by a size biased rate and we introduce

$$\tilde{\pi}(n) := \frac{\pi(n)n}{\sum_{j=1}^{\infty} \pi(j)j}. \quad (2.5)$$

**Corollary 1** (Law of large number for  $(Y_t)_{t \geq 0}$ ). *For any function  $f : \mathbb{N}_+ \rightarrow \mathbb{R}$  of at most polynomial growth, we have that*

$$e^{-\lambda t} \langle Y_t, f \rangle \xrightarrow{t \rightarrow \infty} W \left( \frac{\theta}{\lambda} \right) \left( \sum_{j=1}^{\infty} \pi(j)j \right) \langle \tilde{\pi}, f \rangle, \quad \text{almost surely and in } L^2,$$

and

$$\frac{\langle Y_t, f \rangle}{\langle Y_t, 1 \rangle} \xrightarrow{t \rightarrow \infty} \langle \tilde{\pi}, f \rangle, \quad \text{almost surely on } \{\tau = \infty\}.$$

Regarding our motivations for tracing in epidemics, the data of inactive clusters are observable, while the active clusters are not. The equation (2.5) enlightens a size biased phenomenon in observations and sampling, which is a direct consequence of the modeling. Indeed, each active vertex becomes inactive at the same rate and every active cluster gets isolated at rate proportional to its size. The size-biased sampling results in the size-biased transformation of  $\pi$  to  $\tilde{\pi}$ .

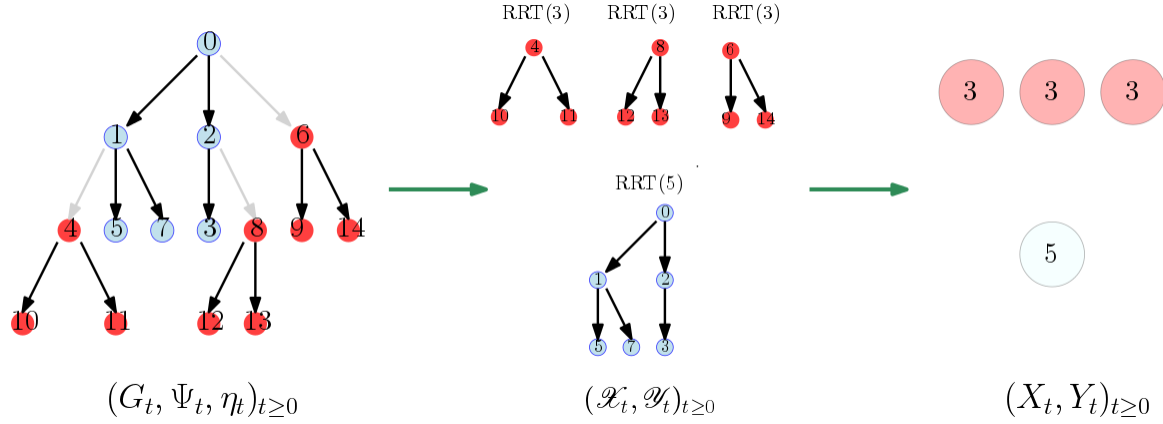


Figure 4: An illustration of the main idea to study GFI process  $(G_t, \Psi_t, \eta_t)_{t \geq 0}$ . We can decompose the graph into clusters and study the cluster process  $(\mathcal{X}_t, \mathcal{Y}_t)_{t \geq 0}$ , where every cluster is a RRT whose law only depends on the size. At the end we only need to study size process  $(X_t, Y_t)_{t \geq 0}$  which captures essential information.

Finally, we can also describe the typical genealogical structure of infections in clusters. Recall that every vertex  $v \in \mathbb{R}$  is labeled by its infection time and every cluster is a recursive tree. We consider these clusters up to an equivalence relation, which consists in keeping the order between vertices but forgetting the their infection times, see Section 3.2 for a rigorous definition. We denote by  $\mathcal{T}$  the space of equivalent classes of all sizes and  $T_\pi$  the random recursive tree with size distribution  $\pi$ . With a slight abuse of notations, for any recursive tree  $t$ , we write  $\mathbf{t} \in \mathcal{T}$  its equivalent class and for any function  $f : \mathcal{T} \rightarrow \mathbb{R}$ , we let  $f(t) = f(\mathbf{t})$ . We also write  $|\mathbf{t}|$  for the number of vertices in the equivalent class  $\mathbf{t}$ . We prove the following counterpart of the asymptotic result for the empirical measure of clusters.

**Theorem 3** (Limit of empirical measure of clusters). *Consider any  $p > 0$  and  $f : \mathcal{T} \rightarrow \mathbb{R}$  such that*

$$\sup_{\mathbf{t} \in \mathcal{T}} \frac{|f(\mathbf{t})|}{|\mathbf{t}|^p} < \infty.$$

*Then on the event  $\{\tau = \infty\}$*

$$\frac{1}{|\mathcal{X}_t|} \sum_{C \in \mathcal{X}_t} f(C) \xrightarrow{t \rightarrow \infty} \mathbb{E}[f(T_\pi)], \quad \frac{1}{|\mathcal{Y}_t|} \sum_{C \in \mathcal{Y}_t} f(C) \xrightarrow{t \rightarrow \infty} \mathbb{E}[f(T_{\tilde{\pi}})] \quad a.s..$$

The rest of the paper is organized as follows: in Section 3 we introduce some notations in this paper, and also recall the key splitting property of the random recursive tree. The existence of Perron's root  $\lambda$  is proved in Section 4 using Lyapunov functions. We then prove Theorem 1. Section 5 is devoted to the strong convergences and we will prove Theorem 2, Corollary 1 and Theorem 3. Finally, we give some further discussions in Section 6.

### 3 Preliminaries

In this part, we introduce notations and the key reduction of the study. Our model is a tree-type branching process and can be seen “a branching within branching”. We reduce the study of this branching process  $(G_t, \Psi_t, \eta_t)_{t \geq 0}$  to that of its sizes  $(X_t, Y_t)_{t \geq 0}$ , so types are now integers.

We denote by  $C$  a positive constant and its value may change in different contexts in the paper. We use the notation  $[x^p]$  for the polynomial function such that  $[x^p](n) = n^p, n \geq 1$ . When  $p = 1$  it is just the identity function and we omit the index, i.e.  $[x](n) = n$ . We introduce  $\mathcal{B}$  the set of functions from  $\mathbb{N}_+$  to  $\mathbb{R}$  with at most polynomial growth:

$$\mathcal{B} := \left\{ f : \mathbb{N}_+ \rightarrow \mathbb{R}, \exists p \geq 1 \text{ such that } \sup_{n \geq 1} |f(n)|/n^p < \infty \right\}. \quad (3.6)$$

Let  $\mathcal{B}_p$  be the set of normalized functions as above with  $p \geq 1$  fixed

$$\mathcal{B}_p := \left\{ f : \mathbb{N}_+ \rightarrow \mathbb{R}, \sup_{n \geq 1} |f(n)|/n^p \leq 1 \right\}. \quad (3.7)$$

#### 3.1 UHN labelling of clusters

We introduce the notation for the genealogical tree of the clusters. Recall that every vertex is labeled by the infection time. For any cluster  $\mathcal{C}$  (active or inactive), we call the vertex with the minimum label *the root* of  $\mathcal{C}$ , and denote it by  $\text{root}(\mathcal{C})$ . We can also label every cluster by the Ulam-Harris-Neveu notation that

$$\mathcal{U} = \bigcup_{n \geq 0} \{1, 2\}^n,$$

where an element  $u \in \mathcal{U}$  is called *label* or *word*. For the initial cluster, we use the label  $\emptyset$  as a convention for it. Then by induction, for any cluster  $\mathcal{C}$  labeled by a word  $u \in \mathcal{U}$ : this label is unchanged during the growth of the clusters (infection); this label dies (it does not belong to set of active clusters any longer) when it is isolated; this label is replaced by two labels  $u1$  and  $u2$  when it is fragmentation. By convention,  $u1$  is the label containing  $\text{root}(\mathcal{C})$  and is called *the first child*, while  $u2$  is for *the second child*.

Every cluster except the initial cluster has a parent and this latter is unique. There exists a partial order  $\preceq$  on  $\mathcal{U}$  defined by the dictionary order, i.e. for two words  $u$  and  $uv$ , the former is an *ancestor* of the later, while the later is an *descendant* of the former, and we note  $u \preceq uv$ . For two words  $u, v \in \mathcal{U}$ , we denote by  $u \wedge v$  the most recent common ancestor of  $u$  and  $v$ .

We denote by  $\mathcal{U}_t$  the collection of labels of active clusters at time  $t$ , while  $\mathcal{U}_t^\dagger$  gathers the labels of inactive clusters at this time. For  $u \in \mathcal{U}$ , we use  $\mathcal{U}(u), \mathcal{U}_t(u), \mathcal{U}_t^\dagger(u)$  to represent respectively the descendants of  $u$ , the active descendants of  $u$  and the inactive descendants of  $u$  at time  $t$ . Finally, if  $u \in \mathcal{U}_t$  (or  $u \in \mathcal{U}_t^\dagger$ ), we denote by  $\mathcal{X}_t^u$  (or  $\mathcal{Y}_t^u$ ) for its associated cluster, and  $X_t^u$  (or  $Y_t^u$ ) its size, i.e.  $X_t^u = |\mathcal{X}_t^u|$  (or  $Y_t^u = |\mathcal{Y}_t^u|$ ). We also use  $X_t(n)$  (or  $Y_t(n)$ ) for the number of active clusters (or inactive clusters) of size  $n$  at time  $t$ . With these notations, we have  $X_t = \sum_{u \in \mathcal{U}_t} \delta_{X_t^u}$  and a useful identity that

$$\langle X_t, f \rangle = \sum_{\mathcal{C} \in \mathcal{X}_t} f(|\mathcal{C}|) = \sum_{u \in \mathcal{U}_t} f(X_t^u) = \sum_{n=1}^{\infty} X_t(n) f(n),$$

for any function  $f$  from  $\mathbb{N}_+$  to  $\mathbb{R}$ .

### 3.2 Random recursive trees

The genealogy in the cluster is given by recursive tree and here we define some notations for this object. Given a set  $V = \{a_1, \dots, a_n\} \subset \mathbb{R}$  with increasing order  $a_1 < a_2 < \dots < a_n$ , a *recursive tree*  $t$  on  $V$  is a rooted tree labeled by  $V$  such that for any  $a_i, 2 \leq i \leq n$ , the path from  $a_1$  to  $a_i$  is increasing. Thus, the descendants of each vertex has a larger label. The minimal element  $a_1$  is called the *root* of  $t$ . The collection of all the recursive trees on  $V$  is denoted by  $\mathcal{T}_V$  and it is clear  $|\mathcal{T}_V| = (|V| - 1)!$ .

We also define the equivalence relation  $\sim$  between the recursive trees on different ordering sets. Denoting by  $t_1$  a recursive tree on  $V_1$  and  $t_2$  a recursive tree on  $V_2$ , then  $t_1 \sim t_2$  if and only if there exists an order-preserving function  $\psi : V_1 \rightarrow V_2$ , such that  $\psi$  is also a bijection between the graphs  $t_1$  and  $t_2$ . We denote by  $\mathcal{T}_n$  the set of recursive trees of size  $n$  up to the equivalence relation  $\sim$ , and use the recursive trees defined on  $\{1, \dots, n\}$  as a representative of the equivalent class; see Figure 5 for an example of  $\mathcal{T}_4$ . Finally, we define the space of finite recursive trees

$$\mathcal{T} := \bigcup_{n=1}^{\infty} \mathcal{T}_n. \quad (3.8)$$

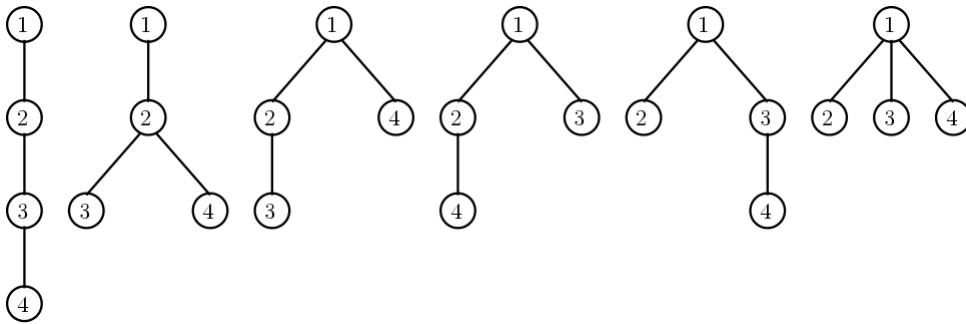


Figure 5: All the recursive trees (as representatives of equivalent classes) in  $\mathcal{T}_4$ .

A (*uniform*) *random recursive tree (RRT)* of size  $n$  is a random element chosen uniformly in  $\mathcal{T}_n$ . We denote by  $T_n$  this random equivalent class. With a slight abuse, RRT can refer both to the equivalent class or a specific labeling (for instance with the first integers). Since  $\mathcal{T}$  defined in eq. (3.8) contains only countably many elements, the space  $\mathcal{T}$  is a Polish space under the trivial distance. We denote by  $C_b(\mathcal{T})$  the bounded function on  $\mathcal{T}$  and we can construct more general probability measures rather than  $\mathbb{P}_{T_n}$  on  $\mathcal{T}$ . For example, for any  $\nu$  a probability measure on  $\mathbb{N}_+$ , we use the notation  $T_\nu$  to represent a random variable on  $\mathcal{T}$ , such that we sample first the size by  $\nu$ , then sample an equivalent class uniformly given its size, i.e. for any  $f \in C_b(\mathcal{T})$ ,

$$\mathbb{E}[f(T_\nu)] = \sum_{n=1}^{\infty} \nu(n) \mathbb{E}[f(T_n)] = \sum_{n=1}^{\infty} \nu(n) \left( \frac{1}{(n-1)!} \sum_{t \in \mathcal{T}_n} f(t) \right). \quad (3.9)$$

This is the rigorous definition for the expression in Theorem 3.

There are many ways to construct  $T_n$ . One classical construction is the recursive approach: let  $T_1$  be the tree with the single vertex 1, and construct  $T_{k+1}$  by attaching the vertex labeled  $(k+1)$  uniformly onto a vertex of  $T_k$ . This construction explains why our infection process (Yule process), conditioned on its size, is a RRT. Indeed each individual

contaminates a new individual with the same rate, which amounts to attaching a new vertex to a uniformly chosen vertex of the tree, independently from the previous choices.

The key property of RRT that we need is the splitting property. Its proof can be found in [20] or [8]. For the sake of completeness, we give it here and explain what role this property plays in our model in the next section.

**Proposition 1** (Splitting property). *Let  $n \geq 2$  and  $T_n$  the canonical random recursive tree of size  $n$ . We choose uniformly one edge in  $T_n$  and remove it. Then  $T_n$  is split into two subtrees  $T_n^0$  and  $T_n^*$ , corresponding to two connected components, where  $T_n^0$  contains the root of  $T_n$  and  $T_n^*$  does not. Then we have*

$$\mathbb{P}[|T_n^*| = j] = \frac{n}{n-1} \frac{1}{j(j+1)}, \quad j = 1, 2, \dots, n-1. \quad (3.10)$$

Furthermore, conditionally on  $|T_n^*| = j$ ,  $T_n^0$  and  $T_n^*$  are two independent RRT's of size respectively  $(n-j)$  and  $j$ .

