Condensation phenomena in preferential attachment trees with neighbourhood influence

Fountoulakis, Nikolaos; Iyer, Tejas

DOI:
10.1214/22-EJP787

Citation for published version (Harvard):
Fountoulakis, N & Iyer, T 2022, 'Condensation phenomena in preferential attachment trees with neighbourhood influence', *Electronic Journal of Probability*, vol. 27, 76. https://doi.org/10.1214/22-EJP787

Electronic Journal of Probability

Electron. J. Probab. **27** (2022), article no. 76, 1–49.
ISSN: 1083-6489 https://doi.org/10.1214/22-EJP787

Condensation phenomena in preferential attachment trees with neighbourhood influence*

Nikolaos Fountoulakis[†] Tejas Iyer[‡]

Abstract

We introduce a model of evolving preferential attachment trees where vertices are assigned weights, and the evolution of a vertex depends not only on its own weight, but also on the weights of its neighbours. We study empirical measures associated with edges with endpoints having certain weights, and degrees of vertices having a given weight. We show that the former exhibits a condensation phenomenon under a certain critical condition, whereas the latter behaves like a mixture of a power law distribution, depending on the weight distribution. Moreover, in the absence of condensation, for any measurable set we prove almost-sure convergence of the measure of that set under the related measure. This generalises existing results on the Bianconi-Barabási tree as well as on an evolving tree model introduced by the second author. Finally, as an application of our results, we provide criteria under which the degree distribution of this model behaves like a power law, and prove a limiting statement about the growth of the neighbourhood of a fixed vertex.

Keywords: preferential attachment trees; random recursive trees; Pólya processes; scale-free.
MSC2020 subject classifications: 90B15; 60J20; 05C80.
Submitted to EJP on January 26, 2021, final version accepted on April 27, 2022.

1 Introduction

1.1 Background

Complex networks appearing in areas as diverse as the internet, social networks and telecommunications are well known for their ubiquitous, non-trivial properties; in particular, they often have a scale free, i.e., a power law, degree distribution, and display a small or ultra-small world phenomenon, that is, having diameter of logarithmic or double logarithmic order with respect to the size of the network. In their seminal paper, Albert and Barabási in [4], observed that these properties emerged naturally in a model where vertices arrive one at a time, and display a "preference" to popular

*Research supported by the EPSRC grant EP/P026729/1.
[†]School of Mathematics, University of Birmingham, United Kingdom. E-mail: n.fountoulakis@bham.ac.uk
[‡]Weierstrass Institute for Applied Analysis and Stochastics, Germany. E-mail: iyer@wias-berlin.de

vertices – more precisely, connect to existing vertices with probability proportional to their degree. This model was later studied rigorously in [9, 37]. In the case where the newly arriving vertex connects to a single existing vertex, this gives rise to a well-known model of random trees that has been studied under various names: first under the name *ordered recursive tree* by Prodinger and Urbanek in [41], *nonuniform recursive trees* by Szymański in [46], random *plane oriented recursive trees* in [33, 34], random *heap ordered recursive trees* [12] and *scale-free trees* [9, 45, 8]. Various other modifications of this model have also been studied, including the case that vertices are chosen according to a *super-linear* function of their degree in [39], or indeed any positive function of the degree [42], assuming a certain technical condition is satisfied. In [24], the latter model is generalised to arbitrary non-negative functions of the degree and is referred to as *generalised preferential attachment*.

Whilst the preferential attachment model is successful in reproducing the properties of complex networks, it is generally the earlier arriving vertices that are more likely to have higher degrees, since (informally) they have more time to acquire new neighbours, which in turn reinforces the growth of their degree. Indeed, a result of [15] shows that, from a certain time point onward, the vertex with maximal degree remains fixed in this model. In contrast, in real world models it is often newly arriving nodes that quickly acquire a large number of links, for example, in the world wide web. Motivated by this, in [6], Bianconi and Barabási introduced their well-known model, sometimes called *preferential attachment with multiplicative fitness*. There, vertices arrive one at a time, and, upon arrival, each vertex is equipped with a random weight sampled independently from a fixed distribution. At each time-step, the newly arriving vertex u connects to an existing vertex v with probability proportional to the product of the weight of v and its degree. Thus, the random weight may be interpreted as a measure of the intrinsic "attractiveness" of a vertex. Bianconi and Barabási postulated the emergence of an interesting dichotomy in this model which they called *Bose-Einstein condensation*, motivated by similar phenomena in statistical physics. This *condensation* phenomenon refers to the fact that under a certain critical condition on the weight distribution, a positive proportion of all the edges in tree accumulate around vertices of maximum weight. This dichotomy was first proved rigorously by Borgs et al. in [10] in the case that the weight distribution is supported on an interval, and absolutely continuous with respect to Lebesgue measure. However, they note that other classes of weight distribution are possible. They also showed that in this model, the degree distribution of vertices with a given weight follows a power law, with exponent depending on the weight of the vertex. A similar condensation phenomenon was observed in a variant of this model by Dereich in [13], and later, in a more general, robust setting (in the sense that the results apply to wide variety of model specifications) in [16].

Two other similar models are the *preferential attachment with additive fitness* introduced by Ergün and Rodgers in [17], where newly arriving vertices now connect to existing vertices with probability proportional to the sum of their weight and degree, and the *weighted recursive tree* introduced in [11]. In [43], Sénizergues showed that the preferential attachment with additive fitness with deterministic weights, is equal in distribution to a particular weighted random recursive tree with random weights. In addition, Lodewijks and Ortgiese in [32, 31] uncovered an interesting dichotomy in the maximal degrees of these models, in a robust, evolving graph setting. In [25], the second author studied a model incorporating the weighted recursive tree as well as preferential attachment trees with both additive and multiplicative fitness: here at each time-step vertex with weight w and degree k is chosen with probability proportional to $g(w)(k-1)+h(w)$, where g, h are non-negative, measurable functions. In this case, the dynamics of the model depend on h in an non-trivial way: under a certain critical condition on the weight distribution, g and h condensation occurs, but does not occur if h takes large enough values on certain parts of its domain.

In the case of evolving trees, many of the above models describe the family tree of associated continuous time branching processes – often *Crump-Mode-Jagers* or multitype branching processes. This perspective has offered some interesting insights into the evolution of these models. For example, the preferential attachment tree of Albert and Barabási was actually first described in the context of evolution by Yule in [47] and in the context of language by Simon in [44]. In addition, the *condensation* phenomenon observed by Bianconi and Barabási was first studied in a similar, yet simpler manner, in the context of evolution by Kingman in [30]. Later, the results of [39, 42, 24, 1] have all exploited the connection to branching processes to derive results related to more general preferential attachment models, and in [25, 5] in relation to inhomogeneous models with a 'fitness' component. Often, the associated branching process with the discrete time model is known as the *continuous time embedding*, or *Arthreya-Karlin embedding*, based on pioneering work by Arthreya and Karlin in [3] who applied this approach in the context of Pólya urns. As shown in [24, 5], when studying 'local' properties such as degrees of vertices, one can observe that the continuous time embedding is a Crump-Mode-Jagers branching process, and apply the results of [38], whilst when studying properties such as the height (which is the same order of magnitude as the diameter), one can apply the results of [29] and an argument of Pittel [40].

In [14], the authors studied condensation in models of reinforced branching processes that generalise the continuous time embedding of the Bianconi-Barabási model, showing that the condensation is *non-extensive*: whilst a positive proportion of edges in the family tree of the process accumulate around vertices of maximal weight, the maximal degree of the tree remains sub-linear. In addition, in [20], the authors studied another generalisation of the continuous time embedding of the Bianconi-Barabási model, incorporating 'aging' effects, and applying this to the study of citation networks; they demonstrated a dichotomy between degree distributions having power law and exponential tails based on the aging parameter.

There are a number of other interesting variations of inhomogeneous preferential attachment models which also incorporate some degree of neighbourhood dependence. In [27], Jordan studies models of preferential attachment where vertices belong to two types and new vertices connect to existing vertices via a fitness mechanism, depending on the type. Newly arriving vertices are then assigned types randomly, depending on the types of their neighbours. Depending on the model specification, the fitness mechanism here can be preferential attachment with either additive or multiplicative fitness. Geometric models have also been considered in [28]: here, new vertices are equipped with a location in a metric space, and connect to existing vertices with probability proportional to the product of their degree, and a positive function of the distance between them. This positive function is known as an attractiveness function. In [28], the authors demonstrate a dichotomy, depending on the attractiveness function, between behaviour according to the model of Albert and Barabási, and a well known geometric model known as the on line nearest neighbour model.

Inhomogeneous models have also been studied in the context of models with *choice* in [19, 21], with the appearance of more fascinating condensation phenomena. In this model vertices are equipped with weights, at each time step r vertices are chosen with probability proportional to their degree, and out of these r vertices, a random vertex is chosen as the neighbour of the new-coming vertex. Here, the probability distribution by which the random vertex is chose, may depend on the weights of the vertices. In [19], the authors showed that, in the case that the maximal weight vertex is chosen, *extensive condensation* may occur, that is, under a critical condition on the weight distribution, a positive proportion of edges accumulate around the vertex of maximal degree. In addition, in [21], the authors showed that in certain cases, with random choice rules, the distribution of edges with endpoint having certain weight converges weakly to a *random measure* where *multiple condensation* can occur with positive probability, that is, positive proportions of edges accumulate around vertices of multiple weights. In addition, they showed that multiple condensation cannot occur when deterministic choice rules are used, and there exist phase transitions for condensation occurring with probability 0 or 1.

1.2 The definition of the neighbourhood influence model

As we discussed above, a number of preferential attachment mechanisms which incorporate inhomogeneity have been considered. However, models where the attachment mechanism depends on the *weights* of the *neighbours* of a vertex have received far less attention. In this direction, the authors in [18] recently incorporated *higher-dimensional interactions* into this notion of preferential attachment, studying a model of evolving simplicial complexes. They proved convergence in probability of the limiting degree distribution to a limiting value, depending on a *companion Markov process* that tracks the evolution of the neighbourhood of a given vertex. In this paper, we study a simplified version of that model, which involves evolving *trees*; as a result, we are able to derive stronger statements.

More precisely, we consider a model of *weighted directed trees* $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$; these are labelled directed trees, where vertices have real valued weights associated to them. Let $\mathbb{T}$ denote the set of all such weighted trees, and given a tree $\mathcal{T} \in \mathbb{T}$ and a vertex $j \in \mathcal{T}$, let $N^+(j,\mathcal{T})$ be the weighted tree consisting of j and all of its *out-neighbours*. In order to define the model, we require a probability measure μ which is supported on a bounded subset of $[0,\infty)$, and a *fitness function* $f : \mathbb{T} \to [0,\infty)$. Without loss of generality, we may assume that μ is supported on a subset of $[0,1]$.

In the model we consider, we start with an initial tree $\mathcal{T}_0$ consisting of a single vertex with random weight W_0 sampled from μ. Then, given $\mathcal{T}_i$, the model proceeds recursively as follows:

(i) Sample a vertex j from $\mathcal{T}_i$ with probability $\frac{f(N^+(j,\mathcal{T}_i))}{\mathcal{Z}_i}$, where $\mathcal{Z}_i := \sum_{k=0}^{i} f(N^+(k,\mathcal{T}_i))$ is the *partition function* associated with the process.

(ii) Form $\mathcal{T}_{i+1}$ by adding the edge $(j, i+1)$, and assigning vertex $i+1$ weight W_{i+1} sampled independently from μ.

In this paper, we define f so that

$$f(N^+(v,\mathcal{T})) = h(W_v) + \sum_{(v,u)\in E(\mathcal{T})} g(W_v, W_u), \tag{1.1}$$

where $h : [0,1] \to [0,\infty)$ and $g : [0,1] \times [0,1] \to [0,\infty)$ are bounded and measurable. Without loss of generality, we assume that h takes values within the interval $[0,1]$. In addition, to ensure that the evolution of the model is well-defined, in all of our results we condition on W_0 satisfying $h(W_0) > 0$, which we assume is an event that has positive probability.

Remark 1.1. The form of the fitness function in (1.1) is sufficiently general to encompass some existing models. In the case where g and h are a single constant, we obtain the classic preferential attachment tree of Albert and Barabási. The case $g(x,y) = h(x) = x$ is the Bianconi-Barabási model, whilst the case $g(x,y) \equiv 1, h(x) = x$ is the preferential attachment tree with additive fitness. Finally, the case $g(x,y) = \hat{g}(x)$, for some bounded

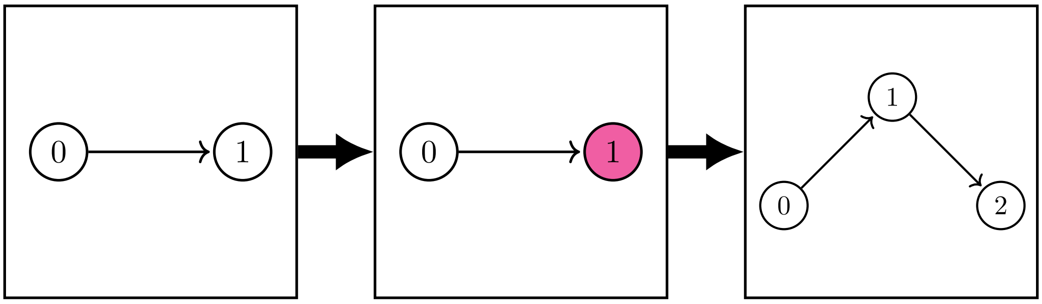

Figure 1: **Dynamics of the Model**. A sample transition from $\mathcal{T}_1$ to $\mathcal{T}_2$. In $\mathcal{T}_1$, 0 is chosen with probability proportional to $f(N^+(0,\mathcal{T}_1)) = h(W_0) + g(W_0, W_1)$, while 1 is chosen with probability proportional to $f(N^+(1,\mathcal{T}_1)) = h(W_1)$. In this evolution, 1 is chosen, so the newcomer 2 arrives as an out-neighbour of 1.

measurable function of a single variable is the generalised preferential attachment with fitness model studied by the second author in [25]. More importantly, this model is general enough to encompass functions $g(x,y)$ with a non-trivial dependence on the second variable y: some more general examples of fitness function to which our first main theorem, Theorem 1.1, applies are provided in Remark 1.4, whilst Example 1.10 provides some examples of fitness functions to which Theorem 1.3 applies.

Remark 1.2. One may interpret $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ in the context of reinforced branching processes as follows: we begin with an *individual* 0 belonging to its own *family* that reproduces after an exponentially distributed amount of time, with parameter $h(W_0)$. We say that the *ancestral weight* of the family is W_0. Then, recursively, when a birth event occurs in the ith family, with ancestral weight W_i, a new individual with random weight W joins the ith family, reproducing after an $\mathrm{Exp}\,(g(W_i, W))$-distributed amount of time; and simultaneously, an individual of weight W begins its own family, with ancestral weight W. The out-neighbourhood of a vertex i in the tree $\mathcal{T}_n$, including the vertex i itself, then represents individuals in the ith family in the branching process, at the time of the nth birth event.

Remark 1.3. One can extend the model from the previous remark further by supplanting it with constants $0 \le \beta, \gamma \le 1$, so that when a birth event occurs, independently with probability β, an individual with random weight W joins the ith family, and with probability γ, an individual with random weight W' (also sampled from μ) initiates its own family with ancestral weight W'. While not immediately clear from the way we have defined the model, our methods also extend to this case – this link becomes clearer when viewing individuals as “loops” and “edges” in a Pólya urn similar to Urn E (see Figure 2 below). In this extended model, the case $g(x,y) = h(x) = x$, and this terminology, was introduced in [14], as a stochastic analogue of the model of Kingman [30].

Let $\mathscr{B}$ denote the Borel σ-algebra on $[0,1]$, and $\mathscr{B} \otimes \mathscr{B}$ the product σ-algebra on $[0,1] \times [0,1]$. In this paper, we study the following quantities:

1. Given $A \in \mathscr{B} \otimes \mathscr{B}$, the quantity $\Xi^{(2)}(n, A)$ denotes the number of edges (v, v') in the tree $\mathcal{T}_n$ such that $(W_v, W_{v'}) \in A$, that is,

$$\Xi^{(2)}(n, A) := \sum_{(v,v')\in\mathcal{T}_n} \mathbf{1}_A(W_v, W_{v'}); \tag{1.2}$$

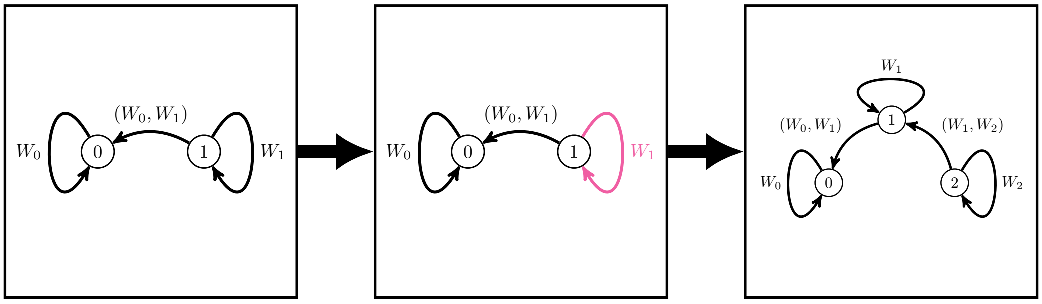

Figure 2: **The Dynamics of Urn E**. The evolution of the tree from $\mathcal{T}_1$ to $\mathcal{T}_2$ from Figure 1 viewed as a transition in Urn E. The event vertex 1 is selected may be interpreted as the event that the 'loop' W_1 is selected in the Pólya urn - and thus the arrival of the vertex 2 corresponds to the arrival of the 'loop' W_2 and the 'edge' (W_1, W_2) in the Pólya urn.

2. Given $B \in \mathscr{B}$, the quantity $N_{\geq k}(n, B)$ denotes the number of vertices v in the tree $\mathcal{T}_n$ with out-degree at least k and weight $W_v \in B$, that is,

$$N_{\geq k}(n, B) := \sum_{v \in \mathcal{T}_n : \deg^+(v, \mathcal{T}_n) \geq k} \mathbf{1}_B(W_v). \tag{1.3}$$

3. For $B \in \mathscr{B}$, we also define $\Xi(n, B)$, so that

$$\Xi(n, B) := \sum_{(v, v') \in \mathcal{T}_n} \mathbf{1}_B(W_v) = \Xi^{(2)}(n, B \times [0, 1]), \tag{1.4}$$

where the latter equality is in the almost sure sense.

1.2.1 Notation

We denote by $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ – i.e. the natural numbers including 0. Also, in general in this paper, W refers to a generic μ distributed random variable on a probability space $(\Omega, \mathbb{F}, \mathbb{P})$ taking values in the measure space $([0, 1], \mathscr{B})$ and $\mathbb{E}[\cdot]$ denotes expectations with respect to this random variable. In addition, we require a probability space with an infinite sequence $W_0, W_1, W_2, \ldots$ of random variables, which are independent and identically distributed; we view these, abusing notation slightly, as belonging to the product space $(\mathbf{\Omega}, \mathbf{F}, \mathbb{P}) := (\prod_{i \in \mathbb{N}_0} (\Omega_i, \mathbb{F}_i, \mathbb{P}_i)$. For brevity, $\mathbb{E}[\cdot]$ also denotes expectations with respect to random variables on this product space.

In addition, for $s \in \mathbb{N}$, we denote by $[s]$ the set $\{1, \ldots, s\}$. In addition, for $\ell \in \mathbb{N}$, we denote by $[s]^\ell$ the ℓ-fold Cartesian product $[s] \times \cdots \times [s]$. Given a set $S \subset \mathcal{S}$, we denote by S^c the complement of this set, and (if $\mathcal{S}$ has a topology made clear from context), we denote by $\overline{S}$ the topological closure of S. We also denote the indicator function associated with S by $\mathbf{1}_S$. Finally, we introduce some extra notation specific to the section in Section 2.1.2.

1.3 Statements of the main results

The results in this paper depend on two sets of conditions; intuitively one set of conditions describes the 'non-condensation' regime, whilst the other describes the 'condensation' regime.

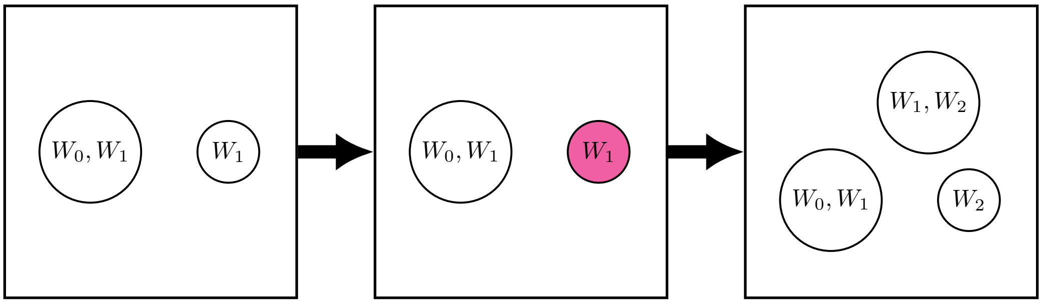

Figure 3: **The Dynamics of Urn D**. The evolution of the tree from $\mathcal{T}_1$ to $\mathcal{T}_2$ from Figure 1 viewed as a transition in Urn D. The event vertex 1 is selected may be interpreted as the event that the ball W_1 is selected in the Pólya urn - and thus the arrival of the vertex 2 corresponds to the addition of the balls W_2 and (W_1, W_2). The latter ball represents the addition of vertex 2 into the neighbourhood of vertex 1.

1.3.1 The non-condensation regime of the model

The first main conditions are the following: recalling g and h as defined in (1.1), assume

C1 There exists some $\lambda^* > \tilde{g}^*$ such that

$$\mathbb{E}\left[\frac{h(W)}{\lambda^* - \tilde{g}(W)}\right] = 1, \tag{1.5}$$

where $\tilde{g}(x) := \mathbb{E}\left[g(x, W)\right]$ and $\tilde{g}^* := \mathbb{E}\left[\sup_{x \in [0,1]} g(x, W)\right]$. We call λ^* the *Malthusian parameter* of the process.

C2 For $N \in \mathbb{N}$, there exist measurable functions $\phi_j^{(i)} : [0,1] \to [0,1]$, $j = 1, 2$, $i \in [N]$, and a bounded continuous function $\kappa : [0,1]^{2N} \to [0, \infty)$ such that

$$g(x, y) = \kappa\left(\phi_1^{(1)}(x), \ldots, \phi_1^{(N)}(x), \phi_2^{(1)}(y), \ldots, \phi_2^{(N)}(y)\right). \tag{1.6}$$

We denote by $g_{\max} := \sup_{x,y \in [0,1]} \{g(x, y)\}$.

Remark 1.4. We expect similar results under the weaker hypothesis that g and h are measurable and bounded rather than Condition **C2**. However, this condition still allows many "reasonable" choices of bounded measurable functions g. This includes the models mentioned in Remark 1.1, the case where g is continuous, as well as functions of the form $g(x, y) = \phi_1(x) + \phi_2(y)$ or $g(x, y) = \phi_1(x)\phi_2(y)$, where ϕ_1, ϕ_2 are bounded and measurable.

In order to prove our first theorem, we will require some definitions. Define $\psi(x) = h(x)/(\lambda^* - \tilde{g}(x))$, denote by $\psi_*\mu$ the pushforward measure of μ under ψ – i.e. the measure such that for $A \in \mathscr{B}$

$$(\psi_*\mu)(A) = \mathbb{E}\left[\frac{h(W)}{\lambda^* - \tilde{g}(W)}\mathbf{1}_A(W)\right]. \tag{1.7}$$

In addition, we define a *companion process* $(S_i(w))_{i \geq 0}$ that describes the evolution of the *fitness* of a vertex with weight w as its neighbourhood changes. First, let $W_1, W_2, \ldots$ be independent μ-distributed random variables and let $w \in [0,1]$. We then define the random process $(S_i(w))_{i \geq 0}$ inductively so that

$$S_0(w) := h(w); \quad S_{i+1}(w) := S_i(w) + g(w, W_{i+1}),\ i \geq 0. \tag{1.8}$$

Also recall from Section 1.2.1, that $\mathbb{E}[\cdot]$ also denotes expectation with respect to the path of $S_i(W_0)$, i.e., expectations with respect to the product measure involving the terms $W_0, W_1, W_2 \ldots$. We then have the following theorem:

Theorem 1.1. Assume Conditions **C1** and **C2**. Then:

1. We have $\lim_{n\to\infty} \frac{\mathcal{Z}_n}{n} \to \lambda^*$ almost surely.
2. With $\Xi^{(2)}(n,\cdot)$ as defined in (1.2), the sequence of random measures

$$\frac{\Xi^{(2)}(n,\cdot)}{n} \to (\psi_*\mu \times \mu)(\cdot), \tag{1.9}$$

almost surely, with respect to the weak topology.

3. For any $A \in \mathscr{B}$, we have

$$\lim_{n\to\infty} \frac{N_{\geq k}(n,A)}{n} = \mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W_0)}{S_i(W_0)+\lambda^*}\right)\mathbf{1}_A(W_0)\right], \tag{1.10}$$

almost surely.

4. For any set $B \in \mathscr{B}$ we have

$$\frac{\Xi(n,B)}{n} \to (\psi_*\mu)(B),$$

almost surely.

Remark 1.5. One may interpret the right hand side of (1.10) as the probability of a sequence of at least k consecutive heads before a first tail when, sampling W_0 at random, and flipping the ith coin heads with probability proportional to $S_{i-1}(W_0)$.

Remark 1.6. By applying the Portmanteau theorem, and the well known fact that any open set in $[0,1]$ may be expressed as a countable disjoint union of half open dyadic intervals, the convergences in Assertions 3 and 4 of Theorem 1.1 are stronger than almost sure convergence of the sequences of random measures in the weak topology, which is the usual form in which these results appear in the literature, (e.g. [10, 14]). However, it is not immediately clear whether one can swap the "for all" quantifier and the "almost sure" statement, so that the respective convergence occurs almost surely for all sets $B \in \mathscr{B}$. This latter convergence would be almost sure setwise convergence of the sequences of random measures, and it may be the case that this convergence occurs as well.

In order to prove Theorem 1.1, we require the following interesting identity, which may be of independent interest.

