Applied Probability Trust (10 October 2023)

SUPPLEMENTARY MATERIAL: WEAK LOCAL LIMIT OF PREFERENTIAL ATTACHMENT RANDOM TREES WITH ADDITIVE FITNESS

TIFFANY Y. Y. LO,* *Uppsala University*

Abstract

This is intended as a complementary article to the published paper [25], where in addition to the content of [25], we collect the results and detailed proofs that are omitted from [25].

We study the local weak limit of the linear preferential attachment trees with additive fitness, where fitness is the random initial vertex attractiveness. We show that when the fitness are i.i.d. and have positive bounded support, the weak local limit can be constructed using a sequence of mixed Poisson point processes. We also provide a rate of convergence for the total variation distance between the r-neighbourhoods of the uniformly chosen vertex in the preferential attachment tree and the root vertex of the weak local limit. The proof uses a Pólya urn representation of the model, for which we give new estimates to the beta and product beta variables in its construction. As applications, we obtain limiting results and convergence rates for the degrees of the uniformly chosen vertex and its ancestors, where the latter are the vertices that are on the path between the uniformly chosen vertex and the initial vertex.

Keywords: graph limit; complex networks; distributional approximation

2020 Mathematics Subject Classification: Primary 90B15

Secondary 05C80, 60F05

1. Introduction

There has been considerable interest in studying the preferential attachment (PA) random graphs since [2] used them to explain the observed power-law degree distribution in some real networks such as the World Wide Web. The primary feature of

Submitted 21 March 2021; accepted 19 September 2023.

* Postal address: Department of Mathematics, Uppsala University, Lägerhyddsvägen 1, 752 37 Uppsala, Sweden. Email: yin_yuan.lo@math.uu.se

the stochastic mechanism consists of adding vertices sequentially over time with some number of edges attached to them, and then connecting these edges to the existing graph in such a way that vertices with higher degrees are more likely to receive them. A general overview of PA random graphs can be found in the books [39, 40].

In the basic models, vertices are born with the same constant 'weight' as their initial vertex attractiveness. To relax this assumption, [15] introduced a class of PA graphs with additive fitness (referred to as Model A in [15]), where *fitness* is defined as the *random* initial attractiveness. When the fitness are i.i.d., this family is the subject of recent works such as [21] and [26], whose results we discuss in Section 1.3.2. In this paper, we study the weak local limit of this family, and provide a rate of convergence for the total variation distance between the local neighbourhoods of the PA tree and its weak local limit. The result extends that of [4], which considered the PA graphs with constant initial attractiveness. As applications, we obtain limiting results for the degree distributions of the uniformly chosen vertex and its ancestors. where rates of convergence are also provided. Another objective of this article is to present the arguments of [4] in more detail, which is the main reason why we consider the PA tree instead of the case where multiple edges are possible.

Before defining the model, note that we view the edges as directed, where a newly-born vertex always sends a single outgoing edge to an existing vertex in the graph. We define the *weight* of a vertex as its in-degree plus its fitness; each time a vertex receives an edge from another vertex, its weight increases by one. As we mostly work conditionally on the fitness sequence, we also introduce a conditional version of the model. Note that the seed graph is a single vertex that has a non-random fitness, but as we explained in Section 1.3.1 below, the choices of the seed graph and its fitness have no effect on the weak local limit. This is chosen purely to streamline the argument.

Definition 1. ($(\mathbf{x}, n)$*-sequential model and PA tree with additive fitness.*) Given a positive integer n and a sequence $\mathbf{x} := (x_i, i \geqslant 1)$, with $x_1 > -1$ and $x_i > 0$ for $i \geqslant 2$, we construct a sequence of random trees $(G_i, 1 \leqslant i \leqslant n)$ as follows. The seed graph G_1 consists of vertex 1 with the initial attractiveness x_1 and has degree 0. The graph G_2 is constructed by joining vertex 2 and 1 with an edge, and equipping vertex 2 with the initial attractiveness x_2. For $3 \leqslant m \leqslant n$, G_m is constructed from G_{m-1} by attaching one edge between vertex m and $k \in [m-1]$, and the edge is directed towards vertex k with probability

$$\frac{D^{(in)}_{m-1,k} + x_k}{m - 2 + \sum_{j=1}^{m-1} x_j} \quad \text{for } 1 \leqslant k \leqslant m-1,$$

where $D^{(in)}_{m,k}$ is the in-degree of vertex k in G_m, and $D^{(in)}_{m,k} = 0$ whenever $k \geqslant m$. The m-th attachment step is completed by assigning vertex m the initial attractiveness x_m. We call the resulting graph G_n an $(\mathbf{x}, n)$-*sequential model*, and its law is denoted by $\mathrm{Seq}(\mathbf{x})_n$. Taking $X_1 = x_1$, the distribution $\mathrm{PA}(\pi, X_1)_n$ of the PA tree with additive fitness follows from mixing $\mathrm{Seq}(\mathbf{x})_n$ with the distribution of the i.i.d. fitness sequence $(X_i, i \geqslant 2)$.

1.1. Local weak convergence

The concept of local weak convergence was independently introduced by [3] and [1]. Here, we follow [40, Section 2.3 and 2.4], and also [3]. Informally, this involves exploring some random graph G_n from vertex o_n, chosen uniformly at random from G_n, and studying the distributional limit of the neighbourhoods of radius r rooted at o_n for each $r < \infty$.

We begin with a few definitions. A rooted graph is a pair (G, o), where $G = (V(G), E(G))$ is a graph with vertex set $V(G)$ and edge set $E(G)$, and $o \in V(G)$ is the designated root in G. Next, let r be a finite, positive integer. For any (G, o), denote by $B_r(G, o)$ the rooted neighbourhood of radius r around o. More formally, $B_r(G, o) = (V(B_r(G, o))), E(B_r(G, o)))$, where

$$V(B_r(G, o)) = \{u \in V(G) : \text{the distance between } u \text{ and } o \text{ is at most } r \text{ edges}\};$$
$$E(B_r(G, o)) = \{\{u, v\} \in E(G) : u, v \in V(B_r(G, o))\}.$$

We refer to $B_r(G, o)$ as the r-*neighbourhood* of vertex o, or simply as the *local neighbourhood* of o when the reference to r is not needed. Finally, two rooted graphs (G, o) and (H, o') are isomorphic, denoted $(G, o) \cong (H, o')$, if there is a bijection $\psi : V(G) \to V(H)$ such that $\psi(o) = o'$ and $\{u, v\} \in E(G)$ if and only if $\{\psi(u), \psi(v)\} \in E(H)$.

Below we define the local weak convergence of a sequence of finite, random graphs $(G_n, n \geqslant 1)$ using a criterion given in [40, Theorem 2.14].

Definition 2. (*Local weak convergence.*) Let $(G_n, n \geqslant 1)$ be a sequence of finite random graphs. The local weak limit of G_n is (G, o) when for all finite rooted graphs (H, v) and all finite r,

$$\frac{1}{n}\sum_{j=1}^{n}\mathbb{P}[(B_r(G_n, j), j) \cong (H, v)] \overset{n\to\infty}{\longrightarrow} \mathbb{P}[(B_r(G, o), o) \cong (H, v)]. \tag{1.1}$$

The left-hand side of (1.1) is the probability that the r-neighbourhood of a randomly chosen vertex is isomorphic to (H, v). Note that the convergence in (1.1) is equivalent to the convergence of the expectations of all bounded and continuous functions with respect to an appropriate metric on rooted graphs; see e.g. [40, Chapter 2].

1.2. Statement of the main results

In the main result, we fix $X_1 > -1$ and assume that $(X_i, i \geqslant 2)$ are i.i.d. positive bounded variables with distribution π. We write $\mu := \mathbb{E}X_2 < \infty$ and

$$\chi := \frac{\mu}{\mu + 1}. \tag{1.2}$$

We first define the local weak limit of the PA tree with additive fitness, which is an infinite rooted random tree that generalises the Pólya point tree introduced in [4]. Hence, we refer to it simply as a π-Pólya point tree, with π being the fitness distribution of the PA tree. Denote this random tree by $(\mathcal{T}, 0)$, so that 0 is its root. We begin by explaining the Ulam-Harris labelling of trees that we use in the construction of $(\mathcal{T}, 0)$. Starting from the root 0, the *children* of any vertex $\bar{v}$ (if any) are generated recursively as $(\bar{v}, j)$, $j \in \mathbb{N}$, and we say that $\bar{v}$ is the *parent* of $(\bar{v}, j)$. With the convention $B_0(\mathcal{T}, 0) = \{0\}$, note that if $\bar{v} := (0, v_1, ..., v_r)$, then $(\bar{v}, i) \in \partial B_{r+1} := V(B_{r+1}(\mathcal{T}, 0)) \setminus V(B_r(\mathcal{T}, 0))$.

Furthermore, each vertex $\bar{v} \in V((\mathcal{T}, 0))$ has a fitness $X_{\bar{v}}$ and a random *age* $a_{\bar{v}}$, where $0 < a_{\bar{v}} \leqslant 1$. We write $a_{\bar{v},i} := a_{(\bar{v},i)}$ for convenience. Apart from the root vertex 0, there are two types of vertices, namely, type L (for left) and R (for right). Vertex $\bar{v}$ belongs to type L if $a_{\bar{v},i} < a_{\bar{v}}$ for some $i \geqslant 1$; and $\bar{v}$ belongs to type R if $a_{\bar{v},i} \geqslant a_{\bar{v}}$ for all $i \geqslant 1$. There is exactly one type L vertex in ∂B_r for all $r \geqslant 1$, and the labels

$(0, 1, 1, ..., 1)$ are designated to these vertices. For any vertex $\bar{v}$, define

$$R_{\bar{v}} = \text{the number of type R children of } \bar{v} \text{ in the } (\mathcal{T}, 0). \tag{1.3}$$

We also label the type R vertices in increasing order of their ages, so that if $\bar{v}$ is the root or belongs to type L, then $a_{\bar{v},1} \leqslant a_{\bar{v}} \leqslant a_{\bar{v},2} \leqslant ... \leqslant a_{\bar{v},1+R_{\bar{v}}}$; and if $\bar{v}$ belongs to type R, then $a_{\bar{v}} \leqslant a_{\bar{v},1} \leqslant ... \leqslant a_{\bar{v},R_{\bar{v}}}$. See Figure 1 below for an illustration of $(\mathcal{T}, 0)$ and the vertex ages.

To understand how the vertex types and the ages above arise in the weak limit, consider the r-neighbourhood of a uniformly chosen vertex k_0 in $G_n \sim \text{PA}(\pi, X_1)_n$. Observe that there is a unique path from k_0 to the initial vertex, unless k_0 is the initial vertex. Apart from k_0, the vertices in the r-neighbourhood that belong to this path are called type L vertices, and the remaining vertices are referred to as type R vertices. The ages in $(\mathcal{T}, 0)$ encode the (rescaled) arrival times of the vertices in the r-neighbourhood of vertex k_0, which determine their degree distributions. A comparison of the 2-neighbourhoods of the PA tree and $(\mathcal{T}, 0)$ is given in Figure 1.

Definition 3. (*π-Pólya point tree.*) A π-Pólya point tree $(\mathcal{T}, 0)$ is defined recursively as follows. The root 0 has an age $a_0 = U_0^{\chi}$, where $U_0 \sim \text{U}[0, 1]$. Assuming that $\bar{v} \in \partial B_r$ and $a_{\bar{v}}$ have been generated, we define its children $(\bar{v}, j) \in \partial B_{r+1}$ for $j = 1, 2, ...$ as follows. Independently of all random variables generated before, let $X_{\bar{v}} \sim \pi$ and

$$Z_{\bar{v}} \sim \begin{cases} \text{Gamma}(X_{\bar{v}}, 1), & \text{if } \bar{v} \text{ is the root or of type R;} \\ \text{Gamma}(X_{\bar{v}} + 1, 1), & \text{if } \bar{v} \text{ is of type L.} \end{cases}$$

If $\bar{v}$ is the root or of type L, let $a_{\bar{v},1} | a_{\bar{v}} \sim \text{U}[0, a_{\bar{v}}]$; and $(a_{\bar{v},i}, 2 \leqslant i \leqslant 1 + R_{\bar{v}})$ be the points of a mixed Poisson point process on $(a_{\bar{v}}, 1]$ with intensity

$$\lambda_{\bar{v}}(y)dy := \frac{Z_{\bar{v}}}{\mu a_{\bar{v}}^{1/\mu}} y^{1/\mu - 1} dy.$$

If $\bar{v}$ is of type R, then $(a_{\bar{v},i}, 1 \leqslant i \leqslant R_{\bar{v}})$ are sampled as the points of a mixed Poisson process on $(a_{\bar{v}}, 1]$ with intensity $\lambda_{\bar{v}}$. We obtain $(\mathcal{T}, 0)$ by continuing this process ad infinitum.

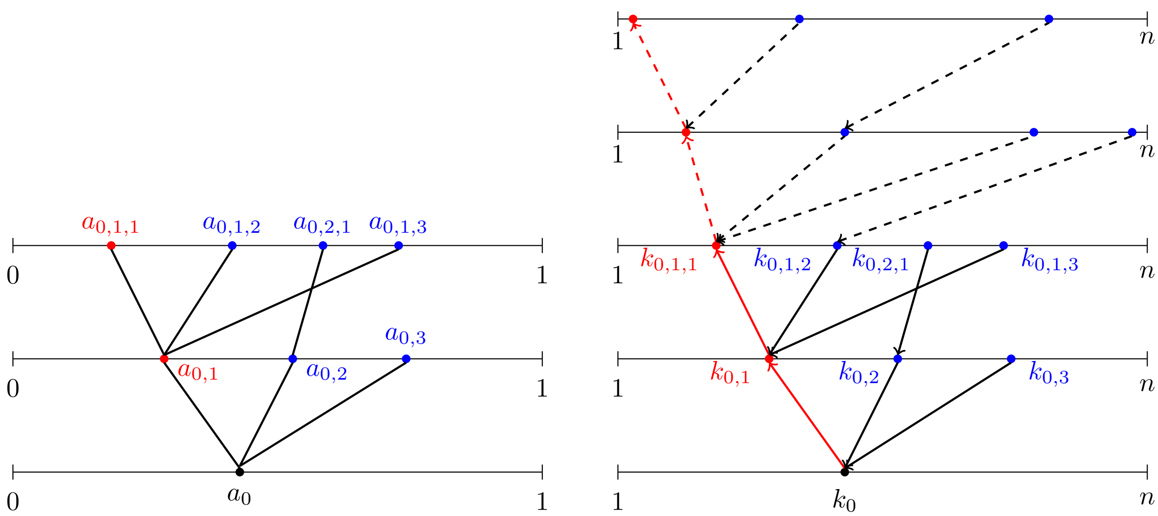

FIGURE 1: A comparison between the 2-neighbourhoods in $(\mathcal{T}, 0)$ *(left)* and the PA tree G_n rooted at the uniformly chosen vertex k_0 *(right)*. We assign Ulam-Harris labels as subscripts $(k_{\bar{v}})$ to the vertices in (G_n, k_0) to better compare the 2-neighbourhoods. In both figures, the vertex location corresponds to either its arrival time or its age. A vertex is coloured red (blue) if it is a type L (R) child. On the left, $R_0 = 2$, $R_{0,1} = 2$, $R_{0,2} = 1$ and $R_{0,3} = 0$. On the right, the red path starting from k_0 leads to the initial vertex. The 2-neighbourhoods are coupled such that they are isomorphic, and the vertex ages and rescaled arrival times are close to each other. The dashed edges and the unlabelled vertices are not coupled.

Remark 1. When exploring the r-neighbourhood of the uniformly chosen vertex in the PA tree, the type L vertices are uncovered via the incoming edge it received from the probed uniformly chosen vertex or type L vertices. To account for the size-biasing effect of these edges, the type L gamma variables in the π-Pólya point tree thus have a unit increment in the shape parameter.

We define the total variation distance between two probability distributions ν_1 and ν_2 as

$$d_{\mathrm{TV}}(\nu_1, \nu_2) = \inf\{\mathbb{P}[V \neq W] : (V, W) \text{ is a coupling of } \nu_1 \text{ and } \nu_2\}. \qquad (1.4)$$

When proving the local weak convergence, we couple the random elements $(B_r(G_n, k_0), k_0)$ and $(B_r(\mathcal{T}, 0), 0)$ in the space of rooted graphs (modulo isomorphisms) $\mathcal{G}$ such that they are isomorphic w.h.p., thus bounding their total variation distance. In the main theorem below, we emphasise that the convergence does not take into account the ages and fitness of the π-Pólya point tree, but they are essential for the graph construction and are used for the couplings later. The results on the asymptotic degree distributions and the respective convergence rates are deferred to Section 8.

Theorem 1. *Suppose that the fitness distribution π is supported on $(0, \kappa]$ for some $\kappa < \infty$. Let $G_n \sim \mathrm{PA}(\pi, X_1)_n$, k_0 be a uniformly chosen vertex in G_n and $(\mathcal{T}, 0)$ be the π-Pólya point tree. Then, given $r < \infty$, there is a positive constant $C := C(X_1, \mu, r, \kappa)$ such that*

$$d_{\mathrm{TV}}(\mathscr{L}((B_r(G_n, k_0), k_0)), \mathscr{L}((B_r(\mathcal{T}, 0), 0)) \leqslant C(\log\log n)^{-\chi} \tag{1.5}$$

for all $n \geqslant 3$. This implies that the local weak limit of G_n is the π-Pólya point tree.

Next, we state the limiting results for some degree statistics of the PA tree. The connection of the following results to [4, 7, 40, 26, 31] are discussed in detail later in Section 8, where we also give the probability mass functions of the limiting distributions. Recall that $D^{(in)}_{n,j}$ is the in-degree of vertex j in $G_n \sim \mathrm{PA}(\pi, X_1)_n$. Define the degree of vertex j in G_n as

$$D_{n,j} := D^{(in)}_{n,j} + 1, \quad \text{with } D^0_n := D_{n,k_0}, \tag{1.6}$$

so that D^0_n is the degree of the uniformly chosen vertex. Let $L(i)$ be the Ulam-Harris labels $(0, 1, 1, \dots, 1)$ such that $|(0, 1, \dots, 1)| = i + 1$, so that $k_{L[i]}$ is the type L vertex that is exactly i edges away from k_0. Type L vertices are commonly known as the ancestors of k_0 in the fringe tree analysis (see e.g. [20]), where their degrees are often of particular interest. Fixing $r \in \mathbb{N}$, let D^0_n be as in (1.6) and

$$D^i_n = D_{n,k_{L[i]}} \quad \text{if } k_{L[i]} \neq 1 \text{ for all } 1 \leqslant i \leqslant r;$$

and if $k_{L[i]} = 1$ for some $i \leqslant r$, let

$$D^j_n = D_{n,k_{L[j]}}, \quad \text{for } 1 \leqslant j < i \quad \text{and} \quad D^j_n = -1 \quad \text{for } i \leqslant j \leqslant r.$$

As the probability that we see vertex 1 in the r-neighbourhood $B_r(G_n, k_0)$ tends to zero as $n \to \infty$, $(D_1^0, ..., D_n^r)$ can be understood as the joint degree sequence of $(k_0, k_{L[1]}, \dots, k_{L[r]})$. The next result concerning this joint degree sequence follows directly from Theorem 1.

Corollary 1. *Retaining the notation above, assume that π is supported on $(0, \kappa]$ for some $\kappa < \infty$. Define $U_0 \sim \mathrm{U}[0,1]$, $a_{L[0]} = U_0^{\chi}$, and given $a_{L[i-1]}$, let $a_{L[i]} \sim \mathrm{U}[0, a_{L[i-1]}]$ for $1 \leqslant i \leqslant r$. Independently from $(a_{L[i]}, 0 \leqslant i \leqslant r)$, let $X_{L[i]}$ be i.i.d. random variables with distribution π, and*

$$Z_{L[i]} \sim \begin{cases} \mathrm{Gamma}(X_{L[i]}, 1), & \text{if } i = 0, \\ \mathrm{Gamma}(X_{L[i]} + 1, 1) & \text{if } 1 \leqslant i \leqslant r. \end{cases}$$

Writing $L[0] = 0$, let $R_{L[i]}$ be conditionally independent variables with $R_{L[i]} \sim \mathrm{Po}\big(Z_{L[i]}(a_{L[i]}^{-1/\mu} - 1)\big)$. Define $\overline{R}^{(r)} := (R_0 + 1, R_{L[1]} + 2, ..., R_{L[r]} + 2)$ and $\overline{D}_n^{(r)} := (D_n^0, D_n^1, ..., D_n^r)$. There is a positive constant $C := C(X_1, \mu, r, \kappa)$ such that

$$d_{\mathrm{TV}}\left(\mathscr{L}(\overline{D}_n^{(r)}), \mathscr{L}(\overline{R}^{(r)})\right) \leqslant C(\log\log n)^{-\chi} \quad \text{for all } n \geqslant 3.$$

We now state a convergence result for D_n^0 in (1.6). In view of Definition 2, the limiting distribution of D_n^0 and the convergence rate can be read from Theorem 1. However, the theorem below holds without the assumption of bounded fitness and has a much sharper rate. The improvement in the rate can be understood as a consequence that k_0 only needs to be large enough so that w.h.p. it has a small degree, in contrast to having a small enough r-neighbourhood for all $r < \infty$, as required when proving Theorem 1.

Theorem 2. *Assume that the p-th moment of the distribution π is finite for some $p > 4$. Let $R_0 \sim \mathrm{Po}(Z_0(a_0^{-1/\mu} - 1))$, where given $X_0 \sim \pi$, $Z_0 \sim \mathrm{Gamma}(X_0, 1)$, and independently of Z_0, $U_0 \sim \mathrm{U}[0,1]$ and $a_0 := U_0^{\chi}$. Writing $\xi_0 = R_0 + 1$, there are positive constants $C := C(X_1, \mu, p)$ and $0 < d < \chi(1/4 - 1/(2p))$ such that*

$$d_{\mathrm{TV}}\left(\mathscr{L}(D_n^0), \mathscr{L}(\xi_0)\right) \leqslant Cn^{-d} \quad \text{for all } n \geqslant 1. \tag{1.7}$$

1.3. Possible extensions and related works

Below we discuss the possible ways of extending Theorem 1, but we refrain from pursuing these directions in the interest of article length. We also give an overview on the recent development in the PA graphs with additive fitness, and collect some results on the local weak convergence of related PA models.

1.3.1. *Possible extensions* The convergence rate in (1.5) roughly follows from $k_0 \geqslant n(\log\log n)^{-1}$ w.p. at least $(\log\log n)^{-1}$, and on this event, we can couple the two graphs such that the probability that $(B_r(G_n, k_0), k_0) \cong (B_r(\mathcal{T}, 0), 0)$ tends to one as $n \to \infty$. It is likely possible to improve the rate by optimising this and similar choices of thresholds, as well as making the dependence on the radius r explicit by carefully keeping track of the coupling errors, but with much added technicality.

When $X_i = 1$ a.s. for all $i \geqslant 2$, [10, Theorem 1] established that the choice of the seed graph has no effect on the local weak limit. This is because w.h.p., the local neighbourhood does not contain any of the seed vertices. By simply replacing the seed graph in the proof, Theorem 1 can be shown to hold for more general seed graphs. With some straightforward modifications to the proof, we can also show that when the fitness is bounded, the π-Pólya point tree is the weak local limit of the PA trees with self-loops, and when each vertex in the PA model sends $m \geqslant 2$ outgoing edges, the limit is a variation of the π-Pólya point tree; see e.g. [4] for the non-random unit fitness case.

By adapting the argument of the recent paper [17], it is possible to show that the weak convergence in Theorem 1 holds for fitness distributions with finite p-th moment for some $p > 1$. The convergence rate should be valid for fitness distributions with at least exponentially decaying tails, but the assumption of bounded fitness greatly simplifies the proof. The i.i.d. assumption is only needed so that the fitness variables that we see in the local neighbourhood are i.i.d., and for applying the standard moment inequality in Lemma 15. Hence, we believe the theorem to at least hold for a fitness sequence $\mathbf{X}$ such that (1) X_i has the same marginal distribution π for all $i \geqslant 2$; and (2) for some $m \geqslant 1$, the variables in the collection $(X_i, i \in A)$ are independent for any A such that $\{i, j \in A : |i - j| > 2m\}$. If in the limit, the vertex labels in the local neighbourhood are at least $2m$ apart from each other w.h.p., then (1) and (2) ensure that these vertices have i.i.d. fitness; while (2) alone may be sufficient for proving a suitable analogue of the standard moment inequality Lemma 15.

The weak convergence in Theorem 1 should hold *in probability*, meaning that as $n \to \infty$, $\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}[(B_r(G_n,j),j) \cong (H,v)] \to \mathbb{P}[(B_r(\mathcal{T},0),0) \cong (H,v)]$. This convergence is valid in the case where the initial attractiveness are equal a.s. [40, Chapter 5], and for certain PA models where each newly added vertex sends a random number of outgoing edges [17]. For the PA tree with additive fitness, this could be proved by adapting the 'second moment' method used in [17, 40], which involves establishing that the r-neighbourhoods of two independently, uniformly chosen vertices in the PA tree are disjoint w.h.p.

1.3.2. *Related works* The fitness variables were assumed to be i.i.d. in [26], [21] and [5]. In [26], martingale techniques were used to investigate the maximum degree for fitness distributions with different tail behaviours, where the results are applicable to PA graphs with additive fitness that allow for multiple edges. The authors also studied the empirical degree distribution (e.d.d.), whose detail we defer to Remark 4. The model is a special case of the PA tree considered in [21], where vertices are chosen with probability proportional to a suitable function of their fitness and degrees at each attachment step. Using Crump-Mode-Jagers (CMJ) branching processes, [21] studied the e.d.d. and the condensation phenomenon. The same method was applied in [5] to investigate the e.d.d., the height and the degree of the initial vertex, assuming that the fitness are bounded. Note that these articles focused on 'global' results, which cannot be deduced from the local weak limit.

As observed in [38], the model is closely related to the weighted random recursive trees introduced in [9]. Due to the Pólya urn representation (Theorem 3 below), the PA tree with additive fitness can be viewed as a special case of this model class, where the weights of the vertices are distributed as $(S^{(X)}_{n,j} - S^{(X)}_{n,j-1}, 2 \leqslant j \leqslant n)$ with $S^{(X)}_{n,0} = 0$ and $S^{(X)}_{n,n} = 1$, and $B^{(X)}_j$ and $S^{(X)}_{n,j}$ being the variables B_j and $S_{n,j}$ in (1.8) and (1.9) mixed over the fitness sequence $\mathbf{X}$. For the weighted random recursive trees, [9] studied the average degree of a fixed vertex and the distance between of a newly added vertex and the initial vertex. The joint degree sequence of fixed vertices, the height and the profile were investigated in [38], and a more refined result on the height was given in

[30]. When $\mathbf{X}$ is a deterministic sequence satisfying a certain growth condition, [38] and [30] showed that their results are applicable to the corresponding PA tree with additive fitness.

We now survey the local weak limit results developed for PA graphs. When $x_1 = 0$ and $x_i = 1$ for all $i \geqslant 2$, the $(\mathbf{x}, n)$-sequential model is the pure 'sequential' model in [4] with no multiple edges; and a special case of the model considered in [37], where the 'weight' function there is the identity function plus one. Using the CMJ embedding method, [37] studied the asymptotic distribution of the subtree rooted at a uniformly chosen vertex, which implies the local weak convergence of the PA family considered in their work. The urn representation of PA models was used in [4], [40, Chapter 5], [16, Chapter 4] and [17] to study local weak limits. In particular, they showed that the weak limit of several PA models with non-random fitness and possibly random out-degrees is a variant of the Pólya point tree [4]. Again using the CMJ method, it can be shown that the local weak convergence result of [18] for a certain 'continuous time branching process tree' implies that the PA tree with additive fitness converges in the *directed* local weak sense.

Finally, a different PA model was considered in [5, 8, 11, 12], where the probability that a new vertex attaches to an existing vertex is proportional to its fitness times its degree. Currently, there are no results concerning the local weak convergence of this model class.

1.4. Proof overview

1.4.1. *Pólya urn representation of the* $(\mathbf{x}, n)$*-sequential model* In the proof of Theorem 1, the key ingredient is an alternative definition of the $(\mathbf{x}, n)$-sequential model in Definition 1, which relies on the fact that the dynamics of PA graphs can be represented as embedded classical Pólya urns. In a classical Pólya urn initially with a red balls and b black balls, a ball is chosen randomly from the urn, and is returned to the urn along with a new ball of the same colour. It is well-known that the a.s. limit of the proportion of red balls has the $\mathrm{Beta}(a, b)$ distribution; see e.g. [28]. Furthermore, by de Finetti's theorem (see e.g. [28, Theorem 1.2, p. 29]), conditional on $\beta \sim \mathrm{Beta}(a, b)$, the indicators that a red ball is chosen at each step are distributed as independent Bernoulli variables with parameter β. In the PA mechanism, an existing vertex i can

FIGURE 2: An example of the $(\mathbf{x}, n)$-Pólya urn tree for $n = 5$, where $U_i \sim \mathrm{U}[0, S_{n,i-1}]$ for $i = 2, ..., 5$ and an outgoing edge is drawn from vertices i to j if $U_i \in [S_{n,j-1}, S_{n,j})$.

be represented by some colour i in an urn, and its weight is given by the total weight of the balls in the urn. At each urn step, we choose a ball w.p. proportional to its weight, and if vertex i is chosen at some step $j > i$, we return the chosen ball, an extra ball of colour i with weight 1, plus a ball of new colour j with weight x_j to the urn. Classical Pólya urns are naturally embedded in this multi-colour urn; see [33]. The attachment steps of the graph can therefore be generated independently when conditioned on the associated beta variables.

Definition 4. ($(\mathbf{x}, n)$-*Pólya urn tree.*) Given $\mathbf{x}$ and n, let $T_j := \sum_{i=1}^{j} x_i$, and $(B_j, 1 \leqslant j \leqslant n)$ be independent random variables such that $B_1 := 1$ and

$$B_j \sim \mathrm{Beta}(x_j, j - 1 + T_{j-1}) \quad \text{for } 2 \leqslant j \leqslant n. \tag{1.8}$$

Moreover, let $S_{n,0} := 0$, $S_{n,n} := 1$ and

$$S_{n,j} := \prod_{i=j+1}^{n} (1 - B_i) \quad \text{for } 1 \leqslant j \leqslant n - 1. \tag{1.9}$$

We connect n vertices with labels $[n] := \{1, \dots, n\}$ as follows. Let $I_j = [S_{n,j-1}, S_{n,j})$ for $1 \leqslant j \leqslant n$. Conditionally on $(S_{n,j}, 1 \leqslant j \leqslant n - 1)$, let $(U_j, 2 \leqslant j \leqslant n)$ be independent variables such that $U_j \sim \mathrm{U}[0, S_{n,j-1}]$. If $j < k$ and $U_k \in I_j$, we attach an outgoing edge from vertex k to vertex j. We say that the resulting graph is an $(\mathbf{x}, n)$-Pólya urn tree and denote its law $\mathrm{PU}(\mathbf{x})_n$.