*Proof.* Recall that  $T_n$  has the canonical representation, i.e. it is defined on  $\{1, \dots, n\}$ . It is clear that after the splitting, the two subtrees are also recursive trees, so it suffices to calculate the joint law of subtrees. Given that  $|T_n^*| = j, 1 \leq j \leq n-1$ , and for any possible two equivalent classes  $\mathbf{t}_1 \in \mathcal{T}_j, \mathbf{t}_2 \in \mathcal{T}_{n-j}$  of sizes respectively  $j$  and  $n-j$ , we calculate the probability of  $\{T_n^* \sim \mathbf{t}_1, T_n^0 \sim \mathbf{t}_2\}$

$$\mathbb{P}[T_n^* \sim \mathbf{t}_1, T_n^0 \sim \mathbf{t}_2] = \sum_{k=1}^{n-j} \mathbb{P}[T_n^* \sim \mathbf{t}_1, T_n^0 \sim \mathbf{t}_2, \{\text{root}(T_n^*), k\} \text{ is removed}].$$

Here the decomposition is due to the edge removal, which is the one between the root of  $T_n^*$  and its parent vertex (some  $k$ ). If this parent vertex is  $k$ , then we know all the vertices  $\{a_1, \dots, a_j\}$  of  $T_n^*$  are chosen from  $\{k+1, \dots, n\}$ . Then  $T_n^*$  should be a concrete recursive tree  $\mathbf{t}_1 \sim \mathbf{t}_1$  defined on  $\{a_1, \dots, a_j\}$  and  $T_n^0$  is a recursive tree  $\mathbf{t}_2 \sim \mathbf{t}_2$  defined on  $\{1, \dots, n\} \setminus \{a_1, \dots, a_j\}$ . Notice that  $T_n$  has  $(n-1)!$  configurations each occurring with equal probability. Moreover there are  $(n-1)$  edges in  $T_n$  each being selected to remove with equal probability. Then we have

$$\begin{aligned} \mathbb{P}[T_n^* \sim \mathbf{t}_1, T_n^0 \sim \mathbf{t}_2] &= \sum_{k=1}^{n-j} \mathbb{P}[T_n^* \sim \mathbf{t}_1, T_n^0 \sim \mathbf{t}_2, \{\text{root}(T_n^*), k\} \text{ is removed}] \\ &= \sum_{k=1}^{n-j} \sum_{k < a_1 < \dots < a_j \leq n} \mathbb{P}[T_n^* = \mathbf{t}_1, T_n^0 = \mathbf{t}_2, \{\text{root}(T_n^*), k\} \text{ is removed}, a_1, \dots, a_j \in \mathbf{t}_1] \\ &= \sum_{k=1}^{n-j} \binom{n-k}{j} \frac{1}{(n-1)!(n-1)}. \end{aligned}$$

Using  $\sum_{k=1}^{n-j} \binom{n-k}{j} = \binom{n}{j+1}$ , we get

$$\mathbb{P}[T_n^* \sim \mathbf{t}_1, T_n^0 \sim \mathbf{t}_2] = \binom{n}{j+1} \frac{1}{(n-1)!(n-1)}. \quad (3.11)$$

This implies that

$$\begin{aligned} \mathbb{P}[|T_n^*| = j] &= \sum_{|\mathbf{t}_1|=j, |\mathbf{t}_2|=n-j} \mathbb{P}[T_n^* \sim \mathbf{t}_1, T_n^0 \sim \mathbf{t}_2] \\ &= (j-1)!(n-j-1)! \binom{n}{j+1} \frac{1}{(n-1)!(n-1)} \\ &= (j-1)!(n-j-1)! \left( \left( \frac{n}{n-1} \frac{1}{j(j+1)} \right) \frac{1}{(j-1)!} \frac{1}{(n-j-1)!} \right) = \frac{n}{n-1} \frac{1}{j(j+1)}, \end{aligned}$$

and hence

$$\mathbb{P}[T_n^* \sim \mathbf{t}_1, T_n^0 \sim \mathbf{t}_2 \mid |T_n^*| = j] = \frac{1}{(j-1)!} \frac{1}{(n-j-1)!}.$$

This yields the desired results.  $\square$

### 3.3 Reduction to the size process

Let us explain more explicitly how the study of GFI process  $(G_t, \Psi_t, \eta_t)_{t \geq 0}$  can be reduced to the study of the size process  $(X_t, Y_t)_{t \geq 0}$ , with the help of the splitting property. We denote by  $\mathcal{M}$  the finite punctual measures on  $\mathbb{N}_+$  and endow it with the weak topology corresponding Borel algebra. Then we treat  $(X_t, Y_t)_{t \geq 0}$  as  $\mathcal{M}^2$ -valued process and denote by  $(\mathcal{F}_t)_{t \geq 0}$  its natural filtration.

**Proposition 2.** *Let  $t \geq 0$ . Conditionally on  $(X_t^u)_{u \in \mathcal{U}_t}$  and  $(Y_t^u)_{u \in \mathcal{U}_t^\dagger}$ , the clusters in  $\mathcal{X}_t \cup \mathcal{Y}_t$  are independent RRT's whose sizes are given by  $(X_t^u)_{u \in \mathcal{U}_t}$  and  $(Y_t^u)_{u \in \mathcal{U}_t^\dagger}$ . Moreover,  $(X_t, Y_t)_{t \geq 0}$  is a branching measure-valued Markov process in  $(\mathcal{M}^2, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ .*

*Proof.* The property on the distribution of clusters is obvious at initial time when there is one single vertex. Let us check that the property remains valid along time and at the same time that the size process satisfies the Markov property. The branching property of the size process is then a direct consequence of the branching process of the initial process.

To prove the remaining part, we consider the three events and corresponding rates. First, a cluster is isolated with a rate depending only on its size, it then simply become inactive. Second, the growth rate of cluster also just depends on its size and the vertex is added independently of the state of the other clusters. Thus, after a growth, the new cluster remains independent from the other ones (conditionally to the sizes). Third, for fragmentation, we invoke the splitting property (Proposition 1), which guarantees that sizes determine the law of the two new clusters and independence with other clusters is preserved. This also ensures the Markov property thanks to absence of memory for each event.  $\square$

The transitions rates of the size process are thus directly inherited from our original process:

- i) becomes an isolated cluster of size  $n$  at rate  $\theta n$ ;
- ii) becomes a RRT of size  $(n+1)$  at rate  $\beta n$ ;
- iii) splits into two RRTs of size  $(n-j, j)$  at rate  $\gamma n \frac{1}{j(j+1)}$ , for  $n \geq 2, 1 \leq j \leq n-1$ .

In particular, each active cluster lives an exponential time of parameter  $(\beta + \theta + \gamma)n - \gamma$ . We introduce now the infinitesimal generator  $\mathcal{A}$  of the Markov process  $(X_t, Y_t)_{t \geq 0}$ . It is defined on a suitable subspace of measurable bounded functions on  $\mathcal{M}^2$ . Consider two functions  $f, g : \mathbb{N} \rightarrow \mathbb{R}_+$  and  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  a bounded Borel function. We set

$$F_{f,g} : (\mu, \nu) \in \mathcal{M}^2 \rightarrow F(\langle \mu, f \rangle, \langle \nu, g \rangle) \in \mathbb{R},$$

and define

$$\begin{aligned}
\mathcal{A}F_{f,g}(\mu, \nu) &= \sum_{n=1}^{\infty} \mu(\{n\}) \beta n (F(\langle \mu + \delta_{n+1} - \delta_n, f \rangle, \langle \nu, g \rangle) - F(\langle \mu, f \rangle, \langle \nu, g \rangle)) \\
&+ \sum_{n=1}^{\infty} \mu(\{n\}) \theta n (F(\langle \mu - \delta_n, f \rangle, \langle \nu + \delta_n, g \rangle) - F(\langle \mu, f \rangle, \langle \nu, g \rangle)) \\
&+ \sum_{n=1}^{\infty} \mu(\{n\}) \gamma(n-1) \times \\
&\quad \sum_{j=1}^{n-1} \left( \frac{n}{n-1} \frac{1}{j(j+1)} \right) (F(\langle \mu + \delta_j + \delta_{n-j} - \delta_n, f \rangle, \langle \nu, g \rangle) - F(\langle \mu, f \rangle, \langle \nu, g \rangle)).
\end{aligned} \tag{3.12}$$

## 4 First moment semigroup and Perron's root

In this part, we study the first moment semigroup associated to the process  $(X_t, Y_t)_{t \geq 0}$ . We will establish the existence of Perron's eigenlements and speed of convergence and prove Theorem 1.

### 4.1 Semigroup and generator

Thanks to Section 3.2, the study of the model is reduced to the long time behavior of the measure-valued branching Markov process  $(X_t, Y_t)_{t \geq 0}$ . We consider now the first moment semigroup  $M = (M_t)_{t \geq 0}$  associated to  $(X_t)_{t \geq 0}$ , which is defined for any non-negative function  $f$  on  $\mathbb{N}_+$  such that for any  $t \geq 0$  and  $n \geq 1$  by

$$M_t f(n) := \mathbb{E}_{\delta_n}[\langle X_t, f \rangle], \tag{4.13}$$

where  $\mathbb{P}_{\delta_n}$  stands for the size process with initial condition  $(X_0, Y_0) = (\delta_n, 0)$  and  $\mathbb{E}_{\delta_n}$  is its associated expectation. In particular we consider for any  $n, m \in \mathbb{N}_+$ ,

$$M_t(n, m) := M_t \mathbf{1}_m(n) = \mathbb{E}_{\delta_n}[\langle X_t, \mathbf{1}_m \rangle] = \mathbb{E}_{\delta_n}[\#\{\mathcal{C} \in \mathcal{X}_t : |\mathcal{C}| = m\}],$$

which is the mean number of clusters of size  $m$  at time  $t$  issued from one single cluster of size  $n$  at time 0.

Let us recall that the functional spaces of polynomial growth  $\mathcal{B}$  and  $\mathcal{B}_p$  and the polynomial function  $[x^p]$  have been introduced at the beginning of Section 3. We extend now the first moment semigroup to these spaces.

**Lemma 1.** (i) For any  $p \geq 1, t \geq 0, n \geq 1$ , we have

$$M_t([x^p])(n) \leq e^{(2^{p-1}p\beta - \theta)t} n^p.$$

(ii) For any  $f \in \mathcal{B}$ , setting  $f_+$  (resp.  $f_-$ ) the positive (resp. negative part) of  $f$ , the functions  $t \in [0, \infty) \rightarrow M_t f_+$  and  $t \in [0, \infty) \rightarrow M_t f_-$  are well defined and finite and we set for any  $t \geq 0$  and  $n \in \mathbb{N}_+$ ,

$$M_t f(n) = \mathbb{E}_{\delta_n}[\langle X_t, f \rangle] := M_t f_+(n) - M_t f_-(n).$$

(iii)  $(M_t)_{t \geq 0}$  is a positive semigroup on  $\mathcal{B}$  and for any  $f \in \mathcal{B}$ , we have

$$\frac{d}{dt} M_t f(n) = M_t(\mathcal{L}f)(n), \quad \forall n \geq 1, \tag{4.14}$$

where the linear operator  $\mathcal{L} : \mathcal{B} \rightarrow \mathcal{B}$  is defined for  $n \geq 1$  by

$$\begin{aligned} \mathcal{L}f(n) = & \underbrace{\beta n(f(n+1) - f(n))}_{\text{I}} \underbrace{- \theta n f(n)}_{\text{II}} \\ & + \underbrace{\gamma(n-1) \sum_{j=1}^{n-1} \frac{n}{n-1} \frac{1}{j(j+1)} (f(j) + f(n-j) - f(n))}_{\text{III}}. \end{aligned} \quad (4.15)$$

The three terms in the linear operator  $\mathcal{L}$  correspond respectively to *the growth*, *the isolation* and *the fragmentation*. This linear operator is the generator of the first moment semigroup of the size process.

*Proof.* We study first  $\mathcal{L}([x^p])$ . Notice that for  $p \geq 1$  and any  $x, y > 0$ ,  $(x+y)^p \geq x^p + y^p$ , so the contribution of the fragmentation term is negative. Thus we have

$$\mathcal{L}([x^p])(n) \leq \beta n((n+1)^p - n^p) - \theta n^{p+1}.$$

We then apply mean-value principle

$$(n+1)^p - n^p = n^p \left( \left(1 + \frac{1}{n}\right)^p - 1 \right) = n^p \times \frac{1}{n} \times p \left(1 + \frac{\xi}{n}\right)^{p-1},$$

where  $\xi \in [0, 1]$ . It gives us

$$\mathcal{L}([x^p]) \leq (2^{p-1}p\beta - \theta[x])[x^p] \leq (2^{p-1}p\beta - \theta)[x^p]. \quad (4.16)$$

Here we use simply  $[x] \geq 1$  for the isolation term.

The rest of the proof follows classical arguments of localization, see e.g. Theorem 1 in [21] and we give only the main lines. We assume that  $X_0 = \delta_n$  for any given  $n \geq 1$ . We consider the stopped process  $(X_t^m, Y_t^m)_{t \geq 0}$  defined by  $X_t^m = X_{t \wedge T_m}$ ,  $Y_t^m = Y_{t \wedge T_m}$ , where

$$T_m = \inf\{t \geq 0 : \langle X_t, [x] \rangle \geq m\}.$$

Note that on the event  $T_m \geq t$ , we have  $\langle X_t, [x] \rangle \leq m$  and  $\langle X_t, [x^p] \rangle \leq m^p$ . The process  $(X_t^m, Y_t^m)_{t \geq 0}$  lives on a finite state space and has bounded rates. Consider three functions  $f, g : \mathbb{N} \rightarrow \mathbb{R}_+$  and  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  a bounded Borel function and recall that  $F_{f,g}(\mu, \nu) = F(\langle \mu, f \rangle, \langle \nu, g \rangle)$ . We get by Kolmogorov forward equation and Dynkin formula (or stopping theorem)

$$\mathbb{E}[F_{f,g}(X_t^m, Y_t^m)] = \mathbb{E}[F_{f,g}(X_0^m, Y_0^m)] + \mathbb{E}\left[\int_0^{t \wedge T_m} \mathcal{A}F_{f,g}(X_s, Y_s) ds\right],$$

where  $\mathcal{A}$  is defined in (3.12). We apply this equation with  $F(x, y) = x \wedge m^p$  and  $f = [x^p]$  to obtain

$$\mathbb{E}[\langle X_t^m, [x^p] \rangle] = \mathbb{E}[\langle X_0^m, [x^p] \rangle] + \mathbb{E}\left[\int_0^{t \wedge T_m} \langle X_s, \mathcal{L}[x^p] \rangle ds\right]. \quad (4.17)$$

Using the above display and (4.16) yields

$$\mathbb{E}[\langle X_t^m, [x^p] \rangle] \leq \mathbb{E}[\langle X_0^m, [x^p] \rangle] + (2^{p-1}p\beta - \theta)\mathbb{E}\left[\int_0^{t \wedge T_m} \langle X_s, [x^p] \rangle ds\right].$$

Since the process  $(X, Y)$  is non explosive,  $T_m$  tends a.s. to infinity as  $m$  tends to infinity. Besides it is increasing in  $m$ . Applying Fatou's lemma on the left hand side and monotone convergence on the right hand side, the above inequality yields

$$M_t[x^p](n) \leq M_0[x^p](n) + (2^{p-1}p\beta - \theta) \int_0^t M_s[x^p](n) ds.$$

Grönwall lemma then ensures *i*) and *ii*) are an immediate consequence.

