A growth-fragmentation-isolation process on random recursive trees and contact tracing

Vincent Bansaye, Chenlin Gu, Linglong Yuan

The Annals of Applied Probability
2023, Vol. 33, No. 6B, 5233–5278
https://doi.org/10.1214/23-AAP1947

A GROWTH-FRAGMENTATION-ISOLATION PROCESS ON RANDOM RECURSIVE TREES AND CONTACT TRACING

BY VINCENT BANSAYE[1,a], CHENLIN GU[2,b] AND LINGLONG YUAN[3,c]

[1]*CMAP, Ecole Polytechnique, IP Paris,* [a]*vincent.bansaye@polytechnique.edu*
[2]*Yau Mathematical Sciences Center, Tsinghua University,* [b]*guchenlin@hotmail.com*
[3]*Department of Mathematical Sciences, University of Liverpool,* [c]*linglong.yuan@liverpool.ac.uk*

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 (or trees) are frozen with a rate proportional to their sizes (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, the number of clusters increases exponentially and we prove that the normalised 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 work is motivated by the control of epidemics and contact tracing where clusters correspond to trees of infected individuals that can be identified and isolated. We complement this work by providing results on the Malthusian exponent to describe the effect of control policies on epidemics.

1. Introduction. The evolution of random trees is motivated by various fields: algorithms, queuing systems, population modelling, etc. The random deletion of edges of a tree has been studied in particular by [11, 31]. Initially, Meir and Moon [31] were interested in the number of steps needed to isolate a distinguished vertex in a random recursive tree, when every deleted edge is chosen uniformly. Bertoin [12] and Marzouk [30] have then studied processes where sets of vertices can be burnt. 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 behaviour of similar dynamics when the random recursive tree grows, following a binary branching process. Our original motivation is to study the effects of control policies of an epidemic. The growth of the infected population is modelled by a Yule process in this work, that is, a binary Markov branching process. The discrete structure of the Yule process is a random recursive tree. The connected new vertices are the new infected individuals. The fragmentation occurs when one edge is removed, interpreted as the loss of infector-infectee information: tracing of this contact becomes impossible. Various reasons may explain the loss or absence of information on contacts, including issues on memory and storage of information or the contact becomes just inaccessible. We restrict ourselves here to a simple model with a single parameter accounting for a rate at which contacts (edges) get lost (removed) independently. Growth and fragmentation will generate connected components, called *clusters*. Each cluster is a set of connected infected individuals that can be isolated all together as soon as one individual in the cluster is detected. An

Received October 2021; revised September 2022.
MSC2020 subject classifications. 60J27, 60J85, 60J80.
Key words and phrases. Branching process, many-to-two formula, Kesten–Stigum theorem, nonconservative semigroup, random recursive tree, strong law of large numbers, epidemic, contact tracing.

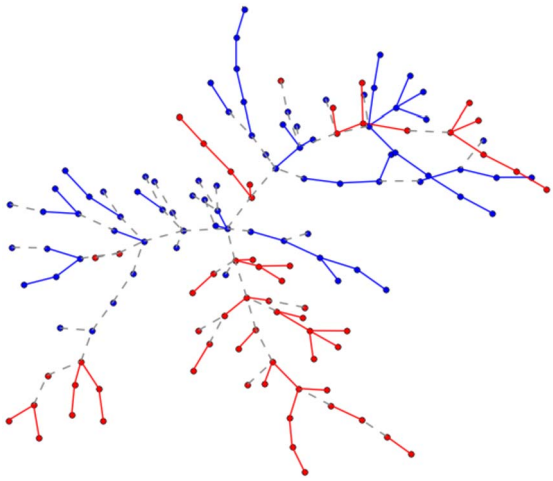


FIG. 1. *An illustration of the growth-fragmentation-isolation process with* 62 *active vertices* (*in red*) *and* 77 *inactive vertices* (*in blue*). *A cluster* (*or a connected component*, *a tree*) *is formed by solid lines*, *while the dashed lines split clusters*.

isolated cluster is frozen, in the sense that no more event will happen to it. It is interpreted as that all the individuals in the cluster are either under treatment or self-isolating. This is the contact tracing strategy which has been for decades the central public health response to control infectious disease outbreaks; see Figure 1 for an illustration.

Numerous research papers highlight the importance of contact-tracing for controlling epidemics, using either simulations or real data, and we mention [17, 19, 25] for Covid-19. For mathematical modelling, we refer to [1, 10, 14, 28] for related studies which are motivated by contact tracing and exploit a branching structure. Let us describe the main differences. In [28], the author provides explosion criterion (threshold for R_0, the reproduction number), for different levels of tracing. The work [1] is more general regarding the characteristics of epidemics but does not allow for backward tracing (i.e., tracing and isolation of ancestors when a descendant is identified and isolated). In [10], a discrete-time model is studied and a Malthusian behaviour is exhibited, with different methods and results. Here we exploit random recursive trees and the growth-fragmentation structure. Our model allows to study the effect of tracing along time, in a Markovian dynamical way. It allows to keep a simple description of the probabilistic structure at any time, to achieve fine asymptotic analysis and to make emerge tractable key quantities for epidemics (by theory or simulations). Moreover, we will exhibit a Malthusian coefficient which characterises the main features of the epidemics, particularly the speed of explosion or extinction (depending on its sign). We will see how it changes with the three parameters of the model, which allows to see the respective effect of social distancing measures (decrease of infection rate β), detection effort (increase of isolation rate θ) and tracing effort (decrease of fragmentation rate γ). The role of β happens to be quite surprising for us and somewhat reinforces the interest of "isolation-tracing" strategy compared to social distancing. Finally, in [14] a model related to ours is considered, where the contact information can be lost when infection occurs (and not later). In this model, fragmentation disappears and more explicit eigenelements and criterion for explosion are given. Combining both effects (instantaneous or continuous loss of contact) would be an interesting direction to explore.

Our growth-fragmentation-isolation model is thus a branching process to study the effect of identification-tracing-isolation strategy in the context of simple epidemics, with loss of contact information along time. The second author also considered a similar model with numerical simulations in [20] to study the propagation of Covid-19. The approximation of the

outbreak by a branching process is classical in the early stages of an epidemic, when the whole population is large and the proportion of susceptible individuals is close to one, see, for instance, [5]. We also obtain in this work estimations on the speed of convergence of infected profiles, which allow to control in which time window the Malthusian exponential growth indeed captures the dynamics of the epidemics. There are more epidemiological and control features that could be incorporated to reflect epidemics like Covid-19. In particular, no recovery happens in our setting, and we consider only large homogeneously mixed populations. We expect some extensions of our results on the long-time behaviour of infected population which would take into account these features, even if the probabilistic structure of clusters becomes more complex. We give a more detailed discussion in Section 8.3.

Random recursive trees (RRT) have a nice splitting property that allows us to characterise a cluster by its size. More precisely, considering the collection of active (nonisolated) clusters, the size process turns out to be a branching process with a countable set of types (i.e., type being the size). At fixed time, conditionally on the sizes of the clusters, the collection of clusters are independent RRTs. We then study the ergodic properties of the first moment semigroup of this branching process and obtain a phase transition depending on the sign of the maximal eigenvalue (Malthusian growth rate). We describe the a.s. behaviour of the process when the active clusters survive, proving a strong law of large numbers for the empirical measure of sizes and a Kesten–Stigum type result for the growth of the population. With the knowledge on the size process, we can deduce the a.s. behaviour of the process of active clusters conditioned on survival. We can also characterise the a.s. behaviour of the process of isolated clusters when the active clusters survive, which is by itself non-Markovian.

The asymptotic analysis of the mean behaviour, the weak convergence and the estimation of the speed of convergence uses now well developed techniques for branching processes with possibly infinitely many types. We refer to [7, 13, 15, 23, 29, 34] and references therein for related works on the asymptotic analysis of growth-fragmentation processes. Roughly, it relies on the ergodic properties of the size of a typical cluster, and the fact that the common ancestor of two clusters uniformly chosen at large times is found at small times, see, for instance, Theorem 2 in [4], see also [22]. More precisely, we exploit the fact that large clusters fragment fast, and with high probability give one small cluster and one large cluster at fragmentation. Together with isolation, it allows one to control the size of a typical cluster and the eigenelements of the first moment semigroup. 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 at the same scale as fragmentation. Once active clusters are well described, we can treat the isolated clusters using an additive functional.

Besides, the fact that the number of types is infinite and the loss of Markov property for the isolated clusters raise some technical difficulties to get a.s. limits (strong convergence). We refer to [2, 3, 18] for classical references on strong law of large numbers of some classes of multitype branching processes. We adapt here the argument of [3] for strong convergence. In the growth-fragmentation context, let us mention respectively [16] and [21, 23] for L^1 and strong convergence.

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

- We identify the vertex set V_t as the set of patients (individuals infected up to time t), and label them with the infection time $v \in [0, \infty)$. The function Ψ_t tells us the state of a vertex at time t, 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 (direct) infection links between patients, and the function η_t tells us the state of an edge at time t. For an edge e, we say e is open if $\eta_t(e) = 1$

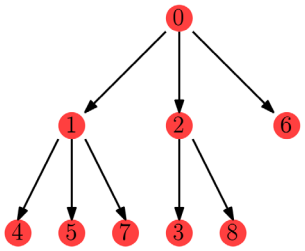
which means the infection link can still be retrieved (i.e., it can be found who infected the infectee), otherwise $\eta_t(e) = 0$ and it is *closed* (i.e., it is not possible to know who infected the infectee). A set of vertices connected by open edges is called a *cluster*.

The growth-fragmentation-isolation model (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 $(\beta, \theta, \gamma) \in \mathbb{R}_+^3 = (0, \infty)^3$ representing three types of events, with the notation $\mathbb{P}$ for the probability and $\mathbb{E}$ for the associated expectation.

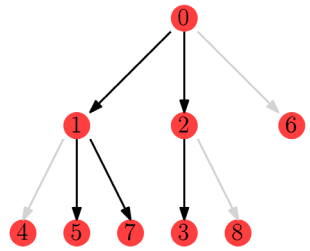
- *Infection* (growth): every active vertex v independently attaches a new vertex in an exponential waiting time with parameter β. When a new vertex u is created and attached to v, it is active (i.e., $\Psi_t(u) = 1$) and the edge $\{u, v\}$ is open (i.e., $\eta_t(\{u, v\}) = 1$).
- *Information decay* (fragmentation): every open edge e independently becomes closed in an exponential waiting time with parameter γ.
- *Confirmation and contact tracing* (isolation): every active vertex independently gets "confirmed" in an exponential waiting time with parameter θ, and once a vertex is confirmed, the associated cluster is isolated and every vertex in 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. As vertices are indexed by infection times, every cluster is a labelled *recursive tree* (see Section 3.2 for a rigorous definition).

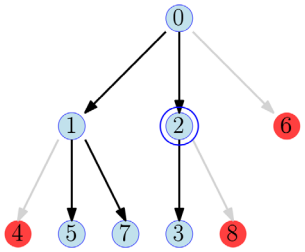
Because the vertices in a cluster have the same state, it is very natural to decompose the dynamical tree G_t into clusters of individuals connected by open edges: for an isolated cluster,



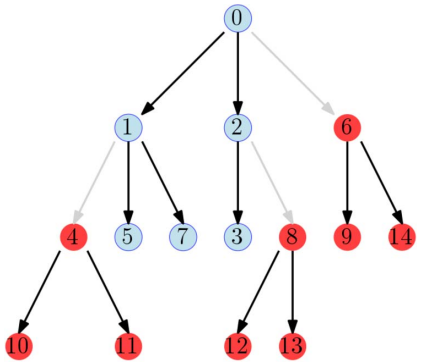
(a) Growth: starting from vertex 0, the vertices are attached one by one, and form a recursive tree.



(b) Fragmentation: links get lost over time; for example the grey links $\{0,6\}, \{1,4\}, \{2,8\}$ in the tree are lost.



(c) Isolation: when the vertex 2 is confirmed, all the vertices in the same cluster are isolated. These are the vertices in blue $\{0, 1, 2, 3, 5, 7\}$.



(d) The isolated vertices are no longer active, while the active vertices continue to attach new vertices.

FIG. 2. *An illustration of GFI process.*

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 $(\mathscr{X}_t, \mathscr{Y}_t)_{t\geq 0}$ the associated *cluster process*, where $\mathscr{X}_t$ is the set of active clusters and $\mathscr{Y}_t$ is the set of inactive clusters:

$$\mathscr{X}_t = \{\mathcal{C} | \mathcal{C} \text{ is a cluster in } G_t; \forall v \in \mathcal{C}, \Psi_t(v) = 1\},$$
$$\mathscr{Y}_t = \{\mathcal{C} | \mathcal{C} \text{ is a cluster in } G_t; \forall v \in \mathcal{C}, \Psi_t(v) = 0\}.$$

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

$$\tau := \inf\{t \geq 0 | \mathscr{X}_t = \varnothing\}, \tag{1}$$

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 what values of the parameters (β, θ, γ) does the epidemic reach extinction almost surely? If the epidemic survives (with positive probability), what is the long-time behaviour of the population of active and inactive clusters? We give some answers to these questions by first determining the asymptotic behaviour of the first moment semigroup associated to the active clusters. It depends on the maximal eigenvalue of this semigroup, which is called Malthusian exponent. When this exponent is negative (subcritical case) or zero (critical case), the population of active clusters reaches extinction almost surely. It corresponds to the fact that the isolation process is strong enough to stop the epidemic. When this exponent is positive (supercritical case), survival occurs with positive probability and on this event, the growth of the population is exponential with rate given by this Malthusian exponent. In that case, we also shed some light on the genealogical structure of clusters and describe the asymptotic behaviour 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, that is, the Malthusian exponent. We also show similar asymptotic behaviours for the inactive clusters on the survival event.

We give the main results and the outline of the paper in the next section. The rest of the paper is dedicated to proofs and some additional results and comments.

2. Main results. We first introduce the *Malthusian exponent* λ which describes the (mean) exponential growth (or decay) of the number of clusters. This growth rate coincides for the active and inactive clusters, whose numbers at time t are respectively denoted by $|\mathscr{X}_t|$ and $|\mathscr{Y}_t|$. As expected in branching structures, its sign gives the global extinction and survival criterion, leading to the classification of subcritical, critical and supercritical phases.

THEOREM 2.1 (Malthusian exponent). *The following limits exist and coincide and are finite*:

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

If $\lambda \leq 0$, *then extinction occurs a.s., that is*, $\mathbb{P}[\tau < \infty] = 1$. *Otherwise, survival occurs with positive probability, that is*, $\mathbb{P}[\tau = \infty] > 0$.

The Malthusian exponent λ corresponds to the maximal eigenvalue of the first moment semigroup of $(X_t)_{t\geq 0}$ and is also called *Perron's root*. The fact that the cluster size can be any positive integer leads us to use techniques for ergodic behaviour in the infinite-dimension setting, where the control of large sizes is crucial. As usual, irreducibility on the state of sizes ensures that the value λ does not depend on the initial condition (although the model

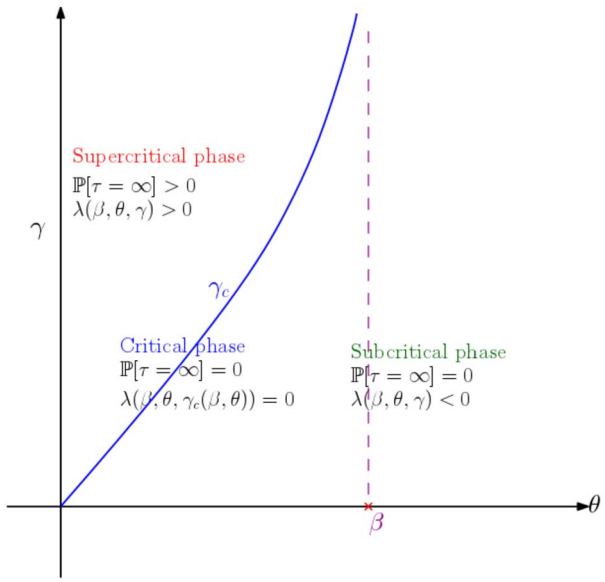


FIG. 3. *An illustration of different phases*; *see Section 6.2 for more details*.

starts with a single patient, that is, a cluster of size 1, the setting of initial condition with a single random recursive tree of size $n \geq 1$ will be used later, especially in the size processes). 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 isolated faster than they contaminate and in the case $\theta \geq \gamma$, isolation is faster than fragmentation.

We want to emphasise that Theorem 2.1 and the forthcoming results do not depend on the initial condition (random or deterministic), see the discussion in Section 8.2. We also refer to Section 8.3 for other possible generalisations with additional epidemic features.

To go further in the analysis of the model, we introduce the following *size process* $(X_t, Y_t)_{t\geq 0}$, where the two empirical measures count the clusters of different sizes:

$$X_t = \sum_{\mathcal{C}\in\mathscr{X}_t} \delta_{|\mathcal{C}|}, \qquad Y_t = \sum_{\mathcal{C}\in\mathscr{Y}_t} \delta_{|\mathcal{C}|}. \tag{2}$$

Here $|\mathcal{C}|$ is the number of vertices in the cluster $\mathcal{C}$, and we call it *the size of cluster*. The process $(X_t)_{t\geq 0}$ is still a branching Markov process with respect to its natural filtration. This comes from the splitting property of RRT, which allows to preserve RRT clusters when fragmentation occurs, see the forthcoming Proposition 3.1. Moreover, for any fixed time, conditionally on cluster sizes, all (active and inactive) clusters are independent, see Proposition 3.2. We can thus reduce the study of our GFI process $(G_t, \Psi_t, \eta_t)_{t\geq 0}$ to that of the size process $(X_t, Y_t)_{t\geq 0}$; see Figure 4.

We prove the following strong law of large numbers in the supercritical case. This provides the asymptotic behaviour of $\langle X_t, f\rangle = \sum_{\mathcal{C}\in\mathscr{X}_t} f(|\mathcal{C}|)$, where f has at most polynomial growth, that is, there exists $p > 0$ such that $\sup_{n\geq 1} |f(n)|/n^p < \infty$. In particular, $f = \mathbf{1}_m$ yields the number of active clusters of size m, while the identity function provides the number of active individuals.