An example of the $(\mathbf{x}, n)$-Pólya urn tree is given in Figure 2. Note that $1 - B_j$ in Definition 4 is β_{j-1} in [38]. As we only work with B_j and $S_{n,j}$ by fixing the sequence $\mathbf{x}$, we omit $\mathbf{x}$ from their notation throughout this article. The $(\mathbf{x}, n)$-Pólya urn tree is related to the $(\mathbf{x}, n)$-sequential model via the following result of [38].

Theorem 3. (Theorem 1.1, [38].) *Let G_n be an $(\mathbf{x}, n)$-Pólya urn tree, then G_n has the same law as the $(\mathbf{x}, n)$-sequential model, that is,* $\mathrm{PU}(\mathbf{x})_n \overset{d}{=} \mathrm{Seq}(\mathbf{x})_n$.

A proof of Theorem 3 is given in Section 10.1, as the argument is needed to prove a variation of the result that we use later. From now on, we work with the $(\mathbf{x}, n)$-Pólya urn tree in place of the $(\mathbf{x}, n)$-sequential model.

1.4.2. *Coupling for the non-random fitness case* When proving Theorem 1, we couple the (randomised) urn tree and the π-Pólya point tree such that for all positive integers r, the probability that their r-neighbourhoods are not isomorphic is of order at most $(\log\log n)^{-\chi}$. To give a brief overview to the coupling, below we suppose that $X_i = 1$ a.s. for $i \geqslant 2$ and only consider the $r = 1$ case. We couple the children of the uniformly chosen vertex in the urn tree G_n and the root in the π-Pólya point tree $(\mathcal{T}, 0)$ such that w.h.p., they have the same number of children, and the ages and the rescaled arrival times of these children are close enough. Note that for non-random fitness, the urn tree is simply an alternative definition of the PA tree with additive fitness. For G_n, we use the terms 'age' and 'rescaled arrival times' interchangeably.

I. The ages of the roots. As mentioned before, the ages in $(\mathcal{T}, 0)$ encode the rescaled arrival times in G_n. Since for any vertex in either G_n or $(\mathcal{T}, 0)$, the number of its children and the ages of its children depend heavily on its own age, we need to couple the uniformly chosen vertex k_0 in G_n and the root of $(\mathcal{T}, 0)$ such that their ages are close enough.

II. The type R children and a Bernoulli-Poisson coupling. Using the urn representation in Definition 4, the ages and the number of the type R children of vertex k_0 can be encoded in a sequence of conditionally independent Bernoulli variables, where the success probabilities are given in terms of the variables B_i and $S_{n,i}$ in (1.8) and (1.9). See Figure 3 for an illustration. To couple this Bernoulli sequence to a suitable discretisation of the mixed Poisson point process in Definition 3, we use the beta-gamma algebra and the law of large numbers to approximate B_i and $S_{n,i}$ for large enough i. Once we use these estimates to swap the success probabilities with simpler

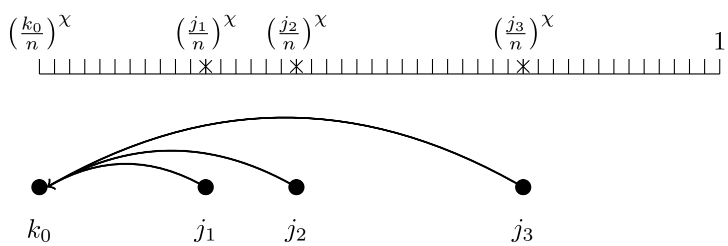

FIGURE 3: An illustration of the relation between the Bernoulli sequence on $((k_0/n)^\chi, 1]$ constructed using $(\mathbb{1}[U_j \in I_{k_0}], k_0+1 \leqslant j \leqslant n)$ and the $(\mathbf{x}, n)$-Pólya urn tree, where k_0 is the uniformly chosen vertex, U_j and I_{k_0} are as in Definition 4. We put a point on $(j/n)^\chi$ if vertex k_0 receives the outgoing edge from vertex j. Here the type R children of k_0 are j_1, j_2 and j_3. The rescaled arrival times $((j/n)^\chi, k_0+1 \leqslant j \leqslant n)$ are later used to discretise the mixed Poisson process in the coupling step.

quantities, we can apply the standard Bernoulli-Poisson coupling.

III. The ages of the type L children. It is clear that the numbers of the type L children of vertex k_0 and the root of $(\mathcal{T}, 0)$ only differ when $k_0 = 1$, which occurs w.p. n^{-1}. To couple their ages such that they are close enough, we use the estimates for B_i and $S_{n,i}$ to approximate the distribution of the type L child of vertex k_0. This completes the coupling of the 1-neighbourhoods.

We reiterate that although the closeness of the ages are not part of the local weak convergence, we need to closely couple the ages of the children of vertex k_0 and the root in $(\mathcal{T}, 0)$, as otherwise we cannot couple the 2-neighbourhoods when we prove the theorem for all finite radii.

1.4.3. *Random fitness and the general local neighbourhood* Here we summarise the additional ingredients needed for handling the random fitness and for coupling the neighbourhoods of any finite radius.

When $\mathbf{x}$ is a realisation of the fitness sequence $\mathbf{X}$, B_i and $S_{n,i}$ can be approximated in the same way as in the non-random fitness case if w.h.p., the sums $\sum_{h=2}^{i} X_h$ are close to $(i-1)\mu$ for all i greater than some suitably chosen function $\phi(n)$. As $\mathbf{X}$ is i.i.d., we can apply standard moment inequalities to show that this event occurs w.h.p. The remaining parts of the coupling are similar to the above.

When coupling general r-neighbourhoods, we use an induction over the neighbourhood radius. For any vertex in the neighbourhood other than the uniformly chosen vertex, the distributions of its degree and the ages of its children cannot be read from Definition 4, due to the conditioning effects of the discovered edges in the breadth-first search that we define later. However, these effects can be quantified when we work conditionally on the fitness sequence. In particular, their distributions can be deduced from an urn representation of the $(\mathbf{x}, n)$-sequential model conditional on the discovered edges. Consequently, the coupling argument for the root vertex can be applied to these non-root vertices as well.

1.5. Article outline

The approximation results for B_i and $S_{n,i}$ in (1.8) and (1.9) are in Section 2, and are proved in Section 9. In Section 3, we define the tree exploration process, and describe the offspring distributions of the root and of any type L or R parent in the subsequent generations of the local neighbourhood in the $(\mathbf{x}, n)$-Pólya urn tree. We also introduce a conditional analogue of the π-Pólya point tree $(\mathcal{T}, 0)$ in Section 4, which we need for coupling the urn tree and $(\mathcal{T}, 0)$. We couple the 1-neighbourhoods in the urn tree and this analogue in Section 5, and their general r-neighbourhoods in Section 6. To prove Theorem 1, we couple the analogue and $(\mathcal{T}, 0)$ in Section 7. We discuss in more detail the connection of Corollary 1 and Theorem 2 to the previous works [4, 7, 26, 40] in Section 8. In Section 10, we construct the urn representation of the $(\mathbf{x}, n)$-sequential model conditional on a set of edges; and Section 11 collects the additional proofs, which includes that of Theorem 2.

2. Approximation of the beta variables

Let B_i and $S_{n,i}$ be as in (1.8) and (1.9), where we treat $\mathbf{x}$ in B_i and $S_{n,i}$ as a realisation of the fitness sequence $\mathbf{X} := (X_i, i \geqslant 1)$. In this section, we state the approximation results for B_i and $S_{n,i}$, assuming that $X_1 := x_1 > -1$ is fixed and $(X_i, i \geqslant 2)$ are i.i.d. positive variables with $\mu := \mathbb{E}X_2 < \infty$ and $\mathbb{E}[X_2^p] < \infty$ for some $p > 2$. These results are later used to derive the limiting degree distribution of the uniformly chosen vertex of the PA tree and the corresponding convergence rate, where the fitness is not necessarily bounded (Theorem 2). The proofs of the upcoming lemmas are deferred to Section 9.

For the approximations, we require that w.h.p., $\mathbf{X}$ is such that $\sum_{h=2}^{i} X_h$ is close enough to its mean $(i-1)\mu$ for all i sufficiently large. So given $0 < \alpha < 1$ and n, we define

$$A_{\alpha,n} = \left\{ \bigcap_{i=\lceil \phi(n) \rceil}^{\infty} \left\{ \left| \sum_{h=2}^{i} X_h - (i-1)\mu \right| \leqslant i^{\alpha} \right\} \right\}, \tag{2.1}$$

where $\phi(n) = \Omega(n^{\chi})$, with χ as in (1.2). The first lemma is due to an application of standard moment inequalities.

Lemma 1. *Assume that* $\mathbb{E}[X_2^p] < \infty$ *for some* $p > 2$*. Given a positive integer* n *and* $1/2 + 1/p < \alpha < 1$*, there is a constant* $C := C(\mu, \alpha, p)$ *such that* $\mathbb{P}[A_{\alpha,n}] \geqslant 1 - Cn^{\chi[-p(\alpha-1/2)+1]}$.

In the remainder of this article, we mostly work with a realisation $\mathbf{x}$ of $\mathbf{X}$ such that $A_{\alpha,n}$ holds, which we denote (abusively) as $\mathbf{x} \in A_{\alpha,n}$. Below we write $\mathbb{P}_{\mathbf{x}}$ and $\mathbb{E}_{\mathbf{x}}$ to indicate the conditioning on a specific realisation of the fitness sequence $\mathbf{x}$. The next lemma, which extends [4, Lemma 3.1], states that $S_{n,k} \approx (k/n)^{\chi}$ for large enough k and n when $\mathbf{x} \in A_{\alpha,n}$. The proof relies on that we can replace $\mathbb{E}_{\mathbf{x}}[S_{n,k}]$ with $(k/n)^{\chi}$ when $\mathbf{x} \in A_{\alpha,n}$, and approximate $S_{n,k}$ with $\mathbb{E}_{\mathbf{x}}[S_{n,k}]$ using a martingale argument. As $A_{\alpha,n}$ occurs w.h.p., the result allows us to substitute these $S_{n,k}$ with $(k/n)^{\chi}$ when we construct a coupling for the $(\mathbf{x}, n)$-Pólya urn tree (Definition 4).

Lemma 2. *Given a positive integer* n *and* $1/2 < \alpha < 1$*, assume that* $\mathbf{x} \in A_{\alpha,n}$*. Then there are positive constants* $C := C(x_1, \mu, \alpha)$ *and* $c := c(x_1, \mu, \alpha)$ *such that*

$$\mathbb{P}_{\mathbf{x}} \left[\max_{\lceil \phi(n) \rceil \leqslant k \leqslant n} \left| S_{n,k} - \left(\frac{k}{n} \right)^{\chi} \right| \leqslant \delta_n \right] \geqslant 1 - \varepsilon_n, \tag{2.2}$$

where $\phi(n) = \Omega(n^{\chi})$*,* $\delta_n := Cn^{-\chi(1-\alpha)/4}$ *and* $\varepsilon_n := cn^{-\chi(1-\alpha)/2}$.

Remark 2. For vertices whose arrival time is of order at most n^{χ}, the upper bound δ_n in Lemma 2 is only meaningful when $\chi > (\alpha+3)/4$, as otherwise δ_n is of order greater than $(k/n)^{\chi}$. However, the bound is still useful for studying the local weak limit of the PA tree, as w.h.p. we do not see vertices with arrival times that are $o(n)$ in the local neighbourhood of the uniformly chosen vertex.

The last lemma is an extension of [4, Lemma 3.2]. It says that on the event $A_{\alpha,n}$, we can construct the urn tree by generating suitable gamma variables in place of B_j in (1.8). These gamma variables are comparable to $Z_{\bar{v}}$ in the construction of the π-Pólya point tree (Definition 3), after assigning Ulam-Harris labels to the vertices of the urn tree. To state the result, we recall a distributional identity. Independently of B_j, let $((\mathcal{Z}_j, \widetilde{\mathcal{Z}}_{j-1}), 2 \leqslant j \leqslant n)$ be conditionally independent variables such that $\mathcal{Z}_j \sim \text{Gamma}(x_j, 1)$ and $\widetilde{\mathcal{Z}}_j \sim \text{Gamma}(T_j + j, 1)$, where $T_j := \sum_{i=1}^{j} x_i$. Then by the beta-gamma algebra; see e.g. [27],

$$\Big(B_j, \widetilde{\mathcal{Z}}_{j-1} + \mathcal{Z}_j\Big) =_d \left(\frac{\mathcal{Z}_j}{\mathcal{Z}_j + \widetilde{\mathcal{Z}}_{j-1}}, \widetilde{\mathcal{Z}}_{j-1} + \mathcal{Z}_j\right) \quad \text{for } 2 \leqslant j \leqslant n,$$

where the two random variables on the right-hand side are independent. Using the law of the large numbers, we prove the following.

Lemma 3. *Given positive integer n and $1/2 < \alpha < 3/4$, let $\mathcal{Z}_j$ and $\widetilde{\mathcal{Z}}_j$ be as above. Define the event*

$$E_{\varepsilon,j} := \left\{ \left| \frac{\mathcal{Z}_j}{\mathcal{Z}_j + \widetilde{\mathcal{Z}}_{j-1}} - \frac{\mathcal{Z}_j}{(\mu+1)j} \right| \leqslant \frac{\mathcal{Z}_j}{(\mu+1)j}\varepsilon \right\} \quad \textit{for } 2 \leqslant j \leqslant n. \tag{2.3}$$

When $\mathbf{x} \in A_{\alpha,n}$, there is a positive constant $C := C(x_1, \alpha, \mu)$ such that

$$\mathbb{P}_{\mathbf{x}}\Bigg[\bigcap_{j=\lceil \phi(n) \rceil}^{n} E_{\varepsilon,j}\Bigg] \geqslant 1 - C(1+\varepsilon)^4 \varepsilon^{-4} n^{\chi(4\alpha-3)}, \tag{2.4}$$

where $\phi(n) = \Omega(n^{\chi})$. In addition,

$$\mathbb{P}_{\mathbf{x}}\Bigg[\bigcap_{j=\lceil \phi(n) \rceil}^{n} \{\mathcal{Z}_j \leqslant j^{1/2}\}\Bigg] \geqslant 1 - \sum_{j=\lceil \phi(n) \rceil}^{n} j^{-2} \prod_{\ell=0}^{3} (x_j + \ell); \tag{2.5}$$

and if $x_i \in (0, \kappa]$ for all $i \geqslant 2$, then there is a positive constant C such that

$$\mathbb{P}_{\mathbf{x}}\Bigg[\bigcap_{j=\lceil \phi(n) \rceil}^{n} \{\mathcal{Z}_j \leqslant j^{1/2}\}\Bigg] \geqslant 1 - C\kappa^4 n^{-\chi}. \tag{2.6}$$

3. Offspring distributions of the type L and R parents in the urn tree

Recall that a vertex in the local neighbourhood of the uniformly chosen vertex k_0 in the PA tree is a type L parent if it lies on the path from the uniformly chosen vertex to the initial vertex; otherwise it is a type R parent. In this section, we state two main lemmas for the $(\mathbf{x}, n)$-Pólya urn tree (Definition 4). In Lemma 5, for the root k_0, any type L or type R parent in the subsequent generations of the local neighbourhood, we encode its type R children in a suitable Bernoulli sequence that we introduce later. In Lemma 6, we construct the distribution of the type L children of the uniformly chosen vertex and any type L parent in the subsequent generations. For vertex k_0, these results follow immediately from the urn representation in Definition 4 and Theorem 3. For the non-root vertices, these lemmas cannot be deduced from Definition 4. Instead, we need the urn representation in Section 10, which accounts for the conditioning effects of the edges uncovered in the neighbourhood exploration.

3.1. Breath-first search

To construct the offspring distributions, we need to keep track of the vertices that we discover in the local neighbourhood. For this purpose, we have to precisely define the exploration process. We start with a definition.

Definition 5. (*Breadth-first order.*) Write $\bar{w} <_{UH} \bar{y}$ if the Ulam-Harris label $\bar{w}$ is smaller than $\bar{y}$ in the breadth-first order. This means that either $|\bar{w}| < |\bar{y}|$, or when $\bar{w} = (0, w_1, ..., w_q)$ and $\bar{y} = (0, v_1, ..., v_q)$, $w_j < v_j$ for $j = \min\{l : v_l \neq w_l\}$. If $\bar{w}$ is either smaller than or equal to $\bar{y}$ in the breadth-first order, then we write $\bar{w} \leqslant_{UH} \bar{y}$.

As examples, we have $(0, 2, 3) <_{UH} (0, 1, 1, 1)$ and $(0, 3, 1, 5) <_{UH} (0, 3, 4, 2)$. We run a breadth-first search on the $(\mathbf{x}, n)$-Pólya urn tree G_n as follows.

Definition 6. (*Breadth-first search (BFS).*) A BFS of G_n splits the vertex set $V(G_n) = [n]$ into the random subsets $(\mathcal{A}_t, \mathcal{P}_t, \mathcal{N}_t)_{t \geqslant 0}$ as follows, where the letters respectively stand for *active, probed* and *neutral.* We initialise the search with

$$(\mathcal{A}_0, \mathcal{P}_0, \mathcal{N}_0) = (\{k_0\}, \varnothing, V(G_n) \setminus \{k_0\}),$$

where k_0 is uniform in $[n]$. Given $(\mathcal{A}_{t-1}, \mathcal{P}_{t-1}, \mathcal{N}_{t-1})$, where each vertex in $\mathcal{A}_{t-1} \cup \mathcal{P}_{t-1}$

already receives an additional Ulam-Harris label of the form $\bar{w}$, $(\mathcal{A}_t, \mathcal{P}_t, \mathcal{N}_t)$ is generated as follows. Let $v[1] = k_0$ and $v[t] \in \mathbb{N}$ be the vertex in $\mathcal{A}_{t-1}$ that is the smallest in the breadth-first order:

$$v[t] = k_{\bar{y}} \in \mathcal{A}_{t-1} \quad \text{if } \bar{y} <_{UH} \bar{w} \text{ for all } k_{\bar{w}} \in \mathcal{A}_{t-1} \setminus \{k_{\bar{y}}\}. \tag{3.1}$$

Note that the Ulam-Harris labels appear as subscripts. Denote $\mathcal{D}_t$ the set of vertices in $\mathcal{N}_{t-1}$ that either receives an incoming edge from $v[t]$ or sends an outgoing edge to $v[t]$:

$$\mathcal{D}_t := \{j \in \mathcal{N}_{t-1} : \{j, v[t]\} \text{ or } \{v[t], j\} \in E(G_n)\},$$

where $\{i, j\}$ is the edge directed from vertex j to vertex $i < j$. Then in the t-th exploration step, we probe vertex $v[t]$ by marking the neutral vertices attached to $v[t]$ as active. That is,

$$(\mathcal{A}_t, \mathcal{P}_t, \mathcal{N}_t) = (\mathcal{A}_{t-1} \setminus \{v[t]\} \cup \mathcal{D}_t, \mathcal{P}_{t-1} \cup \{v[t]\}, \mathcal{N}_{t-1} \setminus \mathcal{D}_t); \tag{3.2}$$

and if $v[t] = k_{\bar{y}}$, we use the Ulam-Harris scheme to label the newly active vertices as $k_{\bar{y},j} := k_{(\bar{y},j)}$, $j \in \mathbb{N}$, in increasing order of their *vertex arrival times*, so that $k_{\bar{y},i} < k_{\bar{y},j}$ for any $i < j$. If $\mathcal{A}_{t-1} = \varnothing$, we set $(\mathcal{A}_t, \mathcal{P}_t, \mathcal{N}_t) = (\mathcal{A}_{t-1}, \mathcal{P}_{t-1}, \mathcal{N}_{t-1})$.

The characterisation of the BFS above is standard, and more details can be found in works such as [39, Chapter 4] and [23]. The vertex labelling $v[t]$ is very useful for the construction here, as the offspring distribution of $v[t]$ depends on the vertex partition $(\mathcal{A}_{t-1}, \mathcal{P}_{t-1}, \mathcal{N}_{t-1})$; whereas the vertex arrival times are helpful for identifying a type L vertex in the BFS (see Lemma 4 below) and constructing the offspring distributions. On the other hand, we shall use the Ulam-Harris labels to match the vertices when coupling the urn tree and the π-Pólya point tree (Definition 3). Hereafter we ignore the possibility that $v[t] = 1$, because when t (or equivalently the number of discovered vertices) is not too large, the probability that $v[t] = o(n)$ tends to zero as $n \to \infty$. For $t \geqslant 1$, let $v^{(op)}[t]$ (resp. $v^{(oa)}[t]$) be the vertex in $\mathcal{P}_{t-1}$ (resp. $\mathcal{A}_{t-1}$) that has the earliest arrival time, where *op* and *oa* stand for *oldest probed* and *oldest active*. That is,

$$v^{(op)}[t] := \min\{j : j \in \mathcal{P}_{t-1}\} \quad \text{and} \quad v^{(oa)}[t] := \min\{j : j \in \mathcal{A}_{t-1}\}. \tag{3.3}$$

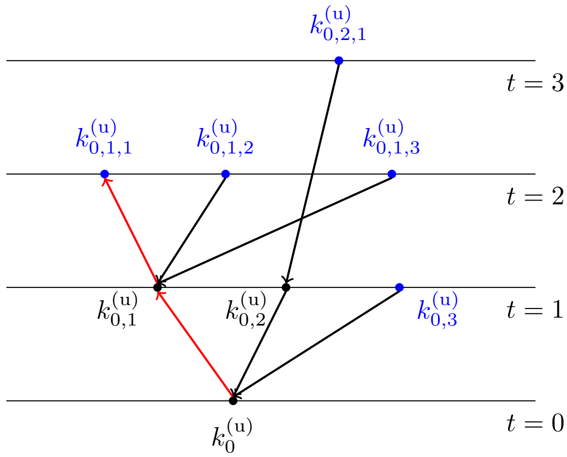

FIGURE 4: Each level corresponds to each time step of the BFS $(\mathbf{x}, n)$-Pólya urn tree G_n. The vertices are arranged from left to right in increasing order of their arrival times. Black and blue vertices correspond to the probed and the active vertices. The red path joins the uniformly chosen vertex and the discovered type L vertices. Here, $\mathcal{P}_3 = \{k^{(\mathrm{u})}_0, k^{(\mathrm{u})}_{0,1}, k^{(\mathrm{u})}_{0,2}\}$, $\mathcal{A}_3 = \{k^{(\mathrm{u})}_{0,3}, k^{(\mathrm{u})}_{0,1,1}, k^{(\mathrm{u})}_{0,1,2}, k^{(\mathrm{u})}_{0,1,3}, k^{(\mathrm{u})}_{0,2,1}\}$, $v[4] = k^{(\mathrm{u})}_{0,3}$, $v^{(op)}[4] = k^{(\mathrm{u})}_{0,1}$ and $v^{(oa)}[4] = k^{(\mathrm{u})}_{0,1,1}$.

3.2. Construction of the offspring distributions

To distinguish the $(\mathbf{x}, n)$-Pólya urn tree G_n from the other models, we apply Ulam-Harris labels of the form $k^{(\mathrm{u})}_{\bar{y}}$ to its vertices, where the superscript $^{(\mathrm{u})}$ stands for urn. An example is given in Figure 4. Define

$$R^{(\mathrm{u})}_{\bar{y}} = \text{the number of type R children of } k^{(\mathrm{u})}_{\bar{y}} \text{ in the } (\mathbf{x}, n)\text{-Pólya urn tree;} \tag{3.4}$$

noting that $R^{(\mathrm{u})}_0$ is the in-degree of the uniformly chosen vertex $k^{(\mathrm{u})}_0$ in G_n. Before we proceed further, we prove a simple lemma to help us identify when $v[t]$ in (3.1) is a type L vertex, which will be useful for constructing the offspring distributions later. The result can be understood as a consequence of the facts that in the tree setting, we uncover a new active type L vertex immediately after we probe an active type L vertex, and that the oldest probed vertex cannot be rediscovered as children of another type R vertex in the subsequent explorations.

Lemma 4. *Assume that $\mathcal{A}_{t-1} \cup \mathcal{P}_{t-1}$ does not contain vertex 1. If $t = 2$, then $v^{(op)}[2] = k^{(\mathrm{u})}_0$ and $v^{(oa)}[2] = k^{(\mathrm{u})}_{0,1}$; while if $2 < i \leqslant t$, $v^{(op)}[i]$ and $v^{(oa)}[i]$ are type L children,* *where $v^{(oa)}[i]$ is the only type L child in $\mathcal{A}_{i-1}$, and it receives an incoming edge from $v^{(op)}[i]$.*

Proof. We prove the lemma by an induction on $2 \leqslant i \leqslant t$. The base case is clear, since $\mathcal{P}_1 = \{k_0^{(\mathrm{u})}\}$ and $\mathcal{A}_1$ consists of its type L and R children. Assume that the lemma holds for some $2 \leqslant i < t$. If we probe a type L child at time i, then $v[i] = v^{(oa)}[i]$, and there is vertex $u \in \mathcal{N}_{i-1}$ that receives the incoming edge emanating from $v[i]$. Hence, vertex u belongs to type L and $v^{(oa)}[i+1] = u$. Furthermore, $v^{(op)}[i+1] = v[i]$, as $v^{(op)}[i]$ sends an outgoing edge to $v[i]$ by assumption, implying $v[i] < v^{(op)}[i]$. If we probe a type R vertex at time i, then $v[i] > v^{(oa)}[i]$ and we uncover vertices in $\mathcal{N}_{i-1}$ that have later arrival times than $v[i]$. Setting $\mathcal{P}_i = \mathcal{P}_{i-1} \cup \{v[i]\}$, we have $v^{(op)}[i+1] = v^{(op)}[i]$ and $v^{(oa)}[i+1] = v^{(oa)}[i]$, which are of type L. $\square$

To construct the offspring distributions for each parent in the local neighbourhood, we also need to define some notation and variables. Let $\mathcal{E}_0 = \varnothing$, and for $t \geqslant 1$, let $\mathcal{E}_t$ be the set of edges connecting the vertices in $\mathcal{A}_t \cup \mathcal{P}_t$. Given a positive integer m, denote the set of vertices and edges in $\mathcal{P}_t$ and $\mathcal{E}_t$ whose arrival time in G_n is earlier than that of vertex m as

$$\mathcal{P}_{t,m} = \mathcal{P}_t \cap [m-1] \quad \text{and} \quad \mathcal{E}_{t,m} = \{\{h,i\} \in \mathcal{E}_t : i < m\}; \tag{3.5}$$

noting that $(\mathcal{A}_0, \mathcal{P}_0, \mathcal{N}_0) = (\{k_0\}, \varnothing, V(G_n) \setminus \{k_0\})$ and $\mathcal{E}_0 = \varnothing$. Below we use $[t]$ in the notation to indicate the exploration step, and omit $\mathbf{x}$ for simplicity. Let $((\mathcal{Z}_i[t], \widetilde{\mathcal{Z}}_i[t]), 2 \leqslant i \leqslant n, i \notin \mathcal{P}_{t-1})$ be independent variables, where

$$\mathcal{Z}_i[1] \sim \mathrm{Gamma}(x_i, 1) \quad \text{and} \quad \widetilde{\mathcal{Z}}_i[1] \sim \mathrm{Gamma}(T_{i-1} + i - 1, 1), \tag{3.6}$$

with $T_i := \sum_{j=1}^{i} x_j$. Now, suppose that $t \geqslant 2$. Due to the size-bias effect of the edge $\{v^{(op)}[t], v^{(oa)}[t]\} \in \mathcal{E}_{t-1}$, we define

$$\mathcal{Z}_i[t] \sim \mathrm{Gamma}(x_i + \mathbb{1}[i = v^{(oa)}[t]], 1), \quad i \in \mathcal{A}_{t-1} \cup \mathcal{N}_{t-1}, \tag{3.7}$$

so that the initial attractiveness of the type L child $v^{(oa)}[t]$ is $x_{v^{(oa)}[t]} + 1$. The shape parameter of $\widetilde{\mathcal{Z}}_i[t]$ defined below is the total weight of the vertices in $\mathcal{A}_{t-1} \cup \mathcal{N}_{t-1}$ whose arrival time is earlier than the i-th time step. This is because when adding a new vertex to the $(\mathbf{x}, n)$-sequential model conditional on having the edges $\mathcal{E}_{t-1}$, the recipient of its outgoing edge cannot be a vertex in $\mathcal{P}_{t-1}$, and is chosen w.p. proportional to the current weights of the vertices in $\mathcal{A}_{t-1} \cup \mathcal{N}_{t-1}$ that arrive before the new vertex. Hence, define

$$\widetilde{\mathcal{Z}}_i[t] \sim \begin{cases} \mathrm{Gamma}(T_{i-1} + i - 1, 1), & \text{if } 2 \leqslant i \leqslant v^{(oa)}[t]; \\ \mathrm{Gamma}(T_{i-1} + i, 1), & \text{if } v^{(oa)}[t] < i < v^{(op)}[t], \\ \mathrm{Gamma}(T_{i-1} + i - \sum_{h \in \mathcal{P}_{t-1,i}} x_h - |\mathcal{E}_{t-1,i}|, 1), & \text{if } v^{(op)}[t] < i \leqslant n. \end{cases} \tag{3.8}$$

Let $B_1[t] := 1$ and $B_i[t] := 0$ for $i \in \mathcal{P}_{t-1}$, as the edges attached to $\mathcal{P}_{t-1}$ are already determined. Furthermore, define

$$B_i[t] := \frac{\mathcal{Z}_i[t]}{\mathcal{Z}_i[t] + \widetilde{\mathcal{Z}}_i[t]} \quad \text{for } i \in \mathcal{A}_{t-1} \cup \mathcal{N}_{t-1}. \tag{3.9}$$

Denote $S_{n,0}[t] := 0$, $S_{n,n}[t] := 1$ and

$$S_{n,i}[t] := \prod_{j=i+1}^{n} (1 - B_j[t]) = \prod_{j=i+1; j \notin \mathcal{P}_{t-1}}^{n} (1 - B_j[t]) \quad \text{for } 1 \leqslant i \leqslant n-1, \tag{3.10}$$

where the second equality is true because $B_i[t] = 0$ for $i \in \mathcal{P}_{t-1}$. Observe that by the beta-gamma algebra, $(B_i[1], 1 \leqslant i \leqslant n) =_d (B_i, 1 \leqslant i \leqslant n)$ and $(S_{n,i}[1], 1 \leqslant i \leqslant n) =_d (S_{n,i}, 1 \leqslant i \leqslant n)$, where B_i and $S_{n,i}$ are as in (1.8) and (1.9). For $t \geqslant 2$, $(B_i[t], 1 \leqslant i \leqslant n)$ are the beta variables in the urn representation of the $(\mathbf{x}, n)$-sequential model conditional on having the edges $\mathcal{E}_{t-1}$; see Section 10.

