Adv. Appl. Probab. **55,** 407–443 (2023)
*doi:*10.1017/apr.2022.40

DEGREE DISTRIBUTIONS IN RECURSIVE TREES WITH FITNESSES

TEJAS IYER ,* *Weierstrass Institute, Berlin*

Abstract

We study a general model of recursive trees where vertices are equipped with independent weights and at each time-step a vertex is sampled with probability proportional to its fitness function, which is a function of its weight and degree, and connects to ℓ new-coming vertices. Under a certain technical assumption, applying the theory of Crump–Mode–Jagers branching processes, we derive formulas for the limiting distributions of the proportion of vertices with a given degree and weight, and proportion of edges with endpoint having a certain weight. As an application of this theorem, we rigorously prove observations of Bianconi related to the evolving Cayley tree (*Phys. Rev. E* **66**, paper no. 036116, 2002). We also study the process in depth when the technical condition can fail in the particular case when the fitness function is affine, a model we call 'generalised preferential attachment with fitness'. We show that this model can exhibit condensation, where a positive proportion of edges accumulates around vertices with maximal weight, or, more drastically, can have a degenerate limiting degree distribution, where the entire proportion of edges accumulates around these vertices. Finally, we prove stochastic convergence for the degree distribution under a different assumption of a strong law of large numbers for the partition function associated with the process.

Keywords: Complex networks; generalised preferential attachment with fitness; random recursive tree; plane-oriented recursive tree; Crump–Mode–Jagers branching processes; condensation

2020 Mathematics Subject Classification: Primary 90B15
Secondary 60J20; 05C80

1. Introduction

Recursive trees are *rooted* labelled trees that are *increasing*; that is, starting at a distinguished *root* vertex labelled 0, nodes are labelled in increasing order away from the root. Recursive trees generated using stochastic processes have attracted widespread study, motivated by, for example, their applications to the evolution of languages [29], the analysis of algorithms [25], and the study of complex networks (see, for example, [38, Chapter 8.1]). Other applications include modelling the spread of epidemics, modelling pyramid schemes, and constructing family trees of ancient manuscripts (e.g. [15, p. 14]).

A common framework for randomly generating recursive trees is to have vertices arrive one at a time and each connect to an existing vertex in the tree, selected according to some probability distribution. In the *uniform recursive tree*, introduced by Na and Rapoport in [28], existing vertices are chosen uniformly at random, whilst the well-known random *ordered recursive tree*,

Received 23 June 2020; revision received 29 June 2022.
* Postal address: Weierstrass Institute, Berlin, Germany. Email address: tejas.iyer@wias-berlin.de

introduced by Prodinger and Urbanek in [32], may be interpreted as having existing vertices chosen with probability proportional to their degree. The latter model has been studied and rediscovered in various guises: under the names *nonuniform recursive trees* by Szymański in [36], random *plane-oriented recursive trees* in [24, 26], random *heap-ordered recursive trees* in [10], and *scale-free trees* in [6, 7, 35]. Random ordered recursive trees, or plane-oriented recursive trees, are so named because the process stopped after n vertices arrive is distributed like a tree chosen at random from the set of rooted labelled trees on n vertices embedded in the plane where descendants of a node are ordered from left to right. However, as first observed by Albert and Barabási in [2] and studied in a mathematically precise way in [7, 27], these trees, and more generally graphs evolving according to a similar mechanism, possess many interesting, non-trivial properties of real-world networks. These properties include having a power law degree distribution with exponent between 2 and 3 and a diameter that scales logarithmically in the number of vertices. The latter may be interpreted as a 'small-world' phenomenon: although the size of the network is large, the diameter of the network remains relatively small. In this context, the fact that vertices are chosen according to their degree may be interpreted as the network showing 'preference' for vertices of high degree; hence the model is often called *preferential attachment*. This model has been generalised in a number of ways, to encompass the cases where vertices are chosen according to a *super-linear* function of their degree [31] and a *sub-linear* function of their degree [13], or indeed any positive function of the degree [33]. In [20], this model is generalised to possibly non-negative functions of the degree and is referred to as *generalised preferential attachment*.

In applications, it is often interesting to add weights to vertices as a measure of the intrinsic 'fitness' of the node the vertex represents. In the *Bianconi–Barabási model*, or *preferential attachment with multiplicative fitness*, introduced in [5] and studied in a mathematically precise way in [3, 8, 11, 12, 14], vertices are equipped with independent, identically distributed (i.i.d.) weights and connect to previous vertices with probability proportional to the product of their weight and their degree. Interestingly, as observed in [5] and confirmed rigorously in [8, 12, 14], there is a critical condition on the weight distribution under which this model undergoes a phase transition, resulting in a *Bose–Einstein condensation*: in the limit, a positive fraction of vertices accumulate around vertices of maximum weight. In a similar model known as *preferential attachment with additive fitness*, introduced in [16] and studied mathematically in [3, 23, 34], vertices connect to previous vertices with probability proportional to the sum of their degree (or degree minus one) and their weight. A number of other interesting preferential attachment models with fitness have been studied, including a model of preferential attachment with both additive and multiplicative weights [16], a related continuous-time model which incorporates *ageing* of vertices [18], and discrete-time models with co-existing additive and multiplicative attachment rules [1, 22].

Adding weights also allows for a generalisation of the uniform recursive tree called the *weighted recursive tree*, where now vertices connect to previous vertices with probability proportional to their weight. This model was introduced in [9] for specific types of weights and in [19] in full generality. It was also introduced independently by Janson in the case that all weights are one except at the root, motivated by applications to infinite-colour Pólya urns [21]. In [34], Sénizergues showed that a preferential attachment tree with additive fitness with deterministic weights is equal in distribution to an associated weighted random recursive tree with random weights, an interesting link between the two classes of models.

Motivated by applications to invasive percolation models in physics, Bianconi [4] introduced a similar model of growing *Cayley trees*. In this model, vertices are equipped with independent weights and are either *active* or *inactive*. At each time-step an active vertex is chosen with probability proportional to its weight, produces ℓ new vertices with weights of their own, and then becomes inactive. Bianconi observed that in this model, the distribution of weights of active vertices converges to a *Fermi–Dirac distribution*, in contrast to the *Bose distribution* that emerges in the Bianconi–Barabási model.

1.1. Notation

Generally in this paper we set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and $\mathbb{R}_+ := [0, \infty)$. We assume that $\mathbb{R}_+$ is equipped with the usual *Borel sigma-algebra*, and μ will denote a fixed probability measure on $\mathbb{R}_+$. Also, in general in this paper, W refers to a generic μ-distributed random variable. Finally, we generally refer to an element of the Borel sigma-field as a *measurable set*. Given such a set A, we denote by $\mathbf{1}_A(x)$ the indicator function associated with the set, so that $\mathbf{1}_A(x) = 1$ if $x \in A$ and 0 otherwise. Moreover, if $\mathbf{1}_A(x)$ is a random variable on a probability space $(\Omega, \mathbb{F}, \mathbb{P})$, we often omit the dependence on $x \in \Omega$, and simply write $\mathbf{1}_A$.

1.2. Description of model

The goal of this paper is to present a unified model that encompasses most of the models described in the introduction above. In order to define the model, we first require a probability measure μ supported on $\mathbb{R}_+$ and a *fitness function*, which is a function $f : \mathbb{N}_0 \times \mathbb{R}_+ \to \mathbb{R}_+$. We consider evolving sequences of *weighted oriented trees* $\mathcal{T} := (\mathcal{T}_t)_{t\in\mathbb{N}_0}$; these are trees with *directed edges*, where vertices have real-valued weights assigned to them. The model also has an additional parameter $\ell \in \mathbb{N}$. We start with an initial tree $\mathcal{T}_0$ consisting of a single vertex 0 with weight W_0 sampled from μ. To ensure that the evolution of the model is well-defined, for brevity, we assume $f(0, W_0) > 0$ almost surely. Then we define $\mathcal{T}_{t+1}$ recursively as follows:

(i) Sample a vertex j from $\mathcal{T}_t$ with probability

$$\frac{f(\deg^+(j, \mathcal{T}_t)/\ell, W_j)}{\mathcal{Z}_t},$$

where $\deg^+(j, \mathcal{T}_t)$ denotes the out-degree of the vertex j in the oriented tree $\mathcal{T}_t$, and $\mathcal{Z}_t := \sum_{j=0}^{\ell t} f(\deg^+(j, \mathcal{T}_t)/\ell, W_j)$ is the *partition function* associated with the process.

(ii) Introduce ℓ new vertices $t+1, t+2, \ldots, t+\ell$ with weights $W_{t+1}, W_{t+2}, \ldots, W_{t+\ell}$ sampled independently from μ and the directed edges $(j, t+1), (j, t+2), \ldots, (j, t+\ell)$ oriented towards the newly arriving vertices. We say that j is the *parent* of the new-coming vertices.

Note that, since ℓ new vertices are connected to a parent at each time-step, for any vertex i in the tree, ℓ divides the out-degree of i. Moreover, the evolution of the out-degree of vertex i with weight W_i is determined by the values $(f(j, W_i))_{j\in\mathbb{N}_0}$. In general, when the distribution μ, fitness function f, and ℓ are specified, we refer to this model as a (μ, f, ℓ)-*recursive tree with independent fitnesses*, often abbreviated as a '(μ, f, ℓ)-RIF tree' for brevity. Here 'independent fitnesses' refers to the fact that the fitness associated with a given vertex does not depend on the weights of its neighbours, in contrast to, for example, the models of dynamical *simplicial complexes* studied in [17].

Remark 1.1. If we adopt the convention that the process terminates when no vertex can be chosen in the next step, the assumption that $f(0, W_0) > 0$ almost surely may be dropped in many places in this paper, if we condition on the event that the number of vertices in the tree tends to infinity as $t \to \infty$.

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

Remark 1.2. In this model, the law of the sequence $(f(k, W))_{k\in\mathbb{N}_0}$ is more important than the function f. It is possible, for example, to define this model so that $(f(k, W))_{k\in\mathbb{N}_0}$ is any law on $\mathbb{R}_+^{\mathbb{N}_0}$ depending on W, even if one cannot write f explicitly, and as long as the sequences associated with different vertices are independent. For example, the sequence could be any stochastic process indexed by the non-negative integers, depending on an initial source of randomness W.

1.3. Quantities of interest studied in this paper

In this section, we will introduce the main quantities we will be interested in studying in this paper, along with some important definitions. Note that the definitions we introduce in this section will depend on the underlying parameters of the tree, μ, f, and ℓ.

In this paper we will generally be concerned with the limiting behaviour of the following quantities:

1. Given a Borel set $B \subseteq \mathbb{R}_+$, the quantity $N_k(t, B)$ denotes the number of vertices v in the tree $\mathcal{T}_t$ with out-degree $k\ell$ and weight $W_v \in B$; that is,
$$N_k(t, B) := \sum_{v\in\mathcal{T}_t:\ \deg^+(v,\mathcal{T}_t)=k\ell} \mathbf{1}_B(W_v). \tag{1}$$
2. Given a Borel set $B \subseteq \mathbb{R}_+$, the quantity $\Xi(t, B)$ denotes the number of directed edges (v, v') in the tree $\mathcal{T}_t$ such that $W_v \in B$; that is,
$$\Xi(t, B) := \sum_{(v,v')\in\mathcal{T}_t} \mathbf{1}_B(W_v). \tag{2}$$

Now, reasoning informally and non-rigorously for a moment, suppose that W takes finitely many values. In addition, suppose that for all $t \geq t'$, where one considers t' to be a 'large' constant, we have $\mathcal{Z}_t = \alpha t$, and for all $k \in \mathbb{N}_0$ we have $N_k(t, \{w\}) = \mathbb{E}[N_k(t, \{w\})] = \ell t \cdot n_k(\{w\})$ for some value $n_k(\{w\})$. The latter assumptions are motivated by the intuition that the respective quantities obey strong laws of large numbers. Then, for $t \geq t'$ and $k \geq 1$, we have

$$\begin{aligned}
\ell n_k(\{w\}) &= \mathbb{E}[\mathbb{E}[N_k(t+1, \{w\}) - N_k(t, \{w\})|\mathcal{T}_t]] \\
&= \mathbb{P}(\text{vertex of out-degree } k \text{ and weight } w \text{ chosen}) \\
&\qquad\qquad - \mathbb{P}(\text{vertex of out-degree } k-1 \text{ and weight } w \text{ chosen}) \\
&= N_{k-1}(t, \{w\}) \cdot \frac{f(k-1, w)}{\mathcal{Z}_t} - N_k(t, \{w\}) \cdot \frac{f(k, w)}{\mathcal{Z}_t} \\
&= \frac{\ell n_{k-1}(\{w\}) \cdot f(k-1, w)}{\alpha} - \frac{\ell n_k(\{w\}) \cdot f(k, w)}{\alpha}.
\end{aligned}$$

Meanwhile, for $k = 0$ we have

$$\begin{aligned}
\ell n_0(\{w\}) &= \mathbb{E}[\mathbb{E}[N_0(t+1, \{w\}) - N_0(t, \{w\})|\mathcal{T}_t]] \\
&= \mathbb{P}(\text{newly arriving vertex with weight } w) \\
&\qquad\qquad - \mathbb{P}(\text{vertex of out-degree 0 and weight } w \text{ chosen}) \\
&= \ell \mu(\{w\}) - N_0(t, \{w\}) \cdot \frac{f(0, w)}{\mathcal{Z}_t} \\
&= \ell \mu(\{w\}) - \frac{\ell n_0(\{w\}) \cdot f(0, w)}{\alpha}.
\end{aligned}$$

Solving the recursion from the above two displays, for all $k \in \mathbb{N}_0$ we have

$$n_k(\{w\}) = \mu(\{w\}) \cdot \frac{\alpha}{f(k, w) + \alpha} \prod_{i=0}^{k-1} \frac{f(i, w)}{f(i, w) + \alpha}$$
$$= \mathbb{E}\left[\frac{\alpha}{f(k, W) + \alpha} \prod_{i=0}^{k-1} \frac{f(i, W)}{f(i, W) + \alpha} \mathbf{1}_{\{w\}}(W)\right].$$

It is therefore reasonable to expect that the limit of $\frac{N_k(t,B)}{\ell t}$ belongs to a one-parameter family $p_k^{\lambda}(\cdot)$ indexed by a positive real number λ such that

$$p_k^{\lambda}(B) := \mathbb{E}\left[\frac{\lambda}{f(k, W) + \lambda} \prod_{i=0}^{k-1} \frac{f(i, W)}{f(i, W) + \lambda} \mathbf{1}_B(W)\right], \tag{3}$$

where the parameter λ can be recovered by the asymptotics of the partition function, so that the limit satisfies $\lambda = \alpha > 0$. It is important to note that in this paper, it may be the case that the limit of $\frac{N_k(t,B)}{\ell t}$ belongs to this family, but we do not necessarily have a strong law for the asymptotics of the partition function. For example, this is the case if the conditions of Theorem 2.1 are satisfied, but not those of Theorem 2.4.

Now, note that for every $t \in \mathbb{N}_0$, by computing the number of directed edges (v, v') in $\mathcal{T}_t$ with $W_v \in B$ in two different ways, we have

$$\Xi(t, B) = \sum_{k=0}^{t} \ell k N_k(t, B). \tag{4}$$

When we normalise by ℓt, if, for $k \in \mathbb{N}_0$ the limit of $\frac{N_k(t,B)}{\ell t}$ is $p_k^{\alpha}(B)$, by Fatou's lemma we get

$$\liminf_{t\to\infty} \frac{\Xi(t, B)}{\ell t} \geq \sum_{k=0}^{\infty} \ell k p_k^{\alpha}(B), \tag{5}$$

which motivates the definition of the following family indexed by a positive real number λ:

$$m(\lambda, B) := \sum_{k=0}^{\infty} \ell k p_k^{\lambda}(B) = \ell \cdot \mathbb{E}\left[\sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{f(i, W)}{f(i, W) + \lambda} \mathbf{1}_B(W)\right]. \tag{6}$$

Now, if the limit exists, since we add ℓ edges at each time-step, the limit of the measures $\Xi(t, \cdot)/\ell t$ is a probability measure. However, if $m(\alpha, \cdot)$ is not a probability distribution, we can show that there exists a measurable set B such that

$$\limsup_{t\to\infty} \frac{\Xi(t, B)}{\ell t} > m(\lambda, B).$$

In this case, the inequality in (5) is strict, so that, after normalising by ℓt, the operations of taking limits in k and in t in (4) do not commute. Thus, the set B has acquired additional 'mass' in the limit. We call this phenomenon *condensation*, motivated by the term used in the network science literature (e.g. [5]). In Section 3.2 we derive an example of this in the case that $f(i, W) = g(W)i + h(W)$, where g is bounded. This generalises the case $f(i, W) = (i + 1)W$ which has already been studied in [8, 12, 14] under the name *preferential attachment with multiplicative fitness*.

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

1.4. Open problems

The discussion in Section 1.3 shows that much of the analysis of this model depends on a parameter α. We conjecture that, in general, this parameter makes $m(\lambda, \cdot)$ 'as close as possible' to a probability distribution, so that

$$\alpha = \inf \{\lambda > 0 : m(\lambda, \mathbb{R}_+) \leq 1\} \qquad \text{if } m(\lambda, \mathbb{R}_+) < \infty \text{ for some } \lambda > 0, \tag{7}$$

where we follow the convention that $\inf \varnothing = \infty$.

Conjecture 1.1. *Let $\mathcal{T}$ be a (μ, f, ℓ)-RIF tree, with α as defined in (7). Then, for each $k \in \mathbb{N}_0$ and measurable set B, almost surely, we have*

$$\frac{N_k(t, B)}{\ell t} \xrightarrow{t \to \infty} \begin{cases} p_k^{\alpha}(B) & \text{if } \alpha < \infty, \\ \mu(B)\mathbf{1}_{\{0\}}(k) & \text{otherwise.} \end{cases}$$

The conjectured limit in the case when $\alpha = \infty$ is obtained by taking the limit of $p_k^{\alpha}(B)$ as $\alpha \to \infty$. This limit is 0 unless $k = 0$, in which case it is $\mu(B)$.