THEOREM 2.2 (Law of large numbers for $(X_t)_{t\geq 0}$). *Assume that $\lambda > 0$. Then there exists a probability distribution π on $\mathbb{N}_+ = \{1, 2, 3, \ldots\}$ and a random variable $W \geq 0$, such that for any function $f : \mathbb{N}_+ \to \mathbb{R}$ of at most polynomial growth, we have $\langle \pi, |f|\rangle < \infty$ and*

$$e^{-\lambda t}\langle X_t, f\rangle \xrightarrow{t\to\infty} W\langle \pi, f\rangle \quad \text{a.s. and in } L^2. \tag{3}$$

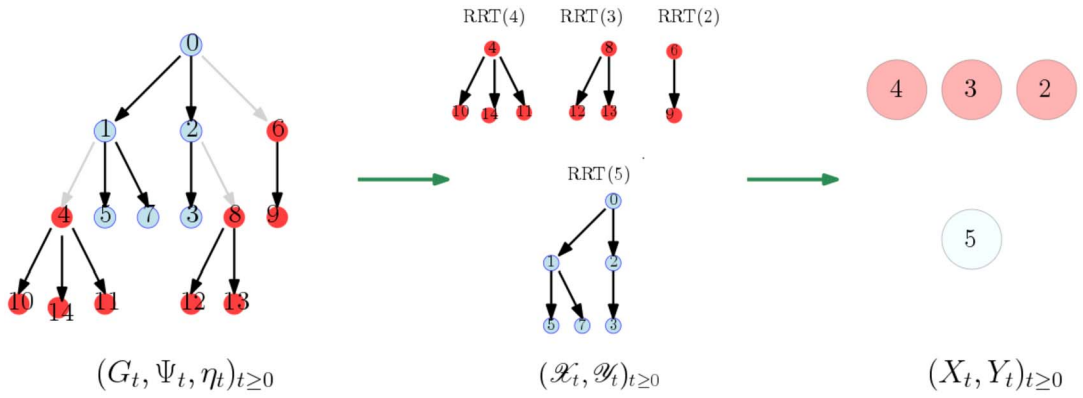


FIG. 4. *Illustration of the reduction of the study of GFI process* $(G_t, \Psi_t, \eta)_{t\geq 0}$. *The first simplification is to decompose the graph into clusters and study the cluster process* $(\mathscr{X}_t, \mathscr{Y}_t)_{t\geq 0}$, *where every cluster is a RRT whose law only depends on its size. The second simplification is to study size process* $(X_t, Y_t)_{t\geq 0}$, *which provides the key information for the law of* $(\mathscr{X}_t, \mathscr{Y}_t)_{t\geq 0}$.

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

$$\frac{\langle X_t, f\rangle}{\langle X_t, \mathbf{1}\rangle} \xrightarrow{t\to\infty} \langle \pi, f\rangle \quad a.s. \tag{4}$$

Recall that λ is defined in Theorem 2.1. It gives the a.s. exponential growth of the number of active clusters and active individuals. The probability distribution π gives the distribution of the size of clusters for large times. This distribution is the (positive normalised) left eigenvector of the first moment semigroup $M = (M_t)_{t\geq 0}$ associated to the size process $(X_t)_{t\geq 0}$. Equivalently, it can be characterised as the positive normalised left eigenvector of the generator $\mathcal{L}$ of M, that is, $\pi\mathcal{L} = \lambda\pi$, where

$$\mathcal{L}f(n) = \beta n\big(f(n+1) - f(n)\big) - \theta n f(n) + \sum_{j=1}^{n-1} \frac{\gamma n}{j(j+1)}\big(f(j) + f(n-j) - f(n)\big),$$

for real valued functions f and $n \geq 0$. Thus $(\pi(n))_{n\geq 0}$ satisfies a linear system given by the dual operator of $\mathcal{L}$. We refer to Section 4.1 for rigorous statements and details on the semigroup and generator. The fact that $\langle \pi, |f|\rangle < \infty$ will also be given in Proposition 4.1.

The random variable W in the statement is the limit of the Malthusian martingale $e^{-\lambda t}\langle X_t, h\rangle$, where h is the right eigenvector of the semigroup: $M_t h = e^{\lambda t}h$, for any $t \geq 0$; again see Proposition 4.1. Because the cluster size can be arbitrarily large, the classical finite-dimensional Perron–Frobenius theorem does not apply and we need a more precise analysis on the semigroup and its generator to ensure the properties of eigenvectors; see Section 4.

Note that the above results hold for functions of at most polynomial growth. This is inherited from our Lyapunov functions which are polynomial; see Lemma 4.2 and Proposition 4.1.

A similar result about the inactive clusters can be derived.

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

$$e^{-\lambda t}\langle Y_t, f\rangle \xrightarrow{t\to\infty} \left(\frac{\theta}{\lambda}\right) W \sum_{n=1}^{\infty} n\pi(n) f(n) \quad a.s. \text{ and in } L^2,$$

and

$$\frac{\langle Y_t, f\rangle}{\langle Y_t, \mathbf{1}\rangle} \xrightarrow{t\to\infty} \langle \widetilde{\pi}, f\rangle \quad a.s. \text{ on } \{\tau = \infty\},$$

where $\widetilde{\pi}$ is a probability law on $\mathbb{N}_+ = \{1, 2, 3, \ldots\}$ given by

$$\widetilde{\pi}(n) := \frac{\pi(n)n}{\sum_{j=1}^{\infty} \pi(j)j}. \tag{5}$$

We observe that interestingly, $\widetilde{\pi}$ is a size-biased version of π. It means that the isolated clusters are of larger size than the active clusters. This phenomenon is also recorded in the recent work of Jean Bertoin [14], Corollary 4.3. The reason for this phenomenon is that every active cluster gets isolated at a rate proportional to its size.

Recall that every vertex $v \in \mathbb{R}$ is labelled by its infection time and every cluster is a tree. We consider these clusters up to an equivalence relation, which consists in keeping the order between vertices but forgetting their infection times, see Section 3.2 for a rigorous definition. We denote by $\mathscr{T}$ the space of equivalence classes of trees of all sizes. We denote by T_π the random tree whose size is distributed as π, and conditionally on its size the tree is a RRT. With a slight abuse of notation, for any recursive tree t, we write $\mathbf{t} \in \mathscr{T}$ for its equivalence class and for any function $f : \mathscr{T} \mapsto \mathbb{R}$, we let $f(\mathrm{t}) = f(\mathbf{t})$. We also write $|\mathbf{t}|$ for the number of vertices in the equivalence class $\mathbf{t}$.

Using the splitting property, we derive the long-time behaviour of the empirical measures on (active and inactive) clusters from the results on the size processes obtained above.

THEOREM 2.3 (Limit of the empirical measure of clusters). *Consider any $f : \mathscr{T} \to \mathbb{R}$ such that there exists $p > 0$ satisfying*

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

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

$$\frac{1}{|\mathscr{X}_t|} \sum_{\mathcal{C} \in \mathscr{X}_t} f(\mathcal{C}) \overset{t\to\infty}{\longrightarrow} \mathbb{E}\big[f(T_\pi)\big], \qquad \frac{1}{|\mathscr{Y}_t|} \sum_{\mathcal{C} \in \mathscr{Y}_t} f(\mathcal{C}) \overset{t\to\infty}{\longrightarrow} \mathbb{E}\big[f(T_{\widetilde{\pi}})\big] \quad a.s.$$

The Malthusian exponent λ is the key characteristic of the model and is fully determined by the model parameters (β, γ, θ). For both mathematical and practical purposes, it is highly important to understand how λ depends on (β, θ, γ) The main results are presented below.

THEOREM 2.4. *The mapping $(\beta, \theta, \gamma) \mapsto \lambda(\beta, \theta, \gamma)$ is continuous, and the sets of parameters resulting in respectively $\lambda > 0, \lambda = 0, \lambda < 0$ are nonempty. The Malthusian exponent depends on the three parameters monotonically that*:

1. *$\theta \mapsto \lambda(\beta, \theta, \gamma)$ is decreasing*;
2. *$\gamma \mapsto \lambda(\beta, \theta, \gamma)$ is increasing*;
3. *$\beta \mapsto \lambda(\beta, \theta, \gamma)$ is increasing if $\gamma > \theta$, constant if $\gamma = \theta$, and decreasing if $\gamma < \theta$.*

In the above statements, the continuity, the existence of phases, and also the monotonicity with respect to θ and γ, are not surprising. However, the change of monotonicity of $\beta \mapsto \lambda(\beta, \theta, \gamma)$ seems not obvious. Indeed, the increment of β makes the clusters grow faster, thus the clusters will split (fragmentation) or be detected (isolation) faster. Faster fragmentation makes the number of clusters increase while faster detection has the opposite effect. The complex competition between the two forces leads to the simple Statement 3 in Theorem 2.4 surprisingly. At the end, the increment of the infection rate β will not necessarily speed up the epidemics, but does make the epidemics weaker when $\gamma < \theta$. In practical terms, in this regime, if the contact tracing policy is effective so that the population is in the subcritical

regime, the emergence of a more contagious variant will make the epidemic decay faster, even if θ and γ are unchanged (it means we use the same contact tracing practice as before).

The rest of the paper is organised as follows. In Section 3 we introduce notation and we recall the key splitting property of the random recursive trees and derive the size process. The study of the first moment semigroup of $(X_t)_{t\geq 0}$ and the associated martingale and L^2 estimates is carried out in Section 4, using in particular Lyapunov functions. We can then prove Theorem 2.1. Section 5 is devoted to the strong convergences and we prove Theorem 2.2, Corollary 1 and Theorem 2.3. Then we prove Theorem 2.4 in Section 6 and present related simulations in Section 7. Finally we discuss possible extensions and generalisations of the model in Section 8.

3. Preliminaries. In this section, we introduce notation and explain why the size process captures the essential information of the cluster process.

3.1. *Ulam–Harris–Neveu labeling of clusters.* We introduce the notation for the genealogical tree of clusters (active and inactive). Recall that every vertex is labelled by its 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 root($\mathcal{C}$). We label every cluster using the Ulam–Harris–Neveu (UHN) notation:

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

where an element $u \in \mathcal{U}$ is called a *label* or *word*. For the initial cluster, we use the label $\varnothing$ as convention. Then by induction, for any cluster $\mathcal{C}$ labelled by a word $u \in \mathcal{U}$: this label is unchanged during the growth of the cluster (infection); this label becomes inactive when the cluster is isolated; this label is replaced by two labels $u1$ and $u2$ when fragmentation occurs. By convention, $u1$ is the label of the subcluster containing root($\mathcal{C}$) and is called *the first child*, while $u2$ is for the other subcluster that we call *the second child*. (See Figure 6 for an illustration.)

Every cluster except the initial one has a unique parent. There exists a partial order $\preccurlyeq$ on $\mathcal{U}$ defined by the genealogy, that is, for two words u and uv with $v \neq \varnothing$, the former is an *ancestor* of the latter, while the latter is a *descendant* of the former, and we denote $u \preccurlyeq uv$. 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 $\mathcal{U}(u)$ to represent the genealogical tree rooted at u

$$\mathcal{U}(u) := \{w \in \mathcal{U} | u \preccurlyeq w \text{ or } w = u\}.$$

We also use $\mathcal{U}_t(u)$ (resp. $\mathcal{U}_t^\dagger(u)$) to denote the set of labels of clusters in $\mathcal{U}(u)$ which are active (resp. inactive) at time t. One should notice that $\mathcal{U}_t(u)$ can be either the set of labels of active descendants of u at time t, or u itself if it is still alive at that moment, or empty if it encounters isolation first at a time no later than t, so we consider $\mathcal{U}_t(u)$ as the "set of labels of active clusters at t issued from the cluster of the label u". Finally, if $u \in \mathcal{U}_t$ (resp. $u \in \mathcal{U}_t^\dagger$), we denote by $\mathscr{X}_t^u$ (resp. $\mathscr{Y}_t^u$) for its associated cluster, and X_t^u (resp. Y_t^u) its size, that is, $X_t^u = |\mathscr{X}_t^u|$ (resp. $Y_t^u = |\mathscr{Y}_t^u|$); we write $X_t^u = 0$ for $u \notin \mathcal{U}_t$. We also use $X_t(n)$ (resp. $Y_t(n)$) for the number of active clusters (resp. inactive clusters) of size n at time t. With these notation, we have $X_t = \sum_{u\in\mathcal{U}_t} \delta_{X_t^u}$ and for any function f from $\mathbb{N}_+$ to $\mathbb{R}$:

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

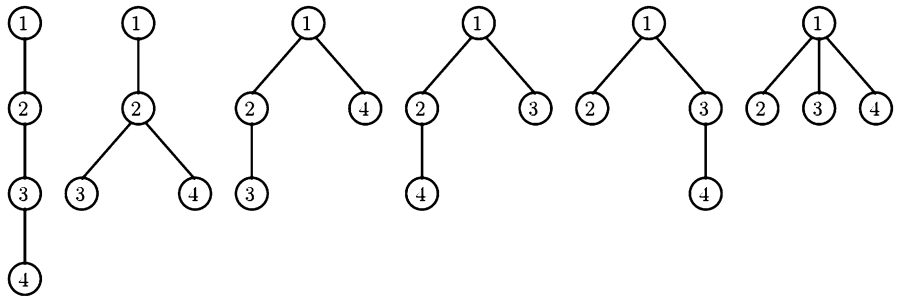


FIG. 5. *All the recursive trees (as representatives of equivalence classes) in* $\mathscr{T}_4$.

3.2. *Random recursive trees.* The genealogy in a cluster is given by a recursive tree and here we define some notation related to the latter. Given $V = \{a_1, \dots, a_n\} \subset \mathbb{R}$ with increasing order $a_1 < a_2 < \cdots < a_n$, a *recursive tree* t on V is a rooted tree with V as the set of vertices, such that for any a_i, $2 \le i \le n$, the path from a_1 to a_i is increasing. Thus, a descendant has a larger label than that of the parent. The minimal element a_1 is called the *root* of t. The collection of all the recursive trees on V has cardinality $(|V| - 1)!$.

We define the equivalence relation $\sim$ between 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 $\mathrm{t}_1 \sim \mathrm{t}_2$ if and only if there exists an order-preserving function $\psi : V_1 \to V_2$, which induces a bijection between the trees t_1 and t_2. We denote by $\mathscr{T}_n$ the set of recursive trees of size n up to the equivalence relation $\sim$, and use the recursive trees defined on $\{1, \dots, n\}$ to represent equivalence classes (canonical representation); see Figure 5 for an example of $\mathscr{T}_4$. Finally, we define the space of finite recursive trees up to equivalence relation $\sim$

$$\mathscr{T} := \bigcup_{n=1}^{\infty} \mathscr{T}_n. \tag{6}$$

A *(uniform) random recursive tree (RRT)* of size n is a random element chosen uniformly in $\mathscr{T}_n$. We denote by T_n this random equivalence class. With a slight abuse, RRT can refer both to the equivalence class or a specific labelling (for instance, with the first integers). Since $\mathscr{T}$ defined in (6) contains only countably many elements, the space $\mathscr{T}$ is a Polish space under

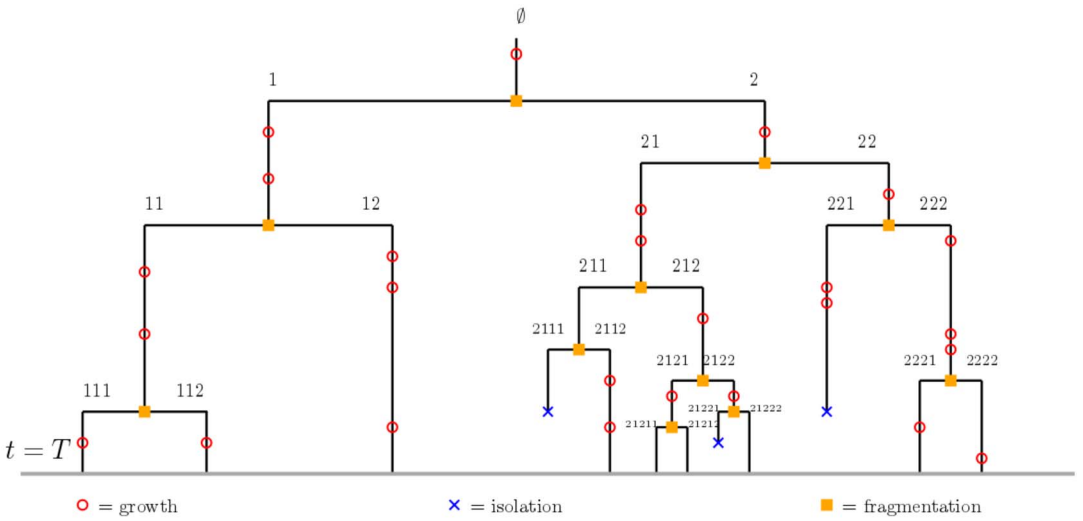


FIG. 6. *An illustration of the genealogy of clusters.*

the trivial distance. We can construct more general probability measures than uniform random recursive tree on $\mathscr{T}$. For example, for any ν a probability measure on $\mathbb{N}_+$, we use the notation T_ν to represent a random variable on $\mathscr{T}$, such that we sample first the size by ν, then sample an equivalence class uniformly given its size, that is, for any bounded function f on $\mathscr{T}$,

$$\mathbb{E}\big[f(T_\nu)\big] = \sum_{n=1}^{\infty} \nu(n)\mathbb{E}\big[f(T_n)\big] = \sum_{n=1}^{\infty} \nu(n)\bigg(\frac{1}{(n-1)!}\sum_{\mathbf{t}\in\mathscr{T}_n} f(\mathbf{t})\bigg). \tag{7}$$

This gives the rigorous definition for $\mathbb{E}[f(T_\pi)]$ and $\mathbb{E}[f(T_{\widetilde{\pi}})]$ in Theorem 2.3.

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 labelled $(k+1)$ 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, independent of the history of construction given the current state.

The key property of RRT that we need is the splitting property. We state this property below and its proof can be found in [31] and [11]. We will explain what role this property plays in our model in the next section.

PROPOSITION 3.1 (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*

$$\mathbb{P}\big[|T_n^*| = j\big] = \frac{n}{n-1}\frac{1}{j(j+1)}, \quad j = 1, 2, \ldots, n-1. \tag{8}$$

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

3.3. *Reduction to the size process.* Let us explain more explicitly how the study of $(\mathscr{X}_t, \mathscr{Y}_t)_{t\geq 0}$ can be reduced to that 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 point measures on $\mathbb{N}_+$ and endow it with the weak topology and corresponding Borel algebra. Then $(X_t, Y_t)_{t\geq 0}$ is a $\mathcal{M}^2$-valued process and we denote by $(\mathcal{F}_t)_{t\geq 0}$ its natural filtration. In fact, $(\mathcal{F}_t)_{t\geq 0}$ is also the natural filtration of $(X_t)_{t\geq 0}$ since Y_t is a deterministic function of $(X_s)_{0\leq s\leq t}$ for any $t \geq 0$.

PROPOSITION 3.2. *Fix arbitrary $t \geq 0$. Conditionally on $(X_t^u)_{u\in\mathcal{U}_t}$ and $(Y_t^u)_{u\in\mathcal{U}_t^\dagger}$, the clusters in $\mathscr{X}_t \cup \mathscr{Y}_t$ are independent RRTs 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 measure-valued Markov branching process in $(\mathcal{M}^2, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.*

PROOF. The RRT-distribution of clusters and conditional independence is obvious at the initial time when there is only one single vertex. Let us check that this 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 a direct consequence of the branching property of the cluster process.