We construct a Bernoulli sequence that encodes the type R children of vertex $v[t]$ as follows. For $j \in \mathcal{N}_{t-1}$ and $v[t] + 1 \leqslant j \leqslant n$, let $\mathbb{1}_{\mathrm{R}}[j \to v[t]]$ be an indicator variable that takes value one if and only if vertex j sends an outgoing edge to $v[t]$; while for $j \notin \mathcal{N}_{t-1}$, let $\mathbb{1}_{\mathrm{R}}[j \to v[t]] = 0$ w.p. one, since the recipient of the incoming edge from vertex j is already in $\mathcal{P}_{t-1}$. Note that if $v[t] = k^{(\mathrm{u})}_{\bar{y}}$, then $R^{(\mathrm{u})}_{\bar{y}}$ in (3.4) is equal to $\sum_{j=v[t]+1}^{n} \mathbb{1}_{\mathrm{R}}[j \to v[t]]$. We also assume $\mathcal{N}_{t-1} \neq \varnothing$, because for large n, w.h.p. the local neighbourhood of vertex k_0 does not contain all the vertices of G_n. To state the distribution of $(\mathbb{1}_{\mathrm{R}}[j \to v[t]], v[t]+1 \leqslant j \leqslant n)$, we use (3.9) and (3.10) to define

$$P_{j\to v[t]} := \begin{cases} \frac{S_{n,v[t]}[t]}{S_{n,j-1}[t]} B_{v[t]}[t], & \text{if } j \in \{v[t]+1,\ldots,n\} \cap \mathcal{N}_{t-1}; \\ 0, & \text{if } j \in \{v[t]+1,\ldots,n\} \setminus \mathcal{N}_{t-1}. \end{cases} \tag{3.11}$$

Definition 7. Given $(\mathcal{A}_{t-1}, \mathcal{P}_{t-1}, \mathcal{N}_{t-1})$ and $(B_j[t], v[t] \leqslant j \leqslant n)$, let $Y_{j\to v[t]}$, $v[t]+1 \leqslant j \leqslant n$ be conditionally independent Bernoulli variables, each with parameter $P_{j\to v[t]}$. Define this Bernoulli sequence by the random vector

$$\mathbf{Y}_{\mathrm{Be}}^{(v[t],n)} := \left(Y_{(v[t]+1)\to v[t]}, Y_{(v[t]+2)\to v[t]}, ..., Y_{n\to v[t]}\right).$$

With the preparations above, we are ready to state the main results of this section. For $t \geqslant 2$, the following lemmas are immediate consequences of the urn representation in Section 10 for the $(\mathbf{x}, n)$-sequential model conditional on the discovered edges. For $t = 1$, they follow directly from Theorem 3. The first lemma states that we can encode the type R children of the uniformly chosen vertex (the root) or a non-root parent in the local neighbourhood in a Bernoulli sequence; see Figure 3 in the case of the root.

Lemma 5. *Assume that $\mathcal{N}_{t-1} \neq \varnothing$ and $\mathcal{A}_{t-1} \cup \mathcal{P}_{t-1}$ does not contain vertex 1. Then given $(\mathcal{A}_{t-1}, \mathcal{P}_{t-1}, \mathcal{N}_{t-1})$, the random vector $(\mathbb{1}_{\mathrm{R}}[j \to v[t]], v[t]+1 \leqslant j \leqslant n)$ is distributed as $\mathbf{Y}_{\mathrm{Be}}^{(v[t],n)}$.*

Remark 3. When $|\mathcal{E}_{t-1}| = o(n)$ for some $t \geqslant 2$, $P_{j\to v[t]}$ in (3.11) is approximately distributed as $(S_{n,v[t]}/S_{n,j-1})B_{v[t]}$ for n sufficiently large, with B_j and $S_{n,j}$ as in (1.8) and (1.9). As we shall see later in the proof, $|\mathcal{E}_{t-1}| = o(n)$ indeed occurs w.h.p.

The next lemma states when $v[t]$ is the uniformly chosen vertex or a type L parent, we can use the beta variables in (3.9) to obtain the distribution of the type L child of $v[t]$. Observe that $(B_j[t], 2 \leqslant j \leqslant v[t]-1)$ does not appear in Definition 7, but are required for this purpose.

Lemma 6. *Assume that $\mathcal{N}_{t-1} \neq \varnothing$, $\mathcal{A}_{t-1} \cup \mathcal{P}_{t-1}$ does not contain vertex 1, and $v[t]$ is either the uniformly chosen vertex or of type L. Given $(\mathcal{A}_{t-1}, \mathcal{P}_{t-1}, \mathcal{N}_{t-1})$, let $S_{n,j}[t]$ be as in (3.10), and $U \sim \mathrm{U}[0, S_{n,v[t]-1}[t]]$. For $1 \leqslant j \leqslant v[t]-1$, the probability that vertex j receives the only incoming edge from $v[t]$ is given by the probability that*

$S_{n,j-1}[t] \leqslant U < S_{n,j}[t]$.

4. An intermediate tree for graph couplings

As the π-Pólya point tree $(\mathcal{T}, 0)$ in Definition 3 does not have vertex labels that are comparable to the vertex arrival times of the $(\mathbf{x}, n)$-Pólya urn tree G_n, we define a suitable conditional analogue of the π-Pólya point tree. We call this analogue the *intermediate Pólya point tree* $(\mathcal{T}_{\mathbf{x},n}, 0)$, where vertex 0 is its root, and the subscripts are the parameters corresponding to $\mathbf{x}$ and n of G_n.

All the variables of $(\mathcal{T}_{\mathbf{x},n}, 0)$ have the superscript $^{(\mathrm{i})}$ (for intermediate). Each vertex of $(\mathcal{T}_{\mathbf{x},n}, 0)$ has an Ulam-Harris label $\bar{v}$, an age $a^{(\mathrm{i})}_{\bar{v}}$ and a type (except for the root), which are defined similarly as in Section 1.2 for $(\mathcal{T}, 0)$. Moreover, vertex $\bar{v}$ has an additional *PA label* $k^{(\mathrm{i})}_{\bar{v}}$, which determines its initial attractiveness by taking $x_{k^{(\mathrm{i})}_{\bar{v}}}$. The distributions of the PA labels are constructed using gamma variables similar to (3.6) (3.7) and (3.8), and as we remark after the definition, $k^{(\mathrm{i})}_{\bar{v}}$ are approximately distributed as vertices (or equivalently the arrival times) $k^{(\mathrm{u})}_{\bar{v}}$ in the local neighbourhood in G_n. Denote

$$R^{(\mathrm{i})}_{\bar{v}} = \text{the number of type R children of vertex } \bar{v} \text{ in } (\mathcal{T}_{\mathbf{x},n}, 0), \tag{4.1}$$

which is analogous to $R_{\bar{v}}$ in (1.3) and $R^{(\mathrm{u})}_{\bar{v}}$ in (3.4). Finally, recall χ in (1.2) and $\bar{w} \leqslant_{UH} \bar{v}$ whenever the Ulam-Harris label $\bar{w}$ is smaller than $\bar{v}$ in the breadth-first order (Definition 5).

Definition 8. (*Intermediate Pólya point tree.*) Given n and $\mathbf{x}$, $(\mathcal{T}_{\mathbf{x},n}, 0)$ is constructed recursively as follows. The root 0 has an age $a^{(\mathrm{i})}_0 = U_0^{\chi}$ and an initial attractiveness $x_{k^{(\mathrm{i})}_0}$, where $U_0 \sim \mathrm{U}[0,1]$ and $k^{(\mathrm{i})}_0 = \lceil nU_0 \rceil$. Assume that $\big((a^{(\mathrm{i})}_{\bar{w}}, k^{(\mathrm{i})}_{\bar{w}}), \bar{w} \leqslant_{UH} \bar{v}\big)$ and $\big((R^{(\mathrm{i})}_{\bar{w}}, a^{(\mathrm{i})}_{\bar{w},j}, k^{(\mathrm{i})}_{\bar{w},j}), \bar{w} <_{UH} \bar{v}\big)$ have been generated, such that $k^{(\mathrm{i})}_{\bar{w}} > 1$ for $\bar{w} \leqslant_{UH} \bar{v}$. If vertex $\bar{v}$ is the root or belongs to type L, we generate $\big((a^{(\mathrm{i})}_{\bar{v},j}, k^{(\mathrm{i})}_{\bar{v},j}), 1 \leqslant j \leqslant 1 + R^{(\mathrm{i})}_{\bar{v}}\big)$ as follows.

1. We sample the age of its type L child $(\bar{v}, 1)$ by letting $U_{\bar{v},1} \sim \mathrm{U}[0,1]$ and $a^{(\mathrm{i})}_{\bar{v},1} = a^{(\mathrm{i})}_{\bar{v}} U_{\bar{v},1}$.
2. Next we choose the PA label $k^{(\mathrm{i})}_{\bar{v},1}$. Suppose that $t = |\{\bar{w} : \bar{w} \leqslant_{UH} \bar{v}\}|$, so that $t = 1$ if $\bar{v} = 0$. We define the conditionally independent variables $((\mathcal{Z}^{(\mathrm{i})}_j[t], \widetilde{\mathcal{Z}}^{(\mathrm{i})}_j[t]), 2 \leqslant$

$j \leqslant n, j \notin \{k^{(\mathrm{i})}_{\bar w} : \bar w <_{UH} \bar v\}$) as follows. Let $T_j := \sum_{\ell=1}^{j} x_\ell$. When $\bar v = 0$, let

$$\mathcal{Z}^{(\mathrm{i})}_j[1] \sim \mathrm{Gamma}(x_j, 1) \quad \text{and} \quad \widetilde{\mathcal{Z}}^{(\mathrm{i})}_j[1] \sim \mathrm{Gamma}(T_{j-1} + j - 1, 1);$$

whereas when $\bar v = (0, 1, 1, ..., 1)$, let $\mathcal{Z}^{(\mathrm{i})}_j[t] \sim \mathrm{Gamma}(x_j + \mathbb{1}[j = k^{(\mathrm{i})}_{\bar v}], 1)$, and

$$\widetilde{\mathcal{Z}}^{(\mathrm{i})}_j[t] \sim \begin{cases} \mathrm{Gamma}(T_{j-1} + j - 1, 1), & 2 \leqslant j \leqslant k^{(\mathrm{i})}_{\bar v}; \\ \mathrm{Gamma}(T_{j-1} + j, 1), & k^{(\mathrm{i})}_{\bar v} < j < k^{(\mathrm{i})}_{\bar v'}; \\ \mathrm{Gamma}\big(T_{j-1} + j + 1 - |\bar v| - W_{\bar v, j}, 1\big), & k^{(\mathrm{i})}_{\bar v'} < j \leqslant n, \end{cases}$$

where $W_{\bar v,j} := \sum_{\{k^{(\mathrm{i})}_{\bar w} < j : \bar w <_{UH} \bar v\}} \{x_{k^{(\mathrm{i})}_{\bar w}} + R^{(\mathrm{i})}_{\bar w}\}$, and $\bar v' = (0, 1, 1, ..., 1)$ with $|\bar v'| = |\bar v| - 1$. For either the root or the type L parent, define $B^{(\mathrm{i})}_1[t] := 1$, $B^{(\mathrm{i})}_j[t] := 0$ for $j \in \{k^{(\mathrm{i})}_{\bar w} : \bar w <_{UH} \bar v\}$ and

$$B^{(\mathrm{i})}_j[t] := \frac{\mathcal{Z}^{(\mathrm{i})}_j[t]}{\mathcal{Z}^{(\mathrm{i})}_j[t] + \widetilde{\mathcal{Z}}^{(\mathrm{i})}_j[t]} \quad \text{for } j \in [n] \setminus \{k^{(\mathrm{i})}_{\bar w} : \bar w <_{UH} \bar v\}; \tag{4.2}$$

then let $S^{(\mathrm{i})}_{n,0}[t] := 0$, $S^{(\mathrm{i})}_{n,n}[t] := 1$ and $S^{(\mathrm{i})}_{n,j}[t] := \prod_{\ell=j+1}^{n} (1 - B^{(\mathrm{i})}_\ell[t])$ for $1 \leqslant j \leqslant n - 1$. Momentarily define $b := k^{(\mathrm{i})}_{\bar v}$ and $c := k^{(\mathrm{i})}_{\bar v,1}$, we choose the PA label c such that $S^{(\mathrm{i})}_{n,c-1}[t] \leqslant U_{\bar v,1} S^{(\mathrm{i})}_{n,b-1}[t] < S^{(\mathrm{i})}_{n,c}[t]$.

3. We generate the ages and the PA labels of the type R children. Let $(a^{(\mathrm{i})}_{\bar v,j}, 2 \leqslant j \leqslant 1 + R^{(\mathrm{i})}_{\bar v})$ be the points of a mixed Poisson process on $(a^{(\mathrm{i})}_{\bar v}, 1]$ with intensity

$$\lambda^{(\mathrm{i})}_{\bar v}(y) dy := \frac{\mathcal{Z}^{(\mathrm{i})}_b[t]}{\mu \big(a^{(\mathrm{i})}_{\bar v}\big)^{1/\mu}} y^{1/\mu - 1} dy. \tag{4.3}$$

Then choose $k^{(\mathrm{i})}_{\bar v,j}, 2 \leqslant j \leqslant 1 + R^{(\mathrm{i})}_{\bar v}$ such that

$$\big((k^{(\mathrm{i})}_{\bar v,j} - 1)/n\big)^{\chi} < a^{(\mathrm{i})}_{\bar v,j} \leqslant \big(k^{(\mathrm{i})}_{\bar v,j}/n\big)^{\chi}. \tag{4.4}$$

If $\bar v$ is of type R, let $\mathcal{Z}_b[t] \sim \mathrm{Gamma}(x_b, 1)$ and apply step 3 only to obtain $\big((a^{(\mathrm{i})}_{\bar v,j}, k^{(\mathrm{i})}_{\bar v,j}), 1 \leqslant j \leqslant R^{(\mathrm{i})}_{\bar v}\big)$. We build $(\mathcal{T}_{\mathbf{x},n}, 0)$ by iterating this process, and terminate the construction whenever there is some vertex $\bar v$ such that $b = 1$.

We now discuss the distributions of the PA labels above. Clearly, $k_0^{(\mathrm{i})}$ is uniform in $[n]$. Observe that t above is essentially the number of breadth-first exploration steps in $(\mathcal{T}_{\mathbf{x},n},0)$, starting from the root 0. Suppose that we have completed $t-1$ exploration steps in G_n, starting from its uniformly chosen vertex $k_0^{(\mathrm{u})}$, and the resulting neighbourhood is coupled to that of vertex $0\in V((\mathcal{T}_{\mathbf{x},n},0))$ such that they are isomorphic and $k_{\bar{w}}^{(\mathrm{i})}=k_{\bar{w}}^{(\mathrm{u})}$ for all $\bar{w}$ in the neighbourhoods. If $v[t]=k_{\bar{v}}^{(\mathrm{u})}$, a moment's thought shows that the total weight of the vertices $\{k_{\bar{w}}^{(\mathrm{i})}<j:\bar{w}<_{UH}\bar{v}\}$ and the vertex set $\mathcal{P}_{t-1,j}$ satisfy

$$W_{\bar{v},j}+|\bar{v}|-1=\sum_{h\in\mathcal{P}_{t-1,j}}x_h+|\mathcal{E}_{t-1,j}|,$$

where $\mathcal{P}_{t-1,j}$ and $\mathcal{E}_{t-1,j}$ are as in (3.5). Hence, recalling the gammas in (3.7) and (3.8), it follows that $\mathcal{Z}_j[t]=_d\mathcal{Z}_j^{(\mathrm{i})}[t]$ and $\widetilde{\mathcal{Z}}_j[t]=_d\widetilde{\mathcal{Z}}_j^{(\mathrm{i})}[t]$ for any j. So by Lemma 6, $k_{\bar{v}}^{(\mathrm{i})}=_d k_{\bar{v}}^{(\mathrm{u})}$ if $\bar{v}$ is a type L child. However, $k_{\bar{v}}^{(\mathrm{i})}$ is only approximately distributed as $k_{\bar{v}}^{(\mathrm{u})}$ if $\bar{v}$ is of type R. The discretised Poisson point process that generates these PA labels will be coupled to the Bernoulli sequence in Lemma 5 that encodes the type R children of $k_{\bar{v}}^{(\mathrm{u})}$.

When $k_{\bar{v}}^{(\mathrm{i})}=1$, it must be the case that $\bar{v}=0$ or $\bar{v}=(0,1,\dots,1)$. We stop the construction in this case to avoid an ill-defined $\mathcal{Z}_1^{(\mathrm{i})}[t]$ when $-1<x_1\leqslant 0$ and $k_0^{(\mathrm{i})}=1$; and also because vertex 1 cannot have a type L neighbour, so steps 1 and 2 are unnecessary in this case. Nevertheless, for $r<\infty$ and any vertex $\bar{v}$ in the r-neighbourhood of vertex $0\in V((\mathcal{T}_{\mathbf{x},n},0))$, the probability that $k_{\bar{v}}^{(\mathrm{i})}=1$ tends to zero as $n\to\infty$.

5. Local weak convergence: the base case

Recall that, as in Section 3.2, $k_{\bar{v}}^{(\mathrm{u})}$ are the vertices in the local neighbourhood of the uniformly chosen vertex $k_0^{(\mathrm{u})}$ of the $(\mathbf{x},n)$-Pólya urn tree G_n (Definition 4). Let χ be as in (1.2), $(\mathcal{T}_{\mathbf{x},n},0)$ be the intermediate Pólya point tree in Definition 8, with $k_{\bar{v}}^{(\mathrm{i})}$ and $a_{\bar{v}}^{(\mathrm{i})}$ being the PA label and the age of vertex $\bar{v}$ in the tree. Here, we couple $(G_n,k_0^{(\mathrm{u})})$ and $(\mathcal{T}_{\mathbf{x},n},0)$ such that w.p. tending to one, $(B_1(G_n,k_0^{(\mathrm{u})}),k_0^{(\mathrm{u})})\cong(B_1(\mathcal{T}_{\mathbf{x},n},0),0)$, $\big(k_{0,j}^{(\mathrm{u})}/n\big)^\chi\approx a_{0,j}^{(\mathrm{i})}$ and $k_{0,j}^{(\mathrm{u})}=k_{0,j}^{(\mathrm{i})}$. For convenience, we also refer to the rescaled arrival time $(k/n)^\chi$ of vertex k in G_n simply as its *age*.

Let $U_0 \sim \mathrm{U}[0,1]$, $a_0^{(\mathrm{i})} = U_0^\chi$ and $k_0^{(\mathrm{u})} = \lceil nU_0 \rceil$, where $a_0^{(\mathrm{i})}$ is the age of vertex 0. By a direct comparison to Definition 8, $a_0^{(\mathrm{i})} \approx \big(k_0^{(\mathrm{u})}/n\big)^\chi$ and the initial attractiveness of vertex 0 is $x_{k_0^{(\mathrm{u})}}$. Under this coupling we define the event

$$\mathcal{H}_{1,0} = \{a_0^{(\mathrm{i})} > (\log\log n)^{-\chi}\}, \tag{5.1}$$

which guarantees that we choose a vertex of low degree in G_n. To prepare for the coupling, let $((\mathcal{Z}_j[1], \widetilde{\mathcal{Z}}_j[1]), 2 \leqslant j \leqslant n)$ and $(S_{n,j}[1], 1 \leqslant j \leqslant n)$ be as in (3.6) and (3.10). We use $(S_{n,j}[1], 1 \leqslant j \leqslant n)$ to construct the distribution of the type R children of $k_0^{(\mathrm{u})}$, and for sampling the initial attractiveness of vertex $(0,1) \in V((\mathcal{T}_{\mathbf{x},n}, 0))$. Furthermore, let the ages $(a_{0,j}^{(\mathrm{i})}, 1 \leqslant j \leqslant 1 + R_0^{(\mathrm{i})})$ and $R_0^{(\mathrm{u})}$ be as in Definition 8 and (3.4). Below we define a coupling of $(G_n, k_0^{(\mathrm{u})})$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$; and for this coupling we define the events that the children of $k_0^{(\mathrm{u})}$ have low degrees, the 1-neighbourhoods are isomorphic and of size at most $(\log n)^{1/r}$, and the ages are close enough as

$$\begin{aligned}
&\mathcal{H}_{1,1} = \{a_{0,1}^{(\mathrm{i})} > (\log\log n)^{-2\chi}\},\\
&\mathcal{H}_{1,2} = \left\{R_0^{(\mathrm{u})} = R_0^{(\mathrm{i})},\ \text{for } \bar{v} \in V(B_1(\mathcal{T}_{\mathbf{x},n}, 0)),\ k_{\bar{v}}^{(\mathrm{u})} = k_{\bar{v}}^{(\mathrm{i})}, \left| a_{\bar{v}}^{(\mathrm{i})} - \left(\frac{k_{\bar{v}}^{(\mathrm{u})}}{n}\right)^\chi \right| \leqslant C_1 b(n) \right\},\\
&\mathcal{H}_{1,3} = \{R_0^{(\mathrm{i})} < (\log n)^{1/r}\},
\end{aligned} \tag{5.2}$$

where $R_0^{(\mathrm{i})}$ and $R_0^{(\mathrm{u})}$ are as in (3.4) and (4.1), $b(n) := n^{-\frac{\chi}{12}}(\log\log n)^\chi$ and $C_1 := C_1(x_1, \mu)$ will be chosen in the proof of Lemma 8 below. Because the initial attractiveness of the vertices $(0,j)$ and $k_{0,j}^{(\mathrm{u})}$ match and their ages are close enough on the event $\mathcal{H}_{1,2}$, we can couple the Bernoulli and the mixed Poisson sequences that encode their type R children, and hence the 2-neighbourhoods. The event $\mathcal{H}_{1,1}$ ensures that the local neighbourhood of $k_0^{(\mathrm{u})}$ grows slowly; and on the event $\mathcal{H}_{1,3}$, the number of vertex pairs that we need to couple for the 2-neighbourhoods are not too large. So by a union bound argument, the probability that any of the subsequent couplings fail tends to zero as $n \to \infty$.

The aim of this section is to show that when $\sum_{i=2}^j x_i \approx (j-1)\mu$ for all j sufficiently large, we can couple the two graphs such that $\bigcap_{i=0}^3 \mathcal{H}_{1,i}$ occurs w.h.p. Thus, let $A_n :=$

$A_{2/3,n}$ be as in (2.1) with $\alpha = 2/3$; this α is chosen to simplify the calculation, and can be chosen under the assumption of bounded fitness. Recall that $\mathbb{P}_{\mathbf{x}}$ indicates the conditioning on a specific realisation of the fitness sequence $\mathbf{x}$. The main result is the lemma below, and is the base case when we inductively prove an analogous result for the general radius $r < \infty$. Because $B_1(\mathcal{T}_{\mathbf{x},n}, 0)$ approximately distributes as the 1-neighbourhood of the π-Pólya point tree after randomisation of $\mathbf{x}$, and Lemma 1 says that $\mathbb{P}[A_n^c] = O(n^{-b})$ for some $b > 0$, it follows from the lemma below and (1.4) that (1.5) of Theorem 1 holds for $r = 1$.

Lemma 7. *Assume that $x_j \in (0, \kappa]$ for $j \geqslant 2$ and $\mathbf{x} \in A_n$. Let $\mathcal{H}_{1,j}$, $j = 0, \dots, 3$ be as in (5.1) and (5.2). Then there is a coupling of $(G_n, k_0^{(\mathrm{u})})$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with a positive constant $C := C(x_1, \mu, \kappa)$ such that*

$$\mathbb{P}_{\mathbf{x}}\left[\left(\bigcap_{j=0}^{3} \mathcal{H}_{1,j}\right)^c\right] \leqslant C(\log\log n)^{-\chi} \quad \textit{for all } n \geqslant 3. \tag{5.3}$$

The proof of Lemma 7 consists of several lemmas which we now develop. From the definitions of $a_0^{(\mathrm{i})}$ and $a_{0,1}^{(\mathrm{i})}$, it is obvious that the probabilities of the events $\mathcal{H}_{1,0}$ and $\mathcal{H}_{1,1}$ tend to one as $n \to \infty$; and by Chebyshev's inequality, we can show that this is also true for the event $\mathcal{H}_{1,3}$. We take care of $\mathcal{H}_{1,2}$ in the lemma below, whose proof is the core of this section.

Lemma 8. *Retaining the assumption and the notation of Lemma 7, there is a coupling of $(G_n, k_0^{(\mathrm{u})})$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with a positive constant $C := C(x_1, \mu, \kappa)$ such that*

$$\mathbb{P}_{\mathbf{x}}\left[\mathcal{H}_{1,0} \cap \mathcal{H}_{1,1} \cap \mathcal{H}_{1,2}^c\right] \leqslant Cn^{-\beta}(\log\log n)^{1-\chi} \quad \textit{for all } n \geqslant 3,$$

where $0 < \gamma < \chi/12$ and $\beta = \min\{\chi/3 - 4\gamma, \gamma\}$.

Given that the vertices $k_0^{(\mathrm{u})} \in V((G_n, k_0^{(\mathrm{u})}))$ and $0 \in V((\mathcal{T}_{\mathbf{x},n}, 0))$ are coupled, we prove Lemma 8 as follows. Note that $v[1] = k_0^{(\mathrm{u})}$, where $v[t]$ is as in (3.1). We first couple $\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)}$ in Definition 7 and a discretisation of the mixed Poisson process that encodes the ages and the PA labels that are of type R. Then, we couple the type L vertices $k_{0,1}^{(\mathrm{u})}$ and $(0,1) \in V((\mathcal{T}_{\mathbf{x},n}, 0))$ such that on the event $\mathcal{H}_{1,0} \cap \mathcal{H}_{1,1}$, $k_{0,1}^{(\mathrm{i})} = k_{0,1}^{(\mathrm{u})}$ and $\left(k_{0,1}^{(\mathrm{u})}/n\right)^{\chi} \approx a_{0,1}^{(\mathrm{i})}$ w.h.p.

For the means of the discretised Poisson process, we define

$$\lambda^{[1]}_{v[1]+1} := \int_{a_0^{(\mathrm{i})}}^{\left(\frac{v[1]+1}{n}\right)^{\chi}} \frac{\mathcal{Z}_{v[1]}[1]}{(a_0^{(\mathrm{i})})^{1/\mu}\mu} y^{1/\mu-1}dy \quad \text{and} \quad \lambda^{[1]}_{j} := \int_{\left(\frac{j-1}{n}\right)^{\chi}}^{\left(\frac{j}{n}\right)^{\chi}} \frac{\mathcal{Z}_{v[1]}[1]}{(a_0^{(\mathrm{i})})^{1/\mu}\mu} y^{1/\mu-1}dy \tag{5.4}$$

for $v[1]+2 \leqslant j \leqslant n$, where $\mathcal{Z}_{v[1]}[1]$ is the gamma variable in (3.6).

Definition 9. Given $v[1] = k_0^{(\mathrm{u})}$, $a_0^{(\mathrm{i})}$ and $\mathcal{Z}_{v[1]}[1]$, let $V_{j\to v[1]}$, $v[1]+1 \leqslant j \leqslant n$, be conditionally independent Poisson random variables, each with parameter $\lambda^{[1]}_j$ as in (5.4). Define this Poisson sequence by the random vector

$$\mathbf{V}^{(v[1],n)}_{\mathrm{Po}} := (V_{(v[1]+1)\to v[1]}, V_{(v[1]+2)\to v[1]}, ..., V_{n\to v[1]}).$$

Next, we define the events that ensure $P_{j\to v[1]}$ is close enough to $\lambda^{[1]}_j$. Let $\phi(n) = \Omega(n^{\chi})$, $C^{\star} := C^{\star}(x_1,\mu)$ be a positive constant such that (2.2) of Lemma 2 holds with $\delta_n = C^{\star} n^{-\frac{\chi}{12}}$. Let $\mathcal{Z}_j[1]$, $B_j[1]$ and $S_j[1]$ be as in (3.6), (3.9) and (3.10). Given $0<\gamma<\chi/12$, define the events

$$\begin{aligned} F_{1,1} &= \left\{ \max_{\lceil \phi(n)\rceil \leqslant j\leqslant n} \left| S_{n,j}[1] - \left(\frac{j}{n}\right)^{\chi} \right| \leqslant C^{\star} n^{-\frac{\chi}{12}} \right\}; \\ F_{1,2} &= \bigcap_{j=\lceil \phi(n)\rceil}^{n} \left\{ \left| B_j[1] - \frac{\mathcal{Z}_j[1]}{(\mu+1)j} \right| \leqslant \frac{\mathcal{Z}_j[1] n^{-\gamma}}{(\mu+1)j} \right\}; \\ F_{1,3} &= \bigcap_{j=\lceil \phi(n)\rceil}^{n} \{\mathcal{Z}_j[1] \leqslant j^{1/2}\}. \end{aligned} \tag{5.5}$$

The next lemma is the major step towards proving Lemma 8, as it implies that we can couple the ages and the initial attractiveness of the type R children of the uniformly chosen vertex $k_0^{(\mathrm{u})} := v[1] \in V(G_n, k_0^{(\mathrm{u})})$ and the root $0 \in V(\mathcal{T}_{\mathbf{x},n}, 0)$.