Lemma 1.2. Let $(S_i(w))_{i\geq 0}$ denote the process defined in (1.8) in terms of bounded, measurable functions g, h, suppose $\tilde{g}(x) := \mathbb{E}\left[g(x,W)\right]$ and $\tilde{g}_+ = \sup_{x\in[0,1]} \tilde{g}(x)$. Then for any $w \in [0,1]$ and $\lambda \geq \tilde{g}_+$ we have

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(w)}{S_i(w)+\lambda}\right)\right] = \frac{h(w)}{\lambda - \tilde{g}(w)}, \tag{1.11}$$

where the right hand side is infinite if $\lambda = \tilde{g}_+ = \tilde{g}(w)$. In particular,

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W_0)}{S_i(W_0)+\lambda}\right)\mathbf{1}_B(W_0)\right] = \mathbb{E}\left[\frac{h(W_0)}{\lambda - \tilde{g}(W_0)}\mathbf{1}_B(W_0)\right].$$

As the proof of this lemma detracts from the main techniques used in this paper, we include its proof in the appendix, Section 4.1.

Remark 1.7. One may interpret (1.11) as a generalisation of the classic geometric series formula: if we set $g(x, y) \equiv 0$, and $q := h(w)/(h(w) + \lambda)$, the left hand side of (1.11) is $\sum_{i=1}^{\infty} q^i = \frac{h(w)}{\lambda} = \frac{q}{1-q}$. Indeed, as Remark 1.5 shows, one may interpret the left hand side as the expected value of a generalised geometrically distributed random variable.

1.3.2 The condensation regime of the model

In this paper, we are able to describe a "condensation" result; we first make precise what "condensation" means.

Definition. Suppose we are given a μ-null set $S \in \mathscr{B}$ and let $\Xi(n, \cdot)$ be as in (1.2). We say that *condensation* occurs around the set S, if for some decreasing collection of sets $(S_\varepsilon)_{\varepsilon \geq 0}$, with $S_\varepsilon \downarrow S$ as $\varepsilon \to 0$ we have

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\Xi(n, S_\varepsilon)}{n} > 0,$$

with positive probability.

Remark 1.8. Informally, condensation means that, in the limit of the random measure $\Xi(n, \cdot)/n$, the set S acquires more mass than one 'would expect'. Indeed, if we swap limits,

$$\lim_{n \to \infty} \lim_{\varepsilon \to 0} \frac{\Xi(n, S_\varepsilon)}{n} = \lim_{n \to \infty} \frac{\Xi(n, S)}{n} = 0,$$

almost surely, since $\mu(S) = 0$.

Our main assumptions in the condensation regime are:

D1 We have

$$\mathbb{E}\left[\frac{h(W)}{\tilde{g}^* - \tilde{g}(W)}\right] < 1. \tag{1.12}$$

D2 The function g satisfies Condition **C2**.

D3 There exists a (maximal) set of points $\mathcal{M} \subseteq [0, 1]$, such that, for any $z \in \mathcal{M}$ for all $w \in [0, 1]$,

$$\operatorname{ess\,sup}(g(\cdot, w)) = g(z, w).$$

D4 For some $\varepsilon_0 \in (0, 1)$, there exists a family of measurable functions $\{u_\varepsilon : [0, 1] \to [0, \infty)\}_{\varepsilon \in (0, \varepsilon_0)}$ such that $\lim_{\varepsilon \to 0} u_\varepsilon = 0$ pointwise and for all $z \in \mathcal{M}$, we have

$$\begin{aligned} \mathcal{M}_\varepsilon &:= \{x : \mathbb{P}\left(g(z, W) - g(x, W) < u_\varepsilon(W)\right) = 1\} \\ &= \{x : \mathbb{P}\left(g(z, W) - g(x, W) < u_\varepsilon(W)\right) > 0\}. \end{aligned} \tag{1.13}$$

In other words, the set $\mathcal{M}_\varepsilon$ of x such that $g(z, W) - g(x, W) < u_\varepsilon(W)$ with positive probability, satisfy $g(z, W) - g(x, W) < u_\varepsilon(W)$ almost surely. We also assume that $\mu(\mathcal{M}_\varepsilon) > 0$.

Remark 1.9. Note that, by the measurability of $g(\cdot, q)$ for any $q \in [0, 1]$, for all $z \in \mathcal{M}$ the function

$$p \mapsto \operatorname{ess\,sup}_{q \in [0,1]} \{g(z, q) - g(p, q) - u_\varepsilon(q)\}$$

is also measurable as the lattice supremum of an uncountable family of measurable functions – see, e.g. [7, Theorem 4.7.1.]. This ensures that the set $\mathcal{M}_\varepsilon \in \mathscr{B}$.

Example 1.10. The conditions **D2-D4** are restrictive enough that it may not be the case that they are satisfied by any continuous function $g(x, y)$. However, they still encompass the other examples stated in Remark 1.4: functions of the form $g(x, y) = \phi_1(x) + \phi_2(y)$ or $g(x, y) = \phi_1(x)\phi_2(y)$, where ϕ_1, ϕ_2 are bounded and measurable. Indeed, in either case, $\mathcal{M} = \phi_1^{-1}(\{\operatorname{ess\,sup}(\phi_1)\})$, and $\mathcal{M}_\varepsilon = \phi_1^{-1}((\operatorname{ess\,sup}(\phi_1) - \varepsilon, \operatorname{ess\,sup}(\phi_1)])$. Indeed, in the case that $g(x, y) = \phi_1(x) + \phi_2(y)$ we may simply choose $u_\varepsilon \equiv \varepsilon$, so that

$$\begin{aligned}\mathcal{M}_\varepsilon &= \{x : \mathbb{P}\left(\phi_1(z) + \phi_2(W) - \phi_1(x) - \phi_2(W) < \varepsilon\right) = 1\} \\ &= \{x : \phi_1(z) - \phi_1(x) < \varepsilon\} = \phi_1^{-1}((\operatorname{ess\,sup}(\phi_1) - \varepsilon, \operatorname{ess\,sup}(\phi_1)]).\end{aligned}$$

while in the case that $g(x, y) = \phi_1(x)\phi_2(y)$ for bounded, measurable ϕ_1, ϕ_2, for $\varepsilon > 0$ and $z \in \mathcal{M}$ we may take $u_\varepsilon = \varepsilon \cdot \phi_2$ and

$$\begin{aligned}\mathcal{M}_\varepsilon &= \{x : \mathbb{P}\left(\phi_1(z)\phi_2(W) - \phi_1(x)\phi_2(W) < \varepsilon\phi_2(W)\right) = 1\} \\ &= \{x : \phi_1(z) - \phi_1(x) < \varepsilon\} = \phi_1^{-1}((\operatorname{ess\,sup}(\phi_1) - \varepsilon, \operatorname{ess\,sup}(\phi_1)]).\end{aligned}$$

By the definition of the essential supremum, $\mathcal{M}_\varepsilon$ has positive measure.

Remark 1.11. Conditions **D1** and **D2** may be interpreted as analogues of Conditions **C1** and **C2** in the condensation regime. One may regard $\mathcal{M}$ from **D3** as a “dominating set”, in the sense that $\mathbb{P}$-a.s., upon arrival of a new vertex into its neighbourhood, the change of the fitness of any vertex is at most the change of the fitness of a vertex with weight with weight in $\mathcal{M}$. Condition **D4** ensures that this “dominating property” is captured by sets $\mathcal{M}_\varepsilon$ of positive measure. Indeed the right hand side of (1.13) implies that the change of the fitness of any vertex with weight in $\mathcal{M}_\varepsilon^c$ is at most the change of the fitness of a vertex having weight in $\mathcal{M}_\varepsilon$. Note that $\mathcal{M}_\varepsilon \downarrow \mathcal{M}$ as $\varepsilon \to 0$. This accounts for the formation of the condensate in Theorem 1.3, since $\tilde{g}$ is maximised on $\mathcal{M}$, by **D1** it must be the case that $\mu(\mathcal{M}) = 0$.

Theorem 1.3. Assume Conditions **D1-D4**. Then:

1. We have $\lim_{n\to\infty} \frac{\mathcal{Z}_n}{n} \to \tilde{g}^*$ almost surely.
2. For any $A \in \mathscr{B}$ such that, for $\varepsilon > 0$ sufficiently small $A \cap \mathcal{M}_\varepsilon = \varnothing$, we have
$$\frac{\Xi(n, A)}{n} \to (\psi_*\mu)(A), \quad \text{almost surely.} \tag{1.14}$$
In addition,
$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} \frac{\Xi(n, \mathcal{M}_\varepsilon)}{n} = 1 - (\psi_*\mu)([0, 1]) > 0, \tag{1.15}$$
so that condensation occurs around $\mathcal{M}$.
3. For any set $B \in \mathscr{B}$, almost surely, we have
$$\lim_{n\to\infty} \frac{N_{\geq k}(n, B)}{n} = \mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W) + \tilde{g}^*}\right)\mathbf{1}_B(W)\right].$$

Remark 1.12. The condensation phenomenon has an interesting evolutionary interpretation in the context of reinforced branching processes (see Remarks 1.2 and 1.3). Here, ‘fitness’ refers to the rate at which a family produces offspring (a natural parameter for reproductive success), and condensation occurs around families of maximally reinforced fitness. Thus, viewing this model as a stochastic analogue of the model of Kingman in [30], one may informally view this as an extreme “survival of the fittest” phenomenon: despite any single individual having zero probability of weight belonging to $\mathcal{M}$, individuals in the population are skewed so much towards weights conferring higher reinforced fitness that a positive proportion of individuals in the population tend towards weights belonging to the set $\mathcal{M}$.

We have the following corollary:

Corollary 1.4. Assume Conditions **D1-D4**, and the sets $\mathcal{M}_\varepsilon$ in **D4** are such that $\overline{\mathcal{M}}_\varepsilon \downarrow \mathcal{M}$ as $\varepsilon \to 0$ (recalling $\overline{\mathcal{M}}_\varepsilon$ denotes the topological closure of $\mathcal{M}_\varepsilon$). Also, suppose that $\mathcal{M} = \{z\}$, and define the measure $\Pi(\cdot)$ such that, for $B \in \mathscr{B}$

$$\Pi(B) = (\psi_* \mu)(B) + (1 - (\psi_* \mu)([0,1]))\, \delta_z(B).$$

Then,

$$\frac{\Xi(n, \cdot)}{n} \to \Pi(\cdot) \quad \text{almost surely},$$

in the weak topology.

Remark 1.13. Recalling the examples from Example 1.10, i.e., $g(x,y) = \phi_1(x)\phi_2(y)$ or $g(x,y) = \phi_1(x)+\phi_2(y)$ for bounded measurable functions ϕ_1, ϕ_2, an additional assumption that meets the requirement that $\overline{\mathcal{M}}_\varepsilon \downarrow \mathcal{M}$ is that ϕ_1 is continuous on some $\overline{\mathcal{M}}_{\varepsilon'}$. Indeed, in this case, for $0 < \varepsilon < \varepsilon'$,

$$\overline{\mathcal{M}}_\varepsilon \subseteq \{x : \phi_1(z) - \phi_1(x) \leq \varepsilon\},$$

so that $\overline{\mathcal{M}}_\varepsilon \downarrow \{z\}$ as $\varepsilon \to 0$. Another assumption that would meet this requirement would be to have ϕ_1 monotone.

1.4 Discussion of some implications of the main results

In this subsection, we provide a discussion of some of the implications of our main results.

1.4.1 Power law degree distributions

We will show the following theorem, which essentially states that the number of vertices of degree k scales like a power law with a certain exponent, for k that is large. First, we define c^* such that

$$c^* := \inf_{\lambda > 0} \left\{ \lambda : 0 < \mathbb{E}\left[\frac{h(W)}{\lambda - \tilde{g}(W)}\right] \leq 1 \right\}$$

Note that this implies that $c^* = \lambda^*$ for the non-condensation regime (Conditions **C1** and **C2**) and $c^* = \tilde{g}^*$ for the condensation regime (Conditions **D1-D4**).

Theorem 1.5. Assume that, for some ε_0 sufficiently small, h is positive on the set

$$\{w : \operatorname{ess\,sup}\{\tilde{g}\}(1 - \varepsilon_0) \leq \tilde{g}(w) \leq \operatorname{ess\,sup}\{\tilde{g}\}\},$$

and that $\inf_{w \in [0,1]} \tilde{g}(w) > 0$. Then, assuming Conditions **C1** and **C2** or Conditions **D1-D4** are satisfied,

$$\lim_{k \to \infty} \log_k(p_k) = -\left(1 + \frac{c^*}{\operatorname{ess\,sup}\{\tilde{g}\}}\right). \tag{1.16}$$

In this subsection, we provide a proof of this theorem.

Remark 1.14. In much of the scientific literature surrounding the study of complex networks (see, for example, [23]), scientists observe that the degree distribution behaves like a power law, with exponent τ between 2 and 3. A weak definition of this power law behaviour is to define $\tau := \lim_{k \to \infty} -\log_k(p_k)$. In this case, since $h(w) \leq 1$, whenever $\operatorname{ess\,sup}\{\tilde{g}\} = 1$ (so that $c^* \leq 2$) this behaviour always emerges in this model. In the condensation regime however, we have $\operatorname{ess\,sup}\{\tilde{g}\} = \tilde{g}^*$, so that the exponent of the power law is always 2.

First note that by Theorem 1.1, if $N_k(n,B)$ denotes the number of vertices with degree k and weight belonging to B at time n, then almost surely

$$\begin{aligned}\lim_{n\to\infty}\frac{N_k(n,B)}{n} &= \lim_{n\to\infty}\left(\frac{N_{\geq k}(n,B)}{n}-\frac{N_{\geq k+1}(n,B)}{n}\right)\\ &= \mathbb{E}\left[\frac{c^*}{S_k(W)+c^*}\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+c^*}\right)\mathbf{1}_B(W)\right]. \qquad (1.17)\end{aligned}$$

In what follows, we define the measures $p_k(\cdot)$ such that, for any set $B\in\mathscr{B}$ we have

$$p_k(B) := \mathbb{E}\left[\frac{c^*}{S_k(W)+c^*}\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+c^*}\right)\mathbf{1}_B(W)\right].$$

The following inequality is well-known.

Lemma 1.6 (Hoeffding's inequality [22], Theorem 2.5 in [36])**.** Suppose $X_1,\dots,X_n$ are random variables such that $a\leq X_i\leq b$ for $i\in\{1,\dots,n\}$; set $X:=\sum_{i=1}^n X_i$ and $\tilde{X}:=\mathbb{E}\left[X\right]$. Then, for any $\delta>0$ we have

1. $\mathbb{P}\left(X\geq(1+\delta)\tilde{X}\right)\leq e^{-\frac{2\delta^2\tilde{X}^2}{n(b-a)^2}}$
2. $\mathbb{P}\left(X\leq(1-\delta)\tilde{X}\right)\leq e^{-\frac{2\delta^2\tilde{X}^2}{n(b-a)^2}}$.

Lemma 1.7. For any $B\in\mathscr{B}$ such that $\mu(B)>0$, and $\inf_{w\in B}\tilde{g}(w)>0$ we have

$$\limsup_{k\to\infty}\log_k p_k(B)\leq-\left(1+\frac{c^*}{\sup_{w\in B}\tilde{g}(w)}\right). \qquad (1.18)$$

and, moreover, if $h(w)>0$ on $A\subseteq B$ with $\mu(A)>0$,

$$\liminf_{k\to\infty}\log_k p_k(B)\geq-\left(1+\frac{c^*}{\inf_{w\in B}\tilde{g}(w)}\right). \qquad (1.19)$$

Proof. As a shorthand, we define the following quantities for a given $B\in\mathscr{B}$:

$$\tilde{g}_-(B):=\inf_{w\in B}\tilde{g}(w),\quad \tilde{g}_+(B):=\sup_{w\in B}\tilde{g}(w). \qquad (1.20)$$

Also, for any $w\in B,\delta>0$, define the event $\mathcal{G}_{w,\delta,k_0}$ such that

$$\mathcal{G}_{w,\delta,k_0}=\left\{\forall j\geq k_0\ \ |S_j(w)-(h(w)+j\tilde{g}(w))|\leq\delta(h(w)+j\tilde{g}(w))\right\}.$$

By Lemma 1.6, for any $\delta>0,k_0\in\mathbb{N}$, with J being an upper bound on $\max\{g,1\}$ we have

$$\begin{aligned}\mathbb{P}\left(\mathcal{G}^c_{w,\delta,k_0}\right)&\leq\sum_{j=k_0}^{\infty}e^{\frac{-2\delta^2\tilde{g}(w)^2j}{J^2}}\leq\left(1+\frac{J^2}{2\delta^2\tilde{g}(w)^2}\right)e^{\frac{-2\delta^2\tilde{g}(w)^2k_0}{J^2}}\\ &\leq\left(1+\frac{J^2}{2\delta^2\tilde{g}_-(B)^2}\right)e^{\frac{-2\delta^2\tilde{g}_-(B)^2k_0}{J^2}}=:ce^{-dk_0}, \qquad (1.21)\end{aligned}$$

where we have used the integral test to bound the infinite series, and define the constants c and d, which are independent of k_0, to simplify notation. We begin with the proof of (1.19). First, we fix k_0 sufficiently large that the right hand side of (1.21) is smaller than one; this implies the events $\{\mathcal{G}_{w,\delta,k_0}\}_{w\in B}$ have probability uniformly bounded from below, so that

$$\inf_{w\in B}\mathbb{P}\left(\mathcal{G}_{w,\delta,k_0}\right)>0. \qquad (1.22)$$

Then,

$$
\begin{aligned}
p_k(B) &\geq \frac{c^*}{(k+1)J+\lambda^*}\mathbb{E}\left[\mathbb{E}_W\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+c^*}\right)\mathbf{1}_{\mathcal{G}_{W,\delta,k_0}}\right]\mathbf{1}_B(W)\right] \\
&\geq \frac{c^*}{(k+1)J+\lambda^*} \\
&\times \mathbb{E}\left[\mathbb{E}_W\left[\left(\prod_{i=0}^{k_0-1}\frac{S_i(W)}{S_i(W)+c^*}\right)\prod_{i=k_0}^{k-1}\left(\frac{(h(W)+i\tilde{g}(W))(1-\delta)}{(h(W)+i\tilde{g}(W))(1-\delta)+c^*}\right)\mathbf{1}_{\mathcal{G}_{W,\delta,k_0}}\right]\mathbf{1}_B(W)\right] \\
&\geq \frac{c^*}{(k+1)J+\lambda^*} \\
&\times \mathbb{E}\left[\mathbb{E}_W\left[\left(\frac{h(W)}{h(W)+c^*}\right)^{k_0}\prod_{i=k_0}^{k-1}\left(\frac{i\tilde{g}_-(B)(1-\delta)}{i\tilde{g}_-(B)(1-\delta)+c^*}\right)\mathbf{1}_{\mathcal{G}_{W,\delta,k_0}}\right]\mathbf{1}_B(W)\right],
\end{aligned}
$$

where in these inequalities we have used the fact that $x \mapsto \frac{x}{x+c^*}$ is increasing. Now, re-writing the product in the last expectation with gamma functions, we have

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{E}_W\left[\left(\frac{h(W)}{h(W)+c^*}\right)^{k_0}\prod_{i=k_0}^{k-1}\left(\frac{i\tilde{g}_-(B)(1-\delta)}{i\tilde{g}_-(B)(1-\delta)+c^*}\right)\mathbf{1}_{\mathcal{G}_{W,\delta,k_0}}\right]\mathbf{1}_B(W)\right] \\
&\quad= \mathbb{E}\left[\left(\frac{h(W)}{h(W)+c^*}\right)^{k_0}\frac{\Gamma(k)\,\Gamma(k_0+c^*/((1-\delta)\tilde{g}_-(B)))}{\Gamma(k+c^*/((1-\delta)\tilde{g}_-(B)))\,\Gamma(k_0)}\mathbb{E}_W\left[\mathbf{1}_{\mathcal{G}_{W,\delta,k_0}}\right]\mathbf{1}_B(W)\right] \\
&\quad\geq \mathbb{E}\left[\left(\frac{h(W)}{h(W)+c^*}\right)^{k_0}\mathbf{1}_B(W)\right]\frac{\Gamma(k)\,\Gamma(k_0+c^*/((1-\delta)\tilde{g}_-(B)))}{\Gamma(k+c^*/((1-\delta)\tilde{g}_-(B)))\,\Gamma(k_0)}\inf_{w\in B}\mathbb{P}\left(\mathcal{G}_{w,\delta,k_0}\right) \\
&\quad= c_1\frac{\Gamma(k)}{\Gamma(k+c^*/((1-\delta)\tilde{g}_-(B)))} = c_1\left(1-\frac{c_2}{k}\right)k^{-c^*/((1-\delta)\tilde{g}_-(B))};
\end{aligned}
$$

where c_1 is a constant depending only on B and k_0, and c_2 is a constant independent of k coming from applying Stirling's approximation to the ratio of gamma functions. Thus,

$$
\begin{aligned}
\liminf_{k\to\infty}\log_k p_k(B) &\geq \liminf_{k\to\infty}\log_k\left(\frac{c_1c^*}{J}\left(1-\frac{c_2}{k}\right)k^{-(1+c^*/((1-\delta)\tilde{g}_-(B))}\right) \\
&= -\left(1+\frac{c^*}{(1-\delta)\tilde{g}_-(B)}\right),
\end{aligned}
$$

and, as δ can be made arbitrarily small, (1.19) follows.

Now, for (1.18), we again use (1.21), but we instead define $k_0 = k_0(k)$ as a function of k; in particular, for a given $\varepsilon > 0$, we set $k_0 := \lfloor k^{\varepsilon}\rfloor$. Then,

$$
\begin{aligned}
p_k(B) &\leq \mathbb{E}\left[\mathbb{E}_W\left[\frac{c^*}{S_k(W)}\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+c^*}\right)\mathbf{1}_{\mathcal{G}_{W,\delta,k_0}}\right]\mathbf{1}_B(W)\right] + \sup_{w\in B}\mathbb{P}\left(\mathcal{G}^c_{w,\delta,k_0}\right) \\
&\leq \mathbb{E}\left[\mathbb{E}_W\left[\frac{c^*}{k\tilde{g}(W)(1-\delta)}\prod_{i=k_0}^{k-1}\left(\frac{h(W)+i\tilde{g}(W)(1+\delta)}{h(W)+i\tilde{g}(W)(1+\delta)+c^*}\right)\mathbf{1}_{\mathcal{G}_{W,\delta,k_0}}\right]\mathbf{1}_B(W)\right] \\
&\qquad + \sup_{w\in B}\mathbb{P}\left(\mathcal{G}^c_{w,\delta,k_0}\right) \\
&\leq \frac{c^*}{k\tilde{g}_+(B)(1-\delta)}\frac{\Gamma(k+1)\,\Gamma(k_0+1+c^*/((1+\delta)\tilde{g}_+(B)))}{\Gamma(k+1+c^*/((1+\delta)\tilde{g}_+(B)))\,\Gamma(k_0+1)} + ce^{-dk_0},
\end{aligned}
$$

where, in the last inequality, we have bounded $h(W)$ above by 1, and $\tilde{g}(W)$ by $\tilde{g}_+(B)$. Now, bounding k_0 above by k^{ε}, and applying Stirling's approximation, there exists a constant c_3, independent of k, such that

$$p_k(B) \le c_3 k^{-(1+(c^*/((1+\delta)\tilde{g}_+(B)))(1-\varepsilon))} + ce^{-dk^{\varepsilon}}.$$

Therefore,

$$\limsup_{k\to\infty} \log_k p_k(B) \le -\left(1 + (c^*/((1+\delta)\tilde{g}_+(B)))\,(1-\varepsilon)\right),$$

and making ε and δ arbitrarily small proves (1.18). □

We are now ready to prove Theorem 1.5.

Proof of Theorem 1.5. First, we set $\tilde{g}' := \operatorname{ess\,sup}\{\tilde{g}\}$, and, for each $n \in \mathbb{N}$, define the following partition $\{\mathcal{P}_i\}_{i\in\{0\}\cup[n-1]}$ of $[0,1]$:

$$\mathcal{P}_0 := \tilde{g}^{-1}\left([0, \tilde{g}'/n]\right), \quad \mathcal{P}_i := \tilde{g}^{-1}\left((\tilde{g}'i/n, \tilde{g}'(i+1)/n]\right), \quad \text{for } i \in [n-1].$$

Now the lower bound for (1.16) follows immediately from (1.19) in Lemma (1.7): for $n \ge \frac{1}{\varepsilon_0}$ the function h is positive on $\mathcal{P}_{n-1}$, thus, by (1.19)

$$\liminf_{k\to\infty} \log_k p_k \ge \liminf_{k\to\infty} \log_k\left(p_k(\mathcal{P}_{n-1})\right) \ge -\left(1 + \frac{c^*}{\tilde{g}'(1-\frac{1}{n})}\right),$$

and sending n to ∞, we deduce the claim. For the upper bound, we first fix n and, by applying (1.18), choose k sufficiently large that for all $i \in \{0\} \cup [n-2]$

$$p_k(\mathcal{P}_i) < k^{\frac{1}{2n} - \left(1+\frac{c^* n}{\tilde{g}'(i+1)}\right)} \quad \text{and} \quad p_k(\mathcal{P}_{n-1}) > k^{-\frac{1}{2n} - \left(1+\frac{c^* n}{\tilde{g}'(n-1)}\right)}. \tag{1.23}$$

Then we have

$$\begin{aligned}
\limsup_{k\to\infty} \log_k p_k &= \limsup_{k\to\infty} \log_k\left(\sum_{i=0}^{n-1} p_k(\mathcal{P}_i)\right) \\
&\le \limsup_{k\to\infty} \log_k\left(p_k(\mathcal{P}_{n-1})\right) + \limsup_{k\to\infty} \log_k\left(\sum_{i=0}^{n-1} p_k(\mathcal{P}_i)/p_k(\mathcal{P}_{n-1})\right)
\end{aligned}$$

Now, by (1.23), we have

$$\log_k\left(\sum_{i=0}^{n-1} p_k(\mathcal{P}_i)/p_k(\mathcal{P}_{n-1})\right) \le \log_k\left(nk^{1/n}\right),$$

so that

$$\limsup_{k\to\infty} \log_k p_k \le \limsup_{k\to\infty} \log_k\left(p_k(\mathcal{P}_{n-1})\right) + \frac{1}{n} \le -\left(1 + \frac{c^*}{\tilde{g}'}\right) + \frac{1}{n}.$$

Sending $n \to \infty$, we deduce (1.16). □

1.4.2 The growth of the neighbourhood of a fixed vertex

In the following proposition, we let $f_n(v) = f(N^+(v, \mathcal{T}_n))$ denote the fitness, as defined in (1.1), of a vertex labelled $v \in \mathbb{N}_0$, with weight w_v in the tree at time n. In addition, let $(R_i)_{i\ge v}$ denote the natural filtration generated by the tree process $(\mathcal{T}_i)_{i\ge v}$ after the arrival of v. Next, set

$$M_n(v) := \frac{f_n(v)}{\prod_{s=v}^{n-1}\left(\frac{\mathcal{Z}_s + \tilde{g}(w_v)}{\mathcal{Z}_s}\right)}.$$

Proposition 1.8. For any vertex $v \in \mathbb{N}_0$, $(M_n(v))_{n \geq v}$ is a non-negative martingale with respect to $(R_i)_{i \geq v}$.