We need to consider three events and their corresponding rates. First, a cluster is isolated with a rate depending only on its size, and then becomes inactive. Second, the growth rate of a cluster also only depends on its size, and the new vertex is added independently of the state of other clusters. Thus, after a growth event, the new cluster remains independent from the other ones (conditionally on the sizes). Moreover, by the construction of RRT by adding a

vertex uniformly at random, we know that the new cluster is also distributed as a RRT. Third, for fragmentation, we invoke the splitting property (Proposition 3.1), which guarantees that the two new clusters are independent RRTs conditionally on their sizes. This also ensures the Markov property thanks to the absence of memory for each event. □

Let us sum up the dynamics of active clusters and the transition for the size process. Independently, each active cluster of size n:

(i) becomes an inactive cluster of size n at rate θn (death of the active cluster and birth of an inactive cluster);
(ii) becomes a RRT of size $(n+1)$ at rate βn (cluster size increases by 1);
(iii) splits into two RRTs of sizes $(n-j, j)$ at rate $\gamma n \frac{1}{j(j+1)}$, for $n \geq 2, 1 \leq j \leq n-1$ (fragmentation). Here $(n-j)$ is the size of the first child (containing the root) and j is the size of the second child.

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

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

and define

$$\begin{aligned}
\mathcal{A}F_{f,g}(\mu,\nu) &= \sum_{n=1}^{\infty} \mu(\{n\})\beta n\big(F\big(\langle \mu+\delta_{n+1}-\delta_n, f\rangle, \langle \nu, g\rangle\big) - F\big(\langle \mu, f\rangle, \langle \nu, g\rangle\big)\big) \\
&\quad + \sum_{n=1}^{\infty} \mu(\{n\})\theta n\big(F\big(\langle \mu-\delta_n, f\rangle, \langle \nu+\delta_n, g\rangle\big) - F\big(\langle \mu, f\rangle, \langle \nu, g\rangle\big)\big) \\
&\quad + \sum_{n=1}^{\infty} \mu(\{n\})\gamma n \sum_{j=1}^{n-1}\Big(\frac{1}{j(j+1)}\Big) \\
&\quad \times \big(F\big(\langle \mu+\delta_j+\delta_{n-j}-\delta_n, f\rangle, \langle \nu, g\rangle\big) - F\big(\langle \mu, f\rangle, \langle \nu, g\rangle\big)\big).
\end{aligned} \tag{9}$$

Although we start from one single active individual to define our model for convenience (see Section 8.2 for a discussion on the initial condition), we remark that the size process can be defined from any finite initial state since it is a measure-valued process on integers.

4. First moment semigroup of $(X_t)_{t\geq 0}$ and Perron's root. In this part, we study the first moment semigroup associated to the process $(X_t)_{t\geq 0}$. We will establish the existence of Perron's eigenelements and speed of convergence and prove Theorem 2.1.

4.1. *Semigroup and generator.* Thanks to Section 3.2, the study of the model is reduced to the long-time behaviour of the measure-valued Markov branching process $(X_t, Y_t)_{t\geq 0}$. We denote the first moment semigroup associated to $(X_t)_{t\geq 0}$ by $M = (M_t)_{t\geq 0}$. It is defined for any nonnegative function f by

$$M_t f(n) := \mathbb{E}_{\delta_n}\big[\langle X_t, f\rangle\big] \quad \forall t \geq 0, n \geq 1, \tag{10}$$

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 $n, m \in \mathbb{N}_+$,

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

which is the mean number of clusters of size m at time t descending from one single RRT of size n at time 0.

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 $[x](n) = n$. We introduce $\mathcal{B}$ the set of functions from $\mathbb{N}_+$ to $\mathbb{R}$ with at most polynomial growth

$$\mathcal{B} := \Big\{ f : \mathbb{N}_+ \to \mathbb{R}, \exists p > 0 \text{ such that } \sup_{n \geq 1} |f(n)|/n^p < \infty \Big\}. \tag{11}$$

Let $\mathcal{B}_p$ be the set of functions with $p > 0$ fixed

$$\mathcal{B}_p := \Big\{ f : \mathbb{N}_+ \to \mathbb{R}, \sup_{n \geq 1} |f(n)|/n^p < \infty \Big\}. \tag{12}$$

Notice that $\mathcal{B}_p \subset \mathcal{B}_{p'} \subset \mathcal{B}$ for $0 < p \leq p'$, and $\mathcal{B} = \bigcup_{p>0} \mathcal{B}_p$. We then extend now the first moment semigroup to the space $\mathcal{B}$.

LEMMA 4.1.

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

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

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

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

(iii) *$(M_t)_{t \geq 0}$ is a positive semigroup on $\mathcal{B}$ and for any $f \in \mathcal{B}$, we have for $t \geq 0$ and $n \geq 1$*

$$\frac{\mathrm{d}}{\mathrm{d}t} M_t f(n) = M_t(\mathcal{L} f)(n), \tag{13}$$

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

$$\begin{aligned} \mathcal{L} f(n) = {} & \underbrace{\beta n \big(f(n+1) - f(n)\big)}_{\text{“growth”}} \underbrace{-\theta n f(n)}_{\text{“isolation”}} \\ & + \underbrace{\sum_{j=1}^{n-1} \frac{\gamma n}{j(j+1)} \big(f(j) + f(n-j) - f(n)\big)}_{\text{“fragmentation”}}. \end{aligned} \tag{14}$$

PROOF. Linear operator $\mathcal{L}$ yields the generator of the first moment semigroup of the size process. We focus on (i) and study $\mathcal{L}([x^p])$ at first. 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 in $\mathcal{L}([x^p])$ is negative. Thus we have

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

We then apply convexity for $n \in \mathbb{N}_+$

$$(n+1)^p - n^p \leq p(n+1)^{p-1} \leq p 2^{p-1} n^{p-1}.$$

It gives us

$$\mathcal{L}\big([x^p]\big) \leq \big(2^{p-1}p\beta - \theta[x]\big)[x^p] \leq \big(2^{p-1}p\beta - \theta\big)[x^p]. \tag{15}$$

Here we use simply the fact that $[x](n) = n \geq 1$ for all n for the isolation term.

The rest of the proof follows classical arguments of localisation, see for example, Theorem 1 in [32], and we give here the main lines only. 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 $X_t^m = X_{t\wedge\tau_m}$, $Y_t^m = Y_{t\wedge\tau_m}$, where

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

Notice that on the event $\{\tau_m \geq t\}$, 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 transition rates. Consider three functions $f, g : \mathbb{N}_+ \to \mathbb{R}_+$ and $F : \mathbb{R}^2 \to \mathbb{R}$ a bounded Borel function and recall that $F_{f,g}(\mu, \nu) = F(\langle \mu, f\rangle, \langle \nu, g\rangle)$. We get by Dynkin's formula

$$\mathbb{E}_{\delta_n}\big[F_{f,g}(X_t^m, Y_t^m)\big] = \mathbb{E}_{\delta_n}\big[F_{f,g}(X_0^m, Y_0^m)\big] + \mathbb{E}_{\delta_n}\bigg[\int_0^{t\wedge\tau_m} \mathcal{A}F_{f,g}(X_s, Y_s)\,\mathrm{d}s\bigg],$$

where $\mathcal{A}$ is defined in (9). We apply this equation with $F(x, y) = x \wedge m^p$ and $f = [x^p]$ and $g = 0$ to obtain

$$\mathbb{E}_{\delta_n}\big[\langle X_t^m, [x^p]\rangle\big] = \mathbb{E}_{\delta_n}\big[\langle X_0^m, [x^p]\rangle\big] + \mathbb{E}_{\delta_n}\bigg[\int_0^{t\wedge\tau_m} \langle X_s, \mathcal{L}[x^p]\rangle\,\mathrm{d}s\bigg]. \tag{16}$$

Using the above display and (15) yields

$$\mathbb{E}_{\delta_n}\big[\langle X_t^m, [x^p]\rangle\big] \leq \mathbb{E}_{\delta_n}\big[\langle X_0^m, [x^p]\rangle\big] + (2^{p-1}p\beta - \theta)\mathbb{E}_{\delta_n}\bigg[\int_0^{t\wedge\tau_m} \langle X_s, [x^p]\rangle\,\mathrm{d}s\bigg].$$

Since the process (X, Y) is nonexplosive, τ_m tends a.s. to infinity as m tends to infinity. Moreover, τ_m 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

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

Grönwall's lemma then ensures that (i) and (ii) are immediate consequences.

For (iii), thanks to the fact $\mathcal{B}_p \subset \mathcal{B}_{p'} \subset \mathcal{B} = \bigcup_{p>0}\mathcal{B}_p$ for any $p' \geq p > 0$, we only need to show that for any $f \in \mathcal{B}_p$ with $p \geq 1$, we have

$$\mathbb{E}_{\delta_n}\big[\langle X_t, f\rangle\big] = \mathbb{E}_{\delta_n}\big[\langle X_0, f\rangle\big] + \mathbb{E}_{\delta_n}\bigg[\int_0^t \langle X_s, \mathcal{L}f\rangle\,\mathrm{d}s\bigg]. \tag{17}$$

Let $C > 0$ be such that $|f(n)| \leq Cn^p$, $|\mathcal{L}f(n)| \leq Cn^{p+1}$ for all n. Similar to (16), we have

$$\mathbb{E}_{\delta_n}\big[\langle X_t^m, f\rangle\big] = \mathbb{E}_{\delta_n}\big[\langle X_0^m, f\rangle\big] + \mathbb{E}_{\delta_n}\bigg[\int_0^{t\wedge\tau_m} \langle X_s, \mathcal{L}f\rangle\,\mathrm{d}s\bigg]. \tag{18}$$

We will show the limits of the left and right terms in above display. We first study the left term which is the sum of the two terms below

$$\mathbb{E}_{\delta_n}\big[\langle X_t^m, f\rangle \mathbf{1}_{\{t<\tau_m\}}\big] \quad \text{and} \quad \mathbb{E}_{\delta_n}\big[\langle X_t^m, f\rangle \mathbf{1}_{\{t\geq\tau_m\}}\big]. \tag{19}$$

For the first term in (19), we have

$$\langle X_t^m, f\rangle \mathbf{1}_{\{t<\tau_m\}} \xrightarrow{m\to\infty} \langle X_t, f\rangle \quad \text{almost surely,}$$

and

$$\big|\langle X_t^m, f\rangle \mathbf{1}_{\{t<\tau_m\}}\big| \leq \langle X_t, |f|\rangle \leq C\langle X_t, [x^p]\rangle.$$

By (i), the last term in the above display has a finite mean. Using dominated convergence theorem, we obtain

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

We prove now that the second term in (19) vanishes. We use a coupling argument below

$$\mathbb{E}_{\delta_n}\big[\big|\langle X_t^m, f\rangle \mathbf{1}_{\{t\geq\tau_m\}}\big|\big] \leq m^p \mathbb{P}_{\delta_n}[\tau_m \leq t] \leq m^p \mathbb{P}_{\delta_n}\big[\langle \widetilde{X}_t, [x]\rangle \geq m\big],$$

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

$$\mathbb{E}_{\delta_n}\big[\big|\langle X_t^m, f\rangle \mathbf{1}_{\{t\geq\tau_m\}}\big|\big] \leq m^p \frac{\mathbb{E}_{\delta_n}[(\langle \widetilde{X}_t, [x]\rangle)^{2p}]}{m^{2p}} \xrightarrow{m\to\infty} 0.$$

Combining the analysis of the two terms in (19), the term on the left-hand side in (18) converges to $\mathbb{E}_{\delta_n}[\langle X_t, f\rangle]$ as $m \to \infty$.

Now we turn to the right-hand side of (18). Note that

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

Due to (i), the last term has finite mean. Moreover $\int_0^{t\wedge\tau_m} \langle X_s, \mathcal{L}f\rangle \,\mathrm{d}s \longrightarrow \int_0^t \langle X_s, \mathcal{L}f\rangle \,\mathrm{d}s$, almost surely as $m \to \infty$. Applying dominated convergence theorem yields

$$\mathbb{E}_{\delta_n}\left[\int_0^{t\wedge\tau_m} \langle X_s, \mathcal{L}f\rangle \,\mathrm{d}s\right] \xrightarrow{m\to\infty} \mathbb{E}_{\delta_n}\left[\int_0^t \langle X_s, \mathcal{L}f\rangle \,\mathrm{d}s\right].$$

Letting $m \to \infty$ in (18) gives (17) and ends the proof. □

4.2. *Perron's root and eigenvectors.* In this part, we study the asymptotic behaviour of the first moment semigroup $(M_t)_{t\geq 0}$ of $(X_t)_{t\geq 0}$. Under general assumptions extending the Perron–Frobenius theory in the finite-dimension setting, the ergodic behaviour of the positive semigroup is given by the unique triplet of eigenelements corresponding to the maximal eigenvalue. We refer in particular to [7, 33] and references therein for general statements and applications to growth-fragmentation. In this work, we apply a general statement of [7] on the ergodic behaviour 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 (22), and the fast splitting or isolation of large clusters which provides a Lyapunov function for a typical cluster (21). The fact that splitting is very asymmetric (such that one child cluster is close to the parent in size) and the fact that a typical active cluster at a given time has avoided isolation make the proof of the required lower bound delicate. It involves a subtle compensation of fragmentation and isolation terms. Moreover, we will show the exponential speed of convergence of the semigroup. 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 (including constants). In this case, we say f is upper bounded by g or g is lower bounded by f. We also use $f \sim g$ to denote the fact that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$. We prove at first Lemma 4.2, the key technical ingredient for Proposition 4.1, the latter is the main result of this subsection. The following space of sublinear functions is useful to control the harmonic function of the semigroup:

$$\mathcal{S} := \Bigg\{ f : \mathbb{N}_+ \to [1,\infty), \text{ such that}$$

$$(a)\, f \text{ is increasing and } \lim_{n\to\infty} f(n) = \infty \tag{20}$$

$$(b)\, f \text{ is sublinear } \frac{f(n+1)}{n+1} \leq \frac{f(n)}{n}, \text{ and } C_f := \sum_{j=1}^{\infty} \frac{f(j)}{j(j+1)} < \infty \Bigg\}.$$

LEMMA 4.2. *There exists a positive function ψ defined on $\mathbb{N}_+$, such that*

$$0 < \inf_{\mathbb{N}_+} \psi < \sup_{\mathbb{N}_+} \psi \le 1$$

and for every $V \in \mathcal{S} \cup \{[x^p] : p \ge 1\}$:

(i) *There exist real constants $a < b$ and $\zeta > 0$ such that*

$$\mathcal{L}V \le aV + \zeta\psi, \qquad b\psi \le \mathcal{L}\psi \le \xi\psi. \tag{21}$$

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

$$M_{t_0}(x, y) > 0. \tag{22}$$

PROOF. To find the Lyapunov-type functions in (i), the main difficulty is to find the lower bound of $\mathcal{L}\psi$ in (21). As we can see in (14), 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 ψ—setup. We set

$$\psi(n) := A - (A - B)q^{n-1}, \tag{23}$$

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

$$\begin{aligned}\mathcal{L}\psi(n) = {} & \underbrace{\beta n\big(\psi(n+1) - \psi(n)\big)}_{\mathbf{I}} - \underbrace{(\theta + \gamma)n\psi(n)}_{\mathbf{II}} + \gamma\psi(n) \\ & \underbrace{+ \gamma n \sum_{j=1}^{n-1} \frac{1}{j(j+1)}\big(\psi(j) + \psi(n-j)\big)}_{\mathbf{III}}.\end{aligned} \tag{24}$$

First

$$|\mathbf{I}| = \big|\beta(A - B)n\big(q^{n-1} - q^n\big)\big| \le C_1\psi(n), \tag{25}$$

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

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

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

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

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

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

Step 2: Construction of ψ—choice of parameters. We add and subtract the term $(1 + C_q^{A,B})\gamma n\psi(n)$ and reformulate (24) as

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

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

$$
\begin{aligned}
R_1(n,q) &:= \gamma n\left(\sum_{j=1}^{n-1}\left(\frac{1}{j(j+1)}\psi(j)\right) - C_q^{A,B}\psi(n)\right),\\
R_2(n,q) &:= \gamma n\left(\sum_{j=1}^{n-1}\left(\frac{1}{j(j+1)}\psi(n-j)\right) - \psi(n)\right).
\end{aligned}
\tag{26}
$$

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

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

$$
\sup_{n\in\mathbb{N}_+}\left\{\left|\frac{R_1(n,q)}{\psi(n)}\right| + \left|\frac{R_2(n,q)}{\psi(n)}\right|\right\} \le C_2. \tag{27}
$$

Together with (25), we obtain that

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

This guarantees that the last two inequalities of (21) hold with the following choice of parameters:

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

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

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

Using the first inequality in (15) and distinguishing if $2^{p-1}p\beta - \theta n$ is larger than a or not, we can write

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

with $\zeta \in (0,\infty)$. The above result holds because ψ is bounded and there exist only finitely many n satisfying $2^{p-1}p\beta - \theta n \ge a$. This ends the proof of (i) for $p \ge 1$. Besides, for any large R, the set K is finite and nonempty. The combination of growth, fragmentation and isolation ensures the irreducibility of $(X_t)_{t\ge 0}$ which allows one to end the proof of (ii). So both (i) and (ii) are proved for $V = [x^p]$ with $p \ge 1$.

Now we treat the case $V \in \mathcal{S}$ and verify the condition (i) and (ii) with ψ given in Step 1. For condition (ii), we only have to show that K is nonempty and finite, which is straightforward to see since V is increasing to infinity while ψ is a bounded function. For condition (i), we only have to show the first half of (21). We calculate $\mathcal{L}V$ and use the decomposition in (14). Note that $\frac{V(n+1)}{n+1} \le \frac{V(n)}{n}$ implies for the growth term that

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

Then the fact that V increases and C_V in (20) is finite yields for the fragmentation term **III**:

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

The above two displays entail that

$$\mathcal{L}V(n)\le(\beta-\theta n)V(n)+C_V\gamma n. \tag{28}$$

Now we pick a real number a such that $a<\min\{\beta-\theta,b\}$. Using $\lim_{n\to\infty}V(n)=\infty$, we notice that $E:=\{n\in\mathbb{N}_+:(\beta-\theta n)V(n)+C_V\gamma n>aV(n)\}$ is a nonempty finite set. Then distinguishing the cases whether n belongs to E or not in (28) yields

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

Here the constant ζ is defined by

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

We conclude that both conditions (i) and (ii) are verified. The whole proof is complete and finished. □

Now we come to the main result of this subsection which gives the existence of eigenelements and asymptotic behaviour of the semigroup, based on Theorem 2.1 in [7].

PROPOSITION 4.1. *There exists a unique triplet* (λ,π,h) *where* $\lambda\in\mathbb{R}$ *and* $\pi=(\pi(n))_{n\in\mathbb{N}_+}$ *is a positive vector of probability distribution and* $h:\mathbb{N}_+\to(0,\infty)$ *is a positive function, such that for all* $t\ge0$,

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

and $0<\inf_{n\ge1}h(n)\le\sup_{n\ge1}h(n)<\infty$ *and* $\sum_{n\ge1}\pi(n)=\sum_{n\ge1}\pi(n)h(n)=1$.

Besides, for every $p>0$ *there exists* $C,\omega>0$ *such that for any* $n,m\ge1$, $t\ge0$,

$$\big|e^{-\lambda t}M_t(n,m)-h(n)\pi(m)\big|\le Cn^p m^{-p}e^{-\omega t},\qquad\sum_{n\ge1}\pi(n)n^p<\infty. \tag{29}$$

PROOF. Lemma 4.2 together with Lemma 4.1(iii) ensures that the semigroup M satisfies the drift and irreducibility conditions given in [7] (Propositions 2.2 and 2.3 therein). More precisely, these conditions are met with $V=[x^p]$ for $p>0$ and $\varphi=\psi$ (defined in (23)), while $\psi\le V$ is guaranteed by the fact that $\psi\le1$. Using these conditions, we can apply Theorem 2.1 in [7], which yields the proposition, except the boundedness of h. In particular, (29) is obtained by specifying the initial condition $\mu=\delta_n$ and using the test function 1_m.