Lemma 9. *Retaining the assumption and the notation in Lemma 8, let* $\mathbf{Y}^{(v[1],n)}_{\mathrm{Be}}$ *and* $\mathbf{V}^{(v[1],n)}_{\mathrm{Po}}$ *be as in Definition 7 and 9; and* $F_{1,i}$, $i=1,2,3$ *be as in (5.5). Then we can couple the random vectors so that there is a positive constant* $C := C(x_1,\mu,\kappa)$ *such that*

$$\mathbb{P}_{\mathbf{x}}\left[\{\mathbf{Y}^{(v[1],n)}_{\mathrm{Be}} \neq \mathbf{V}^{(v[1],n)}_{\mathrm{Po}}\} \cap \bigcap_{i=1}^{3} F_{1,i} \cap \mathcal{H}_{1,0} \right] \leqslant C n^{-\gamma} (\log\log n)^{1-\chi} \quad \textit{for all } n \geqslant 3.$$

We summarise the proof here and defer the detail to Section 11.1. On the event $F_{1,1} \cap F_{1,2} \cap \mathcal{H}_{1,0}$, a little calculation shows that the Bernoulli success probability $P_{j\to v[1]}$ in (3.11) is close enough to

$$\widehat{P}_{j\to v[1]} := \left(\frac{v[1]}{j}\right)^{\chi} \frac{\mathcal{Z}_{v[1]}[1]}{(\mu+1)v[1]}, \tag{5.6}$$

while the event $F_{1,3}$ ensures that $\widehat{P}_{j\to v[1]} \leqslant 1$. The Poisson mean $\lambda_j^{[1]} \approx \widehat{P}_{j\to v[1]}$ as $a_0^{(\mathrm{i})} \approx (v[1]/n)^{\chi}$. Hence, we use $\widehat{P}_{j\to v[1]}$ in (5.6) as means to construct two intermediate Bernoulli and Poisson random vectors, and then explicitly couple the four processes. It is enough to consider the coupling under the event $\bigcap_{j=1}^{3} F_{1,j} \cap \mathcal{H}_{1,0}$, because when $\mathbf{x} \in A_n$, Lemma 2 and 3 imply that $F_{1,1}$, $F_{1,2}$ and $F_{1,3}$ occur w.p. tending to one as $n \to \infty$.

Below we use Lemma 2, 3 and 9 to prove Lemma 8.

Proof of Lemma 8. We bound the right-hand side of

$$\begin{aligned}
&\mathbb{P}_{\mathbf{x}}[\mathcal{H}_{1,2}^c \cap \mathcal{H}_{1,1} \cap \mathcal{H}_{1,0}] \\
&\leqslant \mathbb{P}_{\mathbf{x}}\bigg[\bigcap_{j=1}^{3} F_{1,j} \cap \bigcap_{j=0}^{1} \mathcal{H}_{1,j} \cap \mathcal{H}_{1,2}^c\bigg] + \mathbb{P}_{\mathbf{x}}\bigg[\bigg(\bigcap_{j=1}^{3} F_{1,j}\bigg)^c\bigg] \\
&\leqslant \mathbb{P}_{\mathbf{x}}\bigg[\bigcap_{j=1}^{3} F_{1,j} \cap \bigcap_{j=0}^{1} \mathcal{H}_{1,j} \cap \mathcal{H}_{1,2}^c\bigg] + \mathbb{P}_{\mathbf{x}}[F_{1,3}^c] + \mathbb{P}_{\mathbf{x}}[F_{1,2}^c] + \mathbb{P}_{\mathbf{x}}[F_{1,1}^c],
\end{aligned} \tag{5.7}$$

under a suitable coupling of $(G_n, k_0^{(\mathrm{u})})$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$. We first handle the last three terms. Firstly, apply (2.6) of Lemma 3 to obtain

$$\mathbb{P}_{\mathbf{x}}[F_{1,3}^c] = \mathbb{P}_{\mathbf{x}}\bigg[\bigcup_{j=\phi(n)}^{n} \{\mathcal{Z}_j[1] \geqslant j^{1/2}\}\bigg] \leqslant C\kappa^4 n^{-\chi}, \tag{5.8}$$

where C is a positive constant. Since $\mathbf{x} \in A_n$, we can apply (2.4) of Lemma 3 (with $\varepsilon = n^{-\gamma}$, $0 < \gamma < \chi/12$ and $\alpha = 2/3$) and Lemma 2 to deduce that

$$\begin{aligned}
\mathbb{P}_{\mathbf{x}}[F_{1,2}^c] &= \mathbb{P}_{\mathbf{x}}\bigg[\bigcup_{j=\lceil\phi(n)\rceil}^{n} \bigg\{\bigg|B_j[1] - \frac{\mathcal{Z}_j[1]}{(\mu+1)j}\bigg| \geqslant \frac{\mathcal{Z}_j[1] n^{-\gamma}}{(\mu+1)j}\bigg\}\bigg] \leqslant C n^{4\gamma - \frac{\chi}{3}}; \\
\mathbb{P}_{\mathbf{x}}[F_{1,1}^c] &\leqslant \mathbb{P}_{\mathbf{x}}\bigg[\max_{\lceil\phi(n)\rceil \leqslant k \leqslant n} \bigg|S_{n,k}[1] - \Big(\frac{k}{n}\Big)^{\chi}\bigg| \geqslant C^{\star} n^{-\frac{\chi}{12}}\bigg] \leqslant c n^{-\frac{\chi}{6}},
\end{aligned} \tag{5.9}$$

where $C := C(x_1, \gamma, \mu)$ and $c := c(x_1, \mu)$. Note that $\mathbb{P}_{\mathbf{x}}[F^c_{1,2}] \to 0$ as $n \to \infty$ due to our choice of γ.

Next, we give the appropriate coupling to bound the first probability of (5.7). Let the vertices $k_0^{(\mathrm{u})} \in V(G_n, k_0^{(\mathrm{u})})$ and $0 \in V(\mathcal{T}_{\mathbf{x},n}, 0)$ be coupled as in the beginning of this section. Assume that they are such that the event $\mathcal{H}_{1,0}$ occurs; and the variables $((\mathcal{Z}_j[1], \widetilde{\mathcal{Z}}_j[1]), 2 \leqslant j \leqslant n)$ are such that $\bigcap_{j=1}^3 F_{1,j}$ holds. We first argue that under the event $\bigcap_{j=1}^3 F_{1,j} \cap \mathcal{H}_{1,0}$, their type R children can be coupled such that w.h.p., $R_0^{(\mathrm{u})} = R_0^{(\mathrm{i})}$ and $k_{0,j}^{(\mathrm{u})} = k_{0,j}^{(\mathrm{i})}$ for $j = 2, ..., 1 + R_0^{(\mathrm{i})}$. In view of Definition 4 and 8, this follows readily from Lemma 9. It remains to prove that under this coupling, $|(k_0^{(\mathrm{u})}/n)^\chi - a_0^{(\mathrm{i})}|$ and $|a_{0,j}^{(\mathrm{i})} - (k_{0,j}^{(\mathrm{u})}/n)^\chi|$ are sufficiently small. Since $k_0^{(\mathrm{u})} = \lceil n(a_0^{(\mathrm{i})})^{1/\chi} \rceil$, and for $j = 2, ..., 1 + R_0^{(\mathrm{i})}$, $k_{0,j}^{(\mathrm{u})}$ satisfies $k_{0,j}^{(\mathrm{u})} > k_0^{(\mathrm{u})}$ and $((k_{0,j}^{(\mathrm{u})} - 1)/n)^\chi < a_{0,j}^{(\mathrm{i})} \leqslant (k_{0,j}^{(\mathrm{u})}/n)^\chi$, it is enough to bound $(\ell/n)^\chi - ((\ell-1)/n)^\chi$ for $k_0^{(\mathrm{u})} + 1 \leqslant \ell \leqslant n$. For such ℓ, we use $k_0^{(\mathrm{u})} > n(\log\log n)^{-1}$ on the event $\mathcal{H}_{1,0}$, and the mean value theorem to obtain

$$\mathbb{1}[\mathcal{H}_{1,0}]\left[\left(\frac{\ell}{n}\right)^\chi - \left(\frac{\ell-1}{n}\right)^\chi\right] \leqslant \mathbb{1}[\mathcal{H}_{1,0}]\frac{\chi}{n}\left(\frac{n}{\ell-1}\right)^{1-\chi} \leqslant \frac{\chi(\log\log n)^{1-\chi}}{n}. \tag{5.10}$$

Next, we couple the type L child of $k_0^{(\mathrm{u})} \in V(G_n, k_0^{(\mathrm{u})})$ and $0 \in V(\mathcal{T}_{\mathbf{x},n}, 0)$, such that $k_{0,1}^{(\mathrm{u})} = k_{0,1}^{(\mathrm{i})}$ and $(k_{0,1}^{(\mathrm{u})}/n)^\chi \approx a_{0,1}^{(\mathrm{i})}$. Independently from $a_0^{(\mathrm{i})}$, let $U_{0,1} \sim \mathrm{U}[0,1]$ and $a_{0,1}^{(\mathrm{i})} = U_{0,1} a_0^{(\mathrm{i})}$. Recalling $v[2] = k_{0,1}^{(\mathrm{u})}$ in the BFS, it follows from Definition 4 and 8 that we can define $k_{0,1}^{(\mathrm{u})}$ to satisfy

$$S_{n,v[2]-1}[1] \leqslant U_{0,1} S_{n,v[1]-1}[1] < S_{n,v[2]}[1],$$

or equivalently,

$$S_{n,v[2]-1}[1] \leqslant \frac{a_{0,1}^{(\mathrm{i})}}{a_0^{(\mathrm{i})}} S_{n,v[1]-1}[1] < S_{n,v[2]}[1]; \tag{5.11}$$

and take $k_{0,1}^{(\mathrm{u})} = k_{0,1}^{(\mathrm{i})}$. Now, we show that under this coupling, $a_{0,1}^{(\mathrm{i})} \approx (k_{0,1}^{(\mathrm{u})}/n)^\chi$ on the 'good' event $\mathcal{H}_{1,0} \cap \mathcal{H}_{1,1} \cap F_{1,1}$. Observe that $S_{n,v[2]}[1] = \Omega((\log\log n)^{-3\chi})$ on the good event, because for n large enough,

$$\begin{aligned} S_{n,v[2]}[1] &\geqslant U_{0,1} S_{n,v[1]-1}[1] \geqslant (\log\log n)^{-2\chi}\left[\left(\frac{v[1]-1}{n}\right)^\chi - C^\star n^{-\frac{\chi}{12}}\right] \\ &\geqslant (\log\log n)^{-2\chi}[(\log\log n)^{-\chi} - 2C^\star n^{-\frac{\chi}{12}}]. \end{aligned}$$

Since $S_{n,j}[1]$ increases with j, the last calculation implies that $|S_{n,j}[1] - (j/n)^\chi| \leqslant C^\star n^{-\frac{\chi}{12}}$ for $j = k_{0,1}^{(\mathrm{u})}, k_{0,1}^{(\mathrm{u})} - 1$. Furthermore, a little calculation shows that on the event $\mathcal{H}_{1,0} \cap F_{1,1}$, there is a constant $c := c(x_1, \mu)$ such that

$$\big|(S_{n,v[1]-1}[1]/a_0^{(\mathrm{i})}) - 1\big| = cn^{-\frac{\chi}{12}}(\log\log n)^\chi.$$

Swapping $(S_{n,v[1]-1}[1]/a_0^{(\mathrm{i})})$, $S_{n,v[2]}[1]$ and $S_{n,v[2]-1}[1]$ in (5.11) for one, $(v[2]/n)^\chi$ and $((v[2]-1)/n)^\chi$ at the costs above, we conclude that there exists $\widehat{C} := \widehat{C}(x_1, \mu)$ such that on the good event, $|a_{0,1}^{(\mathrm{i})} - (k_{0,1}^{(\mathrm{u})}/n)^\chi| \leqslant \widehat{C} n^{-\frac{\chi}{12}}(\log\log n)^\chi$.

Hence, we can take $C_1 := \widehat{C} \vee \chi$ for the event $\mathcal{H}_{1,2}$, and apply Lemma 9 to obtain

$$\begin{aligned}
\mathbb{P}_{\mathbf{x}}\Bigg[\bigcap_{j=1}^{3} F_{1,j} \cap \mathcal{H}_{1,2}^c \cap \mathcal{H}_{1,1} \cap \mathcal{H}_{1,0}\Bigg] &= \mathbb{P}_{\mathbf{x}}\Bigg[\{\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)} \neq \mathbf{V}_{\mathrm{Po}}^{(v[1],n)}\} \cap \bigcap_{j=1}^{3} F_{1,j} \cap \bigcap_{j=0}^{1} \mathcal{H}_{1,j}\Bigg] \\
&\leqslant \mathbb{P}_{\mathbf{x}}\Bigg[\{\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)} \neq \mathbf{V}_{\mathrm{Po}}^{(v[1],n)}\} \cap \bigcap_{j=1}^{3} F_{1,j} \cap \mathcal{H}_{1,0}\Bigg] \\
&\leqslant C n^{-\gamma}(\log\log n)^{1-\chi}, \qquad (5.12)
\end{aligned}$$

where $C := C(x_1, \mu, \kappa)$ is as in Lemma 9. The lemma is proved by applying (5.8), (5.9) and (5.12) to (5.7). □

Before proving Lemma 7, we need a result that says that under the graph coupling, the event that vertex $k_0^{(\mathrm{u})}$ has a low degree occurs w.h.p. The next lemma is proved in Section 11.2 using Chebyshev's inequality. Recall that $A_n := A_{2/3,n}$ is the event in (2.1) with $\alpha = 2/3$.

Lemma 10. *Assume that $x_i \in (0, \kappa]$ for $i \geqslant 2$ and $\mathbf{x} \in A_n$. Let $\mathcal{H}_{1,i}$, $0, \dots, 3$ be as in (5.2). There is a coupling of $(G_n, k_0^{(\mathrm{u})})$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with a positive integer p and a positive constant $C := C(p, \kappa)$ such that*

$$\mathbb{P}_{\mathbf{x}}\Bigg[\bigcap_{i=0}^{2} \mathcal{H}_{1,i} \cap \mathcal{H}_{1,3}^c\Bigg] \leqslant C(\log n)^{-\frac{p}{r}}(\log\log n)^{\frac{p}{\mu+1}} \quad \textit{for all } n \geqslant 3.$$

We now complete the proof of Lemma 7 by applying Lemma 8 and 10.

Proof of Lemma 7. The lemma follows from bounding the right-hand side of

$$\mathbb{P}_{\mathbf{x}}\Big[\Big(\bigcap_{i=0}^{3}\mathcal{H}_{1,i}\Big)^c\Big] = \mathbb{P}_{\mathbf{x}}[\mathcal{H}_{1,0}^c] + \sum_{i=1}^{3}\mathbb{P}_{\mathbf{x}}\Big[\bigcap_{j=0}^{i-1}\mathcal{H}_{1,j}\cap\mathcal{H}_{1,i}^c\Big],$$

under the coupling in the proof of Lemma 8. To bound $\mathbb{P}_{\mathbf{x}}[\mathcal{H}_{1,0}^c]$ and $\mathbb{P}_{\mathbf{x}}[\mathcal{H}_{1,0}\cap\mathcal{H}_{1,1}^c]$, recall that $a_0^{(\mathrm{i})} = U_0^{\chi}$ and $a_{0,1}^{(\mathrm{i})} = U_{0,1}a_0^{(\mathrm{i})}$, where U_0 and $U_{0,1}$ are independent standard uniform variables. Hence,

$$\begin{aligned}\mathbb{P}_{\mathbf{x}}[\mathcal{H}_{1,0}\cap\mathcal{H}_{1,1}^c] &= \mathbb{E}\big[\mathbb{1}[\mathcal{H}_{1,0}]\mathbb{P}_{\mathbf{x}}[U_{0,1}\leqslant (a_0^{(\mathrm{i})})^{-1}(\log\log n)^{-2\chi}|U_0]\big]\\ &\leqslant \mathbb{P}_{\mathbf{x}}[U_{0,1}\leqslant(\log\log n)^{-\chi}]\\ &= (\log\log n)^{-\chi};\end{aligned}\tag{5.13}$$

and $\mathbb{P}_{\mathbf{x}}[\mathcal{H}_{1,0}^c] = \mathbb{P}_{\mathbf{x}}[U_0 \leqslant (\log\log n)^{-1}] = (\log\log n)^{-1}$. Bounding the remaining probabilities using Lemma 8 and 10 completes the proof. □

6. Local weak convergence: the general case

Recall the definitions of $k_{\bar{w}}^{(\mathrm{u})}$, $a_{\bar{w}}^{(\mathrm{i})}$ and $k_{\bar{w}}^{(\mathrm{i})}$ in the beginning of Section 3.2 and Definition 8. In this section, we inductively couple the uniformly rooted Pólya urn graph $(G_n, k_0^{(\mathrm{u})})$ and the intermediate Pólya point tree $(\mathcal{T}_{\mathbf{x},n}, 0)$ in Definition 8 such that w.h.p., $(B_r(G_n, k_0^{(\mathrm{u})}), k_0^{(\mathrm{u})}) \cong (B_r(\mathcal{T}_{\mathbf{x},n}, 0), 0)$, and for $k_{\bar{w}}^{(\mathrm{u})} \in V(B_r(G_n, k_0^{(\mathrm{u})}))$ and $\bar{w} \in V(B_r(\mathcal{T}_{\mathbf{x},n}, 0)$, their ages are closed enough $(\big(k_{\bar{w}}^{(\mathrm{u})}/n\big)^{\chi} \approx a_{\bar{w}}^{(\mathrm{i})})$ and their PA labels and arrival times match $(k_{\bar{w}}^{(\mathrm{i})} = k_{\bar{w}}^{(\mathrm{u})})$.

Given a positive integer r, let $L[q] := (0, 1, ..., 1)$ and $|L[q]| = q+1$ for $2 \leqslant q \leqslant r$, so that $L[q]$ and $k_{L[q]}^{(\mathrm{u})}$ are the type L children in $\partial\mathfrak{B}_q := V(B_q(\mathcal{T}_{\mathbf{x},n}, 0)) \setminus V(B_{q-1}(\mathcal{T}_{\mathbf{x},n}, 0))$ and $\partial\mathcal{B}_q := V(B_q(G_n, k_0^{(\mathrm{u})})) \setminus V(B_{q-1}(G_n, k_0^{(\mathrm{u})}))$. Let $\chi/12 := \beta_1 > \beta_2 > \cdots > \beta_r > 0$, so that $n^{-\beta_q} > n^{-\beta_{q-1}}(\log\log n)^{q\chi}$ for n large enough. To ensure that we can couple the $(q+1)$-neighbourhoods for $1 \leqslant q \leqslant r-1$, we define the coupling events analogous to $\mathcal{H}_{1,j}, j = 1, 2, 3$ in (5.2):

$$\begin{aligned}
\mathcal{H}_{q,1} &= \{a^{(\mathrm{i})}_{L[q]} > (\log\log n)^{-\chi(q+1)}\},\\
\mathcal{H}_{q,2} &= \bigg\{(B_q(G_n, k^{(\mathrm{u})}_0), k^{(\mathrm{u})}_0) \cong (B_q(\mathcal{T}_{\mathbf{x},n}, 0), 0),\ \text{with } k^{(\mathrm{u})}_{\bar v} = k^{(\mathrm{i})}_{\bar v}\\
&\qquad \text{and } \bigg|a^{(\mathrm{i})}_{\bar v} - \bigg(\frac{k^{(\mathrm{u})}_{\bar v}}{n}\bigg)^{\chi}\bigg| \leqslant C_q n^{-\beta_q} \text{ for } \bar v \in V(B_q(\mathcal{T}_{\mathbf{x},n}, 0))\bigg\},\\
\mathcal{H}_{q,3} &= \{R^{(\mathrm{i})}_{\bar v} < (\log n)^{1/r} \text{ for } \bar v \in V(B_{q-1}(\mathcal{T}_{\mathbf{x},n}, 0))\},
\end{aligned} \tag{6.1}$$

where $R^{(\mathrm{i})}_{\bar v}$ is as in (4.1) and $C_q := C_q(x_1, \mu)$ will be chosen in the proof of Lemma 12 later.

Let $A_n := A_{2/3,n}$ be the event in (2.1) and $\mathbb{P}_{\mathbf{x}}$ indicates the conditioning on a specific realisation of the fitness sequence $\mathbf{x}$. The next lemma is the key result of this section, which essentially states that if we can couple $(G_n, k^{(\mathrm{u})}_0)$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$ such that $(B_q(G_n, k^{(\mathrm{u})}_0), k^{(\mathrm{u})}_0) \cong (B_q(\mathcal{T}_{\mathbf{x},n}, 0), 0)$ w.h.p., then we can achieve this for the $(q + 1)$-neighbourhoods too.

Lemma 11. *Let $\mathcal{H}_{q,i}$, $1 \leqslant q \leqslant r - 1$, $i = 1, 2, 3$ be as in (5.2) and (6.1). Assume that $x_i \in (0, \kappa]$ for $i \geqslant 2$ and $\mathbf{x} \in A_n$. Given $r < \infty$ and $1 \leqslant q \leqslant r-1$, if there is a coupling of $(G_n, k^{(\mathrm{u})}_0)$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with a positive constant $C := C(x_1, \mu, \kappa, q)$ such that*

$$\mathbb{P}_{\mathbf{x}}\bigg[\bigg(\bigcap_{i=1}^{3} \mathcal{H}_{q,i}\bigg)^c\bigg] \leqslant C(\log\log n)^{-\chi} \quad \textit{for all } n \geqslant 3,$$

then there is a coupling of $(G_n, k^{(\mathrm{u})}_0)$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with a positive constant $C' := C'(x_1, \mu, \kappa, q)$ such that

$$\mathbb{P}_{\mathbf{x}}\bigg[\bigg(\bigcap_{i=1}^{3} \mathcal{H}_{q+1,i}\bigg)^c\bigg] \leqslant C'(\log\log n)^{-\chi} \quad \textit{for all } n \geqslant 3.$$

Since Lemma 7 implies that such a graph coupling exists for $r = 1$, combining Lemma 7 and 11 yields the following corollary, which is instrumental in proving Theorem 1.

Corollary 2. *Retaining the assumption and the notation in Lemma 11, there is a coupling of $(G_n, k^{(\mathrm{u})}_0)$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with a positive constant $C := C(x_1, \mu, \kappa, r)$ such* *that*

$$\mathbb{P}_{\mathbf{x}}\left[\left(\bigcap_{i=1}^{3}\mathcal{H}_{r,i}\right)^{c}\right] \leqslant C(\log\log n)^{-\chi} \quad \textit{for all } n \geqslant 3.$$

The proof of Lemma 11 is in the same vein as that of Lemma 7. Fixing $1 \leqslant q \leqslant r-1$, we may assume that two graphs are already coupled such that the event $\bigcap_{i=1}^{3}\mathcal{H}_{q,i}$ has occurred. On the event $\mathcal{H}_{q,1}$, it is clear from the definitions of $a^{(\mathrm{i})}_{L[q+1]}$ and $\mathcal{H}_{q+1,1}$, that $\mathcal{H}_{q+1,1}$ occurs w.h.p.; while on the event $\mathcal{H}_{q,1} \cap \mathcal{H}_{q,3}$, we can easily show that $\mathcal{H}_{q+1,3}$ occurs w.h.p. So for the most part of the proof, we handle the event $\mathcal{H}_{q+1,2}$.

To couple the graphs such that $\mathcal{H}_{q+1,2}$ occurs w.h.p., we consider the vertices of $\partial\mathcal{B}_q$ and $\partial\mathfrak{B}_q$ in the breath-first order, and couple $k^{(\mathrm{u})}_{\bar{w}} \in \partial\mathcal{B}_q$ and $\bar{w} \in \partial\mathfrak{B}_q$ so that w.h.p., their numbers of children are the same and not too large, and the children's ages are close enough. Recall the definition of $v[t]$ in (3.1). If $v[t] = k^{(\mathrm{u})}_{\bar{w}}$, the associated events are as follows:

$$\begin{aligned}\mathcal{K}_{t,1} = \Bigg\{ & R^{(\mathrm{u})}_{\bar{w}} = R^{(\mathrm{i})}_{\bar{w}}, \text{ and for } 1 \leqslant j \leqslant 1 + R^{(\mathrm{i})}_{\bar{w}}, \\ & \qquad k^{(\mathrm{i})}_{\bar{w},j} = k^{(\mathrm{u})}_{\bar{w},j} \text{ and } \left| a^{(\mathrm{i})}_{\bar{w},j} - \left(\frac{k^{(\mathrm{u})}_{\bar{w},j}}{n}\right)^{\chi} \right| \leqslant C_{q+1} n^{-\beta_{q+1}} \Bigg\}, \qquad (6.2) \\ \mathcal{K}_{t,2} = \{ & R^{(\mathrm{i})}_{\bar{w}} < (\log n)^{1/r} \},\end{aligned}$$

where $R^{(\mathrm{u})}_{\bar{w}}$ and $R^{(\mathrm{i})}_{\bar{w}}$ are as in (3.4) and (4.1), C_{q+1} and β_{q+1} are as in the event $\mathcal{H}_{q+1,2}$. Below we only consider the coupling of the type L children in detail, because the type R case can be proved similarly. Observe that for n large enough, there must be a type L child in $\partial\mathcal{B}_q$ on the event $\bigcap_{j=1}^{3}\mathcal{H}_{q,j}$. Define $v[\tau[q]] = k^{(\mathrm{u})}_{L[q]}$, so that

$$\tau[1] = 2 \quad \text{and} \quad \tau[q] = |V(B_{q-1}(G_n, k^{(\mathrm{u})}_0))| + 1 \quad \text{for } 2 \leqslant q \leqslant r-1$$

are the exploration times of the type L children. Noting that

$$(\mathcal{A}_{\tau[q]-1}, \mathcal{P}_{\tau[q]-1}, \mathcal{N}_{\tau[q]-1}) = (\partial\mathcal{B}_q, V(B_{q-1}(G_n, k^{(\mathrm{u})}_0)), V(G_n) \setminus V(B_q(G_n, k^{(\mathrm{u})}_0))),$$

we let

$$\left(\left(\mathcal{Z}_j[\tau[q]], \widetilde{\mathcal{Z}}_j[\tau[q]]\right), j \in \mathcal{A}_{\tau[q]-1} \cup \mathcal{N}_{\tau[q]-1}\right) \quad \text{and} \quad (S_{n,j}[\tau[q]], 1 \leqslant j \leqslant n)$$

be as in (3.7), (3.8) and (3.10). For convenience, we also denote

$$\zeta_q := \mathcal{Z}_{v[\tau[q]]}[\tau[q]], \quad \text{so that } \zeta_q \sim \mathrm{Gamma}(x_{v[\tau[q]]} + 1, 1). \tag{6.3}$$

We use $(S_{n,j}[\tau[q]], 1 \leqslant j \leqslant n)$ to generate the children of vertex $v[\tau[q]]$, and let $\big(a^{(\mathrm{i})}_{L[q],j}, 2 \leqslant j \leqslant 1+R^{(\mathrm{i})}_{L[q]}\big)$ be the points of the mixed Poisson process on $\big(a^{(\mathrm{i})}_{L[q]}, 1\big]$ with intensity

$$\frac{\zeta_q}{\mu\big(a^{(\mathrm{i})}_{L[q]}\big)^{1/\mu}} y^{1/\mu-1} dy, \tag{6.4}$$

and $a^{(\mathrm{i})}_{L[q],1} \sim \mathrm{U}\big[0, a^{(\mathrm{i})}_{L[q]}\big]$.

In the sequel, we develop lemmas analogous to Lemma 8 and 10 to show that on the event $\bigcap_{j=1}^{3} \mathcal{H}_{q,j}$, we can couple the vertices $v[\tau[q]] \in \partial\mathcal{B}_q$ and $\tau[q] \in \partial\mathfrak{B}_q$ such that the events $\mathcal{K}_{\tau[q],1}$ and $\mathcal{K}_{\tau[q],2}$ occur w.h.p. As in the 1-neighbourhood case, the difficult part is proving the claim for $\mathcal{K}_{\tau[q],1}$. From now on we write

$$\overline{\mathcal{H}}_{q,l} := \bigcap_{j=1}^{3} \mathcal{H}_{q,j} \cap \{|\partial\mathcal{B}_q| = l\} \quad \text{for all } l \geqslant 1. \tag{6.5}$$

Lemma 12. *Retaining the assumption and the notation in Lemma 11, let β_q, $\mathcal{K}_{\tau[q],1}$ and $\overline{\mathcal{H}}_{q,l}$ be as in (6.1), (6.2) and (6.5). Then, there is a coupling of $(G_n, k^{(\mathrm{u})}_0)$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with a positive constant $C := C(x_1, \mu, \kappa, q)$ such that*

$$\mathbb{P}_{\mathbf{x}}(\overline{\mathcal{H}}_{q,l} \cap \mathcal{H}_{q+1,1} \cap \mathcal{K}^c_{\tau[q],1}) \leqslant Cn^{-d}(\log\log n)^{q+1}(\log n)^{q/r} \quad \textit{for all } n \geqslant 3 \textit{ and } l \geqslant 1,$$

where $0 < \gamma < \chi/12$ and $d := \min\{\chi/3 - 4\gamma, \gamma, 1-\chi, \beta_q\}$.

Similar to Lemma 8, the key step is to couple the type R children of vertex $k^{(\mathrm{u})}_{L[q]}$. Let the Bernoulli sequence $\mathbf{Y}^{(v[\tau[q]],n)}_{\mathrm{Be}}$ be as in Definition 7, which by Lemma 5, encodes the type R children of vertex $k^{(\mathrm{u})}_{L[q]}$. Additionally, define

$$M_{L[q]} := \min\big\{j \in [n] : (j/n)^{\chi} \geqslant a^{(\mathrm{i})}_{L[q]}\big\}. \tag{6.6}$$

We want to use the bins $\big(a^{(\mathrm{i})}_{L[q]}, ((j/n)^{\chi})^{n}_{j=M_{L[q]}}\big)$ to construct the discretised mixed Poisson process that encodes the ages and the PA labels, i.e. $\big((a^{(\mathrm{i})}_{L[q],j}, k^{(\mathrm{i})}_{L[q],j}\big), 2 \leqslant j \leqslant 1 + R^{(\mathrm{i})}_{L[q]}\big)$. However, it is possible that $M_{L[q]} \neq v[\tau[q]] + 1$, and in which case the numbers of Bernoulli and Poisson variables do not match and a modification of $\mathbf{Y}^{(v[\tau[q]],n)}_{\mathrm{Be}}$ is needed. If $M_{L[q]} \leqslant v[\tau[q]]$, define $Y_{j\to v[\tau[q]]}$, $M_{L[q]} \leqslant j \leqslant v[\tau[q]]$ as Bernoulli variables with means $P_{j\to v[\tau[q]]} := 0$, and concatenate the vectors $(Y_{j\to v[\tau[q]]}, M_{L[q]} \leqslant j \leqslant v[\tau[q]])$ and $\mathbf{Y}^{(v[\tau[q]],n)}_{\mathrm{Be}}$. This corresponds to the fact that vertex j cannot send an outgoing edge to vertex $v[\tau[q]]$ in $(G_n, k^{(\mathrm{u})}_0)$. If $M_{L[q]} \geqslant v[\tau[q]] + 1$, let $\mathbf{Y}^{(v[\tau[q]],n)}_{\mathrm{Be}}$ be as in Definition 7. Saving notation, we redefine

$$\mathbf{Y}^{(v[\tau[q]],n)}_{\mathrm{Be}} := \left(Y_{\widetilde{M}_{L[q]}\to v[\tau[q]]}, Y_{(\widetilde{M}_{L[q]}+1)\to v[\tau[q]]}, \dots, Y_{n\to v[\tau[q]]}\right), \tag{6.7}$$

where

$$\widetilde{M}_{L[q]} := \min\{M_{L[q]}, v[\tau[q]] + 1\}. \tag{6.8}$$

We also make the following adjustment so that the upcoming Poisson random vector is a discretisation of the mixed Poisson point process on $(a^{(\mathrm{i})}_{L[q]}, 1]$, and is still comparable to the modified $\mathbf{Y}^{(v[\tau[q]],n)}_{\mathrm{Be}}$. When $M_{L[q]} \leqslant v[\tau[q]] + 1$, define the Poisson means

$$\lambda^{[\tau[q]]}_{M_{L[q]}} := \int_{a^{(\mathrm{i})}_{L[q]}}^{\left(\frac{M_{L[q]}}{n}\right)^{\chi}} \frac{\zeta_q}{\mu\big(a^{(\mathrm{i})}_{L[q]}\big)^{1/\mu}} y^{1/\mu-1} dy, \quad \lambda^{[\tau[q]]}_{j} := \int_{\left(\frac{j-1}{n}\right)^{\chi}}^{\left(\frac{j}{n}\right)^{\chi}} \frac{\zeta_q}{\mu\big(a^{(\mathrm{i})}_{L[q]}\big)^{1/\mu}} y^{1/\mu-1} dy \tag{6.9}$$

for $M_{L[q]} + 1 \leqslant j \leqslant n$. When $M_{L[q]} \geqslant v[\tau[q]] + 2$, we let

$$\lambda^{[\tau[q]]}_{j} := 0 \quad \text{for } v[\tau[q]] + 1 \leqslant j < M_{L[q]}, \tag{6.10}$$

in addition (6.9), so that there are no Poisson points outside the interval $(a^{(\mathrm{i})}_{L[q]}, 1]$.