Remark 1.3. Conjecture 1.1 has a natural analogue in the setting, as in Remark 1.2, that the sequence $(f(k, W))_{k \in \mathbb{N}_0}$ is instead given by a general stochastic process indexed by $\mathbb{N}_0$, depending on an initial source of randomness W. In this case, the expectation in (3) is instead taken over evolution over all sequences, with initial $W \in B$. The techniques used in this paper translate without modification to this case, the only important feature being that the sequences corresponding to different vertices are independent of each other.

Remark 1.4. It is important to note the order of the quantifiers in Conjecture 1.1: given a sequence $(\mu_j)_{j \in \mathbb{N}}$ of random measures on $\mathbb{R}_+$, we have that, for any measurable set B, $\lim_{j \to \infty} \mu_j(B) = \mu_\infty(B)$ almost surely. It is not necessarily the case, however, that, almost surely, for all measurable sets B, we have $\lim_{j \to \infty} \mu_j(B) = \mu_\infty(B)$. However, it is the case that almost surely, $\mu_j \to \mu_\infty$ in the weak topology. This uses the fact that there exists a countable family of measurable sets such that any open set in $\mathbb{R}_+$ may be expressed as a disjoint, countable union of elements of this family. For example, one may take the set of all dyadic intervals, with endpoints of the form $j \cdot 2^{-n}$, $(j+1) \cdot 2^{-n}$, where $j, n \in \mathbb{N}_0$, and then apply the portmanteau theorem. This approach is also used in this paper in the proof of Corollary 3.1.

The discussion in Section 1.3 described the quantity α as being closely related to the partition function. As a result, we also conjecture the following.

Conjecture 1.2. *Let $\mathcal{T}$ be a (μ, f, ℓ)-RIF tree, with α as defined in (7). Then we have*

$$\frac{\mathcal{Z}_t}{t} \xrightarrow{t \to \infty} \alpha, \qquad \textit{almost surely.}$$

1.5. Important technical conditions and overview of results

In this paper, we make partial progress towards the proofs of Conjecture 1.1 and Conjecture 1.2. We will refer to the following technical conditions:

C1 With $m(\lambda, \cdot)$ as defined in (6), there exists some $\lambda > 0$ such that

$$1 < m(\lambda, \mathbb{R}_+) < \infty. \tag{8}$$

Under this condition, by monotonicity, there exists a unique $\alpha > 0$ such that $m(\alpha, \mathbb{R}_+) = 1$; we call this the *Malthusian parameter* associated with the process.

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

C2 There exists $\alpha > 0$ such that

$$\lim_{t\to\infty} \frac{\mathcal{Z}_t}{t} = \alpha.$$

Note that in (7) and in Conditions **C1** and **C2** we use the same symbol α. This these coincide in general. In general, as we only assume either **C1** or **C2** at a time, the definition will be clear from context.

The paper will be structured as follows.

Section 2: We analyse the model under Condition **C1**.

- In Theorem 2.1 we prove Conjecture 1.1 under Condition **C1**, and as a consequence, in Theorem 2.2 we show that, for any measurable set B, $\Xi(t, B)/\ell t$ converges almost surely to $m(\alpha, B)$.
- In Theorem 2.4 we derive a condition under which **C1** implies **C2**. In particular, this proves Conjecture 1.2 under this condition and **C1**.
- The approaches used in this section are well established, applying classical results in the theory of *Crump–Mode–Jagers branching processes*, in a similar manner to the approaches taken by the authors of [3, 12, 20, 33]. Nevertheless, these theorems have novel applications: we apply the theorems to the evolving Cayley tree considered by Bianconi in Example 2.4.1 and the *weighted random recursive tree*.

Section 3: We analyse a particular case of the model when the fitness function is such that $f(i, W) = g(W)i + h(W)$, which we call the *generalised preferential attachment tree with fitness* (GPAF-tree). This model, closely related to a model introduced in [16], extends the existing models of preferential attachment with additive fitness, i.e., $f(i, W) = i + 1 + W$, and multiplicative fitness, i.e., $f(i, W) = (i + 1)W$. When the function g is non-decreasing, we also treat the cases where Condition **C1** can fail. Let α be as defined in (7), and also define $\Lambda := \{\lambda > 0 : m(\lambda, \mathbb{R}_+) < \infty\}$.

- We consider the situation in which Condition **C1** fails by having $m(\lambda, \mathbb{R}_+) \leq 1$ for all $\lambda \in \Lambda$. In this case, $m(\lambda, \mathbb{R}_+)$ is finite for some $\lambda > 0$, but never exceeds 1, so that $m(\alpha, \mathbb{R}_+) \leq 1$. In Theorem 3.1 we prove Conjecture 1.1 and Conjecture 1.2 in this case, showing, in particular, that if $m(\alpha, \mathbb{R}_+) < 1$ the GPAF-tree exhibits a *condensation* phenomenon.
- Alternatively, Condition **C1** may fail by having $\alpha = \infty$. Theorem 3.2 also confirms Conjecture 1.1 in this case, showing that the limiting degree distribution is *degenerate*: almost surely the proportion of leaves in the tree tends to 1. Moreover, we show that the *fittest take all* of the mass of the distribution of edges according to weight, in the sense that a proportion of edges tending to 1 accumulates around the sets of vertices with weights conferring higher and higher fitness.
- The techniques in this section are inspired by the coupling techniques exploited in [8, 12], and extend the well-known phase transition associated with the model of preferential attachment with multiplicative fitnesses studied in [8, 12, 14]. This generalisation shows that the phase transition depends on the parameter h too, so that, in some circumstances, condensation occurs, but vanishes if h is increased enough pointwise (see Section 3.2.2). This is interesting because $h(W)$ may be interpreted as the 'initial' popularity of a vertex when it arrives in the tree, showing that in order for the condensation to occur, there need to be sufficiently many vertices of 'low enough' initial popularity.

As far as the author is aware, this effect has not previously been observed in the general scientific literature concerning complex networks.

Section 4: We analyse the model under Condition **C2**, proving general results for the distribution of vertices with a given degree and weight.

- If the term α in Condition **C2** is finite, Theorem 4.1 and Theorem 4.2 confirm a weaker analogue of Conjecture 1.1 under this condition.
- The techniques used in this section are similar to those used in the proof of [17, Theorem 6]; however, in this instance we present a considerably shorter proof.

2. Analysis of (μ, f, ℓ)-RIF trees assuming C1

In order to apply Condition **C1** in this section, we study a branching process with a *family tree* made up of individuals and their offspring whose distribution is identical to the discrete-time model at the times of the branching events. In Section 2.1, we describe this continuous-time model, state Theorem 2.1, and state and prove Theorem 2.2. In Section 2.2 we include the relevant theory of *Crump–Mode–Jagers* branching processes and use this to prove Theorem 2.1. In Section 2.3 we apply the same theory, along with some technical lemmas, to state and prove a strong law of large numbers for the partition function in Theorem 2.4. We conclude the section with some interesting examples in Section 2.4.

2.1. Description of continuous-time embedding

In the continuous-time approach, we begin with a population consisting of a single vertex 0 with weight W_0 sampled from μ and an associated exponential clock with parameter $f(0, W_0)$. Then recursively, when the ith birth event occurs in the population, with the ringing of an exponential clock associated to vertex j, the following occurs:

(i) Vertex j produces offspring $\ell(i-1)+1, \ldots, \ell i$ with independent weights $W_{\ell(i-1)+1}, \ldots, W_{\ell i}$ sampled from μ and exponential clocks with parameters $f(0, W_{\ell(i-1)+1}), \ldots, f(0, W_{\ell i})$.

(ii) Suppose the number of offspring of j before the birth event was m, so that its out-degree in the family tree is m. Then the exponential random variable associated with j is updated to have rate $f(m/\ell+1, W_j)$. If $f(m/\ell+1, W_j)=0$, then j ceases to produce offspring and we say j has *died*.

Now, if we let $\mathcal{Z}_{i-1}$ denote the sum of the rates of the exponential clocks in the population when the population has size $i-1$, the probability that the clock associated with j is the first to ring is $f(m/\ell, W_j)/\mathcal{Z}_{i-1}$. Hence, the family tree of the continuous-time model at the times of the birth events $(\sigma_i)_{i\geq 0}$ has the same distribution as the associated (μ, f, ℓ)-RIF tree. The continuous-time branching process is actually a Crump–Mode–Jagers branching process, which we will describe in more depth in Section 2.2.

To describe the evolution of the degree of a vertex in the continuous-time model, we define the pure birth process with underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and state space $\ell\mathbb{N}$ as follows. First sample a weight W and set $Y(0)=0$. Let $\mathbb{P}_w$ denote the probability measure associated with the process when the weight sampled is w. Then define the birth rates of Y so that

$$\mathbb{P}_w(Y(t+h)=(k+1)\ell \mid Y(t)=k\ell)=f(k, w)h+o(h). \tag{9}$$

In other words, the time taken to jump from $k\ell$ to $(k+1)\ell$ is exponentially distributed with parameter $f(k, w)$.

Let ρ denote the point process corresponding to the jumps in Y, and denote by $\mathbb{E}_w[\rho(\cdot)]$ the intensity measure when the weight $W = w$. Also, denote by $\hat{\rho}_w$ the Laplace–Stieltjes transform, i.e.,

$$\hat{\rho}_w(\lambda) := \int_0^\infty e^{-\lambda t}\mathbb{E}_w[\rho(\mathrm{d}t)].$$

Note that, by Fubini's theorem, we have

$$\begin{aligned}\hat{\rho}_w(\lambda) &= \int_0^\infty \left(\int_t^\infty \lambda e^{-\lambda s}\mathrm{d}s\right)\mathbb{E}_w[\rho(\mathrm{d}t)] = \int_0^\infty \lambda e^{-\lambda s}\left(\int_0^s \mathbb{E}_w[\rho(\mathrm{d}t)]\right)\mathrm{d}s \\ &= \int_0^\infty \lambda e^{-\lambda s}\mathbb{E}_w[Y(s)]\mathrm{d}s.\end{aligned} \tag{10}$$

Moreover, if we write τ_k for the time of the kth jump in Y, we have $\rho = \sum_{k=0}^\infty \ell\delta_{\tau_k}$. Note that if the weight of Y is w, then τ_k is distributed as a sum of independent exponentially distributed random variables with rates $f(0, w), f(1, w), \ldots, f(k-1, w)$, where we follow the convention that an exponentially distributed random variable with rate 0 is ∞. Thus, we have that

$$\hat{\rho}_w(\lambda) = \ell\sum_{n=1}^\infty \mathbb{E}_w\big[e^{-\lambda\tau_n}\big] = \ell\sum_{n=1}^\infty\prod_{i=0}^{n-1}\frac{f(i, w)}{f(i, w)+\lambda}, \tag{11}$$

where in the last equality we have used the facts that a Laplace–Stieltjes transform of a convolution of measures is the product of Laplace–Stieltjes transforms, and the Laplace–Stieltjes transform $\hat{X}(\lambda)$ of an exponentially distributed random variable with parameter s is $\int_0^\infty e^{-\lambda t}se^{-st}\mathrm{d}t = \frac{s}{s+\lambda}$. Therefore, we see that $\mathbb{E}\big[\hat{\rho}_W(\lambda)\big] = m(\lambda, \mathbb{R}_+)$ as defined in (8), and Condition **C1** implies that there exists some $\lambda > 0$ such that $1 < \mathbb{E}\big[\hat{\rho}_W(\lambda)\big] < \infty$. In addition, the Malthusian parameter α appearing in Condition **C1** is the unique positive real number such that

$$\mathbb{E}\big[\hat{\rho}_W(\alpha)\big] = m(\alpha, \mathbb{R}_+) = \ell\cdot\mathbb{E}\left[\sum_{n=1}^\infty\prod_{i=0}^{n-1}\frac{f(i, W)}{f(i, W)+\alpha}\right] = 1. \tag{12}$$

Our first result is the following.

Theorem 2.1. (Convergence of the degree distribution under C1.) *Let $\mathcal{T}$ be a (μ, f, ℓ)-RIF tree satisfying C1 with Malthusian parameter α. Then, for any measurable set $B \subseteq \mathbb{R}_+$, with $N_k(t, B)$ as defined in (1) and $p_k^\alpha(B)$ as defined in (3), we have*

$$\frac{N_k(t, B)}{\ell t} \xrightarrow{t\to\infty} p_k^\alpha(B),$$

almost surely.

The limiting formula for Theorem 2.1 has appeared in a number of contexts, and generalises many known results. Under Condition **C1** this result was proved by Rudas, Tóth and Valkó [33] in the case that W is constant and $\ell = 1$. The cases $f(i, W) = W(i+1)$ and $f(i, W) = i + 1 + W$ with $\ell = 1$ correspond respectively to the preferential attachment models with multiplicative

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

and additive fitness mentioned in the introduction. In the multiplicative model, the result was first proved in [8] and later in [3]. In [3], Bhamidi also first proved the result for the case $f(i, W) = i + 1 + W$. These models are examples of the generalised preferential attachment tree with fitness, which we study in more depth in Section 3. Finally, the case $f(i, W) = W$, $\ell = 1$ corresponds to a model of weighted random recursive trees (see Example 2.4.2). We postpone the proof of Theorem 2.1 to the end of Section 2.2.

Remark 2.1. The limiting value has an interesting interpretation as a generalised geometric distribution. Consider an experiment where W is sampled from μ and, given W, coins are flipped, where the probability of heads in the ith coin flip is proportional to f(i,W) and tails proportional to α. Then the limiting distribution in Theorem 2.1 is the distribution of first occurrence of tails. Note that, by C1, the C1, the probability of infinite sequences of heads is 0.

Remark 2.2. Note that $Y(t) < \infty$ for all $t \geq 0$ almost surely if $\tau_{\infty} := \lim_{k\to\infty} \tau_k = \infty$ almost surely. The latter is satisfied if there exists λ > 0 such that for almost all w

$$\mathbb{E}_w\left[e^{-\lambda\tau_\infty}\right] = \lim_{n\to\infty} \mathbb{E}_w\left[e^{-\lambda\tau_n}\right] = \lim_{n\to\infty} \prod_{i=0}^{n} \frac{f(i, w)}{f(i, w) + \lambda} = 0,$$

which is implied by C1. In the literature concerning pure birth Markov chains, this property is known as non-explosivity.

Remark 2.3. In this paper, we have considered the case where the function f, and thus the birth process Y as defined in (9), depends on a single random variable W taking values in $\mathbb{R}_+$. However, there is no loss of generality in assuming the random variable W takes values in an arbitrary measure space, so long as the function f is measurable. In particular, we may consider the case where the weight is given by a vector (W_1, W_2) where W_1 and W_2 are possibly correlated random variables.

Now, recall the definitions of $\Xi(t, \cdot)$ from (2) and $m(\alpha, \cdot)$ from (6). In the case that $m(\alpha, \cdot)$ is a probability distribution, the almost sure convergence of $N_k(t, B)/\ell n$ to $p^{\alpha}_k(B)$ for any measurable set B is enough to imply that for any measurable set B the quantity $\Xi(t, B)$ converges almost surely to $m(\alpha, B)$. Note that this condition is weaker than directly assuming **C1**. In particular, we have the following.

Theorem 2.2. *Assume $\mathcal{T}$ is a (μ, f, ℓ)-RIF tree with limiting degree distribution of the form $(p^{\alpha}_k(\cdot))_{k\in\mathbb{N}_0}$ and such that $m(\alpha, \mathbb{R}_+) = 1$. Then for any measurable set B we have*

$$\frac{\Xi(t, B)}{\ell t} \xrightarrow{t\to\infty} m(\alpha, B),$$

almost surely.

To prove this theorem, we will apply the following elementary bound: for any two sequences $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ such that either $\liminf_{n\to\infty} a_n > -\infty$ or $\limsup_{n\to\infty} b_n < \infty$, we have

$$\liminf_{n\to\infty} (a_n + b_n) \leq \liminf_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \leq \limsup_{n\to\infty} (a_n + b_n). \tag{13}$$

Proof of Theorem 2.2. Recall that, by (4), for each t, we have $\Xi(t, B) = \sum_{k=1}^{t} k\ell N_k(t, B)$. Also note that

$$\begin{aligned}
\sum_{k=0}^{\infty} k\ell p_k^{\alpha}(B) &= \ell \cdot \mathbb{E}\left[\left(\sum_{k=1}^{\infty} \frac{k\alpha}{f(k, W)+\alpha} \prod_{i=0}^{k-1} \frac{f(i, W)}{f(i, W)+\alpha}\right) \mathbf{1}_B(W)\right] \\
&= \ell \cdot \mathbb{E}\left[\left(\sum_{k=1}^{\infty} k \cdot \left(1 - \frac{f(k, W)}{f(k, W)+\alpha}\right) \prod_{i=0}^{k-1} \frac{f(i, W)}{f(i, W)+\alpha}\right) \mathbf{1}_B(W)\right] \\
&= \ell \cdot \mathbb{E}\left[\sum_{k=1}^{\infty} \left(k \prod_{i=0}^{k-1} \frac{f(i, W)}{f(i, W)+\alpha} - k \prod_{i=0}^{k} \frac{f(i, W)}{f(i, W)+\alpha}\right) \mathbf{1}_B(W)\right] \\
&= \ell \cdot \mathbb{E}\left[\left(\sum_{k=1}^{\infty} \prod_{i=0}^{k-1} \frac{f(i, W)}{f(i, W)+\alpha}\right) \mathbf{1}_B(W)\right] = m(\alpha, B),
\end{aligned}$$

where the second-to-last equality follows from the telescoping nature of the sum inside the expectation. Thus, by Fatou's lemma, almost surely we have

$$m(\alpha, B) = \sum_{k=0}^{\infty} k\ell p_k^{\alpha}(B) = \sum_{k=0}^{\infty} k\ell \liminf_{t\to\infty} \frac{N_k(t, B)}{\ell t} \leq \liminf_{t\to\infty} \frac{\Xi(t, B)}{\ell t}; \tag{14}$$

and likewise, almost surely, $\liminf_{t\to\infty} \frac{\Xi(t,B^c)}{\ell t} \geq m(\alpha, B^c)$. Now, since we add ℓ edges at every time-step, $\Xi(t, \mathbb{R}_+) = \ell t$. Thus, by (13),

$$\begin{aligned}
1 = \liminf_{t\to\infty} \left(\frac{\Xi(t, B)}{\ell t} + \frac{\Xi(t, B^c)}{\ell t}\right) &\leq \liminf_{t\to\infty} \frac{\Xi(t, B^c)}{\ell t} + \limsup_{t\to\infty} \frac{\Xi(t, B)}{\ell t} \\
&\leq \limsup_{t\to\infty} \left(\frac{\Xi(t, B)}{\ell t} + \frac{\Xi(t, B^c)}{\ell t}\right) = 1.
\end{aligned}$$

But $m(\alpha, \cdot)$ is a probability measure; this is only possible if

$$\liminf_{t\to\infty} \frac{\Xi(t, B^c)}{\ell t} = m(\alpha, B^c) \quad \text{and} \quad \limsup_{t\to\infty} \frac{\Xi(t, B)}{\ell t} = m(\alpha, B) \quad \text{almost surely.} \tag{15}$$

Combining (14) and (15) completes the proof. □

2.2. Crump–Mode–Jagers branching processes

In the continuous-time setting, it is convenient not only to identify individuals of the branching process according to the order in which they were born, but also to record their lineage, in such a way that the labelling encodes the structure of the tree. Therefore we also identify individuals of the branching process with elements of the infinite *Ulam–Harris* tree $\mathcal{U} := \bigcup_{n\geq 0} \mathbb{N}^n$, where $\mathbb{N}^0 = \varnothing$ is the *root*. In this case, an individual $u = u_1u_2 \ldots u_k$ is to be interpreted recursively as the u_kth child of $u_1 \ldots u_{k-1}$. For example, $1, 2, \ldots$ represent the offspring of $\varnothing$.