We prove at first that h is upper bounded. Lemma 3.4 in [7] ensures that h is upper bounded by V times a constant (i.e., $h\lesssim V$ in their notation). Adding that Lemma 4.2 guarantees that we can pick a $V\in\mathcal{S}$ that increases arbitrarily slowly, we obtain that h is upper bounded. Finally, we justify that h is lower bounded. Indeed $h(n)=e^{\lambda}M_1h(n)\ge ch(1)$ for all $n\ge1$, where $c>0$. This is because, due to (8), the probability that a cluster of size n produces (by fragmentation) a cluster of size one before unit time 1 and that this latter stays unchanged in the remaining time within the unit time is lower bounded by a positive constant independent of its size n. □

Equation (29) ensures that for any $f \in \mathcal{B}_p$,

$$\left|e^{-\lambda t} M_t f(n) - h(n)\langle \pi, f\rangle\right| \le C n^{p+2} \|f\|_p e^{-\omega t}, \tag{30}$$

where $\|f\|_p := \sum_{m\ge 1} m^{-(p+2)} |f(m)| < \infty$. This result will be useful later.

The fact that the eigenvector h is lower and upper bounded in $(0,\infty)$ ensures that the impact of the size of the initial cluster on the first order approximation of the profile remains bounded. Furthermore, we expect that restricting the set of test functions to bounded functions, the current result (29) can be enhanced as uniform ergodic convergence. Indeed, we may apply Theorem 3.5 in [6] with $\nu = \delta_1$, using again that large clusters produce with high probability clusters of small sizes at fragmentation and are fast isolated. The remaining difficulty lies in controlling uniformly $M_t \mathbf{1}(n)/M_t\mathbf{1}(1)$ in time and size.

At this point one may want to apply [2] to prove strong convergence using the asymptotic behaviour of the first moment semigroup $(M_t)_{t\ge 0}$. But [2] requires stronger assumptions than what is obtained in (29), in particular in terms of the stationary distribution π. Besides, we are interested in finer and more quantitative estimates, with motivations coming from inference and epidemiology. We thus follow another 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 for example, [8, 29] and references therein. The idea is to use the most recent common ancestor of two individuals to decouple their values.

LEMMA 4.3. *For any $x \in \mathbb{N}_+$ and $f \in \mathcal{B}$, we have*

$$\mathbb{E}_{\delta_x}\big[\langle X_t, f\rangle^2\big] = M_t\big(f^2\big)(x) + 2\int_0^t \sum_{n\ge 1} M_s(x,n)\Bigg(\sum_{1\le j\le n-1} \kappa(n,j) M_{t-s} f(j) M_{t-s} f(n-j)\Bigg)\,\mathrm{d}s,$$

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

REMARK 1. Combining this identity with the estimates on the semigroup M obtained in the previous subsection gives L^2 convergence of the empirical measure of clusters, with exponential speed and size dependency; see next section.

PROOF. Recalling notation in Section 3.1, we have

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

Recall also that, for any $u, v \in \mathcal{U}$, $u \wedge v$ is the label of the most recent common ancestor of u and v, and $\mathcal{U}_t(u)$ as the set of active clusters at t issued from u. We first notice that

$$\begin{aligned}\langle X_t, f\rangle^2 &= \sum_{u,v\in\mathcal{U}_t} f\big(X_t^u\big) f\big(X_t^v\big) = \sum_{u\in\mathcal{U}_t} f^2\big(X_t^u\big) + \sum_{w\in\mathcal{U}} \sum_{\substack{u,v\in\mathcal{U}_t,\\ u\ne v, u\wedge v = w}} f\big(X_t^u\big) f\big(X_t^v\big) \\ &= \sum_{u\in\mathcal{U}_t} f^2\big(X_t^u\big) + \sum_{w\in\mathcal{U}} \mathbf{1}_{\{b(w)<t\}} I_t(w),\end{aligned} \tag{31}$$

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

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

Thus $I_t(w)$ is the cross term between the active clusters on the two genealogical subtrees rooted at $w1$ and $w2$.

Concerning the equation (31), we have first

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

Second, 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 to get

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

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

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

Combining these identities, we obtain

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

where we introduce

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

This function involves the fragmentation event and can be written explicitly by recalling that, when a cluster of size n splits, the probability that the size of the first child is $(n-j)$ and the second child in j is $n/((n-1)\cdot j\cdot(j+1))$. That is, we can obtain

$$g(n,s) = \sum_{1\le j\le n-1} \frac{n}{n-1} \frac{1}{j(j+1)} M_{t-s} f(j) M_{t-s} f(n-j). \tag{32}$$

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

$$\begin{aligned} &\mathbb{E}_{\delta_x}\Big[\sum_{w\in\mathcal{U}} \mathbf{1}_{\{b(w)<t\}} g\big(X_{b(w)-}^{w}, b(w)\big) \Big] \\ &\quad = \int_0^t \sum_{w\in\mathcal{U},n\ge1} g(n,s) \mathbb{P}_{\delta_x}\big[w\in\mathcal{U}_{s-}, X_{b(w)-}^{w} = n, b(w) \in \mathrm{d}s \big] \\ &\quad = \int_0^t \sum_{w\in\mathcal{U},n\ge1} g(n,s) \mathbb{P}_{\delta_x}\big[w\in\mathcal{U}_{s-}, X_{b(w)-}^{w} = n \big] \gamma(n-1)\,\mathrm{d}s \\ &\quad = \int_0^t \sum_{n\ge1} g(n,s)\gamma(n-1) M_s(x,n)\,\mathrm{d}s. \end{aligned}$$

This equation and (32) give us the expression of κ. The proof is thus completed. □

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

PROPOSITION 4.2. *The process $(\mathscr{M}_t)_{t\geq 0}$ defined as*

$$\mathscr{M}_t = e^{-\lambda t}\langle X_t, h\rangle, \tag{33}$$

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

PROOF. The martingale property is classical and the proof is given for the sake of completeness. Recall the notation $\mathcal{U}_t$ and X_t^u introduced in Section 3.1. For any $u \in \mathcal{U}_t$, recall that $\mathcal{U}_{t+s}(u)$ is the set of labels of all the clusters active at time $(t+s)$ issued from the cluster labelled by u active at time t. Then we have

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

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

Let us now prove the L^2 convergence. We apply Lemma 4.3 with $x = 1$ and obtain

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

where

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

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 given in Proposition 4.1, 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 (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

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

which is uniformly upper bounded 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

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

and we apply (30) to conclude that $\sup_{t\geq 0}\mathbb{E}[(\mathscr{M}_t)^2] < \infty$. Then by the martingale convergence theorem, we obtain that $\mathbb{E}[W^2] < \infty$ and $(\mathscr{M}_t)_{t\geq 0}$ converges in the L^2 norm to W. □

REMARK 2. In fact, Proposition 4.2 holds for $(X_t)_{t\geq 0}$ under $\mathbb{P}_{\delta_n}$ for any $n \geq 1$, with the limit W depending on n. The proof is essentially the same as for $n = 1$. We state here the result under $\mathbb{P} = \mathbb{P}_{\delta_1}$ so that the limit W is consistent with Theorem 2.2 and Corollary 1.

4.4. *Proof of Theorem* 2.1. With the help of Proposition 4.1 and Proposition 4.2, we are now ready to prove our Theorem 2.1.

PROOF OF THEOREM 2.1. We use the classical notation $\underline{\lim}$ and $\overline{\lim}$ respectively for the limit inferior and limit superior of a sequence defined on discrete or continuous time. We notice that $\mathbb{E}[|\mathscr{X}_t|] = \mathbb{E}[\langle X_t, 1\rangle] = \sum_{j=1}^{\infty} M_t(1, j)$ and we apply (30) with $f \equiv \mathbf{1}$ (constant function) and $n = p = 1$. This ensures that $\lim_{t\to\infty} \log(\mathbb{E}[|\mathscr{X}_t|])/t = \lambda$.

To study the limit of $\log(\mathbb{E}[|\mathscr{Y}_t|])/t$, we use Kolmogorov's equation. More precisely, following the localisation argument in the proof of Lemma 4.1 (ii)–(iii), we check that $F_{f,g}(\mu, \nu) = \langle \nu, 1\rangle$ belongs to the domain of the extended generator of $(X_t, Y_t)_{t\geq 0}$ (see (9)). Then we get

$$\mathbb{E}\big[|\mathscr{Y}_t|\big] = \mathbb{E}\big[\langle Y_t, 1\rangle\big] = \int_0^t \mathbb{E}\big[\langle X_s, \theta[x]\rangle\big]\,\mathrm{d}s = \int_0^t M_s\big(\theta[x]\big)(1)\,\mathrm{d}s,$$

and we conclude using (30).

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

- In the subcritical phase $\lambda < 0$, (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 4.2 and the optional stopping theorem to get

$$h(1) = \mathbb{E}\Big[\lim_{t\to\infty} e^{-\lambda(t\wedge\tau)}\langle X_{t\wedge\tau}, h\rangle\Big] = \mathbb{E}[W\mathbf{1}_{\{\tau=\infty\}}]. \tag{34}$$

  Adding that $h > 0$ from Proposition 4.1 implies that $\mathbb{P}[\tau = \infty] > 0$.
- In the critical phase $\lambda = 0$, we first observe that the probability of extinction starting from one cluster, within a unit time, is greater than a positive constant (uniformly with respect to the cluster size that we started with). Besides, $\underline{\lim}_{t\to\infty} |\mathscr{X}_t| < \infty$ a.s. since Fatou's lemma ensures that

$$\mathbb{E}\Big[\underline{\lim_{t\to\infty}}\,|\mathscr{X}_t|\Big] = \mathbb{E}\Big[\underline{\lim_{t\to\infty}}\,\langle X_t, \mathbf{1}\rangle\Big] \leq \lim_{t\to\infty} \mathbb{E}\big[\langle X_t, h\rangle\big] / \inf_{n\geq 1} h(n) = h(1) / \inf_{n\geq 1} h(n) < \infty.$$

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

5. Strong convergences. The Perron's root $\lambda \in \mathbb{R}$ and associated eigenelements have been characterised in Proposition 4.1. The sign of λ determines if the first moment semigroup goes to 0 or infinity. We turn now to trajectorial results and find first an equivalent statement of being extinct (Kesten–Stigum). 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 $\{W > 0\}$ coincides with survival event $\{\tau = \infty\}$ or not. This is one part of the Kesten–Stigum theorem in branching processes; see, for example, [26, 27]. In our case, the L^2 computation ensures that $\mathbb{P}[W > 0] > 0$ and we will get a positive answer to the question posed.

PROPOSITION 5.1. *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 in Proposition 4.2. Indeed, given (34) and the comments below, we obtain that $\mathbb{P}[W > 0] > 0$. Besides, it is straightforward to see that $\{W > 0\} \subset \{\tau = \infty\}$. Thus if $\mathbb{P}[W > 0] = \mathbb{P}[\tau = \infty]$, the proof is done. The lines of the proof are classical, even though the sizes of clusters being unbounded requires some specific arguments.

First, we use the fact that any cluster can be isolated (before any other event happening to it) during a unit time interval, with the isolation probability lower bounded by a positive value for all sizes. As a result, the number of clusters has to tend to infinity to survive:

$$\{\tau = \infty\} = \Big\{ \lim_{t\to\infty} |\mathscr{X}_t| = \infty \Big\} \quad \text{a.s.}$$

Second, we deduce from the above 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 unchanged and other clusters have a positive probability to create (by fragmentation) one cluster of size one, and the latter probability is lower bounded uniformly with respect to the size $n \geq 2$. By independence of clusters and Markov inequality, this ensures that

$$\overline{\lim_{t\to\infty}} X_t(1) = \infty \quad \text{a.s. on } \{\tau = \infty\}.$$

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

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

We obtain for $t \geq \tau_N$

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

where $\mathcal{A}_N := \{u \in \mathcal{U}_{\tau_N} : X^u_{\tau_N} = 1\}$. By Proposition 4.2, $e^{-\lambda(t-\tau_N)}\sum_{v\in\mathcal{U}_t(u)} h(X_t^v)$ converges to a nonnegative 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

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

As a result, $\mathbb{P}[W = 0] = 1$ or $\mathbb{P}[W = 0, \tau = \infty] = 0$ (by letting $N \to \infty$). Only the latter is possible since we know that $\mathbb{P}[W > 0] > 0$. The proof is thus finished. □

5.2. *Strong law of large numbers for the size process of active clusters.* In this part, we prove Theorem 2.2 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 convergence speed obtained entails strong convergence of subsequences. The strategy is then to control fluctuations to prove the strong convergence along $t \in \mathbb{R}_+$. To that purpose, we follow the idea from [3]. We divide the proof into three steps.

PROOF OF THEOREM 2.2. *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 any $f \in \mathcal{B}_p$ with $p > 0$. We develop the difference as follows

$$\begin{aligned}&e^{-\lambda t}\langle X_t, f\rangle - W\langle\pi, f\rangle \\ &\quad = \underbrace{e^{-\lambda t}\langle X_t, f\rangle - e^{-\lambda t}\langle X_t, h\rangle\langle\pi, f\rangle}_{\mathbf{I}} + \underbrace{e^{-\lambda t}\langle X_t, h\rangle\langle\pi, f\rangle - W\langle\pi, f\rangle}_{\mathbf{II}}.\end{aligned} \tag{35}$$

The second term **II** is nothing but $(\mathscr{M}_t - W)\langle \pi, f\rangle$, which converges in L^2 to 0 by Proposition 4.2. We only have to prove the L^2 convergence of the term **I** to 0. Denoting $g := f - \langle \pi, f\rangle h$, Lemma 4.3 yields

$$e^{2\lambda t}\mathbb{E}\big[|\mathbf{I}|^2\big] = \mathbb{E}\big[\langle X_t, g\rangle^2\big] = M_t(g^2)(1) + J_t, \tag{36}$$

where

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

Recall that $\|f\|_p = \sum_{m\geq 1}|f(m)|m^{-(p+2)} \in (-\infty, \infty)$. Observe that $g \in \mathcal{B}_p$ and let $p' > 2p+8$. By (30), there exists $C' > 0$ such that for any $n \in \mathbb{N}_+$ and $s, t \geq 0$,

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

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

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

Using the second statement in (29) and $p' > 2p+8$, the sum $\sum_{n\geq 1} n^{2p+6}(h(1)\pi(n) + n^{-p'})$ in the above display is finite and there exists $C_2 > 0$ such that

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

Moreover, by (30), $e^{-2\lambda t} M_t(g^2)(1) \leq \langle \pi, g^2\rangle h(1) e^{-\lambda t} + C\|g^2\|_p e^{-(\lambda+\omega)t}$, for any $t \geq 0$. Plugging in these estimates to (36), we see that there exists $C_3 > 0$ such that

$$\mathbb{E}\big[|\mathbf{I}|^2\big] \leq C_3\big(\langle \pi, g^2\rangle + \|g^2\|_p + \|g\|_p\big) t e^{-(\lambda\wedge 2\omega)t} \quad \forall t \geq 0. \tag{37}$$

Note that C_1, C_2, C_3 do not depend on f and t. Then Step 1 is finished.

REMARK 3. A byproduct of (37) and Proposition 4.2 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 $f \in \mathcal{B}_p$,

$$\mathbb{E}\big[\langle X_t, f\rangle^2\big] \leq C_0 e^{2\lambda t}\big(|\langle \pi, f\rangle|^2 + \big(\langle \pi, g^2\rangle + \|g^2\|_p + \|g\|_p\big)e^{-\sigma t}\big) \quad \forall t \geq 0, \tag{38}$$

with the notation $g = f - \langle \pi, f\rangle h$.

Step 2: Almost sure convergence for one type. The main idea is to extend an elegant argument from [3] to our countable-type branching process.

First, we establish an almost sure 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 (35). Then the martingale part **II** converges to 0 almost surely, and for the term **I**, (37) yields

$$\mathbb{E}\big[\big|e^{-\lambda k\Delta}\langle X_{k\Delta}, f\rangle - e^{-\lambda k\Delta}\langle X_{k\Delta}, h\rangle\langle \pi, f\rangle\big|^2\big] \leq Ck\Delta e^{-(\lambda\wedge 2\omega)k\Delta} \quad \forall k \geq 0. \tag{39}$$

By the Borel–Cantelli lemma, we get

$$e^{-\lambda k\Delta}\langle X_{k\Delta}, f\rangle \xrightarrow{k\to\infty} W\langle \pi, f\rangle \quad \text{almost surely.} \tag{40}$$

Recall that on the event $\{W = 0\}$, by Proposition 5.1, the extinction occurs a.s. So we focus on the event $\{W > 0\}$. Let us first prove that

$$e^{-\lambda t} X_t(n) \xrightarrow{t\to\infty} W\pi(n) \quad \text{almost surely.} \tag{41}$$

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 [3] is that we only need to prove the following sufficient (and necessary) condition:

$$\underline{\lim}_{t\to\infty} e^{-\lambda t} X_t(n) \geq W\pi(n) \quad \text{almost surely for all } n \geq 1. \tag{42}$$

We first show that (42) implies (41) using that the martingale convergence controls the dissipation of mass. To this purpose, for any fixed $n \geq 1$, consider any sequence of (random) times $(t_k)_{k\in\mathbb{N}_+}$ such that $\overline{\lim}_{t\to\infty} e^{-\lambda t} X_t(n) = \lim_{k\to\infty} e^{-\lambda t_k} X_{t_k}(n)$. Proposition 4.2 and Fatou's lemma and (42) ensure

$$\begin{aligned}
\overline{\lim_{t\to\infty}} e^{-\lambda t} X_t(n)h(n) &= \lim_{k\to\infty}\Big(\sum_{i\geq 1} e^{-\lambda t_k} X_{t_k}(i)h(i) - \sum_{i\geq 1, i\neq n} e^{-\lambda t_k} X_{t_k}(i)h(i)\Big) \\
&\leq W - \sum_{i\geq 1, i\neq n} \underline{\lim}_{k\to\infty} e^{-\lambda t_k} X_{t_k}(i)h(i) \\
&\leq W - \sum_{i\geq 1, i\neq n} W\pi(i)h(i) = W\pi(n)h(n).
\end{aligned} \tag{43}$$

Then together with (42) we obtain (41).