Definition 10. Given $k^{(\mathrm{u})}_{L[q]}$, $a^{(\mathrm{i})}_{L[q]}$ and ζ_q defined in (6.3), let $\widetilde{M}_{L[q]}$ be as in (6.8); and for $\widetilde{M}_{L[q]}+1 \leqslant j \leqslant n$, let $V_{j\to v[\tau[q]]}$ be conditionally independent Poisson random variables, each with parameters $\lambda_j^{[\tau[q]]}$ given in (6.9) and (6.10). Define this Poisson sequence by the random vector

$$\mathbf{V}_{\mathrm{Po}}^{(v[\tau[q]],n)} := \left(V_{\widetilde{M}_{L[q]}\to v[\tau[q]]}, V_{(\widetilde{M}_{L[q]}+1)\to v[\tau[q]]}, ..., V_{n\to v[\tau[q]]}\right).$$

We proceed to define the events analogous to $F_{1,j}$, $j=1,2,3$ in (5.5). On these events, $P_{j\to v[\tau[q]]}$ in (3.11) and $\lambda_j^{[\tau[q]]}$ are close enough for most j, and so we can couple the point processes as before. Let $\phi(n)=\Omega(n^\chi)$, $0<\gamma<\chi/12$, $\mathcal{Z}_j[\tau[q]]$, $B_j[\tau[q]]$ and and $S_{n,j}[\tau[q]]$ be as in (3.7), (3.9) and (3.10), and $C_q^\star := C_q^\star(x_1,\mu)$ be a positive constant that we specify later. Define

$$\begin{aligned}
F_{\tau[q],1} &= \left\{\max_{\lceil\phi(n)\rceil\leqslant j\leqslant n}\left|S_{n,j}[\tau[q]]-\left(\frac{j}{n}\right)^\chi\right|\leqslant C_q^\star n^{-\frac{\chi}{12}}\right\},\\
F_{\tau[q],2} &= \bigcap_{j=\lceil\phi(n)\rceil; j\notin\mathcal{P}_{\tau[q]-1}}^{n}\left\{\left|B_j[\tau[q]]-\frac{\mathcal{Z}_j[\tau[q]]}{(\mu+1)j}\right|\leqslant\frac{\mathcal{Z}_j[\tau[q]]n^{-\gamma}}{(\mu+1)j}\right\},\\
F_{\tau[q],3} &= \bigcap_{j=\lceil\phi(n)\rceil; j\notin\mathcal{P}_{\tau[q]-1}}^{n}\{\mathcal{Z}_j[\tau[q]]\leqslant j^{1/2}\}.
\end{aligned} \tag{6.11}$$

The following analogue of Lemma 9 is the main ingredient for proving Lemma 12.

Lemma 13. *Retaining the assumption and the notation in Lemma 12, let* $\mathbf{Y}_{\mathrm{Be}}^{(v[\tau[q]],n)}$, $\mathbf{V}_{\mathrm{Po}}^{(v[\tau[q]],n)}$ *and* $F_{\tau[q],j}$, $j=1,2,3$ *be as in Definition 7, 10 and (6.11). Then we can couple the random vectors so that there is a positive constant* $C=C(x_1,\mu,\kappa,q)$ *such that*

$$\mathbb{P}_{\mathbf{x}}\left[\{\mathbf{Y}_{\mathrm{Be}}^{(v[\tau[q]],n)}\neq\mathbf{V}_{\mathrm{Po}}^{(v[\tau[q]],n)}\}\cap\bigcap_{i=1}^{3}F_{\tau[q],i}\cap\overline{\mathcal{H}}_{q,l}\right]\leqslant Cn^{-d}(\log\log n)^{q+1}(\log n)^{q/r}$$

for all $n\geqslant 3$*, where* $d:=\min\{\beta_q,\gamma,1-\chi\}$.

We highlight the main steps here, and leave the detail in Section 11.3. As in the proof of Lemma 9, we construct a Bernoulli and a Poisson sequence using suitable means $\widehat{P}_{j\to v[\tau[q]]}$, $\widetilde{M}_{L[q]}\leqslant j\leqslant n$, with $\widetilde{M}_{L[q]}$ as in (6.8). Then, we couple the four processes. However, this time we need to handle the cases where we couple a Bernoulli variable with mean zero and a Poisson variable with positive mean, or vice versa. We take care of these cases by choosing the appropriate $\widehat{P}_{j\to v[\tau[q]]}$. Firstly, we observe that on the event $(\bigcap_{j=1}^{3} F_{\tau[q],j}) \cap (\bigcap_{j=1}^{3} \mathcal{H}_{q,j})$, $P_{j\to v[\tau[q]]}$ is close enough to

$$\widehat{P}_{j\to v[\tau[q]]} := \left(\frac{v[\tau[q]]}{j}\right)^{\chi} \frac{\zeta_q}{(\mu+1)v[\tau[q]]} \quad \leqslant 1 \tag{6.12}$$

for $j \in \{v[\tau[q]]+1,\dots,n\} \setminus V(B_q(G_n, k_0^{(\mathrm{u})}))$; whereas for $\max\{M_{L[q]}, v[\tau[q]]+1\} \leqslant j \leqslant n$, $\lambda_j^{[\tau[q]]} \approx \widehat{P}_{j\to v[\tau[q]]}$ because $a_{L[q]}^{(\mathrm{i})} \approx (v[\tau[q]]/n)^{\chi}$ on the event $\mathcal{H}_{q,2}$. Hence, we can couple $Y_{j\to v[\tau[q]]}$ and $V_{j\to v[\tau[q]]}$ for $j \in \{\max\{M_{L[q]}, v[\tau[q]]+1\},\dots,n\} \setminus V(B_q(G_n, k_0^{(\mathrm{u})}))$ as in Lemma 9.

When $j \in V(B_q(G_n, k_0^{(\mathrm{u})}))$, the Bernoulli variable $Y_{j\to v[\tau[q]]}$ has mean zero and we couple it to a Bernoulli variable with mean (6.12) as in Lemma 9. By a little computation,

$$|\partial\mathcal{B}_j| < 1 + j(\log n)^{j/r}, \quad |V(B_j(G_n, k_0^{(\mathrm{u})}))| < 1 + j + j^2(\log n)^{j/r} \tag{6.13}$$

for $1 \leqslant j \leqslant q$ on the event $\mathcal{H}_{q,3}$. Because $\widehat{P}_{j\to v[\tau[q]]}$ are sufficiently small and there are at most $O((\log n)^{q/r})$ such pairs of Bernoulli variables, we can use a union bound to show that the probability that $Y_{j\to v[\tau[q]]} \neq V_{j\to v[\tau[q]]}$ for any such j tends to zero as $n \to \infty$.

We now consider the coupling of $Y_{j\to v[\tau[q]]}$ and $V_{j\to v[\tau[q]]}$ for $\widetilde{M}_{L[q]} \leqslant j \leqslant \max\{M_{L[q]}, k_{L[q]}^{(\mathrm{u})}+1\}$. In Table 1 below, we give the possible combinations of the Bernoulli and Poisson means, and our choice of intermediate means. As indicated in the table, we choose $\widehat{P}_{j\to v[\tau[q]]} > 0$ whenever $P_{j\to v[\tau[q]]} > 0$. When $M_{L[q]} \leqslant v[\tau[q]]$, we couple $V_{j\to v[\tau[q]]}$ and a Poisson variable with mean zero; while when $M_{L[q]} \geqslant v[\tau[q]] + 2$, $V_{j\to v[\tau[q]]} := 0$ by construction, and it is coupled with a Poisson variable with mean (6.12). However, the number of these pairs is small because $M_{L[q]} \approx v[\tau[q]] + 1$ when $a_{L[q]}^{(\mathrm{i})} \approx (v[\tau[q]]/n)^{\chi}$. Consequently, another union bound argument shows that the probability that any of these couplings fail tends to zero as $n \to \infty$.

We are now ready to prove Lemma 12.

Proof of Lemma 12. Recall the definition of $\overline{\mathcal{H}}_{q,l}$ in (6.5). We bound the right-hand

	$M_{L[q]} \leqslant k^{(\mathrm{u})}_{L[q]}$	$M_{L[q]} \geqslant k^{(\mathrm{u})}_{L[q]} + 2$	$M_{L[q]} = k^{(\mathrm{u})}_{L[q]} + 1$
$P_{j\to v[\tau[q]]}$	0	as in (3.11)	as in (3.11)
$\lambda_j^{[\tau[q]]}$	as in (6.9)	0	as in (6.9)
$\widehat{P}_{j\to v[\tau[q]]}$	0	as in (6.12)	as in (6.12)

TABLE 1: Combinations of means for $\widetilde{M}_{L[q]} \leqslant j \leqslant \max\{M_{L[q]}, k^{(\mathrm{u})}_{L[q]}+1\}$. Note that $j = k^{(\mathrm{u})}_{L[q]}+1$ when $M_{L[q]} = k^{(\mathrm{u})}_{L[q]} + 1$, and in that case, the coupling is the same as that of Lemma 9.

side of

$$\begin{aligned}
&\mathbb{P}_{\mathbf{x}}[\overline{\mathcal{H}}_{q,l} \cap \mathcal{H}_{q+1,1} \cap \mathcal{K}^c_{\tau[q],1}] \\
&\leqslant \mathbb{P}_{\mathbf{x}}\left[\overline{\mathcal{H}}_{q,l} \cap \bigcap_{j=1}^{3} F_{\tau[q],j} \cap \mathcal{H}_{q+1,1} \cap \mathcal{K}^c_{\tau[q],1}\right] + \mathbb{P}_{\mathbf{x}}\left[\overline{\mathcal{H}}_{q,l} \cap \left(\bigcap_{j=1}^{3} F_{\tau[q],j}\right)^c\right] \\
&\leqslant \mathbb{P}_{\mathbf{x}}\left[\overline{\mathcal{H}}_{q,l} \cap \bigcap_{j=1}^{3} F_{\tau[q],j} \cap \mathcal{H}_{q+1,1} \cap \mathcal{K}^c_{\tau[q],1}\right] + \sum_{j=1}^{3} \mathbb{P}_{\mathbf{x}}[\overline{\mathcal{H}}_{q,l} \cap F^c_{\tau[q],j}], \qquad (6.14)
\end{aligned}$$

starting from the last three terms. Using (6.13), a computation similar to that for Lemma 2 shows that there are positive constants $C^\star_q = C^\star_q(x_1, \mu)$ and $c^\star = c^\star(x_1, \mu, q)$ such that

$$\mathbb{P}_{\mathbf{x}}\left[\overline{\mathcal{H}}_{q,l} \cap \left\{\max_{\lceil\phi(n)\rceil \leqslant j \leqslant n} \left|S_{n,j}[\tau[q]] - \left(\frac{j}{n}\right)^{\chi}\right| \leqslant C^\star_q n^{-\frac{\chi}{12}}\right\}\right] \leqslant c^\star n^{-\frac{\chi}{6}}, \qquad (6.15)$$

which bounds $\mathbb{P}_{\mathbf{x}}[\overline{\mathcal{H}}_{q,l} \cap F^c_{\tau[q],1}]$. Combining (6.13) and the argument of Lemma 3 also yields

$$\mathbb{P}_{\mathbf{x}}[\overline{\mathcal{H}}_{q,l} \cap F^c_{\tau[q],3}] \leqslant C\kappa^4 n^{-\chi} \quad \text{and} \quad \mathbb{P}_{\mathbf{x}}[\overline{\mathcal{H}}_{q,l} \cap F^c_{\tau[q],2}] \leqslant C' n^{4\gamma - \frac{\chi}{3}}, \qquad (6.16)$$

for some positive constants $C := C(q)$, $C' = C'(x_1, \gamma, \mu, q)$ and $0 < \gamma < \chi/12$.

Next, we bound the first term under a suitable coupling of the vertices $k^{(\mathrm{u})}_{L[q]} \in \partial\mathcal{B}_q$ and $L[q] \in \partial\mathfrak{B}_q$, starting from their type R children. Assume that

$$\big((k^{(\mathrm{u})}_{\bar{w}}, k^{(\mathrm{i})}_{\bar{w}}, a^{(\mathrm{i})}_{\bar{w}}), \bar{w} \in V(B_q(\mathcal{T}_{\mathbf{x},n}, 0))\big), \quad \big((R^{(\mathrm{u})}_{\bar{w}}, R^{(\mathrm{i})}_{\bar{w}}), \bar{w} \in V(B_{q-1}(\mathcal{T}_{\mathbf{x},n}, 0))\big),$$
$$\left(\left(\mathcal{Z}_j[\tau[q]], \widetilde{\mathcal{Z}}_j[\tau[q]]\right), j \in [n] \setminus V(B_{q-1}(\mathcal{T}_{\mathbf{x},n}, 0))\right)$$

are coupled such that $\mathrm{E}_q := \bigcap_{j=1}^{3} \mathcal{H}_{q,j} \cap \bigcap_{j=1}^{3} F_{\tau[q],j}$ occurs. In view of Definition 8 and 7, it follows from Lemma 13 that on the event E_q, there is a coupling such that $R^{(\mathrm{u})}_{L[q]} = R^{(\mathrm{i})}_{L[q]}$ and $k^{(\mathrm{u})}_{L[q],j} = k^{(\mathrm{i})}_{L[q],j}$ for $2 \leqslant j \leqslant 1+R^{(\mathrm{i})}_{L[q]}$ w.h.p. Note that $k_{L[q],j} \geqslant M_{L[q]}$ when $\mathbf{Y}^{(v[\tau[q]],n)}_{\mathrm{Be}}$ and $\mathbf{V}^{(v[\tau[q]],n)}_{\mathrm{Po}}$ are coupled. Thus, using $M_{L[q]} \geqslant n(a^{(\mathrm{i})}_{L[q]})^{1/\chi}$, a little calculation shows that there is a constant $\widehat{c} := \widehat{c}(\mu)$ such that

$$\mathbb{1}\left[\bigcap_{i=1}^{2} \mathcal{H}_{q,i}\right]\left[\left(\frac{j}{n}\right)^{\chi} - \left(\frac{j-1}{n}\right)^{\chi}\right] \leqslant \frac{\widehat{c}(\log\log n)^{(1+q)(1-\chi)}}{n}, \quad M_{L[q]} \leqslant j \leqslant n,$$

implying the ages of the type R children are close enough.

We proceed to couple the type L child of $k^{(\mathrm{u})}_{L[q]} \in \partial\mathcal{B}_q$ and $L[q] \in \partial\mathfrak{B}_q$ on the event $\mathrm{E}_q \cap \mathcal{H}_{q+1}$. Independently from all the variables generated so far, let $U_{L[q],1} \sim \mathrm{U}[0,1]$. Set $a^{(\mathrm{i})}_{L[q],1} := a^{(\mathrm{i})}_{L[q+1]} = U_{L[q],1} a^{(\mathrm{i})}_{L[q]}$, so that $a^{(\mathrm{i})}_{L[q],1}$ is the age of vertex $(L[q],1) \in \partial\mathfrak{B}_{q+1}$. Temporarily defining $f := k^{(\mathrm{u})}_{L[q]}$ and $g := k^{(\mathrm{u})}_{L[q],1}$, we define g to satisfy

$$S_{n,g-1}[\tau[q]] \leqslant U_{L[q],1} S_{n,f-1}[\tau[q]] < S_{n,g}[\tau[q]]. \tag{6.17}$$

In light of Definition 8 and Lemma 6, $k^{(\mathrm{i})}_{L[q],1} = g$. To establish $a^{(\mathrm{i})}_{L[q],1} \approx (g/n)^{\chi}$, we first show that we can substitute $S_{n,j}[\tau[q]]$ with $(j/n)^{\chi}$ for $j = g-1, g$ at a small enough cost. A straightforward computation shows that on the event $\mathrm{E}_q \cap \mathcal{H}_{q+1,1}$,

$$U_{L[q],1} \geqslant (\log\log n)^{-(q+2)\chi} \quad \text{and} \quad f \geqslant n(\log\log n)^{-(q+1)} - C_q n^{1-\beta_q/\chi},$$

where C_q is the constant in the event $\mathcal{H}_{q,2}$. Consequently, there is a constant $C := C(x_1,\mu,q)$ such that on the event $\mathrm{E}_q \cap \mathcal{H}_{q+1,1}$,

$$\begin{aligned} S_{n,g}[\tau[q]] \geqslant U_{L[q],1} S_{n,f-1}[\tau[q]] &\geqslant (\log\log n)^{-(q+2)\chi}\left[\left(\frac{f-1}{n}\right)^{\chi} - C^{\star}_q n^{-\frac{\chi}{12}}\right] \\ &\geqslant C(\log\log n)^{-(2q+3)\chi}. \end{aligned} \tag{6.18}$$

This implies that $g = \Omega(n^{\chi})$, and so $|S_{n,j}[\tau[q]] - (j/n)^{\chi}| \leqslant C^{\star}_q n^{-\chi/12}$ for $j = g-1, g$. Additionally, a direct calculation yields

$$\left|\left(S_{n,f-1}[\tau[q]]/a^{(\mathrm{i})}_{L[q]}\right) - 1\right| \leqslant \widetilde{C} n^{-\beta_q} (\log\log n)^{(q+1)\chi}$$

for some $\widetilde{C} := \widetilde{C}(x_1, \mu, q)$. By replacing $S_{n,j}[\tau[q]]$, $j = g-1, g$ in (6.17) with $a^{(\mathrm{i})}_{L[q]}$ and $(j/n)^\chi$ at the costs above, and using $\beta_{q+1} < \beta_q$, we conclude that, on the event $\mathrm{E}_q \cap \mathcal{H}_{q+1,1}$, there is a constant $\widehat{C} := \widehat{C}(x_1, \mu, q)$ such that $|a^{(\mathrm{i})}_{L[q],1} - (g/n)^\chi| \leqslant \widehat{C} n^{-\beta_{q+1}}$.

Now, pick $C_{q+1} := \widehat{C} \vee \widehat{c}$ for the event $\mathcal{K}_{\tau[q],1}$ and $\mathcal{H}_{q+1,2}$. By Lemma 13, there is a positive constant $C := C(x_1, \mu, \kappa, q)$ such that

$$\begin{aligned}\mathbb{P}_{\mathbf{x}}\Big[\overline{\mathcal{H}}_{q,l} \cap \bigcap_{j=1}^{3} F_{\tau[q],j} \cap \mathcal{H}_{q+1,1} \cap \mathcal{K}^c_{\tau[q],1}\Big] &\leqslant \mathbb{P}_{\mathbf{x}}\Big[\{\mathbf{Y}^{(v[\tau[q]],n)}_{\mathrm{Be}} \neq \mathbf{V}^{(v[\tau[q]],n)}_{\mathrm{Po}}\} \cap \bigcap_{j=1}^{3} F_{\tau[q],j} \cap \overline{\mathcal{H}}_{q,l}\Big] \\ &\leqslant C n^{-d} (\log\log n)^{q+1} (\log n)^{q/r}, \qquad (6.19)\end{aligned}$$

where $d = \min\{\beta_q, \gamma, 1-\chi\}$. The lemma follows from applying (6.15), (6.16) and (6.19) to (6.14). □

The following analogue of Lemma 10 shows that under the graph coupling, w.h.p. $v[\tau[q]]$ has at most $(\log n)^{1/r}$ children. We omit the proof as it is similar to Lemma 10.

Lemma 14. *Retaining the assumption and the notation in Lemma 11, let $\mathcal{K}_{\tau[q],j}$ and $\overline{\mathcal{H}}_{q,j}$ be as in (6.2) and (6.5). Then, given positive integers $r < \infty$ and $q < r$, there is a coupling of $(G_n, k^{(\mathrm{u})}_0)$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with an integer $p > 0$ and a constant $C := C(\kappa, p) > 0$ such that*

$$\mathbb{P}_{\mathbf{x}}[\overline{\mathcal{H}}_{q,l} \cap \mathcal{K}_{q,1} \cap \mathcal{K}^c_{q,2}] \leqslant C (\log n)^{-\frac{p}{r}} (\log\log n)^{\frac{p(q+1)}{\mu+1}} \quad \textit{for all } n \geqslant 3.$$

We now apply Lemma 12 and 14 to prove Lemma 11.

Proof of Lemma 11. We begin by stating the type R analogues of Lemma 12 and 14. Suppose that $(G_n, k^{(\mathrm{u})}_0)$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$ are coupled such that for a type R child $v[t] = k^{(\mathrm{u})}_{\bar{w}} \in \partial\mathcal{B}_q$, the event

$$\mathcal{H}_{q+1,1} \cap \mathcal{J}_{q,l,t} := \mathcal{H}_{q+1,1} \cap \overline{\mathcal{H}}_{q,l} \cap \bigcap_{\{s<t: v[s] \in \partial\mathcal{B}_q\}} (\mathcal{K}_{s,1} \cap \mathcal{K}_{s,2}) \qquad (6.20)$$

has occurred for some $l \geqslant 2$, where $\mathcal{H}_{q+1,1}$, $\overline{\mathcal{H}}_{q,l}$, $\mathcal{K}_{s,1}$ and $\mathcal{K}_{s,2}$ are as in (6.1), (6.5) and (6.2). Let $((\mathcal{Z}_j[t], \widetilde{\mathcal{Z}}_j[t]), j \in \mathcal{A}_{t-1} \cup \mathcal{N}_{t-1})$ and $(S_{n,j}[t], 1 \leqslant j \leqslant n)$ be as in (3.7), (3.8) and (3.10). We use these variables to generate the type R children of $k^{(\mathrm{u})}_{\bar{w}}$, and to sample the ages $(a^{(\mathrm{i})}_{\bar{w},j}, 1 \leqslant j \leqslant R^{(\mathrm{i})}_{\bar{w}})$ of the mixed Poisson process with intensity (6.4), where ζ_q and $a^{(\mathrm{i})}_{L[q]}$ are replaced with $\mathcal{Z}_{k^{(\mathrm{i})}_{\bar{w}}}[t] \sim \mathrm{Gamma}(x_{k^{(\mathrm{i})}_{\bar{w}}}, 1)$ and $a^{(\mathrm{i})}_{\bar{w}}$. Note that $v[t] > k^{(\mathrm{u})}_{L[q]}$, and on the event $\mathcal{J}_{q,l,t}$, the number of discovered vertices up to time t can be bounded as

$$|\mathcal{A}_{t-1} \cup \mathcal{P}_{t-1}| \leqslant 2 + q + (q+1)^2 (\log n)^{(q+1)/r},$$

as implied by (6.13). Hence, we can proceed similarly as for Lemma 12 and 13. In particular, there is a coupling of $(G_n, k^{(\mathrm{u})}_0)$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with positive constants $C := C(x_1, \mu, \kappa, q)$ and $c := c(\kappa, p)$ such that for $n \geqslant 3$,

$$\begin{aligned} &\mathbb{P}_{\mathbf{x}}\left[\mathcal{K}^c_{t,1} \cap \mathcal{J}_{q,l,t}\right] \leqslant C n^{-d} (\log\log n)^{q+1} (\log n)^{\frac{q+1}{r}}; \\ &\mathbb{P}_{\mathbf{x}}\left[\mathcal{K}^c_{t,2} \cap \mathcal{K}_{t,1} \cap \mathcal{J}_{q,l,t}\right] \leqslant c (\log n)^{-\frac{p}{r}} (\log\log n)^{\frac{p(q+1)}{\mu+1}}, \end{aligned} \tag{6.21}$$

where $d = \min\{\chi/3 - 4\gamma, 1-\chi, \gamma, \beta_q\}$.

Define $\widetilde{\mathcal{H}}_q := \bigcap_{j=1}^3 \mathcal{H}_{q,j}$. To bound $\mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{H}}^c_{q+1}]$ using Lemma 12, 14 and (6.21), we use

$$\begin{aligned} \mathbb{P}_{\mathbf{x}}[(\widetilde{\mathcal{H}}_{q+1})^c] &\leqslant \mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{H}}^c_{q+1} \cap \widetilde{\mathcal{H}}_q] + \mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{H}}^c_q] \\ &= \mathbb{P}_{\mathbf{x}}\left[\left(\bigcap_{j=2}^3 \mathcal{H}_{q+1,j}\right)^c \cap \mathcal{H}_{q+1,1} \cap \widetilde{\mathcal{H}}_q\right] + \mathbb{P}_{\mathbf{x}}[\mathcal{H}^c_{q+1,1} \cap \widetilde{\mathcal{H}}_q] + \mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{H}}^c_q] \\ &= \mathbb{P}_{\mathbf{x}}\left[\left(\bigcap_{\{s: v[s] \in \partial\mathcal{B}_q\}} (\mathcal{K}_{s,1} \cap \mathcal{K}_{s,2})\right)^c \cap \mathcal{H}_{q+1,1} \cap \widetilde{\mathcal{H}}_q\right] \qquad (6.22) \\ &\qquad + \mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{H}}_q \cap \mathcal{H}^c_{q+1,1}] + \mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{H}}^c_q], \qquad (6.23) \end{aligned}$$

where the last equality follows from

$$\left(\bigcap_{j=2}^3 \mathcal{H}_{q+1,j}\right)^c \cap \widetilde{\mathcal{H}}_q = \left(\bigcap_{\{s: v[s] \in \partial\mathcal{B}_q\}} (\mathcal{K}_{s,1} \cap \mathcal{K}_{s,2})\right)^c \cap \widetilde{\mathcal{H}}_q.$$

The lemma is proved once we show that the probabilities in (6.22) and (6.23) are of order at most $(\log\log n)^{-\chi}$. By assumption, there is a constant $C := C(x_1, \mu, \kappa, q)$ such that $\mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{H}}^c_q] \leqslant C(\log\log n)^{-\chi}$. Recall that $a^{(\mathrm{i})}_{L[q],1} = U_{L[q],1} a^{(\mathrm{i})}_{L[q]}$, where $U_{L[q],1} \sim \mathrm{U}[0,1]$

is independent of $a^{(\mathrm{i})}_{L[q]}$. Thus, similar to (5.13),

$$\mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{H}}_q \cap \mathcal{H}^c_{q+1,1}] \leqslant (\log\log n)^{-\chi}.$$

For (6.22), define $\widetilde{\mathcal{K}}_{q,s,j} = \mathcal{K}_{|V(B_q(G_n,k_0^{(\mathrm{u})}))|+s,j}$ for $j = 1, 2$ and $1 \leqslant s \leqslant |\partial\mathcal{B}_q|$. Noting that $|\partial\mathcal{B}_q| \leqslant c[n,q] := \lfloor 1 + q(\log n)^{\frac{q}{r}} \rfloor$ on the event $\mathcal{H}_{q,3}$,

$$\begin{aligned}
&\mathbb{P}_{\mathbf{x}}\Big[\Big(\bigcap_{\{s:v[s]\in\partial\mathcal{B}_q\}}(\mathcal{K}_{s,1}\cap\mathcal{K}_{s,2})\Big)^c \cap \mathcal{H}_{q+1,1}\cap\widetilde{\mathcal{H}}_q\Big]\\
&\quad= \mathbb{P}_{\mathbf{x}}\Big[\bigcup_{l=1}^{c[n,q]}\Big\{\{|\partial\mathcal{B}_q| = l\}\cap\Big(\bigcap_{s=1}^{l}(\widetilde{\mathcal{K}}_{q,s,1}\cap\widetilde{\mathcal{K}}_{q,s,2})\Big)^c\Big\}\cap\mathcal{H}_{q+1,1}\cap\widetilde{\mathcal{H}}_q\Big].
\end{aligned}$$

By a union bound, and recalling the definitions of $\overline{\mathcal{H}}_{q,l}$ and $\mathcal{J}_{q,l,s}$ in (6.5) and (6.20), the probability above is at most

$$\begin{aligned}
&\sum_{l=1}^{c[n,q]}\Big\{\mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{K}}^c_{q,1,2}\cap\widetilde{\mathcal{K}}_{q,1,1}\cap\mathcal{H}_{q+1,1}\cap\overline{\mathcal{H}}_{q,l}] + \mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{K}}^c_{q,1,1}\cap\mathcal{H}_{q+1,1}\cap\overline{\mathcal{H}}_{q,l}]\\
&\quad+\sum_{s=2}^{l}\mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{K}}^c_{q,s,2}\cap\widetilde{\mathcal{K}}_{q,s,1}\cap\mathcal{H}_{q+1,1}\cap\mathcal{J}_{q,l,s}] + \mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{K}}^c_{q,s,1}\cap\mathcal{H}_{q+1,1}\cap\mathcal{J}_{q,l,s}]\Big\}.
\end{aligned} \tag{6.24}$$

We bound (6.24) using (6.21), Lemma 12 and 14, where we choose $p > 2(r-1)$ in Lemma 14 so that $\sum_{l=1}^{c[n,q]}\sum_{s=2}^{l}\mathbb{P}_{\mathbf{x}}[\widetilde{\mathcal{K}}^c_{q,s,2}\cap\widetilde{\mathcal{K}}_{q,s,1}\cap\mathcal{J}_{q,l,s}] \to 0$ as $n\to\infty$. It follows that there is a constant $C := C(x_1,\mu,\kappa,q)$ such that (6.24) is bounded by $C(\log\log n)^{-\chi}$. □

7. Completion of the proof

Equipped with Corollary 2, we can prove Theorem 1.