Proof. Using the definition of the process, for $n \geq v$ we compute

$$\begin{aligned}\mathbb{E}\left[f_{n+1}(v)|R_n\right] &= \frac{f_n(v)}{\mathcal{Z}_n}\left(f_n(v) + \tilde{g}(w_v)\right) + \left(1 - \frac{f_n(v)}{\mathcal{Z}_n}\right) f_n(v) \\ &= f_n(v)\left(\frac{\mathcal{Z}_n + \tilde{g}(w_v)}{\mathcal{Z}_n}\right).\end{aligned}$$

The result follows from the definition of $(M_n(v))_{n \geq v}$. □

Now, let $\deg_n^+(v)$ denote the out-degree of vertex v at time n in the tree process. We then have the following corollary:

Corollary 1.9. For every $v \in \mathbb{N}_0$, we have

$$\frac{\deg_n^+(v)}{\prod_{s=v}^{n-1}\left(\frac{\mathcal{Z}_s + \tilde{g}(w_v)}{\mathcal{Z}_s}\right)} \to \frac{M_\infty(v)}{\tilde{g}(w_v)} \tag{1.24}$$

almost surely, where $M_\infty(v)$ is an almost surely finite, non-negative random variable, with $\mathbb{E}\left[M_\infty(v)\right] \leq h(w_v)$.

Proof. First note that by the martingale convergence theorem, we have

$$M_n(v) \to M_\infty(v), \tag{1.25}$$

almost surely, where $M_\infty(v)$ is as described in the statement of the corollary. Noting that for each n we have $\mathbb{E}\left[M_n(v)\right] = h(w_v)$, by Fatou's lemma we have $\mathbb{E}\left[M_\infty(v)\right] \leq h(w_v)$ Now, we associate with each fixed vertex v in the process a sequence of stopping times $(\tau_j(v))_{j \in \mathbb{N}}$, describing the time-step in the tree process where v is selected for the jth time. Formally, if $\deg_n^+(v)$ denotes the out-degree of vertex v at time n, we set $\tau_j(v) := \inf\left\{n \geq 0 : \deg_n^+(v) = j\right\}$, following the convention that the infimum of the empty set is $+\infty$. Now, note that for each $n \in \mathbb{N}$, we have the deterministic bound $\mathcal{Z}_n \leq 2Jn$. Therefore, for each $n > v$, on the event that $h(w_v) > 0$, and conditionally on w_v, the event of selecting v at the nth time-step is stochastically bounded below by an independent Bernoulli trial with parameter $h(w_v)/2Jn$, and thus by the converse of the Borel-Cantelli lemma, almost surely, on $\{h(w_v) > 0\}$, v is selected infinitely often. Therefore, on the event $\{h(w_v) > 0\}$, almost surely, for all $j \in \mathbb{N}, \tau_j(v) < \infty$. On the other hand, by the construction of the process, the entire sequence

$$\left(\frac{f_{\tau_j(v)}(v)}{\deg^+_{\tau_j(v)}(v)}\right)_{j \in \mathbb{N}} = \left(\frac{h(w_v) + \sum_{i=1}^{j} g(w_v, W_i)}{j}\right)_{j \in \mathbb{N}}, \tag{1.26}$$

in distribution, where the W_i are i.i.d random variables sampled from μ. It follows from the strong law of large numbers, and the fact that the processes $(f_n(v))_{n \geq v}, (\deg_n^+(v))_{n \geq v}$ are piecewise constant, that (conditionally on the event $\{h(w_v) > 0\}$)

$$\lim_{n \to \infty} \frac{f_n(v)}{\deg_n^+(v)} = \tilde{g}(w_v) \tag{1.27}$$

almost surely. The result follows from (1.25) and (1.27), where we note that we may drop the conditioning $\{h(w_v) > 0\}$ from the statement of the corollary since both sides are 0 on $\{h(w_v) = 0\}$. □

For our next corollary, recall that we define c^* such that

$$c^* := \inf_{\lambda > 0} \left\{ \lambda : 0 < \mathbb{E}\left[\frac{h(W)}{\lambda - \tilde{g}(W)} \right] \le 1 \right\}.$$

Note that Proposition 1.8 is not enough to guarantee that $M_\infty > 0$ with positive probability, although we do conjecture that this is the case in the Non-Condensation regime. Nevertheless, we do have the following result:

Corollary 1.10. Assume that either Conditions **C1** and **C2** or Conditions **D1-D4** are satisfied, and the event $\{M_\infty(v) > 0\}$ occurs with positive probability. Then, almost surely, on $\{M_\infty(v) > 0\}$, we have

$$\lim_{n\to\infty} \frac{\log \deg_n^+(v)}{\log n} = \frac{\tilde{g}(w_v)}{c^*}.$$

Proof. Note that since the event $\{M_\infty(v) > 0\}$ occurs with positive probability, when we condition on this event, almost sure events still occur almost surely. Thus, by the first part of Theorems 1.1 and 1.3, respectively, we have $\frac{1}{i}\mathcal{Z}_i \to c^*$, almost surely under either set of conditions, and by Corollary 1.9, we have

$$\frac{\deg_n^+(v)}{\prod_{s=v}^{n-1}\left(\frac{\mathcal{Z}_s + \tilde{g}(w_v)}{\mathcal{Z}_s}\right)} \to \frac{M_\infty(v)}{\tilde{g}(w_v)} \quad \text{almost surely.} \tag{1.28}$$

Since we condition on the event $\{M_\infty(v) > 0\}$, we may take logarithms on both sides of (1.28), and since $M_\infty < \infty$ almost surely, deduce that

$$\lim_{n\to\infty} \frac{\log \deg_n^+(v)}{\log n} - \frac{\log\left(\prod_{s=v}^{n-1}\left(\frac{\mathcal{Z}_s + \tilde{g}(w_v)}{\mathcal{Z}_s}\right)\right)}{\log n} = 0.$$

By Egorov's theorem, for any $\ell \in \mathbb{N}$, there exists an event $\mathcal{B}_\ell$ such that $\mathbb{P}(\mathcal{B}_\ell) < \frac{1}{\ell}$, and on the complement $\mathcal{B}_\ell^c$, $\frac{\mathcal{Z}_i}{i} \to c^*$ uniformly. In particular, on the event $\mathcal{B}_\ell^c$, for any $\varepsilon > 0$ there exists $i_0 = i_0(\varepsilon, \ell) > 0$ such that for any $i > i_0$ we have $|\mathcal{Z}_i - ic^*| < i\varepsilon$. Then, a computation involving Stirling's approximation, similar to the one displayed in the proof of Lemma 1.7, implies that on $\mathcal{B}_\ell^c$ we have

$$\frac{\tilde{g}(w_v)}{c^* + \varepsilon} < \liminf_{n\to\infty} \frac{\log \deg_n^+(v)}{\log n} \le \limsup_{n\to\infty} \frac{\log \deg_n^+(v)}{\log n} < \frac{\tilde{g}(w_v)}{c^* - \varepsilon}.$$

Sending $\varepsilon \to 0$, we deduce that on each $\mathcal{B}_\ell^c$ we have $\lim_{n\to\infty} \frac{\log \deg_n^+(v)}{\log n} = \frac{\tilde{g}(w_v)}{c^*}$. Thus, the event that this convergence does not occur is contained in the event $\bigcap_{\ell=1}^\infty \mathcal{B}_\ell$, and

$$\mathbb{P}\left(\bigcap_{\ell=1}^\infty \mathcal{B}_\ell\right) < 1/\ell,$$

for each $\ell \in \mathbb{N}$. The result follows. □

1.5 Overview and techniques

1.5.1 Overview

In Section 2 we prove results about the model related to the non-condensation regime. We first review some background theory about *Pólya urns* in Section 2.1, and then, the results of Section 2.2 are used in order to prove Assertions 1 and 2 of Theorem 1.1 in Section 2.2.4. Next, the results of Section 2.3 are used to complete the proof of Theorem 1.1 in Section 2.3.4. In Section 3 we extend the previous results to the condensation regime, proving Theorem 1.3 and Corollary 1.4 in Section 3.1 and Section 3.2 respectively. We prove Lemma 1.2 in the Appendix, in Section 4.1.

1.5.2 Techniques applied in this paper

This paper generalises the techniques used in [10] for the study of the Bianconi-Barabási model – using a *Pólya urn approximation*. However, the generalisation of this model to bounded measurable functions h, functions g satisfying Condition **C2**, and the possibility of arbitrary weight distributions lead to technical challenges, somewhat analogous to those arising from using a measure-theoretic approach to integration as opposed to the Riemann integral. Applying this approach to studying the degree distribution in the case of uncountably supported weight distributions also appears to be novel. In extending the results to the condensation regime we apply a similar coupling to that used in [25].

One might imagine that many of the results here may follow easily from an application of the theory of Crump-Mode-Jagers branching processes, for example as in [14]. However, the dependence between the point processes associated with a parent and its offspring means that the classic theory is not immediately applicable. This in turn raises the question of whether one can develop a theory of C-M-J branching processes with *dependencies* between the point-processes associated with individuals.

2 The non-condensation regime

2.1 A brief review of theory related to generalised Pólya urns

Generalised Pólya urns are a well studied family of stochastic processes representing the composition of an *urn* containing balls with certain *types*. If $\mathscr{T}$ denotes the set of possible types, associated to a ball of type $t \in \mathscr{T}$ is a non-negative *activity* $\mathbf{a}(t)$, which depends on the type. The process then evolves in discrete time so that, at each time-step, a ball of type t is sampled at random from the urn with probability proportional to its activity $\mathbf{a}(t)$, and replaced with balls of a number of different types according to a possibly random *replacement rule*.

In the case that $\mathscr{T}$ is finite, the configuration of the urn after n replacements may be represented as a *composition vector* $(X_n)_{n\in\mathbb{N}_0}$ with entries labelled by type, and the activities encoded in an *activity vector* $\mathbf{a}$. In this vector, the ith entry corresponds to the number of balls of type $i \in \mathscr{T}$. Let $(\xi_{ij})_{i,j\in\mathscr{T}}$ be the matrix whose ijth component denotes the random number of balls of type j added, if a ball of type i is drawn, and (following the notation of Janson in [26]) define the matrix A such that $A_{ij} := a_j\mathbb{E}\left[\xi_{ji}\right]$. The (expected) evolution of the urn in the $(n+1)$st step may therefore be obtained by applying the matrix A to the composition vector X_n. A type $i \in \mathscr{T}$ is said to be *dominating* if, for any $j \in \mathscr{T}$, it is possible to obtain a ball of type j starting with a ball of type i. If we write $i \sim j$ for the equivalence relation where $i \sim j$ if it is possible to obtain j starting from a ball of type i, and vice versa. This partitions the types into equivalence *classes*. A class $\mathscr{C} \subseteq \mathscr{T}$ is *dominating* if, for every $i \in \mathscr{C}$, i is dominating. Moreover, the eigenvalues of A may be obtained by the restriction of A to its classes; we say an eigenvalue belongs to a *dominating class* if it is an eigenvalue of the restriction of A to this class. Finally, we say that the urn, or the matrix A, is *irreducible* if there is only one dominating class. Note the difference when compared to irreducible matrices in the context of Markov chains: here it is possible for diagonal entries to be negative. Now, assume the following conditions are satisfied:

(A1) For all $i, j \in \mathscr{T}$, $\xi_{ij} \geq 0$ if $i \neq j$ and $\xi_{ii} \geq -1$.
(A2) For all $i, j \in \mathscr{T}$, $\mathbb{E}\left[\xi_{ij}^2\right] < \infty$.
(A3) The largest real eigenvalue λ_1 of A is positive.
(A4) The largest real eigenvalue λ_1 is simple.
(A5) We start with at least one ball of a dominating type.
(A6) λ_1 belongs to the dominating class.

The following is a well known result of Janson from 2004 building on previous work by by Athreya and Karlin (for example, [3, Proposition 2] and [2, Theorem 5]):

Theorem 2.1 ([26, Theorem 3.16])**.** Assume Conditions (A1)-(A6), and suppose that v_1 denotes the right eigenvector, corresponding to the leading eigenvalue λ_1 of A, normalised so that $\mathbf{a}^T v_1 = 1$. Then, we have

$$\frac{X_n}{n} \xrightarrow{n\to\infty} \lambda_1 v_1,$$

almost surely, conditional on essential non-extinction (i.e. non-extinction of balls of dominating type).

In addition, the following lemma by Janson provides convenient criteria for satisfying (A1)-(A6):

Lemma 2.2 ([26, Lemma 2.1])**.** If A is irreducible, (A1) and (A2) hold, $\sum_{j\in\mathscr{T}} \mathbb{E}\left[\xi_{ij}\right] \geq 0$ for all $i \in \mathscr{T}$, with the inequality being strict for some $i \in \mathscr{T}$, then (A1)–(A6) are satisfied and essential extinction does not occur.

2.1.1 Analysing the tree using Pólya urns

The idea behind analysing the distribution of edges with a given weight, and the degree distribution in this model, is to consider two different types of Pólya urns, which we call *Urn E* and *Urn D* respectively. We illustrate the evolution of both these urns below. Recall, Figure 1 illustrates a possible evolution of a step of the process $(\mathcal{T}_i)_{i\in\mathbb{N}_0}$; Figures 2 and 3 illustrate the corresponding steps in Urn E and Urn D.

In Urn E, we consider a generalised Pólya urn with balls of two types: singletons x, and ordered pairs (x, y), corresponding to 'loops' and 'edges', respectively. A ball of type (x, y) has activity $g(x, y)$ and a ball of type x has activity $h(x)$. At each step, if a ball of activity x or (x, y) is selected, we introduce two new balls, of which one has random type W, and the other has type (x, W). In relation to the evolving tree, this corresponds to the event that a vertex of weight x has been sampled in the subsequent step. In Urn D, we consider a generalised Pólya urn with balls of types corresponding to tuples of varying lengths. A ball of type $(x_0, \ldots, x_k)$ has activity $h(x_0) + \sum_{i=1}^{k} g(x_0, x_i)$, and at each step, if a ball this type is selected, we remove it and introduce two new balls: one of random type W, and one of type $(x_0, \ldots, x_k, W)$. In relation to the evolving tree, this corresponds to the event that a vertex v of weight x_0 has been sampled when proceeding to the subsequent step, with neighbours of v listed in order of arrival having weights $x_1, \ldots, x_k$.

Note that, in the manner we have described Urns E and D, the set of possible types may be infinite: the measure μ may have infinite support so that W may take on infinite values, and the neighbourhoods of vertices (in Urn D) may be infinite. Whilst there is some theory related to infinite type Pólya urns within the framework of measure-valued Pólya processes (see, for example, [35]), these results are often non-trivial to apply in practice – see, for example, [18, pages 14-21]. As a result, following ideas first used in [10], we instead approximate these infinite urns with urns of finitely many types – enough to approximate the sigma algebras generated by $W, g(W, W')$ and $h(W)$, where W, W' are i.i.d random variables sampled according to μ. In Section 2.2 we apply this analysis to Urn E, and in Section 2.3 we apply it to Urn D. We first introduce some extra notation specific to this section.

2.1.2 Some more notation and terminology used in Section 2

In order to apply the finite Pólya urn theory, given a set of types $\mathscr{T}$, we denote by $\mathbb{V}_{\mathscr{T}}$ the *free vector space* over the field $\mathbb{R}$ generated by $\mathscr{T}$, i.e., the vector space of formal linear combinations of elements of $\mathscr{T}$ with coefficients in $\mathbb{R}$. We generally view an urn with types $\mathscr{T}$ as a stochastic process taking values in $\mathbb{V}_{\mathscr{T}}$. In addition we generally identify vectors $\mathbf{v} \in \mathbb{V}_{\mathscr{T}}$ interchangeably with functions $\mathbf{v} : \mathscr{T} \to \mathbb{R}$. Thus, for $x \in \mathscr{T}$, $\mathbf{v}(x)$ denotes the coordinate of the vector corresponding to x, and for $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{V}_{\mathscr{T}}$, we set $(\mathbf{v}_1\mathbf{v}_2)(x) = \mathbf{v}_1(x)\mathbf{v}_2(x)$. For $x \in \mathscr{T}$, we define $\mathbf{e}_x \in \mathbb{V}_{\mathscr{T}}$ such that $\mathbf{e}_x(y) = 1$ if $y = x$ and 0 otherwise.

For a Borel measurable set $S \subseteq \mathbb{R}$, we say a finite collection of measurable sets $\mathscr{A}$ is a partition of S if the sets in $\mathscr{A}$ are pairwise disjoint and their union is S. Note that, given two partitions $\mathscr{A}_1, \mathscr{A}_2$ of S, the set

$$\{A_1 \cap A_2 : A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2\} \tag{2.1}$$

also forms a partition of S. In addition, if $\mathscr{A}$ is a partition of S, we say that $\mathscr{A}'$ forms a *refined partition* of $\mathscr{A}$, if, $\mathscr{A}'$ is a partition of S, and for any $A' \in \mathscr{A}'$ there exists $A \in \mathscr{A}$ such that $A' \subseteq A$. The following lemma, which is well-known, justifies the use of the word 'refined'.

Lemma 2.3. Suppose $\mathscr{A}$ is a partition of a set S, and $\mathscr{A}'$ is a refined partition of $\mathscr{A}$. Then, for any set $A \in \mathscr{A}$, there exist sets $X_1, \dots, X_s \in \mathscr{A}'$ such that $A = \bigcup_{i=1}^{s} X_i$. In particular, $\{X_i\}_{i\in[s]}$ forms a partition of A.

2.2 Analysing the tree by using urn E

In this subsection we work under the assumption that Conditions **C1** and **C2** hold. We analyse the process under these conditions by coupling the tree process $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ with Pólya urn processes, parametrised by $m \in \mathbb{N}$. These may be interpreted as finite approximations of Urn E. Now, for each $m \in \mathbb{N}$ we consider a particular partition of the interval $[0,1]$ into 2^m intervals, which is the *dyadic partition*: set

$$\mathcal{D}_1^m := [0, 2^{-m}], \quad \text{and} \quad \mathcal{D}_i^m := ((i-1)\cdot 2^{-m}, i \cdot 2^{-m}],\ i \in [2^m] \setminus \{1\}.$$

For $i \in [2^m]$, we also denote the closure of $\mathcal{D}_i^m(x)$ by $\overline{\mathcal{D}}_i^m(x)$, so that

$$\overline{\mathcal{D}}_i^m = [(i-1)\cdot 2^{-m}, i \cdot 2^{-m}].$$

Recalling that the function h takes values in $[0,1]$, and the definitions of the functions $\phi_1^{(j)}, \phi_2^{(j)}, j \in [N]$ from Condition **C2**, for each $i \in [2^m]$, $j \in [N]$ and $k \in [2]$, we set

$$\mathcal{H}_i^m := h^{-1}\left(\mathcal{D}_i^m\right) \quad \text{and } \Phi_k^m(i,j) := \left(\phi_k^{(j)}\right)^{-1}\left(\mathcal{D}_i^m\right).$$

By the measurability assumptions on the functions $\phi_k^{(j)}$ and h, for each $i \in [2^m]$ we have $\mathcal{H}_i^m, \Phi_i^m(j,k) \in \mathscr{B}$, and thus, the collections of sets $\{\mathcal{H}_i^m\}_{i\in[2^m]}$ and $\{\Phi_k^m(i,j)\}_{i\in[2^m]}$ form partitions of $[0,1]$. We now split the latter family of sets to form a refined partition: for $\mathbf{i} = (i_1, \dots, i_N), \mathbf{j} = (j_1, \dots, j_N) \in [2^m]^N$, if we set

$$\begin{aligned} &\Phi_1^m(\mathbf{i}) = \Phi_1^m(i_1, 1) \cap \Phi_1^m(i_2, 2) \cap \cdots \cap \Phi_1^m(i_N, N) \quad \text{and,} \\ &\Phi_2^m(\mathbf{j}) = \Phi_2^m(j_1, 1) \cap \Phi_2^m(j_2, 2) \cap \cdots \cap \Phi_2^m(j_N, N), \end{aligned} \tag{2.2}$$

by iteratively applying (2.1), the families of sets $\{\Phi_1^m(\mathbf{i})\}_{\mathbf{i}\in[2^m]^N}$ and $\{\Phi_2^m(\mathbf{j})\}_{\mathbf{j}\in[2^m]^N}$ also form partitions of $[0,1]$. Now, given $\mathbf{v} = (v_1, \dots, v_N) \in [2^m]^N$, set

$$\overline{\mathcal{D}}_{\mathbf{v}}^m := \overline{\mathcal{D}}_{v_1}^m \times \overline{\mathcal{D}}_{v_2}^m \times \cdots \times \overline{\mathcal{D}}_{v_N}^m,$$

and observe that, given $\mathbf{i}, \mathbf{j} \in [2^m]^N$, the construction of the sets in (2.2) are such that $(x,y) \in \Phi_1^m(\mathbf{i}) \times \Phi_2^m(\mathbf{j})$ implies that

$$\left(\phi_1^{(1)}(x), \dots, \phi_1^{(N)}(x), \phi_2^{(1)}(y), \dots, \phi_2^{(N)}(y)\right) \in \overline{\mathcal{D}}_{\mathbf{i}}^m \times \overline{\mathcal{D}}_{\mathbf{j}}^m$$

Now, recalling the function $\kappa : [0,1]^{2N} \to [0, g_{\max}]$ from Condition **C2**, for each $\mathbf{i}, \mathbf{j} \in [2^m]^N$, by continuity on the compact set $\overline{\mathcal{D}}_{\mathbf{i}}^m \times \overline{\mathcal{D}}_{\mathbf{j}}^m$, for $(x,y) \in \Phi_1^m(\mathbf{i}) \times \Phi_2^m(\mathbf{j})$ we have

$$\begin{aligned} \kappa\left(\phi_1^{(1)}(x), \dots, \phi_1^{(N)}(x), \phi_2^{(1)}(y), \dots, \phi_2^{(N)}(y)\right) &\geq \inf_{\mathbf{u},\mathbf{v} \in \overline{\mathcal{D}}_{\mathbf{i}}^m \times \overline{\mathcal{D}}_{\mathbf{j}}^m} \{\kappa(\mathbf{u},\mathbf{v})\} \\ &= \min_{\mathbf{u},\mathbf{v} \in \overline{\mathcal{D}}_{\mathbf{i}}^m \times \overline{\mathcal{D}}_{\mathbf{j}}^m} \{\kappa(\mathbf{u},\mathbf{v})\} =: \kappa_-(\mathbf{i},\mathbf{j}), \end{aligned} \tag{2.3}$$

and likewise,

$$\begin{aligned} \kappa\left(\phi_1^{(1)}(x), \dots, \phi_1^{(N)}(x), \phi_2^{(1)}(y), \dots, \phi_2^{(N)}(y)\right) &\leq \sup_{\mathbf{u},\mathbf{v} \in \overline{\mathcal{D}}_{\mathbf{i}}^m \times \overline{\mathcal{D}}_{\mathbf{j}}^m} \{\kappa(\mathbf{u},\mathbf{v})\} \\ &= \max_{\mathbf{u},\mathbf{v} \in \overline{\mathcal{D}}_{\mathbf{i}}^m \times \overline{\mathcal{D}}_{\mathbf{j}}^m} \{\kappa(\mathbf{u},\mathbf{v})\} =: \kappa_+(\mathbf{i},\mathbf{j}). \end{aligned} \tag{2.4}$$

Now, set

$$g_-(x,y) := \sum_{\mathbf{i},\mathbf{j} \in [2^m]^N} \kappa_-(\mathbf{i},\mathbf{j}) \mathbf{1}_{\Phi_1^m(\mathbf{i}) \times \Phi_2^m(\mathbf{j})}(x,y),$$

$$g_+(x,y) := \sum_{\mathbf{i},\mathbf{j} \in [2^m]^N} \kappa_+(\mathbf{i},\mathbf{j}) \mathbf{1}_{\Phi_1^m(\mathbf{i}) \times \Phi_2^m(\mathbf{j})}(x,y),$$

and

$$h_-(x) := \sum_{i=1}^{2^m} (i-1) \cdot 2^{-m} \mathbf{1}_{\mathcal{H}_i}(x), \quad h_+(x) := \sum_{i=1}^{2^m} i \cdot 2^{-m} \mathbf{1}_{\mathcal{H}_i}(x).$$

One should interpret these functions as *lower* and *upper* approximations to g and h, indeed, by construction, we now have the following lemma:

Lemma 2.4. We have $g_- \uparrow g$, $h_- \uparrow h$, $g_+ \downarrow g$ and $h_+ \downarrow h$ uniformly, as $m \to \infty$.