We need now to prove (42) following the arguments in [3]. Recall $\Delta > 0$ is the time step size. The proof relies on the following lower bound:

$$\forall t \in \big[k\Delta, (k+1)\Delta\big), \quad X_t(n) \geq X_{k\Delta}(n) - N_{k,\Delta}(n), \tag{44}$$

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 $(k\Delta, (k+1)\Delta)$. Indeed, to prove (42), we can find a lower bound for $\underline{\lim}_{t\to\infty} e^{-\lambda t} X_t(n)$ using the above display. More precisely,

$$\underline{\lim}_{t\to\infty} e^{-\lambda t} X_t(n) \geq \underline{\lim}_{k\to\infty} e^{-\lambda(k+1)\Delta} X_{k\Delta}(n) - \overline{\lim}_{k\to\infty} e^{-\lambda k\Delta} N_{k,\Delta}(n).$$

Using (40) for the first term on the right-hand side, we obtain

$$\underline{\lim}_{t\to\infty} e^{-\lambda t} X_t(n) \geq e^{-\lambda\Delta}\pi(n)W - \overline{\lim}_{k\to\infty} e^{-\lambda k\Delta} N_{k,\Delta}(n).$$

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

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

By branching property, we know that

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

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 Δ small such that $\mathbb{P}[\xi_i = 1] < \varepsilon$. Then,

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

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

Recalling now from (40) that $X_{k\Delta}(n)$ grows exponentially on the event $\{W > 0\}$, so $\{X_{k\Delta} \le k\}$ also happens a.s. a finite number of times. As a result, a.s. on the event $\{W > 0\}$, we have $N_{k,\Delta}(n) \le \varepsilon X_{k\Delta}(n)$ for k large enough. We conclude that $\lim_{k\to\infty} e^{-\lambda k\Delta} N_{k,\Delta} = 0$ a.s. on the event $\{W > 0\}$ since ε can be arbitrarily small. This ends the proof of (42) and we obtain (41). We have thus also proved Theorem 2.2 for functions with bounded support.

Step 3: Almost surely convergence—general test function. It suffices to prove the results for functions in $\mathcal{B}_p$ with $p \ge 1$, thanks to the fact $\mathcal{B} = \bigcup_{p>0}\mathcal{B}_p$. For convenience, we consider instead

$$\overline{\mathcal{B}}_p := \Big\{ f : \mathbb{N}_+ \to \mathbb{R}, \sup_{n\ge 1} |f(n)|/n^p < 1 \Big\}, \quad p \ge 1. \tag{45}$$

Indeed, $\overline{\mathcal{B}}_p \subset \mathcal{B}_p$ and for any $f \in \mathcal{B}_p$ there exists $g \in \overline{\mathcal{B}}_p$ and $c \in \mathbb{R}$ such that $f = cg$.

We define the cutoff operator at some level $K \in \mathbb{N}_+$

$$f_{\le K}(n) := f(n)\mathbf{1}_{\{n\le K\}}, \qquad f_{>K}(n) := f(n)\mathbf{1}_{\{n>K\}}. \tag{46}$$

First, using (41), we obtain

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

Second, recall the definition of $[x^p]$ and $[x]$ introduced above (11). Then

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

Combining these estimates and

$$\begin{aligned} &|e^{-\lambda t}\langle X_t, f\rangle - W\langle \pi, f\rangle| \\ &\quad \le |e^{-\lambda t}\langle X_t, f_{>K}\rangle| + |e^{-\lambda t}\langle X_t, f_{\le K}\rangle - W\langle \pi, f_{\le K}\rangle| + |W\langle \pi, [x^p]_{>K}\rangle|, \end{aligned}$$

it yields

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

for any $K \ge 1$. To show that the right-hand side in the above display goes to 0 as K goes to infinity, it suffices to prove:

$$\lim_{K\to\infty} \overline{\lim_{t\to\infty}} e^{-\lambda t}\langle X_t, [x^p]_{>K}\rangle = 0. \tag{47}$$

Since (40) ensures

$$\lim_{k\to\infty} e^{-\lambda k\Delta}\langle X_{k\Delta}, [x^p]_{>K}\rangle = W\langle \pi, [x^p]_{>K}\rangle, \tag{48}$$

we just need to control what happens on 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 $\widetilde{X}_t$ starting at time $k\Delta$ with the same value $\widetilde{X}_{k\Delta} := X_{k\Delta}$; we let the rates of fragmentation and isolation be zero in $\widetilde{X}_t$, while for any cluster at $k\Delta$, the growth events occurring to it on $(k\Delta, (k+1)\Delta)$ are constructed by the common exponential clocks in $\widetilde{X}_t$ and X_t, until it gets isolated or fragmented in the latter. Notice that the isolation

events make negative contribution to $\mathcal{L}([x^p])$, $p \geq 1$, 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

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

The term on the right-hand side is monotone in t and we get

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

As a result, setting

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

it suffices to prove

$$\mathbb{P}\big[\{\text{i.o. } B_k\} \cap \{W > 0\}\big] = 0, \tag{49}$$

to get that $\overline{\lim}_{t\to\infty} e^{-\lambda t}\langle X_t, [x^p]_{>K}\rangle \leq 2\lim_{k\to\infty} e^{-\lambda k\Delta}\langle X_{k\Delta}, [x^p]_{>K}\rangle$ a.s. Together with (48), we can conclude that (47) holds.

To this purpose, we use a truncation technique. Define

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

for $\varepsilon > 0$. We do a split

$$\mathbb{P}\big[\{\text{i.o. } B_k\} \cap \{W > 0\}\big] \leq \mathbb{P}\big[\{\text{i.o. } B_k \cap C_k\}\big] + \mathbb{P}\big[\{\text{i.o. } B_k \cap (C_k)^c\} \cap \{W > 0\}\big]. \tag{50}$$

The second term on the right-hand side has the following upper bound thanks to (40) and dominated convergence theorem:

$$\mathbb{P}\big[W\langle \pi, [x^p]_{>K}\rangle \in (0, 2\varepsilon)\big] + \mathbb{P}\big[W\langle \pi, [x^{2p}]\rangle > 1/(2\varepsilon)\big],$$

which converges to 0 as $\varepsilon \to 0$.

We now deal with the first term on the right-hand side in (50). Defining

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

and using the Markov inequality,

$$\begin{aligned}\mathbb{P}[B_k|\mathcal{F}_{k\Delta}] &= \mathbb{P}\big[\langle \widetilde{X}_{(k+1)\Delta-}, [x^p]_{>K}\rangle - \langle \widetilde{X}_{k\Delta}, [x^p]_{>K}\rangle > \langle \widetilde{X}_{k\Delta}, [x^p]_{>K}\rangle | \mathcal{F}_{k\Delta}\big] \\ &\leq \frac{\operatorname{var}[Z_{k,p,K}|\mathcal{F}_{k\Delta}]}{(\langle \widetilde{X}_{k\Delta}, [x^p]_{>K}\rangle - \mathbb{E}[Z_{k,p,K}|\mathcal{F}_{k\Delta}])^2}.\end{aligned} \tag{51}$$

We need now to evaluate the conditional expectation and variance. We will apply the following lemma whose proof will be provided in the Appendix.

LEMMA 5.1. *For any $k, K \in \mathbb{N}_+$ and $p \geq 1$, we have*

$$\mathbb{E}[Z_{k,p,K}|\mathcal{F}_{k\Delta}] \leq C_\Delta \langle X_{k\Delta}, [x^p]\rangle, \tag{52}$$

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

$$\operatorname{var}[Z_{k,p,K}|\mathcal{F}_{k\Delta}] \leq 2\beta\Delta(4^p p + K^{2p+1})\langle X_{k\Delta}, [x^{2p}]\rangle. \tag{53}$$

Plugging in the estimates (52) and (53) to (51), we obtain

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

We pick $\Delta = \varepsilon^2$ with ε small enough. Then we obtain $\mathbb{P}[B_k|\mathcal{F}_{k\Delta}] \leq Ce^{-\lambda k\Delta}$ conditional on C_k. Adding that $\mathbb{P}[B_k|C_k] = \mathbb{E}[\mathbb{P}[B_k|\mathcal{F}_{k\Delta}]|C_k]$, we obtain $\sum_{k\geq 1}\mathbb{P}[B_k|C_k] < \infty$. By the Borel–Cantelli lemma, $\mathbb{P}[\text{i.o. } B_k \cap C_k] = 0$ for ε small enough. Therefore we have proved the term on the right hand side in (50) is equal to 0. This implies (49) and the proof for general test functions is finished. □

5.3. *Strong law of large numbers for the size process of inactive clusters.* We prove Corollary 1 in this part. A heuristic argument to obtain the asymptotic limit is to use the generator (9) and the convergence of X_t in Theorem 2.2:

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

with $\widetilde{\pi}$ defined in (5). In the sequel, we prove the result, with a suitable set of test functions using in particular martingale analysis.

PROOF OF COROLLARY 1. Let $f \in \mathcal{B}_p$ for some fixed $p > 0$ throughout the proof. The proof can be divided into 3 steps. In Step 1, we control the value $\langle Y_t, f\rangle$. In Step 2 we prove the result with a specific function $f = h/[x]$ which gives us a martingale. In Step 3, we generalise this result to general $f \in \mathcal{B}_p$. Let C_0 be a constant, independent of f, which may change from line to line.

Step 1: L^2 estimate. We will use again the estimation (38). We can check that $F_{g,f}(\mu, \nu) = \langle\nu, f\rangle$ belongs to the domain of the extended generator of $(X_t, Y_t)_{t\geq 0}$ (see (9)) using the same localisation argument as in the proof of Lemma 4.1. So we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\langle Y_t, f\rangle^2] = 2\theta\mathbb{E}[\langle Y_t, f\rangle\langle X_t, [x]f\rangle] + \theta\mathbb{E}[\langle X_t, [x]f^2\rangle] \quad \forall t \geq 0.$$

Using Young's inequality with $\alpha > 0$ (to be chosen 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 lemma to get

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

Combining the L^2 estimate of $\langle X_s, [x]f\rangle$ obtained in (38) and the L^1 estimate of $\langle X_t, [x]f^2\rangle$ in (30), we get

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

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

$$\mathbb{E}[\langle Y_t, f\rangle^2] \leq C'\left(\langle\pi, [x]f\rangle^2 e^{2\lambda t} + \langle\pi, [x]f^2\rangle e^{\lambda t} + \|f\|_p\left(e^{(2\lambda-\sigma)t} + e^{(\lambda-w)t}\right)\right). \tag{54}$$

Step 2: A martingale for Y which tends to 0. We introduce the function

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

We can again check that it belongs to the domain of the extended generator of $(X_t, Y_t)_{t\geq 0}$. It turns out to be a harmonic function: $\mathcal{A}F_{h,g} = 0$. Then we obtain that

$$H_t := \langle X_t, h\rangle - \left(\frac{\lambda}{\theta}\right)\langle Y_t, h/[x]\rangle \tag{55}$$

is a martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. Let us prove that $e^{-\lambda t}H_t$ converges to 0 as $t \to \infty$ and thus Corollary 1 holds for the specific test function $h/[x]$. This vanishing property is due

to the fact that the two parts in H compensate each other. We prove first L^2 convergence using (9):

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

We develop this equation and recognise the generator $\mathcal{L}$ defined in (14). As $\mathcal{L}h = \lambda h$, we get

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

Since h is bounded (see Proposition 4.1), we obtain

$$\mathbb{E}\big[|H_t|^2\big] \le C_0\int_0^t \mathbb{E}\big[\langle X_s, [x]\rangle\big]\,\mathrm{d}s \le C_0 e^{\lambda t}. \tag{56}$$

This implies the L^2 convergence of $e^{-\lambda t}H_t$ to 0 as $t \to \infty$. For the almost sure convergence, we set the step size $\Delta > 0$. As $(H_t)_{t\ge 0}$ is a martingale, for any $\varepsilon > 0$, we combine Markov's inequality, Doob's inequality for H_t and the estimate (56) to obtain that

$$\begin{aligned}
\mathbb{P}\Big[\sup_{t\in[k\Delta,(k+1)\Delta)}\big|e^{-\lambda t}H_t\big| > \varepsilon\Big] &\le \mathbb{P}\Big[e^{-\lambda k\Delta}\sup_{t\in[k\Delta,(k+1)\Delta)}|H_t| > \varepsilon\Big] \\
&\le \varepsilon^{-2}e^{-2\lambda k\Delta}\mathbb{E}\Big[\Big(\sup_{t\in[k\Delta,(k+1)\Delta)}|H_t|\Big)^2\Big] \\
&\le 4\varepsilon^{-2}e^{-2\lambda k\Delta}\mathbb{E}\big[|H_{(k+1)\Delta}|^2\big] \\
&\le C_0\varepsilon^{-2}e^{-\lambda k\Delta}.
\end{aligned}$$

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

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

$$H_t^f := \langle X_t, f\rangle - \left(\frac{\lambda}{\theta}\right)\langle Y_t, f/[x]\rangle = \langle \pi, f\rangle H_t + A_t + B_t, \tag{57}$$

where $A_t = \langle X_t, f - \langle \pi, f\rangle h\rangle$ and $B_t = (\frac{\lambda}{\theta})\langle Y_t, (f - \langle \pi, f\rangle h)/[x]\rangle$. We apply L^2 estimates from (38) and (54) to ensure that

$$e^{-2\lambda t}\mathbb{E}\big[A_t^2 + B_t^2\big] \le C_0\left(1 + \sup_{n\ge 1}\frac{|f(n)|}{n^p}\right)^2\big(e^{-\lambda t} + e^{-\sigma t} + e^{-(\lambda+\sigma/2)t}\big).$$

Here C_0 may depend on p but not on (the specific choice of) f. By the Borel–Cantelli lemma, this implies the convergence of $e^{-\lambda t}H_t^f$ along a subsequence $\{k\Delta\}_{k\geq 1}$ with $\Delta > 0$

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

We get

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

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

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

We obtain then

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

We take $\Delta \searrow 0$ to get

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

A similar argument also works for f^-. We combine these two terms and use $\widetilde{\pi}$ to prove the almost sure convergence in Corollary 1. The L^2 convergence can be done similarly and we skip the details. □

5.4. *Limit on recursive trees.* In this part, we prove the convergence of the empirical measure on clusters.

PROOF OF THEOREM 2.3. We will prove only the convergence on active clusters $\mathscr{X}_t$, as the proof for $\mathscr{Y}_t$ follows in 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 $\mathbf{t} \in \mathscr{T}$, $|f(\mathbf{t})| \leq |\mathbf{t}|^p$ for some $p > 0$.

Step 1: Cut-off argument. We do the following decomposition on the event $\mathscr{X}_t$ being not empty:

$$\frac{1}{|\mathscr{X}_t|}\sum_{\mathcal{C}\in\mathscr{X}_t} f(\mathcal{C}) - \mathbb{E}[f(T_\pi)] = \underbrace{\sum_{n=1}^{\infty} A_t(n)}_{\mathbf{I}} + \underbrace{\sum_{n=1}^{\infty} B_t(n)}_{\mathbf{II}}, \tag{58}$$

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

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

Theorem 2.2 implies the a.s. convergence of the second term **II**:

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

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

$$\left|\sum_{n=1}^{\infty} A_t(n)\right| \le \left|\sum_{n=1}^{K} A_t(n)\right| + 2\left|\sum_{n=K+1}^{\infty} \frac{X_t(n)n^p}{\langle X_t, 1\rangle}\right|. \tag{59}$$

For the first term on the right-hand side in (59), we admit the following convergence (to be proved in Step 2):

$$\forall n \in \mathbb{N}_+, \quad \lim_{t\to\infty} \frac{1}{X_t(n)} \sum_{\mathcal{C}\in\mathscr{X}_t, |\mathcal{C}|=n} f(\mathcal{C}) = \mathbb{E}\big[f(T_n)\big] \quad \text{almost surely on } \{\tau = \infty\}, \tag{60}$$

which can also be seen as a generalised law of large numbers. By Theorem 2.2, for any $n \in \mathbb{N}_+$, $\frac{X_t(n)}{\langle X_t,1\rangle}$ converges a.s. as $t \to \infty$. Then the above display ensures that $\lim_{t\to\infty} |\sum_{n=1}^{K} A_t(n)| = 0$, almost surely on $\{\tau = \infty\}$. The second term on the right-hand side in (59) also converges a.s. due to Theorem 2.2:

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

where $[x^p]_{>K}(n) = n^p \mathbf{1}_{\{n>K\}}$. We put these results back into (58) and obtain that

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

Then we let $K \to \infty$ and prove Theorem 2.3.

Step 2: Convergence for one type. It remains to prove (60). We follow the same spirit as in Step 2 in the proof of Theorem 2.2, with some minor technical differences. We recall $\mathscr{T}_n$ the space of equivalence classes of RRT of size n, and denote by

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

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

$$\forall n \in \mathbb{N}_+, \forall \mathbf{t} \in \mathscr{T}_n, \quad \lim_{t\to\infty} e^{-\lambda t} X_t(\mathbf{t}) = W\frac{\pi(n)}{(n-1)!} \quad \text{a.s.}, \tag{61}$$

and this can be dealt with using the same technique in the proof of Theorem 2.2: we only need to prove the following:

$$\forall n \in \mathbb{N}_+, \forall \mathbf{t} \in \mathscr{T}_n, \quad \underline{\lim}_{t\to\infty} e^{-\lambda t} X_t(\mathbf{t}) \ge W\frac{\pi(n)}{(n-1)!}. \tag{62}$$

This is similar to proving (43). Let $(t_k)_{k\in\mathbb{N}_+}$ be any (random) time sequence such that $\overline{\lim}_{t\to\infty} e^{-\lambda t} X_t(\mathbf{t}) = \lim_{k\to\infty} e^{-\lambda t_k} X_{t_k}(\mathbf{t})$. Then we have

$$\begin{aligned}
\overline{\lim_{t\to\infty}} e^{-\lambda t} X_t(\mathbf{t}) &= \lim_{k\to\infty}\Bigg(\sum_{\mathbf{t}'\in\mathscr{T}_n} e^{-\lambda t_k} X_{t_k}(\mathbf{t}') - \sum_{\mathbf{t}'\in\mathscr{T}_n, \mathbf{t}'\neq\mathbf{t}} e^{-\lambda t_k} X_{t_k}(\mathbf{t}')\Bigg) \\
&\le W\pi(n) - \sum_{\mathbf{t}'\in\mathscr{T}_n, \mathbf{t}'\neq\mathbf{t}} \underline{\lim}_{k\to\infty} e^{-\lambda t_k} X_{t_k}(\mathbf{t}') \\
&\le W\pi(n) - \sum_{\mathbf{t}'\in\mathscr{T}_n, \mathbf{t}'\neq\mathbf{t}} W\frac{\pi(n)}{(n-1)!} = W\frac{\pi(n)}{(n-1)!}.
\end{aligned}$$

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 (62). This equation controls the upper bound of $\overline{\lim}_{t\to\infty} e^{-\lambda t} X_t(\mathbf{t})$. Using also (62), we conclude that (61) is true.