Proof of Theorem 1. In view of (1.1), it is enough to establish (1.5) of Theorem 1. Let $G'_n \sim \mathrm{PA}(\pi, X_1)_n$ (Definition 1), k_0 be its randomly chosen vertex, $(\mathcal{T}_{\mathbf{X},n}, 0)$ be the intermediate Pólya point tree $(\mathcal{T}_{\mathbf{x},n}, 0)$ randomised over $\mathbf{X}$, and $(\mathcal{T}, 0)$ be the π-Pólya point tree in Definition 3. By the triangle inequality for the total variation distance, for any $r < \infty$ we have

$$\begin{aligned}
&d_{\mathrm{TV}}\left(\mathscr{L}\left((B_r(G'_n, k_0), k_0)\right), \mathscr{L}\left((B_r(\mathcal{T}, 0), 0)\right)\right) \\
&\quad \leqslant d_{\mathrm{TV}}\left(\mathscr{L}\left((B_r(G'_n, k_0), k_0)\right), \mathscr{L}\left((B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0)\right)\right) \\
&\qquad\qquad + d_{\mathrm{TV}}\left(\mathscr{L}\left((B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0)\right), \mathscr{L}\left((B_r(\mathcal{T}, 0), 0)\right)\right).
\end{aligned}$$

Let $A_n := A_{2/3,n}$ be as in (2.1). Applying Jensen's inequality to the total variation distance, the above is bounded by

$$\begin{aligned}
&\mathbb{E}[d_{\mathrm{TV}}\left(\mathscr{L}\left((B_r(G'_n, k_0), k_0)|\mathbf{X}\right), \mathscr{L}\left((B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0)|\mathbf{X}\right)\right)] \\
&\qquad + d_{\mathrm{TV}}\left(\mathscr{L}\left((B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0)\right), \mathscr{L}\left((B_r(\mathcal{T}, 0), 0)\right)\right) \\
&\quad \leqslant \mathbb{E}[\mathbb{1}[A_n] d_{\mathrm{TV}}\left(\mathscr{L}\left((B_r(G'_n, k_0), k_0)|\mathbf{X}\right), \mathscr{L}\left((B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0)|\mathbf{X}\right)\right)] \\
&\qquad + \mathbb{P}[A_n^c] + d_{\mathrm{TV}}\left(\mathscr{L}\left((B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0)\right), \mathscr{L}\left((B_r(\mathcal{T}, 0), 0)\right)\right).
\end{aligned}$$

We prove that each term on the right-hand side is of order at most $(\log\log n)^{-\chi}$, starting from the expectation. Let $\mathcal{H}_{r,j}$, $j = 1, 2, 3$ be as in (6.1). For the $(\mathbf{x}, n)$-Pólya urn tree G_n with $\mathbf{x} \in A_n$, Corollary 2 implies that there is a coupling of $(G_n, k_0^{(\mathrm{u})})$ and $(\mathcal{T}_{\mathbf{x},n}, 0)$, with a positive constant $C := C(x_1, \mu, \kappa, r)$ such that

$$\mathbb{P}_{\mathbf{x}}\left[B_r(\mathcal{T}_{\mathbf{x},n}, 0), 0 \ncong (B_r(G_n, k_0^{(\mathrm{u})}), k_0^{(\mathrm{u})})\right] \leqslant \mathbb{P}_{\mathbf{x}}\left[\left(\bigcap_{j=1}^{3} \mathcal{H}_{r,j}\right)^c\right] \leqslant C(\log\log n)^{-\chi}; \quad (7.1)$$

and following from the definition (1.4) of the total variation distance, the expectation is at most $C(\log\log n)^{-\chi}$.

The probability $\mathbb{P}[A_n^c]$ can be bounded using Lemma 1 (with $\alpha = 2/3$). We now couple $(B_r(\mathcal{T}, 0), 0)$ and $(B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0)$ to bound the last term. For this coupling, denote by E the event that the PA labels in $B_r(\mathcal{T}_{\mathbf{X},n}, 0)$ are distinct, that is, $k_{\bar{w}}^{(\mathrm{i})} \neq k_{\bar{v}}^{(\mathrm{i})}$ for any $\bar{w} \neq \bar{v}$. We first construct $B_r(\mathcal{T}_{\mathbf{X},n}, 0)$, then on the event E, we set $B_r(\mathcal{T}, 0)$ as $B_r(\mathcal{T}_{\mathbf{X},n}, 0)$, inheriting the Ulam-Harris labels, ages, types and fitness from the latter; and if the PA labels are not distinct, we generate $B_r(\mathcal{T}, 0)$ independently from

$B_r(\mathcal{T}_{\mathbf{X},n}, 0)$. Under this coupling,

$$\mathbb{P}[(B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0) \ncong (B_r(\mathcal{T}, 0), 0)] = \mathbb{P}[\{(B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0) \ncong (B_r(\mathcal{T}, 0), 0)\} \cap \mathrm{E}^c]. \quad (7.2)$$

From the definition of $\mathcal{H}_{r,2}$ in (6.1), we can use (7.1) and Lemma 1 to bound the second term in (7.2) by $\mathbb{P}[\mathrm{E}^c] \leqslant \mathbb{P}\left[\left(\bigcap_{j=1}^3 \mathcal{H}_{r,j}\right)^c\right] \leqslant C(\log\log n)^{-\chi}$. So, once again by (1.4),

$$d_{\mathrm{TV}}\left(\mathscr{L}\left((B_r(\mathcal{T}_{\mathbf{X},n}, 0), 0)\right), \mathscr{L}\left((B_r(\mathcal{T}, 0), 0)\right)\right) \leqslant C(\log\log n)^{-\chi},$$

which concludes the proof. □

8. Remarks on the limiting distributions of the degree statistics

In this section, we discuss the connection of Corollary 1 and Theorem 2 to [4, 7, 40, 26, 31]. To this end, we state and prove the probability mass function (p.m.f.) of the limiting distributions of the degrees of the uniformly chosen vertex k_0 in $\mathrm{PA}(\pi, X_1)_n$ and its type L neighbour $k_{L[1]}$.

Below we write $a_n \sim b_n$ to indicate $\lim_{n\to\infty} a_n/b_n = 1$. We start with the degree of the uniformly chosen vertex.

Proposition 1. *Retaining the assumption and the notation in Theorem 2, let ξ_0 be as in the theorem. The p.m.f. of the distribution of ξ_0 is*

$$p_\pi(j) = (\mu+1)\int_0^\infty \frac{\Gamma(x+j-1)\Gamma(x+\mu+1)}{\Gamma(x)\Gamma(x+\mu+j+1)} d\pi(x), \qquad j \geqslant 1. \quad (8.1)$$

Furthermore, if $\mathbb{E}X_2^{\mu+1} < \infty$, then as $j \to \infty$,

$$p_\pi(j) \sim C_\pi j^{-(\mu+2)}, \qquad C_\pi := (\mu+1)\int_0^\infty \frac{\Gamma(x+\mu+1)}{\Gamma(x)} d\pi(x). \quad (8.2)$$

Remark 4. We use Proposition 1 to relate Theorem 2 to some known results.

(i) When π is a point mass at 1, $(p_\pi(j), j \geqslant 1)$ is the p.m.f. of $\mathrm{Geo}_1(\sqrt{U})$ (also known as the Yule-Simon distribution), which is the mixture of geometric distributions on positive integers, with the parameter sampled according to the distribution of $\sqrt{U}$, where $U \sim \mathrm{U}(0,1)$. For such fitness sequence and the model that allows for self-loops, [7] established that the limiting empirical degree distribution is $(p_\pi(j), j \geqslant 1)$; and using Stein's method, [31, Theorem 6.1] showed that the total variation distance is at most $n^{-1}\log n$.

(ii) In the multiple-edge setting, and assuming only the fitness has a finite mean, [26] used stochastic approximation to obtain the a.s. limit of the empirical degree distribution. Theorem 2 and Proposition 1 are special cases of [26, Theorem 2.4 and 2.6], but the distributional representation and the rate in Theorem 2 are new.

Proof of Proposition 1. The p.m.f. in (8.1) is an exercise of integration. The steps are similar to the case where the fitness are non-random, and can be found in [4, Lemma 5.2]. The following proof of (8.2) is the same as [26]. By [22, Theorem 1, equation (5)], $j^{\mu+2}\Gamma(x+j-1)/\Gamma(x+\mu+j+1) \leqslant 1$ for all $x>0$ and $j \geqslant 1$. Hence,

$$f_{\mu,j}(x) = j^{\mu+2}\frac{\Gamma(x+j-1)}{\Gamma(x+\mu+j+1)}\frac{\Gamma(x+\mu+1)}{\Gamma(x)}$$

is dominated by $\Gamma(x+\mu+1)/\Gamma(x)$. Thus, if $\mathbb{E}X_2^{\mu+1} < \infty$, the dominated convergence theorem ([14, Theorem 1.5.8, p. 24]) implies that

$$\lim_{j\to\infty} j^{\mu+2}p_\pi(j) = (\mu+1)\int_0^\infty \lim_{j\to\infty} f_{\mu,j}(x)d\pi(x),$$

and so (8.2) follows from $\lim_{j\to\infty} j^{\mu+2}\Gamma(x+j-1)/\Gamma(x+j+\mu+1) = 1$. □

In the following, we give the limiting p.m.f. of the degree of the type L neighbour $k_{L[1]}$ of the uniformly chosen vertex k_0, which shows that the distribution also exhibits a power-law behaviour. When the PA tree has constant initial attractiveness, the proposition is a special case of [4, Lemma 5.2] and [40, Lemma 5.9]. The proof is similar to that of Proposition 1, hence omitted.

Proposition 2. *Retaining the assumption and the notation in Theorem 1, let $R_{L[1]}$ be as in the theorem, the p.m.f. of the random variable $R_{L[1]}+2$ is given by*

$$q_\pi(j) = \mu(\mu+1)(j-1)\int_0^\infty \frac{\Gamma(x+j-1)\Gamma(x+\mu+1)}{\Gamma(x+1)\Gamma(x+\mu+j+1)}d\pi(x), \qquad j \geqslant 2. \tag{8.3}$$

Furthermore, if $\mathbb{E}X_2^\mu < \infty$*, then as* $j \to \infty$*,*

$$q_\pi(j) \sim C_\pi j^{-(\mu+1)}, \qquad C_\pi = \mu(\mu+1)\int_0^\infty \frac{\Gamma(x+\mu+1)}{\Gamma(x+1)} d\pi(x).$$

Comparing Proposition 1 and 2, we see that in the limit, the degree of vertex $k_{L[1]}$ has a heavier tail than the degree of k_0. This is due to the fact that $k_{L[1]}$ has received an incoming edge from k_0, and so $k_{L[1]}$ is more likely to have a higher degree than k_0.

9. Proofs for the approximation results

Here we prove the lemmas in Section 2. The argument is adapted from [32] and has a similar flavour as [6], both of which studied different PA models. For the proofs below, we recall that $\phi(n) = \Omega(n^\chi)$, where χ is as in (1.2).

Proof of Lemma 1. Given $p > 2$, choose $1/2 + 1/p < \alpha < 1$. Let $A_{\alpha,n}$ be as in (2.1), $T_m^\star := \sum_{i=2}^m X_i$ and C_p be the positive constant given in Lemma 15 below, which bounds the moment of a sum of variables in terms of the moments of the summands. Then,

$$\begin{aligned}
\mathbb{P}[A_{\alpha,n}^c] &= \mathbb{P}\Big[\bigcup_{j=\lceil\phi(n)\rceil}^\infty \{|T_j^\star - (j-1)\mu| > j^\alpha\}\Big] \\
&\leqslant \sum_{j=\lceil\phi(n)\rceil}^\infty \mathbb{P}[|T_j^\star - (j-1)\mu| > j^\alpha] \quad \text{by a union bound,} \\
&\leqslant \sum_{j=\lceil\phi(n)\rceil}^\infty \mathbb{E}[|T_j^\star - (j-1)\mu|^p] j^{-\alpha p} \quad \text{by Chebyshev's inequality,} \\
&\leqslant C_p \mathbb{E}[|X_2 - \mu|^p] \sum_{j=\lceil\phi(n)\rceil}^\infty j^{-p(\alpha-1/2)} \quad \text{by Lemma 15,} \\
&\leqslant C_p \mathbb{E}[|X_2 - \mu|^p] \int_{\lceil\phi(n)\rceil-1}^\infty y^{-p(\alpha-1/2)} dy \\
&= C_p[p(\alpha-1/2)-1]^{-1}\mathbb{E}[|X_2-\mu|^p](\lceil\phi(n)\rceil-1)^{1-p(\alpha-1/2)},
\end{aligned}$$

where $p(\alpha - 1/2) > 1$ and $\mathbb{E}[|X_2 - \mu|^p] < \infty$. The lemma follows from $\phi(n) = \Omega(n^\chi)$. □

The next lemma can be found in [35, Item 16, p. 60], where [13] is credited.

Lemma 15. *Let $Y_1, ..., Y_n$ be independent random variables such that for $i = 1, ..., n$, $\mathbb{E}Y_i = 0$ and $\mathbb{E}|Y_i|^p < \infty$ for some $p \geqslant 2$. Let $W_n := \sum_{i=1}^n Y_i$, then*

$$\mathbb{E}[|W_n|^p] \leqslant C_p n^{p/2-1} \sum_{i=1}^n \mathbb{E}[|Y_i|^p],$$

where

$$C_p := \frac{1}{2}p(p-1)\max(1, 2^{p-3})\left[1 + \frac{2}{p}K_{2m}^{(p-2)/2m}\right],$$

and the integer m satifies the condition $2m \leqslant p \leqslant 2m+2$ and $K_{2m} = \sum_{i=1}^m \frac{i^{2m-1}}{(i-1)!}$.

Keeping the notation $T_i := \sum_{h=1}^i x_h$ and $\phi(n) = \Omega(n^\chi)$, we now prove Lemma 2 under the assumption that the event $A_{\alpha,n}$ holds for the realisation $\mathbf{x}$ of the fitness sequence $\mathbf{X}$ (written as $\mathbf{x} \in A_{\alpha,n}$). We use the subscript $\mathbf{x}$ in $\mathbb{P}_\mathbf{x}$ and $\mathbb{E}_\mathbf{x}$ to indicate the conditioning on $\mathbf{X} = \mathbf{x}$. The first step is to derive an expression for $\mathbb{E}_\mathbf{x}[S_{n,k}]$ that holds for any realisation of the fitness sequence, where we modify a moment formula used in proving [38, Proposition 1.3].

Lemma 16. *Let T_i be as above and B_i, $S_{n,i}$ be as in (1.8) and (1.9). Then for $1 \leqslant k < n$ and a positive integer p,*

$$\mathbb{E}_\mathbf{x}[S_{n,k}^p] = \left[\prod_{h=0}^{p-1} \frac{T_k + k + h}{T_n + n - 1 + h}\right] \prod_{j=0}^{p-1} \prod_{i=k+1}^{n-1} \left[1 + \frac{1}{T_i + i - 1 + j}\right]. \tag{9.1}$$

Proof. Since $(B_i, 1 \leqslant i \leqslant n)$ are independent beta random variables, we use the moment formula of the beta distribution to show that for $p \geqslant 1$,

$$\begin{aligned}
\mathbb{E}_\mathbf{x}[S_{n,k}^p] &= \prod_{i=k+1}^n \mathbb{E}\left[(1-B_i)^p\right] = \prod_{i=k+1}^n \prod_{j=0}^{p-1} \frac{T_{i-1} + i - 1 + j}{T_i + i - 1 + j}, \\
&= \prod_{j=0}^{p-1} \left\{ \frac{(T_k + k + j)}{(T_n + n - 1 + j)} \frac{(T_n + n - 1 + j)}{(T_k + k + j)} \prod_{i=k+1}^n \frac{T_{i-1} + i - 1 + j}{T_i + i - 1 + j} \right\}.
\end{aligned}$$

Noting that $T_k + k + j$ and $T_n + n - 1 + j$ in the second product above cancel with $(T_n + n - 1 + j)/(T_k + k + j)$, we can rewrite the final term as

$$\mathbb{E}_\mathbf{x}[S_{n,k}^p] = \left[\prod_{h=0}^{p-1} \frac{T_k + k + h}{T_n + n - 1 + h}\right] \prod_{i=k+1}^{n-1} \prod_{j=0}^{p-1} \frac{T_i + i + j}{T_i + i - 1 + j}$$

$$= \left[\prod_{h=0}^{p-1} \frac{T_k + k + h}{T_n + n - 1 + h}\right] \prod_{j=0}^{p-1} \prod_{i=k+1}^{n-1} \left[1 + \frac{1}{T_i + i - 1 + j}\right],$$

hence concluding the proof. □

Taking $k = 1$ in (9.1) recovers the original formula of [38], where T_i here is A_i in [38]. Next, we use the moment formula to show that when $\mathbf{x} \in A_{\alpha,n}$, the difference between the mean of $S_{n,k}$ and $(k/n)^\chi$ is small enough for large n and $k \geqslant \lceil \phi(n) \rceil$.

Lemma 17. *Given $1/2 < \alpha < 1$ and a positive integer n, assume that $\mathbf{x} \in A_{\alpha,n}$. Then there is a positive constant $C := C(x_1, \mu, \alpha)$ such that for all $\lceil \phi(n) \rceil \leqslant k \leqslant n$,*

$$\left| \mathbb{E}_{\mathbf{x}}[S_{n,k}] - \left(\frac{k}{n}\right)^\chi \right| \leqslant C n^{\chi(\alpha-1)}. \tag{9.2}$$

Proof. We first prove the upper bound for $\mathbb{E}_{\mathbf{x}}[S_{n,k}]$, using the techniques for proving [32, Lemma 4.4]. Applying the formula (9.1) (with $p = 1$), for $\mathbf{x} \in A_{\alpha,n}$ and $k \geqslant \lceil \phi(n) \rceil$, we obtain

$$\mathbb{E}_{\mathbf{x}}[S_{n,k}] \leqslant \frac{k(\mu+1) + k^\alpha + b}{n(\mu+1) - n^\alpha + b - 1} \prod_{i=k+1}^{n-1} \left[1 + \frac{1}{i\mu - i^\alpha + i + b - 1}\right]. \tag{9.3}$$

where $b := x_1 - \mu$. We rewrite the first term on the right-hand side of (9.3) as follows.

$$\begin{aligned} \left(\frac{k}{n}\right) \frac{\mu + 1 + k^{-1+\alpha} + k^{-1} b}{\mu + 1 - n^{-1+\alpha} + n^{-1}(b-1)} &= \left(\frac{k}{n}\right) \left[1 + \frac{k^{\alpha-1} + n^{\alpha-1} + (k^{-1} - n^{-1})b + n^{-1}}{\mu + 1 - n^{\alpha-1} + n^{-1}(b-1)}\right] \\ &\leqslant \frac{k}{n}(1 + \overline{C} k^{\alpha-1}) \\ &\leqslant \frac{k}{n}(1 + \overline{C} n^{\chi(\alpha-1)}), \end{aligned}$$

where $\overline{C} := \overline{C}(x_1, \mu, \alpha)$ is some positive constant. To bound the product term on the right-hand side of (9.3), we take logarithm and bound

$$\left| \sum_{i=k+1}^{n-1} \log\left(1 + \frac{1}{i(\mu+1) - i^\alpha + b - 1}\right) - \frac{1}{\mu+1} \log\left(\frac{n}{k}\right) \right|.$$

By the triangle inequality, we have

$$
\begin{aligned}
&\left|\sum_{i=k+1}^{n-1} \log\left(1+\frac{1}{i\mu - i^{\alpha}+i+b-1}\right) - \frac{1}{\mu+1}\log\left(\frac{n}{k}\right)\right| \\
&\quad \leqslant \left|\sum_{i=k+1}^{n-1} \log\left(1+\frac{1}{i\mu - i^{\alpha}+i+b-1}\right) - \frac{1}{i(\mu+1)-i^{\alpha}+b-1}\right| \qquad (9.4)\\
&\qquad + \left|\sum_{i=k+1}^{n-1} \frac{1}{i(\mu+1)-i^{\alpha}+b-1} - \frac{1}{\mu+1}\log\left(\frac{n}{k}\right)\right|. \qquad (9.5)
\end{aligned}
$$

We use $y-\log(1+y) \leqslant y^2$ in $y \geqslant 0$ to bound (9.4). Letting $y_i = (i(\mu+1)-i^{\alpha}+b-1)^{-1}$, this implies that (9.4) is bounded by $\sum_{i=k+1}^{n} y_i^2$ for $k \geqslant \lceil \phi(n) \rceil$ and n large enough; and by an integral comparison, $\sum_{i=k+1}^{n} y_i^2 = O(n^{-\chi})$. For (9.5), we have

$$
\begin{aligned}
&\left|\sum_{i=k+1}^{n-1} \frac{1}{i(\mu+1)-i^{\alpha}+b-1} - \frac{1}{(\mu+1)}\log\left(\frac{n}{k}\right)\right| \\
&\quad = \left|\sum_{i=k+1}^{n-1} \left(\frac{1}{i(\mu+1)-i^{\alpha}+b-1} - \frac{1}{(\mu+1)i}\right) + O(k^{-1})\right| \\
&\quad \leqslant \sum_{i=k+1}^{n-1} \left|\frac{i^{\alpha}-b+1}{i(\mu+1)(i(\mu+1)-i^{\alpha}+b-1)}\right| + O(k^{-1}) \\
&\quad \leqslant C' \sum_{i=k+1}^{n-1} i^{-2+\alpha} + O(k^{-1}) \\
&\quad \leqslant C'(1-\alpha)^{-1}[\phi(n)]^{\alpha-1} + O(n^{-\chi}),
\end{aligned}
$$

where $C' := C'(x_1, \mu, \alpha)$ is some positive constant. Combining the bounds above, a little calculation shows that there are positive constants $\overline{C} := \overline{C}(x_1, \mu, \alpha)$ and $\widetilde{C} := \widetilde{C}(x_1, \mu, \alpha)$ such that for $\mathbf{x} \in A_{\alpha,n}$ and $\lceil \phi(n) \rceil \leqslant k \leqslant n$,

$$
\mathbb{E}_{\mathbf{x}}[S_{n,k}] \leqslant (k/n)^{\chi}(1+\overline{C}n^{\chi(\alpha-1)})\exp\{\widetilde{C}n^{\chi(\alpha-1)}\}.
$$

Since $e^x = 1 + x + O(x^2)$ for x near zero, there is a positive constant $C := C(x_1, \mu, \alpha)$ such that for $\lceil \phi(n) \rceil \leqslant k \leqslant n$ and n large enough,

$$
\mathbb{E}_{\mathbf{x}}[S_{n,k}] \leqslant (k/n)^{\chi}(1+Cn^{\chi(\alpha-1)}) \leqslant (k/n)^{\chi} + Cn^{\chi(\alpha-1)},
$$

hence proving the desired upper bound. The lower bound can be proved by first noting that for $\lceil \phi(n) \rceil \leqslant k \leqslant n$,

$$\mathbb{E}_{\mathbf{x}}[S_{n,k}] \geqslant \frac{k(\mu+1) - k^{\alpha} + b}{n(\mu+1) + n^{\alpha} + b - 1} \prod_{i=k+1}^{n} \left[1 + \frac{1}{i\mu + i^{\alpha} + i + b - 1}\right].$$

and repeating the calculations above. The detail is omitted. □

With Lemma 17 and a martingale argument, we can prove Lemma 2.

Proof of Lemma 2. Let $\hat{\delta}_n = Cn^{\chi(\alpha-1)/4}$, where $C := C(x_1, \mu, \alpha)$ is the positive constant in (9.2) of Lemma 17. Writing $K := \lceil \phi(n) \rceil$, define

$$\widetilde{E}_{n,\hat{\delta}_n} := \left\{ \max_{K \leqslant k \leqslant n} \left| S_{n,k} - \left(\frac{k}{n}\right)^{\chi} \right| \geqslant 2\hat{\delta}_n \right\}.$$

The lemma follows from bounding $\mathbb{P}_{\mathbf{x}}[\widetilde{E}_{n,\hat{\delta}_n}]$ under the assumption $\mathbf{x} \in A_{\alpha,n}$. By the triangle inequality, we have

$$\mathbb{P}_{\mathbf{x}}[\widetilde{E}_{n,\hat{\delta}_n}] \leqslant \mathbb{P}_{\mathbf{x}}\left[\max_{K \leqslant k \leqslant n} |S_{n,k} - \mathbb{E}_{\mathbf{x}}[S_{n,k}]| + \max_{K \leqslant j \leqslant n} |\mathbb{E}_{\mathbf{x}}[S_{n,j}] - (j/n)^{\chi}| \geqslant 2\hat{\delta}_n\right].$$

Applying Lemma 17 to bound the difference between $(j/n)^{\chi}$ and $\mathbb{E}_{\mathbf{x}}[S_{n,j}]$, we obtain

$$\mathbb{P}_{\mathbf{x}}[\widetilde{E}_{n,\hat{\delta}_n}] \leqslant \mathbb{P}_{\mathbf{x}}\left[\max_{K \leqslant k \leqslant n} |S_{n,k} - \mathbb{E}_{\mathbf{x}}[S_{n,k}]| \geqslant \hat{\delta}_n\right] \leqslant \mathbb{P}_{\mathbf{x}}\left[\max_{K \leqslant k \leqslant n} \left|S_{n,k}(\mathbb{E}_{\mathbf{x}}[S_{n,k}])^{-1} - 1\right| \geqslant \hat{\delta}_n\right],$$

where the second inequality is due to $\mathbb{E}_{\mathbf{x}}[S_{n,k}] \leqslant 1$. We bound the right-hand side of the above using a martingale argument. Recall that $\mathbb{E}_{\mathbf{x}}[S_{n,k}] = \prod_{j=k+1}^{n} \mathbb{E}[1 - B_j]$. Define $M_0 := 1$ and for $j = 1, ..., n-K$, let

$$M_j := \prod_{i=n-j+1}^{n} \frac{1 - B_i}{\mathbb{E}[1 - B_i]} = \frac{S_{n,n-j}}{\mathbb{E}[S_{n,n-j}]};$$

Let $\mathcal{F}_j$ be the σ-algebra generated by $(B_i, n-j+1 \leqslant i \leqslant n)$ for $1 \leqslant j \leqslant n-K$, with $\mathcal{F}_0 = \varnothing$. It follows that $((M_j, \mathcal{F}_j), 0 \leqslant j \leqslant n-K)$ is a martingale with $\mathbb{E}[M_j] = 1$. Since $(M_j - 1)^2$ is a submartingale, Doob's inequality [14, Theorem 4.4.2, p. 204] yields

$$\mathbb{P}_{\mathbf{x}}[\widetilde{E}_{n,\hat{\delta}_n}] \leqslant \mathbb{P}_{\mathbf{x}}\left[\max_{0 \leqslant j \leqslant n-K} |M_j - 1| \geqslant \hat{\delta}_n\right] \leqslant \hat{\delta}_n^{-2} \mathrm{Var}_{\mathbf{x}}(M_{n-K}), \tag{9.6}$$

where $\mathrm{Var}_{\mathbf{x}}(M_{n-K})$ is the variance conditional on $\mathbf{x}$. We use the formulas for the first and second moments of the beta distribution to bound the variance:

$$\begin{aligned}\mathrm{Var}_{\mathbf{x}}(M_{n-K}) = \mathbb{E}_{\mathbf{x}}[(M_{n-K})^2] - 1 &= \left[\prod_{j=K+1}^{n} \frac{(T_{j-1}+j)}{(T_{j-1}+j-1)} \frac{(T_j+j-1)}{(T_j+j)}\right] - 1 \\ &= \prod_{j=K+1}^{n} \left[1 + \frac{T_j - T_{j-1}}{(T_j+j)(T_{j-1}+j-1)}\right] - 1.\end{aligned}$$

Below we allow the positive constant $C' = C'(x_1, \mu, \alpha)$ to vary from line to line. As $|\sum_{i=2}^{j} x_i - (j-1)\mu| \leqslant j^\alpha$ for all $K+1 \leqslant j \leqslant n$ when $\mathbf{x} \in A_{\alpha,n}$, a little computation yields

$$\begin{aligned}\mathrm{Var}_{\mathbf{x}}(M_{n-K}) &\leqslant \prod_{j=K+1}^{n} \left[1 + \frac{\mu + j^\alpha + (j-1)^\alpha}{\{(\mu+1)j - \mu - j^\alpha + x_1\}\{(\mu+1)j - 2\mu - j^\alpha + x_1 - 1\}}\right] - 1 \\ &\leqslant \prod_{j=K+1}^{n} (1 + C' j^{\alpha-2}) - 1 \\ &\leqslant C' K^{\alpha-1}, \qquad (9.7)\end{aligned}$$

Applying (9.7) to (9.6) completes the proof. □

We conclude this section with the proof of Lemma 3, recalling that $\mathcal{Z}_j \sim \mathrm{Gamma}(x_j, 1)$ and $\widetilde{\mathcal{Z}}_j \sim \mathrm{Gamma}(T_j + j, 1)$.