For *iii*), we need to show that for any  $f \in \mathcal{B}_p$  with  $p \geq 1$ , we have

$$\mathbb{E}[\langle X_t, f \rangle] = \mathbb{E}[\langle X_0, f \rangle] + \mathbb{E} \left[ \int_0^t \langle X_s, \mathcal{L}f \rangle ds \right]. \quad (4.18)$$

Let  $C > 0$  such that  $|\mathcal{L}f(n)| \leq Cn^{p+1}$  for all  $n$ . Similarly to (4.17), we have

$$\mathbb{E}[\langle X_t^m, f \rangle] = \mathbb{E}[\langle X_0^m, f \rangle] + \mathbb{E} \left[ \int_0^{t \wedge T_m} \langle X_s, \mathcal{L}f \rangle ds \right]. \quad (4.19)$$

and we distinguish the events  $t < T_m$  and  $t \geq T_m$ . For the first case, we have

$$\langle X_t^m, f \rangle \mathbf{1}_{\{t < T_m\}} \longrightarrow \langle X_t, f \rangle, \quad \text{almost surely, as } m \rightarrow \infty$$

and  $|\langle X_t^m, f \rangle \mathbf{1}_{\{t < T_m\}}| \leq \langle X_t, |f| \rangle \leq \langle X_t, [x^p] \rangle$ . By *i*), the last term having a finite mean. Using dominated convergence theorem, we obtain

$$\lim_{m \rightarrow \infty} \mathbb{E}[\langle X_t^m, f \rangle \mathbf{1}_{\{t < T_m\}}] = \mathbb{E}[\langle X_t, f \rangle].$$

We prove now that the term corresponding to the second case vanishes. We use a coupling argument and

$$\mathbb{E}[\langle X_t^m, f \rangle \mathbf{1}_{\{t \geq T_m\}}] \leq m^p \mathbb{P}[T_m \leq t] \leq m^p \mathbb{P}[\langle \tilde{X}_t, [x] \rangle \geq m],$$

where  $(\tilde{X}_t)_{t \geq 0}$  is the (increasing) size process with only growth term (i.e.  $\beta > 0, \theta = \gamma = 0$ ) and  $\tilde{X}_0 = \delta_n$ . Then  $(\langle \tilde{X}_t, [x] \rangle)_{t \geq 0}$  is a Yule process with initial value  $n$ . Thus for fixed  $t$ ,  $\langle \tilde{X}_t, [x] \rangle$  follows the negative binomial distribution with parameters  $n, 1 - e^{-\lambda}$ . Using the above display, we get

$$\mathbb{E}[\langle X_t^m, f \rangle \mathbf{1}_{\{t \geq T_m\}}] \leq m^p \frac{\mathbb{E}[(\langle \tilde{X}_t, [x] \rangle)^{2p}]}{m^{2p}} \longrightarrow 0, \quad \text{as } m \rightarrow \infty.$$

Combining the two cases, the term on the left hand side in (4.19) converges to  $\mathbb{E}[\langle X_t, f \rangle]$ .

Now we turn to the right hand side of (4.19). Note that

$$\left| \int_0^{t \wedge T_m} \langle X_s, \mathcal{L}f \rangle ds \right| \leq \int_0^t \langle X_s, |\mathcal{L}f| \rangle ds \leq C \int_0^t \langle X_s, [x^{p+1}] \rangle ds, \quad \forall m \in \mathbb{N}_+.$$

Due to *i*), the last term has finite mean. Moreover  $\int_0^{t \wedge T_m} \langle X_s, \mathcal{L}f \rangle ds \longrightarrow \int_0^t \langle X_s, \mathcal{L}f \rangle ds$ , almost surely as  $m \rightarrow \infty$ . Applying bounded convergence theorem yields

$$\mathbb{E} \left[ \int_0^{t \wedge T_m} \langle X_s, \mathcal{L}f \rangle ds \right] \longrightarrow \mathbb{E} \left[ \int_0^t \langle X_s, \mathcal{L}f \rangle ds \right], \quad \text{as } m \rightarrow \infty.$$

Letting  $m \rightarrow \infty$  in (4.19) gives (4.18) and ends the proof.  $\square$

## 4.2 Perron's root and eigenvectors

Under general assumptions extending the Perron-Frobenius theory in finite dimension, the ergodic behavior of positive semigroups is given by the unique triplet of eigenelements corresponding to the maximal eigenvalue. We refer in particular to [5, 22] and references therein for general statements and applications to growth fragmentation. In this work, we apply a general statement of [5] on the ergodic behavior of positive semigroups. It allows us

to exploit practical sufficient conditions which are satisfied by our process : irreducibility properties of the dynamic of the cluster sizes (4.22) and the fast splitting of large clusters which provide a Lyapunov function for a typical cluster (4.21). Besides, this result ensures exponential speed of convergence of the profile. This will be useful in particular for the proof of the a.s. convergences in the next section.

In what follows, notation “ $f \leq g$ ” means the point-wise comparison for functions. We prove at first Lemma 2, which is the key technical ingredient for Proposition 3. The following space of sublinear functions is useful to control the harmonic function:

$$\begin{aligned} \mathcal{S} := \left\{ f : \mathbb{N}_+ \rightarrow [1, \infty), \text{ such that} \right. \\ \left. \begin{aligned} & a) f \text{ is increasing and } \lim_{n \rightarrow \infty} f(n) = \infty \\ & b) f \text{ is sublinear } \frac{f(n+1)}{n+1} \leq \frac{f(n)}{n}, \text{ and } C_f := \sum_{j=1}^{\infty} \frac{f(j)}{j(j+1)} < \infty \right\}. \end{aligned} \right. \end{aligned} \quad (4.20)$$

**Lemma 2.** *There exists a positive function  $\psi$  and  $b, \xi \in \mathbb{R}$  such that  $0 < \inf_{\mathbb{N}_+} \psi < \sup_{\mathbb{N}_+} \psi \leq 1$  and for every  $V \in \mathcal{S} \cup \{[x^p], p \geq 1\}$*

*i) there exist  $a < b$  and  $\zeta > 0$  such that*

$$\mathcal{L}V \leq aV + \zeta\psi, \quad b\psi \leq \mathcal{L}\psi \leq \xi\psi. \quad (4.21)$$

*ii) for any  $R$  large enough, the set  $K = \{x \in \mathbb{N}_+ : \psi(x) \geq V(x)/R\}$  is a non-empty finite set and for any  $x, y \in K$  and  $t_0 > 0$ ,*

$$M_{t_0}(x, y) > 0. \quad (4.22)$$

*Proof.* To find the Lyapunov-type function  $\psi$ , the main difficulty is to ensure the lower bound of  $\mathcal{L}\psi$  in (4.21) exists. As we can see in eq. (4.15), the isolation term  $-\theta n f(n)$  cannot be bounded from below uniformly in  $n$  by  $f$  times a constant. The strategy is to use the growth term and fragmentation term to compensate the isolation term.

*Step 1: Construction of  $\psi$  - setup.* We set

$$\psi(n) = A - (A - B)q^{n-1},$$

with  $A, B \in (0, \infty)$  and  $q \in (0, 1)$  to be chosen later. Then  $\psi$  is bounded between  $A$  and  $B$  and  $\lim_{n \rightarrow \infty} \psi(n) = A$ . We decompose  $\mathcal{L}\psi$  as follows

$$\mathcal{L}\psi(n) = \underbrace{\beta n(\psi(n+1) - \psi(n))}_{\text{I}} - \underbrace{(\theta + \gamma)n\psi(n)}_{\text{II}} + \underbrace{\gamma\psi(n) + \gamma n \sum_{j=1}^{n-1} \frac{1}{j(j+1)} (\psi(j) + \psi(n-j))}_{\text{III}}. \quad (4.23)$$

First

$$|\text{I}| = |\beta(A - B)n(q^{n-1} - q^n)| \leq C_1\psi(n), \quad (4.24)$$

for some  $C_1 > 0$ , since  $\psi \geq \min\{A, B\} > 0$ . Second, we observe that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\sum_{j=1}^{n-1} \left( \frac{1}{j(j+1)} \psi(j) \right)}{\psi(n)} &= \frac{\sum_{j=1}^{\infty} \frac{1}{j(j+1)} (A - (A - B)q^{j-1})}{A} \\ &= 1 - \left( 1 - \frac{B}{A} \right) q^{-1} \left( 1 + (q^{-1} - 1) \ln(1 - q) \right) =: C_q^{A, B}, \end{aligned}$$

Here we used  $\sum_{j=1}^{\infty} \frac{1}{j(j+1)} = 1$  and  $\sum_{j=1}^{\infty} \frac{q^{j-1}}{j(j+1)} = q^{-1} \left( 1 + (q^{-1} - 1) \ln(1 - q) \right)$ . Moreover,

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^{n-1} \frac{1}{j(j+1)} \psi(n-j)}{\psi(n)} = \frac{\lim_{n \rightarrow \infty} \sum_{j=1}^{n-1} \frac{1}{j(j+1)} (A - (A-B)q^{n-j-1})}{A} = 1,$$

since  $\sum_{j=1}^{n-1} \frac{1}{j(j+1)} q^{n-j-1}$  goes to 0 as  $n \rightarrow \infty$ . Combining the above two displays, we obtain

$$\mathbf{III} \sim (1 + C_q^{A,B}) \gamma n \psi(n), \quad \text{as } n \rightarrow \infty.$$

*Step 2: Construction of  $\psi$  - choice of parameters.* We add and subtract the term  $(1 + C_q^{A,B}) \gamma n \psi(n)$  and reformulate eq. (4.23) as

$$\mathcal{L}\psi(n) = \gamma\psi(n) + \underbrace{(C_q^{A,B} \gamma - \theta) n \psi(n)}_{\mathbf{II}'} + \underbrace{R_1(n, q) + R_2(n, q) + \beta n (\psi(n+1) - \psi(n))}_{\mathbf{III}'}, \quad (4.25)$$

where the term  $\mathbf{III}'$  is the remainder term and

$$\begin{aligned} R_1(n, q) &:= \gamma n \left( \sum_{j=1}^{n-1} \left( \frac{1}{j(j+1)} \psi(j) \right) - C_q^{A,B} \psi(n) \right), \\ R_2(n, q) &:= \gamma n \left( \sum_{j=1}^{n-1} \left( \frac{1}{j(j+1)} \psi(n-j) \right) - \psi(n) \right). \end{aligned} \quad (4.26)$$

We choose  $q, A, B$  such that the term  $\mathbf{II}'$  is 0 (i.e.  $C_q^{A,B} = \theta/\gamma$ ) and  $0 < A, B \leq 1$  (then  $0 < \inf_{\mathbb{N}_+} \psi < \sup_{\mathbb{N}_+} \psi \leq 1$ ). More concretely, we can distinguish three cases:

- If  $\gamma = \theta$ , we can choose  $A = B = 1$ .
- If  $\gamma > \theta$ , we can choose  $q$  close to 1 such that  $q^{-1} \left( 1 + (q^{-1} - 1) \ln(1 - q) \right) \in \left( 1 - \frac{\theta}{\gamma}, 1 \right)$  and then choose  $0 < B < A \leq 1$  such that  $C_q^{A,B} = \theta/\gamma$ .
- If  $\gamma < \theta$ , it suffices to fix some  $q \in (0, 1)$  and then choose  $0 < A < B \leq 1$  such that such that  $C_q^{A,B} = \theta/\gamma$ .

Besides, the convergences in Step 1 ensure that there exists  $C_2 \in (0, \infty)$  such that

$$\sup_{n \in \mathbb{N}_+} \left\{ \left| \frac{R_1(n, q)}{\psi(n)} \right| + \left| \frac{R_2(n, q)}{\psi(n)} \right| \right\} \leq C_2. \quad (4.27)$$

Together with (4.24), we obtain that

$$(\gamma - C_1 - C_2) \psi \leq \mathcal{L}\psi \leq (\gamma + C_1 + C_2) \psi.$$

This guarantees that the two last inequalities of eq. (4.21) hold with the following choices of parameters:

$$b := \gamma - C_1 - C_2, \quad \xi := \gamma + C_1 + C_2.$$

*Step 3: Find  $a, \zeta$ .* For  $V = [x^p]$  with  $p \geq 1$ , we pick a real number  $a$  such that

$$a < \min\{2^{p-1} p \beta - \theta, b\}.$$

Recalling (4.16) and distinguishing if  $2^{p-1}p\beta - \theta n$  is larger than  $a$  or not, we can write

$$\begin{aligned}\mathcal{L}[x^p](n) &\leq an^p + (2^{p-1}p\beta - \theta n - a)n^p \mathbf{1}_{\{2^{p-1}p\beta - \theta n \geq a\}} \\ &\leq a[x^p] + \zeta\psi,\end{aligned}$$

where  $\zeta \in (0, \infty)$  since  $\psi$  is bounded and there exist finitely many  $n$  satisfying  $2^{p-1}p\beta - \theta n \geq a$ . This ends the proof of *i*) for  $p \geq 1$ . Besides, for any large  $R$ , the set  $K$  is finite and non-empty. The combination of growth, fragmentation and isolation ensures the irreducibility of  $(X_t)_{t \geq 0}$  which allows to end the proof of *ii*).

Now we treat the case  $V \in \mathcal{S}$  and verify the condition *i*) and *ii*). Property *ii*) is easy to verify since  $V$  is increasing to infinity while  $\psi$  is a bounded function. For condition *i*), we calculate  $\mathcal{L}V$  and use the decomposition in eq. (4.15). Then  $\frac{V(n+1)}{n+1} \leq \frac{V(n)}{n}$  implies for the growth term that

$$\mathbf{I} = \beta n(V(n+1) - V(n)) \leq \beta n \left( \frac{n+1}{n} V(n) - V(n) \right) \leq \beta V(n).$$

Then the fact that  $V$  increases and  $C_V$  is finite yield for the fragmentation term **III**:

$$\gamma(n-1) \sum_{j=1}^{n-1} \frac{n}{n-1} \frac{1}{j(j+1)} (V(j) + V(n-j) - V(n)) \leq \gamma n \sum_{j=1}^{n-1} \frac{V(j)}{j(j+1)} < C_V \gamma n.$$

The above two displays entail that

$$\mathcal{L}V(n) \leq (\beta - \theta n)V(n) + C_V \gamma n. \quad (4.28)$$

Now we pick a real number  $a$  such that  $a < \min\{\beta - \theta, b\}$ . Using the fact  $\lim_{n \rightarrow \infty} V(n) = \infty$ , we notice that  $E := \{n \in \mathbb{N}_+ : (\beta - \theta n)V(n) + C_V \gamma n > aV(n)\}$  is a non-empty finite set. Then distinguishing the cases when  $n$  belongs to  $E$  or not in eq. (4.28) yields

$$\begin{aligned}\mathcal{L}V(n) &\leq aV(n) \mathbf{1}_{\{n \in E^c\}} + ((\beta - \theta n)V(n) + C_V \gamma n) \mathbf{1}_{\{n \in E\}} \\ &= aV(n) + ((\beta - a - \theta n)V(n) + C_V \gamma n) \mathbf{1}_{\{n \in E\}} \\ &\leq aV(n) + \zeta\psi(n).\end{aligned}$$

Here the constant  $\zeta$  is defined by

$$\zeta := \max_{n \in E} \frac{(\beta - a - \theta n)V(n) + C_V \gamma n}{\psi(n)} \in (0, \infty).$$

□

The following result giving the existence of eigenelements and asymptotic behavior of the semigroup is based on Theorem 2.1 in [5].

**Proposition 3.** *There exists a unique triplet  $(\lambda, \pi, h)$  where  $\lambda \in \mathbb{R}$  and  $\pi = (\pi(n))_{n \in \mathbb{N}_+}$  is a positive vector and  $h : \mathbb{N}_+ \rightarrow (0, \infty)$  is a positive function, such that for all  $t \geq 0$ ,*

$$\pi M_t = e^{\lambda t} \pi, \quad M_t h = e^{\lambda t} h$$

and  $0 < \inf_{n \geq 1} h(n) \leq \sup_{n \geq 1} h(n) < \infty$  and  $\sum_{n \geq 1} \pi(n) = \sum_{n \geq 1} \pi(n) h(n) = 1$ .