Finally, we prove (62), which requires convergence along discrete time sequences and the control of fluctuation. We calculate the L^2 moment as follows

$$\begin{aligned}
&\mathbb{E}\Big[\Big(e^{-\lambda t}\sum_{\mathcal{C}\in\mathscr{X}_t,|\mathcal{C}|=n}\Big(\mathbf{1}_{\{\mathcal{C}\sim\mathbf{t}\}}-\frac{1}{(n-1)!}\Big)\Big)^2\Big]\\
&\quad= e^{-2\lambda t}\mathbb{E}\Big[\sum_{\mathcal{C}\in\mathscr{X}_t,|\mathcal{C}|=n}\mathbb{E}\Big[\Big(\mathbf{1}_{\{\mathcal{C}\sim\mathbf{t}\}}-\frac{1}{(n-1)!}\Big)^2\Big|\mathcal{F}_t\Big]\Big]\\
&\quad\le e^{-2\lambda t}\mathbb{E}\big[X_t(n)\big]=O\big(e^{-\lambda t}\big)\xrightarrow{t\to\infty}0.
\end{aligned}$$

In the first line, we use the fact that the clusters are i.i.d. RRTs of size n, see Proposition 3.2. In the second line, we use (38). Notice that the L^2 moment decreases at least exponentially, we can thus take a discrete time sequence $\{k\Delta\}_{k\ge1}$ with $\Delta>0$ and use the Borel–Cantelli lemma to obtain that for any $\Delta>0$

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

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 event during $(k\Delta,(k+1)\Delta)$. Then like (44), we have

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

and it suffices to prove $\lim_{k\to\infty}e^{-\lambda k\Delta}N_{k,\Delta}(\mathbf{t})=0$ to conclude (62). We skip the details as it follows exactly the same lines starting from (44) in Step 2 in the proof of Theorem 2.2. □

6. Characterisation of phases and regularity. Recall the three parameters β, θ, γ introduced when describing the GFI process in Section 1. The Malthusian exponent $\lambda=\lambda(\beta,\theta,\gamma)$ allows us to define the phases in terms of the sign of λ (i.e., $\lambda>0,=0,<0$). So far, we have not shown how many phases exist and how they are related. The main objective of this section is to justify the phase diagram shown in Figure 3 and prove Theorem 2.4. The proof strategy is organised as follows: in Section 6.1, we prove some regularity properties about the mapping $(\beta,\theta,\gamma)\mapsto\lambda(\beta,\theta,\gamma)$. In Section 6.2, we justify Figure 3. In Section 6.3, we give a brief proof for the change of monotonicity of $\beta\mapsto\lambda(\beta.\theta,\gamma)$ and finally prove Theorem 2.4.

6.1. *Regularity of Perron's root.* The occurrence of different phases relies on the parameters (β,θ,γ). We will study how the Perron's root $\lambda(\beta,\theta,\gamma)$ depends on the three parameters in the following proposition. Recall that $(\beta,\theta,\gamma)\in\mathbb{R}_+^3$.

PROPOSITION 6.1 (Regularity of Perron's root). *For the Perron's root λ associated to $\mathcal{L}$ in (14), the mapping $(\beta,\theta,\gamma)\mapsto\lambda(\beta,\theta,\gamma)$ satisfies the following properties*:

1. *Monotonicity*: $\lambda(\beta,\theta,\gamma)$ *is increasing on* γ *and decreasing on* θ.
2. *Regime*:

- *if* $\theta\ge\beta$, *or* $\theta\in(0,\beta)$ *and* $0<\gamma<\theta(2^{1-\frac{\theta}{\beta}}-1)^{-1}$, *then* $\lambda<0$;
- *if* $\theta\in(0,\beta)$ *and* $\gamma>\max\{2\theta(\frac{1+\frac{\theta}{\beta}}{1-\frac{\theta}{\beta}}),\frac{3}{2}\theta(1+\frac{\theta}{\beta})\}$, *then* $\lambda>0$.

3. *Homogeneity*: *for any* $\rho > 0$, *we have* $\lambda(\rho\beta, \rho\theta, \rho\gamma) = \rho\lambda(\beta, \theta, \gamma)$.
4. *Continuity*: $\lambda(\beta, \theta, \gamma)$ *is continuous with respect to the three parameters*.

PROOF. *Monotonicity.* If the clusters split faster, then the number of active clusters will increase. If the clusters become inactive faster, then the number of active clusters will decrease. This intuition is easy to prove by providing a coupling argument for each of the parameters. Then we have the monotonicity property.

Regime. To find a regime where $\lambda < 0$, we make a test with $f = [x^p]$ with $p \in (0, 1)$. Jensen's inequality implies that

$$\frac{x^p + y^p}{2} \le \left(\frac{x+y}{2}\right)^p \quad \forall x, y \ge 0.$$

This entails that

$$(j)^p + (n-j)^p - n^p \le \big(2^{1-p} - 1\big)n^p.$$

We apply the above inequality into $\mathcal{L}f$ and get that

$$\begin{aligned} \mathcal{L}f(n) &\le \beta n\big((n+1)^p - n^p\big) - \theta n^{p+1} + \gamma\big(2^{1-p} - 1\big)(n-1)n^p \\ &\le \big(p\beta - \big(2^{1-p} - 1\big)\gamma\big)f(n) + \big(\big(2^{1-p} - 1\big)\gamma - \theta\big)nf(n) \\ &= \big(p\beta - \theta + \gamma\big(2^{1-p} - 1 - \theta\big)(n-1)\big)f(n). \end{aligned} \tag{63}$$

Now, we distinguish two situations. If $\theta \in (0, \beta)$, the condition $\gamma < \theta/(2^{1-\frac{\theta}{\beta}} - 1)$ allows us to find some $p < \theta/\beta$ such that $\mathcal{L}f(n) \le (p\beta - \theta)f(n)$ for all n. Consequently $M_t f \le \exp((p\beta - \theta)t)f$. Then Proposition 4.1 entails $\lambda < 0$. If $\theta \ge \beta$, it suffices to choose some p close to 1 such that $\gamma(2^{1-p} - 1) - \theta < 0$ and we get also $\lambda < 0$.

To find a regime where $\lambda > 0$, we consider the test function

$$f(n) = \mathbf{1}_{\{n=1\}} + \kappa \mathbf{1}_{\{n \ge 2\}},$$

with constant $\kappa \in (0, 2)$ to be fixed. Recalling (14), it is clear that

$$\begin{aligned} &n = 1, \quad \mathcal{L}f(1) = \beta(\kappa - 1) - \theta, \\ &n = 2, \quad \mathcal{L}f(2) = -2\theta\kappa + \gamma(2 - \kappa), \\ &n \ge 3, \quad \mathcal{L}f(n) \ge -n\theta\kappa + \gamma(n-1). \end{aligned}$$

For

$$\gamma > \max\left\{2\theta\left(\frac{1 + \frac{\theta}{\beta}}{1 - \frac{\theta}{\beta}}\right), \frac{3}{2}\theta\left(1 + \frac{\theta}{\beta}\right)\right\},$$

we get $\mathcal{L}f(n) > 0$ for any $n \ge 1$. Taking $\kappa = \frac{1}{2}((1 + \frac{\theta}{\beta}) + (\frac{2\gamma}{3\theta}))$, we obtain $\mathcal{L}f(n) \ge Cf(n)$ for all n, with $C > 0$. It implies $\lambda > 0$.

Homogeneity. This is due to the fact that the generator $\mathcal{L}$ is linear with respect to the three parameters. Let $\widetilde{\mathcal{L}}$ and $(\widetilde{M}_t)_{t \ge 0}$ be respectively the generator and semigroup associated to $(\rho\beta, \rho\theta, \rho\gamma)$, while we denote by λ, h, π the eigenelements associated to $\mathcal{L}$. Then by the definition of $(X_t)_{t \ge 0}$, it is clear that $\widetilde{\mathcal{L}} = \rho\mathcal{L}$, so $\widetilde{M}_t = M_{\rho t}$. This also implies

$$\pi\widetilde{\mathcal{L}} = \rho\pi\mathcal{L} = \rho\lambda\pi, \qquad \widetilde{\mathcal{L}}h = \rho\mathcal{L}h = \rho\lambda h,$$

so $\rho\lambda$, h, π are eigenelements associated to $\widetilde{\mathcal{L}}$. Then by Proposition 4.1, $\rho\lambda$ is Perron's root for $\widetilde{\mathcal{L}} = \rho\mathcal{L}$.

Continuity. We will use three steps to prove it.

Step 1: $\lambda(\beta,\theta,\gamma)$ is locally bounded. For any $\varepsilon>0$ such that

$$\frac{1}{1+\varepsilon}<\frac{\beta_0}{\beta}<1+\varepsilon,\qquad \frac{1}{1+\varepsilon}<\frac{\theta_0}{\theta}<1+\varepsilon,\qquad \frac{1}{1+\varepsilon}<\frac{\gamma_0}{\gamma}<1+\varepsilon,$$

we have

$$\begin{aligned}&\big|\lambda(\beta_0,\theta_0,\gamma_0)\big|\\&\quad\le(1+\varepsilon)\max\big\{\big|\lambda\big(\beta,(1+\varepsilon)^2\theta,(1+\varepsilon)^{-2}\gamma\big)\big|,\big|\lambda\big(\beta,(1+\varepsilon)^{-2}\theta,(1+\varepsilon)^2\gamma\big)\big|\big\}.\end{aligned}$$

The above result is due to the homogeneity of λ on all parameters and the monotonicity of λ on θ and γ. Then the local boundedness of λ is proved.

Step 2: $\langle\pi(\beta,\theta,\gamma),[x]\rangle$ is locally bounded. We prove it by contradiction. Assume that $(\beta_n,\theta_n,\gamma_n)$ converges to $(\beta_0,\theta_0,\gamma_0)$. Let $(\beta,\theta,\gamma)=(\beta_0,\theta_0/2,2\gamma_0)$. Let M_t, π, h, λ be associated to (β,θ,γ) and $\widetilde{M}_t$, $\widetilde{\pi}$, $\widetilde{h}$, $\widetilde{\lambda}$ to $(\beta_n,\theta_n,\gamma_n)$. We assume that $\langle\widetilde{\pi},[x]\rangle$ converges to infinity as $n\to\infty$. Due to homogeneity of λ on its parameters, we can assume $\beta_n=\beta_0$ without loss of generality.

By the assumption above, for any $K>0$, let $n_K>0$ such that for any $n\ge n_K$, we have $\langle\widetilde{\pi},[x]\rangle>K$ and $\theta_n>3\theta_0/4$ and $\gamma_n<7\gamma_0/4$. By Proposition 4.1, there exist $0<c<C$ such that $c\le h(n)\le C$ for any $n\ge1$. Then using (iii) in Lemma 4.1, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{M}_t h&=\widetilde{M}_t(\widetilde{\mathcal{L}}h)\\&\le\widetilde{M}_t(Lh)-c(\theta_n-\theta_0/2)\widetilde{M}_t\big([x]\big)+2c(\gamma_n-2\gamma_0)\widetilde{M}_t\big([x]\big)\\&=\lambda\widetilde{M}_t h-c(\theta_n-\theta_0/2)\widetilde{M}_t\big([x]\big)+2c(\gamma_n-2\gamma_0)\widetilde{M}_t\big([x]\big).\end{aligned}\tag{64}$$

Using again Proposition 4.1, as $t\to\infty$,

$$\widetilde{M}_t h(n)\sim\widetilde{h}\langle\widetilde{\pi},h\rangle e^{\widetilde{\lambda}t},\qquad \widetilde{M}_t\big([x]\big)(n)\sim\widetilde{h}\langle\widetilde{\pi},[x]\rangle e^{\widetilde{\lambda}t}.\tag{65}$$

Then for any fixed $n\ge n_K$, and t large enough, we have

$$\widetilde{M}_t\big([x]\big)(n)\ge\frac{K}{2C}\widetilde{M}_t h(n).$$

Then we obtain that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{M}_t h(n)&\le\lambda\widetilde{M}_t h(n)-c(\theta_n-\theta_0/2)\widetilde{M}_t\big([x]\big)(n)+2c(\gamma_n-2\gamma_0)\widetilde{M}_t\big([x]\big)(n)\\&\le\left(\lambda-cK\frac{\theta_0+2\gamma_0}{8C}\right)\widetilde{M}_t h(n).\end{aligned}\tag{66}$$

The above display implies that

$$\widetilde{\lambda}\le\lambda-cK\frac{\theta_0+2\gamma_0}{8C}.$$

As K can be arbitrarily large, we obtain that $\widetilde{\lambda}\to-\infty$ as $n\to\infty$, which is contradictory to the fact that λ is locally bounded as proved in Step 1. Therefore we conclude that $\langle\pi(\beta,\theta,\gamma),[x]\rangle$ is locally bounded.

Step 3: λ is continuous with respect to the three parameters. The approach is very similar to that in Step 2. Assume that $(\beta_n,\theta_n,\gamma_n)$ converges to (β,θ,γ). Let M_t,π,h,λ be associated to (β,θ,γ) and $\widetilde{M}_t,\widetilde{\pi},\widetilde{h},\widetilde{\lambda}$ to $(\beta_n,\theta_n,\gamma_n)$. We use again that there exist $0<c<C$ such that $c\le h(n)\le C$ for any $n\ge1$. Due to homogeneity, we can assume that $\gamma_n=\gamma$. We prove next $\lim_{n\to\infty}\widetilde{\lambda}=\lambda$.

By the assumption above, for any $K > 0$, let $n_K > 0$ such that for any $n \geq n_K$, we have

$$|\beta_n - \beta| \leq \beta/K, \qquad |\theta_n - \theta| \leq \theta/K.$$

Since $\langle \widetilde{\pi}, [x] \rangle$ is locally bounded, using (65) and $c \leq h \leq C$, there exists $l = l(\beta, \theta, \gamma) > 1$ such that for fixed $n \geq n_K$ and t large enough, we have

$$\widetilde{M}_t h(n)/l \leq \widetilde{M}_t\big([x]\big)(n) \leq l \widetilde{M}_t h(n).$$

Then similar to (66), we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \widetilde{M}_t(h)(n) \leq \left(\lambda + Cl\frac{\beta + \theta}{K}\right) \widetilde{M}_t(h)(n). \tag{67}$$

Since K can be arbitrarily large, we obtain that $\overline{\lim}_{n\to\infty} \widetilde{\lambda} \leq \lambda$. The other direction can be proved similarly. Then the proof for continuity is finished. □

6.2. *Existence of phases.* The main objective of this part is to prove the following result, which describes the phases and justifies Figure 3.

PROPOSITION 6.2 (Classification of phases). *The regime of the three phases of the GFI process depends on the parameters (β, θ, γ) in the following way*:

1. *Critical phase*: *the regime is a critical surface defined as*

$$\big\{\lambda(\beta, \theta, \gamma) = 0\big\} = \big\{(\beta, \theta, \gamma) \in \mathbb{R}_+^3 | 0 < \theta < \beta, \gamma = \gamma_c(\beta, \theta)\big\}.$$

Here $\gamma_c(\beta, \theta)$ is a function, such that for fixed β, the mapping $\theta \mapsto \gamma_c(\beta, \theta)$ is strictly increasing, continuous and satisfies

$$\lim_{\theta \searrow 0} \gamma_c(\beta, \theta) = 0, \qquad \lim_{\theta \nearrow \beta} \gamma_c(\beta, \theta) = \infty.$$

2. *Subcritical phase*: *the regime stays below the critical surface*

$$\begin{aligned}\big\{\lambda(\beta, \theta, \gamma) < 0\big\} &= \big\{(\beta, \theta, \gamma) \in \mathbb{R}_+^3 | \theta \geq \beta\big\} \\ &\quad \cup \big\{(\beta, \theta, \gamma) \in \mathbb{R}_+^3 | 0 < \theta < \beta, 0 < \gamma < \gamma_c(\beta, \theta)\big\}.\end{aligned}$$

3. *Supercritical phase*: *the regimes stays above the critical surface*

$$\big\{\lambda(\beta, \theta, \gamma) > 0\big\} = \big\{(\beta, \theta, \gamma) \in \mathbb{R}_+^3 | 0 < \theta < \beta, \gamma > \gamma_c(\beta, \theta)\big\}.$$

The key step in the proof of Proposition 6.2 relies on some strict increment estimates which are proved in the following two lemmas.

LEMMA 6.1. *Perron's root λ associated to the generator $\mathcal{L}$ in (14) satisfies the following estimates*:

$$\begin{aligned}\forall \delta > 0, \quad & \lambda(\beta, \theta + \delta, \gamma) \leq \lambda(\beta, \theta, \gamma) - \delta, \\ \forall \delta \in (0, \theta), \quad & \lambda(\beta, \theta - \delta, \gamma) \geq \lambda(\beta, \theta, \gamma) + \delta.\end{aligned} \tag{68}$$

PROOF. Let $\widetilde{\mathcal{L}}$ and $(\widetilde{M}_t)_{t\geq 0}$ be the generator and semigroup associated to $(\beta, \theta + \delta, \gamma)$, let h be the eigenvector associated to $\mathcal{L}$ of parameters (β, θ, γ). Using a similar development like (64) and using also $[x] \geq 1$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} \widetilde{M}_t h = \widetilde{M}_t(\mathcal{L}h) - \delta \widetilde{M}_t\big([x]h\big) = \lambda \widetilde{M}_t h - \delta \widetilde{M}_t\big([x]h\big) \leq (\lambda - \delta) \widetilde{M}_t h.$$

Using Grönwall's inequality, the above display implies that $\widetilde{M}_t h \leq e^{(\lambda-\delta)t} h$, which proves the first statement in (68). For the second, the proof is similar and we skip it. □

LEMMA 6.2. *Let h be the eigenvector of Perron's root associated to the generator $\mathcal{L}$ in (14). Then in the critical phase $\lambda = 0$, h is subadditive, that is,*

$$\forall n, m \in \mathbb{N}_+, \quad h(m+n) \le h(m) + h(n). \tag{69}$$

PROOF. The proof relies on the probabilistic representation of h. Let N_t^n be the number of clusters at time t with the initial state being a single cluster of size n. Then we have

$$\mathbb{E}\big[N_t^n\big] = \langle \delta_n M_t, 1\rangle.$$

Using (30), in the case $\lambda = 0$, we have

$$\lim_{t\to\infty} \mathbb{E}\big[N_t^n\big] = \lim_{t\to\infty} \langle \delta_n M_t, 1\rangle = h(n)\langle \pi, 1\rangle. \tag{70}$$

Then we think of a coupling as follows: consider two processes of $(\mathscr{X}_t)_{t\ge 0}$, one with initial state being a single RRT of size $(m+n)$, and the other one with the same initial state but has a uniformly selected edge removed which results in two smaller clusters of sizes respectively m and n. By Proposition 3.1, these two clusters are independent RRTs with sizes respectively m, n. Assume that the two processes have the same random events in their life times, except that a splitting may happen in the particular edge that exists in the first process and does not exist in the second process. In this coupling, it is easy to see that the number of clusters at any time $t > 0$ in the first process will not be larger than the number in the second process. Since the clusters evolve independently, we have

$$\mathbb{E}\big[N_t^{n+m}\big] \le \mathbb{E}\big[N_t^n\big] + \mathbb{E}\big[N_t^m\big]. \tag{71}$$

Combining (70) and (71), we obtain the desired result (69). □

REMARK 4. In the critical case, since $\mathcal{L}h = 0h = 0$, we solve $\mathcal{L}h(1) = 0$ to obtain that $h(2) = (1 + \frac{\theta}{\beta})h(1)$. The criticality also implies $\beta > \theta$. Therefore

$$h(1) + h(1) - h(2) = \left(1 - \frac{\theta}{\beta}\right)h(1) > 0.$$

So the "=" in (69) is not always established.

PROOF OF PROPOSITION 6.2. We divide the proof into 4 steps. Steps 1–3 are about the critical phase, and the Step 4 proves the regime of the other phases.