In *Crump–Mode–Jagers (CMJ)* branching processes, individuals $u \in \mathcal{U}$ are equipped with independent copies of a random point process ξ on $\mathbb{R}_+$. The point process ξ associates *birth times* to the offspring of a given individual, and we also may assume that ξ has some dependence on a random weight W associated with that individual. The process, together

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

with birth times, may be regarded as a random variable in the probability space $(\Omega, \Sigma, \mathbb{P}) = \prod_{x\in\mathcal{U}} (\Omega_x, \Sigma_x, \mathbb{P}_x)$, where each $(\Omega_x, \Sigma_x, \mathcal{P}_x)$ is a probability space with (ξ_x, W_x) having the same distribution as (ξ, W). We denote by $(\sigma_i^x)_{i\in\mathbb{N}}$ points ordered in the point process ξ_x and, for brevity, assume that $\xi(\{0\}) = 0$. We also drop the superscript when referring to the point process associated to $\varnothing$, so that $\sigma_i := \sigma_i^{\varnothing}$. Now, we set $\sigma_\varnothing := 0$ and recursively, for $x \in \mathcal{U}$, $\sigma_{xi} := \sigma_x + \sigma_i^x$. Finally, we set $\mathbb{T}_t = \{x \in \mathcal{U} : \sigma_x \leq t\}$ and note that for each $t \geq 0$, $\mathbb{T}_t$ may be identified with the *family tree* of the process in the natural way. Informally, $\mathbb{T}_t$ can be described as follows: at time zero, there is one vertex $\varnothing$, which reproduces according to $(\xi_\varnothing, W_\varnothing)$. Thereafter, at times corresponding to points in $\xi_\varnothing$, descendants of $\varnothing$ are formed, which in turn produce offspring according to the same law. A crucial aspect of the study of CMJ processes is the *characteristics* ϕ_x associated to each element $x \in \mathcal{U}$. For $x \in \mathcal{U}$, let $\mathcal{U}_x := \{xu : u \in \mathcal{U}\}$. Then the processes ϕ_x are identically distributed, non-negative stochastic processes on the space $(\Omega, \Sigma, \mathbb{P})$ associated with individuals x, which may depend on $(\xi_z, W_z)_{z\in\mathcal{U}_x}$. Intuitively, these are processes that track 'characteristics' not only of the individual x, but also of its potential offspring $\{xy : y \in \mathcal{U}\}$. We then define the *general branching process counted with characteristic* as

$$Z^\phi(t) := \sum_{x\in\mathcal{U} : \sigma_x \leq t} \phi_x\big(t - \sigma_x\big);$$

thus this function keeps a 'score' of characteristics of individuals in the family tree associated with the process up to time t.

Let ν be the intensity measure of ξ, that is, $\nu(B) := \mathbb{E}[\xi(B)]$ for measurable sets $B \subseteq \mathbb{R}_+$. A crucial parameter in the study of CMJ processes is the *Malthusian parameter* α, which is defined as the solution (if it exists) of

$$\mathbb{E}\left[\int_0^\infty e^{-\alpha u}\xi(\mathrm{d}u)\right] = 1.$$

Assume that ν is not supported on any lattice, i.e., for any $h > 0$, $\mathrm{Supp}(\nu) \subsetneq \{0, h, 2h, \ldots\}$, and that the first moment of $e^{-\alpha u}\nu(\mathrm{d}u)$ is finite, i.e., $\int_0^\infty ue^{-\alpha u}\nu(\mathrm{d}u) < \infty$. Nerman [30] proved the following theorem.

Theorem 2.3. ([30, Theorem 6.3]) *Suppose that there exists* $\lambda < \alpha$ *satisfying*

$$\mathbb{E}\left[\int_0^\infty e^{-\lambda s}\xi(\mathrm{d}s)\right] < \infty. \tag{16}$$

Then, for any two càdlàg characteristics $\phi^{(1)}, \phi^{(2)}$ *such that* $\mathbb{E}\big[\sup_{t\geq 0} e^{-\lambda t}\phi^{(i)}(t)\big] < \infty$, $i = 1, 2$, *we have*

$$\lim_{t\to\infty} \frac{Z^{\phi^{(1)}}(t)}{Z^{\phi^{(2)}}(t)} = \frac{\int_0^\infty \mathrm{e}^{-\alpha s}\mathbb{E}\big[\phi^{(1)}(s)\big]\mathrm{d}s}{\int_0^\infty \mathrm{e}^{-\alpha s}\mathbb{E}\big[\phi^{(2)}(s)\big]\mathrm{d}s},$$

almost surely on the event $\{|\mathbb{T}_t| \to \infty\}$.

Recall the definition of ρ as the point process associated with the jumps in the process Y defined in (9). Then the continuous-time model outlined in Section 2.1 is a CMJ process having ρ as its associated random point process and weight W. In this case, the Malthusian parameter is given by α in (12), and moreover, Condition **C1** implies that the first moment $\int_0^\infty te^{-\alpha t}\hat{\rho}_\mu(\mathrm{d}t)$ is finite.

Theorem 2.1 is now an immediate application of Theorem 2.3.

Proof of Theorem 2.1. Consider the continuous-time branching process outlined in Section 2.1 and denote by $\sigma'_1 < \sigma'_2 \cdots$ the times of births of individuals in the process. Then $\mathcal{T}_n$ has the same distribution as the family tree $\mathbb{T}_{\sigma'_n}$. For any measurable set $B \subseteq \mathbb{R}$, define the characteristics $\phi^{(1)}(t) = \mathbf{1}_{\{Y(t)=k\ell, W\in B\}}$ and $\phi^{(2)}(t) = \mathbf{1}_{\{t\geq 0\}}$, where W denotes the weight of the process Y. Note that $Z^{\phi^{(1)}}(t)$ is the number of individuals with $k\ell$ offspring and weight belonging to B up to time t, while $Z^{\phi^{(2)}}(t) = |\mathbb{T}_t|$. Thus,

$$\lim_{t\to\infty} \frac{Z^{\phi^{(1)}}(t)}{Z^{\phi^{(2)}}(t)} = \lim_{t\to\infty} \frac{N_k(t, B)}{\ell t}.$$

Note that both $\phi^{(1)}(t)$ and $\phi^{(2)}(t)$ are càdlàg and bounded, and moreover, Condition **C1** implies that (16) is satisfied. In addition, the assumption that $f(0, W) > 0$ almost surely implies that $|\mathbb{T}_t| \to \infty$ almost surely. Thus, by applying Theorem 2.3, we have

$$\begin{aligned}\lim_{t\to\infty} \frac{Z^{\phi^{(1)}}(t)}{Z^{\phi^{(2)}}(t)} &= \alpha \int_0^\infty \mathrm{e}^{-\alpha s}\mathbb{E}\big[\mathbf{1}_{\{Y(s)=k\ell, W\in B\}}\big]\mathrm{d}s \\ &= \mathbb{E}\big[\mathbb{E}_W\big[\big(e^{-\alpha\tau_k} - e^{-\alpha\tau_{k+1}}\big)\big]\mathbf{1}_B(W)\big],\end{aligned} \tag{17}$$

where the last equality follows from Fubini's theorem and we recall that τ_k is the time of the kth event in the process $Y_W(t)$. Now, since, when $W = w$, τ_k is distributed as a sum of independent exponentially distributed random variables with rates $f(0, w), f(1, w) \ldots$, we have

$$\mathbb{E}\big[\mathbb{E}_W\big[e^{-\alpha\tau_k}\big]\mathbf{1}_B(W)\big] = \mathbb{E}\left[\left(\prod_{i=0}^{k-1} \frac{f(i, W)}{f(i, W) + \alpha}\right)\mathbf{1}_B(W)\right]. \tag{18}$$

The result follows from combining (17) and (18). □

Remark 2.4. As noted by the authors of [33], Theorem 2.3 can be applied to deduce a number of other properties of the tree; in particular, the analogue of [33, Theorem 1] applies in this case as well.

2.3. A strong law for the partition function

We can also apply Theorem 2.3 to show that the Malthusian parameter α emerges as the almost sure limit of the partition function, under certain conditions on the fitness function f.

Theorem 2.4. *Let $(\mathcal{T}_t)_{t\geq 0}$ be a (μ, f, ℓ)-RIF tree satisfying C1 with Malthusian parameter α. Moreover, assume that there exists a constant $C < \alpha$ and a non-negative function φ with $\mathbb{E}[\varphi(W)] < \infty$ such that, for all $k \in \mathbb{N}_0$, $f(k, W) \leq Ck + \varphi(W)$ almost surely. Then, almost surely,*

$$\frac{\mathcal{Z}_t}{t} \xrightarrow{t\to\infty} \alpha.$$

To apply Theorem 2.3, we need to bound $\mathbb{E}\big[\sup_{t\geq 0} e^{-\lambda t}\phi^{(1)}(t)\big]$ for an appropriate choice of $\lambda < \alpha$ and characteristic $\phi^{(1)}$ that tracks the evolution of the partition function associated with the process. In order to do so, using the assumptions on $f(i, W)$, we will couple the process Y defined in (9) with an appropriate pure birth process $(\mathcal{Y}(t))_{t\geq 0}$ (Lemma 2.3) and apply Doob's maximal inequality to a martingale associated with $(\mathcal{Y}(t))_{t\geq 0}$ (Lemma 2.2). As we will see, our choice of λ will be given by C, and this is the reason for the assumption that $C < \alpha$ in Theorem 2.4.

In order to define $\mathcal{Y}(t)$, first sample a weight W and set $\mathcal{Y}(0) = 0$. Then, if $\mathbb{P}_w$ denotes the probability measure associated with the process when the weight is w, define the rates so that

$$\mathbb{P}_w(\mathcal{Y}(t+h) = k+1 \mid \mathcal{Y}(t) = k) = (Ck + \varphi(w))h + o(h).$$

We also let $\mathcal{Y}_w$ denote the process with the same transition rates, but deterministic weight w.

It will be beneficial to state a more general result, about pure birth processes $(\mathcal{X}(t))_{t\geq 0}$ with linear rates, from the paper by Holmgren and Janson [20]. For brevity, we adapt the notation and only include some specific statements from both theorems.

Lemma 2.1. ([20, Theorem A.6 and Theorem A.7]) *Let $(\mathcal{X}(t))_{t\geq 0}$ be a pure birth process with $\mathcal{X}(0) = x_0$ and rates such that*

$$\mathbb{P}(\mathcal{X}(t+h) = k+1 \mid \mathcal{X}(t) = k) = (c_1 k + c_2)h + o(h),$$

for some constants $c_1, c_2 > 0$. Then, for each $t \geq 0$,

$$\mathbb{E}[\mathcal{X}(t)] = \left(x_0 + \frac{c_2}{c_1}\right) e^{c_1 t} - \frac{c_2}{c_1}. \tag{19}$$

Moreover, if $x_0 = 0$, the probability generating function is given by

$$\mathbb{E}\left[z^{\mathcal{X}(t)}\right] = \left(\frac{e^{-c_1 t}}{1 - z\left(1 - e^{-c_1 t}\right)}\right)^{c_2/c_1}. \tag{20}$$

Finally, we will require Lemma 2.2 and Lemma 2.3.

Lemma 2.2. *For any $w > 0$, the process $(e^{-Ct}\,(\mathcal{Y}_w(t) + \varphi(w)/C)\,)_{t\geq 0}$ is a martingale with respect to its natural filtration $(\mathcal{F}_t)_{t\geq 0}$. Moreover,*

$$\mathbb{E}\left[\sup_{t\geq 0}\left(e^{-Ct}\mathcal{Y}(t)\right)\right] < \infty.$$

Proof. The process $(\mathcal{Y}_w(t))_{t\geq 0}$ is a pure birth process satisfying the assumptions of Lemma 2.1, with $c_1 = C$ and $c_2 = \varphi(w)$. Therefore, by (19) and the Markov property, for any $t > s > 0$ we have

$$\mathbb{E}[\mathcal{Y}_w(t) \mid \mathcal{F}_s] = \mathbb{E}[\mathcal{Y}_w(t) \mid \mathcal{Y}_w(s)] = \left(\mathcal{Y}_w(s) + \frac{\varphi(w)}{C}\right) e^{C(t-s)} - \frac{\varphi(w)}{C},$$

which implies the martingale statement.

Moreover, applying (20) for the probability generating function, differentiating twice and evaluating at $z = 1$, we obtain

$$\mathbb{E}[\mathcal{Y}_w(t)\,(\mathcal{Y}_w(t) - 1)] = \frac{\varphi(w)\,(C + \varphi(w))}{C^2}\left(e^{Ct} - 1\right)^2,$$

and thus after some manipulations, we find that for all $t \geq 0$

$$\mathbb{E}\left[e^{-2Ct}\,(\mathcal{Y}_w(t) + \varphi(w)/C)^2\right] \leq \frac{\varphi(w)^2}{C^2} + \frac{\varphi(w)}{C}\left(1 - e^{-Ct}\right).$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

Combining this L^2 quadratic bound with Doob's maximal inequality, we have

$$\mathbb{E}\left[\sup_{t\geq 0}\left(e^{-Ct}\mathcal{Y}_w(t)\right)\right]\leq \mathbb{E}\left[\sup_{t\geq 0}\left(e^{-Ct}\left(\mathcal{Y}_w(t)+\varphi(w)/C\right)\right)\right]$$
$$\leq A+B\varphi(w),$$

for constants A, B depending only on C. Thus,

$$\mathbb{E}\left[\sup_{t\geq 0}\left(e^{-Ct}\mathcal{Y}(t)\right)\right]=\mathbb{E}\left[\sup_{t\geq 0}\left(e^{-Ct}\mathcal{Y}_W(t)\right)\right]\leq A+B\mathbb{E}[\varphi(W)]<\infty. \qquad \square$$

Lemma 2.3. *Recall the definition of Y in (9) and assume that there exists a constant $C<\alpha$ and a non-negative function φ with $\mathbb{E}[\varphi(W)]<\infty$ such that, for all $k\in\mathbb{N}_0$, $f(k, W)\leq Ck+\varphi(W)$ almost surely. Then there exists a coupling $(\hat{Y}(t), \hat{\mathcal{Y}}(t))_{t\geq 0}$ of $(Y(t))_{t\geq 0}$ and $(\mathcal{Y}(t))_{t\geq 0}$ such that, for all $t\geq 0$,*

$$\hat{Y}(t)\leq \ell\cdot\hat{\mathcal{Y}}(t).$$

In the following proof, we denote by $\mathrm{Exp}(r)$ the exponential distribution with parameter r.

Proof. First, we sample $\hat{W}$ from μ and use this as a common weight for $\hat{Y}$ and $\hat{\mathcal{Y}}$. Now, let $(\varsigma_i)_{i\geq 0}$ be independent $\mathrm{Exp}\big(f(i, \hat{W})\big)$-distributed random variables. Then, for all $k>0$, set $\hat{\tau}_k=\sum_{i=0}^{k-1}\varsigma_i$ and

$$\hat{Y}(t)=\sum_{k=1}^{\infty}k\ell\mathbf{1}_{\{\hat{\tau}_k\leq t<\hat{\tau}_{k+1}\}}.$$

The ς_i can be interpreted as the intermittent time between jumps from state i to $i+\ell$. For all $t>0$ construct the jump times of $(\hat{\mathcal{Y}}(t))_{t\geq 0}$ iteratively as follows:

- Note that by assumption $f(0, \hat{W})\leq\varphi(\hat{W})$. Let $e_0\sim\mathrm{Exp}\big(\varphi(\hat{W})-f(0, \hat{W})\big)$ and set $\varsigma_0'=\min\{e_0, \varsigma_0\}$. We may interpret ς_0' as the time for $\hat{\mathcal{Y}}$ to jump from 0 to 1.
- Given $\varsigma_0', \dots, \varsigma_j'$, let $q_j:=\sum_{i=0}^{j}\varsigma_i'$ and define $m_j:=\hat{Y}(q_j)/\ell$, i.e., the value of $\hat{Y}/\ell$ once $\hat{\mathcal{Y}}$ has reached $j+1$. Assume inductively that $m_j\leq j+1$ and set

$$e_{j+1}\sim\mathrm{Exp}\Big(C(j+1)+\varphi(\hat{W})-f\big(m_j, \hat{W}\big)\Big) \quad \text{and} \quad \varsigma_{j+1}'=\min\big\{e_j, \varsigma_{m_j}\big\}.$$

Observe that, since $\varsigma_{j+1}'\leq\varsigma_{m_j+1}$, we have $m_{j+1}\leq j+2$, so we may iterate this procedure.