Besides, for every  $p > 0$  there exists  $C, \omega > 0$  such that for any  $n, m \geq 1$ ,  $t \geq 0$ ,

$$|e^{-\lambda t} M_t(n, m) - h(n)\pi(m)| \leq C n^p m^{-p} e^{-\omega t}, \quad \sum_{n \geq 1} \pi(n) n^p < \infty. \quad (4.29)$$

*Proof.* We can check directly that the sufficient conditions given in Propositions 2.2 and 2.3 in [5] are satisfied using Lemma 2, together with Lemma 1 iii) which ensures that our pointwise inequalities ensures the weak version of drift conditions required in [5]. More precisely, these conditions are met with  $V = [x^p]$  and  $\varphi = \psi$  for  $p > 0$ , while  $\psi \leq [x^p]$  is guaranteed by the fact that  $\psi \leq 1$ . Using these sufficient conditions, we can apply Theorem 2.1 in [5], which yields the result. In particular, eq. (4.29) is obtained by specifying the initial condition  $\mu = \delta_n$  and using a test function  $1_m$ .

We now prove that  $h$  is upperbounded. It is a consequence of Lemma 3.4 in [5] which ensures that  $h$  is dominated by  $V$  (i.e.  $h \leq V$ ). Adding that Lemma 2 guarantees that we can pick a  $V \in \mathcal{S}$  that increases arbitrarily slowly, we obtain that  $h$  is bounded.

Finally, we justify that  $h$  is lowerbounded. Indeed  $h(n) = e^\lambda M_1 h(n) \geq ch(1)$ , where  $c > 0$  since we recall that the probability that one cluster of size  $n$  brings one cluster of size one before unit time 1 is lowerbounded by a positive constant with respect to its size.  $\square$

Equation (4.29) ensures that for any  $f : \mathbb{N}_+ \mapsto \mathbb{R}$  such that  $\|f\|_p := \sum_{m \geq 1} m^{-(p+2)} |f(m)| < \infty$ ,

$$\left| e^{-\lambda t} M_t f(n) - h(n) \langle \pi, f \rangle \right| \leq C n^{p+2} \|f\|_p e^{-\omega t}. \quad (4.30)$$

Let us also mention that the fact that the eigenfunction  $h$  is (lower and upper bounded) in  $(0, \infty)$  ensures that for any fixed  $t$

$$0 < \inf_{n \geq 1} M_t \mathbf{1}(n) \leq \sup_{n \geq 1} M_t \mathbf{1}(n) < \infty.$$

This implies uniform exponential convergence for bounded tests function:

$$\sup_{n, m \geq 1} \left| e^{-\lambda t} M_t(n, m) - h(n) \pi(m) \right| \leq C' e^{-\omega' t}, \quad t \geq 0. \quad (4.31)$$

Here  $C' > 0, \omega' > 0$ . This latter fact can be obtained by applying Theorem 3.5 in [4] with  $\nu = \delta_n$ . In words, the impact of the initial size of the cluster is bounded, both on the first order approximation and the control of the gap with this approximation.

At this point one may want to apply [1] to prove strong convergence using the asymptotic behavior of the first moment semigroup. But [1] requires stronger assumptions than what is obtained in eq. (4.29) and eq. (4.31), in particular in terms of control of this gap by the stationary distribution  $\pi(m)$  (instead of  $m^{-p}$ ). Besides, we are interested by finer and more quantitative estimates, with motivations in inference and epidemiology. We thus follow another classical approach via  $L^2$  estimates and control of fluctuations.

### 4.3 $L^2$ martingale

Using the first moment semigroup, we can compute the second moment of  $\langle X_t, f \rangle$  for  $f \in \mathcal{B}$ , which consists in the so called formula for forks or many-to-two formula, see e.g. [6, 18] and references therein. The idea is to use the most recent common ancestor of two individuals to decouple their values.

**Lemma 3.** *For any  $x \in \mathbb{N}_+$  and  $f \in \mathcal{B}$ , we have*

$$\mathbb{E}_{\delta_x} [\langle X_t, f \rangle^2] = M_t(f^2)(x) + 2 \int_0^t \sum_{n \geq 1} M_s(x, n) \left( \sum_{1 \leq j \leq n-1} \kappa(n, j) M_{t-s} f(j) M_{t-s} f(n-j) \right) ds,$$

where  $\kappa(n, j) = \frac{\gamma_n}{j(j+1)}$  is the rate at which a cluster of size  $n$  breaks into a cluster of size  $n-j$  (first child) and  $j$  (second child).

*Proof.* We follow [6] and adapt the computations to our case where the branching rate depends on the size of the cluster. Recalling notation of Section 3.1, we have

$$\langle X_t, f \rangle = \sum_{u \in \mathcal{U}_t} f(X_t^u).$$

Recall also that, for any  $u, v \in \mathcal{U}$ ,  $u \wedge v$  is the label of the most recent ancestor of  $u$  and  $v$  and  $u \succcurlyeq v$  means that  $v$  is an ancestor of  $u$ . We first notice that

$$\begin{aligned} \langle X_t, f \rangle^2 &= \sum_{u, v \in \mathcal{U}_t} f(X_t^u) f(X_t^v) = \sum_{u \in \mathcal{U}_t} f^2(X_t^u) + \sum_{w \in \mathcal{U}} \sum_{\substack{u, v \in \mathcal{U}_t, \\ u \neq v, u \wedge v = w}} f(X_t^u) f(X_t^v) \\ &= \sum_{u \in \mathcal{U}_t} f^2(X_t^u) + \sum_{w \in \mathcal{U}} \mathbf{1}_{\{b(w) < t\}} I_t(w), \end{aligned} \quad (4.32)$$

where for any  $w \in \mathcal{U}$ ,  $b(w)$  is the time at which the cluster labeled by  $w$  branches (i.e. the time when it splits into two clusters, labeled  $w1$  and  $w2$ ; potentially infinite if that does not happen) and

$$I_t(w) = \sum_{\substack{u, v \in \mathcal{U}_t \\ i, j \in \{1, 2\}, i \neq j \\ u \succcurlyeq wi, v \succcurlyeq wj}} f(X_t^u) f(X_t^v) = 2 \left( \sum_{u \in \mathcal{U}_t, u \succcurlyeq w1} f(X_t^u) \times \sum_{v \in \mathcal{U}_t, v \succcurlyeq w2} f(X_t^v) \right).$$

Firstly, we have

$$\mathbb{E}_{\delta_x} \left[ \sum_{u \in \mathcal{U}_t} f^2(X_t^u) \right] = M_t(f^2)(x).$$

Secondly, we deal with  $\mathbb{E}_{\delta_x} [\sum_{w \in \mathcal{U}} \mathbf{1}_{\{b(w) < t\}} I_t(w)]$ . For any  $w \in \mathcal{U}$  and for any  $i \in \{1, 2\}$ , we use strong Markov property and have

$$\mathbf{1}_{\{b(w) < t\}} \mathbb{E}_{\delta_x} \left[ \sum_{u \in \mathcal{U}_t, u \succcurlyeq wi} f(X_t^u) \mid b(w), X_{b(w)}^{wi} \right] = \mathbf{1}_{\{b(w) < t\}} M_{t-b(w)} f(X_{b(w)}^{wi}).$$

For any  $w \in \mathcal{U}$ , the branching property then yields

$$\mathbf{1}_{\{b(w) < t\}} \mathbb{E}_{\delta_x} [I_t(w) \mid \mathcal{F}_{b(w)}, b(w)] = 2 \mathbf{1}_{\{b(w) < t\}} M_{t-b(w)} f(X_{b(w)}^{w1}) M_{t-b(w)} f(X_{b(w)}^{w2}).$$

Combining these identities, we obtain

$$\begin{aligned} \mathbb{E}_{\delta_x} \left[ \sum_{w \in \mathcal{U}} \mathbf{1}_{\{b(w) < t\}} I_t(w) \right] &= 2 \mathbb{E}_{\delta_x} \left[ \sum_{w \in \mathcal{U}} \mathbf{1}_{\{b(w) < t\}} M_{t-b(w)} f(X_{b(w)}^{w1}) M_{t-b(w)} f(X_{b(w)}^{w2}) \right] \\ &= 2 \mathbb{E}_{\delta_x} \left[ \sum_{w \in \mathcal{U}} \mathbf{1}_{\{b(w) < t\}} g(X_{b(w)-}^w, b(w)) \right], \end{aligned}$$

where we introduce

$$g(X_{b(w)-}^w, b(w)) := \mathbb{E}_{\delta_x} [M_{t-b(w)} f(X_{b(w)}^{w1}) M_{t-b(w)} f(X_{b(w)}^{w2}) \mid X_{b(w)-}^w, b(w)].$$

This function involves the fragmentation event and can be explicitated recalling that, when a cluster of size  $n$  splits, the probability that the sizes of the two new clusters are  $(j, n-j)$  is  $n/((n-1) \cdot j \cdot (j+1))$ . We obtain

$$g(n, s) = \sum_{1 \leq j \leq n-1} \frac{n}{n-1} \frac{1}{j(j+1)} M_{t-s} f(j) M_{t-s} f(n-j). \quad (4.33)$$

Adding that the branching rate of a cluster of size  $n$  is  $\gamma(n-1)$ , we obtain

$$\begin{aligned}\mathbb{E}_{\delta_x} \left[ \sum_{w \in \mathcal{U}} \mathbf{1}_{\{b(w) < t\}} g(X_{b(w)-}^w, b(w)) \right] &= \int_0^t \sum_{w \in \mathcal{U}, n \geq 1} g(n, s) \mathbb{P}_{\delta_x} [w \in \mathcal{U}_{s-}, X_{b(w)-}^w = n, b(w) \in ds] \\ &= \int_0^t \sum_{w \in \mathcal{U}, n \geq 1} g(n, s) \mathbb{P}_{\delta_x} [w \in \mathcal{U}_{s-}, X_{b(w)-}^w = n] \gamma(n-1) ds \\ &= \int_0^t \sum_{n \geq 1} g(n, s) \gamma(n-1) M_s(x, n) ds.\end{aligned}$$

This equation and eq. (4.33) give us the expression of  $\kappa$ . It ends the proof.  $\square$

With the help of this  $L^2$  expression, we can deal with the martingale associated to the harmonic function  $h$ .

**Proposition 4.** *The process  $(\mathcal{M}_t)_{t \geq 0}$  defined as*

$$\mathcal{M}_t = e^{-\lambda t} \langle X_t, h \rangle, \quad (4.34)$$

*is a non-negative martingale, which converges almost surely to a non-negative finite random variable  $W$  as  $t$  tends to infinity. Moreover, if  $\lambda > 0$ ,  $(\mathcal{M}_t)_{t \geq 0}$  converges in  $L^2$  norm to  $W$ .*

*Proof.* The martingale property is classical and the proof is given for sake of completeness. Recall the notation  $\mathcal{U}_t$  and  $X_t^u$ . For any  $u \in \mathcal{U}_t$ , let  $\mathcal{U}_{t+s}(u)$  be the set of labels of all the clusters active at time  $t+s$  that are descendants of the cluster labeled by  $u$  active at time  $t$ ; let  $X_{t+s}^{uv}$  be the size of the active cluster labeled by  $uv$  at time  $t+s$  which is descendant of the cluster labeled by  $u$  active at time  $t$ . Then we have

$$\begin{aligned}\mathbb{E}[\mathcal{M}_{t+s} | \mathcal{F}_t] &= e^{-\lambda(t+s)} \mathbb{E} \left[ \sum_{u \in \mathcal{U}_{t+s}} h(X_{t+s}^u) \middle| \mathcal{F}_t \right] \\ &= e^{-\lambda(t+s)} \sum_{u \in \mathcal{U}_t} \mathbb{E}_{\delta_{X_t^u}} \left[ \sum_{v \in \mathcal{U}_{t+s}(u)} h(X_{t+s}^{uv}) \middle| \mathcal{F}_t \right] \\ &= e^{-\lambda(t+s)} \sum_{u \in \mathcal{U}_t} M_s h(X_t^u) = \mathcal{M}_t.\end{aligned}$$

since  $M_s h = e^{\lambda s} h$ . As  $\mathcal{M}$  is non-negative, it converges almost surely to a finite random variable.

Let us now prove the  $L^2$  convergence. We apply Lemma 3 with  $x = 1$  and obtain

$$\begin{aligned}\mathbb{E}[\langle X_t, h \rangle^2] &= M_t(h^2)(1) + 2 \int_0^t \sum_{n \geq 1} M_s(1, n) \left( \sum_{1 \leq j \leq n-1} \kappa(n, j) M_{t-s} h(j) M_{t-s} h(n-j) \right) ds \\ &= M_t(h^2)(1) + 2e^{2\lambda t} J_t,\end{aligned}$$

where

$$J_t = \int_0^t \sum_{n \geq 1} e^{-2\lambda s} M_s(1, n) \left( \sum_{1 \leq j \leq n-1} \kappa(n, j) h(j) h(n-j) \right) ds.$$

Using that  $\kappa(n, j) = \gamma n / (j(j+1))$  for all  $n \geq 1$  and  $1 \leq j \leq n-1$  and that  $h$  is bounded from Proposition 3, we get that  $\sum_{1 \leq j \leq n-1} \kappa(n, j) h(j) h(n-j)$  grows at most linearly with  $n$ .

Moreover we can apply eq. (4.30) to control the gap between  $e^{-\lambda s} M_s(1, n)$  and  $h(1)\pi(n)$ . Combining these estimates ensures that for any  $p > 2$ , there exists  $C > 0$  such that

$$0 \leq J_t \leq C \int_0^t e^{-\lambda s} \sum_{n \geq 1} n \gamma(h(1)\pi(n) + n^{-p}) ds, \quad \forall n \geq 1, t \geq 0,$$

which is uniformly upperbounded for all  $n \geq 1, t \geq 0$ . Adding that  $\pi(n)$  decreases to 0 faster than  $n^{-3}$  ensures that  $\sup_{t \geq 0} J_t < \infty$ . Finally

$$\mathbb{E}[(\mathcal{M}_t)^2] = e^{-2\lambda t} \mathbb{E}[\langle X_t, h \rangle^2] = e^{-2\lambda t} M_t(h^2)(1) + J_t,$$

and we apply (4.30) and conclude that  $\sup_{t \geq 0} \mathbb{E}[(\mathcal{M}_t)^2] < \infty$ . Then by the martingale convergence theorem, we obtain that  $\mathbb{E}[W^2] < \infty$  and  $(\mathcal{M}_t)_{t \geq 0}$  converges in  $L^2$  norm to  $W$ .  $\square$

*Remark 1.* Proposition 4 is also valid for  $(X_t)_{t \geq 0}$  under  $\mathbb{P}_{\delta_n}$ . However, to not confuse the notation, we state the result under  $\mathbb{P} = \mathbb{P}_{\delta_1}$  and then  $W$  is consistent with Theorem 2 and Corollary 1.

#### 4.4 Proof of Theorem 1

With the help of Proposition 3 and Proposition 4, we are now ready to prove our Theorem 1.

*Proof of Theorem 1.* We notice that  $\mathbb{E}[|\mathcal{X}_t|] = \mathbb{E}[\langle X_t, 1 \rangle] = \sum_{j=1}^{\infty} M_t(1, j)$  and we apply eq. (4.30) with  $f = 1$  (constant function),  $n = p = 1$ . This ensures that  $\lim_{t \rightarrow \infty} \log(\mathbb{E}[|\mathcal{X}_t|])/t = \lambda$ .