Proof of Lemma 3. Writing $T_j := \sum_{i=1}^{j} x_i$ and $Y_j \sim \mathrm{Gamma}(T_j + j - 1, 1)$, we first prove (2.4). Let $E_{\varepsilon,j}$ be as in (2.3), we have

$$\begin{aligned}\mathbb{P}_{\mathbf{x}}[E^c_{\varepsilon,j}] &= \mathbb{P}_{\mathbf{x}}\left[\left|\frac{\mathcal{Z}_j}{\mathcal{Z}_j + \widetilde{\mathcal{Z}}_{j-1}} - \frac{\mathcal{Z}_j}{(\mu+1)j}\right| \geqslant \frac{\mathcal{Z}_j}{(\mu+1)j}\varepsilon\right] \\ &= \mathbb{P}_{\mathbf{x}}\left[\left|\frac{(\mu+1)j}{\mathcal{Z}_j + \widetilde{\mathcal{Z}}_{j-1}} - 1\right| \geqslant \varepsilon\right] \\ &\leqslant \mathbb{P}_{\mathbf{x}}\left[\left|\frac{\mathcal{Z}_j + \widetilde{\mathcal{Z}}_{j-1}}{(\mu+1)j} - 1\right| \geqslant \frac{\varepsilon}{1+\varepsilon}\right] \\ &= \mathbb{P}_{\mathbf{x}}\left[\left|\frac{Y_j}{(\mu+1)j} - 1\right| \geqslant \frac{\varepsilon}{1+\varepsilon}\right];\end{aligned}$$

and by Chebyshev inequality,

$$\mathbb{P}_{\mathbf{x}}[E^c_{\varepsilon,j}] \leqslant \left(\frac{1+\varepsilon}{\varepsilon}\right)^4 \mathbb{E}_{\mathbf{x}}\left[\left(\frac{Y_j}{(\mu+1)j}-1\right)^4\right],$$

and so (2.4) follows from bounding the moment above under the assumption $\mathbf{x} \in A_{\alpha,n}$, and applying a union bound. Let $a_j := T_j + j - 1$. Using the moment formula for the standard gamma distribution, a little calculation shows that the moment is equal to

$$\begin{aligned}&\frac{\mathbb{E}_{\mathbf{x}}[Y_j^4]}{(\mu+1)^4 j^4} - \frac{4\mathbb{E}_{\mathbf{x}}[Y_j^3]}{(\mu+1)^3 j^3} + \frac{6\mathbb{E}_{\mathbf{x}}[Y_j^2]}{(\mu+1)^2 j^2} - \frac{4\mathbb{E}_{\mathbf{x}}[Y_j]}{(\mu+1)j} + 1\\ &= \frac{\prod_{k=0}^3 (a_j+k)}{(\mu+1)^4 j^4} - \frac{4\prod_{k=0}^2 (a_j+k)}{(\mu+1)^3 j^3} + \frac{6\prod_{k=0}^1 (a_j+k)}{(\mu+1)^2 j^2} - \frac{4a_j}{(\mu+1)j} + 1.\end{aligned}$$

Noting that $|a_j - (\mu+1)j| \leqslant j^\alpha + x_1 + \mu + 1$ for $j \geqslant \phi(n)$, a direct computation shows that there is a positive constant $C := C(x_1, \mu, \alpha)$ such that

$$\mathbb{E}_{\mathbf{x}}\left[\left(\frac{Y_j}{(\mu+1)j}-1\right)^4\right] \leqslant C j^{4\alpha-4}. \tag{9.8}$$

We now prove (2.4) using (9.8). Let $C := C(x_1, \alpha, \mu)$ be a positive constant that may vary at each step of the calculation. Then,

$$\begin{aligned}\mathbb{P}_{\mathbf{x}}\Bigg[\bigcup_{j=\lceil\phi(n)\rceil}^{n} E^c_{\varepsilon,j}\Bigg] &\leqslant \sum_{j=\lceil\phi(n)\rceil}^{n} \mathbb{P}_{\mathbf{x}}[E^c_{\varepsilon,j}] \leqslant C\left(\frac{1+\varepsilon}{\varepsilon}\right)^4 \sum_{j=\lceil\phi(n)\rceil}^{n} j^{4\alpha-4}\\ &\leqslant C\left(\frac{1+\varepsilon}{\varepsilon}\right)^4 \int_{\lceil\phi(n)\rceil-1}^{\infty} y^{4\alpha-4}dy \quad \leqslant C(1+\varepsilon)^4\varepsilon^{-4} n^{\chi(4\alpha-3)},\end{aligned}$$

as required. Next, we use a union bound and Chebyshev's inequality to prove (2.5) as follows:

$$\mathbb{P}_{\mathbf{x}}\Bigg[\bigcup_{j=\lceil\phi(n)\rceil}^{n} \{\mathcal{Z}_j \geqslant j^{1/2}\}\Bigg] \leqslant \sum_{j=\lceil\phi(n)\rceil}^{n} \mathbb{E}_{\mathbf{x}}[\mathcal{Z}_j^4] j^{-2} = \sum_{j=\lceil\phi(n)\rceil}^{n} j^{-2}\prod_{\ell=0}^{3}(x_j+\ell).$$

If we further assume $x_2 \in (0, \kappa]$, then there are positive numbers C' and C'' such that

$$\mathbb{P}_{\mathbf{x}}\Bigg[\bigcup_{j=\lceil\phi(n)\rceil}^{n} \{\mathcal{Z}_j \geqslant j^{1/2}\}\Bigg] \leqslant C'\kappa^4 \sum_{j=\lceil\phi(n)\rceil}^{n} j^{-2} \leqslant C'\kappa^4 \int_{\phi(n)-1}^{\infty} y^{-2}dy \leqslant C''\kappa^4 n^{-\chi},$$

hence proving (2.6). □

10. The urn representation of the sequential model conditional on a set of edges

Here we develop the urn representation for the $(\mathbf{x}, n)$-sequential model in Definition 1, but conditional on a set of edges. This is used in Section 3 to describe the offspring distributions of the type L and R parents. We first reproduce the proof of Theorem 3 in [38], as the argument for the $(\mathbf{x}, n)$-sequential model conditional on a set of edges is similar after some modifications.

10.1. Proof of Theorem 3

We use an urn embedding to prove Theorem 3; see e.g. [34, 38]. Let $G'_n \sim \mathrm{Seq}(\mathbf{x})_n$ be an $(\mathbf{x}, n)$-sequential model, and $D^{(in)}_{n,j}$ be the in-degree of vertex j in G'_n. Additionally, let $M_k(n) := \sum_{j=1}^{k}(x_j + D^{(in)}_{n,j})$ be the total *weight* of the first k vertices in G'_n, with $M_0(n) = 0$, and $U_k(n) = M_k(n) - M_{k-1}(n) = D^{(in)}_{n,k} + x_k$ being the weight of vertex k after n completed attachment steps, with $U_k(n) = 0$ whenever $k > n$. Furthermore, denote by $\mathrm{Polya}(b, w; n)$ the law of the number of white balls after the n-th draw in a classical Pólya urn, initially with w white balls and b black balls. The following lemma is an easy modification of [34, Lemma 2] that relates $U_k(n)$ to the number of white balls in a classical Pólya urn.

Lemma 18. *Retaining the notations above, let $T_k := \sum_{j=1}^{k} x_j$. Then given $n \geqslant 2$,*

$$U_{n-1}(n) \sim \mathrm{Polya}(T_{n-2} + n - 2, x_{n-1}; 1), \tag{10.1}$$

and conditional on $M_k(n)$ and the events $(\{U_j(n) = U_j(n-1)\}, k+1 \leqslant j \leqslant n-1)$,

$$U_k(n) \sim \mathrm{Polya}(T_{k-1} + k - 1, x_k; M_k(n) - T_k - k + 1). \tag{10.2}$$

Proof. To prove (10.1), note that when adding vertex n to the existing graph G'_{n-1}, the probability that vertex n sends an outgoing edge to $n-1$ is $x_{n-1}/(T_{n-1} + n - 2)$. This implies that $U_{n-1}(n)$ evolves like $\mathrm{Polya}(T_{n-2} + n - 2, x_{n-1}; 1)$.

For (10.2), given that one of the vertices in $[k]$ is chosen when adding vertex n to

G'_{n-1}, a straightforward computation, using the definition of the conditional probability and the PA rules, shows that the conditional probability that $U_k(n) = U_k(n-1)+1$ is $U_k(n-1)/M_k(n-1)$. This implies that $U_k(n)$ behaves like Polya$(b, w; m)$, where $b = T_{k-1} + k - 1$ is the total weight of the first $k-1$ vertices after k steps, $w = x_k$, and $m = M_k(n) - M_k(k) = M_k(n) - T_k - k + 1$ is the number of times the vertices $1, ..., k$ are picked after the k-th step. □

We now use Lemma 18 to prove Theorem 3. Recall that the subscript $\mathbf{x}$ in $\mathbb{P}_{\mathbf{x}}$ and $\mathbb{E}_{\mathbf{x}}$ indicates the conditioning on the fitness sequence $\mathbf{x}$.

Proof of Theorem 3. It is enough to consider the attachment steps $3 \leqslant m \leqslant n$, as the first two PA steps are deterministic. To add vertex m to G'_{m-1}, we consider the vertices in $[m-1]$ in the decreasing order of their labels. By Lemma 18, $U_{m-1}(m) \sim$ Polya$(T_{m-2} + m - 2, x_{m-1}; 1)$; and when $1 \leqslant k \leqslant m-2$, given that vertex m attaches to a vertex in $[k]$, $U_k(m) \sim$ Polya$(T_{k-1} + k - 1, x_k; M_k(m) - T_k - k + 1)$. Hence, for any $1 \leqslant k \leqslant m-1$, conditional on the event that vertex m attaches to one of the vertices in $[k]$, $U_k(m) = U_k(m-1)+1$ (resp. $U_k(m) = U_k(m-1)$) has the same probability as a trial in a sequence of conditionally independent Bernoulli variables with success probability B_k (resp. $1 - B_k$). We emphasise that B_k only depends on the initial attractiveness $(x_j, 1 \leqslant j \leqslant k)$ and not $M_k(m)$, since $M_k(m) - M_k(k)$ is the number of draws from the Pólya urn.

To conclude the proof, observe that

$$\begin{aligned}
&\mathbb{P}_{\mathbf{x}}[U_k(m) = U_k(m-1)+1|G'_{m-1}] \\
&= \mathbb{P}_{\mathbf{x}}\left[U_k(m) = U_k(m-1)+1, \bigcap_{i=k+1}^{m-1}\{U_i(m) = U_i(m-1)\}\middle| G'_{m-1}\right] \\
&= \mathbb{P}_{\mathbf{x}}\left[U_k(m) = U_k(m-1)+1\middle| \bigcap_{i=k+1}^{m-1}\{U_i(m) = U_i(m-1)\}, G'_{m-1}\right] \\
&\quad \times \prod_{h=k+1}^{m-2} \mathbb{P}_{\mathbf{x}}\left[U_h(m) = U_h(m-1)\middle| \bigcap_{i=h+1}^{m-1}\{U_i(m) = U_i(m-1)\}, G'_{m-1}\right] \\
&\quad \times \mathbb{P}_{\mathbf{x}}[U_{m-1}(m) = U_{m-1}(m-1)|G'_{m-1}].
\end{aligned}$$

Hence, given $(B_k, 2 \leqslant k \leqslant n)$, the edges in the $(\mathbf{x}, n)$-Pólya urn tree are independent, and the probability that vertex m attaches to vertex k in G'_n is $B_k \prod_{j=k+1}^{m-1}(1-B_j)$. Noting that

$$S_{n,k} - S_{n,k-1} = B_k \prod_{j=k+1}^{n} (1-B_j) \quad \text{and} \quad \frac{S_{n,k} - S_{n,k-1}}{S_{n,m-1}} = B_k \prod_{j=k+1}^{m-1} (1-B_j),$$

concludes the proof. □

10.2. Constructing the sequential model conditional on a set of edges

In this section, we study the distribution of the $(\mathbf{x}, n)$-sequential model when conditional on a finite collection of edges. Importantly, this model has an urn representation that is analogous to the $(\mathbf{x}, n)$-Pólya urn tree, thus enabling us to identify the offspring distribution of non-root vertices discovered in the BFS (Definition 6) of an $(\mathbf{x}, n)$-sequential model.

10.2.1. *The attachment rules* Given $\mathbf{x}$ and n, let $G'_n \sim \mathrm{Seq}(\mathbf{x})_n$ be an $(\mathbf{x}, n)$-sequential model. To specify the edges that we condition on, let $Q_{j\to k}$ be a zero-one variable that takes value one if vertex j sends an outgoing edge to k in G'_n, and $D^{(in)}_{n,k} := \sum_{j=k+1}^{n} Q_{j\to k}$ be the in-degree of vertex k in G'_n, with $D^{(in)}_{n,k} = 0$ if $k \geqslant n$. In view of the PA rules, if $Q_{k\to i} = 1$ for some $1 \leqslant i < k$, then $Q_{k\to j} = 0$ for any $j \neq i$. Furthermore, we write $\{i, j\}$ to denote the edge directed from j to $i < j$.

Now, let $\mathrm{V_d}$ be a subset of $V(G'_n)$ such that for any vertex $i \in \mathrm{V_d}$, its degree and the arrival times of *all* its neighbours are *determined*, hence the subscript d. Moreover, for each $i \in \mathrm{V_d}$, we require that $Q_{i\to k} = 1$ for some $k < i$ and $Q_{i\to j} = 0$ for all $j \neq k$, so that vertex k is the recipient of the only outgoing edge from vertex i. Denote $\mathrm{E_d}$ *all* the edges of the vertices in $\mathrm{V_d}$. Given $\mathrm{V_d}$ and $\mathrm{E_d}$, we investigate how conditioning on all the edges of the vertices in $\mathrm{V_d}$:

$$\mathscr{I} := \mathscr{I}(\mathrm{V_d}, \mathrm{E_d}) := \bigcap_{\substack{\{j,k\} \notin \mathrm{E_d};\\ j \in \mathrm{V_d}, j<k\leqslant n}} \{Q_{k\to j} = 0\} \cap \bigcap_{\substack{\{j,k\} \in \mathrm{E_d};\\ 1\leqslant j<k\leqslant n}} \{Q_{k\to j} = 1\},$$

changes the attachment rules of G'_n. Note that $\mathbb{P}_{\mathbf{x}}(\mathscr{I}) > 0$, since the vertices of $\mathrm{V_d}$ obey the attachment rules.

To prepare for the subsequent arguments, denote $\mathrm{V_s} := \{j \in [n]\backslash \mathrm{V_d} : \{i, j\}$ or $\{j, i\} \in$

$\mathrm{E_d}\}$ the vertices that are not in $\mathrm{V_d}$, but *shares* at least one edge in $\mathrm{E_d}$ with a vertex in $\mathrm{V_d}$, hence the subscript s in $\mathrm{V_s}$. We assume that $\mathrm{V_d} \cup \mathrm{V_s} \subset [n]$, as otherwise the whole tree can be constructed deterministically from the event $\mathscr{I}$. Moreover, let $\mathrm{u_d^{(o)}} := \mathrm{u_d^{(o)}}(\mathrm{V_d})$ (resp. $\mathrm{u_s^{(o)}} := \mathrm{u_s^{(o)}}(\mathrm{V_s})$) be the vertex in $\mathrm{V_d}$ (resp. $\mathrm{V_s}$) that has the earliest vertex arrival time:

$$\mathrm{u_d^{(o)}} := \min_{i \in \mathrm{V_d}} i, \quad \text{and} \quad \mathrm{u_s^{(o)}} := \min_{i \in \mathrm{V_s}} i, \tag{10.3}$$

where the superscript $^{(\mathrm{o})}$ stands for *oldest*. On the event $\mathscr{I}$, $\mathrm{u_s^{(o)}}$ must be a recipient of at least one incoming edge from the vertices in $\mathrm{V_d}$, and it does not send an outgoing edge to a vertex in $\mathrm{V_d}$. Because if $\mathrm{u_s^{(o)}}$ sends an outgoing edge to some vertex $i \in \mathrm{V_d}$, then by the definition of $\mathrm{V_s}$, there is some vertex $j < \mathrm{u_s^{(o)}}$ in $\mathrm{V_s}$ that receives the edge emanating from i. Since $\mathbb{P}_{\mathbf{x}}(\mathscr{I}) > 0$, i cannot send two outgoing edges to j and $\mathrm{u_s^{(o)}}$.

We also impose the following assumption on the collection of edges, which greatly simplifies the upcoming computation. As we discuss below, this assumption includes the case where the collection is the discovered vertices and edges in the BFS (Definition 6).

The outgoing edge sent by any vertex in $\mathrm{V_d} \setminus \{\mathrm{u_d^{(o)}}\}$ is received by another vertex in $\mathrm{V_d}$ and $\mathrm{u_d^{(o)}}, \mathrm{u_s^{(o)}} > 1$. $(\triangle)$

A moment's thought reveals that under $(\triangle)$, $\mathrm{u_d^{(o)}}$ must send an outgoing edge to $\mathrm{u_s^{(o)}}$, and any vertex in $\mathrm{V_s} \setminus \{\mathrm{u_s^{(o)}}\}$ must be sending an outgoing edge to a vertex in $\mathrm{V_d}$. To see how $(\triangle)$ fits into the BFS, we first prove a simple lemma. Recall the definitions of $v[t]$, $v^{(op)}[t]$ and $v^{(oa)}[t]$ in (3.1) and (3.3), and that $\mathcal{P}_{t-1}$ and $\mathcal{A}_{t-1}$ are the sets of probed and active vertices in the BFS.

Lemma 19. *Given $t \geqslant 2$, assume that $\mathcal{A}_{t-1} \cup \mathcal{P}_{t-1}$ does not contain vertex 1. Then the recipient of the incoming edge from vertex $i \in \mathcal{P}_{t-1}$ is also in $\mathcal{P}_{t-1}$, unless $i = v^{(op)}[t]$, in which case the recipient is $v^{(oa)}[t] \in \mathcal{A}_{t-1}$.*

Proof. The vertex we probe at exploration time $2 \leqslant j \leqslant t$, $v[j]$, either belongs to type L or R. If it belongs to type R, then it must have been discovered at some step $i < j$ via the outgoing edge it sends to $v[i] \in \mathcal{P}_{j-1}$. If $v[j]$ is a type L vertex, by Lemma 4, $v[s] = v^{(op)}[j+1]$ and vertex $v^{(oa)}[j+1] \in \mathcal{A}_j$ receives the incoming edge from $v[j]$.

□

In view of Lemma 19, $\mathrm{V_d}$ and $\mathrm{V_s}$ correspond to $\mathcal{P}_{t-1}$ and $\mathcal{A}_{t-1}$ for any $t \geqslant 2$ under the assumption ($\triangle$); noting that w.h.p. $\mathcal{P}_{t-1}$ and $\mathcal{A}_{t-1}$ do not contain vertex 1. It follows that $\mathrm{u_d^{(o)}}$ and $\mathrm{u_s^{(o)}}$ correspond to $v^{(op)}[t]$ and $v^{(oa)}[t]$ respectively. Under ($\triangle$), the event $\mathscr{I}$ can be written as

$$\begin{aligned} \mathscr{I} &= \{Q_{\mathrm{u_d^{(o)}} \to \mathrm{u_s^{(o)}}} = 1\} \cap \mathcal{K}(\mathrm{V}_d, \mathrm{E_d}) \qquad (10.4) \\ &:= \{Q_{\mathrm{u_d^{(o)}} \to \mathrm{u_s^{(o)}}} = 1\} \cap \bigcap_{\substack{\{j,k\} \notin \mathrm{E_d}; \\ j \in \mathrm{V_d}, j<k \leqslant n}} \{Q_{k\to j} = 0\} \cap \bigcap_{\substack{\{j,k\} \in \mathrm{E_d} \setminus \{\mathrm{u_s^{(o)}}, \mathrm{u_d^{(o)}}\}; \\ \mathrm{u_d^{(o)}} \leqslant j < k \leqslant n}} \{Q_{k\to j} = 1\} \end{aligned}$$

We now show that conditioning on $\mathscr{I}$, G'_n has a modified PA rules. Firstly, note that conditioning on $\mathscr{I}$ does not change the attachment rules for the first $\mathrm{u_s^{(o)}} - 1$ steps for constructing G'_n. Next, we study how it changes the rule for constructing G'_i from G'_{i-1} for $\mathrm{u_s^{(o)}} < i < \mathrm{u_d^{(o)}}$. Under the assumption ($\triangle$), the edges in $\mathrm{E_d} \setminus \{\mathrm{u_s^{(o)}}, \mathrm{u_d^{(o)}}\}$ are born after step $\mathrm{u_d^{(o)}}$, so they do not affect step $\mathrm{u_s^{(o)}} < i < \mathrm{u_d^{(o)}}$. Hence it is enough to consider how conditioning on the edge $\{\mathrm{u_s^{(o)}}, \mathrm{u_d^{(o)}}\}$ changes the rules of these attachment steps. The upcoming lemma is a variation of [36, Lemma 3.5]. We state the lemma in slightly greater generality, but is clearly applicable to our case by taking $\ell = \mathrm{u_s^{(o)}}$ and $p = \mathrm{u_d^{(o)}}$ in what follows. Choose any two positive integers ℓ and p, with $\ell < p$. The lemma below implies that given $Q_{p\to\ell} = 1$ and G'_{m-1}, where $\ell < m < p$, we attach vertex m to ℓ according to the same PA rule, but also include the edge $\{\ell, p\}$ in the vertex weight of ℓ. In other words, we can think of the initial attractiveness of vertex ℓ as $x_\ell + 1$ instead of x_ℓ.

Lemma 20. *Let $Q_{m\to j}$, G'_m and $D^{(in)}_{m,j}$ be as above and $T_m := \sum_{j=1}^{m} x_j$. Then for $j, \ell < m < p$,*

$$\mathbb{P}_{\mathbf{x}}[Q_{m\to j} = 1 | G'_{m-1}, Q_{p\to\ell} = 1] = \frac{D^{(in)}_{m-1,j} + x_j + \mathbb{1}[j = \ell]}{T_{m-1} + m - 1}. \qquad (10.5)$$

Proof. We follow the proof of [36, Lemma 3.5], but with some minor modifications.

First, we use the definition of conditional probability to write

$$\begin{aligned}&\mathbb{P}_{\mathbf{x}}[Q_{m\to j}=1|G'_{m-1},Q_{p\to\ell}=1]\\&\quad=\frac{\mathbb{P}_{\mathbf{x}}[Q_{m\to j}=1|G'_{m-1}]\mathbb{P}_{\mathbf{x}}[Q_{p\to\ell}=1|G'_{m-1},Q_{m\to j}=1]}{\mathbb{P}_{\mathbf{x}}[Q_{p\to\ell}=1|G'_{m-1}]}.\end{aligned}\tag{10.6}$$

We compute the probabilities above as follows. Evidently,

$$\mathbb{P}_{\mathbf{x}}[Q_{m\to j}=1|G'_{m-1}]=\frac{D^{(in)}_{m-1,j}+x_j}{T_{m-1}+m-2}.$$

This in turn implies that

$$\mathbb{P}_{\mathbf{x}}[Q_{p\to\ell}=1|G'_{m-1}]=\frac{\mathbb{E}_{\mathbf{x}}[D^{(in)}_{p-1,\ell}+x_\ell|G'_{m-1}]}{T_{p-1}+p-2},$$

and

$$\mathbb{P}_{\mathbf{x}}[Q_{p\to\ell}=1|G'_{m-1},Q_{m\to j}=1]=\frac{\mathbb{E}_{\mathbf{x}}[D^{(in)}_{p-1,\ell}+x_\ell|G'_{m-1},Q_{m\to j}=1]}{T_{p-1}+p-2}.$$

Moreover, Lemma 22 at the end of this section implies that for $\ell<m<p$,

$$\mathbb{E}_{\mathbf{x}}[D^{(in)}_{p-1,\ell}+x_\ell|G'_{m-1}]=(D^{(in)}_{m-1,\ell}+x_\ell)\prod_{h=m-1}^{p-2}\frac{T_h+h}{T_h+h-1}.$$

Note that

$$\mathbb{E}_{\mathbf{x}}[D^{(in)}_{m,\ell}+x_\ell|G'_{m-1},Q_{m\to j}=1]=D^{(in)}_{m-1,\ell}+x_\ell+\mathbb{1}[\ell=j],$$

and so by another application of Lemma 22,

$$\mathbb{E}_{\mathbf{x}}[D^{(in)}_{p-1,\ell}+x_\ell|G'_{m-1},Q_{m\to j}=1]=(D^{(in)}_{m-1,\ell}+x_\ell+\mathbb{1}[j=\ell])\prod_{h=m}^{p-2}\frac{T_h+h}{T_h+h-1}.$$

Applying these results to (10.6) and simplifying yields

$$\mathbb{P}_{\mathbf{x}}[Q_{m\to j}=1|G'_{m-1},Q_{p\to\ell}=1]=\frac{(D^{(in)}_{m-1,j}+x_j)(D^{(in)}_{m-1,\ell}+\mathbb{1}[j=\ell]+x_\ell)}{(T_{m-1}+m-1)(D^{(in)}_{m-1,\ell}+x_\ell)},\tag{10.7}$$

and (10.7) is equal to (10.5) by considering the cases $\ell = j$ and $\ell \neq j$ separately. □

Given that $\mathrm{u}_\mathrm{d}^{(\mathrm{o})} < m \leqslant n$ and $m \notin \mathrm{V}_\mathrm{s} \cup \mathrm{V}_\mathrm{d}$, we proceed to prove that on the event $\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})$, defined in (10.4), we attach vertex m to $j \in [m-1] \setminus \mathrm{V}_\mathrm{d}$ w.p. proportional to the weight of vertex j. Below we exclude the attachment steps of the vertices in $\mathrm{V}_\mathrm{s} \setminus \{\mathrm{u}_\mathrm{s}^{(\mathrm{o})}\}$ in the lemma, because under the assumption ($\triangle$), vertex $m \in \mathrm{V}_\mathrm{s} \setminus \{\mathrm{u}_\mathrm{s}^{(\mathrm{o})}\}$ necessarily sends an outgoing edge to a vertex in V_d. To precisely state the lemma, denote the sets of vertices and edges in V_d and E_d that are born before vertex m as

$$\mathrm{V}_{\mathrm{d},m} = \mathrm{V}_\mathrm{d} \cap [m-1] \quad \text{and} \quad \mathrm{E}_{\mathrm{d},m} = \{\{i,j\} \in \mathrm{E}_\mathrm{d} : j < m\}.$$

When compared to the BFS, $\mathrm{V}_{\mathrm{d},m}$ and $\mathrm{E}_{\mathrm{d},m}$ respectively correspond to $\mathcal{P}_{t,m}$ and $\mathcal{E}_{t,m}$ in (3.5) for any $t \geqslant 2$.

Lemma 21. *Retaining the notations in Lemma 20, let* V_d *and* E_d *be such that* ($\triangle$) *holds, with* $\mathrm{V}_\mathrm{d} \cup \mathrm{V}_\mathrm{s} \subset [n]$*. Let* $\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})$ *be as in (10.4). For* $m \in \{\mathrm{u}_\mathrm{d}^{(\mathrm{o})} + 1, ..., n\} \setminus (\mathrm{V}_\mathrm{s} \cup \mathrm{V}_\mathrm{d})$ *and* $j \in [m-1] \setminus \mathrm{V}_{\mathrm{d},m}$*,*

$$\mathbb{P}(Q_{m\to j} = 1 | G'_{m-1}, \mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})) = \frac{D_{m-1,j}^{(in)} + x_j}{T_{m-1} + m - 1 - \sum_{k \in \mathrm{V}_{\mathrm{d},m}} x_k - |\mathrm{E}_{\mathrm{d},m}|}. \tag{10.8}$$

Before proving the lemma, note that the edge count in the normalising constant in (10.8) is $m-1-|\mathrm{E}_{\mathrm{d},m}|$ instead of $m-2-|\mathrm{E}_{\mathrm{d},m}|$, as we need to include $\{\mathrm{u}_\mathrm{s}^{(\mathrm{o})}, \mathrm{u}_\mathrm{d}^{(\mathrm{o})}\} \in \mathrm{E}_{\mathrm{d},m}$. When attaching vertex m to vertex $j \in \mathrm{V}_\mathrm{s} \setminus \{\mathrm{u}_\mathrm{s}^{(\mathrm{o})}\}$, ($\triangle$) ensures that vertex j does not receive any incoming edges from the vertices of V_d, as otherwise these edges have a size-biasing effect on the initial attractiveness of vertex j, in that case (10.8) no longer holds.

Proof of Lemma 21. We use the definition of the conditional probability again, this time to rewrite the the left-hand side of (10.8) in terms of probabilities conditional on the events occurring before step m. For $m \in \{\mathrm{u}_\mathrm{d}^{(\mathrm{o})} + 1, ..., n\} \setminus (\mathrm{V}_\mathrm{s} \cup \mathrm{V}_\mathrm{d})$, let $\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d}) := \mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{<m} \cap \mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{\geqslant m}$, where $\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{<m}$ collects the events in

$\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})$ that occur between step $\mathrm{u}_\mathrm{d}^{(\mathrm{o})}$ and step m:

$$\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{<m} := \bigcap_{\{k,\ell\}\in \mathrm{E}_{\mathrm{d},m}\setminus\{\mathrm{u}_\mathrm{s}^{(\mathrm{o})},\mathrm{u}_\mathrm{d}^{(\mathrm{o})}\}} \{Q_{\ell\to k}=1\} \cap \bigcap_{\substack{\mathrm{u}_\mathrm{d}^{(\mathrm{o})}\leqslant k<\ell<m;\\ k\in \mathrm{V}_\mathrm{d}, \{k,\ell\}\notin \mathrm{E}_{\mathrm{d},m}}} \{Q_{\ell\to k}=0\};$$

and $\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{\geqslant m}$ collects the events that occur after or at step m:

$$\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{\geqslant m} := \bigcap_{\{k,\ell\}\in \mathrm{E}_\mathrm{d}\setminus \mathrm{E}_{\mathrm{d},m}} \{Q_{\ell\to k}=1\} \cap \bigcap_{\substack{(m\vee k)\leqslant \ell\leqslant n;\\ \ell\in \mathrm{V}_\mathrm{d}, \{k,\ell\}\notin \mathrm{E}_\mathrm{d}\setminus \mathrm{E}_{\mathrm{d},m}}} \{Q_{\ell\to k}=0\}.$$

Hence we have the following expression:

$$\begin{aligned}&\mathbb{P}_\mathbf{x}[Q_{m\to j}=1|\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d}), G'_{m-1}]\\ &\quad= \frac{\mathbb{P}_\mathbf{x}[Q_{m\to j}=1|\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{<m}, G'_{m-1}]\mathbb{P}_\mathbf{x}[\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{\geqslant m}|Q_{m\to j}=1, \mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{<m}, G'_{m-1}]}{\mathbb{P}_\mathbf{x}[\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{\geqslant m}|\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{<m}, G'_{m-1}]}.\end{aligned}$$

It is straightforward to see that for $j\notin \mathrm{V}_\mathrm{d}$,

$$\mathbb{P}_\mathbf{x}[Q_{m\to j}=1|\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{<m}, G'_{m-1}] = \frac{D^{(in)}_{m-1,j}+x_j}{T_{m-1}+m-2}.$$

Using the PA rules and simplifying, we can compute

$$\frac{\mathbb{P}_\mathbf{x}[\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{\geqslant m}|\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{<m}, G'_{m-1}]}{\mathbb{P}_\mathbf{x}[\mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{\geqslant m}|Q_{m\to j}=1, \mathcal{K}(\mathrm{V}_d, \mathrm{E}_\mathrm{d})_{<m}, G'_{m-1}]} = \frac{T_{m-1}+m-1-\sum_{k\in \mathrm{V}_{\mathrm{d},m}} x_k - |\mathrm{E}_{\mathrm{d},m}|}{T_{m-1}+m-2}.$$

Combining the last equations completes the proof. □

The following lemma is applied in the proof of Lemma 20, which is a slight modification of Lemma 4.1 in [33] and Theorem 2.1 of [29].