Proof. We prove the statements regarding h_- and g_-; the others follow analogously (in the case of g_+ using (2.4) instead of (2.3)). Since the sets $(\mathcal{H}_i^m)_{i \in [2^m]}$ form a partition of $[0,1]$, for each $m \in \mathbb{N}$, given $x \in [0,1]$, we have $x \in \mathcal{H}_j^m$ for some $j \in [2^m]$, and thus

$$h_-(x) = (j-1) \cdot 2^{-m} \leq h(x) \leq h_-(x) + 2^{-m}.$$

The convergence result for h_- follows. Now, note that by uniform continuity of κ on the compact set $[0,1]^{2N}$, for $\varepsilon > 0$, let M be sufficiently large so that for all $\mathbf{u}, \mathbf{v} \in [0,1]^{2N}$

$$\|\mathbf{u} - \mathbf{v}\| < \sqrt{2N} \cdot 2^{-M} 1 \quad \Longrightarrow \quad |\kappa(\mathbf{u}) - \kappa(\mathbf{v})| < \varepsilon. \tag{2.5}$$

Now, for any $m > M$, given $(x,y) \in [0,1] \times [0,1]$, there exists a unique set $\Phi_1^m(\mathbf{i}) \times \Phi_2^m(\mathbf{j})$ containing (x,y), which implies that

$$\left(\phi_1^{(1)}(x), \dots, \phi_1^{(N)}(x), \phi_2^{(1)}(y), \dots, \phi_2^{(N)}(y)\right) \in \overline{\mathcal{D}}_{\mathbf{i}}^m \times \overline{\mathcal{D}}_{\mathbf{j}}^m.$$

Thus, for each $j \in [N]$, combining this equation with the definition of $\kappa_-(\mathbf{i},\mathbf{j})$ from (2.3), we have

$$\kappa_-(\mathbf{i},\mathbf{j}) \leq \kappa\left(\phi_1^{(1)}(x), \dots, \phi_1^{(N)}(x), \phi_2^{(1)}(y), \dots, \phi_2^{(N)}(y)\right) \leq \kappa_-(\mathbf{i},\mathbf{j}) + \varepsilon,$$

and thus

$$g_-(x,y) \leq g(x,y) \leq g_-(x,y) + \varepsilon.$$

The result now follows. ☐

Now, using the partitions $\{\mathcal{H}_i^m\}_{i\in[2^m]}$, $\{\Phi_1^m(\mathbf{i})\}_{\mathbf{i}\in[2^m]^N}$, $\{\Phi_2^m(\mathbf{j})\}_{\mathbf{j}\in[2^m]^N}$ and $\{\mathcal{D}_i^m\}_{i\in[2^m]}$, we form an even more refined partition, which we use as the "building blocks" of the evolution of the Pólya urn approximations. For each m, define the partition $\mathscr{I}^m$ such that

$$\mathscr{I}^m := \left\{ I \in \mathscr{B} : I = \mathcal{H}_p^m \cap \mathcal{D}_q^m \cap \Phi_1^m(\mathbf{i}) \cap \Phi_2^m(\mathbf{j}),\ p, q \in [2^m], \mathbf{i}, \mathbf{j} \in [2^m]^N \right\}. \tag{2.6}$$

Intuitively, this family of sets is such that the finite σ-algebra $\sigma(\mathscr{I}^m)$, is "fine enough" to approximate $\mathscr{B}$, and also capture the behaviour of g and h. Observe that, for $m_1 < m_2$, $\mathscr{I}^{m_2}$ is a refined partition of $\mathscr{I}^{m_1}$.

Suppose $|\mathscr{I}^m| = D_m$; then we label the sets in $\mathscr{I}^m$ arbitrarily as $(\mathcal{I}_i^m)_{i\in[D_m]}$. Now, for each $(x,y) \in \mathcal{I}_i^m \times \mathcal{I}_j^m$, $g_-(x,y)$ and $g_+(x,y)$ are constant, depending only on (i,j), and likewise, for each $x \in \mathcal{I}_\ell^m$, $h_-(x)$ and $h_+(x)$ are constant, depending on ℓ. Motivated by this, for each $(i,j) \in [D_m] \times [D_m]$, we define the following quantities:

$$g_{\min}(i,j) := g_-(x,y), \quad g_{\max}(i,j) := g_+(x,y), \quad (x,y) \in \mathcal{I}_i^m \times \mathcal{I}_j^m, \tag{2.7}$$

and likewise, for each $\ell \in [D_m]$, we define

$$h_{\min}(\ell) := h_-(x), \quad h_{\max}(\ell) := h_+(x), \quad x \in \mathcal{I}_\ell^m, \tag{2.8}$$

We also set

$$r(x) := \sum_{i=1}^{D_m} i \mathbf{1}_{\mathcal{I}_i^m}(x), \tag{2.9}$$

so that $r(x) = i$ if $x \in \mathcal{I}_i^m$. In addition, set

$$p_i^m := \mu(\mathcal{I}_i^m), i \in [D_m], \quad g^*(j) := \max_{i\in[D_m]} \{g_{\max}(i,j)\},$$
$$\tilde{g}_-(i) := \sum_{j=1}^{D_m} p_j^m g_{\min}(i,j), \quad \tilde{g}_+(i) := \sum_{j=1}^{D_m} p_j^m g_{\max}(i,j), \quad \text{and} \quad \tilde{g}_+^* := \sum_{j=1}^{D_m} p_j^m g^*(j). \tag{2.10}$$

Recall that $\tilde{g}(x) = \mathbb{E}[g(x,W)]$, and note that $\tilde{g}_-(r(x)) = \mathbb{E}[g_-(x,W)]$, $\tilde{g}_+(r(x)) = \mathbb{E}[g_+(x,W)]$ and $\tilde{g}_+^* = \mathbb{E}\left[\max_{x\in[0,1]} g_+(x,W)\right]$. Then, observe that by Lemma 2.4 and dominated convergence, $\tilde{g}_-(r(x)) \uparrow \tilde{g}(x)$, $\tilde{g}_+(r(x)) \downarrow \tilde{g}(x)$ and

$$\tilde{g}_+^* \downarrow \mathbb{E}\left[\sup_{x\in[0,1]} g(x,W)\right] = \tilde{g}^*, \quad \text{as } m \to \infty.$$

2.2.1 Definitions of the urn schemes associated with urn E

We are now ready to define the urn process $(\mathcal{U}_n)_{n\in\mathbb{N}_0}$. For $i \in \mathbb{N}$, set

$$[D_m]^i := [D_m] \times [D_m] \cdots \times [D_m] = \{(u_0, \dots u_{i-1}) : u_0, \dots, u_{i-1} \in [D_m]\},$$

and

$$\mathcal{B} := [D_m] \cup [D_m]^2 \cup (\{D_m + 1\} \times [D_m]);$$

this is the set of types in Urn E. We now define parameters $\boldsymbol{\gamma}$ such that, for $x \in [D_m] \cup [D_m] \times [D_m]$,

$$\boldsymbol{\gamma}(x) = \begin{cases} \frac{g_{\min}(i,j)}{g_{\max}(i,j)}, & x = (i,j) \in [D_m]^2, g_{\max}(i,j) > 0; \\ \frac{h_{\min}(i)}{h_{\max}(i)}, & x = i \in [D_m], h_{\max}(i) > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

Then, we define the urn process $(\mathcal{U}_n^m)_{n \in \mathbb{N}_0}$ as the urn process with *activities* $\mathbf{a}$ such that

$$\mathbf{a}(x) = \begin{cases} g_{\max}(i,j) & \text{if } x = (i,j),\ i,j \in [D_m] \\ g^*(j) & \text{if } x = (i,j),\ i = D_m + 1, j \in [D_m] \\ h_{\max}(i) & \text{if } x = i \in [D_m]; \end{cases} \tag{2.12}$$

and a *replacement matrix* M such that, for $x, x' \in \mathbb{V}_{\mathcal{B}}$,

$$M_{x',x} = \begin{cases} (\boldsymbol{\gamma}\mathbf{a})(x) p_\ell^m, & \text{if } x' = (i,\ell), x \in (\{i\} \times [D_m]) \cup \{i\}, i, \ell \in [D_m]; \\ (\mathbf{a} - \boldsymbol{\gamma}\mathbf{a})(x) p_\ell^m, & \text{if } x' = (D_m + 1, \ell)\,, x \in \mathcal{B}; \\ \mathbf{a}(x) p_\ell^m, & \text{if } x' = \ell, x \in \mathcal{B}; \\ 0 & \text{otherwise.} \end{cases}$$

Note that it is not necessarily the case that M is irreducible: it may be the case that $\mathbf{a}(x) = 0$ for certain $x \in \mathcal{B}$ (this is possible if $h_{\max}(i) = 0$ or $g_{\max}(i,j) = 0$), or it may be the case that $p_\ell^m = 0$ for certain choices of ℓ. We therefore define the following subsets of $\mathcal{B}$:

$$\mathscr{U}_1 := \{x \in \mathcal{B} : M_{x'x} = 0\ \forall x' \in \mathcal{B}\} = \{x \in \mathcal{B} : \mathbf{a}(x) = 0\}\,,$$

and

$$\mathscr{U}_2 := \{x' \in \mathcal{B} : M_{x'x} = 0\ \forall x \in \mathcal{B}\}\,.$$

Also assume that $\mathscr{U}_1 \cap \mathscr{U}_2 = \varnothing$; if not, we replace $\mathscr{U}_1$ by $\mathscr{U}_1 \setminus \mathscr{U}_2$. We then set $R = \mathcal{B} \setminus (\mathscr{U}_1 \cup \mathscr{U}_2)$, and let M_R be the restriction of M to R. It is easy to check that M_R is irreducible, and thus, by Lemma 2.2, has a unique largest positive eigenvalue λ_m with corresponding eigenvector $\mathbf{u}_R$. But then, writing M in block form (with columns and rows labelled by $R, \mathscr{U}_1, \mathscr{U}_2$) for suitable matrices A, B, C, we have

$$M = \begin{array}{cc} \begin{array}{ccc} R & \mathscr{U}_1 & \mathscr{U}_2 \end{array} & \\ \begin{pmatrix} M_R & 0 & B \\ A & 0 & C \\ 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} R \\ \mathscr{U}_1 \\ \mathscr{U}_2 \end{array} \end{array}.$$

Thus, M has the same largest positive eigenvalue as M_R, with corresponding right eigenvector given (in block form) by

$$\mathbf{u}_m = \begin{bmatrix} \mathbf{u}_R \\ \left(\lambda_R^{-1}\right) A\mathbf{u}_R \\ 0 \end{bmatrix}.$$

Here, we assume $\mathbf{u}_m$ is normalised so that $\mathbf{a} \cdot \mathbf{u}_m = 1$. In addition, assuming we begin with a single ball $x \in R$, one readily verifies that the restriction of M to R and $\mathscr{U}_1$ satisfies conditions (A1)-(A6) of Section 2.1. Note also, that at each time-step the probability of adding a ball of type $x \in \mathscr{U}_2$ is 0 and thus, for each $n \in \mathbb{N}_0$, $\mathcal{U}_n(x) = 0$ almost surely. Therefore, combining this fact with Theorem 2.1, we have the following corollary.

Corollary 2.5. With $\mathbf{u}_m, \lambda_m$ and R as defined above, assuming we begin with a ball $x \in R$, we have

$$\frac{\mathcal{U}_n^m}{n} \xrightarrow{n\to\infty} \lambda_m \mathbf{u}_m \tag{2.13}$$

almost surely. In particular, almost surely

$$\frac{\mathbf{a} \cdot \mathcal{U}_n^m}{n} \xrightarrow{n\to\infty} \lambda_m. \tag{2.14}$$

In the coupling below, the assumption of a ball $x \in R$ is met by the tree process being initiated by a vertex 0 with weight W_0 sampled at random from μ and satisfying $h(W_0) > 0$.

2.2.2 Coupling the tree process with the urns associated with urn E

For $A \in \mathscr{B} \otimes \mathscr{B}$, recall the definition of $\Xi^{(2)}(A,n)$ from (1.2): this is the number of directed edges (v,v') of $\mathcal{T}_n$ where $(W_v, W_{v'}) \in A$.

Proposition 2.6. There exists a coupling $((\hat{\mathcal{U}}_n^m)_{m\in\mathbb{N}}, \hat{\mathcal{T}}_n)_{n\in\mathbb{N}_0}$ of the Pólya urn processes $\{(\mathcal{U}_n^m)_{n\in\mathbb{N}_0}, m \in \mathbb{N}\}$ and the tree process $(\mathcal{T}_n)_{n\in\mathbb{N}_0}$ such that, for each $m \in \mathbb{N}$, almost surely (on the coupling space), $\hat{\mathcal{U}}_0^m = \mathbf{e}_\ell$ for an initial ball $\ell \in R$ and, in addition, for $(i,j) \in [D_m]^2$, we have

$$\hat{\mathcal{U}}_n^m((i,j)) \le \Xi^{(2)}(n, \mathcal{I}_i^m \times \mathcal{I}_j^m), \tag{2.15}$$

$$\sum_{(i,j)\in[D_m]^2} \Big(\Xi^{(2)}(n, \mathcal{I}_i^m \times \mathcal{I}_j^m) - \hat{\mathcal{U}}_n^m((i,j))\Big) = \sum_{j=1}^{D_m} \hat{\mathcal{U}}_n^m((D_m+1, j)), \tag{2.16}$$

and

$$(\boldsymbol{\gamma}\mathbf{a}) \cdot \hat{\mathcal{U}}_n^m \le \mathcal{Z}_n \le \mathbf{a} \cdot \hat{\mathcal{U}}_n^m. \tag{2.17}$$

for all $n \in \mathbb{N}_0$.

Proof. First sample the entire tree process $(\hat{\mathcal{T}}_n)_{n\in\mathbb{N}_0}$; we use this to define the evolution of the urn processes. Moreover, for $i \in [D_m]$ let

$$\eta_n(i) := \sum_{v\in\mathcal{T}_n : r(v)=i} f(N^+(v, \mathcal{T}_n));$$

i.e., the sum of fitnesses of vertices with weight belonging to $\mathcal{I}_i^m$. Also, for $i \in [D_m]$ define

$$\theta_n(i) := (\boldsymbol{\gamma}\,\mathbf{a}\hat{\mathcal{U}}_n^m)(i) + \sum_{j=1}^{D_m} (\boldsymbol{\gamma}\,\mathbf{a}\hat{\mathcal{U}}_n^m)((i,j)).$$

Finally, recall that $\mathcal{Z}_n$ denotes the partition function associated with the tree at time n. Assume that at time 0 the tree consists of a single vertex 0 such that $r(W_0) = \ell \in [D_m]$. Then, set $\hat{\mathcal{U}}_0^m = \mathbf{e}_\ell$. Using the definition of r, since $W_0 \in \mathcal{I}_\ell^m$

$$0 < \mathcal{Z}_0 = h(W_0) \le h_{\max}(\ell) = \mathbf{a} \cdot \hat{\mathcal{U}}_0^m,$$

and by the choice of $\boldsymbol{\gamma}$, we have

$$\eta_0(\ell) = h(W_0) \ge h_{\min}(\ell) = (\boldsymbol{\gamma}\,\mathbf{a}\hat{\mathcal{U}}_0^m)(\ell) = \theta_0(\ell).$$

In this case, (2.15) and (2.16) are trivially satisfied since both sides of both equations are 0. Now, assume inductively that after n steps in the urn process, (2.15) and (2.16) are satisfied, we have

$$\eta_n(k) \ge \theta_n(k) \quad \text{for each} \quad k \in [D_m]\,, \tag{2.18}$$

and moreover, $\mathcal{Z}_n \le \mathbf{a} \cdot \hat{\mathcal{U}}_n^m$. Note that (2.18) implies the left hand side of (2.17), since

$$(\boldsymbol{\gamma}\mathbf{a}) \cdot \hat{\mathcal{U}}_n^m = \sum_{k=1}^{D_m} \theta_n(k) \le \sum_{k=1}^{D_m} \eta_n(k) = \mathcal{Z}_n.$$

Let s be the vertex sampled from $\mathcal{T}_n$ in the $(n+1)$st step, and assume that $r(W_s) = \ell'$, $r(W_{n+1}) = k$. Then, for the $(n+1)$th step in the urn: sample an independent random variable U_{n+1} uniformly distributed on $[0,1]$. Then:

- If $U_{n+1} \le \frac{\theta_n(\ell')\mathcal{Z}_n}{\eta_n(\ell')\mathbf{a}\cdot\hat{\mathcal{U}}_n^m}$, add balls of type (ℓ', k) and k to the urn (i.e. set $\hat{\mathcal{U}}_{n+1}^m = \hat{\mathcal{U}}_n^m + \mathbf{e}_{(\ell',k)} + \mathbf{e}_k$).
- Otherwise, add two balls of type (D_m+1, k) and k, respectively.

Note that, in the first case, we have

$$\Xi^{(2)}(n+1, \mathcal{I}_{\ell'}^m \times \mathcal{I}_k^m) = \Xi^{(2)}(n, \mathcal{I}_{\ell'}^m \times \mathcal{I}_k^m) + 1 \ge \hat{\mathcal{U}}_n^m((\ell',k)) + 1 = \hat{\mathcal{U}}_{n+1}^m((\ell',k))$$

and for $i \neq \ell'$ or $j \neq k$

$$\Xi^{(2)}(n+1, \mathcal{I}_i^m \times \mathcal{I}_j^m) = \Xi^{(2)}(n, \mathcal{I}_i^m \times \mathcal{I}_j^m) \ge \hat{\mathcal{U}}_n^m((i,j)) = \hat{\mathcal{U}}_{n+1}^m((i,j)).$$

Also, in this case

$$\eta_{n+1}(\ell') = \eta_n(\ell') + g(W_s, W_{n+1}) \ge \theta_n(\ell') + g_{\min}(\ell', k) = \theta_{n+1}(\ell'),$$

and similarly,

$$\eta_{n+1}(k) = \eta_n(k) + h(W_{n+1}) \ge \theta_n(k) + h_{\min}(k) = \theta_{n+1}(k),$$

so that (2.18) is satisfied. Finally, in this case,

$$\mathcal{Z}_{n+1} = \mathcal{Z}_n + g(W_s, W_{n+1}) + h(W_{n+1}) \le \mathbf{a}\cdot\hat{\mathcal{U}}_n^m + g_{\max}(\ell', k) + h_{\max}(k) = \mathbf{a}\cdot\hat{\mathcal{U}}_{n+1}^m.$$

Meanwhile, in the second case $\Xi^{(2)}(n, \mathcal{I}_{\ell'}^m \times \mathcal{I}_k^m)$ and $\eta_n(\ell')$ increase, while $\sum_{j=1}^{D_m} \hat{\mathcal{U}}_n^m((\ell', j))$ and $\theta_n(\ell')$ remain the same, and thus (2.15) is satisfied and $\eta_{n+1}(\ell') \ge \theta_{n+1}(\ell')$. As this is the only case when $\Xi^{(2)}(n, \mathcal{I}_{\ell'}^m \times \mathcal{I}_k^m) - \hat{\mathcal{U}}_n^m((\ell', k))$ increases, and we add a ball of type (D_m+1, k), (2.16) also follows. Both $\eta_n(k)$ and $\theta_n(k)$ increase as in the first case. Next,

$$\mathcal{Z}_{n+1} = \mathcal{Z}_n + g(W_s, W_{n+1}) + h(W_{n+1}) \le \mathbf{a}\cdot\hat{\mathcal{U}}_n^m + g^*(k) + h_{\max}(k) = \mathbf{a}\cdot\hat{\mathcal{U}}_{n+1}^m.$$

As all other quantities remain the same, (2.18) is satisfied, and moreover, $\mathcal{Z}_{n+1} \le \mathbf{a}\cdot\hat{\mathcal{U}}_{n+1}^m$. To complete the proof, it remains to prove the following claim.

Claim 2.6.1. For each $m \in \mathbb{N}$, almost surely (on the coupling space), the urn process $\hat{\mathcal{U}}^m = (\hat{\mathcal{U}}_n^m)_{n\in\mathbb{N}_0}$ is distributed like the Pólya urn process $(\mathcal{U}_n^m)_{n\in\mathbb{N}_0}$ with $\mathcal{U}_0^m$ consisting of an initial ball $\ell \in R$.

Proof. First note that, since W_0 is sampled from μ, conditionally on the positive probability event $\{h(W_0) > 0\}$, we have

$$\mathbb{P}\left(W_0 \in \mathcal{I}_\ell^m, h(W_0) > 0\right) \le \mathbb{P}\left(W_0 \in \mathcal{I}_\ell^m\right) = p_\ell^m,$$

and thus, $\mathbb{P}$-a.s., we have $W_0 \in \mathcal{I}_\ell^m$ with $p_\ell^m > 0$. This, combined with the fact that $0 < h(W_0) \le h_{\max}(\ell)$, implies that $\mathbb{P}$-a.s., the initial ball $\ell \in R$.

Now, note that in every step in $(\hat{\mathcal{U}}_n^m)_{n\in\mathbb{N}_0}$, we add a ball of type k for $k \in [D_m]$ with probability p_k^m, which is the same as in $(\mathcal{U}_n^m)_{n\in\mathbb{N}_0}$. Moreover, given $\hat{\mathcal{U}}_n^m$, the probability of adding balls of type (k, ℓ) is

$$p_\ell^m \left(\frac{\eta_n(k)}{\mathcal{Z}_n} \times \frac{\theta_n(k)\mathcal{Z}_n}{\eta_n(k)\mathbf{a}\cdot\hat{\mathcal{U}}_n^m}\right) = p_\ell^m \frac{\theta_n(k)}{\mathbf{a}\cdot\hat{\mathcal{U}}_n^m},$$

which also agrees with the Pólya urn scheme. Finally, the probability of adding a ball of type (D_m+1, ℓ) is

$$p_\ell^m \sum_{j=1}^{D_m}\left[\left(1 - \frac{\theta_n(j)\mathcal{Z}_n}{\eta_n(j)\mathbf{a}\cdot\hat{\mathcal{U}}_n^m}\right)\frac{\eta_n(j)}{\mathcal{Z}_n}\right] = p_\ell^m\left(1 - \sum_{j=1}^{D_m}\frac{\theta_n(j)}{\mathbf{a}\cdot\hat{\mathcal{U}}_n^m}\right),$$

as required. ☐

□

Note also that since the functions h_+, g_+ are non-increasing pointwise in m, on the coupling we have that for any fixed n, $\mathbf{a} \cdot \mathcal{U}_n^m$ is non-increasing in m. Combining this result with Corollary 2.5, we have the following corollary.

Corollary 2.7. The sequence $(\lambda_m)_{m\in\mathbb{N}}$ is non-increasing in m. In particular, there exists a limit $\lambda_\infty \geq 0$ such that

$$\lambda_m \downarrow \lambda_\infty$$

as $m \to \infty$.

2.2.3 The limiting compositions of the urn schemes associated with urn E

We now calculate the limiting vector $\mathbf{u}_m$ and the limiting eigenvalue λ_m. First note that by the definition of the urn process, for each $n \in \mathbb{N}_0$, $\ell \in [D_m]$ we have that $\mathcal{U}_{n+1}^m(\ell) - \mathcal{U}_n^m(\ell)$ is Bernoulli distributed with parameter p_ℓ^m. Thus, by the strong law of large numbers and Corollary 2.5, we have, for each $\ell \in [D_m]$,

$$\mathbf{u}_m(\ell) = \frac{p_\ell^m}{\lambda_m}. \tag{2.19}$$

Next, for any $i, j \in [D_m]$ using the definitions of $\boldsymbol{\gamma}$ and $\mathbf{a}$ ((2.11) and (2.12)) we have

$$\begin{aligned} \lambda_m \mathbf{u}_m((i,j)) &= p_j^m \sum_{\ell=1}^{D_m} (\boldsymbol{\gamma}\, \mathbf{a}\mathbf{u}_m)((i,\ell)) + p_j^m (\boldsymbol{\gamma}\, \mathbf{a}\mathbf{u}_m)(i) \\ &= p_j^m \sum_{\ell=1}^{D_m} g_{\min}(i,\ell)\mathbf{u}_m((i,\ell)) + p_j^m h_{\min}(i)\mathbf{u}_m(i) \\ &\overset{(2.19)}{=} p_j^m \sum_{\ell=1}^{D_m} g_{\min}(i,\ell)\mathbf{u}_m((i,\ell)) + \frac{p_j^m p_i^m h_{\min}(i)}{\lambda_m}. \end{aligned} \tag{2.20}$$

We now define

$$\mathcal{A}_i := \sum_{\ell=1}^{D_m} g_{\min}(i,\ell)\mathbf{u}_m((i,\ell)).$$

Multiplying both sides of (2.20) by $g_{\min}(i,j)$ and taking the sum over $j \in [D_m]$, recalling the definition of $\tilde{g}_-(i)$ in (2.10), we get

$$\begin{aligned} \lambda_m \mathcal{A}_i &= \left(\mathcal{A}_i + \frac{p_i^m h_{\min}(i)}{\lambda_m}\right) \sum_{j=1}^{D_m} p_j^m g_{\min}(i,j) \\ &= \left(\mathcal{A}_i + \frac{p_i^m h_{\min}(i)}{\lambda_m}\right) \tilde{g}_-(i). \end{aligned}$$

Thus, solving for $\mathcal{A}_i$

$$\mathcal{A}_i = \frac{p_i^m h_{\min}(i)\tilde{g}_-(i)}{\lambda_m(\lambda_m - \tilde{g}_-(i))}. \tag{2.21}$$

Substituting (2.21) into (2.20), we have

$$\begin{aligned} \lambda_m \mathbf{u}_m((i,j)) &= p_j^m \left(\frac{p_i^m h_{\min}(i)\tilde{g}_-(i)}{\lambda_m(\lambda_m - \tilde{g}_-(i))} + \frac{p_i^m h_{\min}(i)}{\lambda_m}\right) \\ &= p_j^m \frac{p_i^m h_{\min}(i)}{\lambda_m - \tilde{g}_-(i)}. \end{aligned} \tag{2.22}$$

Meanwhile, for each $j \in [D_m]$ we have

$$\begin{aligned}\lambda_m \mathbf{u}_m((D_m{+}1,j)) &= p_j^m \left(\sum_{\ell=1}^{D_m}(\mathbf{a}\mathbf{u}_m)((D_m+1,\ell)) + \sum_{i=1}^{D_m}\sum_{\ell=1}^{D_m}(\mathbf{a}-\boldsymbol{\gamma}\,\mathbf{a})((i,\ell)) + \sum_{i=1}^{D_m}(\mathbf{a}-\boldsymbol{\gamma}\,\mathbf{a})(i)\right) \\ &= p_j^m \left(\sum_{\ell=1}^{D_m} g^*(\ell)\mathbf{u}_m((D_m+1,\ell)) + \sum_{i=1}^{D_m}\sum_{\ell=1}^{D_m}(g_{\max}(i,\ell) - g_{\min}(i,\ell))\mathbf{u}_m((i,\ell)) \right. \\ &\qquad \left. + \sum_{i=1}^{D_m}(h_{\max}(i) - h_{\min}(i))\mathbf{u}_m(i)\right) \\ &=: p_j^m \left(\mathcal{B}_m + \mathcal{E}_m\right); \end{aligned} \tag{2.23}$$

where, in the last equation we set

$$\mathcal{B}_m := \sum_{\ell=1}^{D_m} g^*(\ell)\mathbf{u}_m((D_m+1,\ell))$$

and

$$\mathcal{E}_m := \sum_{i=1}^{D_m}\sum_{\ell=1}^{D_m}(g_{\max}(i,\ell) - g_{\min}(i,\ell))\mathbf{u}_m((i,\ell)) + \sum_{i=1}^{D_m}(h_{\max}(i) - h_{\min}(i))\mathbf{u}_m(i).$$

Multiplying both sides of (2.23) by $g^*(j)$ and taking the sum over j, we have

$$\lambda_m \mathcal{B}_m = \left(\sum_{j=1}^{D_m} p_j^m g^*(j)\right)(\mathcal{B}_m + \mathcal{E}_m) = \tilde{g}_+^*(\mathcal{B}_m + \mathcal{E}_m)$$

and thus

$$\mathcal{B}_m = \frac{\tilde{g}_+^*}{\lambda_m - \tilde{g}_+^*}\mathcal{E}_m. \tag{2.24}$$

We now use Condition **C1** in the following lemma.