Step 1: Unique critical point. Fix $\beta > 0$ and $0 < \theta < \beta$. Using the regimes discovered for subcritical and critical phases in Proposition 6.1 and also the continuity and monotonicity property of λ therein, we know that $\{\gamma : \lambda(\beta, \theta, \gamma) = 0\}$ is a nonempty closed interval. We aim to prove that this interval is a single point by contradiction. Suppose that there exist $\gamma > 0$, $\delta > 0$ such that $\lambda(\beta, \theta, \gamma) = \lambda(\beta, \theta, \gamma + \delta) = 0$. Let $\widetilde{\mathcal{L}}$ and $(\widetilde{M}_t)_{t\ge 0}$ be the generator and semigroup associated to $(\beta, \theta, \gamma + \delta)$, while we denote by $\mathcal{L}$ the generator and h the eigenvector for Perron's root associated to (β, θ, γ). We also define the function G

$$G(n) := \sum_{j=1}^{n-1} \frac{n}{j(j+1)}\big(h(j) + h(n-j) - h(n)\big).$$

Since h is subadditive (Lemma 6.2) and nonnegative (Proposition 4.1), we have

$$G(n) \le \sum_{j=1}^{n-1} \frac{n}{j(j+1)}\big(jh(1) + (n-j)h(1)\big) \le n^2 h(1).$$

Moreover by Remark 4, G is a nonzero positive function. Then by a similar argument like (64) and using $\mathcal{L}h = \lambda h = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{M}_t h = \widetilde{M}_t(\widetilde{\mathcal{L}}h) = \widetilde{M}_t(\mathcal{L}h + \delta G) = \widetilde{M}_t(\mathcal{L}h) + \delta \widetilde{M}_t G = \delta \widetilde{M}_t G. \tag{72}$$

By Proposition 4.1, there exist Perron's eigenelements $\widetilde{h}$, $\widetilde{\pi}$, and also a constant $\widetilde{\omega} > 0$, corresponding to $\widetilde{\mathcal{L}}$, and the estimate (30) holds with the corresponding terms. Then we apply (30) and $G(n) \le n^2 h(1)$ to obtain

$$\widetilde{M}_t G \ge \langle \widetilde{\pi}, G\rangle \widetilde{h} - Ce^{-\omega t}.$$

Here $C > 0$ is a constant that does not depend on t. We plug in the above display to (72) and obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{M}_t h \ge \delta\big(\langle \widetilde{\pi}, G\rangle \widetilde{h} - Ce^{-\omega t}\big).$$

If $\lambda(\beta, \theta, \gamma + \delta) = 0$, then by (30), $\widetilde{M}_t h$, as a vector, converges to $\langle \widetilde{\pi}, h\rangle \widetilde{h}$ as $t \to \infty$ for every element. However since G is a nonzero positive function (which implies $\langle \widetilde{\pi}, G\rangle > 0$), the above display shows that $\widetilde{M}_t h$ increases at least linearly to infinity as $t \to \infty$ for every element. This is a contradiction, so by the monotonicity of λ in γ, we have $\lambda(\beta, \theta, \gamma + \delta) > 0$. We conclude that for fixed $0 < \theta < \beta$, there exists only one γ which makes $\lambda(\beta, \theta, \gamma) = 0$. We denote it by $\gamma_c = \gamma_c(\beta, \theta)$.

Step 2: Strictly increasing and continuous function. Using Lemma 6.1 (first inequality), we have

$$\lambda\big(\beta, \theta + \delta, \gamma_c(\beta, \theta)\big) \le \lambda\big(\beta, \theta, \gamma_c(\beta, \theta)\big) - \delta = -\delta.$$

Since $\lambda(\beta, \theta, \gamma)$ is increasing with respect to γ, we obtain that $\gamma_c(\beta, \theta + \delta) > \gamma_c(\beta, \theta)$.

Next we prove the continuity of $\theta \mapsto \gamma_c(\beta, \theta)$. By the monotonicity of λ in γ, we have for any $\varepsilon > 0$, $\lambda(\beta, \theta, \gamma_c(\beta, \theta) + \varepsilon) > 0$. Then by the continuity of λ on θ, there exists some $\delta_0 > 0$ such that for all $\delta \in (0, \delta_0)$, $\lambda(\beta, \theta + \delta, \gamma_c(\beta, \theta) + \varepsilon) > 0$. This implies, using the uniqueness of γ_c and monotonicity of λ on γ, that for any $\delta \in (0, \delta_0)$

$$\gamma_c(\beta, \theta + \delta) < \gamma_c(\beta, \theta) + \varepsilon,$$

which proves the right continuity of $\gamma_c(\beta, \theta)$ on θ. The left continuity can be proved exactly in the same way.

Step 3: Asymptotic behaviour. By the monotonicity property of λ on γ and regimes for subcriticality and supercriticality given in Proposition 6.1, we have:

$$\theta\big(2^{1-\frac{\theta}{\beta}} - 1\big)^{-1} \le \gamma_c(\beta, \theta) \le \max\left\{2\theta\left(\frac{1 + \frac{\theta}{\beta}}{1 - \frac{\theta}{\beta}}\right), \frac{3}{2}\theta\left(1 + \frac{\theta}{\beta}\right)\right\}. \tag{73}$$

Then it suffices to let $\theta \searrow 0$ and $\theta \nearrow \beta$ to prove the asymptotic limits.

Step 4: Regime. Once we have proved that the critical phase is the single point for β, θ fixed in an admissible domain, we apply the monotonicity about γ in Proposition 6.1 and justify the regime for other phases. □

6.3. *Monotonicity of h and λ.* The dependence of λ on θ, γ has been discussed in the previous subsections, while the dependence on β is more complicated. Indeed, in terms of number of active clusters, isolation and fragmentation have the opposite effect and increasing the cluster size accelerate both. This problem is closely related to the monotonicity of the eigenfunction h. The main result in this part is the following one.

PROPOSITION 6.3. *The monotonicity of the mappings $h(.,\theta,\gamma)$ and $\lambda(.,\theta,\gamma)$ depends on the values of γ,θ:*

- *if $\gamma>\theta$, the two functions are increasing*;
- *if $\gamma=\theta$, the two functions are constants*;
- *if $\gamma<\theta$, the two functions are decreasing.*

The case $\gamma=\theta$ is special: using (14), we find that $h\equiv 1$ is an eigenvector and $\lambda=-\theta$. The other cases are more delicate. Roughly, increasing β makes the infection progress faster and accelerates the process. If $\gamma>\theta$, this benefits the fragmentation more and thus makes the number of clusters and λ increase. Otherwise, isolation is more impacted and the converse happens.

This effect of acceleration due to the increase of β can be understood via the following simplified problem. Let us consider a linear birth and death process $(S_t)_{t\geq 0}$ starting from a single particle: that is, each particle (representing a cluster) lives during an exponential time of parameter $1/\rho$ and is replaced by two particles (resp. zero) with probability $p_2=\gamma/(\theta+\gamma)$ (resp. $p_0=\theta/(\theta+\gamma)$), corresponding respectively to fragmentation and isolation. This birth and death process $(S_t)_{t\geq 0}$ satisfies $\mathbb{E}[S_t]=\exp(\lambda_{\mathrm{BD}}t)$ with

$$\lambda_{\mathrm{BD}}=\rho(p_2-p_0)=\rho\left(\frac{\gamma-\theta}{\gamma+\theta}\right).$$

The monotonicity of λ_{BD} in function of ρ thus depends on the sign of $(\gamma-\theta)$, that is, on the fact that the process is subcritical or supercritical.

Such a dichotomy also exists in our model, and will be clear in a modified process, which is close to our original GFI process and the Malthusian coefficients are explicitly linked. This modification yields a simpler process, since the probability for a cluster to be isolated or fragmented does not depend on its size any longer and $\gamma=\theta$ provides the critical case for this latter. This key property will allow us to make a coupling for the modified processes with different β and then verify the monotonicities.

We divide the proof into three steps. First, we introduce a *modified GFI process*, by slightly changing the isolation term. Second, we obtain coupling results for times and sizes of our modified GFI process. Finally, we use this coupling to prove Proposition 6.3. We skip some technical details in the second and the third step, which can be found in the Appendix B of the long version [9] of this paper on arXiv.

6.3.1. *A modified GFI process.* In this part, we introduce our modified GFI process. Its dynamic of growth and fragmentation is the same as the original one. For the dynamic of isolation, we let each edge (instead of each vertex) have rate θ to be detected independently and then the cluster is isolated. This minor modification of the isolation rate does not change the splitting property, thus the study of the modified GFI process can also be reduced to its size process $(\overline{X}_t,\overline{Y}_t)_{t\geq 0}$. For the parameters (β,θ,γ), the associated generator $\overline{\mathcal{L}}$ of the first moment is defined as

$$\begin{aligned}\overline{\mathcal{L}}f(n)&=\beta n\big(f(n+1)-f(n)\big)-\theta(n-1)f(n)\\&\quad+\sum_{j=1}^{n-1}\frac{\gamma n}{j(j+1)}\big(f(j)+f(n-j)-f(n)\big).\end{aligned}\tag{74}$$

Compared to the generator $\mathcal{L}$ defined in (14) with the same parameters, the isolation rate is $\theta(n-1)$ instead of θn for a cluster of size n, and it is clear that we have

$$\overline{\mathcal{L}}=\mathcal{L}-\theta\mathrm{Id}.\tag{75}$$

Therefore, Lemma 4.2 also applies to $\overline{\mathcal{L}}$. The couple (π, h) associated to $\mathcal{L}$ is also the eigenelement of $\overline{\mathcal{L}}$, with the Perron's root / Malthusian exponent

$$\overline{\lambda}h = \overline{\mathcal{L}}h = (\lambda - \theta)h \quad \Longrightarrow \quad \overline{\lambda} = \lambda - \theta. \tag{76}$$

For fixed $\gamma, \theta > 0$, the mapping $\beta \mapsto \lambda(\beta, \theta, \gamma)$ shares the same monotonicity with $\beta \mapsto \overline{\lambda}(\beta, \theta, \gamma)$, so it suffices to study the the latter one.

The modified GFI process provides some advantages. We observe that for a cluster of size n, the fragmentation rate and the isolation rate are respectively $\gamma(n-1)$ and $\theta(n-1)$. So the probability that isolation happens before fragmentation (and conversely) does not depend on its size and is equal to $\theta/(\gamma+\theta)$ (resp. $\gamma/(\gamma+\theta)$). This nice property has inspired the idea of coupling in the next step.

We finish this part by proving the following lemma.

LEMMA 6.3. *The modified GFI process has the following criteria of phases*: *it is supercritical* ($\overline{\lambda} > 0$) *when* $\gamma > \theta$, *critical* ($\overline{\lambda} = 0$) *when* $\gamma = \theta$ *and subcritical* $\overline{\lambda} < 0$ *when* $\gamma < \theta$.

PROOF. For the original GFI process, we know that the case $\gamma = \theta$ gives us $\lambda = -\theta$. Moreover, the monotonicity of the function $\gamma \mapsto \lambda(\beta, \theta, \gamma)$ in Lemma 6.1 tells us

$$\gamma > \theta \quad \Longrightarrow \quad \lambda > -\theta, \qquad \gamma < \theta \quad \Longrightarrow \quad \lambda < -\theta.$$

Combing this with (75) proves Lemma 6.3. □

6.3.2. *Monotone coupling for the modified GFI processes*. Let us fix the rates θ and γ, and consider an initial cluster of size n in the modified GFI process with the infection rate β. This cluster may grow by infection for a random time τ_n^β, called *lifetime*, when (independently of this time) it makes fragmentation with probability $\gamma/(\gamma+\theta)$ or isolation with probability $\theta/(\gamma+\theta)$. When it splits, it leaves two child clusters whose sizes are distributed as the random variable $(Z_{n,1}^\beta, Z_{n,2}^\beta)$ following Proposition 3.1. That is, $Z_{n,1}^\beta$ is the size of the first child, and $Z_{n,2}^\beta$ is for the second child. Here we write $(Z_{n,1}^\beta, Z_{n,2}^\beta) = (0,0)$ by convention if this cluster is isolated.

We say there is a coupling for two random variables A and B if we can find A' distributed as A, B' distributed as B, and A', B' are defined in a common probability space. We sometimes skip the superscripts for these coupled random variables when the context of coupling is clear. Then we have the following lemma for $(Z_{n,1}^\beta, Z_{n,2}^\beta)$ and τ_n^β defined above.

LEMMA 6.4. *For any* $1 \le n \le n'$ *and* $0 < \beta \le \beta'$, *there exists a coupling for* $(\tau_n^\beta, Z_{n,1}^\beta, Z_{n,2}^\beta)$ *and* $(\tau_{n'}^{\beta'}, Z_{n',1}^{\beta'}, Z_{n',2}^{\beta'})$ *such that*

$$\tau_n^\beta \ge \tau_{n'}^{\beta'}, \qquad Z_{n,1}^\beta \le Z_{n',1}^{\beta'}, \qquad Z_{n,2}^\beta \le Z_{n',2}^{\beta'} \quad a.s.,$$

and in this coupling the two clusters either both reach fragmentation or both reach isolation.

This lemma implies that the larger cluster size or infection rate will result in shorter lifetime and larger sizes for child clusters. Concerning the second statement, since the probability of either event is independent of the cluster size, we can use a common Bernoulli random variable to determine whether they split or are isolated in both processes. For (6.4), it is also natural as larger cluster size or infection rate speed up the growth events.

We now consider two modified GFI processes with common θ, γ but different infection rates $\beta \le \beta'$ and initial cluster sizes $n_0 \le n_0'$. We denote the active cluster processes by $(\overline{\mathscr{X}}^{\beta}_{n_0,t})_{t\ge 0}$, $(\overline{\mathscr{X}}^{\beta'}_{n_0',t})_{t\ge 0}$ and their size processes by $(\overline{X}^{\beta}_{n_0,t})_{t\ge 0}$ and $(\overline{X}^{\beta'}_{n_0',t})_{t\ge 0}$.

We can construct a coupling of the two modified GFI processes in a common probability space as follows. Notice that Lemma 6.4 allows us to couple the lifetimes and cluster sizes of any two clusters. In particular, the lemma tells us that either fragmentation or isolation occurs to both of them. We can continue to apply the coupling to their first child clusters and second child clusters if the fragmentation takes place. If we start from the initial clusters of both process, we can couple the whole processes in a common probability space denoted by $(\Omega, \mathcal{G}, \mathbb{P}^{\beta,\beta'}_{n_0,n_0'})$. In particular, we can use the UHN labelling to describe the one-to-one mapping or coupling between clusters from the two processes. For instance, the label 12, if it exists, maps the second child of the first child of the initial cluster from one process to another.

In this coupling, due to Lemma 6.4, any active cluster in $(\overline{\mathscr{X}}^{\beta}_{n_0,t})_{t\ge 0}$ has a longer lifetime and smaller child cluster sizes (if they exist) compared to the corresponding cluster in $(\overline{\mathscr{X}}^{\beta'}_{n_0',t})_{t\ge 0}$, almost surely.

6.3.3. *Proof of Proposition* 6.3. Let us prove the monotonicity result on the modified GFI, which will immediately give the result on the original process using the link (76) between the maximal eigenvalues.

To compare the Malthusian exponent and the eigenfunction, it suffices to compare $\mathbb{E}[\langle \overline{X}^{\beta}_{n_0,T}, 1\rangle]$ and $\mathbb{E}[\langle \overline{X}^{\beta'}_{n_0',T}, 1\rangle]$ for a large T thanks to the Malthusian behaviour in (29).

We compare them in the common probability space $(\Omega, \mathcal{G}, \mathbb{P}^{\beta,\beta'}_{n_0,n_0'})$ constructed previously. It can be done using a stopping line as follows. Let us write $\overline{\mathcal{U}}^{\beta}_{T}$ the set of UHN labels of the active clusters in $\overline{\mathscr{X}}^{\beta}_{n,T}$. The corresponding clusters in $(\overline{\mathscr{X}}^{\beta'}_{n_0',t})_{t\ge 0}$ labelled by $\overline{\mathcal{U}}^{\beta}_{T}$ are active or have been active before time T due to the coupling. Conversely, the active clusters in $\overline{\mathscr{X}}^{\beta'}_{n_0',T}$ are issued from active clusters whose labels are in $\overline{\mathcal{U}}^{\beta}_{T}$, which allows us to use the branching property (see Figure 7):

$$\langle \overline{X}^{\beta'}_{n_0',T}, 1\rangle = \sum_{u\in\overline{\mathcal{U}}^{\beta}_{T}} N^{\beta'}_{T}(u).$$

Here $N^{\beta'}_{T}(u)$ is the number of active clusters in $\overline{\mathscr{X}}^{\beta'}_{n_0',T}$ that are issued from the cluster labelled by u (either its descendants, or itself if it is still alive up to T) in $(\overline{\mathscr{X}}^{\beta'}_{n_0',t})_{0<t\le T}$, and we also notice that $\langle \overline{X}^{\beta}_{n_0,T}, 1\rangle = \sum_{u\in\overline{\mathcal{U}}^{\beta}_{T}} 1$.

The comparison of $\langle \overline{X}^{\beta'}_{n_0',T}, 1\rangle$ and $\langle \overline{X}^{\beta}_{n_0,T}, 1\rangle$ is reduced to that between the mean of $N^{\beta'}_{T}(u)$ and 1. For any active cluster labelled by u that has evolved to time T in $(\overline{\mathscr{X}}^{\beta}_{n_0,t})_{t\ge 0}$, the coupled evolution of the corresponding active cluster in $(\overline{\mathscr{X}}^{\beta'}_{n_0',t})_{t\ge 0}$ is at a random time $\tau_T(u) < T$. Thus $N^{\beta'}_{T}(u)$ is the number of clusters issued from the cluster labelled u in $(\overline{\mathscr{X}}^{\beta'}_{n_0',t})_{t\ge 0}$ that evolves for the extra time $(T - \tau_T(u))$. We recall that each active cluster either fragments or be isolated with fixed probabilities whose difference is $(\gamma-\theta)/(\gamma+\theta)$. Then, if $\gamma > \theta$ (resp. $\gamma < \theta$), $N^{\beta'}_{T}(u)$ as the mean number of clusters issued from the cluster labelled by u during the extra time $(T - \tau_T(u))$ is larger than 1 (resp. smaller than 1). This finishes the proof of Proposition 6.3.