To study the limit of  $\log(\mathbb{E}[|\mathcal{Y}_t|])/t$ , we use Kolmogorov equation. More precisely, following the localization argument of the proof of Lemma 1 (ii – iii), we check that  $F_{f,g}(\mu, \nu) = \langle \nu, 1 \rangle$  belongs to the domain of the extended generator defined in eq. (3.12). We get

$$\mathbb{E}[|\mathcal{Y}_t|] = \mathbb{E}[\langle Y_t, 1 \rangle] = \int_0^t \mathbb{E}[\langle X_s, \theta[x] \rangle] ds = \int_0^t M_s(\theta[x])(1) ds,$$

and we conclude using eq. (4.30).

Lastly, we study the survival probability  $\mathbb{P}[\tau = \infty]$ .

- In the subcritical phase ( $\lambda < 0$ ), eq. (4.30) and the classical first moment estimate prove that extinction is almost sure.
- In the supercritical phase ( $\lambda > 0$ ), we use the  $L^2$  martingale of Proposition 4 and the stopping time theorem to get

$$h(1) = \mathbb{E} \left[ \lim_{t \rightarrow \infty} e^{-\lambda(t \wedge \tau)} \langle X_{t \wedge \tau}, h \rangle \right] = \mathbb{E} [W \mathbf{1}_{\{\tau = \infty\}}].$$

Adding that  $h > 0$  from Proposition 3 implies that  $\mathbb{P}[\tau = \infty] > 0$  and  $\mathbb{P}[W > 0] > 0$ .

- In the critical phase  $\lambda = 0$ , we first observe that the probability of extinction (isolation) of clusters, within a unit time, is greater than a positive constant (uniformly with respect to their size). Besides  $\lim_{t \rightarrow \infty} |\mathcal{X}_t| < \infty$  a.s. since Fatou's lemma ensures that

$$\mathbb{E}[\lim_{t \rightarrow \infty} |\mathcal{X}_t|] \leq \lim_{t \rightarrow \infty} \mathbb{E}[\langle X_t, h \rangle] = h(1) < \infty.$$

This ensures that extinction occurs a.s. in finite time by a classical argument of Markov process with accessible absorbing point. Indeed, for any  $K \geq 1$ , on the event  $\lim_{t \rightarrow \infty} |\mathcal{X}_t| \leq K$ , extinction occurs a.s. since we can construct an infinite sequence of stopping times  $T_n$  (separated by a unit time) such that  $|\mathcal{X}_{T_n}| \leq K$  and for each  $n$ , extinction occurs with a positive (lower bounded) probability during  $[T_n, T_n + 1]$ .

$\square$

## 5 Strong convergences

We recall that the Perron's root  $\lambda \in \mathbb{R}$  and associated eigenelements have been characterized in Proposition 3. The sign of  $\lambda$  determines if the first moment semigroup goes to 0 or infinity. We turn now to trajectorial results and first check that it yields the extinction criterion. Then we focus on the supercritical regime  $\lambda > 0$  and prove strong law of large numbers for the distribution of clusters.

### 5.1 Kesten-Stigum limit theorem

A fundamental and classical question is whether  $\{W > 0\}$  coincide with survival event  $\{\tau = \infty\}$  or not. This is the Kesten-Stigum theorem in branching process, see e.g. [15, 16]. In our case, the  $L^2$  computation ensure that  $\mathbb{P}[W > 0] > 0$  and we get the expected result.

**Proposition 5.** *Assume  $\lambda > 0$ . Then  $\mathbb{P}[W > 0] > 0$  and  $\{W > 0\} = \{\tau = \infty\}$  a.s..*

*Proof.* The fact that  $\mathbb{P}[W > 0] > 0$  comes from the  $L^2$  martingale convergence of Proposition 4, see the proof there. Besides  $\{W > 0\} \subset \{\tau = \infty\}$ . Thus if  $\mathbb{P}[W > 0] = \mathbb{P}[\tau = \infty]$ , the proof is complete. The lines of the proof are classical, even if the fact that sizes of clusters is non bounded requires some specific argument.

First, we use the fact that any cluster can be isolated (before any other event happens to it) during a unit time interval, with a positive probability uniform with respect to its size. As a consequence, the number of clusters has to tend to infinity to survive :

$$\{\tau = \infty\} = \left\{ \lim_{t \rightarrow \infty} |\mathcal{X}_t| = \infty \right\} \quad \text{a.s.}$$

Second, we derive from this result that the number of clusters of size 1 tends to infinity on the survival event. Indeed, during a unit time interval, clusters of size one have a positive probability to stay cluster of size one and other clusters have a positive probability to create (by fragmentation) one cluster of size one, and this probability is lower bounded with respect to the size  $n \geq 2$ . By independence of clusters and Markov inequality, this ensures that

$$\overline{\lim}_{t \rightarrow \infty} X_t(1) = +\infty \quad \text{a.s. on } \{\tau = \infty\}.$$

On this event  $\{\tau = \infty\}$ , we can thus define a sequence of (finite) stopping times for  $N \geq 1$

$$\tau_N := \inf\{t : X_t(1) \geq N\}.$$

We obtain for  $t \geq \tau_N$

$$e^{-\lambda t} \langle X_t, h \rangle \geq e^{-\lambda \tau_N} \sum_{u \in \mathcal{A}_N} e^{-\lambda(t-\tau_N)} \sum_{v \in \mathcal{U}_t, v \geq u} h(X_t^v),$$

where  $\mathcal{A}_N := \{u \in \mathcal{U}_{\tau_N} : X_{\tau_N}^u = 1\}$ . By Proposition 4,  $e^{-\lambda(t-\tau_N)} \sum_{v \in \mathcal{U}_t, v \geq u} h(X_t^v)$  converges to a non-negative random variable denoted by  $W(u)$  which is equal in law to  $W$ . Besides  $\{W(u)\}_{u \in \mathcal{A}_N}$  are i.i.d. random variables. Thus we have

$$\begin{aligned} \mathbb{P}[W = 0, \tau = \infty] &\leq \mathbb{P}[W = 0, \tau_N < \infty] \\ &\leq \mathbb{P}[\{\tau_N < \infty\} \cap \{W(u) = 0, \forall u \in \mathcal{A}_N\}] = (\mathbb{P}[W = 0])^N. \end{aligned}$$

As a conclusion,  $\mathbb{P}[W = 0] = 1$  or  $\mathbb{P}[W = 0, \tau = \infty] = 0$  (by letting  $N \rightarrow \infty$ ), which ends the proof.  $\square$

## 5.2 Strong law of large numbers for the size of active clusters

In this part, we prove Theorem 2 using the estimates of the first moment semigroup, the  $L^2$  estimates and the martingale associated to the harmonic function. The  $L^2$  estimates ensure weak convergence and the speed obtained allows for strong convergence of subsequences. Some additional work is needed to control fluctuations and prove the strong convergence and we follow [2]. We divide the proof into three steps.

*Proof of Theorem 2. Step 1:  $L^2$  convergence.* We prove first the  $L^2$  convergence of  $e^{-\lambda t} \langle X_t, f \rangle$  to  $W \langle \pi, f \rangle$  for  $f \in \mathcal{B}_p$ . We develop the difference as follows

$$\begin{aligned} e^{-\lambda t} \langle X_t, f \rangle - W \langle \pi, f \rangle &= \underbrace{e^{-\lambda t} \langle X_t, f \rangle - e^{-\lambda t} \langle X_t, h \rangle \langle \pi, f \rangle}_{\mathbf{I}} + \underbrace{e^{-\lambda t} \langle X_t, h \rangle \langle \pi, f \rangle - W \langle \pi, f \rangle}_{\mathbf{II}}. \end{aligned} \quad (5.35)$$

The second term  $\mathbf{II}$  is nothing but  $(\mathcal{M}_t - W) \langle \pi, f \rangle$ , which converges in  $L^2$  to 0 by Proposition 4. We only have to prove the  $L^2$  convergence of the term  $\mathbf{I}$  to 0. Denoting  $g := f - \langle \pi, f \rangle h$ , Lemma 3 yields

$$e^{2\lambda t} \mathbb{E}[\mathbf{I}^2] = \mathbb{E}[\langle X_t, g \rangle^2] = M_t(g^2)(1) + J_t, \quad (5.36)$$

where

$$J_t := 2 \int_0^t \sum_{n \geq 1} M_s(1, n) \left( \sum_{1 \leq j \leq n-1} \kappa(n, j) M_{t-s}g(j) M_{t-s}g(n-j) \right) ds.$$

Recalling that  $\|f\|_p := \sum_{m \geq 1} |f(m)| m^{-(p+2)} \in (-\infty, \infty)$ . Observe that  $g \in \mathcal{B}_p$  and let  $p' \geq 2p + 8$ . By (4.30), there exists  $C' > 0$  such that for any  $n \in \mathbb{N}_+$  and  $s, t \geq 0$ ,

$$\begin{aligned} |e^{-\lambda t} M_t g(n) - h(n) \langle \pi, g \rangle| &\leq C' n^{p+2} \|g\|_p e^{-wt}, \\ |e^{-\lambda s} M_s(1, n) - h(1) \pi(n)| &\leq C' n^{-p'} e^{-ws}. \end{aligned}$$

Since  $\langle \pi, g \rangle = 0$  and  $\kappa(n, j) \leq \gamma n$ , using the above two displays, there exists  $C_1 > 0$  such that

$$|J_t| \leq C_1 e^{2(\lambda-\omega)t} \int_0^t e^{(2\omega-\lambda)s} \sum_{n \geq 1} n^{2p+6} (h(1) \pi(n) + n^{-p'}) ds.$$

Using eq. (4.29) (second statement) and  $p' > 2p + 8$ , the sum  $\sum_{n \geq 1} n^{2p+6} (h(1) \pi(n) + n^{-p'})$  in last line is finite and there exists  $C_2 > 0$  such that

$$|J_t| \leq C_2 e^{2(\lambda-\omega)t} \int_0^t e^{(2\omega-\lambda)s} ds, \quad \forall n \geq 1, s \geq 0.$$

Moreover, by Proposition 3,  $e^{-2\lambda t} M_t(g^2)(1) = o(1), t \rightarrow \infty$ . We plug these estimates in (5.36) and there exists  $C_3 > 0$  such that

$$\mathbb{E}[\mathbf{I}^2] \leq C_3 t e^{-(\lambda \wedge 2\omega)t}, \quad \forall t \geq 0. \quad (5.37)$$

Note that  $C_1, C_2, C_3$  may depend on  $f$  but not on  $t$ . Then the proof is finished.

*Remark 2.* A byproduct of eq. (5.37) and Proposition 4 is that, for the case  $\lambda > 0$  there exists a constant  $C_0 > 0$  and an exponent  $\sigma \in (0, \lambda)$ , such that for any  $f \in \mathcal{B}_p$ ,

$$\mathbb{E}[\langle X_t, f \rangle^2] \leq C_0 e^{2\lambda t} (|\langle \pi, f \rangle|^2 + \|f\|_p e^{-\sigma t}). \quad (5.38)$$

*Step 2: Almost sure convergence for one type.* We use an elegant argument from [2] and we extend it to our countable-type branching process.

First, we establish an almost convergence for a discrete scheme, using the speed of convergence obtained from the  $L^2$  estimates. We can pick a step size  $\Delta > 0$  and apply the decomposition eq. (5.35). Then the martingale part **II** converges to 0 almost surely, and for the term **I**, eq. (5.37) yields

$$\mathbb{E} \left[ |e^{-\lambda k \Delta} \langle X_{k\Delta}, f \rangle - e^{-\lambda k \Delta} \langle X_{k\Delta}, h \rangle \langle \pi, f \rangle|^2 \right] \leq C k \Delta e^{-(\lambda \wedge 2\omega)k\Delta}, \quad \forall k \geq 0. \quad (5.39)$$

By Borel-Cantelli lemma, we get

$$e^{-\lambda k \Delta} \langle X_{k\Delta}, f \rangle \xrightarrow{k \rightarrow \infty} W \langle \pi, f \rangle, \quad \text{almost surely.} \quad (5.40)$$

Let us observe that on the event  $\{W = 0\}$ , by Proposition 5, the extinction occurs a.s. in finite time and we focus on the event  $W > 0$ . Let us first prove that

$$e^{-\lambda t} X_t(n) \xrightarrow{t \rightarrow \infty} W \pi(n), \quad \text{almost surely.} \quad (5.41)$$

Given the almost sure convergence in discrete times, we need to control the fluctuations in the intervals  $[k\Delta, (k+1)\Delta)$ . A nice observation in [2] is that we only need to prove the following sufficient (and necessary) condition

$$\varliminf_{t \rightarrow \infty} e^{-\lambda t} X_t(n) \geq W \pi(n), \quad \text{almost surely for all } n \geq 1. \quad (5.42)$$

We first show that (5.42) implies (5.41) using that the martingale convergence controls the dissipation of mass. Indeed, for any  $n \geq 1$ , using a subsequence of times  $(t_k)_{k \in \mathbb{N}_+}$  such that  $\varliminf_{t \rightarrow \infty} e^{-\lambda t} X_t(n) = \lim_{k \rightarrow \infty} e^{-\lambda t_k} X_{t_k}(n)$ . Proposition 4 and Fatou's lemma and (5.42) ensure

$$\begin{aligned} \varliminf_{t \rightarrow \infty} e^{-\lambda t} X_t(n) h(n) &= \lim_{k \rightarrow \infty} \left( \sum_{i \geq 1} e^{-\lambda t_k} X_{t_k}(i) h(i) - \sum_{i \geq 1, i \neq n} e^{-\lambda t_k} X_{t_k}(i) h(i) \right) \\ &\leq W - \sum_{i \geq 1, i \neq n} \varliminf_{k \rightarrow \infty} e^{-\lambda t_k} X_{t_k}(i) h(i) \\ &\leq W - \sum_{i \geq 1, i \neq n} W \pi(i) h(i) = W \pi(n) h(n). \end{aligned} \quad (5.43)$$

So (5.42) implies (5.41).

We need now to prove eq. (5.42) and follow the argument of [2]. Let  $\Delta > 0$  be the time step size. The proof relies on the following lower bound:

$$\forall t \in [k\Delta, (k+1)\Delta), \quad X_t(n) \geq X_{k\Delta}(n) - N_{k,\Delta}(n), \quad (5.44)$$

where  $N_{k,\Delta}(n)$  is the number of active clusters of size  $n$  at time  $k\Delta$  that will encounter at least one event within  $(k\Delta, (k+1)\Delta)$ . Thus,

$$\varliminf_{t \rightarrow \infty} e^{-\lambda t} X_t(n) \geq \varliminf_{k \rightarrow \infty} e^{-\lambda(k+1)\Delta} X_{k\Delta}(n) - \varliminf_{k \rightarrow \infty} e^{-\lambda k \Delta} N_{k,\Delta}(n).$$

Using eq. (5.40) for the first term of the right hand side, we obtain

$$\varliminf_{t \rightarrow \infty} e^{-\lambda t} X_t(n) \geq e^{-\lambda \Delta} \pi(n) W - \varliminf_{k \rightarrow \infty} e^{-\lambda k \Delta} N_{k,\Delta}(n).$$

It remains to prove that  $\lim_{k \rightarrow \infty} e^{-\lambda k \Delta} N_{k,\Delta}(n) = 0$  a.s. and let  $\Delta$  go to 0. We introduce

$$D_k = D_{\Delta,n,k,\varepsilon} := \{N_{k,\Delta}(n) > \varepsilon X_{k\Delta}(n), \quad X_{k\Delta}(n) > k\}, \quad k \geq 1.$$

By branching property, we know that

$$N_{k,\Delta}(n) \stackrel{d}{=} \sum_{i=1}^{X_{k\Delta}(n)} \xi_i,$$

where  $\{\xi_i\}_{i \geq 1}$  are i.i.d. Bernoulli random variables, independent of  $X_{k\Delta}(n)$  and

$$\mathbb{P}[\xi_i = 0] = 1 - \mathbb{P}[\xi_i = 1] = \exp(-r_n \Delta), \quad r_n = (\beta + \theta + \gamma)n - \gamma.$$

Indeed,  $r_n$  is the total jump rate of an active cluster of size  $n$ . Choose  $\Delta$  small such that  $\mathbb{P}[\xi_i = 1] < \varepsilon$ . Then,

$$\sum_{k \geq 1} \mathbb{P}[D_k] \leq \sum_{k \geq 1} \mathbb{P}[N_{k,\Delta}(n) > \varepsilon X_{k\Delta}(n) \mid X_{k\Delta}(n) > k] < \infty,$$

using that  $\mathbb{P}[\sum_{i=1}^k \xi_i > \varepsilon k]$  decreases exponentially as  $k$  grows thanks to Hoeffding inequality. Borel-Cantelli lemma then ensures that a.s.  $D_k$  happens a finite number of times.