Lemma 2.8. Assume Conditions **C1** and **C2**. Then $\lambda_\infty := \lim_{m\to\infty}\lambda_m > \tilde{g}^*$.

Proof. Note that, since we add two balls to the urn at each time-step, we have

$$\|\mathcal{U}_{n+1}^m\|_1 - \|\mathcal{U}_n^m\|_1 = 2.$$

Thus, by (2.13), we have $\|\lambda_m \mathbf{u}_m\|_1 = 2$. Now, by (2.19), we have $\lambda_m \sum_{\ell=1}^{D_m}\mathbf{u}_m(\ell) = 1$, and thus, by (2.22), we have

$$\sum_{j=1}^{D_m}\sum_{i=1}^{D_m}\lambda_m \mathbf{u}_m((i,j)) = \mathbb{E}\left[\frac{h_{\min}(r\,(W))}{\lambda_m - \tilde{g}_-\,(r\,(W))}\right] \leq 1. \tag{2.25}$$

Note that for any $x > 0$ as $m \to \infty$, $h_{\min}(r\,(x)) \uparrow h(x)$ and $\tilde{g}_-(r\,(x)) \uparrow \tilde{g}(x)$. Thus, by the monotone convergence theorem,

$$\mathbb{E}\left[\frac{h(W)}{\lambda_\infty - \tilde{g}(W)}\right] = \lim_{m\to\infty}\mathbb{E}\left[\frac{h_{\min}(r\,(W))}{\lambda_m - \tilde{g}_-\,(r\,(W))}\right] \leq 1.$$

Now, since the eigenvectors $\mathbf{u}_m$ are non-negative, by (2.24), $\lambda_m \geq \tilde{g}_+^*$, and thus, $\lambda_\infty = \lim_{m\to\infty}\lambda_m \geq \lim_{m\to\infty}\tilde{g}_+^* = \tilde{g}^*$. But, if $\lambda_\infty = \tilde{g}^*$, since the expression in (1.5) is decreasing in λ^*, this implies a contradiction to Condition **C1**. The result follows. □

Lemma 2.9. Assume Conditions **C1** and **C2**. Then, we have $\mathcal{B}_m \downarrow 0$ and $\mathcal{E}_m \downarrow 0$ as $m \to \infty$. In particular,

$$\mathbb{E}\left[\frac{h(W)}{\lambda_\infty - \tilde{g}(W)}\right] = 1, \tag{2.26}$$

so that $\lambda_\infty = \lambda^*$.

Proof. First, note that by Corollary 2.7 and Lemma 2.8, for each $m \in \mathbb{N}$, $\lambda_m \geq \lambda_\infty > \tilde{g}^*$. Combining this fact with the boundedness of g and h,

$$\sup_{x\in[0,1]}\left\{\frac{h(x)}{\lambda_m\left(\lambda_m - \tilde{g}(x)\right)}, \frac{1}{\lambda_m}\right\} < \sup_{x\in[0,1]}\left\{\frac{1}{\tilde{g}^*\left(\lambda_\infty - \tilde{g}(x)\right)}, \frac{1}{\lambda_\infty}\right\} =: C < \infty,$$

where the bound on the right is independent of m. Now, given $\varepsilon > 0$, by applying Lemma 2.4, let m be sufficiently large that

$$\sup_{(x,y)\in[0,1]\times[0,1]} (g_+(x,y) - g_-(x,y)) < \frac{\varepsilon}{2C} \quad \text{and} \quad \sup_{x\in[0,1]} (h_+(x) - h_-(x)) < \frac{\varepsilon}{2C}.$$

Then

$$\begin{aligned}
\mathcal{E}_m &= \sum_{i=1}^{D_m}\sum_{j=1}^{D_m} \left(g_{\max}(i,j) - g_{\min}(i,j)\right)\mathbf{u}_m((i,j)) + \sum_{\ell=1}^{D_m}\left(h_{\max}(\ell) - h_{\min}(\ell)\right)\mathbf{u}_m(\ell) \\
&\overset{(2.19),(2.22)}{=} \sum_{i=1}^{D_m}\sum_{j=1}^{D_m}\left(g_{\max}(i,j) - g_{\min}(i,j)\right)\frac{h_{\min}(i)p_i^m p_j^m}{\lambda_m(\lambda_m - \tilde{g}_-(i))} + \sum_{\ell=1}^{D_m}\left(h_{\max}(\ell) - h_{\min}(\ell)\right)\frac{p_\ell^m}{\lambda_m} \\
&< \frac{\varepsilon}{2C}\cdot C\left(\sum_{i=1}^{D_m}\sum_{j=1}^{D_m} p_i^m p_j^m\right) + \frac{\varepsilon}{2C}\cdot C\left(\sum_{\ell=1}^{D_m} p_\ell^m\right) = \varepsilon.
\end{aligned}$$

The result for $\mathcal{B}_m$ then follows from the fact that $\tilde{g}_+^* \downarrow \tilde{g}^*$, and Lemma 2.8. □

We are now ready to prove the main results of this subsection.

2.2.4 Proof of assertions 1 and 2 of Theorem 1.1

Proof. For Assertion 1, note that, by (2.17) from Proposition 2.6, we have

$$0 \leq \mathbf{a}\cdot\mathcal{U}_n^m - \mathcal{Z}_n \leq (\mathbf{a} - \boldsymbol{\gamma}\mathbf{a})\cdot\mathcal{U}_n^m.$$

Dividing by n and taking limits as $n \to \infty$, by (2.14) we have

$$0 \leq \lambda_m - \limsup_{n\to\infty}\frac{\mathcal{Z}_n}{n} \leq \lambda_m - \liminf_{n\to\infty}\frac{\mathcal{Z}_n}{n} \leq \limsup_{n\to\infty}\left((\mathbf{a} - \boldsymbol{\gamma}\mathbf{a})\cdot\frac{\mathcal{U}_n^m}{n}\right) = \mathcal{B}_m + \mathcal{E}_m.$$

The result follows by applying Lemma 2.9.

In addition, recalling the definition of $\mathscr{I}^m$ from (2.6), note that

$$\sigma(\mathscr{I}^m) = \left\{S \subseteq [0,1] : S = \bigcup_{i\in I}\mathcal{I}_i^m, I \subseteq [D_m]\right\}. \tag{2.27}$$

In other words, the σ-algebra generated by $\mathscr{I}^m$ is the set of finite unions of sets in $\mathscr{I}^m$. Recalling that $\mathscr{I}^{m_2}$ is a refined partition of $\mathscr{I}^{m_1}$ for $m_1 < m_2$, by Lemma 2.3 we have

$$\sigma(\mathscr{I}^{m_1}) \subseteq \sigma(\mathscr{I}^{m_2}). \tag{2.28}$$

We now prove Assertion 2. We begin by proving the result for Cartesian products of the form $S \times S'$ with $S, S' \in \sigma(\mathscr{I}^{m'})$, for $m' \in \mathbb{N}$. Note that, by the definition of $\Xi^{(2)}(n, \cdot)$, we clearly have finite *additivity*, that is, for any $S_1, S_2, S_3 \in \mathscr{B}$ if $S_1 \cap S_2 = \varnothing$, we have

$$\Xi^{(2)}(n, (S_1 \cup S_2) \times S_3) = \Xi^{(2)}(n, S_1 \times S_3) + \Xi^{(2)}(n, S_2 \times S_3), \quad \text{and similarly,}$$

$$\Xi^{(2)}(n, S_3 \times (S_1 \cup S_2)) = \Xi^{(2)}(n, S_3 \times S_1) + \Xi^{(2)}(n, S_3 \times S_2).$$

Combining these facts with Proposition 2.6, Corollary 2.5 and (2.22), for sets $S \times S'$ with $S, S' \in \sigma(\mathscr{I}^{m'})$ we have, for each $m > m'$,

$$\begin{aligned}
\mathbb{E}\left[\frac{h_-(W)}{\lambda_m - \tilde{g}_-(r(W))}\mathbf{1}_S\right]\mu(S') &\leq \liminf_{n\to\infty} \frac{\Xi^{(2)}(n, S \times S')}{n} \\
&\leq \limsup_{n\to\infty} \frac{\Xi^{(2)}(n, S \times S')}{n} \\
&\leq \mathbb{E}\left[\frac{h_-(W)}{\lambda_m - \tilde{g}_-(r(W))}\mathbf{1}_S\right]\mu(S') + \mathcal{B}_m + \mathcal{E}_m.
\end{aligned}$$

Taking limits as $m \to \infty$ and applying Lemma 2.9, this proves the result for this family of sets.

Now, by the Portmanteau Theorem, we need only prove that for all sets $U \in \mathcal{O}$, where $\mathcal{O}$ denotes the class of open subsets of $[0,1] \times [0,1]$, we have

$$\liminf_{n\to\infty} \frac{\Xi^{(2)}(n, U)}{n} \geq (\psi_*\mu \times \mu)(U). \tag{2.29}$$

Now, let

$$\mathcal{I}^m(U) := \bigcup_{i,j \in [D_m]: \mathcal{I}_i^m \times \mathcal{I}_j^m \subseteq U} \mathcal{I}_i^m \times \mathcal{I}_j^m. \tag{2.30}$$

Note that, since U is open, and $\mathscr{I}^m$ is fine enough that the set of dyadic intervals $\{\mathcal{D}_i^m\}_{i \in [2^m]} \subseteq \sigma(\mathscr{I}^m)$, we have

$$\mathbf{1}_{\mathcal{I}^m(U)} \uparrow \mathbf{1}_U \quad \text{pointwise as } m \to \infty. \tag{2.31}$$

In addition, since $\mathcal{I}^m(U) \subseteq U$, for each $m \in \mathbb{N}$

$$(\psi_*\mu \times \mu)(\mathcal{I}^m(U)) = \liminf_{n\to\infty} \frac{\Xi^{(2)}(n, \mathcal{I}^m(U))}{n} \leq \liminf_{n\to\infty} \frac{\Xi^{(2)}(n, U)}{n}.$$

Then, (2.29) follows by taking limits as $m \to \infty$. □

2.3 Analysing the tree by using urn D

2.3.1 Definitions of the urn schemes associated with urn D

In order to analyse the degree distribution in this model under Conditions **C1** and **C2**, we introduce another collection of Pólya urns $(\mathcal{V}_n^K)_{n \in \mathbb{N}_0}$, which not only depend on m, but also depend on a parameter $K \in \mathbb{N}$. Recall that in Urn E, the parameter m was used to approximate the infinite number of types of weight a vertex may have, as the support of the measure μ may be uncountable. With Urn D, we also need to account for the fact that the neighbourhood of a vertex can have unbounded size. Here, for some specific value of the parameter K, we keep track of the number of nodes in the tree process which have out-degree at most K using $K+1$ types, whereas all nodes with larger out-degree (with the first K nodes in their neighbourhood having certain types of weight) are represented by a single type. For brevity of notation, wherever possible in this subsection we omit the dependence of these parameters on m, and often also on K. For $i \in \mathbb{N}$, define $[D_m]^i$ so that

$$[D_m]^i := \{(u_0, \dots u_{i-1}) : u_0, \dots, u_{i-1} \in [D_m]\}.$$

Now, we set

$$\mathcal{B}' := \left(\bigcup_{i=1}^{K+1} [D_m]^i \right) \cup (\{D_m + 1\} \times [D_m]).$$

The urn process $(\mathcal{V}_n^K)_{n \geq 0}$ is then a vector-valued stochastic process taking values in $\mathbb{V}_{\mathcal{B}'}$. We now define the vectors $\mathbf{a}'$, $\boldsymbol{\gamma}'$ associated with the urn process such that

$$\mathbf{a}'(x) = \begin{cases} h_{\max}(u_0) + \sum_{j=1}^{k} g_{\max}(u_0, u_j) & \text{if } x = (u_0, \dots, u_k) \in [D_m]^{k+1} \\ g^*(\ell) & \text{if } x = (D_m + 1, \ell); \end{cases} \tag{2.32}$$

and,

$$\boldsymbol{\gamma}'(x) = \begin{cases} \frac{h_{\min}(u_0) + \sum_{j=1}^{k} g_{\min}(u_0, u_j)}{h_{\max}(u_0) + \sum_{j=1}^{k} g_{\max}(u_0, u_j)}, & \text{if } x = (u_0, \dots, u_k) \in [D_m]^{k+1}, k < K, \mathbf{a}'(x) > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2.33}$$

Now, given $\mathbf{u} = (u_0, \dots, u_k) \in [D_m]^{k+1}$, $k < K$, and $\ell \in [D_m]$, we define their *concatenation* $(\mathbf{u}, \ell) \in [D_m]^{k+2}$ such that

$$(\mathbf{u}, \ell) := (u_0, \dots, u_k, \ell).$$

Then, we define the replacement matrix M' of the urn $(\mathcal{V}_n^K)_{n \in \mathbb{N}_0}$ such that, given $x, x' \in \mathcal{B}'$,

$$M'_{x',x} = \begin{cases} -(\boldsymbol{\gamma}'\mathbf{a}')(x) & \text{if } x' = x, x \in [D_m]^k, k \leq K; \\ (\boldsymbol{\gamma}'\mathbf{a}')(x) p_\ell^m, & \text{if } x' = (x, \ell), \ell \in [D_m], x \in [D_m]^k, k \leq K; \\ (\mathbf{a}' - \boldsymbol{\gamma}'\mathbf{a}')(x) p_\ell^m, & \text{if } x' = (D_m + 1, \ell), \ell \in [D_m], x \in \mathcal{B}'; \\ \mathbf{a}'(x) p_\ell^m, & \text{if } x' = \ell, x \in \mathcal{B}'; \\ 0 & \text{otherwise.} \end{cases}$$

Again, note that it may be the case that M' is not irreducible, if either $\mathbf{a}'(x) = 0$ for certain $x \in \mathcal{B}'$ or $p_\ell^m = 0$ for certain choices of ℓ. Nevertheless, we define the sets

$$\mathscr{U}_1' := \{x \in \mathcal{B}' : M'_{x'x} = 0 \; \forall x' \in \mathcal{B}'\} = \{x \in \mathcal{B}' : \mathbf{a}'(x) = 0\},$$

and

$$\mathscr{U}_2' := \{x' \in \mathcal{B}' : M'_{x'x} = 0 \; \forall x \in \mathcal{B}' \setminus \{x'\}\}.$$

Again, we assume that $\mathscr{U}_1' \cap \mathscr{U}_2' = \varnothing$; if not, we replace $\mathscr{U}_1'$ by $\mathscr{U}_1' \setminus \mathscr{U}_2'$. We then set $R' = \mathcal{B}' \setminus (\mathscr{U}_1' \cup \mathscr{U}_2')$, and let $M'_{R'}$ be the restriction of M' to R'. As before, $M'_{R'}$ satisfies the conditions of Lemma 2.2, and thus has a unique largest positive eigenvalue $\lambda'_{R'}$ with corresponding eigenvector $\mathbf{V}_{R'}$. But then, writing M' in block form in a manner analogous to the previous subsection, M has the same largest positive eigenvalue, with corresponding right eigenvector given, in block form, by

$$\mathbf{V}_K = \begin{bmatrix} \mathbf{V}_{R'} \\ (\lambda'_{R'})^{-1} A' \mathbf{V}_{R'} \\ 0 \end{bmatrix}.$$

Here, we assume $\mathbf{V}_K$ is normalised so that $\mathbf{a}' \cdot \mathbf{V}_K = 1$. Also in a manner similar to the previous subsection, assuming we begin with a ball of type $x \in R'$, one readily verifies that the restriction of M' to R' and $\mathscr{U}_1'$ satisfies conditions (A1)-(A6) of Section 2.1, and also, that for each $x \in \mathscr{U}_2'$ and $n \in \mathbb{N}_0$, $\mathcal{U}_n(x) = 0$ almost surely. Therefore, applying Theorem 2.1 again, we have the following corollary:

Corollary 2.10. With $\mathbf{V}_K, \lambda_K'$ and R' as defined above, assuming we begin with a ball $x \in R'$, we have

$$\frac{\mathcal{V}_n^K}{n} \xrightarrow{n \to \infty} \lambda_K' \mathbf{V}_K \tag{2.34}$$

almost surely. In particular, we have

$$\frac{\mathbf{a} \cdot \mathcal{V}_n^K}{n} \xrightarrow{n \to \infty} \lambda_K'. \tag{2.35}$$

As in the previous subsection, in the coupling below, the assumption of a ball $x \in R'$ is met by the tree process being initiated by a vertex 0 with weight W_0 sampled at random from μ and satisfying $h(W_0) > 0$.

2.3.2 Coupling the tree process with the urns associated with urn D

Recall that we denote by $N_{\geq k}(n, B)$ the number of vertices of out-degree at least k having weight belonging to $B \in \mathscr{B}$. We also define the analogue $\mathscr{D}_{\geq k}(n, j)$ for $n \in \mathbb{N}_0$ and $j \in [D_m]$ such that

$$\mathscr{D}_{\geq k}(n, j) := \sum_{i=k}^{K+1} \sum_{\mathbf{u}_i \in [D_m]^i} \mathcal{V}_n^K(\mathbf{u}_i) \mathbf{1}_{\{j\}}(u_0). \tag{2.36}$$

This represents the number of balls in the urn $\mathcal{V}_n^K$ with type $\mathbf{u} = (u_0, \dots)$ having dimension at least $k+1$, with $u_0 = j$. We then have the following analogue of Proposition 2.6:

Proposition 2.11. There exists a coupling $(\hat{\mathcal{V}}_n^K, \hat{\mathcal{T}}_n)_{n \in \mathbb{N}_0}$ of the Pólya urn process $(\mathcal{V}_n^K)_{n \in \mathbb{N}_0}$ and the tree process $(\mathcal{T}_n)_{n \in \mathbb{N}_0}$ such that, almost surely (on the coupling space), $\mathcal{V}_0^K$ consists of a single ball $\ell \in R'$ and for all $n \in \mathbb{N}_0$, $k \in \{0\} \cup [K]$, we have

$$\mathscr{D}_{\geq k}(n, j) \leq N_{\geq k}\left(n, \mathcal{I}_j^m\right) \quad \text{and} \tag{2.37}$$

$$\sum_{j=1}^{D_m} \left(N_{\geq k}\left(n, \mathcal{I}_j^m\right) - \mathscr{D}_{\geq k}(n, j)\right) \leq \sum_{j=1}^{D_m} \hat{\mathcal{V}}_n^K((D_m + 1, j)). \tag{2.38}$$

In addition, we have

$$(\boldsymbol{\gamma}' \mathbf{a}') \cdot \hat{\mathcal{V}}_n^K \leq \mathcal{Z}_n \leq \mathbf{a}' \cdot \hat{\mathcal{V}}_n^K. \tag{2.39}$$

Proof. We proceed in a somewhat similar manner to Proposition 2.6, however, in this case, we first introduce a "labelled" Pólya urn $\left(\mathcal{L}_n^K\right)_{n \geq 0}$ where balls carry *integer labels* from $\{-D_m, \dots, 0, \dots, n\}$. In addition, for $j \in \{0\} \cup [n]$, the label is independent of the *type* of the ball: we denote by $b_n(j)$ the *type* of a ball with label j at time n. One may interpret the ball with label j as representing the evolution of vertex j in the tree process – in this sense, the label may be interpreted as a "time-stamp". Balls of type $(D_m + 1, j)$, $j \in [D_m]$, however, are labelled $-j$ – we denote by $d_n(j)$ the number of balls with this label, since here, multiple balls may share the same label. We describe the labelled urn process $\mathcal{L}_n^K$

as an evolving vector in $\mathcal{B}' \times \mathbb{Z}$, so that $\mathcal{L}_n^K = \sum_{j=1}^{D_m} d_n(j) \cdot \mathbf{e}_{(b_n(j),j)} + \sum_{i=0}^{n} \mathbf{e}_{(b_n(i),i)}$. We set

$$\mathbf{a}'(\mathcal{L}_n^K) = \sum_{j=-D_m}^{-1} d_n(j) \cdot \mathbf{a}'(b_n(j)) + \sum_{i=0}^{n} \mathbf{a}'(b_n(i)), \text{ and } (\boldsymbol{\gamma}'\mathbf{a}')(\mathcal{L}_n^K) = \sum_{i=0}^{n} (\boldsymbol{\gamma}'\mathbf{a}')(b_n(i)).$$

Now, we use $\mathcal{L}_{n+1}^K$ to define $\hat{\mathcal{V}}_{n+1}^K$ by "forgetting" labels, so that,

$$\text{if } \mathcal{L}_{n+1}^K = \sum_{j=-D_m}^{-1} d_n(j) \cdot \mathbf{e}_{(b_{n+1}(j),j)} + \sum_{i=0}^{n+1} \mathbf{e}_{(b_{n+1}(j),i)},$$
$$\text{we set } \hat{\mathcal{V}}_{n+1}^K = \sum_{j=-D_m}^{-1} d_n(j) \cdot \mathbf{e}_{b_{n+1}(j)} + \sum_{i=0}^{n+1} \mathbf{e}_{b_{n+1}(i)}.$$

Sample the entire tree process $(\hat{\mathcal{T}}_n)_{n \in \mathbb{N}_0}$. If, at time 0, the tree consists of a single vertex 0 with weight $W_0 \in I_\ell^m$ then, we set $\mathcal{L}_0^K = \mathbf{e}_{(\ell,0)}$, and note that we have

$$(\boldsymbol{\gamma}'\mathbf{a}')(\mathcal{L}_0^K) = h_{\min}(\ell) \leq h(W_0) = \mathcal{Z}_0 \leq \mathbf{a}'(\mathcal{L}_0^K) = h_{\max}(\ell),$$

and

$$f(N^+(0, \hat{\mathcal{T}}_0)) = h(W_0) \geq (\boldsymbol{\gamma}'\,\mathbf{a}')\,(b_0(0)) = h_{\min}(\ell).$$

Now, assume inductively that after n steps in the process, for each $i \in \{0\} \cup [n]$ we have

$$f(N^+(i, \hat{\mathcal{T}}_n)) \geq (\boldsymbol{\gamma}'\,\mathbf{a}')\,(b_n(i)), \quad \deg^+(i, \mathcal{T}_n) \geq \dim(b_n(i)) - 1, \tag{2.40}$$

$$\sum_{i=0}^{n} \left(\deg^+(i, \mathcal{T}_n) - \dim(b_n(i)) + 1\right) = \sum_{j=1}^{D_m} \hat{\mathcal{V}}_n^K((D_m+1, j), \tag{2.41}$$

and (2.39) is satisfied.

Let s be the vertex sampled in the tree in the $(n+1)$st step, assume that $r(s) = \ell'$ and that $r(n+1) = k$. Then, for the $(n+1)$th step in the urn: sample an independent random variable U_{n+1} uniformly distributed on $[0,1]$. Then:

- If $\dim(b_n(s)) \leq K$ and $U_{n+1} \leq \frac{(\boldsymbol{\gamma}'\mathbf{a}')(b_n(s))\mathcal{Z}_n}{f(N^+(s,\hat{\mathcal{T}}_n))\mathbf{a}'(\mathcal{L}_n^K)}$, remove the ball $(b_n(s), s)$ from the urn, and add balls $((b_n(s), k), s)$ and $(k, n+1)$ to the urn (i.e. set $\mathcal{L}_{n+1}^K = \mathcal{L}_n^K + \mathbf{e}_{((b_n(s),\ell),s)} + \mathbf{e}_{(k,n+1)} - \mathbf{e}_{(b_n(s),s)}$). We call this step Case 1.
- Otherwise, add balls of type $((D_m+1, k), -k)\,,(k, n+1)$ – we call this Case 2.

First note that

$$\begin{aligned}(\boldsymbol{\gamma}'\mathbf{a}')(b_{n+1}(s)) - (\boldsymbol{\gamma}'\mathbf{a}')(b_n(s)) &= \begin{cases} g_{\min}(\ell', k), & \text{in Case 1} \\ 0, & \text{in Case 2} \end{cases} \\ &\leq g(W_s, W_{n+1}) = f(N^+(s, \hat{\mathcal{T}}_{n+1})) - f(N^+(s, \hat{\mathcal{T}}_n)),\end{aligned}$$

and likewise

$$(\boldsymbol{\gamma}'\mathbf{a}')(b_{n+1}(n+1)) = h_{\min}(\ell) \leq h(W_{n+1}) = f(N^+(n+1, \hat{\mathcal{T}}_{n+1})).$$

Additionally, in Case 1 the dimension of $b_n(s)$ and the degree of s in $\hat{\mathcal{T}}_n$ both increase, whilst in Case 2 only the degree of s increases whilst the dimension of $b_n(s)$ remains the same. This proves (2.40) at time $n+1$. In addition, Case 2 coincides with the addition of a ball of type (D_m+1,ℓ), which yields (2.41). Finally,

$$
\begin{aligned}
(\boldsymbol{\gamma}'\mathbf{a}')\cdot\left(\hat{\mathcal{V}}^K_{n+1}-\hat{\mathcal{V}}^K_n\right) &= \begin{cases} h_{\min}(k)+g_{\min}(\ell',k), & \text{in Case 1} \\ h_{\min}(k), & \text{in Case 2} \end{cases} \\
&\le h(W_{n+1})+g(W_s,W_{n+1}) = \mathcal{Z}_{n+1}-\mathcal{Z}_n \\
&\le \begin{cases} h_{\max}(k)+g_{\max}(\ell',k), & \text{in Case 1} \\ h_{\max}(k)+g^*_{\max}(k), & \text{in Case 2} \end{cases} \\
&\le (\mathbf{a}')\cdot(\hat{\mathcal{V}}^K_{n+1}-\hat{\mathcal{V}}^K_n);
\end{aligned}
$$

which shows that (2.39) is also satisfied at time $n+1$.

Claim 2.11.1. Almost surely (on the coupling space), the urn process $\hat{\mathcal{V}}^K=(\hat{\mathcal{V}}^K_n)_{n\in\mathbb{N}_0}$ is distributed like the Pólya urn $(\mathcal{V}^K_n)_{n\in\mathbb{N}_0}$ with $\mathcal{V}^K_0$ consisting of an initial ball $\ell\in R'$.