It is clear that $(\hat{Y}(t))_{t\geq 0}$ is distributed like $(Y(t))_{t\geq 0}$, and using the properties of the exponential distribution one readily confirms that $(\hat{\mathcal{Y}}(t))_{t\geq 0}$ is distributed like $(\mathcal{Y}(t))_{t\geq 0}$. Finally, the desired inequality follows from the fact that $\hat{\mathcal{Y}}(t)$ always jumps before or at the same time as $\hat{Y}(t)$. $\square$

Proof of Theorem 2.4. Consider the continuous-time embedding of the (μ, f, ℓ)-RIF tree and define the characteristics $\phi^{(1)}(t):=\sum_{k=0}^{\infty}f(k, W)\mathbf{1}_{\{Y(t)=k\ell\}}$ and $\phi^{(2)}(t):=\mathbf{1}_{\{t\geq 0\}}$. Recall that we denote by $(\tau_i)_{i\geq 1}$ the times of the jumps in Y and that, for all $k\geq 0$, $f(k, W)\leq Ck+\varphi(W)$. Then, by Lemma 2.3, Lemma 2.2, and the assumptions of the theorem,

$$\mathbb{E}\left[\sup_{t\geq 0}\left(e^{-Ct}\phi^{(1)}(t)\right)\right]\overset{\text{Lem. 2.3}}{\leq}\mathbb{E}\left[\sup_{t\geq 0}\left(e^{-Ct}\left(C\mathcal{Y}_W(t)+\varphi(W)\right)\right)\right]\overset{\text{Lem. 2.2}}{<}\infty.$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

Now, in this case $Z^{\phi^{(1)}}(t)$ is the total sum of fitnesses of individuals born up to time t, while $Z^{\phi^{(2)}}(t) = |\mathcal{T}_t|$. Thus, by Theorem 2.3 and Fubini's theorem in the second equality, almost surely we have

$$\lim_{n\to\infty} \frac{\mathcal{Z}_n}{\ell n} = \alpha \int_0^\infty e^{-\alpha s} \mathbb{E}\left[\sum_{k=0}^{\infty} f(k, W) \mathbf{1}_{\{Y(s)=k\ell\}}\right] \mathrm{d}s = \mathbb{E}\left[\sum_{k=0}^{\infty} f(k, W) \left(e^{-\alpha \tau_k} - e^{-\alpha \tau_{k+1}}\right)\right] \tag{21}$$

$$= \mathbb{E}\left[\sum_{k=1}^{\infty} \frac{\alpha f(k, W)}{f(k, W)+\alpha} \prod_{i=0}^{k-1} \frac{f(i, W)}{f(i, W)+\alpha}\right].$$

Now, recall that by (12) we have

$$\mathbb{E}\left[\sum_{k=1}^{\infty} \frac{f(k, W)}{f(k, W)+\alpha} \prod_{i=0}^{k-1} \frac{f(i, W)}{f(i, W)+\alpha}\right] = \frac{1}{\ell},$$

and combining this with (21) proves the result. □

2.4. Examples of applications of Theorem 2.1

2.4.1. *Weighted Cayley trees.* Consider the model where $f(k, W) = 0$ for $k \geq 1$ and $f(0, W) = g(W)$. Thus, at each step, a vertex with degree 0 is chosen and produces ℓ children, and thus this model produces an $(\ell + 1)$-*Cayley* tree, i.e., a tree in which each node that is not a leaf has degree $\ell + 1$. Without loss of generality, by considering the pushforward of μ under g if necessary, we may assume that $g(W) = W$. In this case, $\hat{\rho}_\mu(\lambda) = \ell \cdot \mathbb{E}\left[\frac{W}{W+\lambda}\right]$ and thus **C1** is satisfied as long as $\ell \geq 2$. Thus, $p_k^\alpha(B) = 0$ for all $k \geq 2$, and

$$p_0(B) = \mathbb{E}\left[\frac{\alpha}{W+\alpha} \mathbf{1}_B(W)\right], \quad p_1(B) = \mathbb{E}\left[\frac{W}{W+\alpha} \mathbf{1}_B(W)\right].$$

This rigorously confirms a result of Bianconi [4]. Note, however, that in [4] α is described as the almost sure limit of the partition function, and we may only apply Theorem 2.4 under the assumption that $\mathbb{E}[W] < \infty$.

In the notation of [4], the weights W are called 'energies', using the symbol ϵ, the function $g(\epsilon) := e^{\beta\epsilon}$, where $\beta > 0$ is a parameter of the model, and $\alpha := e^{\beta \mu_F}$ is described as the limit of the partition function. Thus, the proportion of vertices with out-degree 0 with 'energy' belonging to some measurable set B is

$$\mathbb{E}\left[\frac{1}{e^{\beta(\epsilon-\mu_F)}+1} \mathbf{1}_B(W)\right],$$

which is known as a *Fermi–Dirac* distribution in physics.

2.4.2. *Weighted random recursive trees.* In the case that $f(k, W) = W$, we obtain a model of *weighted random recursive trees* with independent weights and **C1** is satisfied with $\alpha = \mathbb{E}[W]$ provided $\mathbb{E}[W] < \infty$. Theorem 2.1 then implies that

$$\frac{N_k(t, B)}{\ell t} \xrightarrow{t\to\infty} \mathbb{E}\left[\frac{\ell \mathbb{E}[W] W^k}{(W + \ell \mathbb{E}[W])^{k+1}} \mathbf{1}_B(W)\right],$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

almost surely. This was observed in the case $\ell = 1$ by the authors of [17, Proposition 3]. Note also that in this case Theorem 2.4 coincides with the usual strong law of large numbers.

The weighted random recursive tree has a natural generalisation to affine fitness functions. This is the topic of the next section.

3. Generalised preferential attachment trees with fitnesses

In this section, we study (μ, f, ℓ)-RIF trees in the specific case when the function f takes an affine form, that is, $f(i, W) = ig(W) + h(W)$, for positive, measurable functions g, h. We call this particular case of the model a *generalised preferential attachment tree with fitness* (which we abbreviate as a GPAF-tree). The affine form of this model means that it is tractable to apply the coupling methods outlined in Section 3.2.3, when Condition **C1** fails. Moreover, this model is general enough to be an extension not only of the weighted random recursive tree, but also of the additive and multiplicative models studied in [3, 8].

Below, in Section 3.1, we apply the theory of the previous section to this model when **C1** is satisfied. In satisfied. In Section 3.2, we analyse the model when Condition **C1** fails by having $m(\lambda, \mathbb{R}_+) \leq 1$ for all $\lambda > 0$ such that $m(\lambda, \mathbb{R}_+) < \infty$, stating and proving Theorem 3.1. Then, in Section 3.3, we analyse the model when Condition **C1** fails by having $m(\lambda, \mathbb{R}_+) = \infty$ for all $\lambda > 0$, stating and proving Theorem 3.2.

Note that in this section, we formulate our results in terms of functions g and h of a random variable W taking values in $\mathbb{R}_+$. However, in the vein of Remark 2.3, we expect these results to extend to cases where g and h are replaced respectively with possibly correlated random variables W_1 and W_2 assigned to a given vertex, a model first analysed in [16]. In this case, the coupling technique applied in Section 3.2.3 needs to be adjusted accordingly, with appropriate 'truncations' of the vector (W_1, W_2).

3.1. When the GPAF-tree satisfies Condition C1

In the context of the GPAF-tree, Condition **C1** states that there exists $\lambda > 0$ such that

$$m(\lambda, \mathbb{R}_+) = \ell \cdot \mathbb{E}\left[\sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{g(W)i + h(W)}{g(W)i + h(W) + \lambda}\right] > 1.$$

First recall the definition of the birth process Y from (9) in Section 2, with $f(k, W) = g(W)k + h(W)$. By applying (19) from Lemma 2.1 and the initial condition $Y(0) = 0$, for any $w \in \mathbb{R}_+$ we have

$$\mathbb{E}_w\left[Y(t)\right] = \left(\frac{h(w)}{g(w)}\right) e^{\ell g(w)t} - \frac{h(w)}{g(w)}.$$

Now, (10) and (11) in Section 2 showed that

$$\ell \cdot \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{g(W)i + h(W)}{g(W)i + h(W) + \lambda} = \int_0^{\infty} \lambda e^{-\lambda s} \mathbb{E}_w[Y(s)] \mathrm{d}s = \begin{cases} \frac{h(w)}{\lambda/\ell - g(w)} & \text{if } \lambda/\ell > g(w); \\ \infty & \text{otherwise.} \end{cases} \quad (22)$$

For a measurable function $g : \mathbb{R}_+ \to \mathbb{R}_+$ we define $\operatorname{ess\,sup}(g)$ by

$$\operatorname{ess\,sup}(g) := \inf\left\{a \in \mathbb{R}_+ : \mu(\{x : g(x) > a\}) = 0\right\}.$$

Therefore by (22), for $\lambda \geq \ell \cdot \operatorname{ess\,sup}(g)$ we have

$$m(\lambda, \mathbb{R}_+) = \mathbb{E}\left[\frac{h(W)}{\lambda/\ell - g(W)}\right],$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

while if $\lambda < \ell \cdot \operatorname{ess\,sup}(g)$ we have $m(\lambda, \mathbb{R}_+) = \infty$. Thus, Condition **C1** is satisfied if $\operatorname{ess\,sup}(g) < \infty$, $\mathbb{E}[h(W)] < \infty$, and for some $\lambda \geq \ell \cdot \operatorname{ess\,sup}(g)$

$$1 < \mathbb{E}\left[\frac{h(W)}{\lambda/\ell - g(W)}\right] < \infty.$$

As a result, the Malthusian parameter α appearing in Condition **C1** is given by the unique $\alpha > 0$ such that

$$\mathbb{E}\left[\frac{h(W)}{\alpha/\ell - g(W)}\right] = 1. \tag{23}$$

Note that the parameter ℓ in the model has the effect of re-scaling the Malthusian parameter α. Also, since $\alpha \geq \ell \cdot \operatorname{ess\,sup}(g)$, if $\mathbb{E}[h(W)] < \infty$, Theorem 2.4 applies and α may also be interpreted as the almost sure limit of the partition function associated with the process. Now, in the context of this model, the limiting value $p_k^{\alpha}(\cdot)$ from Theorem 2.1 is such that

$$p_k^{\alpha}(B) = \mathbb{E}\left[\frac{\alpha}{g(W)k + h(W) + \alpha} \prod_{i=0}^{k-1} \frac{g(W)i + h(W)}{g(W)i + h(W) + \alpha} \mathbf{1}_B(W)\right]. \tag{24}$$

Now, recall Stirling's approximation, which states that

$$\Gamma(z) = (1 + O(1/z))\, z^{z-\frac{1}{2}} e^{-z} \tag{25}$$

as $z \to \infty$. If $g(W) > 0$ on B, by dividing the numerator and denominator of terms inside the product in (24), we obtain a ratio of gamma functions. Thus, by applying Stirling's approximation, on any measurable set B on which g, h are bounded, we have

$$p_k^{\alpha}(B) = (1 + O(1/k))\, \mathbb{E}\left[c_B k^{-\left(1+\frac{\alpha}{g(W)}\right)} \mathbf{1}_B(W)\right],$$

where c_B, which comes from the term outside the product in (24), depends on g and h but not k. Thus, the distribution of $(p_k^{\alpha}(B))_{k\in\mathbb{N}_0}$ follows what one might describe as an 'averaged' power law. Moreover, in the case that $\ell = 1$, we have $\alpha \geq \operatorname{ess\,sup}(g)$; thus,

$$\mathbb{E}\left[c_B k^{-\left(1+\frac{\alpha}{g(W)}\right)} \mathbf{1}_B(W)\right] \geq c' k^{-2}$$

for some $c' > 0$. It has been observed that real-world complex networks have power law degree distributions where the observed power law exponent lies between 2 and 3 (see, for example, [37]). Note that by (23), α depends on both h and g, so that keeping g fixed and making h smaller has the effect of reducing the exponent of the power law.

In the remainder of this section we set $\ell = 1$, for brevity. The arguments may be adapted in a similar manner to the case $\ell > 1$.

3.2. A condensation phenomenon in the GPAF-tree when Condition C1 fails Condition C1 fails

Recall that, in the GPAF-tree, if $\lambda \geq \operatorname{ess\,sup}(g)$ we have

$$m(\lambda, \mathbb{R}_+) = \mathbb{E}\left[\frac{h(W)}{\lambda - g(W)}\right],$$

and if $\lambda < \text{ess sup}(g)$ we have $m(\lambda, \mathbb{R}_+) = \infty$. If we define

$$\Lambda := \{\lambda > 0 : m(\lambda, \mathbb{R}_+) < \infty\},$$

in this subsection, we consider the case that the GPAF-tree fails to satisfy Condition **C1** by having $m(\lambda, \mathbb{R}_+) \leq 1$ for all $\lambda \in \Lambda$. We show that in this case the GPAF-tree satisfies a formula for the degree distribution of the same form as (3). Moreover, if $\lambda^* := \inf(\Lambda)$ and $m(\lambda^*, \mathbb{R}_+) < 1$, this model exhibits a condensation phenomenon, as described in Theorem 3.1. We remark that such results have been proved for the case of the preferential attachment tree with multiplicative fitness, i.e., the case $h \equiv g$, in [14], in a more general framework; that is to say, encompassing other models apart from a tree.

In Section 3.2.1 we state our main result, Theorem 3.1; we discuss interesting implications in Section 3.2.2. In Section 3.2.3 we state and prove Lemma 3.1, which is the crucial tool used in proofs of the theorem. The proof of Theorem 3.1 is deferred to Section 3.2.4.

Remark 3.1. If $\mathcal{T}$ does not satisfy **C1** and $\mathbb{E}[h(W)] < \infty$, we must have $\mu(\{x : g(x) = \text{ess sup} g\}) = 0$, since otherwise, for each $\lambda > \lambda^*$, we have $m(\lambda, \mathbb{R}_+) < \infty$, and by monotone convergence $\lim_{\lambda \downarrow \lambda^*} m(\lambda, \mathbb{R}_+) \uparrow \infty$.

3.2.1. *Theorem 3.1: condensation in the GPAF-tree.* Our main result in this subsection is the following theorem, which demonstrates the possibility of condensation in this model. Recall that in this section, we have $\lambda^* = \text{ess sup}(g)$. We then define the following family of sets of positive μ-measure, depending on a parameter $\varepsilon > 0$:

$$\mathcal{M}_\varepsilon := \left\{x : g(x) \geq \lambda^* - \varepsilon\right\}. \tag{26}$$

Theorem 3.1. *Suppose $\mathcal{T} = (\mathcal{T}_t)_{t \geq 0}$ is a GPAF-tree, with associated functions g, h, where g is bounded, $\mathbb{E}[h(W)] < \infty$, and Condition C1 fails. Then we have the following assertions:*

1. *For any measurable set B such that for some $\varepsilon > 0$ we have $B \subseteq \mathcal{M}_\varepsilon^c$,*

$$\frac{\Xi(t, B)}{\ell t} \xrightarrow{t \to \infty} \mathbb{E}\left[\frac{h(W)}{\lambda^* - g(W)} \mathbf{1}_B(W)\right], \quad \textit{almost surely.}$$

In particular, if

$$\mathbb{E}\left[\frac{h(W)}{g(w^*) - g(W)}\right] < 1,$$

for $\varepsilon > 0$ sufficiently small we have

$$\frac{\Xi(t, \mathcal{M}_\varepsilon)}{\ell t} \xrightarrow{t \to \infty} 1 - \mathbb{E}\left[\frac{h(W)}{\lambda^* - g(W)} \mathbf{1}_{\mathcal{M}_\varepsilon^c}(W)\right] > \mathbb{E}\left[\frac{h(W)}{\lambda^* - g(W)} \mathbf{1}_{\mathcal{M}_\varepsilon}(W)\right] = m(\lambda^*, \mathcal{M}_\varepsilon),$$

so that this model exhibits a condensation phenomenon, as described before Conjecture 1.1 in Section 1.3.

2. *For any measurable set $B \subseteq \mathbb{R}_+$, almost surely we have*

$$\frac{N_k(t, B)}{t} \xrightarrow{t \to \infty} \mathbb{E}\left[\frac{\lambda^*}{g(W)k + h(W) + \lambda^*} \prod_{i=0}^{k-1} \frac{g(W)i + h(W)}{g(W)i + h(W) + \lambda^*} \mathbf{1}_B(W)\right] = p_k^{\lambda^*}(B).$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

3. *The partition function satisfies*

$$\frac{\mathcal{Z}_t}{t} \xrightarrow{t\to\infty} \lambda^*, \quad \text{almost surely.}$$

Suppose that $w^* := \sup(\mathrm{Supp}(\mu)) < \infty$, $\mathrm{Supp}(\mu) \subseteq [0, w^*]$, and g is increasing. Define the measure $\pi(\cdot)$ so that, for any measurable set B,

$$\pi(B) = \mathbb{E}\left[\frac{h(W)}{g(w^*) - g(W)}\mathbf{1}_B(W)\right] + \left(1 - \mathbb{E}\left[\frac{h(W)}{g(w^*) - g(W)}\right]\right)\delta_{w^*}(B).$$

Then Assertion 1 of Theorem 3.1 leads to the following result.