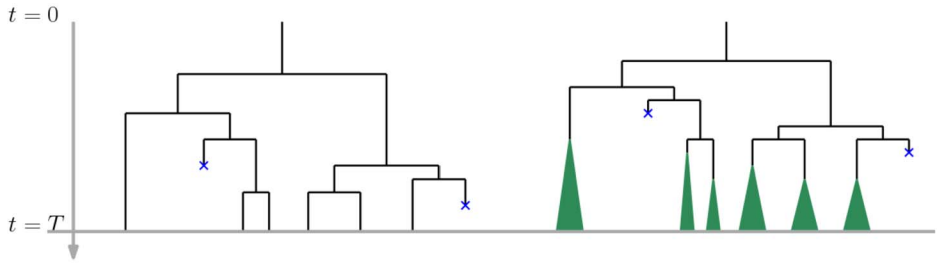


FIG. 7. *An illustration of the monotone coupling argument and the proof of Proposition 6.3. The blue crosses represent isolation. The figure on the left is for* $(\overline{\mathscr{X}}^{\beta}_{n_0,t})_{0\le t\le T}$, *and the one on the right is for* $(\overline{\mathscr{X}}^{\beta'}_{n'_0,t})_{0\le t\le T}$. *With a monotone coupling, the evolution on the right is quicker than the one on the left, so each cluster at time T on the left can find its coupled one on the right at a random moment before T. The rest of the evolution on the right (the green cones) is not coupled with the left up to time T, thus the monotonicity depends on whether the extra time will make the population grow or diminish. That is, the monotonicity depends on the phase (supercricial, critical and subcritical) that the population belongs to.*

6.3.4. *Proof of Theorem 2.4.* We combine Proposition 6.1, Proposition 6.2 and Proposition 6.3 to obtain Theorem 2.4.

7. Examples and simulations. In this part, we give some examples about how our model and theorems can be applied to perform the numerical simulation of epidemics. The key quantity of our model is the Malthusian exponent λ. It dictates whether the infection will grow or vanish, and moreover serves as the growth or decay rate. We described how λ depends on (β, θ, γ) in Proposition 6.1 and 6.3. However, the function $\lambda(\beta, \theta, \gamma)$ is not explicit, so it is useful to investigate its values numerically. We use the finite difference method applied to (13) to approximate M_t and then use (29) to approximate λ.

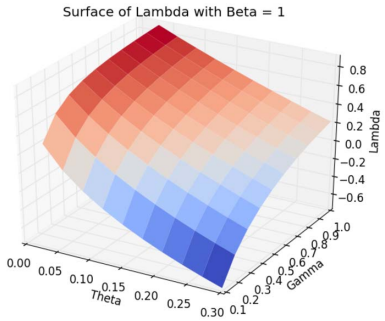
We simulated λ for different values of the three parameters in Figure 8. Figure 8a gives a surface of the function λ with $\beta = 1$, which confirms the monotonicity of λ on θ and γ (see also Figure 8b and Figure 8c). Thus to reduce infection, we should conduct more contact tracing (increase θ) and avoid loss of contact information (decrease γ). These two approaches are natural and intuitive.

Figure 8d confirms the monotonicity of λ on β, see Proposition 6.3. When $\theta = 0.03, 0.06, 0.09$ (thus $\theta < \gamma$), λ is increasing on β; when $\theta = 0.12, 0.15$ (thus $\theta > \gamma$), λ is a decreasing function of β.

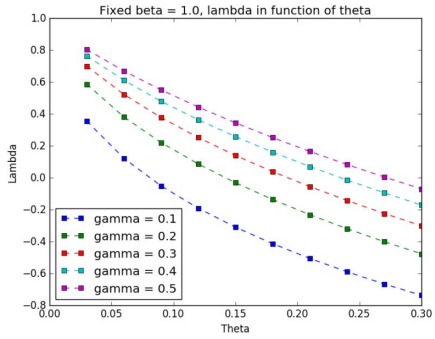
8. Complementary results and further discussions. In this part, we discuss some other properties and generalisations of our model.

8.1. *Fragmentation by removing vertices.* One can also consider a similar GFI process where the fragmentation is generated by removing vertices. We call the new process *v-GFI process* for short. More precisely, this process has the same dynamics of growth and isolation, but every vertex rings independently with an exponential clock of parameter γ, and once it rings, the associated vertex is removed from the cluster to generate several subclusters.

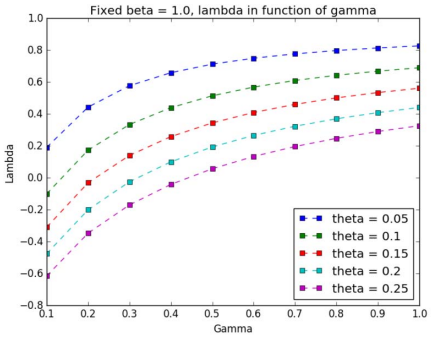
For the model of fragmentation by removing vertices on RRT (without isolation), some interesting properties about the size distribution have been discussed in [24]. Especially, a generalised splitting property should hold for our v-GFI process, so we can also follow the same strategy to study its size process $(X_t, Y_t)_{t\ge 0}$. Moreover, by [24], eq.(7), we deduce that



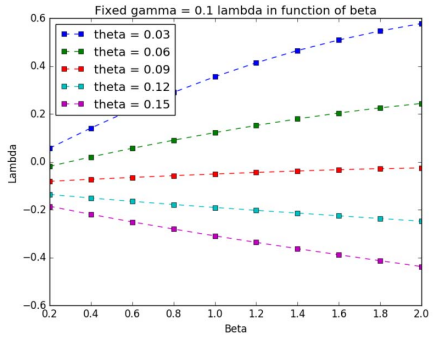
(a) The values of λ with $\beta = 1$.



(b) Each curve describes the dependence of λ on θ, with γ, β fixed.



(c) Each curve describes the dependence of λ on γ, with θ, β fixed.



(d) Each curve describes the dependence of λ on β, with θ, γ fixed.

FIG. 8. *The dependence of* λ *on* (β, θ, γ).

the generator for the first moment semigroup of $(X_t)_{t\geq 0}$ is

$$\begin{aligned}\mathcal{L}_\bullet f(n) &= \beta n\big(f(n+1) - f(n)\big) - \theta n f(n) \\ &\quad - \gamma n f(n) + \gamma n \sum_{j=1}^{n-1} \left(\frac{1}{j(j+1)} + \frac{1}{(n-j)(n-j+1)}\right) f(j).\end{aligned} \tag{77}$$

Compared to (14), we have $\mathcal{L}_\bullet f = \mathcal{L} f - \gamma f$, which implies that $e^{\mathcal{L}_\bullet t} = e^{-\gamma t} e^{\mathcal{L} t} = e^{-\gamma t} M_t$. Therefore we conclude that, quite surprisingly, $\mathcal{L}$ and $\mathcal{L}_\bullet$ have exactly the same eigenelements (π, h) given in Proposition 4.1, with a modification of Perron's root $\lambda \mapsto \lambda - \gamma$. We can expect to recover other similar properties for the v-GFI process.

8.2. *Role of the initial condition.* Although the main results are proved with the initial condition $G_0 = \{0\}$ (a patient zero), the proofs and results will not change much if the initial cluster is a RRT of any size. Indeed the reduction of the study to a growth-fragmentation-isolation process for sizes works similarly.

Besides, extension of the results to an initial condition given by a collection of RRTs can be achieved using that each initial cluster evolves independently by branching property.

Finally, one may wish to start from one (or a collection of) deterministic finite tree(s). We believe that our work can be adapted or used in that purpose, for instance, using the stopping line when the clusters have size 1 to exploit the results of this paper.

8.3. *Model generalisations*. We can use the same method to deal with generalisations such as allowing a cluster to lose several edges at the same time, or infect several individuals at the same time. More delicate generalisations could be interesting too:

- Adding recovering;
- Breaking the Markov property (adding an age would register the time for infection and would affect the individual rates in the epidemics: contamination, loss of contact information, and recovery);
- Improving and enriching the tracing procedure;
- Adding heterogeneity within the population (for tracing and infection).

The reduction we have exploited (from collections of trees to collection of sizes, due to the nice splitting property of RRT) should fail in general. We hope there is still some nice probabilistic constructions (even if more complex) to find out but it is still speculative. We are confident on the extension of asymptotic results on the Malthusian growth and empirical measures. It could be achieved without going through a reduction of the state space and without describing the distribution of clusters at fixed times, but rather exploiting the general statement on semigroups that we have used here on a more complex state space (space of trees). In particular, the Lyapunov functions that we have exhibited here may still be relevant to deal with processes in more general state spaces.

8.4. *Finite-distance contact tracing*. In our model, the contact tracing is implemented for the whole cluster despite of its size, which is very rare in reality. It is natural to ask whether the properties observed in our model also hold for the finite-distance contact tracing. The answer is probably true, but the justification requires some new mathematical techniques.

APPENDIX: CONDITIONAL EXPECTATION AND VARIANCE

We prove here Lemma 5.1 for the conditional expectation and variance estimates of

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

used in (51). Here $(\widetilde{X}_t)_{t\in[k\Delta,(k+1)\Delta)}$, containing only growth and no fragmentation nor isolation, is a coupling of $(X_t)_{t\in[k\Delta,(k+1)\Delta)}$.

PROOF OF LEMMA 5.1. For the conditional expectation, we decompose it using the genealogy of clusters

$$\mathbb{E}[Z_{k,p,K}|\mathcal{F}_{k\Delta}] = \sum_{u\in\mathcal{U}_{k\Delta}} \mathbb{E}\big[[x^p]_{>K}(\widetilde{X}^u_{(k+1)\Delta}) - [x^p]_{>K}(\widetilde{X}^u_{k\Delta})|\mathcal{F}_{k\Delta}\big]. \tag{78}$$

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

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

On the right-hand side, the second term can be bounded by $K^p\mathbf{1}_{\{a\le K<b\}}$ and the third term is zero, so we have

$$[x^p]_{>K}(b) - [x^p]_{>K}(a) \le (b^p - a^p) + K^p\mathbf{1}_{\{a\le K<b\}}. \tag{79}$$

Since there is only growth in the process $\widetilde{X}_t$ on $[k\Delta,(k+1)\Delta)$, $\widetilde{X}^u_{(k+1)\Delta-} \ge \widetilde{X}^u_{k\Delta}$ for any $u \in \mathcal{U}_{k\Delta}$. So we can apply (79) with $a = \widetilde{X}^u_{k\Delta}$ and $b = \widetilde{X}^u_{(k+1)\Delta-}$ to obtain that

$$\begin{aligned}\mathbb{E}[Z_{k,p,K}|\mathcal{F}_{k\Delta}] &\le \mathbb{E}\big[\langle \widetilde{X}_{(k+1)\Delta-}, [x^p]\rangle - \langle \widetilde{X}_{k\Delta}, [x^p]\rangle|\mathcal{F}_{k\Delta}\big] \\ &\quad + K^p \sum_{u\in\mathcal{U}_{k\Delta}} \mathbb{E}[\mathbf{1}_{\{\widetilde{X}^u_{k\Delta}\le K<\widetilde{X}^u_{(k+1)\Delta}\}}|\mathcal{F}_{k\Delta}].\end{aligned} \tag{80}$$

For the first term on the right-hand side in (80), we follow the proof of Lemma 4.1(i) to get

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

For the second term in (80), we can control it by the total number of active clusters of size smaller than K at $k\Delta$ that grow within $[k\Delta, (k+1)\Delta)$:

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

Plugging in the two inequalities to (80) yields (52).

For the conditional variance, we have

$$\operatorname{var}[Z_{k,p,K} | \mathcal{F}_{k\Delta}] = \operatorname{var}\Big[\sum_{u\in\mathcal{U}_{k\Delta}} \big([x^p]_{>K}(\widetilde{X}^u_{(k+1)\Delta-}) - [x^p]_{>K}(\widetilde{X}^u_{k\Delta})\big) | \mathcal{F}_{k\Delta} \Big].$$

By branching property,

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

From the second line to the third line we use $(a-b)^2 \le a^2 - b^2$ for any $a > b > 0$. The rest is the same as in the computation of conditional expectation and we obtain (53). □

Acknowledgments. Parts of this work were finished when C.G. was affiliated to NYU Shanghai.

Funding. 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] AKIAN, M., GANASSALI, L., GAUBERT, S. and MASSOULIÉ, L. (2020). Probabilistic and mean-field model of COVID-19 epidemics with user mobility and contact tracing. arXiv preprint arXiv:2009.05304.

[2] ASMUSSEN, S. and HERING, H. (1976). Strong limit theorems for general supercritical branching processes with applications to branching diffusions. *Z. Wahrsch. Verw. Gebiete* **36** 195–212. MR0420889 https://doi.org/10.1007/BF00532545

[3] ATHREYA, K. B. (1968). Some results on multitype continuous time Markov branching processes. *Ann. Math. Stat.* **39** 347–357. MR0221600 https://doi.org/10.1214/aoms/1177698395

[4] ATHREYA, K. B. (2012). Coalescence in the recent past in rapidly growing populations. *Stochastic Process. Appl.* **122** 3757–3766. MR2965924 https://doi.org/10.1016/j.spa.2012.06.015

[5] BALL, F. and DONNELLY, P. (1995). Strong approximations for epidemic models. *Stochastic Process. Appl.* **55** 1–21. MR1312145 https://doi.org/10.1016/0304-4149(94)00034-Q

[6] BANSAYE, V., CLOEZ, B. and GABRIEL, P. (2020). Ergodic behavior of non-conservative semigroups via generalized Doeblin's conditions. *Acta Appl. Math.* **166** 29–72. MR4077228 https://doi.org/10.1007/s10440-019-00253-5

[7] BANSAYE, V., CLOEZ, B., GABRIEL, P. and MARGUET, A. (2022). A non-conservative Harris ergodic theorem. *J. Lond. Math. Soc.* (2) **106** 2459–2510. MR4498558 https://doi.org/10.1112/jlms.12639

[8] BANSAYE, V., DELMAS, J.-F., MARSALLE, L. and TRAN, V. C. (2011). Limit theorems for Markov processes indexed by continuous time Galton–Watson trees. *Ann. Appl. Probab.* **21** 2263–2314. MR2895416 https://doi.org/10.1214/10-AAP757

[9] BANSAYE, V., GU, C. and YUAN, L. (2021). A growth-fragmentation-isolation process on random recursive trees and contact tracing. arXiv preprint arXiv:2109.05760.

[10] BARLOW, M. T. (2020). A branching process with contact tracing. Preprint, available at https://www.math.ubc.ca/~barlow/preprints/112-bpct5.pdf.

[11] BAUR, E. and BERTOIN, J. (2014). Cutting edges at random in large recursive trees. In *Stochastic Analysis and Applications* 2014. *Springer Proc. Math. Stat.* **100** 51–76. Springer, Cham. MR3332709 https://doi.org/10.1007/978-3-319-11292-3_3

[12] BERTOIN, J. (2012). Fires on trees. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 909–921. MR3052398 https://doi.org/10.1214/11-AIHP435

[13] BERTOIN, J. (2017). Markovian growth-fragmentation processes. *Bernoulli* **23** 1082–1101. MR3606760 https://doi.org/10.3150/15-BEJ770

[14] BERTOIN, J. (2022). A model for an epidemic with contact tracing and cluster isolation, and a detection paradox. arXiv preprint arXiv:2201.01924.

[15] BERTOIN, J. and WATSON, A. R. (2018). A probabilistic approach to spectral analysis of growth-fragmentation equations. *J. Funct. Anal.* **274** 2163–2204. MR3767431 https://doi.org/10.1016/j.jfa.2018.01.014

[16] BERTOIN, J. and WATSON, A. R. (2020). The strong Malthusian behavior of growth-fragmentation processes. *Ann. Henri Lebesgue* **3** 795–823. MR4149826 https://doi.org/10.5802/ahl.46

[17] DU, M. (2022). Contact tracing as a measure to combat Covid-19 and other infectious diseases. *Am. J. Infect. Control* **50** 638–644.

[18] ENGLÄNDER, J., HARRIS, S. C. and KYPRIANOU, A. E. (2010). Strong law of large numbers for branching diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 279–298. MR2641779 https://doi.org/10.1214/09-AIHP203

[19] FETZER, T. and GRAEBER, T. (2021). Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment. *Proc. Natl. Acad. Sci. USA* **118** e2100814118.

[20] GU, C., JIANG, W., ZHAO, T. and ZHENG, B. (2020). Mathematical recommendations to fight against COVID-19. Available at SSRN 3551006.

[21] HARRIS, S. C., HORTON, E. and KYPRIANOU, A. E. (2020). Stochastic methods for the neutron transport equation II: Almost sure growth. *Ann. Appl. Probab.* **30** 2815–2845. MR4187129 https://doi.org/10.1214/20-AAP1574

[22] HARRIS, S. C., JOHNSTON, S. G. G. and ROBERTS, M. I. (2020). The coalescent structure of continuous-time Galton–Watson trees. *Ann. Appl. Probab.* **30** 1368–1414. MR4133376 https://doi.org/10.1214/19-AAP1532

[23] HORTON, E. and WATSON, A. R. (2022). Strong laws of large numbers for a growth-fragmentation process with bounded cell sizes. *ALEA, Lat. Am. J. Probab. Math. Stat.* **19** 1799–1826. https://doi.org/10.30757/ALEA.v19-68

[24] KALAY, Z. and BEN-NAIM, E. (2015). Fragmentation of random trees. *J. Phys. A* **48** 0405001, 15. MR3300251 https://doi.org/10.1088/1751-8113/48/4/045001

[25] KEELING, M. J., HOLLINGSWORTH, T. D. and READ, J. M. (2020). Efficacy of contact tracing for the containment of the 2019 novel coronavirus (Covid-19). *J. Epidemiol. Community Health* **74** 861–866. https://doi.org/10.1136/jech-2020-214051

[26] KESTEN, H. and STIGUM, B. P. (1966). A limit theorem for multidimensional Galton–Watson processes. *Ann. Math. Stat.* **37** 1211–1223. MR0198552 https://doi.org/10.1214/aoms/1177699266

[27] KURTZ, T., LYONS, R., PEMANTLE, R. and PERES, Y. (1997). A conceptual proof of the Kesten–Stigum theorem for multi-type branching processes. In *Classical and Modern Branching Processes* (*Minneapolis, MN*, 1994). *IMA Vol. Math. Appl.* **84** 181–185. Springer, New York. MR1601737 https://doi.org/10.1007/978-1-4612-1862-3_14

[28] LAMBERT, A. (2021). A mathematical assessment of the efficiency of quarantining and contact tracing in curbing the COVID-19 epidemic. *Math. Model. Nat. Phenom.* **16** Paper No. 53, 23. MR4320883 https://doi.org/10.1051/mmnp/2021042

[29] MARGUET, A. (2019). A law of large numbers for branching Markov processes by the ergodicity of ancestral lineages. *ESAIM Probab. Stat.* **23** 638–661. MR4011569 https://doi.org/10.1051/ps/2018029

[30] MARZOUK, C. (2016). Fires on large recursive trees. *Stochastic Process. Appl.* **126** 265–289. MR3426519 https://doi.org/10.1016/j.spa.2015.08.006

[31] MEIR, A. and MOON, J. (1974). Cutting down recursive trees. *Bellman Prize Math. Biosci.* **21** 173–181.

[32] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. MR1234295 https://doi.org/10.2307/1427522

[33] MISCHLER, S. and SCHER, J. (2016). Spectral analysis of semigroups and growth-fragmentation equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **33** 849–898. MR3489637 https://doi.org/10.1016/j.anihpc.2015.01.007

[34] TOMAŠEVIĆ, M., BANSAYE, V. and VÉBER, A. (2022). Ergodic behaviour of a multi-type growth-fragmentation process modelling the mycelial network of a filamentous fungus. *ESAIM Probab. Stat.* **26** 397–435. MR4516847 https://doi.org/10.1051/ps/2022013