Proof. The fact that, $\mathbb{P}$-a.s., the initial ball $\ell\in R'$ follows immediately from the fact that the initial weight W_0 is sampled from μ conditionally on the event $\{h(W_0)>0\}$ (analogous to in Claim 2.6.1). Moreover, in every step in $\hat{\mathcal{V}}^K$, we add a ball of type k for $k\in[D_m]$ with probability p^m_k, which is the same as in $\mathcal{V}^K$. Furthermore, given $\hat{\mathcal{V}}^K_n$ the probability of removing a ball of type $\mathbf{u}$ with $\dim\mathbf{u}\le K$ and adding a ball of type $(\mathbf{u},\ell)$ is

$$
\begin{aligned}
p^m_\ell\sum_{s\in\mathcal{L}^K_n:b_n(s)=\mathbf{u}}\frac{(\boldsymbol{\gamma}'\mathbf{a}')(b_n(s))\mathcal{Z}_n}{f(N^+(s,\hat{\mathcal{T}}_n))\mathbf{a}'(\mathcal{L}^K_n)}\times\frac{f(N^+(s,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n} &= p^m_\ell\sum_{s\in\mathcal{L}^K_n:b_n(s)=\mathbf{u}}\frac{(\boldsymbol{\gamma}'\mathbf{a}')(b_n(s))}{\mathbf{a}'(\mathcal{L}^K_n)} \\
&= p^m_\ell\frac{\hat{\mathcal{V}}^K_n(\mathbf{u})}{\mathcal{Z}_n},
\end{aligned}
$$

which also agrees with the transition law of the Pólya urn scheme $\mathcal{V}$. Finally, the probability of adding a ball of type (D_m+1,ℓ) is

$$
\begin{aligned}
& p^m_\ell\sum_{s\in\mathcal{L}^K_n:\dim b_n(s)>K}\frac{f(N^+(s,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n} \\
&\qquad + p^m_\ell\sum_{s\in\mathcal{L}^K_n:\dim b_n(s)\le K}\left(1-\frac{(\boldsymbol{\gamma}'\mathbf{a}')(b_n(s))\mathcal{Z}_n}{f(N^+(s,\hat{\mathcal{T}}_n))\mathbf{a}'(\mathcal{L}^K_n)}\right)\times\frac{f(N^+(s,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n} \\
&= p^m_\ell\sum_{s\in\mathcal{L}^K_n}\left(\frac{f(N^+(s,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n}\right)-p^m_\ell\sum_{s\in\mathcal{L}^K_n:\dim b_n(s)\le K}\frac{(\boldsymbol{\gamma}'\mathbf{a}')(b_n(s))}{\mathbf{a}'(\mathcal{L}^K_n)} \\
&= p^m_\ell\left(1-\sum_{\mathbf{u}\in\hat{\mathcal{V}}^K_n:\dim\mathbf{u}\le K}\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\hat{\mathcal{V}}^K(\mathbf{u}))}{\mathbf{a}'(\hat{\mathcal{V}}^K_n)}\right),
\end{aligned}
$$

which agrees with the transition rule of $\mathcal{V}^K$. □

Finally, to complete the proof, we verify the following claim.

Claim 2.11.2. For all $n\in\mathbb{N}_0$, (2.37) and (2.38) are satisfied for all $k\in\{0\}\cup[K]$.

Proof. If we define $b_n(i)|_0$ such that $b_n(i)|_0=x_0$ if $b_n(i)=(x_0,\dots,x_k)$, then, by construction of the labelled urn process $(\mathcal{L}^K_n)_{n\in\mathbb{N}_0}$, $b_n(i)|_0=x_0 \implies r(W_i)=x_0$, so that $W_i\in\mathcal{I}^m_{x_0}$. Therefore, for each $k\in\{0\}\cup[K], j\in[D_m]$,

$$
\mathscr{D}_{\ge k}(n,j)=\sum_{b_n(i):\dim(b_n(i))\ge k+1}\mathbf{1}_{\{j\}}(b_n(i)|_0)\overset{(2.40)}{\le}\sum_{i:\deg^+(i,\hat{\mathcal{T}}_n)\ge k}\mathbf{1}_{\mathcal{I}^m_j}(W_i)=N_{\ge k}(n,\mathcal{I}^m_j).
$$

Moreover, by (2.41),

$$\sum_{j=1}^{D_m} \hat{\mathcal{V}}_n^K((D_m+1,j) = \sum_{i=0}^{n} \Big(\deg^+(i,\hat{\mathcal{T}}_n) - \dim(b_n(i)) + 1\Big)$$
$$= \sum_{k=0}^{n}\sum_{j=1}^{D_m} \left(\left(N_{\geq k}\left(n, \mathcal{I}_j^m\right) - \mathscr{D}_{\geq k}(n,j)\right)\right),$$

which implies (2.38). □

□

2.3.3 The limiting compositions of the urn schemes associated with urn D

We now calculate the limiting vector $\mathbf{V}_K$ and limiting eigenvalue λ'_K of the Pólya urn scheme $(\mathcal{V}_n^K)_{n\geq 0}$. We first introduce some more notation: for any vector $\mathbf{u} = (u_0,\dots,u_{k-1}) \in [D_m]^k$, and $i \in \{0\}\cup[k-1]$, denote by $\mathbf{u}|_i := (u_0,\dots,u_i) \in [D_m]^{i+1}$. We also define the following quantities:

$$\mathcal{R}_K := \sum_{\ell=1}^{D_m} \mathbf{a}'((D_m+1,\ell))\mathbf{V}_K((D_m+1,\ell)), \tag{2.42}$$

$$\mathcal{E}_K := \sum_{\mathbf{u}:\dim \mathbf{u}\leq K} (\mathbf{a}' - \boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u})\mathbf{V}_K(\mathbf{u}), \quad \text{and} \tag{2.43}$$

$$\mathcal{F}_K := \sum_{\mathbf{v}:\dim \mathbf{v}=K+1} \mathbf{a}'(\mathbf{v})\mathbf{V}_K(\mathbf{v}). \tag{2.44}$$

Proposition 2.12. Let λ'_K and $\mathbf{V}_K$ denote the limiting leading eigenvalue and corresponding right eigenvector of M', respectively. Then, denoting the components of a vector $\mathbf{u}$ by $u_0, u_1, \dots$, the eigenvector $\mathbf{V}_K$ satisfies

$$\lambda'_K\mathbf{V}_K(x) = \begin{cases} \frac{p^m_{u_k}\lambda'_K}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u})+\lambda'_K}\prod_{i=0}^{k-1}\left[p^m_{u_i}\left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)+\lambda'_K}\right)\right], & x = \mathbf{u}\in[D_m]^{k+1}, 0\leq k<K; \\ p^m_{u_K}\prod_{i=0}^{K-1}\left[p^m_{u_i}\left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)+\lambda'_K}\right)\right], & x=\mathbf{u}\in[D_m]^{K+1}, \end{cases} \tag{2.45}$$

where we set the empty product of terms, when $k=0$ equal to 1. In addition, we have

$$\mathcal{R}_K = \frac{\mathcal{E}_K + \mathcal{F}_K}{\lambda'_K - g^*_+}. \tag{2.46}$$

Proof. First note that, for each $u_0 \in [D_m]$, since we add a ball of type u_0 with probability $p^m_{u_0}$ at each time-step, and remove such a ball with probability proportional to $(\boldsymbol{\gamma}'\mathbf{a}')(u_0)$, we have

$$\lambda'_K\mathbf{V}_K(u_0) = p^m_{u_0} - (\boldsymbol{\gamma}'\mathbf{a}')(u_0)\mathbf{V}_K(u_0), \tag{2.47}$$

this implies the case $k=0$ in (2.45). Next, for $k>0$, we have

$$\lambda'_K\mathbf{V}_K(\mathbf{u}) = \begin{cases} p^m_{u_k}(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_{k-1})\mathbf{V}_K(\mathbf{u}|_{k-1}) - (\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u})\mathbf{V}_K(\mathbf{u}), & \mathbf{u}\in[D_m]^{k+1},\ k<K; \\ p^m_{u_K}(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_{K-1})\mathbf{V}_K(\mathbf{u}|_{K-1}); & \mathbf{u}\in[D_m]^{K+1}; \end{cases} \tag{2.48}$$

so that, if $\mathbf{u}\in[D_m]^{k+1}$, $1\leq k\leq K-1$,

$$\mathbf{V}_K(\mathbf{u}) = \frac{p^m_{u_k}(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_{k-1})\mathbf{V}_K(\mathbf{u}|_{k-1})}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u})+\lambda'_K}. \tag{2.49}$$

Applying (2.48) and (2.49), recursing backwards, and using the fact that $\mathbf{V}_K(u_0) = p^m_{u_0}/((\boldsymbol{\gamma}'\mathbf{a}')(u_0)+\lambda'_K)$ from (2.47), completes the proof of (2.45). Finally, for each $j \in [D_m]$, we have

$$\begin{aligned}\lambda'_K \mathbf{V}_K((D_m+1,j)) =& p^m_j \Bigg(\sum_{\ell=1}^{D_m} \mathbf{a}'((D_m+1,\ell))\mathbf{V}_K((D_m+1,\ell)) \\ &+ \sum_{\mathbf{u}:\dim \mathbf{u} \le K} (\mathbf{a}' - \boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u})\mathbf{V}_K(\mathbf{u}) + \sum_{\mathbf{v}:\dim \mathbf{v} = K+1} \mathbf{a}'(\mathbf{v})\mathbf{V}_K(\mathbf{v}) \Bigg) \\ =& p^m_j \left(\mathcal{R}_K + \mathcal{E}_K + \mathcal{F}_K\right); \end{aligned} \tag{2.50}$$

where, in the last equation we recall the definitions in (2.42) and (2.44). Now, multiplying both sides of (2.50) by $\mathbf{a}'((D_m+1,j)) = g^*(j)$ and taking the sum over j, we have

$$\lambda'_K \mathcal{R}_K = \left(\sum_{j=1}^{D_m} p^m_j g^*(j) \right) \left(\mathcal{R}_K + \mathcal{E}_K + \mathcal{F}_K\right) = \tilde{g}^*_+ \left(\mathcal{R}_K + \mathcal{E}_K + \mathcal{F}_K\right).$$

Rearranging this proves (2.46), thus completing the proof of the proposition. □

Now, we recall the definition of the *companion process* $(S_i(w))_{i\ge 0}$ from Section 1.3 in (1.8): Recall that $W_1, W_2, \dots$ were defined to be independent μ-distributed random variables and let $w \in [0,1]$. We then defined the random process $(S_i(w))_{i\ge 0}$ inductively so that $S_0(w) = h(w)$ and for all $i \ge 0$, we have $S_{i+1}(w) = S_i(w) + g(w, W_{i+1})$. Now, we also define the *lower companion process* $(S^-_i(w))_{i\ge 0}$ in a similar way, but instead with the functions h_-, g_- respectively, so that

$$S^-_0(w) := h_-(w); \quad S^-_{i+1}(w) := S^-_i(w) + g_-(w, W_{i+1}),\ i \ge 0. \tag{2.51}$$

Lemma 2.13. Assume Conditions **C1** and **C2**. Then we have

$$\lim_{K\to\infty} \lim_{m\to\infty} \mathcal{F}_K = 0.$$

Proof. Note that by (2.45), with J being an upper bound on $\max\{1, g_{\max}\}$ (recalling $g_{\max}$ from **C2**), we have

$$\begin{aligned}\mathcal{F}_K &= \sum_{\mathbf{u}:\dim \mathbf{u} = K+1} \mathbf{a}'(\mathbf{u})\mathbf{V}_K(\mathbf{u}) \\ &= \sum_{\mathbf{u}:\dim \mathbf{u} = K+1} \mathbf{a}'(\mathbf{u})\, p^m_{u_K} \prod_{i=0}^{K-1} \left[p^m_{u_i} \left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i) + \lambda'_K} \right) \right] \\ &\le J(K+1) \cdot \sum_{\mathbf{u}:\dim \mathbf{u} = K+1} p^m_{u_K} \prod_{i=0}^{K-1} \left[p^m_{u_i} \left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i) + \lambda'_K} \right) \right] \\ &= J(K+1) \cdot \sum_{\mathbf{u}:\dim \mathbf{u} = K} \left(\sum_{u_K \in [D_m]} p^m_{u_K} \right) \prod_{i=0}^{K-1} \left[p^m_{u_i} \left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i) + \lambda'_K} \right) \right] \\ &= J(K+1) \cdot \sum_{\mathbf{u}:\dim \mathbf{u} = K} \prod_{i=0}^{K-1} \left[p^m_{u_i} \left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i) + \lambda'_K} \right) \right] \\ &= J(K+1) \cdot \mathbb{E}\left[\prod_{i=0}^{K-1} \left(\frac{S^-_i(W)}{S^-_i(W) + \lambda'_K} \right) \right], \end{aligned}$$

where we recall the definition of $(S_i^-(w))_{i\geq 0}$ from (2.51). Now, note that for all $m \in \mathbb{N}$, and for all $w \in [0,1]$, $S^-(w)$ is stochastically bounded above by $S(w)$, and by (2.35) and (2.39), λ'_K is bounded below by λ^* uniformly in m and K. Therefore, since the function $x \mapsto \frac{x}{x+\lambda}$ is increasing in x and decreasing in λ, we may bound the previous display above so that

$$\begin{aligned} J(K+1)\cdot\mathbb{E}\left[\prod_{i=0}^{K-1}\left(\frac{S_i^-(W)}{S_i^-(W)+\lambda'_K}\right)\right] &\leq J(K+1)\cdot\mathbb{E}\left[\prod_{i=0}^{K-1}\left(\frac{S_i(W)}{S_i(W)+\lambda'_K}\right)\right] \\ &\leq J(K+1)\cdot\mathbb{E}\left[\prod_{i=0}^{K-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right]. \end{aligned}$$

We complete the proof by proving the following claim.

Claim 2.13.1. We have

$$\lim_{k\to\infty} k\cdot\mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] = 0$$

Proof. First observe that

$$\mathbb{E}\left[\prod_{i=0}^{\infty}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] \leq \prod_{i=1}^{\infty}\left(\frac{Ji}{Ji+\lambda^*}\right) = \prod_{i=0}^{\infty}\left(1-\frac{\lambda^*}{Ji+\lambda^*}\right) \leq e^{-\sum_{i=1}^{\infty}\frac{\lambda^*}{Ji+\lambda^*}} = 0.$$

Therefore, we have

$$\begin{aligned} k\cdot\mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] &= k\cdot\sum_{j=k}^{\infty}\mathbb{E}\left[\left(1-\frac{S_j(W)}{S_j(W)+\lambda^*}\right)\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] \\ &= k\cdot\sum_{j=k}^{\infty}\mathbb{E}\left[\frac{\lambda^*}{S_j(W)+\lambda^*}\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] \\ &\leq \sum_{j=k}^{\infty} j\cdot\mathbb{E}\left[\frac{\lambda^*}{S_j(W)+\lambda^*}\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right]. \end{aligned}$$

The series on the right of the previous display consists of non-negative terms, and for each $N \in \mathbb{N}$, we have

$$\begin{aligned} &\sum_{j=1}^{N} j\cdot\mathbb{E}\left[\frac{\lambda^*}{S_j(W)+\lambda^*}\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] \qquad (2.52) \\ &= \sum_{j=1}^{N}\left(j\cdot\mathbb{E}\left[\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] - j\cdot\mathbb{E}\left[\prod_{i=0}^{j}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right]\right) \\ &= \sum_{j=1}^{N}\mathbb{E}\left[\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] - N\cdot\mathbb{E}\left[\prod_{i=0}^{N}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] \\ &\leq \sum_{j=1}^{N}\mathbb{E}\left[\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right]. \end{aligned}$$

Now, note that by Lemma 1.2, we have

$$\sum_{j=1}^{\infty}\mathbb{E}\left[\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] < \infty,$$

and thus by (2.52) and the monotone convergence theorem, we also have

$$\sum_{j=1}^{\infty} j \cdot \mathbb{E}\left[\frac{\lambda^*}{S_j(W)+\lambda^*}\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] < \infty.$$

Therefore,

$$\lim_{k\to\infty} k \cdot \mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] \le \lim_{k\to\infty}\sum_{j=k}^{\infty} j \cdot \mathbb{E}\left[\frac{\lambda^*}{S_j(W)+\lambda^*}\prod_{i=0}^{j-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\right] = 0.$$

□

Lemma 2.14. Assume Conditions **C1** and **C2**. Then we have

$$\lim_{K\to\infty}\lim_{m\to\infty}\mathcal{E}_K = 0, \quad \text{and} \quad \lim_{K\to\infty}\lim_{m\to\infty}\mathcal{R}_K = 0. \tag{2.53}$$

In addition,

$$\lim_{K\to\infty}\lim_{m\to\infty}\lambda'_K = \lambda^*. \tag{2.54}$$

Proof. The proof is similar to that of Lemma 2.9. First, let $\varepsilon > 0$ be given, and, by Lemma 2.4, let m be sufficiently large that for all $x, y \in [0,1]$

$$(g_+(x,y) - g_-(x,y)) < \frac{\varepsilon\lambda'_K}{K} \text{ and } (h_+(x) - h_-(x)) < \frac{\varepsilon\lambda'_K}{K}. \tag{2.55}$$

The inequalities in (2.55) now imply that for any $\mathbf{u} = (u_0, \dots, u_{K-1}) \in [D_m]^K$, and each $i \in \{0\} \cup [K-1]$ we have (taking the empty sum to be zero when $i = 0$)

$$\begin{aligned}(\mathbf{a}' - \boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i) &= h_{\max}(u_0) - h_{\min}(u_0) + \sum_{j=1}^{i-1}(g_{\max}(u_0,u_j) - g_{\min}(u_0,u_j)) \\ &< \frac{\varepsilon\lambda'_K}{K}\cdot K = \varepsilon\lambda'_K. \end{aligned} \tag{2.56}$$

Now, using the $\mathbf{u}|_i$ notation as a shorthand, we can write

$$\begin{aligned}\mathcal{E}_K &= \sum_{\mathbf{u}\in[D_m]^K}\sum_{i=0}^{K-1}\left((\mathbf{a}' - \boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)\right)\mathbf{V}_K(\mathbf{u}|_i) \\ &\overset{(2.45)}{=} \sum_{\mathbf{u}\in[D_m]^K}\sum_{i=0}^{K-1}\frac{\left((\mathbf{a}' - \boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)\right)p^m_{u_i}}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i) + \lambda'_K}\prod_{j=0}^{i-1}\left[p^m_{u_j}\left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_j)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_j)+\lambda'_K}\right)\right] \\ &\overset{(2.56)}{\le} \varepsilon \cdot \sum_{\mathbf{u}\in[D_m]^K}\sum_{i=0}^{K-1}\frac{\lambda'_K p^m_{u_i}}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i) + \lambda'_K}\prod_{j=0}^{i-1}\left[p^m_{u_j}\left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_j)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_j)+\lambda'_K}\right)\right] \\ &= \varepsilon\cdot\mathbb{E}\left[\sum_{i=0}^{K-1}\frac{\lambda'_K}{S^-_i(W)+\lambda'_K}\prod_{j=0}^{i-1}\frac{S^-_j(W)}{S^-_j(W)+\lambda'_K}\right] < \varepsilon,\end{aligned}$$

where we recall the definition of $(S^-_j(w))_{j\ge 0}$ from (2.51), and observe that the sum in the final line of the display telescopes. The first equation in (2.53) follows. Next, (2.46), Lemma 2.13, and the facts that $\lambda'_K \ge \lambda^*$ and $\lim_{m\to\infty}\tilde{g}^*_+ = \tilde{g}^* < \lambda^*$ together imply the second limit in (2.53). Finally, by (2.39) in Proposition 2.11 we have

$$\lambda'_K - \lambda^* = \lim_{n\to\infty}\left(\frac{\mathbf{a}'\cdot\hat{\mathcal{V}}^K_n - \mathcal{Z}_n - (\boldsymbol{\gamma}'\mathbf{a}')\cdot\hat{\mathcal{V}}^K_n}{n}\right) \le \lim_{n\to\infty}\frac{(\mathbf{a}' - \boldsymbol{\gamma}'\mathbf{a}')\cdot\hat{\mathcal{V}}^K_n}{n} = \mathcal{E}_K + \mathcal{F}_K + \mathcal{R}_K,$$

so that (2.54) follows by taking limits as $m\to\infty$ and $K\to\infty$. □

2.3.4 Proof of Assertions 3 and 4 of Theorem 1.1

Proof. We begin with the proof of Assertion 3. First, recalling the definition of $\mathscr{D}_{\geq k}(n, \cdot)$ from (2.36), by Proposition 2.12 for any $\ell \in [D_m]$ we have

$$\begin{aligned}\lim_{n \to \infty} \frac{\mathscr{D}_{\geq k}(n, \ell)}{n} &= \sum_{j=k}^{K} \sum_{\mathbf{u} \in [D_m]^{K+1}} \mathbf{V}_K(\mathbf{u}|_j) \mathbf{1}_{\{\ell\}}(u_0) \\ &= \sum_{\mathbf{u} \in [D_m]^{K+1}} \Bigg(p^m_{u_K} \prod_{i=0}^{K-1} \left[p^m_{u_i} \left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i) + \lambda'_K} \right) \right] \\ &\qquad + \sum_{j=k}^{K-1} \frac{p^m_{u_j} \lambda'_K}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_j) + \lambda'_K} \prod_{i=0}^{j-1} \left[p^m_{u_i} \left(\frac{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i)}{(\boldsymbol{\gamma}'\mathbf{a}')(\mathbf{u}|_i) + \lambda'_K} \right) \right] \Bigg) \mathbf{1}_{\{\ell\}}(u_0).\end{aligned}$$

Now, as with the proofs of Lemma 2.13 and Lemma 2.14, recalling the definition of $(S_i^-(w))_{i \geq 0}$ from (2.51), we may write the last equation as

$$\begin{aligned}&= \mathbb{E}\left[\prod_{i=0}^{K-1} \left(\frac{S_i^-(W)}{S_i^-(W) + \lambda'_K} \right) \mathbf{1}_{\mathcal{I}^m_\ell}(W) \right] \\ &\qquad + \sum_{j=k}^{K-1} \mathbb{E}\left[\frac{\lambda'_K}{S_j^-(W) + \lambda'_K} \prod_{i=0}^{j-1} \left(\frac{S_i^-(W)}{S_i^-(W) + \lambda'_K} \right) \mathbf{1}_{\mathcal{I}^m_\ell}(W) \right] \\ &= \mathbb{E}\left[\prod_{i=0}^{k-1} \left(\frac{S_i^-(W)}{S_i^-(W) + \lambda'_K} \right) \mathbf{1}_{\mathcal{I}^m_\ell}(W) \right]. \qquad (2.57)\end{aligned}$$

For $m' \in \mathbb{N}$, (2.57) allows us to prove the result for sets $S \in \sigma(\mathscr{I}^{m'})$, where we recall the definition of $\mathscr{I}^{m'}$ in (2.6), and (2.27) and (2.28). Since $N(n, \cdot)$ is finitely additive, if $S \in \sigma(\mathscr{I}^m)$, by (2.37) and (2.57) we have

$$\begin{aligned}\mathbb{E}\left[\prod_{i=0}^{k-1} \left(\frac{S_i^-(W)}{S_i^-(W) + \lambda'_K} \right) \mathbf{1}_S(W) \right] &\leq \liminf_{n \to \infty} \frac{N_{\geq k}(n, S)}{n} \leq \limsup_{n \to \infty} \frac{N_{\geq k}(n, S)}{n} \\ &\leq \mathbb{E}\left[\prod_{i=0}^{k-1} \left(\frac{S_i^-(W)}{S_i^-(W) + \lambda'_K} \right) \mathbf{1}_S(W) \right] + \mathcal{R}_K + \mathcal{E}_K + \mathcal{F}_K.\end{aligned}$$

Taking limits as $m \to \infty$ and then as $K \to \infty$, and applying Lemma 2.13 and Lemma 2.14 now proves the result for sets in $\sigma(\mathscr{I}^{m'})$. Now, note that for each $k \in \mathbb{N}_0$, and measurable sets $S' \subseteq [0, 1]$, we have

$$\limsup_{n \to \infty} \frac{N_{\geq k}(n, S')}{n} \leq \limsup_{n \to \infty} \frac{N_{\geq 0}(n, S')}{n} = \mu(S') \quad \text{almost surely,} \qquad (2.58)$$

where the last equality applies the strong law of large numbers.

We now prove the result for sets $U \in \mathcal{O}$ where $\mathcal{O}$ denotes the class of all open subsets of $[0, 1]$. For a fixed open set $U \in \mathcal{O}$, and $m \in \mathbb{N}$, recall that $\mathcal{I}^m(U) := \bigcup_{j \in [D_m]: \mathcal{I}^m_j \subseteq U} \mathcal{I}^m_j$. Also recall (2.31), which states that $\mathbf{1}_{\mathcal{I}^m(U)} \uparrow \mathbf{1}_U$ pointwise as $m \to \infty$. Now, since each $\mathcal{I}^m(U) \in \sigma(\mathscr{I}^m)$, by applying (2.58) for each $k \leq K$ we have

$$\begin{aligned}\mathbb{E}\left[\prod_{i=0}^{k-1} \left(\frac{S_i(W)}{S_i(W) + \lambda'_K} \right) \mathbf{1}_{\mathcal{I}^m(U)}(W) \right] &\leq \liminf_{n \to \infty} \frac{N_{\geq k}(n, U)}{n} \leq \limsup_{n \to \infty} \frac{N_{\geq k}(n, U)}{n} \\ &\leq \mathbb{E}\left[\prod_{i=0}^{k-1} \left(\frac{S_i(W)}{S_i(W) + \lambda'_K} \right) \mathbf{1}_{\mathcal{I}^m(U)}(W) \right] + \mu(U \setminus \mathcal{I}^m(U)).\end{aligned}$$

Taking limits as $m \to \infty$ and then $K \to \infty$ now proves the result for sets belonging to $\mathcal{O}$.

Finally, note that since μ is a *regular* measure, for any measurable set $A \subseteq [0,1]$ we have

$$\mu(A) = \inf_{U \in \mathcal{O}: A \subseteq U} \{\mu(U)\}.$$

Thus, for a given measurable set A, and any $\varepsilon > 0$, there exists an open set U_ε such that

$$\mu(U_\varepsilon \setminus A) \leq \varepsilon.$$

Therefore by finite additivity and (2.58)

$$\lim_{n\to\infty} \frac{N_{\geq k}(n, U_\varepsilon)}{n} - \varepsilon \leq \liminf_{n\to\infty} \frac{N_{\geq k}(n,A)}{n} \leq \limsup_{n\to\infty} \frac{N_{\geq k}(n,A)}{n} \leq \lim_{n\to\infty} \frac{N_{\geq k}(n,U_\varepsilon)}{n}.$$

Assertion 3 now follows by applying the result for the class $\mathcal{O}$, and sending $\varepsilon \to 0$.