Recalling now from (5.40) that  $X_{k\Delta}(n)$  grows exponentially on the event  $\{W > 0\}$ , so  $X_{k\Delta} \leq k$  also happens a.s. a finite number of times. As a consequence, a.s. on the event  $\{W > 0\}$ , we have  $N_{k,\Delta}(n) \leq \varepsilon X_{k\Delta}(n)$  for  $k$  large enough, and we conclude that  $\lim_{k \rightarrow \infty} e^{-\lambda k \Delta} N_{k,\Delta} = 0$  a.s. on the event  $\{W > 0\}$  by letting  $\varepsilon$  go to 0. This ends the proof of (5.42) and we get eq. (5.41), that is Theorem 2 for functions with bounded support.

*Step 3: Almost surely convergence - general test function.* We now extend the space of test functions and get uniform estimates for  $\mathcal{B}_p$ . To this purpose, we define the cutoff operator at some level  $K \in \mathbb{N}_+$

$$f_{\leq K}(n) := f(n) \mathbf{1}_{\{n \leq K\}}, \quad f_{> K}(n) := f(n) \mathbf{1}_{\{n > K\}}. \quad (5.45)$$

First, using eq. (5.41), we obtain

$$\sup_{f \in \mathcal{B}_p} |e^{-\lambda t} \langle X_t, f_{\leq K} \rangle - W \langle \pi, f_{\leq K} \rangle| \leq K^p \sum_{n=1}^K |e^{-\lambda t} X_t(n) - W \pi(n)| \xrightarrow{t \rightarrow \infty} 0, \quad a.s..$$

Second,

$$\sup_{f \in \mathcal{B}_p} |e^{-\lambda t} \langle X_t, f_{> K} \rangle| \leq e^{-\lambda t} \langle X_t, [x^p]_{> K} \rangle, \quad \sup_{f \in \mathcal{B}_p} |W \langle \pi, f_{> K} \rangle| \leq W \langle \pi, [x^p]_{> K} \rangle,$$

Combining these estimates and  $|e^{-\lambda t} \langle X_t, f \rangle - W \langle \pi, f \rangle| \leq |e^{-\lambda t} \langle X_t, f_{> K} \rangle| + |e^{-\lambda t} \langle X_t, f_{\leq K} \rangle - W \langle \pi, f_{\leq K} \rangle| + |W \langle \pi, [x^p]_{> K} \rangle|$  yields

$$\overline{\lim}_{t \rightarrow \infty} \sup_{f \in \mathcal{B}_p} |e^{-\lambda t} \langle X_t, f \rangle - W \langle \pi, f \rangle| \leq \overline{\lim}_{t \rightarrow \infty} e^{-\lambda t} \langle X_t, [x^p]_{> K} \rangle + W \langle \pi, f_{> K} \rangle,$$

for any  $K \geq 1$ . We show now that the right hand side goes to 0 as  $K$  goes to infinity:

$$\lim_{K \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} e^{-\lambda t} \langle X_t, [x^p]_{> K} \rangle = 0. \quad (5.46)$$

Indeed (5.40) ensures that

$$\lim_{k \rightarrow \infty} e^{-\lambda k \Delta} \langle X_{k\Delta}, [x^p]_{> K} \rangle = W \langle \pi, [x^p]_{> K} \rangle$$

and the right hand side goes to 0 as  $K$  goes to infinity.

So we just need to control what happens between the time intervals  $[k\Delta, (k+1)\Delta)$ . For that purpose, we use a coupling argument. On every interval  $[k\Delta, (k+1)\Delta)$ , we consider a size process  $\tilde{X}_t$  starting the coupling at time  $k\Delta$  with the same value  $\tilde{X}_{k\Delta} := X_{k\Delta}$ ; we let the rates of fragmentation and isolation be zero in  $\tilde{X}_t$ , while the growth process in  $\tilde{X}_t$  and  $X_t$  are constructed by the common exponential clocks. Notice the isolation events make negative contribution when testing  $[x^p]$ ,  $p \geq 1$ , so are the fragmentation events because  $(a+b)^p \geq a^p + b^p$  for all  $a, b > 0, p \geq 1$ . Therefore, we obtain

$$\sup_{t \in [k\Delta, (k+1)\Delta)} \langle X_t, [x^p]_{>K} \rangle \leq \sup_{t \in [k\Delta, (k+1)\Delta)} \langle \tilde{X}_t, [x^p]_{>K} \rangle.$$

The right hand side is monotone in  $t$  and we get

$$\sup_{t \in [k\Delta, (k+1)\Delta)} \langle X_t, [x^p]_{>K} \rangle \leq \langle \tilde{X}_{(k+1)\Delta-}, [x^p]_{>K} \rangle.$$

As a consequence, setting

$$B_k = B_{\Delta, n, k}^K := \{ \langle \tilde{X}_{(k+1)\Delta-}, [x^p]_{>K} \rangle > 2 \langle X_{k\Delta}, [x^p]_{>K} \rangle \},$$

it suffices to prove that

$$\mathbb{P}[\{\text{i.o. } B_k\} \cap \{W > 0\}] = 0, \quad (5.47)$$

to get that  $\overline{\lim}_{t \rightarrow \infty} e^{-\lambda t} \langle X_t, [x^p]_{>K} \rangle \leq 2 \lim_{k \rightarrow \infty} e^{-\lambda k\Delta} \langle X_{k\Delta}, [x^p]_{>K} \rangle$  a.s. and conclude by letting  $K \rightarrow 0$ . To this purpose, we use a truncation and define

$$C_k = C_{\Delta, n, k, \varepsilon} := \{e^{-\lambda k\Delta} \langle X_{k\Delta}, [x^p]_{>K} \rangle \geq \varepsilon\} \cap \{e^{-\lambda k\Delta} \langle X_{k\Delta}, [x^{2p}] \rangle \leq 1/\varepsilon\},$$

for  $\varepsilon > 0$ . We split

$$\mathbb{P}[\{\text{i.o. } B_k\} \cap \{W > 0\}] \leq \mathbb{P}[\{\text{i.o. } B_k \cap C_k\}] + \mathbb{P}[\{\text{i.o. } B_k \cap (C_k)^c\} \cap \{W > 0\}].$$

The second term on the right side has the following upper bound thanks to (5.40) and dominated convergence theorem

$$\mathbb{P}\left[W \langle \pi, [x^p]_{>K} \rangle \in (0, 2\varepsilon)\right] + \mathbb{P}\left[W \langle \pi, [x^{2p}] \rangle > 1/(2\varepsilon)\right] \longrightarrow 0, \quad \text{as } \varepsilon \rightarrow 0.$$

Defining  $Z_{k,p,K} := \langle \tilde{X}_{(k+1)\Delta-}, [x^p]_{>K} \rangle - \langle \tilde{X}_{k\Delta}, [x^p]_{>K} \rangle$ , and using Markov inequality,

$$\begin{aligned} \mathbb{P}[B_k \mid \mathcal{F}_{k\Delta}] &= \mathbb{P}\left[\langle \tilde{X}_{(k+1)\Delta-}, [x^p]_{>K} \rangle - \langle \tilde{X}_{k\Delta}, [x^p]_{>K} \rangle > \langle \tilde{X}_{k\Delta}, [x^p]_{>K} \rangle \mid \mathcal{F}_{k\Delta}\right] \\ &\leq \frac{\text{var}[Z_{k,p,K} \mid \mathcal{F}_{k\Delta}]}{(\langle \tilde{X}_{k\Delta}, [x^p]_{>K} \rangle - \mathbb{E}[Z_{k,p,K} \mid \mathcal{F}_{k\Delta}])^2}. \end{aligned} \quad (5.48)$$

We need now to evaluate the conditional expectation and variance carefully. The computations are deferred to appendix. Plugging the estimates (1.58) and (1.59) of appendix into eq. (5.48), we obtain

$$1_{C_k} \mathbb{P}[B_k \mid \mathcal{F}_{k\Delta}] \leq \frac{2\beta\Delta(4^p p + K^{2p+1}) \times \varepsilon^{-1} e^{\lambda k\Delta}}{(\varepsilon e^{\lambda k\Delta} - \beta\Delta(2^p p + K^{p+1}) \times \varepsilon^{-1} e^{\lambda k\Delta})^2}.$$

We pick  $\Delta = \varepsilon^2$  small enough and obtain  $\mathbb{P}[B_k \mid \mathcal{F}_{k\Delta}] \leq C e^{-\lambda k\Delta}$  on  $C_k$ . Adding that  $\mathbb{P}[B_k \mid C_k] = \mathbb{E}[\mathbb{P}[B_k \mid \mathcal{F}_{k\Delta}] \mid C_k]$ , we obtain  $\sum_{k \geq 1} \mathbb{P}[B_k \mid C_k] < \infty$ . By Borel-Cantelli lemma,  $\mathbb{P}[\text{i.o. } B_k \cap C_k] = 0$  for every  $\varepsilon$ . This implies (5.47) and ends the proof.  $\square$

### 5.3 Strong law of large numbers for the size of inactive clusters

In this part, we prove Corollary 1. A heuristic argument to obtain the asymptotic limit is to use the generator eq. (3.12) and the convergence of  $X_t$  ins Theorem 2 :

$$\lim_{s \searrow t} \frac{\mathbb{E}[\langle Y_s, f \rangle - \langle Y_t, f \rangle | \mathcal{F}_t]}{s - t} = \theta \langle X_t, [x]f \rangle \sim_{t \rightarrow \infty} \theta e^{\lambda t} W \langle \pi, [x] \rangle \langle \tilde{\pi}, f \rangle,$$

with  $\tilde{\pi}$  defined in eq. (2.5). In the following paragraphs, we prove the expected result, with a suitable set of tests functions, using in particular martingale analysis.

*Proof of Corollary 1.* We suppose  $f \in \mathcal{B}_p$  for some fixed  $p > 0$  throughout the proof. The proof can be divided into 3 step. In Step 1, we control the value  $\langle Y_t, f \rangle$ . In Step 2 we prove the expected result with a specific function  $f = h/[x]$  which gives us a martingale. In Step 3, we generalize this result to general  $f \in \mathcal{B}_p$ .

*Step 1:  $L^2$  estimate.* We use (5.38) several times and we mention that  $C_0$  will be a constant, independent of  $f$ , which may change from line to line.

We use the extended generator (3.12) and justify that  $F_{g,f}(\mu, \nu) = \langle \nu, f \rangle$  belongs to its domain using the same localization argument as in the proof of Lemma 1. We get

$$\frac{d}{dt} \mathbb{E}[\langle Y_t, f \rangle^2] = 2\theta \mathbb{E}[\langle Y_t, f \rangle \langle X_t, [x]f \rangle] + \theta \mathbb{E}[\langle X_t, [x]f^2 \rangle].$$

Using Young's inequality with  $\alpha > 0$  to be fixed later,  $\theta \mathbb{E}[\langle Y_t, f \rangle \langle X_t, [x]f \rangle] \leq \alpha \mathbb{E}[\langle Y_t, f \rangle^2] + \left(\frac{\theta^2}{\alpha}\right) \mathbb{E}[\langle X_t, [x]f \rangle^2]$ . Then we use Grönwall's inequality to get

$$\mathbb{E}[\langle Y_t, f \rangle^2] \leq \int_0^t e^{\alpha(t-s)} \left( \left( \frac{\theta^2}{\alpha} \right) \mathbb{E}[\langle X_s, [x]f \rangle^2] + \theta \mathbb{E}[\langle X_s, [x]f^2 \rangle] \right) ds.$$

Combining  $L^2$  estimate of  $\langle X_s, [x]f \rangle$  obtained in (5.38) and  $L^1$  estimate of  $\langle X_t, [x]f^2 \rangle$  in (4.30), we obtain

$$\begin{aligned} \mathbb{E}[\langle Y_t, f \rangle^2] &\leq C_0 \int_0^t e^{\alpha(t-s)} \left( \frac{\theta^2}{\alpha} \right) \left( \langle \pi, [x]f \rangle^2 e^{2\lambda s} + \|f\|_p e^{(2\lambda-\sigma)s} \right) ds \\ &\quad + \int_0^t e^{\alpha(t-s)} \theta \left( \langle \pi, [x]f^2 \rangle e^{\lambda s} + C \|f\|_p e^{(\lambda-w)s} \right) ds. \end{aligned}$$

We choose  $\alpha \in (0, \lambda - \max(\sigma/2, w))$  and conclude that there exist  $C''$  such that

$$\mathbb{E}[\langle Y_t, f \rangle^2] \leq C'' \left( \langle \pi, [x]f \rangle^2 e^{2\lambda t} + \langle \pi, [x]f^2 \rangle e^{\lambda t} + \|f\|_p \left( e^{(2\lambda-\sigma)t} + e^{(\lambda-w)t} \right) \right). \quad (5.49)$$

*Step 2: A martingale for  $Y$  which tends to 0.* We use the function

$$F_{h,h/[x]}(\mu, \nu) = \langle \mu, h \rangle - \left( \frac{\lambda}{\theta} \right) \langle \nu, h/[x] \rangle,$$

where  $F(x, y) = x - \left( \frac{\lambda}{\theta} \right) y$ . It belongs to the domain of the extended generator of  $(X_t, Y_t)_{t \geq 0}$ , whose expression can be found in (3.12). It provides a harmonic function :  $\mathcal{A}F_{h,g} = 0$ . We obtain that

$$H_t := \langle X_t, h \rangle - \left( \frac{\lambda}{\theta} \right) \langle Y_t, h/[x] \rangle. \quad (5.50)$$

is a martingale. Let us prove that  $e^{-\lambda}H_t$  converges to 0 as  $t \rightarrow \infty$  and thus Corollary 1 for the specific test function  $h/[x]$ . This vanishing property is due to the fact that two parts in  $H$  compensate. We prove  $L^2$  estimates using again eq. (3.12) (or the quadratic variation) and

$$\begin{aligned} & \frac{d}{dt} \mathbb{E}[|H_t|^2] \\ &= \mathbb{E} \left[ \sum_{n=1}^{\infty} X_t(n) \beta n (|H_t + h(n+1) - h(n)|^2 - |H_t|^2) \right] \\ &+ \mathbb{E} \left[ \sum_{n=1}^{\infty} X_t(n) \theta n \left( \left| H_t - h(n) - \left( \frac{\lambda}{\theta} \right) h(n)/n \right|^2 - |H_t|^2 \right) \right] \\ &+ \mathbb{E} \left[ \sum_{n=1}^{\infty} \left( X_t(n) \gamma(n-1) \sum_{j=1}^{n-1} \left( \frac{n}{n-1} \frac{1}{j(j+1)} \right) (|H_t + h(j) + h(n-j) - h(n)|^2 - |H_t|^2) \right) \right]. \end{aligned}$$