Corollary 3.1. *Under the above assumptions, with respect to the weak topology,*

$$\frac{\Xi(t, \cdot)}{\ell t} \xrightarrow{t\to\infty} \pi(\cdot), \quad \textit{almost surely.}$$

Remark 3.2. Corollary 3.1 is the form in which condensation results usually appear in the literature, showing that the limit of the sequence $\frac{\Xi(t,\cdot)}{\ell t}$ is no longer absolutely continuous with respect to μ. In this regard, Corollary 3.1 generalises the case $f(i, W) = (i+1)W$ which has already been proved in [8].

Proof of Corollary 3.1. In view of the portmanteau theorem, it suffices to prove that, almost surely, for any open set O of $[0, w^*]$ we have

$$\liminf_{t\to\infty} \frac{\Xi(t, O)}{\ell t} \geq \pi(O).$$

Now, it is well known that there exists a countable family of measurable sets $D_1, D_2, \ldots$ such that any open subset of $[0, w^*]$ may be expressed as a countable disjoint union of elements of this family. For example, one may take the set of all *dyadic intervals*, with endpoints of the form $j \cdot 2^{-n}w^*, (j+1)\cdot 2^{-n}w^*$, where $j, n \in \mathbb{N}_0$. Fix such a family. Now, by Assertion 1 of Theorem 3.1, it is the case that, almost surely,

$$\lim_{t\to\infty} \frac{\Xi(t, S)}{t} = \pi(S) \quad \forall S \in \mathcal{C},$$

where $\mathcal{C}$ is the countable collection of sets

$$\mathcal{C} := \left\{D_i \cap \mathcal{M}_{1/j}^c, \ \mathcal{M}_{1/j} : i, j \in \mathbb{N}\right\}.$$

Now, let O be an arbitrary open set. First, suppose that $w^* \notin O$. Then, for a pairwise disjoint collection $D_{i_1}, D_{i_2}, \ldots$ such that $O = \bigcup_{\ell\in\mathbb{N}} D_{i_\ell}$, for each $j, k \in \mathbb{N}$ we have

$$\liminf_{t\to\infty} \frac{\Xi(t, O)}{t} \geq \sum_{\ell=1}^{k} \liminf_{t\to\infty} \frac{\Xi\left(t, D_{i_\ell} \cap \mathcal{M}_{1/j}^c\right)}{t} \geq \sum_{\ell=1}^{k} \pi\left(D_{i_\ell} \cap \mathcal{M}_{1/j}^c\right).$$

Taking limits in j and k, the right-hand side converges to $\pi(O)$, as required. On the other hand, if $w^* \in O$, since g is increasing, for each $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that $\mathcal{M}_\varepsilon \subseteq [w^* - \delta, w^*]$. Therefore, because O is open, for all j sufficiently large, we have $\mathcal{M}_{1/j} \subseteq O$. But then, for a pairwise disjoint collection $D_{i_1}, D_{i_2}, \ldots$ such that $O = \bigcup_{\ell\in\mathbb{N}} D_{i_\ell}$, we have

$$O = \mathcal{M}_{1/j} \cup \left(\bigcup_{\ell\in\mathbb{N}} D_{i_\ell} \cap \mathcal{M}_{1/j}^c\right).$$

Therefore,

$$\liminf_{t\to\infty} \frac{\Xi(t, O)}{t} \geq \liminf_{t\to\infty} \frac{\Xi(t, \mathcal{M}_{1/j})}{t} + \sum_{\ell=1}^{k} \liminf_{t\to\infty} \frac{\Xi\big(t, D_{i_\ell} \cap \mathcal{M}_{1/j}^c\big)}{t}$$
$$\geq \pi(\mathcal{M}_{1/j}) + \sum_{\ell=1}^{k} \pi\big(D_{i_\ell} \cap \mathcal{M}_{1/j}^c\big),$$

so that, by again taking limits in j and k, the right-hand side converges to $\pi(O)$. The result follows. □

3.2.2. *Some interesting implications of the condensation phenomenon.* The condensation result in Theorem 3.1 has interesting implications for the GPAF-tree. Informally, the parameter $g(w)$ measures the extent to which the 'popularity' of a vertex with weight w is reinforced by the number of its neighbours, while the parameter $h(w)$ represents its 'initial popularity'. The condensation phenomenon then depends on both μ and h, in the sense that condensation occurs if vertices of high weight are 'rare enough' and the initial popularity is 'low enough'. More precisely, if $\mathbb{E}[h(W)] < \infty$ and $\lambda^* < \infty$ we see that the tree displays the following interesting features:

1. By Remark 3.1, if μ and g are such that $\mathbb{E}\Big[\frac{1}{\lambda^* - g(W)}\Big] = \infty$, Condition **C1** is satisfied in this model, and thus, the model does not demonstrate a condensation phenomenon.

2. Otherwise, if μ and g are such that $\mathbb{E}\Big[\frac{1}{\lambda^* - g(W)}\Big] = C' < \infty$, then either

$$\mathbb{E}\left[\frac{h(W)}{\lambda^* - g(W)}\right] > 1 \quad \text{or} \quad \mathbb{E}\left[\frac{h(W)}{\lambda^* - g(W)}\right] \leq 1.$$

   Condition **C1** is satisfied in the first case but fails in the second case. However, in the second case, if the inequality is strict, condensation arises. Therefore, for fixed g, condensation in this model arises if we reduce h sufficiently pointwise, for example, by replacing h by $K \cdot h$ where $K < 1/C'$ is a constant.

3. Informally, if one considers h to be a factor representing 'initial popularity' and g to be a factor representing 'reinforcement', the condensation occurs around nodes of maximum 'reinforcement', rather than those of maximal 'initial popularity'. It may even be the case, for example, that h is minimised on the sets $\mathcal{M}_\varepsilon$ where the condensation occurs.

3.2.3. *A coupling lemma.* In order to prove Theorem 3.1, we first prove an additional lemma. For each $\varepsilon > 0$ such that $\varepsilon < \lambda^*$, let $\mathcal{T}^{+\varepsilon} = (\mathcal{T}_t^{+\varepsilon})_{t\geq 0}$ and $\mathcal{T}^{-\varepsilon} = (\mathcal{T}_t^{-\varepsilon})_{t\geq 0}$ denote GPAF-trees with the same set of weights, but with the function g modified to $g_{+\varepsilon}$ and $g_{-\varepsilon}$ respectively, on the set $\mathcal{M}_\varepsilon$ from (26), with

$$g_{+\varepsilon} := g\mathbf{1}_{\mathcal{M}_\varepsilon^c} + \lambda^*\mathbf{1}_{\mathcal{M}_\varepsilon} \quad \text{and} \quad g_{-\varepsilon} := g\mathbf{1}_{\mathcal{M}^c} + (\lambda^* - \varepsilon)\mathbf{1}_{\mathcal{M}_\varepsilon}.$$

The motivation behind these choices of $\mathcal{T}^{+\varepsilon}$ and $\mathcal{T}^{-\varepsilon}$ is that, because $\mathcal{T}$ does not satisfy Condition **C1**, $g_{+\varepsilon}$ and $g_{-\varepsilon}$ attain their essential suprema on sets of positive measure. Thus, because $\mathbb{E}[h(W)] < \infty$, by Remark 3.1 these auxiliary trees satisfy Condition **C1**, and we may apply the theorems from Section 2 with regard to these trees. Then, using the fact that these

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

trees provide sufficiently good 'approximations' of the tree $\mathcal{T}$, we may deduce our results by sending ε to 0.

In this vein, let $N^{+\varepsilon}_{\geq k}(t, B)$, $N_{\geq k}(t, B)$, and $N^{-\varepsilon}_{\geq k}(t, B)$ denote the number of vertices with out-degree $\geq k$ and weight belonging to the set B in $\mathcal{T}^{+\varepsilon}_t$, $\mathcal{T}_t$, and $\mathcal{T}^{-\varepsilon}_t$, respectively. In their respective trees, we also denote by $\mathcal{Z}^{+\varepsilon}_t$, $\mathcal{Z}_t$, and $\mathcal{Z}^{-\varepsilon}_t$ the partition functions at time t. Finally, for brevity, we write $f^{(+\varepsilon)}_t(v)$, $f_t(v)$, and $f^{(-\varepsilon)}_t(v)$ for the fitness of a vertex v at time t in each of these models. For example, $f_t(v) = g(W_v)\deg^+(v, \mathcal{T}_t) + h(W_v)$.

Lemma 3.1. *For every $\varepsilon > 0$, there exists a coupling $(\hat{\mathcal{T}}^{+\varepsilon}, \hat{\mathcal{T}}, \hat{\mathcal{T}}^{-\varepsilon})$ of these processes such that, on the coupling, for all $t \in \mathbb{N}_0$, the following hold:*

1. *For all measurable sets $B \subseteq \mathcal{M}^c_\varepsilon$ we have $\Xi^{+\varepsilon}(t, B) \leq \Xi(t, B) \leq \Xi^{-\varepsilon}(t, B)$.*
2. *For all measurable sets $B \subseteq \mathcal{M}^c_\varepsilon$ and $k \in \mathbb{N}_0$, we have*

$$N^{+\varepsilon}_{\geq k}(t, B) \leq N_{\geq k}(t, B) \leq N^{-\varepsilon}_{\geq k}(t, B).$$

3. *We have the inequalities $\mathcal{Z}^{-\varepsilon}_t \leq \mathcal{Z}_t \leq \mathcal{Z}^{+\varepsilon}_t$.*

Proof of Lemma 3.1. We construct the trees having the same sequence of weights $(W_i)_{i\geq 0}$, so that the dynamics of the models are only influenced by differences in the function g in the respective models. Thus, at time 0 each model consists of a single vertex labelled 0 with weight W_0 and having fitness given by $h(W_0)$. Assume, that at the tth time-step,

$$(\hat{\mathcal{T}}^{+\varepsilon}_n)_{0\leq n\leq t} \sim (\mathcal{T}^{+\varepsilon}_n)_{0\leq n\leq t}, \quad (\hat{\mathcal{T}}_n)_{0\leq n\leq t} \sim (\mathcal{T}_n)_{0\leq n\leq t}, \quad \text{and} \quad (\hat{\mathcal{T}}^{-\varepsilon}_n)_{0\leq n\leq t} \sim (\mathcal{T}^{-\varepsilon}_n)_{0\leq n\leq t}.$$

In addition, assume, by induction, that we have $\mathcal{Z}^{-\varepsilon}_t \leq \mathcal{Z}_t \leq \mathcal{Z}^{+\varepsilon}_t$ and for each vertex v with $W_v \in \mathcal{M}^c_\varepsilon$

$$\deg^+(v, \hat{\mathcal{T}}^{+\varepsilon}_t) \leq \deg^+(v, \hat{\mathcal{T}}_t) \leq \deg^+\left(v, \hat{\mathcal{T}}^{-\varepsilon}_t\right). \tag{27}$$

Note that (27) and the fact that the trees have the same number of directed edges imply the first and the second assertions of the lemma up to time t. As a result, for each vertex v with $W_v \in \mathcal{M}^c_\varepsilon$ we have $f^{(+\varepsilon)}_t(v) \leq f_t(v) \leq f^{(-\varepsilon)}_t(v)$. Now, for the $(t + 1)$th step, we do the following:

- Introduce a vertex $t + 1$ with weight W_{t+1} sampled independently from μ.
- Form $\hat{\mathcal{T}}^{-\varepsilon}_{t+1}$ by sampling the parent v of $t + 1$ independently according to the law of $\mathcal{T}^{-\varepsilon}$, i.e., with probability proportional to $f^{(-\varepsilon)}_t(v)$. Then, in order to form $\hat{\mathcal{T}}_{t+1}$, sample an independent uniformly distributed random variable U_1 on [0, 1].

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

  - If
$$U_1 \leq \frac{\mathcal{Z}_t^{-\varepsilon} f_t(v)}{\mathcal{Z}_t f_t^{(-\varepsilon)}(v)}$$
and $W_v \in \mathcal{M}_\varepsilon^c$, select v as the parent of $t+1$ in $\hat{\mathcal{T}}_{t+1}$ as well.
  - Otherwise, form $\hat{\mathcal{T}}_{t+1}$ by selecting the parent v' of $t+1$ with probability proportional to $f_t(v')$ out of all the vertices with weight $W_{v'} \in \mathcal{M}_\varepsilon$.
- Then form $\hat{\mathcal{T}}^{+\varepsilon}_{t+1}$ from $\hat{\mathcal{T}}_{t+1}$ in an identical manner to the way $\hat{\mathcal{T}}_{t+1}$ is formed from $\hat{\mathcal{T}}^{-\varepsilon}$, with another, independent uniform random variable U_2 on $[0, 1]$.

It is clear that $\hat{\mathcal{T}}^{-\varepsilon}_{t+1} \sim \mathcal{T}^{-\varepsilon}_{t+1}$. On the other hand, in $\hat{\mathcal{T}}_{t+1}$ the probability of choosing a parent v of $t+1$ with weight $W_v \in \mathcal{M}_\varepsilon^c$ is

$$\frac{\mathcal{Z}_t^{-\varepsilon} f_t(v)}{\mathcal{Z}_t f_t^{(-\varepsilon)}(v)} \times \frac{f_t^{(-\varepsilon)}(v)}{\mathcal{Z}_t^{-\varepsilon}} = \frac{f_t(v)}{\mathcal{Z}_t},$$

whilst the probability of choosing a parent v' with weight $W_{v'} \in \mathcal{M}_\varepsilon$ is

$$\begin{aligned}
&\frac{f_t(v')}{\sum_{v\,:\,W_v \geq w^*-\varepsilon} f_t(v)} \left(\sum_{v\,:\,W_v < w^*-\varepsilon} \left(1 - \frac{\mathcal{Z}_t^{-\varepsilon} f_t(v)}{\mathcal{Z}_t f_t^{(-\varepsilon)}(v)}\right) \frac{f_t^{(-\varepsilon)}(v)}{\mathcal{Z}_t^{-\varepsilon}} \right) \\
&\quad + \frac{f_t(v')}{\sum_{v\,:\,W_v \geq w^*-\varepsilon} f_t(v)} \left(\sum_{v\,:\,W_v \geq w^*-\varepsilon} \frac{f_t^{(-\varepsilon)}(v)}{\mathcal{Z}_t^{-\varepsilon}} \right) \\
&= \frac{f_t(v')}{\sum_{v\,:\,W_v \geq w^*-\varepsilon} f_t(v)} \left(\sum_{v} \frac{f_t^{(-\varepsilon)}(v)}{\mathcal{Z}_t^{-\varepsilon}} - \sum_{v\,:\,W_v < w^*-\varepsilon} \frac{f_t(v)}{\mathcal{Z}_t} \right) \\
&= \frac{f_t(v')}{\sum_{v\,:\,W_v \geq w^*-\varepsilon} f_t(v)} \left(1 - \frac{\sum_{v\,:\,W_v < w^*-\varepsilon} f_t(v)}{\mathcal{Z}_t} \right) = \frac{f_t(v')}{\mathcal{Z}_t},
\end{aligned}$$

where we use the fact that $\sum_v f_t(v) = \mathcal{Z}_t$. Thus, we have $\hat{\mathcal{T}}_{t+1} \sim \mathcal{T}_{t+1}$. Moreover, either the same vertex is chosen as the parent of $t+1$ in both $\hat{\mathcal{T}}^{-\varepsilon}_{t+1}$ and $\hat{\mathcal{T}}_{t+1}$, or a vertex of weight belonging to $\mathcal{M}_\varepsilon$ is chosen as the parent of $t+1$ in $\hat{\mathcal{T}}_{t+1}$. This implies the left inequality in (27); in addition, when combined with the fact that $g_{-\varepsilon}(W_{t+1}) \leq g(W_{t+1})$, it guarantees that $\mathcal{Z}^{-\varepsilon}_{t+1} \leq \mathcal{Z}_{t+1}$. The proofs of the fact that $\hat{\mathcal{T}}^{+\varepsilon}_{t+1} \sim \mathcal{T}^{+\varepsilon}_{t+1}$, the right inequality in (27), and the fact that $\mathcal{Z}_{t+1} \leq \mathcal{Z}^{+\varepsilon}_{t+1}$ are similar, so we may thus iterate the coupling. □

3.2.4. *Proof of Theorem 3.1.* In order to prove Theorem 3.1, we use the auxiliary GPAF-trees $\mathcal{T}^{+\varepsilon}$ and $\mathcal{T}^{-\varepsilon}$ according to Lemma 3.1.

Proof of Theorem 3.1. For the first assertion, suppose that B is measurable, with $B \subseteq \mathcal{M}_\varepsilon^c$. Then, if we define the corresponding quantities $\Xi^{+\varepsilon}(t, \cdot)$, $\Xi^{-\varepsilon}(t, \cdot)$ associated with $\mathcal{T}^{+\varepsilon}$ and $\mathcal{T}^{-\varepsilon}$, from the coupling in Lemma 3.1, we have

$$\frac{\Xi^{+\varepsilon}(t, B)}{t} \leq \frac{\Xi(t, B)}{t} \leq \frac{\Xi^{-\varepsilon}(t, B)}{t}.$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

Recall that the auxiliary trees $\mathcal{T}^{+\varepsilon}$ and $\mathcal{T}^{-\varepsilon}$ have associated functions $g_{+\varepsilon}$ and $g_{-\varepsilon}$ which attain their maxima on a set of positive measure and thus satisfy Condition **C1**, with Malthusian parameters $\alpha^{(+\varepsilon)} > \lambda^*$ and $\alpha^{(-\varepsilon)} > \lambda^* - \varepsilon$. Moreover, note that, by the definition of $g_{-\varepsilon}$,

$$\mathbb{E}\left[\frac{h(W)}{\lambda^* - g_{-\varepsilon}(W)}\right] \leq \mathbb{E}\left[\frac{h(W)}{\lambda^* - g(W)}\right] \leq 1,$$

so that, recalling (23), $\alpha^{(-\varepsilon)} \leq \lambda^*$. As a result, $\lambda^* - \varepsilon < \alpha^{(-\varepsilon)} \leq \lambda^*$, so $\lim_{\varepsilon \downarrow 0} \alpha^{(-\varepsilon)} = \lambda^*$. Thus, by Lemma 3.1, Theorem 2.2, and dominated convergence, almost surely we have

$$\limsup_{t\to\infty} \frac{\Xi(t, B)}{t} \leq \lim_{\varepsilon\to 0} \mathbb{E}\left[\frac{h(W)}{\alpha^{(-\varepsilon)} - g_{-\varepsilon}(W)} \mathbf{1}_B(W)\right] = \mathbb{E}\left[\frac{h(W)}{\lambda^* - g(W)} \mathbf{1}_B(W)\right].$$

Now, we also have $\lim_{\varepsilon\to 0} \alpha^{(+\varepsilon)} = \lambda^*$. Indeed, suppose for the sake of contradiction that $\lim_{\varepsilon\to 0} \alpha^{(+\varepsilon)} = \alpha' > \lambda^*$. Then, because $\mathbb{E}[h(W)] < \infty$, by dominated convergence we have

$$1 = \lim_{\varepsilon\to 0} \mathbb{E}\left[\frac{h(W)}{\alpha^{(+\varepsilon)} - g_{+\varepsilon}(W)}\right] = \mathbb{E}\left[\frac{h(W)}{\alpha' - g(W)}\right].$$

But then, (23) is satisfied for λ such that $\lambda^* < \lambda < \alpha'$, contradicting the assumption that Condition **C1** fails for $\mathcal{T}$.