Finally, Assertion 3 allows us to prove Assertion 4 of Theorem 1.1. Note that, if $N_k(n, A)$ denotes the number of vertices of out-degree k in the tree at time n having weight in A, by counting the edges in the tree in two ways we have

$$\Xi(n, A) = \sum_{k=1}^{n} k N_k(n, A) = \sum_{k=1}^{n} N_{\geq k}(n, A).$$

But now, Lemma 1.2 and using Fatou's Lemma in the last inequality, we have,

$$\begin{aligned}(\psi_*\mu)(A) = \mathbb{E}\left[\frac{h(W)}{\lambda^* - \tilde{g}(W)} \mathbf{1}_A\right] &= \sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+\lambda^*}\right)\mathbf{1}_A\right] \\ &\leq \sum_{k=1}^{\infty} \liminf_{n\to\infty} \frac{N_{\geq k}(n,A)}{n} \leq \liminf_{n\to\infty} \frac{\Xi(n,A)}{n};\end{aligned}$$

and likewise, $\liminf_{n\to\infty} \frac{\Xi(n,A^c)}{n} \geq (\psi_*\mu)(A^c)$. Now, since we add one edge at each time-step, it follows that $\Xi(n, [0,1]) = n$. Thus, by finite additivity,

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

But, since (1.5) implies that $(\psi_*\mu)(\cdot)$ is a probability measure, this is only possible if

$$\limsup_{n\to\infty} \frac{\Xi(n,A)}{n} = (\psi_*\mu)(A) \text{ and } \liminf_{n\to\infty}\frac{\Xi(n,A^c)}{n} = (\psi_*\mu)(A^c) \text{ almost surely.} \tag{2.59}$$

The result follows. □

3 The condensation regime

Here, we extend the results of the previous section to the condensation regime. The techniques used in this section are closely related to those of [25].

The results of this section depend on sequences of auxiliary trees $\mathcal{T}^{(\varepsilon)}, \mathcal{T}^{(-\varepsilon)}, \varepsilon > 0$. Given $\varepsilon > 0$, and $\mathcal{M}_\varepsilon$ as defined in (1.13), define the functions $g_\varepsilon, g_{-\varepsilon}$ such that, for $z \in \mathcal{M}$,

$$g_\varepsilon(p,q) := \mathbf{1}_{\mathcal{M}_\varepsilon^c}(p) g(p,q) + \mathbf{1}_{\mathcal{M}_\varepsilon}(p) g(z,q)$$

and

$$g_{-\varepsilon}(p,q) := \mathbf{1}_{\mathcal{M}_\varepsilon^c}(p) g(p,q) + \mathbf{1}_{\mathcal{M}_\varepsilon}(p)(g(z,q) - u_\varepsilon(q));$$

and let $\mathcal{T}^{(\varepsilon)}, \mathcal{T}^{(-\varepsilon)}$ be the evolving trees with measure μ, and associated functions g_{ε}, h and $g_{-\varepsilon}, h$ respectively. We also denote by $(\mathcal{Z}_n^{(\varepsilon)})_{n \geq 0}$ and $(\mathcal{Z}_n^{(-\varepsilon)})_{n \geq 0}$ the partition functions associated with $\mathcal{T}^{(\varepsilon)}, \mathcal{T}^{(-\varepsilon)}$, respectively.

Lemma 3.1. Assume Conditions **D1-D4**. Then, for each $\varepsilon > 0$ sufficiently small, $\mathcal{T}^{(\varepsilon)}$ and $\mathcal{T}^{(-\varepsilon)}$ satisfy Conditions **C1** and **C2**. In addition, if $\lambda_{\varepsilon}, \lambda_{-\varepsilon}$ denote the Malthusian parameters associated with $\mathcal{T}^{(\varepsilon)}, \mathcal{T}^{(-\varepsilon)}$, then $\lambda_{\varepsilon} \downarrow \tilde{g}^*$ and $\lambda_{-\varepsilon} \uparrow \tilde{g}^*$ as $\varepsilon \downarrow 0$.

Proof. First, since by **D2** g satisfies Condition **C2**, we have

$$g(x,y) = \kappa\left(\phi_1^{(1)}(x), \dots, \phi_1^{(N)}(x), \phi_2^{(1)}(y), \dots, \phi_2^{(N)}(y)\right),$$

for measurable functions $\phi_j^{(i)} : [0,1] \to [0,1]$, $j = 1,2$, $i \in [N]$ and a bounded continuous function $\kappa : [0,1]^{2N} \to [0,\infty)$. Now, if we set $\phi_1^{(N+1)}(x) := \mathbf{1}_{\mathcal{M}_{\varepsilon}}(x), \phi_1^{(N+2)}(x) := \mathbf{1}_{\mathcal{M}_{\varepsilon}^c}(x)$, $\phi_2^{(N+1)}(y) := g(x^*, y) - u_{\varepsilon}(y)$ and define κ' such that

$$\kappa'(c_1, \dots, c_{N+2}, d_1, \dots d_{N+1}) := c_{N+2}\kappa(c_1, \dots, c_N, d_1, \dots, d_N) + c_{N+1}d_{N+1},$$

we clearly have that $\phi_1^{(N+1)}, \phi_1^{(N+2)}, \phi_2^{(N+1)}$ are bounded, non-negative measurable functions, and κ' is bounded and continuous, taking values in $[0, \infty)$. Noting that

$$g_{-\varepsilon}(x,y) = \kappa'\left(\phi_1^{(1)}(x), \dots, \phi_1^{(N+2)}(x), \phi_2^{(1)}(y), \dots, \phi_2^{(N+1)}(y)\right),$$

it follows that $g_{-\varepsilon}$ satisfies Condition **C2**. The proof of **C2** for g_{ε} is similar.

For **C1**, since h is bounded, for sufficiently large $\lambda > \tilde{g}^*$, we have

$$\mathbb{E}\left[\frac{h(W)}{\lambda - \tilde{g}_{\varepsilon}(W)}\right] < 1.$$

Meanwhile, since, by Condition **D4**, $\mu(\mathcal{M}_{\varepsilon}) > 0$ and $\tilde{g}_{\varepsilon}(x) = \tilde{g}^*$ for any $x \in \mathcal{M}_{\varepsilon}$, by monotone convergence

$$\lim_{\lambda \downarrow \tilde{g}^*} \mathbb{E}\left[\frac{h(W)}{\lambda - \tilde{g}_{\varepsilon}(W)}\right] = \mathbb{E}\left[\frac{h(W)}{\tilde{g}^* - \tilde{g}_{\varepsilon}(W)}\right] = \infty.$$

Thus, by continuity in λ, Condition **C1** is satisfied for $\mathcal{T}^{(\varepsilon)}$. A similar argument also works for $\mathcal{T}^{(-\varepsilon)}$: if $\tilde{g}_{-\varepsilon}^*$ denotes the maximum value of $\tilde{g}_{-\varepsilon}(x)$, then this value is also attained on $\mathcal{M}_{\varepsilon}$ which has positive measure. If $\lambda_{\varepsilon}, \lambda_{-\varepsilon}$ denote the associated Malthusian parameters associated with the trees, then, for each $\varepsilon > 0$, $\lambda_{\varepsilon} > \tilde{g}^*$ and $\lambda_{-\varepsilon} > \tilde{g}_{-\varepsilon}^*$. Moreover, since g_{ε} is non-increasing pointwise as ε decreases, λ_{ε} is non-increasing in ε; likewise, $\lambda_{-\varepsilon}$ is non-decreasing in ε. Now, suppose $\lim_{\varepsilon \downarrow 0} \lambda_{\varepsilon} = \lambda_+ > \tilde{g}^*$. Then we may apply dominated convergence, and

$$1 = \lim_{\varepsilon \downarrow 0} \mathbb{E}\left[\frac{h(W)}{\lambda_{\varepsilon} - \tilde{g}_{\varepsilon}(W)}\right] = \mathbb{E}\left[\lim_{\varepsilon \downarrow 0} \frac{h(W)}{\lambda_{\varepsilon} - \tilde{g}_{\varepsilon}(W)}\right] = \mathbb{E}\left[\frac{h(W)}{\lambda_+ - \tilde{g}(W)}\right],$$

contradicting (1.12). The case for $\lambda_{-\varepsilon}$ follows identically. $\square$

Lemma 3.2. There exists a coupling $(\hat{\mathcal{T}}^{(-\varepsilon)}, \hat{\mathcal{T}}, \hat{\mathcal{T}}^{(\varepsilon)})$ of these processes such that, almost surely (on the coupling space), for all $n \in \mathbb{N}_0$,

$$\mathcal{Z}_n^{(-\varepsilon)} \leq \mathcal{Z}_n \leq \mathcal{Z}_n^{(\varepsilon)}, \tag{3.1}$$

and, for each vertex v with $W_v \in \mathcal{M}_{\varepsilon}^c$, we have

$$f(N^+(v, \hat{\mathcal{T}}_n^{(\varepsilon)})) \leq f(N^+(v, \hat{\mathcal{T}}_n)) \leq f(N^+(v, \hat{\mathcal{T}}_n^{(-\varepsilon)})) \tag{3.2}$$

and

$$\deg(v, \hat{\mathcal{T}}_n^{(\varepsilon)}) \leq \deg(v, \hat{\mathcal{T}}_n) \leq \deg(v, \hat{\mathcal{T}}_n^{(-\varepsilon)}). \tag{3.3}$$

Proof. We initialise the trees with a single vertex 0 having weight W_0 sampled independently from μ, conditioned on $\{h(W_0) > 0\}$ and construct copies of these three tree processes on the same vertex set, which is identified with $\mathbb{N}_0$. Now, assume that at the nth time-step,

$$(\hat{\mathcal{T}}_j^{(-\varepsilon)})_{0\le j\le n} \sim (\mathcal{T}_j^{(-\varepsilon)})_{0\le j\le n}, \quad (\hat{\mathcal{T}}_j)_{0\le j\le n} \sim (\mathcal{T}_j)_{0\le j\le n} \quad \text{and} \quad (\hat{\mathcal{T}}_j^{(\varepsilon)})_{0\le j\le n} \sim (\mathcal{T}_j^{(\varepsilon)})_{0\le j\le n}.$$

In addition, assume that (3.1) and (3.2) are satisfied up to time n.

Now, for the $(n+1)$st step:

- Introduce vertex $n+1$ with weight W_{n+1} sampled independently from μ in $\hat{\mathcal{T}}_n^{(-\varepsilon)}$, $\hat{\mathcal{T}}_n$ and $\hat{\mathcal{T}}_n^{(\varepsilon)}$.
- Form $\hat{\mathcal{T}}_{n+1}^{(-\varepsilon)}$ by sampling the parent v of $n+1$ independently according to the law of $\mathcal{T}^{(-\varepsilon)}$ (i.e. with probability proportional to $f(N^+(v, \hat{\mathcal{T}}_n^{(-\varepsilon)})))$. Then, in order to form $\hat{\mathcal{T}}_{n+1}$ sample an independent uniformly distributed random variables U_1 on $[0,1]$.
  - If $U_1 \le \frac{\mathcal{Z}_n^{(-\varepsilon)} f(N^+(v,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n f(N^+(v,\hat{\mathcal{T}}_n^{(-\varepsilon)}))}$ and $W_v \in \mathcal{M}_\varepsilon^c$, select v as the parent of $n+1$ in $\hat{\mathcal{T}}_{n+1}$ as well.
  - Otherwise, form $\hat{\mathcal{T}}_{n+1}$ by selecting the parent v' of $n+1$ with probability proportional to $f(N^+(v', \hat{\mathcal{T}}_n))$ out of all all the vertices with weight $W_{v'} \in \mathcal{M}_\varepsilon$.
- Then form $\hat{\mathcal{T}}_{n+1}^{(\varepsilon)}$ in a similar manner. Sample an independent uniform random variable U_2 on $[0,1]$.
  - If vertex v (with weight $W_v \in \mathcal{M}_\varepsilon^c$) was chosen as the parent of $n+1$ in $\hat{\mathcal{T}}_{n+1}$ and $U_2 \le \frac{\mathcal{Z}_n f(N^+(v,\hat{\mathcal{T}}_n^{(\varepsilon)}))}{\mathcal{Z}_n^{(\varepsilon)} f(N^+(v,\hat{\mathcal{T}}_n))}$, also select v as the parent of $n+1$ in $\hat{\mathcal{T}}_{n+1}^{\varepsilon}$.
  - Otherwise, form $\hat{\mathcal{T}}_{n+1}^{(\varepsilon)}$ by selecting the parent v'' of $n+1$ with probability proportional to $f(N^+(v'', \mathcal{T}_n^{(\varepsilon)}))$ out of all the vertices with weight $W_{v''} \in \mathcal{M}_\varepsilon$.

Clearly $\hat{\mathcal{T}}_{n+1}^{(-\varepsilon)} \sim \mathcal{T}_{n+1}^{(-\varepsilon)}$. On the other hand, in $\hat{\mathcal{T}}_{n+1}$ the probability of choosing a certain parent v of $n+1$ with weight $W_v \in \mathcal{M}_\varepsilon^c$ is

$$\frac{\mathcal{Z}_n^{(-\varepsilon)} f(N^+(v,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n f(N^+(v,\hat{\mathcal{T}}_n^{(-\varepsilon)}))} \times \frac{f(N^+(v,\hat{\mathcal{T}}_n^{(-\varepsilon)}))}{\mathcal{Z}_n^{(-\varepsilon)}} = \frac{f(N^+(v,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n},$$

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

$$\begin{aligned}
&\frac{f(N^+(v',\hat{\mathcal{T}}_n))}{\sum_{v':W_{v'}\in\mathcal{M}_\varepsilon} f(N^+(v',\hat{\mathcal{T}}_n))} \left(\sum_{v:W_v\in\mathcal{M}_\varepsilon^c} \left(1 - \frac{\mathcal{Z}_n^{(-\varepsilon)} f(N^+(v,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n f(N^+(v,\hat{\mathcal{T}}_n^{(-\varepsilon)}))}\right) \frac{f(N^+(v,\hat{\mathcal{T}}_n^{(-\varepsilon)}))}{\mathcal{Z}_n^{(-\varepsilon)}} \right) \\
&\qquad + \frac{f(N^+(v',\hat{\mathcal{T}}_n))}{\sum_{v':W_{v'}\in\mathcal{M}_\varepsilon} f(N^+(v',\hat{\mathcal{T}}_n))} \left(\sum_{v:W_v\in\mathcal{M}_\varepsilon} \frac{f(N^+(v,\hat{\mathcal{T}}_n^{(-\varepsilon)}))}{\mathcal{Z}_n^{(-\varepsilon)}} \right) \\
&= \frac{f(N^+(v',\hat{\mathcal{T}}_n))}{\sum_{v':W_{v'}\in\mathcal{M}_\varepsilon} f(N^+(v',\hat{\mathcal{T}}_n))} \left(\sum_{v} \frac{f(N^+(v,\hat{\mathcal{T}}_n^{(-\varepsilon)})))}{\mathcal{Z}_n^{(-\varepsilon)}} - \sum_{v:W_v\in\mathcal{M}_\varepsilon^c} \frac{f(N^+(v,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n} \right) \\
&= \frac{f(N^+(v',\hat{\mathcal{T}}_n))}{\sum_{v':W_{v'}\in\mathcal{M}_\varepsilon} f(N^+(v',\hat{\mathcal{T}}_n))} \left(1 - \frac{\sum_{v:W_v\in\mathcal{M}_\varepsilon^c} f(N^+(v,\hat{\mathcal{T}}_n))}{\mathcal{Z}_n}\right) = \frac{f(N^+(v',\hat{\mathcal{T}}_n))}{\mathcal{Z}_n},
\end{aligned}$$

where we use the fact that $\sum_v f(N^+(v,\hat{\mathcal{T}}_n)) = \mathcal{Z}_n$. Thus, we have $\hat{\mathcal{T}}_{n+1} \sim \mathcal{T}_{n+1}$. Now, note that if the parent v of $n+1$ in $\hat{\mathcal{T}}_{n+1}^{(-\varepsilon)}$ is such that $W_v \in \mathcal{M}_\varepsilon^c$, the same parent is chosen in $\hat{\mathcal{T}}_{n+1}$. Since $W_v \in \mathcal{M}_\varepsilon^c$, we have

$$\begin{aligned} f(N^+(v, \hat{\mathcal{T}}_{n+1}^{(-\varepsilon)})) - f(N^+(v, \hat{\mathcal{T}}_n^{(-\varepsilon)})) &= g_{-\varepsilon}(W_v, W_{n+1}) = g(W_v, W_{n+1}) \\ &= f(N^+(v, \hat{\mathcal{T}}_{n+1})) - f(N^+(v, \hat{\mathcal{T}}_n)). \end{aligned}$$

Otherwise, the parent of $n+1$ in $\hat{\mathcal{T}}_{n+1}$ has weight which belongs to $\mathcal{M}_\varepsilon$, and thus $f(N^+(v, \hat{\mathcal{T}}_n^{(-\varepsilon)}))$ increases whilst $f(N^+(v, \hat{\mathcal{T}}_n))$ stays the same. An increase in $f(N^+(v, \hat{\mathcal{T}}_n^{(-\varepsilon)}))$ coincides with the increase of $\deg(v, \hat{\mathcal{T}}_n^{(-\varepsilon)})$, and thus the right hand sides of (3.2) and (3.3) are satisfied for time $n+1$.

Now, note that

$$\mathcal{Z}_{n+1}^{(-\varepsilon)} - \mathcal{Z}_n^{(-\varepsilon)} = h(W_{n+1}) + g_{-\varepsilon}(W_v, W_{n+1}), \text{ and } \mathcal{Z}_{n+1} - \mathcal{Z}_n = h(W_{n+1}) + g(W_{v'}, W_{n+1})$$

where v, v' denote the parent of $n+1$ in $\hat{\mathcal{T}}_n$ and $\hat{\mathcal{T}}_n^{(\varepsilon)}$ respectively. Then we either have:

- $v = v'$, so that $g_{-\varepsilon}(W_v, W_{n+1}) = g(W_{v'}, W_{n+1})$.
- $v \in \mathcal{M}_\varepsilon^c$ and $v' \in \mathcal{M}_\varepsilon$, in which case, $\mathbb{P}$-a.s, using **D4**

$$g_{-\varepsilon}(W_v, W_{n+1}) = g(W_v, W_{n+1}) \leq g(x^*, W_{n+1}) - u_\varepsilon(W_{n+1}) < g(W_{v'}, W_{n+1}).$$

- Both $v, v' \in \mathcal{M}_\varepsilon$, in which case, $\mathbb{P}$-a.s.,

$$g_{-\varepsilon}(W_v, W_{n+1}) = g(x^*, W_{n+1}) - u_\varepsilon(W_{n+1}) < g(W_{v'}, W_{n+1}).$$

In every case we have $\mathcal{Z}_{n+1}^{(-\varepsilon)} - \mathcal{Z}_n^{(-\varepsilon)} \leq \mathcal{Z}_{n+1} - \mathcal{Z}_n$, and thus (3.1) is also satisfied at time $n+1$.

Each of the statements concerning $\hat{\mathcal{T}}^{(\varepsilon)}$ follow in an analogous manner, applying Condition **D3**. □

3.1 Proof of Theorem 1.3

The proof of Theorem 1.3 uses the auxiliary trees $\mathcal{T}^{(\varepsilon)}$ and $\mathcal{T}^{(-\varepsilon)}$, and Lemma 3.2.

Proof of Theorem 1.3. For the first statement, note that by (3.1) in Lemma 3.2 and Theorem 1.1, for each $\varepsilon > 0$ we have, $\mathbb{P}$-a.s.,

$$\lambda_{-\varepsilon} = \lim_{n\to\infty} \frac{\mathcal{Z}_n^{(-\varepsilon)}}{n} \leq \liminf_{n\to\infty} \frac{\mathcal{Z}_n}{n} \leq \limsup_{n\to\infty} \frac{\mathcal{Z}_n}{n} = \lim_{n\to\infty} \frac{\mathcal{Z}_n^{(\varepsilon)}}{n} = \lambda_\varepsilon.$$

The statement follows by sending $\varepsilon \to 0$, using Lemma 3.1.

Next, by assumption, for each $\varepsilon > 0$ sufficiently small, we have $A \subseteq \mathcal{M}_\varepsilon^c$. Next, applying (3.3), if $\Xi^{(\varepsilon)}$ and $\Xi^{(-\varepsilon)}$ denote the edge distributions in the coupled trees $\hat{\mathcal{T}}^{(\varepsilon)}, \hat{\mathcal{T}}^{(-\varepsilon)}$, respectively, then for each $n \in \mathbb{N}_0$

$$\Xi^{(\varepsilon)}(n, A) \leq \Xi(n, A) \leq \Xi^{(-\varepsilon)}(n, A),$$

and thus, by Theorem 1.1, we have

$$\begin{aligned} \mathbb{E}\left[\frac{h(W)}{\lambda_\varepsilon - \tilde{g}_\varepsilon(W)} \mathbf{1}_A(W)\right] &\leq \liminf_{n\to\infty} \frac{\Xi(n, A)}{n} \\ &\leq \limsup_{n\to\infty} \frac{\Xi(n, A)}{n} \leq \mathbb{E}\left[\frac{h(W)}{\lambda_{-\varepsilon} - \tilde{g}_{-\varepsilon}(W)} \mathbf{1}_A(W)\right]. \end{aligned} \tag{3.4}$$

Now, noting that $\tilde{g}_{-\varepsilon} = \tilde{g} = \tilde{g}_{\varepsilon}$ on A, and $\lambda_{-\varepsilon} > \tilde{g}^*_{-\varepsilon} \geq \sup_{x \in A} \tilde{g}(x)$ and is non-decreasing in ε, by applying Lemma 3.1 and dominated convergence we have

$$\begin{aligned}\lim_{\varepsilon \to 0} \mathbb{E}\left[\frac{h(W)}{\lambda_{\varepsilon} - \tilde{g}_{\varepsilon}(W)} \mathbf{1}_A(W)\right] &= \lim_{\varepsilon \to 0} \mathbb{E}\left[\frac{h(W)}{\lambda_{-\varepsilon} - \tilde{g}_{-\varepsilon}(W)} \mathbf{1}_A(W)\right] \\ &= \mathbb{E}\left[\frac{h(W)}{\tilde{g}^* - \tilde{g}(W)} \mathbf{1}_A(W)\right]. \end{aligned} \tag{3.5}$$

Then, (1.14) follows by combining (3.4) and (3.5). Moreover, for each $\varepsilon' > 0$, by setting $A = \mathcal{M}^c_{\varepsilon'}$,

$$\lim_{n \to \infty} \frac{\Xi(n, \mathcal{M}_{\varepsilon'})}{n} = \lim_{n \to \infty} \left(1 - \frac{\Xi(n, \mathcal{M}^c_{\varepsilon'})}{n}\right) = 1 - \mathbb{E}\left[\frac{h(W)}{\tilde{g}^* - \tilde{g}(W)} \mathbf{1}_{\mathcal{M}^c_{\varepsilon'}}(W)\right]. \tag{3.6}$$

But then, again by dominated convergence,

$$\lim_{\varepsilon' \to 0} \mathbb{E}\left[\frac{h(W)}{\tilde{g}^* - \tilde{g}(W)} \mathbf{1}_{\mathcal{M}^c_{\varepsilon'}}(W)\right] = \mathbb{E}\left[\frac{h(W)}{\tilde{g}^* - \tilde{g}(W)}\right],$$

and (1.15) follows.