We develop this equation and recognize the generator  $\mathcal{L}$  defined in eq. (4.15). As  $\mathcal{L}h = \lambda h$ , we get

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[|H_t|^2] &= \mathbb{E} \left[ \sum_{n=1}^{\infty} X_t(n) \left( \beta n |h(n+1) - h(n)|^2 + \theta n \left| h(n) + \left( \frac{\lambda}{\theta} \right) h(n)/n \right|^2 \right) \right] \\ &+ \mathbb{E} \left[ \sum_{n=1}^{\infty} X_t(n) \left( \gamma(n-1) \sum_{j=1}^{n-1} \left( \frac{n}{n-1} \frac{1}{j(j+1)} \right) |h(j) + h(n-j) - h(n)|^2 \right) \right]. \end{aligned}$$

Since  $h$  is bounded (see eq. (4.29)), we obtain

$$\mathbb{E}[|H_t|^2] \leq C \int_0^t \mathbb{E}[\langle X_s, [x] \rangle] ds \leq C e^{\lambda t}. \quad (5.51)$$

This implies the  $L^2$  convergence of  $e^{-\lambda t}H_t$  to 0 as  $t \rightarrow \infty$ . For the pathwise convergence, we set step size  $\Delta > 0$ , then for any  $\varepsilon > 0$ , we combine Markov inequality, Doob's inequality for  $H_t$  and the estimate eq. (5.51) that

$$\begin{aligned} \mathbb{P} \left[ \sup_{t \in [k\Delta, (k+1)\Delta)} |e^{-\lambda t} H_t| > \varepsilon \right] &\leq \mathbb{P} \left[ e^{-\lambda k\Delta} \sup_{t \in [k\Delta, (k+1)\Delta)} |H_t| > \varepsilon \right] \\ &\leq \varepsilon^{-2} e^{-2\lambda k\Delta} \mathbb{E} \left[ \left( \sup_{t \in [k\Delta, (k+1)\Delta)} |H_t| \right)^2 \right] \\ &\leq 4\varepsilon^{-2} e^{-2\lambda k\Delta} \mathbb{E}[|H_{(k+1)\Delta}|^2] \\ &\leq C\varepsilon^{-2} e^{-\lambda k\Delta}. \end{aligned}$$

By Borel-Cantelli lemma, we obtain the a.s. convergence of  $e^{-\lambda t}H_t$  to 0.

*Step 3: Convergence of general test function.* Now, we need to obtain the result of a general test function  $f \in \mathcal{B}_p$ . The idea is similar: we define

$$H_t^f := \langle X_t, f \rangle - \left( \frac{\lambda}{\theta} \right) \langle Y_t, f/[x] \rangle = \langle \pi, f \rangle H_t + A_t + B_t, \quad (5.52)$$

where  $A_t = \langle X_t, f - \langle \pi, f \rangle h \rangle$  and  $B_t = \left( \frac{\lambda}{\theta} \right) \langle Y_t, (f - \langle \pi, f \rangle h)/[x] \rangle$ . We use again  $L^2$  estimate and eq. (5.38) and eq. (5.49) to ensure that

$$e^{-2\lambda t} \mathbb{E}[A_t^2 + B_t^2] \leq C_f \left( e^{-\lambda t} + e^{-\sigma t} + e^{-(\lambda+\sigma/2)t} \right).$$

As above, this implies the convergence of  $e^{-\lambda t} H_t^f$  along a subsequence  $\{k\Delta\}_{k \geq 1}$  with  $\Delta > 0$

$$e^{-\lambda k\Delta} H_{k\Delta}^f \xrightarrow{k \rightarrow \infty} 0, \quad \text{in } L^2 \text{ and almost surely.}$$

and

$$\lim_{k \rightarrow \infty} e^{-\lambda k\Delta} \langle Y_{k\Delta}, f/[x] \rangle = \lim_{k \rightarrow \infty} \left( \frac{\theta}{\lambda} \right) e^{-\lambda k\Delta} \langle X_t, f \rangle = \left( \frac{\theta}{\lambda} \right) \langle \pi, f \rangle W, \quad \text{in } L^2 \text{ and almost surely.}$$

Finally, to obtain the convergence along  $t \in \mathbb{R}_+$ , we decompose  $f$  into the difference of two positive functions  $f = f^+ - f^-$  and use that  $Y_t$  is increasing with respect to  $t$

$$\forall t \in [k\Delta, (k+1)\Delta), \quad e^{-\lambda(k+1)\Delta} \langle Y_{k\Delta}, f^+ \rangle \leq e^{-\lambda t} \langle Y_t, f^+ \rangle \leq e^{-\lambda k\Delta} \langle Y_{(k+1)\Delta}, f^+ \rangle,$$

and obtain that

$$\begin{aligned} e^{-\lambda\Delta} \left( \frac{\theta}{\lambda} \right) \langle \pi, [x]f^+ \rangle W &\leq \lim_{k \rightarrow \infty} e^{-\lambda(k+1)\Delta} \langle Y_{k\Delta}, f^+ \rangle \leq \varliminf_{t \rightarrow \infty} e^{-\lambda t} \langle Y_t, f^+ \rangle \\ &\leq \overline{\lim}_{t \rightarrow \infty} e^{-\lambda t} \langle Y_t, f^+ \rangle \leq \lim_{k \rightarrow \infty} e^{-\lambda k\Delta} \langle Y_{(k+1)\Delta}, f^+ \rangle = e^{\lambda\Delta} \left( \frac{\theta}{\lambda} \right) \langle \pi, [x]f^+ \rangle W. \end{aligned}$$

We take  $\Delta \searrow 0$  and obtain that

$$\lim_{t \rightarrow \infty} e^{-\lambda t} \langle Y_t, f^+ \rangle = \left( \frac{\theta}{\lambda} \right) \langle \pi, [x]f^+ \rangle W, \quad a.s.$$

Similar argument also works for  $f^-$ . We combine two terms and use a normalization with the notation  $\tilde{\pi}$  to prove the almost surely convergence in Corollary 1. The  $L^2$  convergence can be done similarly and we skip the details.  $\square$

## 5.4 Limit on recursive tree

In this part, we prove the convergence of the empirical measure on clusters.

*Proof of Theorem 3.* We prove the convergence on active clusters  $\mathcal{X}_t$ , and the proof for the statement on  $\mathcal{Y}_t$  follows the same manner. The main idea is similar to the size process  $(X_t, Y_t)_{t \geq 0}$ , which involves one type convergence and the cut-off argument. Without loss of generality, we suppose that for any  $\mathbf{t} \in \mathcal{T}$ ,  $|f(\mathbf{t})| \leq |\mathbf{t}|^p$  for some  $p > 0$ .

*Step 1: Cut-off argument.* We do the following decomposition on the event  $\mathcal{X}_t \neq \emptyset$

$$\frac{1}{|\mathcal{X}_t|} \sum_{\mathcal{C} \in \mathcal{X}_t} f(\mathcal{C}) - \mathbb{E}[f(T_\pi)] = \sum_{n=1}^{\infty} A_t(n) + \sum_{n=1}^{\infty} B_t(n), \quad (5.53)$$

where we write  $g(n) = \mathbb{E}[f(T_n)]$  and

$$A_t(n) = \frac{X_t(n)}{\langle X_t, 1 \rangle} \left( \frac{1}{X_t(n)} \sum_{\mathcal{C} \in \mathcal{X}_t, |\mathcal{C}|=n} f(\mathcal{C}) - g(n) \right), \quad B_t(n) = \left( \frac{X_t(n)}{\langle X_t, 1 \rangle} - \pi(n) \right) g(n).$$

Theorem 2 implies the a.s. convergence of the second term **II**

$$\lim_{t \rightarrow \infty} \sum_{n=1}^{\infty} B_t(n) = \lim_{t \rightarrow \infty} \left( \frac{\langle X_t, g \rangle}{\langle X_t, 1 \rangle} - \langle \pi, g \rangle \right) = 0, \quad \text{on } \{\tau = \infty\}.$$

For the first term  $\mathbf{I}$ , we use a cut-off and  $|f(\mathcal{C})| \leq |\mathcal{C}|^p$

$$\left| \sum_{n=1}^{\infty} A_t(n) \right| \leq \left| \sum_{n=1}^K A_t(n) \right| + 2 \left| \sum_{n=K+1}^{\infty} \frac{X_t(n)n^p}{\langle X_t, 1 \rangle} \right|. \quad (5.54)$$

For the first part in eq. (5.54), we admit the following equation right now

$$\forall n \in \mathbb{N}_+, \quad \lim_{t \rightarrow \infty} \frac{1}{X_t(n)} \sum_{\mathcal{C} \in \mathcal{X}_t, |\mathcal{C}|=n} f(\mathcal{C}) = \mathbb{E}[f(T_n)], \quad \text{almost surely on } \{\tau = \infty\}, \quad (5.55)$$

which can also be seen as a generalized law of large number and will be proved in Step 2. By Theorem 2, for any  $n \in \mathbb{N}_+$ ,  $\frac{X_t(n)}{\langle X_t, 1 \rangle}$  converges a.s. as  $t \rightarrow \infty$ . Then the above display ensures that  $\lim_{t \rightarrow \infty} \left| \sum_{n=1}^K A_t(n) \right| = 0$ , almost surely on  $\{\tau = \infty\}$ . The second part in eq. (5.54) also converges a.s. by Theorem 2:

$$\lim_{t \rightarrow \infty} \left| \sum_{n=K+1}^{\infty} \frac{X_t(n)n^p}{\langle X_t, 1 \rangle} \right| = \langle \pi, [x^p]_{>K} \rangle, \quad \text{almost surely on } \{\tau = \infty\},$$

where  $[x^p]_{>K}(n) = n^p \mathbf{1}_{\{n > K\}}$ . We put these results back to eq. (5.53) and obtain that

$$\lim_{t \rightarrow \infty} \left| \frac{1}{|\mathcal{X}_t|} \sum_{\mathcal{C} \in \mathcal{X}_t} f(\mathcal{C}) - \mathbb{E}[f(T_\pi)] \right| \leq 2 \langle \pi, [x^p]_{>K} \rangle, \quad \text{almost surely on } \{\tau = \infty\}.$$

Then we let  $K \rightarrow \infty$  and prove Theorem 3.

*Step 2: One type convergence.* It remains to prove eq. (5.55). For the same statement on  $Y_t$ , it is exactly the classical law of large number, as  $(Y_t(n))_{t \geq 1}$  is an increasing process and every time the increment is of size is 1. To prove eq. (5.55) for  $X_t$ , we follow the same spirit in Step 2 of Theorem 2 with some minor technical differences. We recall that  $\mathcal{T}_n$  the space of equivalent class of RRT of size  $n$ , and denote by

$$\forall \mathbf{t} \in \mathcal{T}_n, \quad X_t(\mathbf{t}) := \sum_{\mathcal{C} \in \mathcal{X}_t} \mathbf{1}_{\{\mathcal{C} \sim \mathbf{t}\}},$$

the number of active cluster of type  $\mathbf{t}$ . Because the space  $|\mathcal{T}_n| = (n-1)!$  is finite, it suffices to prove that

$$\forall n \in \mathbb{N}_+, \forall \mathbf{t} \in \mathcal{T}_n, \quad \lim_{t \rightarrow \infty} e^{-\lambda t} X_t(\mathbf{t}) = W \frac{\pi(n)}{(n-1)!}, \quad a.s., \quad (5.56)$$

and this can be reduced once again by the trick from [2] that

$$\forall n \in \mathbb{N}_+, \forall \mathbf{t} \in \mathcal{T}_n, \quad \varliminf_{t \rightarrow \infty} e^{-\lambda t} X_t(\mathbf{t}) \geq W \frac{\pi(n)}{(n-1)!}. \quad (5.57)$$

We explain here how to prove eq. (5.57) with eq. (5.56), which is similar to proving eq. (5.43). The main idea is the fact that  $\sum_{\mathbf{t} \in \mathcal{T}_n} e^{-\lambda t} X_t(\mathbf{t}) = e^{-\lambda t} X_t(n)$  has a proper limit and controls the mass dissipation. Let  $(t_k)_{k \in \mathbb{N}_+}$  be the subsequence such that  $\varliminf_{t \rightarrow \infty} e^{-\lambda t} X_t(\mathbf{t}) = \lim_{k \rightarrow \infty} e^{-\lambda t_k} X_{t_k}(\mathbf{t})$ , then we have

$$\begin{aligned} \varliminf_{t \rightarrow \infty} e^{-\lambda t} X_t(\mathbf{t}) &= \lim_{k \rightarrow \infty} \left( \sum_{\mathbf{t}' \in \mathcal{T}_n} e^{-\lambda t_k} X_{t_k}(\mathbf{t}') - \sum_{\mathbf{t}' \in \mathcal{T}_n, \mathbf{t}' \neq \mathbf{t}} e^{-\lambda t_k} X_{t_k}(\mathbf{t}') \right) \\ &\leq W \pi(n) - \sum_{\mathbf{t}' \in \mathcal{T}_n, \mathbf{t}' \neq \mathbf{t}} \varliminf_{k \rightarrow \infty} e^{-\lambda t_k} X_{t_k}(\mathbf{t}') \\ &\leq W \pi(n) - \sum_{\mathbf{t}' \in \mathcal{T}_n, \mathbf{t}' \neq \mathbf{t}} W \frac{\pi(n)}{(n-1)!} = W \frac{\pi(n)}{(n-1)!}. \end{aligned}$$

Here from the first line to the second line, we use Fatou's lemma, and from the second line to the third line we use eq. (5.57). This equation controls the upper bound of  $\overline{\lim}_{t \rightarrow \infty} e^{-\lambda t} X_t(\mathbf{t})$  and eq. (5.56) is established.

Finally, we prove eq. (5.57), which requires a convergence along discrete subsequence and the control of fluctuation. We calculate the  $L^2$  moment

$$\begin{aligned} \mathbb{E} \left[ \left( e^{-\lambda t} \sum_{\mathcal{C} \in \mathcal{X}_t, |\mathcal{C}|=n} \left( \mathbf{1}_{\{\mathcal{C} \sim \mathbf{t}\}} - \frac{1}{(n-1)!} \right) \right)^2 \right] &= e^{-2\lambda t} \mathbb{E} \left[ \sum_{\mathcal{C} \in \mathcal{X}_t, |\mathcal{C}|=n} \mathbb{E} \left[ \left( \mathbf{1}_{\{\mathcal{C} \sim \mathbf{t}\}} - \frac{1}{(n-1)!} \right)^2 \mid \mathcal{F}_t \right] \right] \\ &\leq e^{-2\lambda t} \mathbb{E} [X_t(n)] = O(e^{-\lambda t}) \xrightarrow{t \rightarrow \infty} 0. \end{aligned}$$

In the first line, we use the i.i.d. RRT Proposition 2. In the second line, we use eq. (5.38). Notice the convergence rate is exponential, thus we take a discrete time  $\{k\Delta\}_{k \geq 1}$  with  $\Delta > 0$  and use Borel-Cantelli lemma to obtain that for any  $\Delta > 0$

$$\forall n \in \mathbb{N}_+, \forall \mathbf{t} \in \mathcal{T}_n, \quad \lim_{k \rightarrow \infty} e^{-\lambda k \Delta} X_{k\Delta}(\mathbf{t}) = W \frac{\pi(n)}{(n-1)!}.$$

Let us now control fluctuations and let  $N_{k,\Delta}(\mathbf{t})$  be the number of active clusters of type  $\mathbf{t}$  at time  $k\Delta$ , on which occurs some growth/isolation/fragmentation event during  $(k\Delta, (k+1)\Delta)$ . Then like eq. (5.44), we have

$$\forall t \in [k\Delta, (k+1)\Delta), \quad X_t(\mathbf{t}) \geq X_{k\Delta}(\mathbf{t}) - N_{k,\Delta}(\mathbf{t}),$$

and it suffices to prove  $\lim_{k \rightarrow \infty} e^{-\lambda k \Delta} N_{k,\Delta}(\mathbf{t}) = 0$  to conclude eq. (5.57). We skip the details as it is exactly the same proof starting from eq. (5.44) in the Step 2 of Theorem 2, which only involves a single type branching.  $\square$

## 6 Further discussions

In the last part, we give some remarks of our results and mention some possible questions in future work.