It follows that $\lim_{\varepsilon\to 0} \alpha^{(+\varepsilon)} = \lambda^*$, and thus, by Lemma 3.1 and dominated convergence, almost surely we have

$$\limsup_{t\to\infty} \frac{\Xi(t, B)}{t} \leq \lim_{\varepsilon\to 0} \mathbb{E}\left[\frac{h(W)}{\alpha^{(-\varepsilon)} - g_{-\varepsilon}(W)} \mathbf{1}_B(W)\right] = \mathbb{E}\left[\frac{h(W)}{\lambda^* - g(W)} \mathbf{1}_B(W)\right].$$

The first assertion follows.

For the second assertion, given a measurable set B, for each $\varepsilon > 0$, set $B^\varepsilon := B \cap \mathcal{M}_\varepsilon$. Then, by Lemma 3.1, almost surely we have

$$\begin{aligned}
\limsup_{t\to\infty} \frac{N_{\geq k}(t, B)}{t} &\leq \liminf_{\varepsilon\to 0} \left(\mathbb{E}\left[\prod_{i=0}^{k-1} \frac{g_{-\varepsilon}(W)i + h(W)}{g_{-\varepsilon}(W)i + h(W) + \alpha^{(-\varepsilon)}} \mathbf{1}_{B^\varepsilon}(W)\right] + \mu(\mathcal{M}_\varepsilon) \right) \\
&= \liminf_{\varepsilon\to 0} \mathbb{E}\left[\prod_{i=0}^{k-1} \frac{g(W)i + h(W)}{g(W)i + h(W) + \alpha^{(-\varepsilon)}} \mathbf{1}_{B^\varepsilon}(W)\right] \\
&= \mathbb{E}\left[\prod_{i=0}^{k-1} \frac{g(W)i + h(W)}{g(W)i + h(W) + \lambda^*} \mathbf{1}_B(W)\right].
\end{aligned}$$

Similarly, almost surely,

$$\begin{aligned}
\liminf_{t\to\infty} \frac{N_{\geq k}(t, B)}{t} &\geq \limsup_{\varepsilon\to 0} \mathbb{E}\left[\prod_{i=0}^{k-1} \frac{g_{+\varepsilon}(W)i + h(W)}{g_{+\varepsilon}(W)i + h(W) + \alpha^{(+\varepsilon)}} \mathbf{1}_{B^\varepsilon}(W)\right] \\
&= \limsup_{\varepsilon\to 0} \mathbb{E}\left[\prod_{i=0}^{k-1} \frac{g(W)i + h(W))}{g(W)i + h(W) + \alpha^{(+\varepsilon)}} \mathbf{1}_{B^\varepsilon}(W)\right] \\
&= \mathbb{E}\left[\prod_{i=0}^{k-1} \frac{g(W)i + h(W))}{g(W)i + h(W) + \lambda^*} \mathbf{1}_B(W)\right].
\end{aligned}$$

Finally, for the last assertion, by Lemma 3.1, for each $t \in \mathbb{N}_0$ we have

$$\frac{\mathcal{Z}_t^{-\varepsilon}}{t} \leq \frac{\mathcal{Z}_t}{t} \leq \frac{\mathcal{Z}_t^{+\varepsilon}}{t}.$$

Taking limits as t goes to infinity and applying Theorem 2.4, the result follows similarly to the previous assertions. □

3.3. Degenerate degrees in the GPAF-tree when Condition C1 fails

In this subsection, we show that if the GPAF-tree fails to satisfy Condition **C1** by having $m(\lambda, \mathbb{R}_+) = \infty$ for all $\lambda > 0$, almost surely the proportion of vertices that are leaves tends to 1. Consequently, the limiting mass of edges 'escapes to infinity', as described in Theorem 3.2 below. Note that Condition **C1** fails in this manner in the GPAF tree if $\operatorname{ess\,sup}(g) = \infty$ or $\mathbb{E}[h(W)] = \infty$. We remark that results similar to Theorem 3.2 have been shown in preferential attachment models with multiplicative fitness with μ having finite support [8, Theorem 6] and preferential attachment models with additive fitness (the *extreme disorder* regime in [23, Theorem 2.6]). These cases correspond to $h(x) \equiv 0$ and $g(x) \equiv 1$, respectively. In a similar vein to the start of Section 3.2.1, in this section we will require the following families of sets: for each $m \in \mathbb{N}$, we set

$$\mathscr{G}_m := \{x : g(x) \geq m\}\,, \qquad \mathscr{H}_m := \{x : h(x) \geq m\}\,, \qquad \text{and} \qquad \mathscr{M}_m := \mathscr{G}_m \cup \mathscr{H}_m.$$

Theorem 3.2. *Suppose $\mathcal{T} = (\mathcal{T}_t)_{t\geq 0}$ is a GPAF-tree, with associated functions g,h, such that* $\operatorname{ess\,sup}(g) = \infty$ *or* $\mathbb{E}[h(W)] = \infty$. *Then we have the following assertions:*

1. *For any measurable set B such that for some $m' \in \mathbb{N}$ we have $B \subseteq \mathscr{M}^c_{m'}$,*

$$\frac{\Xi(t, B)}{t} \xrightarrow{t\to\infty} 0, \quad \textit{almost surely.}$$

2. *For any measurable set $B \subseteq \mathbb{R}_+$, we have*

$$\frac{N_0(t, B)}{t} \xrightarrow{t\to\infty} \mu(B), \quad \textit{almost surely.}$$

3. *The partition function satisfies*

$$\frac{\mathcal{Z}_t}{t} \xrightarrow{t\to\infty} \infty, \quad \textit{almost surely.}$$

Proof. This is similar to the proof of Theorem 3.1; however, we require some different notation. For each $m \in \mathbb{N}$, let $\mathcal{T}^m = (\mathcal{T}_t^m)_{t\geq 0}$ and $\mathcal{T}^{m,m} = (\mathcal{T}^{m,m}_{t\geq 0})$ denote analogues of the tree process modified on the sets $\mathscr{G}_m$ and $\mathscr{H}_m$. In particular, if we define g_m, h_m so that

$$g_m := g\mathbf{1}_{\mathscr{G}^c_m} + m\mathbf{1}_{\mathscr{G}_m} \qquad \text{and} \qquad h_m := h\mathbf{1}_{\mathscr{H}^c_m} + m\mathbf{1}_{\mathscr{H}_m},$$

we define $\mathcal{T}^m$ with the associated functions g_m, h, and $\mathcal{T}^{m,m}$ with the associated functions g_m, h_m. Then, by mimicking the approach from the coupling in Lemma 3.1, for each $m \in \mathbb{N}$ we may couple the processes $(\hat{\mathcal{T}}^{m,m}, \hat{\mathcal{T}}^m, \hat{\mathcal{T}})$ so that, for all $t \in \mathbb{N}_0$, their respective partition functions satisfy $\mathcal{Z}_t^{m,m} \leq \mathcal{Z}_t^m \leq \mathcal{Z}_t$; for each vertex v' with $W_{v'} \in \mathscr{H}^c_m$,

$$\deg^+\big(v', \hat{\mathcal{T}}_t^m\big) \leq \deg^+\big(v, \hat{\mathcal{T}}_t^{m,m}\big);$$

and for each vertex v with $W_v \in \mathscr{G}_m^c$,

$$\deg^+(v, \hat{\mathcal{T}}_t) \leq \deg^+\big(v, \hat{\mathcal{T}}_t^m\big).$$

In this coupling, at each time-step t, one samples $\hat{\mathcal{T}}_t^{m,m}$ first, uses this (with another uniformly distributed random variable) to construct $\hat{\mathcal{T}}_t^m$, and then uses this to construct $\hat{\mathcal{T}}_t$. Therefore, we have the following claim. For a measurable set B, let $\Xi^{m,m}(t, B)$ and $N_{\geq k}^{m,m}(t, B)$ denote the counterparts of $\Xi(t, B)$ and $N_{\geq k}(t, B)$ with respect to the tree $\mathcal{T}^{m,m}$. □

Claim 3.1. *For all $m \in \mathbb{N}$, there exists a coupling $(\hat{\mathcal{T}}^{m,m}, \hat{\mathcal{T}})$ of $\mathcal{T}^{m,m}$ and $\mathcal{T}$ such that, on the coupling, for all $t \in \mathbb{N}_0$ we have the following:*

1. *For all measurable sets $B \subseteq \mathscr{M}_m^c$ we have $\Xi(t, B) \leq \Xi^{m,m}(t, B)$.*
2. *For all measurable sets $B \subseteq \mathscr{M}_m^c$ and $k \in \mathbb{N}_0$ we have $N_{\geq k}(t, B) \leq N_{\geq k}^{m,m}(t, B)$.*
3. *We have the inequality $\mathcal{Z}_t^{m,m} \leq \mathcal{Z}_t$.*

Now note that for all m sufficiently large, $\mathcal{T}^{m,m}$ satisfies **C1**. Indeed, if $\operatorname{ess\,sup}(g) = \infty$, then because $\mathbb{E}[h_m(W)] \leq m$ and g_m attains its maximum m on a set of positive measure, this follows from Remark 3.1. Otherwise, for m sufficiently large we have $g_m = g$, and for any $\lambda > \operatorname{ess\,sup}(g)$,

$$\mathbb{E}\left[\frac{h_m(W)}{\lambda - g(W)}\right] < \infty, \quad \text{and, by monotone convergence,} \quad \lim_{m\uparrow\infty} \mathbb{E}\left[\frac{h_m(W)}{\lambda - g(W)}\right] = \infty.$$

Thus, making m larger if necessary, we have that **C1** is satisfied for this choice of λ. In either case, let $\alpha^{(m)}$ denote the Malthusian parameter associated with $\mathcal{T}^{m,m}$. Then $\alpha^{(m)} > \operatorname{ess\,sup}(g_m)$ increases as m increases, and even if $\operatorname{ess\,sup}(g_m) < \infty$ we must have

$$\lim_{m\uparrow\infty} \alpha^{(m)} = \infty.$$

Indeed, suppose this were not the case, and $\lim_{m\uparrow\infty} \alpha^{(m)} = \alpha' < \infty$. Then, by monotone convergence,

$$1 = \lim_{m\to\infty} \mathbb{E}\left[\frac{h_m(W)}{\alpha^{(m)} - g(W)}\right] = \mathbb{E}\left[\frac{h(W)}{\alpha' - g(W)}\right] = \infty,$$

since $\mathbb{E}[h(W)] = \infty$. Now, the assertions of Theorem 3.2 follow from the claim in a similar manner to the way the assertions of Theorem 3.1 follow from Lemma 3.1.

Now, as in the previous subsection, suppose that g and h are increasing, and $\mathrm{Supp}(\mu) \subseteq [0, w^*]$, where $w^* := \sup(\mathrm{Supp}(\mu))$. The proof of the following corollary is similar to that of Corollary 3.1, and we therefore again leave it to the reader.

Corollary 3.2. *Under the above assumptions, with regard to the weak topology,*

$$\frac{\Xi(t, \cdot)}{t} \xrightarrow{t\to\infty} \delta_{w^*}(\cdot), \quad \textit{almost surely.}$$

4. Analysis of (μ, f, ℓ)-RIF trees assuming C2

By Theorem 2.4, under certain conditions on the fitness function f and **C1**, Condition **C2** is satisfied, i.e.,

$$\frac{\mathcal{Z}_t}{t} \xrightarrow{t\to\infty} \alpha, \quad \text{almost surely.}$$

However, Theorem 3.1 shows that this condition may be satisfied despite Condition **C1** failing. Therefore, in this in this section, we analyse the model under Condition **C2**. We state and prove Theorem 4.1 below and Theorem 4.2, leaving the details to the reader. These proofs rely on Proposition 4.1, proved in Section 4.3 and Section 4.4, and on Proposition 4.2, proved in Section 4.5.

4.1. Main results: convergence in probability of $N_k(n, B)/\ell n$ under C2

Theorem 4.1. *Assume Condition C2. Then, for any measurable set B, we have*

$$\frac{N_k(t, B)}{\ell t} \xrightarrow{t\to\infty} \mathbb{E}\left[\frac{\alpha}{f(k, W)+\alpha}\prod_{s=0}^{k-1}\frac{f(s, W)}{f(s, W)+\alpha}\mathbf{1}_B(W)\right] = p_k^{\alpha}(B), \quad \textit{in probability.}$$

In order to prove Theorem 4.1, we define the following family of sets:

$$\mathscr{F} := \{B : B \text{ is measurable and } \forall s\in\mathbb{N}_0,\ f(s, w) \text{ is bounded for } w\in B\}. \tag{28}$$

We also require Proposition 4.1 and Proposition 4.2, proved in Section 4.4.1 and Section 4.5.1. These proofs rely on the results stated in Section 4.2 and Section 4.3.

Proposition 4.1. *For any set $B\in\mathscr{F}$, for each $k\in\mathbb{N}_0$ we have*

$$\lim_{t\to\infty}\frac{\mathbb{E}[N_k(t, B)]}{\ell t} = p_k^{\alpha}(B).$$

Proposition 4.2. *For any $B\in\mathscr{F}$ and $k\in\mathbb{N}_0$ we have*

$$\lim_{t\to\infty}\mathbb{E}\left[\frac{(N_k(t, B))^2}{\ell^2 t^2}\right] = (p_k^{\alpha}(B))^2.$$

Proof of Theorem 4.1. The result follows for all $B\in\mathscr{F}$ by combining Proposition 4.1 and Proposition 4.2 and applying Chebyshev's inequality.

Now, let B be an arbitrary measurable set and let $\varepsilon>0$ be given. Then, since for each $s\in\{1, \ldots, k\}$ the map $w\mapsto f(s, w)$ is measurable, by Lusin's theorem, we can find a compact set $E\subseteq B$ such that $\mu(B\cap E^c)<\varepsilon/3$ and for each $s\in\{1, \ldots, k\}$ the restriction of the map $w\mapsto f(s, w)$ to E is continuous. Moreover, note that $p_k^{\alpha}(B)-p_k^{\alpha}(B\cap E)\le\mu(B\cap E^c)<\varepsilon/3$. Then,

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{N_k(t, B)}{\ell t}-p_k^{\alpha}(B)\right|>\varepsilon\right) &\le \mathbb{P}\Bigg(\Bigg(\left|\frac{N_k(t, B)}{\ell t}-\frac{N_k(t, B\cap E)}{\ell t}\right| \\
&\qquad+\left|\frac{N_k(t, B\cap E)}{\ell t}-p_k^{\alpha}(B\cap E)\right| \\
&\qquad+\left|p_k^{\alpha}(B\cap E)-p_k^{\alpha}(B)\right|\Bigg)>\varepsilon\Bigg) \\
&\le \mathbb{P}\left(\left|\frac{N_k(t, B\cap E)}{\ell t}-p_k^{\alpha}(B\cap E)\right|>\varepsilon/3\right) \\
&\qquad+\mathbb{P}\left(\left|\frac{N_k(t, B)}{\ell t}-\frac{N_k(t, B\cap E)}{\ell t}\right|>\varepsilon/3\right).
\end{aligned} \tag{29}$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

Now, note that by the strong law of large numbers applied to $N_{\geq 0}(t, B \cap E^c)/\ell t$, i.e., the proportion of all vertices with weight belonging to $B \cap E^c$, and Egorov's theorem, for any $\delta > 0$ there exists an event G with $\mathbb{P}(G) < \delta$ such that

$$\limsup_{t\to\infty} \left(\frac{N_k(t, B)}{\ell t} - \frac{N_k(t, B\cap E)}{\ell t}\right) = \limsup_{t\to\infty} \frac{N_k(t, B\cap E^c)}{\ell t} \leq \mu(B\cap E^c)$$

uniformly on the complement of G. Therefore, the result follows from (29), Proposition 4.1, and Proposition 4.2 by taking limits as t tends to infinity. □

Using the approach to the upper bound for the mean in the next subsection, and applying Corollary 4.1 stated below with $k = 1$ and $e_0, e_1 = 0$, if $N_{\geq 1}(t, B)$ denotes the number of vertices of out-degree at least 1 in the tree with weight belonging to B, we actually have

$$\limsup_{t\to\infty} \frac{\mathbb{E}\big[N_{\geq 1}(t, B)\big]}{\ell t} \leq \frac{1}{\alpha'}\mathbb{E}\big[f(0, W)\mathbf{1}_B(W)\big],$$

as long as $\liminf_{t\to\infty} \frac{\mathcal{Z}_t}{t} \geq \alpha'$. By sending α' to infinity, this yields the following analogue of Theorem 3.2.