Finally, for the last statement, recall the definition of the companion process $(S_i)_{i \geq 0}$ in (1.8), and that, for any measurable $B \subseteq [0,1]$, $N_{\geq k}(n, B)$ denotes the number of vertices of out-degree at least k with weight belonging to B at time n. Then, for $\varepsilon > 0$, note that

$$\frac{N_{\geq k}(n, B \cap \mathcal{M}^c_{\varepsilon})}{n} \leq \frac{N_{\geq k}(n, B)}{n} \leq \frac{N_{\geq k}(n, B \cap \mathcal{M}^c_{\varepsilon})}{n} + \frac{N_{\geq 0}(n, \mathcal{M}_{\varepsilon})}{n}.$$

Now, by the strong law of large numbers, in the limit as $n \to \infty$, as in (2.58), the second summand in the right-hand side tends to $\mu(\mathcal{M}_{\varepsilon})$, and thus,

$$\liminf_{n \to \infty} \frac{N_{\geq k}(n, B \cap \mathcal{M}^c_{\varepsilon})}{n} \leq \limsup_{n \to \infty} \frac{N_{\geq k}(n, B)}{n} \tag{3.7}$$

$$\leq \limsup_{n \to \infty} \frac{N_{\geq k}(n, B \cap \mathcal{M}^c_{\varepsilon})}{n} + \mu(\mathcal{M}_{\varepsilon}). \tag{3.8}$$

Now, let $N^{(-\varepsilon)}_{\geq k}(n, \cdot), N^{(\varepsilon)}_{\geq k}(n, \cdot)$ denote the associated quantities in the trees $\mathcal{T}^{(-\varepsilon)}, \mathcal{T}^{(\varepsilon)}$, and denote by $(S^{(-\varepsilon)}_i)_{i \geq 0}$ and $(S^{(\varepsilon)}_i)_{i \geq 0}$ the companion processes defined in terms of the functions $h, g_{-\varepsilon}$ and $h, g_{+\varepsilon}$ respectively. Then, by (3.3), on the coupling in Lemma 3.2, we have

$$N^{(\varepsilon)}_{\geq k}(n, B \cap \mathcal{M}^c_{\varepsilon}) \leq N_{\geq k}(n, B \cap \mathcal{M}^c_{\varepsilon}) \leq N^{(-\varepsilon)}_{\geq k}(n, B \cap \mathcal{M}^c_{\varepsilon}).$$

Therefore, by Theorem 1.1, recalling the definitions of $\lambda_{\varepsilon}, \lambda_{-\varepsilon}$ in Lemma 3.1,

$$\begin{aligned}\mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S^{(\varepsilon)}_i(W)}{S^{(\varepsilon)}_i(W) + \lambda_{\varepsilon}}\right) \mathbf{1}_{B \cap \mathcal{M}^c_{\varepsilon}}(W)\right] &\leq \liminf_{n \to \infty} \frac{N_{\geq k}(n, B \cap \mathcal{M}^c_{\varepsilon})}{n} \\ &\leq \limsup_{n \to \infty} \frac{N_{\geq k}(n, B \cap \mathcal{M}^c_{\varepsilon})}{n} \\ &\leq \mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S^{(-\varepsilon)}_i(W)}{S^{(-\varepsilon)}_i(W) + \lambda_{-\varepsilon}}\right) \mathbf{1}_{B \cap \mathcal{M}^c_{\varepsilon}}(W)\right],\end{aligned}$$

and thus, by (3.7), we have

$$\begin{aligned}\mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i^{(\varepsilon)}(W)}{S_i^{(\varepsilon)}(W)+\lambda_\varepsilon}\right)\mathbf{1}_{B\cap\mathcal{M}_\varepsilon^c}(W)\right] &\leq \liminf_{n\to\infty}\frac{N_{\geq k}(n,B)}{n} \\ &\leq \limsup_{n\to\infty}\frac{N_{\geq k}(n,B)}{n} \\ &\leq \mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i^{(-\varepsilon)}(W)}{S_i^{(-\varepsilon)}(W)+\lambda_{-\varepsilon}}\right)\mathbf{1}_{B\cap\mathcal{M}_\varepsilon^c}(W)\right]+\mu(\mathcal{M}_\varepsilon).\end{aligned} \tag{3.9}$$

Now, by dominated convergence, as $\varepsilon \to 0$

$$\mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i^{(\varepsilon)}(W)}{S_i^{(\varepsilon)}(W)+\lambda_\varepsilon}\right)\mathbf{1}_{B\cap\mathcal{M}_\varepsilon}(W)\right]\to\mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+\tilde{g}^*}\right)\mathbf{1}_B(W)\right], \text{ and}$$

$$\mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i^{(-\varepsilon)}(W)}{S_i^{(-\varepsilon)}(W)+\lambda_{-\varepsilon}}\right)\mathbf{1}_{B\cap\mathcal{M}_\varepsilon}(W)\right]\to\mathbb{E}\left[\prod_{i=0}^{k-1}\left(\frac{S_i(W)}{S_i(W)+\tilde{g}^*}\right)\mathbf{1}_B(W)\right],$$

and, since, by (1.12), $\mathcal{M}$ is a μ-null set, $\mu(\mathcal{M}_\varepsilon) \to 0$. Combining these statements with (3.9) completes the proof. □

3.2 Proof of Corollary 1.4

Proof. By the Portmanteau theorem, it suffices to show that, $\mathbb{P}$-a.s. $\liminf_{n\to\infty}\frac{\Xi(n,O)}{n} \geq \Pi(O)$, for any open set $O \subseteq [0,1]$. Note that, in view of Theorem 1.3 and (3.6), we have that, $\mathbb{P}$-a.s.,

$$\lim_{n\to\infty}\frac{\Xi(n,S)}{n} = \Pi(S) \qquad \forall S \in \mathscr{S},$$

where $\mathscr{S}$ is the countable family of sets defined as

$$\mathscr{S} := \left\{\mathcal{D}_j^m \cap \mathcal{M}_{1/n}^c,\, \mathcal{M}_{1/n} : j \in [2^m], m \in \mathbb{N}, n \in \mathbb{N}\right\}.$$

It is well known that any open subset of $[0,1]$ is a disjoint countable union of half open dyadic cubes (belonging to the family $\mathscr{D} := \left\{\mathcal{D}_j^m : j \in [2^m], m \in \mathbb{N}\right\}$). Suppose $z \notin O$. If we fix such a countable collection of dyadic cubes $D_1, D_2, \ldots$, we have, for each $k, j \in \mathbb{N}$

$$\liminf_{n\to\infty}\frac{\Xi(n,O)}{n} \geq \sum_{i=1}^{k}\lim_{n\to\infty}\frac{\Xi\left(n, D_i\cap\mathcal{M}_{1/j}^c\right)}{n} = \sum_{i=1}^{k}\Pi(D_i\cap\mathcal{M}_{1/j}^c).$$

Taking limits in j and k, by the monotone convergence theorem, the right hand side converges to $\mathbb{E}\left[\frac{h(W)}{\tilde{g}^*-\tilde{g}(W)}\mathbf{1}_O(W)\right] = \Pi(O)$, as required. Now, suppose $z \in O$. A similar argument to the previous case suffices if we can prove that $\mathcal{M}_{1/n} \subseteq O$ for n sufficiently large. But then, if not, $(O^c \cap \overline{\mathcal{M}}_{1/n})_{n\in\mathbb{N}}$ is a nested sequence of non-empty closed sets, thus, by Cantor's intersection theorem,

$$\varnothing \neq \bigcap_{n\in\mathbb{N}}\left(O^c \cap \overline{\mathcal{M}}_{1/n}\right) = O^c \cap \bigcap_{n\in\mathbb{N}}\overline{\mathcal{M}}_{1/n} = O^c \cap \{z\},$$

a contradiction. □

4 Appendix

4.1 Proof of Lemma 1.2

In order to prove Lemma 1.2 we first introduce an auxiliary, piecewise constant continuous time Markov process $(\mathcal{Y}_w(t), r_w(t))_{t\geq 0}$ taking values in $\mathbb{N}\times[0,\infty)$. The idea is to compute the expected value of $\mathcal{Y}_w(t)$ at an independent, exponentially distributed stopping time in two different ways. Let $(W_i)_{i\geq 0}$ be independent μ-distributed random variables, and define $(S_i(w))_{i\geq 0}$ according to (1.8), that is,

$$S_0(w) := h(w); \quad S_{i+1}(w) := S_i(w) + g(w, W_{i+1}),\ i \geq 0.$$

In addition, set $\tau_0 = 0$, and define $(\tau_i)_{i\geq 1}$ recursively so that, given $S_i(w)$,

$$\tau_{i+1} - \tau_i \sim \mathrm{Exp}(S_i(w)); \tag{4.1}$$

where $\mathrm{Exp}(r_w(\tau_{i-1}))$ denotes an exponentially distributed random variable with parameter $r_w(\tau_{i-1})$. Then, we set

$$\mathcal{Y}_w(t) := \sum_{n=1}^{\infty} \mathbf{1}_{[\tau_n,\infty)}(t), \quad \text{and} \quad r_w(t) := \sum_{n=0}^{\infty} S_n(w)\mathbf{1}_{[\tau_n,\tau_{n+1})}(t).$$

Claim 4.0.1. For all $t \in [0,\infty)$, we have $\mathbb{E}\left[\mathcal{Y}_w(t)\right] < \infty$ almost surely.

Proof. Let α be an independent exponentially distributed random variable with parameter $a > 0$, and set $\mathcal{Y}_w(\alpha) := \inf_{t\geq\alpha}(\mathcal{Y}_w(t))$. Then,

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k\}}|S_{k-1}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-1\}}\right] &= \mathbb{E}\left[\mathbf{1}_{\{\alpha\geq\tau_k\}}|S_{k-1}(w)\right]\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-1\}} &(4.2)\\
&= \frac{S_{k-1}(w)}{a + S_{k-1}(w)}\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-1\}}, &(4.3)
\end{aligned}$$

where in the last equality we have used (4.1) and the memory-less property of the exponential distribution. Note also, that for any $j \leq k-1$, the random variables $(S_j(w), \ldots, S_{k-1}(w))$ and $\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq j\}}$ are conditionally independent given the random variables $S_{j-1}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq j-1\}}$. Indeed, for each $\ell \in \{j, \ldots, k-1\}$,

$$S_\ell(w) = S_{j-1}(w) + \sum_{i=j}^{\ell} g(w, W_i),$$

where $W_j, \ldots, W_{k-1}$ are independent random variables sampled from μ, while (again using the memory-less property)

$$\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq j\}} = \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq j-1\}} \times \mathbf{1}_{\{\alpha\geq\tau_j-\tau_{j-1}\}}$$

where we recall $\tau_j - \tau_{j-1}$ is an exponentially distributed random variable with parameter $S_{j-1}(w)$ and thus conditionally independent of $(S_j(w), \ldots, S_{k-1}(w))$. As a result, we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\prod_{i=j}^{k-1}\frac{S_i(w)}{S_i(w)+a}\right)\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq j\}}\middle|S_{j-1}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq j-1\}}\right]\\
&= \mathbb{E}\left[\left(\prod_{i=j}^{k-1}\frac{S_i(w)}{S_i(w)+a}\right)\middle|S_{j-1}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq j-1\}}\right]\mathbb{E}\left[\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq j\}}\middle|S_{j-1}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq j-1\}}\right].
\end{aligned} \tag{4.4}$$

Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\mathcal{Y}_w(\alpha) \geq k\right) = \mathbb{E}\left[\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k\}}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k\}} | S_{k-1}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-1\}}\right]\right] \\
&\overset{(4.2)}{=} \mathbb{E}\left[\frac{S_{k-1}(w)}{a + S_{k-1}(w)} \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-1\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{S_{k-1}(w)}{a + S_{k-1}(w)} \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-1\}} \middle| S_{k-2}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-2\}}\right]\right] \\
&\overset{(4.4)}{=} \mathbb{E}\left[\mathbb{E}\left[\frac{S_{k-1}(w)}{a + S_{k-1}(w)} \middle| S_{k-2}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-2\}}\right] \mathbb{E}\left[\mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-1\}} | S_{k-2}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-2\}}\right]\right] \\
&\overset{(4.2)}{=} \mathbb{E}\left[\mathbb{E}\left[\frac{S_{k-1}(w)}{a + S_{k-1}(w)} \times \frac{S_{k-2}(w)}{a + S_{k-2}(w)} \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-2\}} \middle| S_{k-2}(w), \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-2\}}\right]\right] \\
&= \mathbb{E}\left[\frac{S_{k-1}(w)}{a + S_{k-1}(w)} \times \frac{S_{k-2}(w)}{a + S_{k-2}(w)} \mathbf{1}_{\{\mathcal{Y}_w(\alpha)\geq k-2\}}\right].
\end{aligned}
$$

Iterating in this manner and noting that $\mathcal{Y}_w(\alpha) \geq 0$ almost surely, we deduce that the previous expression is $\mathbb{E}\left[\prod_{i=0}^{k-1} \frac{S_i(w)}{a+S_i(w)}\right]$. This now implies that

$$
\mathbb{E}\left[\mathcal{Y}_w(\alpha)\right] = \sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=0}^{k-1} \frac{S_i(w)}{a + S_i(w)}\right]. \tag{4.5}
$$

Now, the display on the right is increasing in $S_i(w)$, and using the fact that g and h are bounded by $J = \max\{1, g\}$, we may bound this above by

$$
\sum_{k=1}^{\infty} \prod_{i=1}^{k} \frac{Ji}{Ji + a} < \infty \quad \text{for all } a > J\text{, by applying, for example, Stirling's approximation.}
$$

Thus, for a suitable choice of a, $\mathbb{E}\left[\mathcal{Y}_w(\alpha)\right]$ is finite, so that, in particular, for each $t \in [0, \infty)$, since the random variable $\mathcal{Y}_w(t)$ is independent of the event $\{\alpha \geq t\}$ which occurs with positive probability,

$$
\mathbb{E}\left[\mathcal{Y}_w(t)\right] \leq \frac{\mathbb{E}\left[Y_w(\alpha)\mathbf{1}_{\{\alpha\geq t\}}\right]}{\mathbb{P}\left(\alpha \geq t\right)} < \infty. \qquad \square
$$

In addition,

Claim 4.0.2. For each $t \in [0, \infty)$,

$$
\mathbb{E}\left[\mathcal{Y}_w(t)\right] = \int_0^t \mathbb{E}\left[r_w(s)\right] \mathrm{d}s. \tag{4.6}
$$

Proof. Note that by Fubini-Tonelli,

$$
\int_0^t \mathbb{E}\left[r_w(s)\right] \mathrm{d}s = \mathbb{E}\left[\int_0^t r_w(s) \mathrm{d}s\right].
$$

Note also that by Claim 4.0.1, we have $\tau_n \uparrow \infty$ almost surely as $n \uparrow \infty$. Thus, by re-arranging and applying the monotone convergence theorem (taking limits as $n \to \infty$), it suffices to show that for each $n \in \mathbb{N}$ and any $t \in [0, \infty)$

$$
\mathbb{E}\left[\mathcal{Y}_w(t \wedge \tau_n) - \int_0^{t\wedge\tau_n} r_w(s) \mathrm{d}s\right] = 0.
$$

We re-write the left hand side as

$$\begin{aligned}
\mathbb{E}&\left[\mathcal{Y}_w(t\wedge\tau_n)-\int_0^{t\wedge\tau_n} r_w(s)\mathrm{d}s\right]= \\
&\qquad\mathbb{E}\left[\sum_{k=0}^{n-1}\left(\mathcal{Y}_w(t\wedge\tau_{k+1})-\mathcal{Y}_w(\tau_k)-\int_{\tau_k}^{t\wedge\tau_{k+1}} r_w(s)\mathrm{d}s\right)\mathbf{1}_{\{t\geq\tau_k\}}\right] \\
&\qquad=\mathbb{E}\left[\sum_{k=0}^{n-1}\left(\mathbf{1}_{\{t\geq\tau_{k+1}\}}-\int_{\tau_k}^{t\wedge\tau_{k+1}} r_w(s)\mathrm{d}s\right)\mathbf{1}_{\{t\geq\tau_k\}}\right] \\
&\qquad=\sum_{k=0}^{n-1}\mathbb{E}\left[\left(\mathbf{1}_{\{t\geq\tau_{k+1}\}}-\int_{\tau_k}^{t\wedge\tau_{k+1}} r_w(s)\mathrm{d}s\right)\mathbf{1}_{\{t\geq\tau_k\}}\right].
\end{aligned}$$

But for each k, with $\mathcal{F}_{\tau_k}$ denoting the stopped σ-algebra, recall that $\tau_{k+1}-\tau_k$ is distributed like $\mathrm{Exp}\,(S_k(w))$, and note that obviously the event $\{t\geq\tau_{k+1}\}=\{t-\tau_k\geq\tau_{k+1}-\tau_k\}$. Therefore, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{\{t\geq\tau_{k+1}\}}-\int_{\tau_k}^{t\wedge\tau_{k+1}} r_w(s)\mathrm{d}s|\mathcal{F}_{\tau_k}\right]\mathbf{1}_{\{t\geq\tau_k\}} \\
&=\mathbb{E}\left[\mathbf{1}_{\{t\geq\tau_{k+1}\}}-(\tau_{k+1}-\tau_k)S_k(w)\mathbf{1}_{\{t\geq\tau_{k+1}\}}-(t-\tau_k)S_k(w)\mathbf{1}_{\{t<\tau_{k+1}\}}|S_k(w),\tau_k\right]\mathbf{1}_{\{t\geq\tau_k\}} \\
&=\Big(1-e^{-S_k(w)(t-\tau_k)}-\Big(1-e^{-S_k(w)(t-\tau_k)}\left((t-\tau_k)S_w(k)+1\right)\Big) \\
&\qquad\qquad-(t-\tau_k)S_w(k)\left(e^{-S_k(w)(t-\tau_k)}\right)\Big)\mathbf{1}_{\{t\geq\tau_k\}} \\
&=0.
\end{aligned}$$

The result follows by applying the tower property. □

Claim 4.0.3. We have

$$\mathbb{E}\left[r_w(t)\right]=h(w)+\mathbb{E}\left[g(w,W)\right]\mathbb{E}\left[\mathcal{Y}_w(t)\right]=h(w)+\tilde{g}(w)\mathbb{E}\left[\mathcal{Y}_w(t)\right]. \tag{4.7}$$

Proof. First note that, since $r_w(t)$ jumps by $g(w,W)$ whenever $\mathcal{Y}_w(t)$ jumps, we have

$$\mathbb{E}\left[r_w(t)\right]-h(w)=\mathbb{E}\left[\sum_{i=1}^{\mathcal{Y}_w(t)} g(w,W_i)\right].$$

In addition, for each $n\in\mathbb{N}$,

$$\begin{aligned}
\mathbb{E}\left[g(w,W_n)\mathbf{1}_{\{\mathcal{Y}_w(t)\geq n\}}\right]&=\mathbb{E}\left[g(w,W_n)\right]-\mathbb{E}\left[g(w,W_n)\mathbf{1}_{\{\mathcal{Y}_w(t)<n\}}\right] \\
&=\mathbb{E}\left[g(w,W_n)\right]\left(1-\mathbb{P}\left(\mathcal{Y}_w(t)<n\right)\right)=\mathbb{E}\left[g(w,W_n)\right]\mathbb{P}\left(\mathcal{Y}_w(t)\geq n\right),
\end{aligned}$$

where the second to last equality follows from the fact that the event $\{\mathcal{Y}_w(t)<n\}$ depends only on $(S_i(w))_{i=0,\dots,n-1}$, and is thus independent of W_n. Finally, by Claim 4.0.1, $\mathbb{E}\left[Y_w(t)\right]<\infty$, and thus the result follows by applying Wald's Lemma. □

Proof of Lemma 1.2. First note that by (4.6) and (4.7), and continuity of $t\mapsto\mathbb{E}\left[\mathcal{Y}_w(t)\right]$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\mathcal{Y}_w(t)\right]=\tilde{g}(w)\mathbb{E}\left[\mathcal{Y}_w(t)\right]+h(w),$$

and solving this differential equation, with initial condition $\mathbb{E}\left[\mathcal{Y}_w(0)\right]=0$, we have

$$\mathbb{E}\left[\mathcal{Y}_w(t)\right]=\frac{h(w)}{\tilde{g}(w)}(e^{\tilde{g}(w)t}-1). \tag{4.8}$$

Now, let Λ be an exponentially distributed random variable with parameter λ. Then, on the one hand, by (4.5)

$$\mathbb{E}\left[\mathcal{Y}_w(\Lambda)\right] = \sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=0}^{k-1} \frac{S_i(w)}{S_i(w)+\lambda}\right].$$

On the other hand,

$$\mathbb{E}\left[\mathcal{Y}_w(\Lambda)\right] = \int_0^\infty \lambda e^{-\lambda u} \mathbb{E}\left[\mathcal{Y}_w(\Lambda) \middle| \Lambda = u\right] \mathrm{d}u = \int_0^\infty \lambda e^{-\lambda u} \mathbb{E}\left[\mathcal{Y}_w(u)\right] \mathrm{d}u \overset{(4.8)}{=} \frac{h(w)}{\lambda - \tilde{g}(w)}$$

where, in order to evaluate the integral to get the last equality, we have used the fact that $\lambda > \tilde{g}_+$. The result follows. □

References

[1] Krishna B. Athreya, Arka P. Ghosh, and Sunder Sethuraman, *Growth of preferential attachment random graphs via continuous-time branching processes*, Proc. Indian Acad. Sci. Math. Sci. **118** (2008), no. 3, 473–494. MR2450248

[2] Krishna B. Athreya and Samuel Karlin, *Limit theorems for the split times of branching processes*, J. Math. Mech. **17** (1967), 257–277. MR0216592

[3] Krishna B. Athreya and Samuel Karlin, *Embedding of urn schemes into continuous time Markov branching processes and related limit theorems*, Ann. Math. Statist. **39** (1968), 1801–1817. MR232455

[4] Albert-László Barabási and Réka Albert, *Emergence of scaling in random networks*, Science **286** (1999), no. 5439, 509–512. MR2091634

[5] Shankar Bhamidi, *Universal techniques to analyze preferential attachment trees: global and local analysis*, 2007, Preprint available at https://pdfs.semanticscholar.org/e7fb/8c999ff62a5f080e4c329a7a450f41fb1528.pdf.

[6] Ginestra Bianconi and Albert-László Barabási, *Bose-Einstein condensation in complex networks.*, Phys. Rev. Lett. **86 24** (2001), 5632–5.

[7] V. I. Bogachev, *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007. MR2267655

[8] Béla Bollobás and Oliver Riordan, *Shortest paths and load scaling in scale-free trees.*, Phys. Rev. E (3) **69** (2004), 036114.

[9] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády, *The degree sequence of a scale-free random graph process*, Random Structures Algorithms **18** (2001), no. 3, 279–290. MR1824277

[10] Christian Borgs, Jennifer Chayes, Constantinos Daskalakis, and Sebastien Roch, *First to market is not everything: an analysis of preferential attachment with fitness*, STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 135–144. MR2402437

[11] K. A. Borovkov and V. A. Vatutin, *On the asymptotic behaviour of random recursive trees in random environments*, Adv. in Appl. Probab. **38** (2006), no. 4, 1047–1070. MR2285693

[12] Wen-Chin Chen and Wen-Chun Ni, *Internal path length of the binary representation of heap-ordered trees*, Inform. Process. Lett. **51** (1994), no. 3, 129 – 132. MR1290206

[13] Steffen Dereich, *Preferential attachment with fitness: unfolding the condensate*, Electron. J. Probab. **21** (2016), Paper No. 3, 38. MR3485345

[14] Steffen Dereich, Cécile Mailler, and Peter Mörters, *Nonextensive condensation in reinforced branching processes*, Ann. Appl. Probab. **27** (2017), no. 4, 2539–2568. MR3693533

[15] Steffen Dereich and Peter Mörters, *Random networks with sublinear preferential attachment: degree evolutions*, Electron. J. Probab. **14** (2009), no. 43, 1222–1267. MR2511283

[16] Steffen Dereich and Marcel Ortgiese, *Robust analysis of preferential attachment models with fitness*, Combin. Probab. Comput. **23** (2014), no. 3, 386–411. MR3189418

[17] G. Ergün and G.J. Rodgers, *Growing random networks with fitness*, Phys. A. **303** (2002), no. 1, 261–272.

[18] Nikolaos Fountoulakis, Tejas Iyer, Cécile Mailler, and Henning Sulzbach, *Dynamical models for random simplicial complexes*, Ann. Appl. Probab., to appear.

[19] Nic Freeman and Jonathan Jordan, *Extensive condensation in a model of preferential attachment with fitness*, Electron. J. Probab. **25** (2020), Paper No. 68, 42 pp. MR4115737

[20] Alessandro Garavaglia, Remco van der Hofstad, and Gerhard Woeginger, *The dynamics of power laws: Fitness and aging in preferential attachment trees*, J. Stat. Phys. **168** (2017), 1137–1179. MR3691245

[21] John Haslegrave, Jonathan Jordan, and Mark Yarrow, *Condensation in preferential attachment models with location-based choice*, Random Structures Algorithms **56** (2020), no. 3, 775–795. MR4084189

[22] W.J. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 713–721. MR0144363

[23] Remco van der Hofstad, *Random graphs and complex networks volume 1*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2016. MR3617364

[24] Cecilia Holmgren and Svante Janson, *Fringe trees, Crump-Mode-Jagers branching processes and m-ary search trees*, Probab. Surv. **14** (2017), 53–154. MR3626585

[25] Tejas Iyer, *Degree distributions in recursive trees with fitnesses*, arxiv preprint arXiv:2005.02197, 2020.

[26] Svante Janson, *Functional limit theorems for multitype branching processes and generalized Pólya urns*, Stochastic Process. Appl. **110** (2004), no. 2, 177–245. MR2040966

[27] Jonathan Jordan, *Preferential attachment graphs with co-existing types of different fitnesses*, J. Appl. Probab. **55** (2018), no. 4, 1211–1227. MR3899937

[28] Jonathan Jordan and Andrew R. Wade, *Phase transitions for random geometric preferential attachment graphs*, Adv. in Appl. Probab. **47** (2015), no. 2, 565–588. MR3360390

[29] J. F. C. Kingman, *The first birth problem for an age-dependent branching process*, Ann. Probab. **3** (1975), no. 5, 790–801. MR0400438

[30] J. F. C. Kingman, *A simple model for the balance between selection and mutation*, J. Appl. Probability **15** (1978), no. 1, 1–12. MR465272

[31] Bas Lodewijks and Marcel Ortgiese, *The maximal degree in random recursive graphs with random weights*, arxiv preprint arXiv:2007.05438, 2020.

[32] Bas Lodewijks and Marcel Ortgiese, *A phase transition for preferential attachment models with additive fitness*, Electron. J. Probab. **25** (2020), Paper No. 146, 54. MR4193887

[33] Hosam M. Mahmoud, *Distances in random plane-oriented recursive trees*, J. Comput. Appl. Math. **41** (1992), no. 1-2, 237–245, Asymptotic methods in analysis and combinatorics. MR1181723

[34] Hosam M. Mahmoud, R. T. Smythe, and Jerzy Szymański, *On the structure of random plane-oriented recursive trees and their branches*, Random Structures Algorithms **4** (1993), no. 2, 151–176. MR1206674

[35] Cécile Mailler and Denis Villemonais, *Stochastic approximation on non-compact measure spaces and application to measure-valued Pólya processes*, Ann. Appl. Probab. **20** (2020), no. 5, 2393–2438. MR4149532

[36] C. McDiarmid, *Concentration*, Probabilistic Methods for Algorithmic Discrete Mathematics, Algorithms and Combinatorics, vol. 16, Springer, Berlin, Heidelberg, 1998, pp. 195–248. MR1678578

[37] T. F. Móri, *On random trees*, Studia Sci. Math. Hungar. **39** (2002), no. 1-2, 143–155. MR1909153

[38] Olle Nerman, *On the convergence of supercritical general (C-M-J) branching processes*, Z. Wahrsch. Verw. Gebiete **57** (1981), no. 3, 365–395. MR0629532

[39] Roberto Oliveira and Joel Spencer, *Connectivity transitions in networks with super-linear preferential attachment*, Internet Math. **2** (2005), no. 2, 121–163. MR2193157

[40] B. Pittel, *Note on the heights of random recursive trees and random m-ary search trees*, Random Structures Algorithms **5** (1994), no. 2, 337–347. MR1262983

[41] Helmut Prodinger and Friedrich J. Urbanek, *On monotone functions of tree structures*, Discrete Appl. Math. **5** (1983), no. 2, 223–239. MR683514

[42] Anna Rudas, Bálint Tóth, and Benedek Valkó, *Random trees and general branching processes*, Random Structures Algorithms **31** (2007), no. 2, 186–202. MR2343718

[43] Delphin Sénizergues, *Geometry of weighted recursive and affine preferential attachment trees*, Electron. J. Probab. **26** (2021), Paper No. 80, 56. MR4269210

[44] Herbert A. Simon, *On a class of skew distribution functions*, Biometrika **42** (1955), 425–440. MR73085

[45] Gábor Szabó, Mikko Alava, and János Kertész, *Shortest paths and load scaling in scale-free trees*, Phys. Rev. E (3) **66** (2002), 026101.

[46] Jerzy Szymański, *On a nonuniform random recursive tree*, Random graphs '85 (Poznań, 1985), North-Holland Math. Stud., vol. 144, North-Holland, Amsterdam, 1987, pp. 297–306. MR930497

[47] G. Udny Yule, *A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S.*, Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character **213** (1925), 21–87. MR0017488

Acknowledgments. We would like to thank two anonymous referees for their comments, which helped us improve the presentation of our results.