### 6.1 Existence of phases

We have done the classification of the phases of our GFI process and proved the results on different phases. It is also important to point out that:

**Proposition 6.** *All the three phases exist in our model for some parameters  $(\beta, \theta, \gamma) \in \mathbb{R}_+^3$ .*

We just give a sketch of the proof. The case  $\theta \geq \beta$  is subcritical and we focus on the case  $\theta \in (0, \beta)$  in order to illustrate the diagram in Figure 3. The idea is to use Proposition 3 and find some specific test function  $f$  to show  $\mathcal{L}f < \varepsilon f$  (or  $> \varepsilon f$ ) to prove the subcritical (or supercritical) phase. More precisely, one can use  $[x^p]$  with  $p \in (0, 1)$  in Proposition 3 as a test function for subcritical phase, and  $f(n) = \mathbf{1}_{\{n=1\}} + \kappa \mathbf{1}_{\{n \geq 2\}}$  with a careful choice of  $\kappa$  for supercritical phase. For the existence of critical phase, one need to prove the continuity and monotone property with respect to the parameters, which is quite natural in our model with some perturbation analysis.

## 6.2 Process and its initial condition

We prove our main results with initial condition  $G_0 = \{0\}$  a patient zero, and one may think the RRT structure at the starting point is essential in our model. This is partially true, but RRT structure is a method, and our results Theorem 1, Theorem 2, Corollary 1 and Theorem 3 in fact also apply to a general initial condition that  $G_0$  is a deterministic finite graph. We sketch its main idea of proof: set the vertices at moment 0 negative values for their birth time and we aim to verify the case with  $G_0 = (V_0, E_0)$  as one cluster, since different clusters evolve independently. We denote by  $\mathbb{P}_{G_0}$  this probability space. We recall that  $\mathcal{T}_{V_0}$  the space of recursive trees on concrete vertices  $V_0$  and let  $T_{V_0}$  be a uniform random recursive tree on  $\mathcal{T}_{V_0}$ . Let  $\mathbb{P}_{T_{V_0}}$  be the process with a randomized initial condition  $T_{V_0}$ , then Theorem 1, Theorem 2, Corollary 1 and Theorem 3 are valid under  $\mathbb{P}_{T_{V_0}}$  since the initial condition is RRT and the key property Proposition 2 is established.

Then, we observe that we can couple two probability spaces that

$$\mathbb{P}_{G_0} \stackrel{d}{=} \mathbb{P}_{T_{V_0}}[\cdot \mid T_{V_0} = G_0],$$

and by this coupling  $\mathbb{P}_{G_0}$  is absolutely continuous with respect to  $\mathbb{P}_{T_{V_0}}$ . Therefore, all the results proved in this paper are also valid under  $\mathbb{P}_{G_0}$ .

## 6.3 Generalizations of model

Several features could be added, in particular to describe epidemics and tracing in a more realistic way. In particular, a recovery (death) rate should be added. But the splitting property fails. Our results could still be partially generalized, in particular the parts concerning the mean behavior and some (weaker) description of a.s. long time behavior are expected. Among other extensions, one can think to describe more finely the contact tracing procedure and isolation, or to consider structured population or non-Markovian dynamics.

## A Conditional expectation and variance

We prove here the conditional expectation and variance estimates of

$$Z_{k,p,K} := \langle \tilde{X}_{(k+1)\Delta-}, [x^p]_{>K} \rangle - \langle \tilde{X}_{k\Delta}, [x^p]_{>K} \rangle,$$

which is used in eq. (5.48). Here  $(\tilde{X}_t)_{t \in [k\Delta, (k+1)\Delta]}$ , containing only growth and no fragmentation nor isolation, is a coupling process of  $X_t$ .

**Lemma 4.** *For any  $k, K \in \mathbb{N}_+$  and  $p \geq 1$ , we have*

$$\mathbb{E}[Z_{k,p,K} \mid \mathcal{F}_{k\Delta}] \leq C_\Delta \langle X_{k\Delta}, [x^p] \rangle, \quad (1.58)$$

where  $C_\Delta = e^{2^{p-1}p\beta\Delta} - 1 + (1 - e^{-\beta\Delta K})K^p$  and

$$\text{var}[Z_{k,p,K} \mid \mathcal{F}_{k\Delta}] \leq 2\beta\Delta(4^p p + K^{2p+1}) \langle X_{k\Delta}, [x^{2p}] \rangle. \quad (1.59)$$

*Proof.* For the conditional expectation, we decompose it in the genealogy of cluster over  $\mathcal{U}_t$  that

$$\mathbb{E}[Z_{k,p,K} \mid \mathcal{F}_{k\Delta}] = \sum_{u \in \mathcal{U}_{k\Delta}} \mathbb{E} \left[ [x^p]_{>K}(\tilde{X}_{(k+1)\Delta}^u) - [x^p]_{>K}(\tilde{X}_{k\Delta}^u) \mid \mathcal{F}_{k\Delta} \right]. \quad (1.60)$$

We observe that for any  $0 < a \leq b$ ,

$$[x^p]_{>K}(b) - [x^p]_{>K}(a) = (b^p - a^p) \mathbf{1}_{\{b > K\}} + a^p \mathbf{1}_{\{a \leq K < b\}} \mathbf{1}_{\{a \leq K\}} + a^p \mathbf{1}_{\{a \leq K < b\}} \mathbf{1}_{\{a > K\}}.$$

In the last line, the second term can be bounded by  $K^p \mathbf{1}_{\{a \leq K < b\}}$  and the third term is zero, so we have

$$[x^p]_{>K}(b) - [x^p]_{>K}(a) \leq (b^p - a^p) + K^p \mathbf{1}_{\{a \leq K < b\}}. \quad (1.61)$$

Since there is only growth in the process  $\tilde{X}_t$  on  $[k\Delta, (k+1)\Delta)$ ,  $\tilde{X}_{(k+1)\Delta-}^u \geq \tilde{X}_{k\Delta}^u$  for any  $u \in \mathcal{U}_{k\Delta}$ . So we can apply eq. (1.61) with  $a = \tilde{X}_{k\Delta}^u$  and  $b = \tilde{X}_{(k+1)\Delta-}^u$  to obtain that

$$\begin{aligned} \mathbb{E}[Z_{k,p,K} | \mathcal{F}_{k\Delta}] &\leq \mathbb{E}[\langle \tilde{X}_{(k+1)\Delta-}, [x^p] \rangle - \langle \tilde{X}_{k\Delta}, [x^p] \rangle | \mathcal{F}_{k\Delta}] \\ &\quad + K^p \sum_{u \in \mathcal{U}_{k\Delta}} \mathbb{E} \left[ \mathbf{1}_{\{\tilde{X}_{k\Delta}^u \leq K < \tilde{X}_{(k+1)\Delta-}^u\}} | \mathcal{F}_{k\Delta} \right]. \end{aligned} \quad (1.62)$$

For the first term in eq. (1.62), we follow the proof of Lemma 1 i) to get

$$\mathbb{E}[\langle \tilde{X}_{(k+1)\Delta-}, [x^p] \rangle - \langle \tilde{X}_{k\Delta}, [x^p] \rangle | \mathcal{F}_{k\Delta}] \leq (e^{2^{p-1}p\beta\Delta} - 1) \langle \tilde{X}_{k\Delta}, [x^p] \rangle.$$

For the second term in eq. (1.62), we can control it by the total number of active clusters of size smaller than  $K$  at  $k\Delta$  that grow within  $[k\Delta, (k+1)\Delta)$ :

$$\begin{aligned} &K^p \sum_{u \in \mathcal{U}_{k\Delta}} \mathbb{E} \left[ \mathbf{1}_{\{\tilde{X}_{(k+1)\Delta-}^u > K\}} - \mathbf{1}_{\{\tilde{X}_{k\Delta}^u \leq K\}} | \mathcal{F}_{k\Delta} \right] \\ &\leq K^p \sum_{u \in \mathcal{U}_{k\Delta}} \mathbb{E} \left[ \mathbf{1}_{\{\tilde{X}_{k\Delta}^u \leq K, \text{ the cluster labeled by } u \text{ grows in } \tilde{X}_t \text{ within } [k\Delta, (k+1)\Delta)\}} | \mathcal{F}_{k\Delta} \right] \\ &\leq (1 - e^{-\beta\Delta K}) K^p \langle \tilde{X}_{k\Delta}, 1 \rangle. \end{aligned}$$

Plugging the two inequalities in (1.62) yields (1.58).

For the conditional variance, we have

$$\text{var}[Z_{k,p,K} | \mathcal{F}_{k\Delta}] = \text{var} \left[ \sum_{u \in \mathcal{U}_{k\Delta}} \left( [x^p]_{>K}(\tilde{X}_{(k+1)\Delta-}^u) - [x^p]_{>K}(\tilde{X}_{k\Delta}^u) \right) | \mathcal{F}_{k\Delta} \right].$$

By branching property,

$$\begin{aligned} \text{var}[Z_{k,p,K} | \mathcal{F}_{k\Delta}] &= \sum_{u \in \mathcal{U}_{k\Delta}} \text{var} \left[ \left( [x^p]_{>K}(\tilde{X}_{(k+1)\Delta-}^u) - [x^p]_{>K}(\tilde{X}_{k\Delta}^u) \right) | \mathcal{F}_{k\Delta} \right] \\ &\leq \sum_{u \in \mathcal{U}_{k\Delta}} \mathbb{E} \left[ \left( [x^p]_{>K}(\tilde{X}_{(k+1)\Delta-}^u) - [x^p]_{>K}(\tilde{X}_{k\Delta}^u) \right)^2 | \mathcal{F}_{k\Delta} \right] \\ &\leq \sum_{u \in \mathcal{U}_{k\Delta}} \mathbb{E} \left[ [x^p]_{>K}^2(\tilde{X}_{(k+1)\Delta-}^u) - [x^p]_{>K}^2(\tilde{X}_{k\Delta}^u) | \mathcal{F}_{k\Delta} \right] \\ &= \mathbb{E}[\langle \tilde{X}_{(k+1)\Delta-}, [x^{2p}] \rangle - \langle \tilde{X}_{k\Delta}, [x^{2p}] \rangle | \mathcal{F}_{k\Delta}]. \end{aligned}$$

Here from the second line to the third line we use  $(a-b)^2 \leq a^2 - b^2$  for all  $a > b > 0$ . The rest is the same as in the computation of conditional expectation and we obtain (1.59).  $\square$

**Acknowledgement.** This work was partially funded by the Chair “Modélisation Mathématique et Biodiversité” of VEOLIA-Ecole Polytechnique-MNHN-F.X and ANR ABIM 16-CE40-0001 and ANR NOLO 20-CE40-0015. L.Y. acknowledges the support of the National Natural Science Foundation of China (Youth Programme, Grant: 11801458).

# References

- [1] S. r. Asmussen and H. Hering. Strong limit theorems for general supercritical branching processes with applications to branching diffusions. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 36(3):195–212, 1976.
- [2] K. B. Athreya. Some results on multitype continuous time Markov branching processes. Ann. Math. Statist., 39:347–357, 1968.
- [3] F. Ball and P. Donnelly. Strong approximations for epidemic models. Stochastic Process. Appl., 55(1):1–21, 1995.
- [4] V. Bansaye, B. Cloez, and P. Gabriel. Ergodic behavior of non-conservative semigroups via generalized Doeblin’s conditions. Acta Appl. Math., 166:29–72, 2020.
- [5] V. Bansaye, B. Cloez, P. Gabriel, and A. Marguet. A non-conservative harris’ ergodic theorem. Preprint, arXiv :1903.03946 (2021).
- [6] V. Bansaye, J.-F. Delmas, L. Marsalle, and V. C. Tran. Limit theorems for Markov processes indexed by continuous time Galton-Watson trees. Ann. Appl. Probab., 21(6):2263–2314, 2011.
- [7] M. T. Barlow. A branching process with contact tracing. Preprint 2020, pages available via <https://www.math.ubc.ca/~barlow/preprints/112-bpct5.pdf>.
- [8] E. Baur and J. Bertoin. Cutting edges at random in large recursive trees. In Stochastic analysis and applications 2014, volume 100 of Springer Proc. Math. Stat., pages 51–76. Springer, Cham, 2014.
- [9] J. Bertoin. Fires on trees. Ann. Inst. Henri Poincaré Probab. Stat., 48(4):909–921, 2012.
- [10] J. Bertoin. Markovian growth-fragmentation processes. Bernoulli, 23(2):1082–1101, 2017.
- [11] J. Bertoin and A. R. Watson. A probabilistic approach to spectral analysis of growth-fragmentation equations. J. Funct. Anal., 274(8):2163–2204, 2018.
- [12] J. Engländer, S. C. Harris, and A. E. Kyprianou. Strong law of large numbers for branching diffusions. Ann. Inst. Henri Poincaré Probab. Stat., 46(1):279–298, 2010.
- [13] C. Gu, W. Jiang, T. Zhao, and B. Zheng. Mathematical recommendations to fight against covid-19. Available at SSRN 3551006, 2020.
- [14] E. Horton and A. R. Watson. Strong laws of large numbers for a growth-fragmentation process with bounded cell sizes. arXiv preprint arXiv:2012.03273, 2020.
- [15] H. Kesten and B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. Ann. Math. Statist., 37:1211–1223, 1966.
- [16] T. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes. In Classical and modern branching processes (Minneapolis, MN, 1994), volume 84 of IMA Vol. Math. Appl., pages 181–185. Springer, New York, 1997.
- [17] A. Lambert. A mathematical assessment of the efficiency of quarantining and contact tracing in curbing the covid-19 epidemic. Preprint on MedArxiv (2020).

- [18] A. Marguet. A law of large numbers for branching Markov processes by the ergodicity of ancestral lineages. ESAIM Probab. Stat., 23:638–661, 2019.
- [19] C. Marzouk. Fires on large recursive trees. Stochastic Process. Appl., 126(1):265–289, 2016.
- [20] A. Meir and J. Moon. Cutting down recursive trees. Bellman Prize in Mathematical Biosciences, 21:173–181, 1974.
- [21] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. Adv. in Appl. Probab., 25(3):518–548, 1993.
- [22] S. Mischler and J. Scher. Spectral analysis of semigroups and growth-fragmentation equations. Ann. Inst. H. Poincaré Anal. Non Linéaire, 33(3):849–898, 2016.
- [23] M. Tomasevic, A. Véber, and V. Bansaye. Ergodic behaviour of a multi-type growth-fragmentation process modelling the mycelial network of a filamentous fungus. Available via <https://hal.inria.fr/hal-03087196>, 2020.