Theorem 4.2. *Suppose $\mathcal{T}$ is a (μ, f, ℓ)-RIF tree such that $\lim_{t\to\infty} \frac{\mathcal{Z}_t}{t} = \infty$. Then for any measurable set $B \subseteq [0, \infty)$, we have*

$$\frac{N_0(t, B)}{\ell t} \xrightarrow{t\to\infty} \mu(B), \quad \textit{in probability.}$$

4.2. Summation arguments

Here we state some summation arguments required for the subsequent proofs. The following lemma and corollary are taken from [17]. We include them here, with minor changes in notation, for completeness. For $e_0, \ldots, e_k \geq 0$, $0 \leq \eta < 1$, let

$$\mathcal{S}_t(e_0, \ldots, e_k, \eta) := \frac{1}{t} \sum_{\eta t < i_0 < \cdots < i_k \leq t} \prod_{s=0}^{k-1} \left(\left(\frac{i_s}{i_{s+1}}\right)^{e_s} \cdot \frac{1}{i_{s+1} - 1}\right)\left(\frac{i_k}{t}\right)^{e_k}.$$

Lemma 4.1. ([17, Lemma 4].) *Uniformly in $e_0, \ldots, e_k \geq 0$, $0 \leq \eta \leq 1/2$, we have*

$$\mathcal{S}_t(e_0, \ldots, e_k, \eta) = \prod_{s=0}^{k} \frac{1}{e_s + 1} + \theta(\eta) + O\left(\frac{1}{t^{1/(k+2)}} + \frac{\sum_{s=0}^k e_s \log^{k+1}(t)}{t}\right).$$

Here, $\theta(\eta)$ is a term satisfying $|\theta(\eta)| \leq M\eta^{1/(k+2)}$ for some universal constant M depending only on k.

Corollary 4.1. ([17, Corollary 5].) *For $e_0, \ldots, e_k, f_0, \ldots, f_{k-1} \geq 0$, $0 \leq \eta \leq 1/2$, we have*

$$\frac{1}{t} \sum_{\eta t < i_0 \leq t} \sum_{\mathcal{I}_k \in \binom{\{i_0+1, \ldots, t\}}{k}} \prod_{s=0}^{k-1} \left(\left(\frac{i_s}{i_{s+1}}\right)^{e_s} \cdot \frac{f_s}{i_{s+1} - 1}\right)\left(\frac{i_k}{t}\right)^{e_k}$$
$$= \frac{1}{e_k + 1} \prod_{s=0}^{k-1} \frac{f_s}{e_s + 1} + \theta'(\eta) + O\left(\frac{1}{t^{1/(k+2)}}\right).$$

Here, $\theta'(\eta)$ is a term satisfying $|\theta'(\eta)| \le M'\eta^{1/(k+2)}$ for some universal constant M' depending only on k and $f_0, \ldots, f_{k-1}$, and the constant in the big-O term may depend on $e_0, \ldots, e_k, f_0, \ldots, f_k$.

4.3. Upper bound for the mean of $N_k(t, B)/\ell t$

In the following subsections, unless otherwise specified, we let B denote an arbitrary element of the family $\mathscr{F}$ defined in (4.1). Let $N_{\eta,k}(t, B)$ be the number of vertices of degree $k\ell$ with weight in B that arrived after time ηt. Then, since $N_{\eta,k}(t, B) \le N_k(t, B) \le N_{\eta,k}(t, B) + \eta\ell t$, we have

$$\mathbb{E}\left[\left|\frac{N_{\eta,k}(t, B)}{\ell t} - \frac{N_k(t, B)}{\ell t}\right|\right] \le \eta. \tag{30}$$

Thus, to obtain an upper bound for the convergence of the mean, it suffices to prove that

$$\limsup_{\eta\to 0} \limsup_{t\to\infty} \mathbb{E}\left[\frac{N_{\eta,k}(t, B)}{\ell t}\right] = p_k^{\alpha}(B).$$

In what follows, we use the notation $d_i(t)$ to denote the out-degree at time t of the vertex i born at time $i_0 := \lfloor i/\ell \rfloor$. We then have

$$\mathbb{E}\left[N_{\eta,k}(t, B)\right] = \sum_{\eta t < i_0 \le t-k} \ell \cdot \mathbb{P}(d_i(t) = k,\, W_i \in B),$$

since the probability is identical for each of the ℓ vertices born at each time i_0. In what follows, for a given i we denote by $\mathcal{I}_k := \{i_1, \ldots, i_k\}$ a collection of natural numbers $i_0 < i_1 < \ldots < i_k \le t$. For ease of notation we exclude the dependence of $\mathcal{I}_k$ on i.

For a natural number $s > i_0$, we use the notation $i \to s$ to denote that i is the vertex chosen at the sth time-step; hence i gains ℓ new neighbours at time s. Likewise, the notation $i \not\to s$ denotes that i is not chosen at the sth time-step. Then, let $\mathcal{E}_i(\mathcal{I}_k, B)$ denote the event that $W_i \in B$ and for all $s \in \{i_0 + 1, \ldots, t\}$, $i \to s$ if and only if $s \in \mathcal{I}_k$. Clearly, we have

$$\mathbb{P}(d_i(t) = k,\, W_i \in B) = \sum_{\mathcal{I}_k \in \binom{\{i_0+1, \ldots, t\}}{k}} \mathbb{P}(\mathcal{E}_i(\mathcal{I}_k, B)),$$

where $\binom{\{i_0+1, \ldots, t\}}{k}$ denotes the set of all subsets of $\{i_0 + 1, \ldots, t\}$ of size k. For $\varepsilon > 0$ and $t \ge 0$ and natural numbers $N_1 \le N_2$, we let

$$\mathcal{G}_{\varepsilon}(t) = \{|\mathcal{Z}_t - \alpha t| < \varepsilon\alpha t\}, \text{ and } \mathcal{G}_{\varepsilon}(N_1, N_2) = \bigcap_{t=N_1}^{N_2} \mathcal{G}_{\varepsilon}(t). \tag{31}$$

Moreover, for $t \ge 1$, we denote by $\mathscr{T}_t$ the σ-field generated by $(\mathcal{T}_s)_{1\le s\le t}$, containing all the information generated by the process up to time t. By the assumption of almost sure

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

convergence and Egorov's theorem, for any $\delta, \varepsilon > 0$, there exists $N' = N'(\varepsilon, \delta)$ such that, for all $t \geq N'$, $\mathbb{P}\big(\mathcal{G}_\varepsilon(N', t)\big) \geq 1 - \delta$. Thus, for $t \geq N'/\eta$, we have

$$\begin{aligned} \mathbb{E}\big[N_{\eta,k}(t, B)\big] &\leq \mathbb{E}\big[N_{\eta,k}(t, B)\mathbf{1}_{\mathcal{G}_\varepsilon(N',t)}\big] + \ell t\big(1 - \mathbb{P}\big(\mathcal{G}_\varepsilon(N', t)\big)\big) \\ &\leq \ell \left(\sum_{\eta t < i_0 \leq t} \sum_{\mathcal{I}_k \in \binom{\{i_0+1, \ldots, t\}}{k}} \mathbb{P}\left(\mathcal{E}_i(\mathcal{I}_k, B) \cap \mathcal{G}_\varepsilon(i_0, t)\right) + \delta t \right). \end{aligned} \tag{32}$$

We use the shorthand $\alpha_{\pm\varepsilon} := (1 \pm \varepsilon)\alpha$.

Proposition 4.3. *Let $B \in \mathscr{F}$ and $0 < \varepsilon, \eta \leq 1/2$. As $t \to \infty$, uniformly in $\eta t < i_0 \leq t - k$, $\mathcal{I}_k = \{i_1, \ldots, i_k\} \in \binom{\{i_0+1, \ldots, t\}}{k}$, and the choice of ε, we have*

$$\begin{aligned} &\mathbb{P}(\mathcal{E}_i(\mathcal{I}_k, B) \cap \mathcal{G}_\varepsilon(i_0, t)) \\ &\leq (1 + O(1/t))\, \mathbb{E}\left[\left(\frac{i_k}{t}\right)^{f(k,W)/\alpha_{+\varepsilon}} \prod_{s=0}^{k-1} \left(\frac{i_s}{i_{s+1}}\right)^{f(s,W)/\alpha_{+\varepsilon}} \frac{f(s, W)}{\alpha_{-\varepsilon}(i_{s+1} - 1)} \mathbf{1}_B(W)\right]. \end{aligned}$$

Corollary 4.2. *Let $B \in \mathscr{F}$ and $0 < \delta, \varepsilon, \eta \leq 1/2$. Then there exists $N = N(\delta, \varepsilon, \eta)$ such that, for all $t \geq N$,*

$$\begin{aligned} \frac{\mathbb{E}\big[N_{\eta,k}(t, B)\big]}{\ell t} &\leq (1 + \delta) \left(\frac{1 + \varepsilon}{1 - \varepsilon}\right)^k \mathbb{E}\left[\frac{\alpha_{+\varepsilon}}{f(k, W) + \alpha_{+\varepsilon}} \prod_{s=0}^{k-1} \frac{f(s, W)}{f(s, W) + \alpha_{+\varepsilon}} \mathbf{1}_B(W)\right] \\ &\quad + C\eta^{1/(k+2)} + \delta, \end{aligned}$$

where the constant C may depend on k and B but not on n and not on the choices of $\delta, \varepsilon, \eta$. In particular, for each $B \in \mathscr{F}$ and $k \in \mathbb{N}_0$,

$$\limsup_{t \to \infty} \mathbb{E}[N_k(t, B)]/\ell t \leq p_k^\alpha(B).$$

Proof. This follows from applying (32) and Proposition 4.3 and then applying Corollary 4.1 with $e_j = f(j, W)/\alpha_{+\varepsilon}$ and $f_j = f(j, W)/\alpha_{-\varepsilon}$ to bound the sum over the collection of indices. Note that the term $\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^k$ comes from replacing $\alpha_{-\varepsilon}$ by $\alpha_{+\varepsilon}$. □

We proceed towards the proof of Proposition 4.3. Let ε, η be given such that $0 < \varepsilon, \eta \leq 1/2$. For $\eta t < i_0 \leq t$ and $\mathcal{I}_k = \{i_1, \ldots, i_k\} \in \binom{\{i_0+1, \ldots, t\}}{k}$, for each $s \in \{i_0 + 1, \ldots, t\}$ we define

$$\mathcal{D}_s := \begin{cases} \{i \to s\} & \text{if } s \in \mathcal{I}_k, \\ \{i \not\to s\} & \text{otherwise,} \end{cases}$$

and $\tilde{\mathcal{D}}_s = \mathcal{D}_s \cap \mathcal{G}_\varepsilon(s)$. We also define $\tilde{\mathcal{D}}_{i_0} = \mathcal{G}_\varepsilon(i_0) \cap \{W_i \in B\}$, and for simplicity of notation we write D_j and $\tilde{D}_j$ for the indicator random variables $\mathbf{1}_{\mathcal{D}_j}$ and $\mathbf{1}_{\tilde{\mathcal{D}}_j}$, respectively. Note that $\mathcal{E}_i(\mathcal{I}_k, B) \cap \mathcal{G}_\varepsilon(i_0, t) = \bigcap_{j=i_0}^t \tilde{\mathcal{D}}_j$. To bound the probability of this event, we define

$$X_s = \mathbb{E}\left[\prod_{j=i_s+1}^{t} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_s}\right]\tilde{D}_{i_s}, \qquad s \in \{0, \ldots, k\}$$

and observe that $\mathbb{E}[X_0] = \mathbb{P}\left(\bigcap_{s=i_0}^t \tilde{\mathcal{D}}_s\right)$ is the probability we seek.

Lemma 4.2. *For $s \in \{0, \ldots, k\}$, we have*

$$X_s \le \prod_{u=i_k+1}^{n}\left(1 - \frac{f(k, W)}{\alpha_{+\varepsilon}(u-1)}\right)\left(\prod_{j=s}^{k-1} \frac{f(j, W)}{\alpha_{-\varepsilon}(i_{j+1}-1)} \prod_{j'=i_j+1}^{i_{j+1}-1}\left(1 - \frac{f(j, W)}{\alpha_{+\varepsilon}(j'-1)}\right)\right)\tilde{D}_{i_s}, \tag{33}$$

where we interpret any empty products (for example when $i_k = n$) as equal to 1. In particular,

$$\mathbb{E}[X_0] \le \mathbb{E}\left[\prod_{u=i_k+1}^{n}\left(1 - \frac{f(k, W)}{\alpha_{+\varepsilon}(u-1)}\right)\left(\prod_{j=0}^{k-1} \frac{f(j, W)}{\alpha_{-\varepsilon}(i_{j+1}-1)} \prod_{j'=i_j+1}^{i_{j+1}-1}\left(1 - \frac{f(j, W)}{\alpha_{+\varepsilon}(j'-1)}\right)\right)\mathbf{1}_B(W)\right]. \tag{34}$$

Proof. We prove (33) by backwards induction. For the base case, $s = k$, if $i_k = n$, the inequality is trivial, as $X_k = \tilde{D}_{i_k}$. Thus, assuming $i_k < n$, by the tower property,

$$\begin{aligned}
\mathbb{E}\left[\prod_{j=i_k+1}^{n} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_k}\right] &= \mathbb{E}\left[\mathbb{E}\left[\tilde{D}_n \,\middle|\, \mathscr{T}_{n-1}\right]\prod_{j=i_k+1}^{n-1} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_k}\right] \\
&\le \mathbb{E}\left[\mathbb{E}\left[D_n \,\middle|\, \mathscr{T}_{n-1}\right]\prod_{j=i_k+1}^{n-1} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_k}\right] \\
&= \mathbb{E}\left[\left(1 - \frac{f(k, W)}{\mathcal{Z}_{n-1}}\right)\prod_{j=i_k+1}^{n-1} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_k}\right] \\
&\le \left(1 - \frac{f(k, W)}{\alpha_{+\varepsilon}(n-1)}\right)\mathbb{E}\left[\prod_{j=i_k+1}^{n-1} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_k}\right],
\end{aligned}$$

and iterating this argument with the conditional expectation on the right-hand side proves the base case. Now, note that for $s \in \{0, \ldots, k-1\}$

$$X_s = \mathbb{E}\left[X_{s+1}\prod_{j=i_s+1}^{i_{s+1}-1} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_s}\right]\tilde{D}_{i_s}.$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

Applying the induction hypothesis, it suffices to bound the term $\mathbb{E}\left[\prod_{j=i_s+1}^{i_{s+1}} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_s}\right]$, and, similarly to the base case, we may assume $i_s < i_{s+1} - 1$. But then we have

$$\begin{aligned}
\mathbb{E}\left[\prod_{j=i_s+1}^{i_{s+1}} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_s}\right] &= \mathbb{E}\left[\mathbb{E}\left[\tilde{D}_{i_{s+1}} \,\middle|\, \mathscr{T}_{i_{s+1}-1}\right] \prod_{j=i_s+1}^{i_{s+1}-2} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_s}\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[D_{i_{s+1}} \,\middle|\, \mathscr{T}_{i_{s+1}-1}\right] \prod_{j=i_s+1}^{i_{s+1}-2} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_s}\right] \\
&\leq \frac{f(s, W)}{\alpha_{-\varepsilon}(i_{s+1}-1)} \mathbb{E}\left[\prod_{j=i_s+1}^{i_{s+1}-2} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_s}\right] \\
&\leq \frac{f(s, W)}{\alpha_{-\varepsilon}(i_{s+1}-1)} \mathbb{E}\left[\mathbb{E}\left[D_{i_{s+1}-1} \,\middle|\, \mathscr{T}_{i_{s+1}-1}\right] \prod_{j=i_s+1}^{i_{s+1}-2} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_s}\right] \\
&\leq \frac{f(s, W)}{\alpha_{-\varepsilon}(i_{s+1}-1)} \left(1 - \frac{f(s, W)}{\alpha_{+\varepsilon}\,(i_{s+1}-2)}\right) \mathbb{E}\left[\prod_{j=i_s+1}^{i_{s+1}-2} \tilde{D}_j \,\middle|\, \mathscr{T}_{i_s}\right].
\end{aligned}$$

Iterating these bounds, the inductive step follows in a similar manner to the base case. Finally, noting that $\mathbf{1}_{\tilde{\mathcal{D}}_i} \leq \mathbf{1}_B(W)$ proves (34). □

The next lemma follows from a simple application of Stirling's formula, i.e., (25).

Lemma 4.3. *Let η, $C > 0$. Then, uniformly over $\eta t \leq a \leq b$ and $0 \leq \beta \leq C$, we have*

$$\prod_{j=a+1}^{b-1} \left(1 - \frac{\beta}{j-1}\right) = \left(\frac{a}{b}\right)^{\beta} \left(1 + o\left(\frac{1}{t}\right)\right).$$

Proof of Proposition 4.3. We take the upper bound $\mathbb{E}[X_0]$ from Lemma 4.2 and bound each of the products by applying Lemma 4.3. □

4.4. Deducing convergence of the mean of $N_k(t, B)/\ell t$

In this subsection we deduce a lower bound on $\liminf_{t\to\infty} \mathbb{E}[N_k(t, B)]/\ell t$ on measurable sets $B \in \mathscr{F}$. In what follows, denote by $N_{\geq M}(t, B)$ the number of vertices of out-degree $\geq \ell M$ with weight belonging to B. Moreover, let $N(t, B) = N_{\geq 0}(t, B)$ denote the total number of vertices at time t with weight belonging to B.

Lemma 4.4. *For any measurable set B, we have*

$$\limsup_{t\to\infty} \frac{N_{\geq M}(t, B)}{\ell t} \leq \frac{1}{M}$$

almost surely.

Proof. Since we add ℓ vertices at each time-step, we have $\limsup_{t\to\infty} \frac{|\mathcal{T}_t|}{\ell t} = 1$. However, $|\mathcal{T}_t| \geq M N_{\geq M}(t, \mathbb{R})$, since the right-hand side only provides a lower bound for the number of vertices in the tree incident to those with out-degree at least M. The result follows from dividing both sides by $M\ell t$ and sending t to infinity. □

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

4.4.1. *Proof of Proposition 4.1* *Proof.* Recall that Corollary 4.2 showed that for each $B \in \mathscr{F}$ and $k \in \mathbb{N}_0$,

$$\limsup_{t\to\infty} \mathbb{E}[N_k(t, B)]/\ell t \le p_k^{\alpha}(B).$$

Now, suppose that Proposition 4.1 fails, so that in particular there exist some set $B' \in \mathscr{F}$ and some $k' \in \mathbb{N}_0$ such that

$$\liminf_{t\to\infty} \frac{\mathbb{E}\big[N_{k'}(t, B')\big]}{\ell t} < p_{k'}^{\alpha}(B').$$

Thus, for some $\epsilon' > 0$, we have

$$\liminf_{t\to\infty} \frac{\mathbb{E}\big[N_{k'}(t, B')\big]}{\ell t} \le p_{k'}^{\alpha}(B) - \epsilon'.$$

Now, using Lemma 4.4, choose $M > \max\left\{k', \frac{2}{\epsilon'}\right\}$, so that

$$\limsup_{t\to\infty} \frac{N_{\ge M}(t, B')}{\ell t} < \epsilon'/2.$$

Then, recalling (13),

$$\begin{aligned}\liminf_{t\to\infty} \mathbb{E}\left[\sum_{k=0}^{M} \frac{N_k(t, B')}{\ell t}\right] &\le \liminf_{t\to\infty} \mathbb{E}\left[\frac{N_{k'}(t, B')}{\ell t}\right] + \sum_{k\ne k'} \limsup_{t\to\infty} \mathbb{E}\left[\frac{N_k(t, B')}{\ell t}\right] \\ &\le \left(\sum_{k=0}^{\infty} p_k^{\alpha}(B')\right) - \epsilon' \le \mu(B') - \epsilon'.\end{aligned} \tag{35}$$

On the other hand, by Fatou's lemma, we have

$$\begin{aligned}\liminf_{t\to\infty} \mathbb{E}\left[\sum_{k=0}^{M} \frac{N_k(t, B')}{\ell t}\right] &\ge \mathbb{E}\left[\liminf_{t\to\infty} \sum_{k=0}^{M} \frac{N_k(t, B')}{\ell t}\right] \\ &= \mathbb{E}\left[\liminf_{t\to\infty} \left(\frac{N(t, B')}{\ell t} - \frac{N_{\ge M}(t, B')}{\ell t}\right)\right] \ge \mu(B') - \epsilon'/2,\end{aligned} \tag{36}$$

where the last inequality follows from the strong law of large numbers. But then, combining (35) and (36), we have $\mu(B') - \epsilon' \ge \mu(B') - \epsilon'/2$, a contradiction. □

4.5. Second moment calculations

In order to bound the second moment, we apply calculations similar to those at the start of the section to compute asymptotically the number of pairs of vertices of out-degree $k\ell$ born after time ηt. For vertices i and j, as in Section 4.3, we set $i_0 := \lfloor i/\ell \rfloor$ and $j_0 := \lfloor j/\ell \rfloor$, and note that

$$\mathbb{E}\Big[\big(N_{\eta,k}(t, B)\big)^2\Big] = \sum_{\eta t < i_0, j_0 \le t-k} \;\sum_{j\,:\,\lfloor j/\ell \rfloor = j_0} \;\sum_{i\,:\,\lfloor i/\ell \rfloor = i_0} \mathbb{P}\big(d_i(t) = k,\, W_i \in B,\, d_j(t) = k,\, W_j \in B\big). \tag{37}$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

Note that, in a similar manner to (30), we have

$$\mathbb{E}\left[\left|\frac{\left(N_{\eta,k}(t,B)\right)^2}{\ell^2 t^2} - \frac{(N_k(t,B))^2}{\ell^2 t^2}\right|\right] \le \eta,$$

so that it suffices to prove that

$$\limsup_{\eta\to 0}\limsup_{t\to\infty}\mathbb{E}\left[\frac{\left(N_{\eta,k}(t,B)\right)^2}{\ell^2 t^2}\right] \le (p_k^{\alpha}(B))^2.$$

Recall that, for a given i, we denote by $\mathcal{I}_k$ a collection of natural numbers $i_0 < i_1 < \cdots < i_k \le t$. Moreover, for a given j, we denote by $\mathcal{J}_k$ a collection of natural numbers $j_0 < j_1 < \cdots < j_k \le t$. Similarly to Section 4.3, for $s > j$ we use the notation $j \to s$ to denote that j is the vertex chosen at the sth time-step, and likewise, we let $\mathcal{E}_j(\mathcal{J}_k, B)$ denote the event that $W_j \in B$ and for all $s \in \{j_0+1, \ldots, t\}$, $j \to s$ if and only if $s \in \mathcal{J}_k$. Then we have

$$\mathbb{P}\left(d_i(t)=k,\, W_i \in B,\, d_j(t)=k,\, W_j \in B\right) = \sum_{\mathcal{J}_k \in \binom{\{j_0+1,\ldots,t\}}{k}} \sum_{\mathcal{I}_k \in \binom{\{i_0+1,\ldots,t\}}{k}} \mathbb{P}\big(\mathcal{E}_i(\mathcal{I}_k,B) \cap \mathcal{E}_j(\mathcal{J}_k,B)\big).$$

Note that the contribution to the above sum corresponding to terms with $\mathcal{I}_k \cap \mathcal{J}_k \neq \varnothing$, and $i \neq j$, is zero, since it is impossible for distinct vertices to be chosen in a single time-step. But then, the terms corresponding to $i=j$ contribute at most $\mathbb{E}\big[N_{\eta,k}(n,B)\big] \le \ell n$ to the right side of (37). Next, for any choice of indices with $\mathcal{I}_k \cap \mathcal{J}_k = \emptyset$, there are at most ℓ^2 pairs of vertices (i,j) born at respective times (i_0, j_0) contributing to the sum in (37). Recalling the definitions of $\mathcal{G}_{\varepsilon}(t)$, $\mathcal{G}_{\varepsilon}(N_1, N_2)$, and $N' = N'(\varepsilon, \delta)$ from (31) and below in the previous subsection, in a similar manner to (32) we have, for $t \ge N'/\eta$,

$$\mathbb{E}\left[\left(N_{\eta,k}(t,B)\right)^2\right] \le \ell^2\left(\sum_{\eta t < i_0, j_0 \le t-k}\ \sum_{\mathcal{I}_k \cap \mathcal{J}_k = \emptyset} \mathbb{P}\big(\mathcal{E}_i(\mathcal{I}_k,B) \cap \mathcal{E}_j(\mathcal{J}_k,B) \cap \mathcal{G}_{\varepsilon}(i_0,t)\big) + \delta t^2\right) + \ell t. \tag{38}$$

We then have the following.

Proposition 4.4. *Let $B \in \mathscr{F}$ and $0 < \varepsilon, \eta \le 1/2$. As $t \to \infty$, uniformly in $\eta t < i_0 \le j_0 \le t-k$, in $\mathcal{I}_k \in \binom{\{i_0+1,\ldots,t\}}{k}$, $\mathcal{J}_k \in \binom{\{j_0+1,\ldots,t\}}{k}$ such that $\mathcal{I}_k \cap \mathcal{J}_k = \emptyset$, and in the choice of ε, we have*

$$\begin{aligned}\mathbb{P}\big(&\mathcal{E}_i(\mathcal{I}_k,B) \cap \mathcal{E}_j(\mathcal{J}_k,B) \cap \mathcal{G}_{\varepsilon}(i_0,t)\big)\\ &\le (1+O(1/t))\mathbb{E}\left[\left(\frac{i_k}{t}\right)^{f(k,W)/\alpha_{+\varepsilon}} \cdot \prod_{s=0}^{k-1}\left(\left(\frac{i_s}{i_{s+1}}\right)^{f(s,W)/\alpha_{+\varepsilon}} \frac{f(s,W)}{\alpha_{-\varepsilon}(i_{s+1}-1)}\right)\mathbf{1}_B(W)\right]\\ &\quad\times \mathbb{E}\left[\left(\frac{j_k}{t}\right)^{f(k,W)/\alpha_{+\varepsilon}} \cdot \prod_{s=0}^{k-1}\left(\left(\frac{j_s}{j_{s+1}}\right)^{f(s,W)/\alpha_{+\varepsilon}} \frac{f(s,W)}{\alpha_{-\varepsilon}(j_{s+1}-1}\right)\mathbf{1}_B(W)\right].\end{aligned} \tag{39}$$

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

We leave the details of the proof of this proposition to the reader, as it follows an analogous approach to the proof of Proposition 4.3, using a backwards induction argument.

Proof. sketch. Let $u_1, \ldots, u_{2k}$ denote the indices in $\mathcal{I}_k \cup \mathcal{J}_k$, and let $f_x(i), f_x(j)$ denote the fitnesses associated with vertex i and vertex j at time x. Then, when we bound the probabilities $\{i \not\to x\} \cap \{j \not\to x\}$ for all $x \in \{u_s + 1, \ldots, u_{s+1} - 1\}$ from above, we obtain terms of the form

$$\prod_{x=u_s+1}^{u_{s+1}-1} \left(1 - \frac{f_x(i) + f_x(j)}{\alpha_{+\varepsilon}(x-1)}\right) = \left(\frac{u_s}{u_{s+1}}\right)^{f_x(i)+f_x(j)} \left(1 + O\left(\frac{1}{t}\right)\right),$$

where the right side follows from Lemma 4.3. Then, when we evaluate the expectation analogous to the expectation appearing in (34), we obtain an expectation involving products of terms dependent on W_i and W_j, i.e., the weights associated with vertex i and vertex j. These terms separate into a product of expectations by the independence of the random variables W_i, W_j, and finally, many of the products telescope to yield the right side of (39). □

4.5.1. *Proof of Proposition 4.2 Proof.* We apply Proposition 4.4 to bound the summands in (38). Then, as we are looking for an upper bound, we may drop the condition $\mathcal{I}_k \cap \mathcal{J}_k = \emptyset$ when evaluating the sum. But then, by Corollary 4.1, we have, uniformly in ε and η,

$$\begin{aligned}
&\sum_{\eta t < i_0, j_0 \le t} \sum_{\mathcal{I}_k, \mathcal{J}_k} \mathbb{E}\left[\left(\frac{i_k}{t}\right)^{f(k,W)/\alpha_{+\varepsilon}} \cdot \prod_{s=0}^{k-1} \left(\frac{i_s}{i_{s+1}}\right)^{f(s,W)/\alpha_{+\varepsilon}} \frac{f(s, W)}{\alpha_{-\varepsilon}(i_{s+1}-1)} \mathbf{1}_B(W)\right] \\
&\qquad \times \mathbb{E}\left[\left(\frac{j_k}{t}\right)^{f(k,W)/\alpha_{+\varepsilon}} \cdot \prod_{s=0}^{k-1} \left(\frac{j_s}{j_{s+1}}\right)^{f(s,W)/\alpha_{+\varepsilon}} \frac{f(s, W)}{\alpha_{-\varepsilon}(j_{s+1}-1)} \mathbf{1}_B(W)\right] \\
&\le \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{2k} \left(\mathbb{E}\left[\frac{\alpha_{+\varepsilon}}{f(k, W) + \alpha_{+\varepsilon}} \prod_{s=0}^{k-1} \frac{f(s, W)}{f(s, W) + \alpha_{+\varepsilon}} \mathbf{1}_B(W)\right]\right)^2 + O\left(t^{-1/(k+2)}\right) + C'\eta^{1/k+2},
\end{aligned}$$

for some universal constant $C' > 0$, depending only on B, f. The result follows. □

Acknowledgements

I would like to thank my supervisor, Nikolaos Fountoulakis, for his guidance and useful feedback on earlier drafts. I would also like to thank Cécile Mailler and Henning Sulzbach, whose collaborative work (along with Nikolaos) on a previous project introduced me to some of the techniques used in this paper. Finally, I would like to thank the anonymous referees for their helpful comments, which greatly improved the presentation of this paper, in particular leading to the removal of some unnecessary assumptions in the statements of Theorem 3.1 and Theorem 3.2.

Funding information

There are no funding bodies to thank in relation to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

References

[1] Antunović, T., Mossel, E. and Rácz, M. Z. (2016). Coexistence in preferential attachment networks. *Combinatorics Prob. Comput.* **25**, 797–822.
[2] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286**, 509–512.
[3] Bhamidi, S. (2007). *Universal techniques to analyze preferential attachment trees: global and local analysis*. Preprint. Available at https://pdfs.semanticscholar.org/e7fb/8c999ff62a5f080e4c329a7a450f41fb1528.pdf.
[4] Bianconi, G. (2002). Growing Cayley trees described by a Fermi distribution. *Phys. Rev. E* **66**, paper no. 036116.
[5] Bianconi, G. and Barabási, A.-L. (2001). Bose–Einstein condensation in complex networks. *Phys. Rev. Lett.* **86**, 5632–5635.
[6] Bollobás, B. and Riordan, O. (2004). Shortest paths and load scaling in scale-free trees. *Phys. Rev. E* **69**, paper no. 036114.
[7] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18**, 279–290.
[8] Borgs, C., Chayes, J., Daskalakis, C. and Roch, S. (2007). First to market is not everything: an analysis of preferential attachment with fitness. In *STOC '07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, Association for Computing Machinery, New York, pp. 135–144.
[9] Borovkov, K. A. and Vatutin, V. A. (2006). On the asymptotic behaviour of random recursive trees in random environments. *Adv. Appl. Prob.* **38**, 1047–1070.
[10] Chen, W.-C. and Ni, W.-C. (1994). Internal path length of the binary representation of heap-ordered trees. *Inf. Process. Lett.* **51**, 129–132.
[11] Dereich, S. (2016). Preferential attachment with fitness: unfolding the condensate. *Electron. J. Prob.* **21**, paper no. 3, 38 pp.
[12] Dereich, S., Mailler, C. and Mörters, P. (2017). Nonextensive condensation in reinforced branching processes. *Ann. Appl. Prob.* **27**, 2539–2568.
[13] Dereich, S. and Mörters, P. (2009). Random networks with sublinear preferential attachment: degree evolutions. *Electron. J. Prob.* **14**, 1222–1267.
[14] Dereich, S. and Ortgiese, M. (2014). Robust analysis of preferential attachment models with fitness. *Combinatorics Prob. Comput.* **23**, 386–411.
[15] Drmota, M. (2009). *Random Trees: An Interplay between Combinatorics and Probability*. Springer, Vienna.
[16] Ergün, G. and Rodgers, G. J. (2002). Growing random networks with fitness. *Physica A* **303**, 261–272.
[17] Fountoulakis, N., Iyer, T., Mailler, C. and Sulzbach, H. (2019). *Dynamical models for random simplicial complexes*. Preprint. Available at https://arxiv.org/abs/1910.12715.
[18] Garavaglia, A., van der Hofstad, R. and Woeginger, G. (2017). The dynamics of power laws: fitness and aging in preferential attachment trees. *J. Statist. Phys.* **168**, 1137–1179.
[19] Hiesmayr, E. and Işlak, Ü. (2020). Asymptotic results on Hoppe trees and their variations. *J. Appl. Prob.* **57**, 441–457.
[20] Holmgren, C. and Janson, S. (2017). Fringe trees, Crump–Mode–Jagers branching processes and m-ary search trees. *Prob. Surveys* **14**, 53–154.
[21] Janson, S. (2021). A.s. convergence for infinite colour Pólya urns associated with random walks. *Ark. Mat.* **59**, 87–123.
[22] Jordan, J. (2018). Preferential attachment graphs with co-existing types of different fitnesses. *J. Appl. Prob.* **55**, 1211–1227.
[23] Lodewijks, B. and Ortgiese, M. (2020). A phase transition for preferential attachment models with additive fitness. *Electron. J. Prob.* 25, paper no. 146, 54 pp.
[24] Mahmoud, H. M. (1992). Distances in random plane-oriented recursive trees. *J. Comput. Appl. Math.* **41**, 237–245.
[25] Mahmoud, H. M. (1992). *Evolution of Random Search Trees*. John Wiley, New York.
[26] Mahmoud, H. M., Smythe, R. T. and Szymański, J. (1993). On the structure of random plane-oriented recursive trees and their branches. *Random Structures Algorithms* **4**, 151–176.
[27] Móri, T. F. (2002). On random trees. *Studia Sci. Math. Hung.* **39**, 143–155.
[28] Na, H. S. and Rapoport, A. (1970). Distribution of nodes of a tree by degree. *Math. Biosci.* **6**, 313–329.
[29] Najock, D. and Heyde, C. C. (1982). On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *J. Appl. Prob.* **19**, 675–680.
[30] Nerman, O. (1981). On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrscheinlichkeitsth.* **57**, 365–395.
[31] Oliveira, R. and Spencer, J. (2005). Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.* **2**, 121–163.

https://doi.org/10.1017/apr.2022.40 Published online by Cambridge University Press

[32] PRODINGER, H. AND URBANEK, F. J. (1983). On monotone functions of tree structures. *Discrete Appl. Math.* **5**, 223–239.

[33] RUDAS, A., TÓTH, B. AND VALKÓ, B. (2007). Random trees and general branching processes. *Random Structures Algorithms* **31**, 186–202.

[34] SÉNIZERGUES, D. (2021). Geometry of weighted recursive and affine preferential attachment trees. *Electron. J. Prob.* **26**, paper no. 80, 56 pp.

[35] SZABÓ, G., ALAVA, M. AND KERTÉSZ, J. (2002). Shortest paths and load scaling in scale-free trees. *Phys. Rev. E* **66**, paper no. 026101.

[36] SZYMAŃSKI, J. (1987). On a nonuniform random recursive tree. In *Random Graphs '85 (Poznań, 1985)*, North-Holland, Amsterdam, pp. 297–306.

[37] VAN DER HOFSTAD, R. (2016). *Random Graphs and Complex Networks*, Vol. **1**. Cambridge University Press.

[38] VAN DER HOFSTAD, R. (2021). *Random Graphs and Complex Networks*, Vol. **2**. Preprint. Available at https://www.win.tue.nl/rhofstad/NotesRGCNII.pdf.