Lemma 22. *Retaining the notations in Lemma 20, let k, ℓ and m be positive integers such that $k\leqslant \ell\leqslant m$, then*

$$\mathbb{E}_\mathbf{x}[D^{(in)}_{m,k}+x_k|G'_\ell] = (D^{(in)}_{\ell,k}+x_k)\prod_{j=\ell}^{m-1}\frac{T_j+j}{T_j+j-1}.$$

Proof. At attachment step $m>k$, $D^{(in)}_{m,k}$ either increases by exactly one or stays the same, and $D^{(in)}_{m+1,k} = D^{(in)}_{m,k} + 1$ w.p. proportional to $D^{(in)}_{m,k} + x_k$. Hence,

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}}[D^{(in)}_{m,k} + x_k | G'_{m-1}] \\
&= \frac{D^{(in)}_{m-1,k} + x_k}{T_{m-1} + m - 2}(D^{(in)}_{m-1,k} + 1 + x_k) + \left(1 - \frac{D^{(in)}_{m-1,k} + x_k}{T_{m-1} + m - 2}\right)(D^{(in)}_{m-1,k} + x_k) \\
&= \frac{T_{m-1} + m - 1}{T_{m-1} + m - 2}(D^{(in)}_{m-1,k} + x_k).
\end{aligned}$$

The lemma follows from iterating the result above. □

10.2.2. *The model construction* Assume that $\mathrm{V_d}$ and $\mathrm{E_d}$ are such that ($\triangle$) holds, with $\mathrm{V_d} \cup \mathrm{V_s} \subset [n]$. Let $\mathscr{I}$ be as in (10.4), and $G_n(\mathscr{I})$ be a graph that is distributed as $\mathrm{Seq}(\mathbf{x})_n$ conditional on $\mathscr{I}$, where $\mathrm{Seq}(\mathbf{x})_n$ is as in Definition 1. Lemma 20 and 21 imply that the random variables $D^{(in)}_{n,j}(\mathscr{I}) := (D^{(in)}_{n,j} | \mathscr{I})$ can be generated as follows, and has a similar flavour as the construction in [36]. Let $\mathrm{u}^{(\mathrm{o})}_{\mathrm{d}}$ and $\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}$ be as in (10.3). Initially, we generate $G_{\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}-1}(\mathscr{I})$ using the usual attachment rules. At step $\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}$, add the vertices $\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}$ and $\mathrm{u}^{(\mathrm{o})}_{\mathrm{d}}$ to $G_{\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}-1}(\mathscr{I})$, such that $\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}$ receives an incoming edge from $\mathrm{u}^{(\mathrm{o})}_{\mathrm{d}}$. Then, the recipient $j \in [\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}} - 1]$ of the outgoing edge of $\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}$ is chosen w.p. proportional to $D^{(in)}_{\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}-1,j}(\mathscr{I}) + x_j$. After the attachment step, assign vertex $\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}$ the initial attractiveness $x_{\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}}$, and set $D^{(in)}_{\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}},\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}}(\mathscr{I}) = 1$. At step $\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}} < m < \mathrm{u}^{(\mathrm{o})}_{\mathrm{d}}$, vertex m sends an outgoing edge to vertex $j \in [m-1]$, w.p.

$$\frac{D^{(in)}_{m-1,j}(\mathscr{I}) + x_j}{T_{m-1} + m - 1},$$

and we equip vertex m with the initial attractiveness x_m. In step $\mathrm{u}^{(\mathrm{o})}_{\mathrm{d}}$, assign vertex $\mathrm{u}^{(\mathrm{o})}_{\mathrm{d}}$ the initial attractiveness $x_{\mathrm{u}^{(\mathrm{o})}_{\mathrm{d}}}$. At step $m \in \{\mathrm{u}^{(\mathrm{o})}_{\mathrm{d}} + 1, \ldots, n\} \setminus (\mathrm{V_s} \cup \mathrm{V_d})$, vertex m sends an outgoing edge to vertex $j \in [m-1] \setminus \mathrm{V}_{\mathrm{d},m}$ w.p.

$$\frac{D^{(in)}_{m-1,j}(\mathscr{I}) + x_j}{T_{m-1} + m - 1 - \sum_{k \in \mathrm{V}_{\mathrm{d},m}} x_k - |\mathrm{E}_{\mathrm{d},m}|}.$$

At steps m such that $\{j, m\} \in \mathrm{E_d} \setminus \{\mathrm{u}^{(\mathrm{o})}_{\mathrm{s}}, \mathrm{u}^{(\mathrm{o})}_{\mathrm{d}}\}$, we draw an edge directed from vertex m to j, and set the initial attractiveness of vertex m as x_m.

10.2.3. *The urn representation* Using an urn argument analogous to that of Theorem 3, below we give an alternative definition of $G_n(\mathscr{I})$ that has conditionally independent edges. To simplify notation, we drop $\mathbf{x}$ in the definitions of the variables below, which depend on the sequence of initial attractiveness.

Definition 11. (*($\mathbf{x}, \mathscr{I}, n$)-Pólya urn tree.*) Given the sequence $\mathbf{x}$, and that $\mathrm{V_d}$ and $\mathrm{E_d}$ are such that ($\triangle$) holds, with $\mathrm{V_d} \cup \mathrm{V_s} \subset [n]$. Let $\mathscr{I}$ be as in (10.4), and $(B_j(\mathscr{I}), j \in [n])$ be conditionally independent random variables such that $B_1(\mathscr{I}) := 1$, $B_j(\mathscr{I}) := 0$ if $j \in \mathrm{V_d}$, and for $j \notin \mathrm{V_d}$,

$$B_j(\mathscr{I}) \sim \begin{cases} \mathrm{Beta}(x_j + \mathbb{1}[j = \mathrm{u_s^{(o)}}], T_{j-1} + j - 1), & 2 \leqslant j \leqslant \mathrm{u_s^{(o)}}; \\ \mathrm{Beta}(x_j, T_{j-1} + j), & \mathrm{u_s^{(o)}} < j < \mathrm{u_d^{(o)}}; \\ \mathrm{Beta}(x_j, T_{j-1} + j - \sum_{k \in \mathrm{V}_{\mathrm{d},j}} x_k - |\mathrm{E}_{\mathrm{d},j}|), & \mathrm{u_d^{(o)}} < j \leqslant n. \end{cases}$$

Furthermore, let $S_{0,n}(\mathscr{I}) := 0$, $S_{n,n}(\mathscr{I}) := 1$ and

$$S_{n,j}(\mathscr{I}) := \prod_{i=j+1}^{n} (1 - B_i(\mathscr{I})) \quad \text{for } 1 \leqslant j \leqslant n-1.$$

Starting with n vertices and the edges in $\mathrm{E_d}$, we connect the vertices as follows. Let $I_j = [S_{n,j-1}(\mathscr{I}), S_{n,j}(\mathscr{I}))$ for $j \in [n] \setminus \mathrm{V_d}$. Conditionally on $(S_{n,j}(\mathscr{I}), j \in [n-1])$, we generate $U_j \sim \mathrm{U}[0, S_{n,j-1}(\mathscr{I})]$ for $j \in \{2, ..., n\} \setminus (\mathrm{V_d} \cup \mathrm{V_s} \setminus \{\mathrm{u_s^{(o)}}\})$. If $j < k$ and $U_k \in I_j$, we attach vertex k to j. We say that the resulting graph is an $(\mathbf{x}, \mathscr{I}, n)$-Pólya urn graph.

Theorem 4. *Assume that* $\mathrm{V_d}$ *and* $\mathrm{E_d}$ *are such that* ($\triangle$) *holds. If* $\widetilde{G}_n(\mathscr{I})$ *is an* $(\mathbf{x}, \mathscr{I}, n)$*-Pólya urn tree, it has the same distribution as* $\mathrm{Seq}(\mathbf{x})_n$ *conditional on* $\mathscr{I}$*.*

The proof of Theorem 4 is similar to that of Theorem 3. Firstly, we need notation analogous to those in Lemma 18. Let $G_n(\mathscr{I})$ and $D^{(in)}_{n,k}(\mathscr{I})$ be as in Section 10.2.2. Define the total weight of the vertices $[k] \setminus \mathrm{V}_{\mathrm{d},k}$ and the weight of vertex $k \notin \mathrm{V_d}$ in $G_m(\mathscr{I})$ as

$$M'_k(m) = \sum_{j \in [k] \setminus \mathrm{V}_{\mathrm{d},k}} (x_i + D^{(in)}_{m,j}(\mathscr{I})), \qquad U'_k(m) = M'_k(m) - M'_{k-1}(m),$$

and for $k \in \mathrm{V_d}$, $U'_k(m) = U'_k(m-1)+1$ if and only if $\{k,m\} \in \mathrm{E_d}$.

For the discussion and the proof below, suppose that $m \notin \mathrm{V_d} \cup \mathrm{V_s} \setminus \{\mathrm{u_s^{(o)}}\}$. It is clear from the construction of $G_m(\mathscr{I})$ that we can obtain an analogue of Lemma 18 for $U'_k(m)$. The differences here are that we have to use the adjusted total vertex weights, and vertex m can only attach to a vertex in $[m-1] \setminus \mathrm{V}_{\mathrm{d},m}$. If $m-1 \notin \mathrm{V_d}$, arguing as for Lemma 18, we have $U'_{m-1}(m) \sim \mathrm{Polya}(b, w; 1)$, where $w = x_{m-1} + \mathbb{1}[m = \mathrm{u_s^{(o)}}+1]$ and b is the total weight of the vertices $[m-1] \setminus \mathrm{V}_{\mathrm{d},m-1}$ with

$$b = T_{m-2} + m - 2 \quad \text{if } m \leqslant \mathrm{u_s^{(o)}},$$
$$b = T_{m-2} + m - 2 + \mathbb{1}[m \neq \mathrm{u_s^{(o)}}+1] \quad \text{if } \mathrm{u_s^{(o)}} < m < \mathrm{u_d^{(o)}},$$
$$b = T_{m-2} + m - 1 - \sum_{j \in \mathrm{V}_{\mathrm{d},m-1}} x_j - |\mathrm{E}_{\mathrm{d},m-1}| \quad \text{if } \mathrm{u_d^{(o)}} < m \leqslant n \text{ and } (m-1) \notin \mathrm{V_d}.$$

Let $k \in [m-1] \setminus \mathrm{V}_{\mathrm{d},m}$. Conditional on $M'_k(m)$ and the event that vertex m does not attach to any vertex in $\{k+1, ..., m-1\} \setminus \mathrm{V}_{\mathrm{d},m-1}$, $(\{U'_j(m) = U'_j(m-1)\}, j \in \{k+1, ..., m-1\} \setminus \mathrm{V}_{\mathrm{d},m-1})$, we have $U'_k(m) \sim \mathrm{Polya}(b', w'; q')$, where $q' = M'_k(m) - b' - w'$ is the number of draws in the Pólya urn. The parameters are $w' = x_k + \mathbb{1}[k = \mathrm{u_s^{(o)}}]$,

$$b' = T_{k-1} + k - 1, \quad q' = M'_k(m) - T_k - k + 1, \quad \text{for } m \leqslant \mathrm{u_s^{(o)}};$$
$$b' = T_{k-1} + k - 1 + \mathbb{1}[k > \mathrm{u_s^{(o)}}], \quad q' = M'_k(m) - T_k - k + 1 - \mathbb{1}[k \geqslant \mathrm{u_s^{(o)}}], \quad \text{for } \mathrm{u_s^{(o)}} < m < \mathrm{u_d^{(o)}};$$

and for $m > \mathrm{u_d^{(o)}}$,

$$b' = T_{k-1} + k - 1 - \sum_{j \in \mathrm{V}_{\mathrm{d},k}} x_j - |\mathrm{E}_{\mathrm{d},k}| + \mathbb{1}[k > \mathrm{u_s^{(o)}}],$$
$$q' = M'_k(m) - T_k - k + 1 + \sum_{j \in \mathrm{V}_{\mathrm{d},k}} x_j + |\mathrm{E}_{\mathrm{d},k}| - \mathbb{1}[k \geqslant \mathrm{u_s^{(o)}}].$$

Proof of Theorem 4. We only consider the attachment step $m \notin \mathrm{V_d} \cup \mathrm{V_s} \setminus \{\mathrm{u_s^{(o)}}\}$, because step $m \in \mathrm{V_d} \cup \mathrm{V_s} \setminus \{\mathrm{u_s^{(o)}}\}$ is deterministic under the assumption ($\triangle$). To prove the theorem, we replace $U_k(n)$ in the proof of Theorem 3 by $U'_k(n)$ for $k \notin \mathrm{V_d}$ and argue similarly. Note that for $k \in \mathrm{V_d}$ and $m \notin \mathrm{V_d} \cup \mathrm{V_s} \setminus \{\mathrm{u_s^{(o)}}\}$, $U'_k(m) = U'_k(m-1)$ is reflected by $B_k(\mathscr{I}) = 0$ in the construction of the $(\mathbf{x}, \mathscr{I}, n)$-Pólya urn graph. $\square$

With these preparations, it is now possible to prove Lemma 5 and 6.

Proof of Lemma 5 and 6. Apply Theorem 3 for $t = 1$. For $t \geqslant 2$, take $\mathrm{V_d}$, $\mathrm{V_s}$, $\mathrm{u_d^{(o)}}$ and $\mathrm{u_s^{(o)}}$ in Theorem 4 respectively as $\mathcal{P}_{t-1}$, $\mathcal{A}_{t-1}$, $v^{(op)}[t]$ and $v^{(oa)}[t]$ in Definition 6, and $\mathrm{E_d}$ as $\mathcal{E}_{t-1}$, which is the set of edges joining the vertices in $\mathcal{P}_{t-1}$ and $\mathcal{A}_{t-1}$. □

11. Supplementary proofs

In this section, we prove some of the lemmas in Section 5 and 6, as well as Theorem 2. Recall that the subscript $\mathbf{x}$ in $\mathbb{P}_{\mathbf{x}}$ and $\mathbb{E}_{\mathbf{x}}$ indicates the conditioning on $\mathbf{X} = \mathbf{x}$.

11.1. Proof of Lemma 9

In preparation, we use the probabilities $\widehat{P}_{j\to v[1]}$ in (5.6) to construct a Bernoulli sequence and a Poisson sequence that appear in the intermediate coupling steps below. Recall that $U_0 \sim \mathrm{U}[0,1]$, $v[1] := k_0^{(\mathrm{u})} = \lceil nU_0 \rceil$ is the uniformly chosen vertex in the $(\mathbf{x}, n)$-Pólya urn tree (Definition 4), $a_0^{(\mathrm{i})} = U_0^{\chi}$ is the age of vertex 0 in the intermediate Pólya point tree $(\mathcal{T}_{\mathbf{x},n}, 0)$ (Definition 8), $\mathcal{Z}_j[1]$ and $\widetilde{\mathcal{Z}}_j[1]$ are the gamma variables in (3.6). Define $\zeta_0 := \mathcal{Z}_{v[1]}[1]$ and

$$\Xi_{\mathbf{x}} := (U_0, ((\mathcal{Z}_j[1], \widetilde{\mathcal{Z}}_j[1]), 2 \leqslant j \leqslant n)); \tag{11.1}$$

noting that the event $\bigcap_{j=1}^{3} F_{1,j} \cap \mathcal{H}_{1,0}$, defined in (5.2) and (5.5), is measurable with respect to $\Xi_{\mathbf{x}}$.

Definition 12. Given $a_0^{(\mathrm{i})}$, $v[1]$ and ζ_0, let $\widehat{Y}_{j\to v[1]}$, $v[1]+1 \leqslant j \leqslant n$, be conditionally independent Bernoulli variables, each with parameter $\widehat{P}_{j\to v[1]}$. Define a Bernoulli sequence by the random vector

$$\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[1],n)} := (\widehat{Y}_{(v[1]+1)\to v[1]}, \widehat{Y}_{(v[1]+2)\to v[1]}, ..., \widehat{Y}_{n\to v[1]}).$$

Definition 13. Given $a_0^{(\mathrm{i})}$, $v[1]$ and ζ_0, let $\widehat{V}_{j\to v[1]}$, $v[1]+1 \leqslant j \leqslant n$, be conditionally independent Poisson random variables, each with parameter $\widehat{P}_{j\to v[1]}$. Define a Poisson sequence by the random vector

$$\widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[1],n)} := (\widehat{V}_{(v[1]+1)\to v[1]}, \widehat{V}_{(v[1]+2)\to v[1]}, ..., \widehat{V}_{n\to v[1]}).$$

We also require two lemmas for our proofs. The first is a simple result that turns the problem of coupling two random vectors into the problem of coupling two random variables. The proof is an easy application of the union bound.

Lemma 23. *Given a positive integer d, let $\mathbf{V} = (V_1, \dots, V_d)$ and $\mathbf{W} = (W_1, ..., W_d)$ be vectors of independent random variables. Let $(\widetilde{V}_i, \widetilde{W}_i)$ be a coupling of V_i and W_i, where $(\widetilde{V}_i, \widetilde{W}_i)$ is independent of $(\widetilde{V}_j, \widetilde{W}_j)$ for any $i \neq j$. Denote $\widetilde{\mathbf{V}} = (\widetilde{V}_1, ..., \widetilde{V}_d)$ and $\widetilde{\mathbf{W}} = (\widetilde{W}_1, ..., \widetilde{W}_d)$, so that $(\widetilde{\mathbf{V}}, \widetilde{\mathbf{W}})$ is a coupling of $\mathbf{V}$ and $\mathbf{W}$. Then,*

$$\mathbb{P}[\widetilde{\mathbf{V}} \neq \widetilde{\mathbf{W}}] = \mathbb{P}\Big[\bigcup_{i=1}^{d} \{\widetilde{V}_i \neq \widetilde{W}_i\}\Big] \leqslant \sum_{i=1}^{d} \mathbb{P}[\widetilde{V}_i \neq \widetilde{W}_i].$$

The next lemma shows that for three random vectors, it is enough to consider the pairwise couplings, and is easily extended to the more general cases.

Lemma 24. *Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ be random vectors taking values in $\mathbb{R}^d$ for some positive integer d. Suppose that there is a coupling $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$ of $\mathbf{X}$ and $\mathbf{Y}$, and a coupling $(\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}})$ of $\mathbf{Y}$ and $\mathbf{Z}$. Then there is a coupling $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ such that*

$$\mathbb{P}[\mathbf{X}' \neq \mathbf{Z}'] \leqslant \mathbb{P}[\widetilde{\mathbf{X}} \neq \widetilde{\mathbf{Y}}] + \mathbb{P}[\widehat{\mathbf{Y}} \neq \widehat{\mathbf{Z}}].$$

Proof. We construct the coupling $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ as follows. Let (X', Y') be the same coupling as $(\widetilde{X}, \widetilde{Y})$. Due to the existence of regular conditional probability measures [14, Theorem 4.1.18, p. 186], we may define $\mathscr{L}(\mathbf{Z}'|\mathbf{Y}', \mathbf{X}') = \mathscr{L}(\widehat{\mathbf{Z}}|\widehat{\mathbf{Y}})$, so that $\mathbf{Z}'$ is independent of $\mathbf{X}'$ when given $\mathbf{Y}'$. Then, by a union bound,

$$\begin{aligned}
\mathbb{P}[\mathbf{X}' \neq \mathbf{Z}'] &= \mathbb{P}[\{\mathbf{X}' \neq \mathbf{Y}'\} \cup \{\mathbf{Y}' \neq \mathbf{Z}'\}] &(11.2)\\
&\leqslant \mathbb{P}[\mathbf{X}' \neq \mathbf{Y}'] + \mathbb{P}[\mathbf{Y}' \neq \mathbf{Z}'] &(11.3)\\
&= \mathbb{P}[\widetilde{\mathbf{X}} \neq \widetilde{\mathbf{Y}}] + \mathbb{P}[\widehat{\mathbf{Y}} \neq \widehat{\mathbf{Z}}]. &(11.4)
\end{aligned}$$

□

The proof of Lemma 9 consists of two main components. The first is to use Lemma 23 and standard techniques to couple $(\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)}, \widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[1],n)})$, $(\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[1],n)}, \widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[1],n)})$ and $(\widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[1],n)}, \mathbf{V}_{\mathrm{Po}}^{(v[1],n)})$ under the event $\bigcap_{i=1}^{3} F_{1,i} \cap \mathcal{H}_{1,0}$, where $\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)}$ and $\mathbf{V}_{\mathrm{Po}}^{(v[1],n)}$

are as in Definition 7 and 9. These results are given in the next three lemmas. The second is to combine them using Lemma 24.

Lemma 25. *Let $\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)}$ and $\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[1],n)}$ be as in Definition 7 and 12, and the events $\mathcal{H}_{1,0}$, $F_{1,i}$, $i=1,2,3$ and $\Xi_{\mathbf{x}}$ be as in (5.1), (5.5) and (11.1). We can couple the random vectors so that on the event $\bigcap_{i=1}^{3} F_{1,i} \cap \mathcal{H}_{1,0}$, there is a positive constant $C := C(x_1, \mu)$ such that*

$$\mathbb{P}_{\mathbf{x}}\big[\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)} \neq \widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[1],n)} \big| \Xi_{\mathbf{x}}\big] \leqslant C\zeta_0 n^{-\gamma} (\log\log n)^{1-\chi} \quad \textit{for all } n \geqslant 3. \tag{11.5}$$

Proof. We first show that on the event $\bigcap_{i=1}^{3} F_{1,i} \cap \mathcal{H}_{1,0}$, there is a positive constant $C := C(x_1, \mu)$ such that for $j > v[1]$,

$$(1 - Cn^{-\gamma})\widehat{P}_{j\to v[1]} \leqslant P_{j\to v[1]} \leqslant (1 + Cn^{-\gamma})\widehat{P}_{j\to v[1]}, \tag{11.6}$$

and then couple the random vectors. We only prove the upper bound in (11.6), as the lower bound follows from a similar calculation. Pick n large enough so that $C^* n^{-\chi/12} (\log\log n)^{\chi} < 1/2$, where C^* is the constant in the event $F_{1,1}$. Since $v[1] > n(\log\log n)^{-1}$ on the event $\mathcal{H}_{1,0}$, for $j \geqslant v[1]$,

$$\begin{aligned}
\frac{S_{n,v[1]}[1]}{S_{n,j}[1]} &\leqslant \left\{ \left(\frac{v[1]}{n}\right)^{\chi} + C^* n^{-\chi/12} \right\} \left\{ \left(\frac{j}{n}\right)^{\chi} - C^* n^{-\chi/12} \right\}^{-1} \\
&\leqslant \left(\frac{n}{j}\right)^{\chi} \left\{ \left(\frac{v[1]}{n}\right)^{\chi} + C^* n^{-\chi/12} \right\} \left\{ 1 - C^* n^{-\chi/12} (\log\log n)^{\chi} \right\}^{-1} \\
&\leqslant \left\{ \left(\frac{v[1]}{j}\right)^{\chi} + C^* n^{-\chi/12} (\log\log n)^{\chi} \right\} \sum_{i\geqslant 0} (C^* n^{-\chi/12} (\log\log n)^{\chi})^i \\
&\leqslant \left\{ \left(\frac{v[1]}{j}\right)^{\chi} + C^* n^{-\chi/12} (\log\log n)^{\chi} \right\} \{1 + 2C^* n^{-\chi/12} (\log\log n)^{\chi}\}
\end{aligned}$$

on the event $F_{1,1} \cap \mathcal{H}_{1,0}$, where we have used $(n/j)^{\chi} \leqslant (n/v[1])^{\chi} < (\log\log n)^{\chi}$ and the geometric series. Thus, there is a positive constant $C' := C'(x_1, \mu)$ such that

$$\frac{S_{n,v[1]}[1]}{S_{n,j-1}[1]} \leqslant \left(\frac{v[1]}{j}\right)^{\chi} + C' n^{-\frac{\chi}{12}} (\log\log n)^{\chi} \quad \text{for } v[1] < j \leqslant n.$$

Hence, on the event $\bigcap_{j=1}^{3} F_{1,j} \cap \mathcal{H}_{1,0}$, we can bound $P_{j\to v[1]}$ in terms of $\widehat{P}_{j\to v[1]}$:

$$\begin{aligned} P_{j\to v[1]} &\leqslant \left[\frac{\zeta_0}{(\mu+1)v[1]} + \frac{\zeta_0 n^{-\gamma}}{(\mu+1)v[1]}\right]\left[\left(\frac{v[1]}{j}\right)^{\chi} + C' n^{-\frac{\chi}{12}} (\log\log n)^{\chi}\right] \\ &= \widehat{P}_{j\to v[1]}\left[1 + n^{-\gamma} + \left(\frac{j}{v[1]}\right)^{\chi} C' (\log\log n)^{\chi} (n^{-\gamma-\frac{\chi}{12}} + n^{-\frac{\chi}{12}})\right]. \end{aligned} \tag{11.7}$$

Using $(j/v[1])^{\chi} \leqslant (n/v[1])^{\chi} \leqslant (\log\log n)^{\chi}$ and $0 < \gamma < \chi/12$, we deduce that there is a constant $C := C(x_1, \mu)$ such that

$$\begin{aligned} P_{j\to v[1]} &\leqslant \widehat{P}_{j\to v[1]}(1 + n^{-\gamma} + C' n^{-\gamma-\frac{\chi}{12}} (\log\log n)^{3\chi} + C' n^{-\frac{\chi}{12}} (\log\log n)^{3\chi}) \\ &= \widehat{P}_{j\to v[1]}(1 + Cn^{-\gamma}). \end{aligned}$$

For the coupling, we use independent, standard uniform variables U_j, $v[1]+1 \leqslant j \leqslant n$ to define

$$Y_{j\to v[1]} = \mathbb{1}[U_j \leqslant \ P_{j\to v[1]}] \quad \text{and} \quad \widehat{Y}_{j\to v[1]} = \mathbb{1}[U_j \leqslant \widehat{P}_{j\to v[1]}].$$

It follows that on the event $\bigcap_{j=1}^{3} F_{1,j} \cap \mathcal{H}_{1,0}$,

$$\mathbb{P}_{\mathbf{x}}\big[Y_{j\to v[1]} \neq \widehat{Y}_{j\to v[1]} \big| \Xi_{\mathbf{x}}\big] \leqslant Cn^{-\gamma} \widehat{P}_{j\to v[1]}.$$

By Lemma 23, we have

$$\mathbb{P}_{\mathbf{x}}\big[\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)} \neq \widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[1],n)} \big| \Xi_{\mathbf{x}}\big] \leqslant Cn^{-\gamma} \sum_{j=v[1]+1}^{n} \widehat{P}_{j\to v[1]}$$

on the event $\bigcap_{j=1}^{3} F_{1,j} \cap \mathcal{H}_{1,0}$. To bound the sum above, we use $v[1] > n(\log\log n)^{-1}$ and an integral comparison to get

$$\begin{aligned} \mathbb{1}\left[\bigcap_{i=1}^{3} F_{1,i} \cap \mathcal{H}_{1,0}\right] \sum_{j=v[1]+1}^{n} \widehat{P}_{j\to v[1]} &= \mathbb{1}\left[\bigcap_{i=1}^{3} F_{1,i} \cap \mathcal{H}_{1,0}\right] \frac{\zeta_0}{(\mu+1)k_0^{1-\chi}} \sum_{j=v[1]+1}^{n} j^{-\chi} \\ &\leqslant \frac{\zeta_0}{(\mu+1)} \left[\frac{\log\log n}{n}\right]^{1-\chi} \int_{n(\log\log n)^{-1}}^{n} y^{-\chi} dy \\ &\leqslant \zeta_0 (\log\log n)^{1-\chi}. \end{aligned} \tag{11.8}$$

Combining the last two inequalities gives (11.5). □

Lemma 26. *Let* $\widehat{\mathbf{Y}}^{(v[1],n)}_{\mathrm{Be}}$ *and* $\widehat{\mathbf{V}}^{(v[1],n)}_{\mathrm{Po}}$ *be as in Definition 12 and 13, and the events* $\mathcal{H}_{1,0}$*,* $F_{1,i}$*,* $i = 1,2,3$ *and* $\Xi_{\mathbf{x}}$ *be as in (5.1), (5.5) and (11.1). We can couple the random vectors so that on the event* $\bigcap_{i=1}^{3} F_{1,i} \cap \mathcal{H}_{1,0}$*,*

$$\mathbb{P}_{\mathbf{x}}\big[\widehat{\mathbf{Y}}^{(v[1],n)}_{\mathrm{Be}} \neq \widehat{\mathbf{V}}^{(v[1],n)}_{\mathrm{Po}} \big| \Xi_{\mathbf{x}}\big] \leqslant \frac{\zeta_0^2(\log\log n)^{2-\chi}}{(\mu+1)n} \quad \textit{for all } n \geqslant 3.$$

Proof. First, we use the standard Poisson-Bernoulli coupling [24, equation (1.11), p. 5] to couple $Y_{j\to v[1]}$ and $\widehat{Y}_{j\to v[1]}$. Then, applying Lemma 23 and $(v[1]/j)^{\chi} \leqslant 1$, we obtain

$$\mathbb{P}_{\mathbf{x}}\big[\widehat{\mathbf{Y}}^{(v[1],n)}_{\mathrm{Be}} \neq \widehat{\mathbf{V}}^{(v[1],n)}_{\mathrm{Po}} \big| \Xi_{\mathbf{x}}\big] \leqslant \sum_{j=v[1]+1}^{n} \widehat{P}^2_{j\to v[1]} \leqslant \frac{\zeta_0}{(\mu+1)v[1]} \sum_{j=v[1]+1}^{n} \widehat{P}_{j\to v[1]},$$

so the lemma follows from applying (11.8) to the sum above, and noting that $v[1] > n(\log\log n)^{-1}$ on the event $\mathcal{H}_{1,0}$. □

Lemma 27. *Let* $\mathbf{V}^{(v[1],n)}_{\mathrm{Po}}$ *and* $\widehat{\mathbf{V}}^{(v[1],n)}_{\mathrm{Po}}$ *be as in Definition 9 and 13, and the events* $\mathcal{H}_{1,0}$*,* $F_{1,i}$*,* $i = 1,2,3$ *and* $\Xi_{\mathbf{x}}$ *be as in (5.1), (5.5) and (11.1). We can couple the random vectors so that on the event* $\bigcap_{i=1}^{3} F_{1,i} \cap \mathcal{H}_{1,0}$*, there is a positive constant* $C := C(\mu)$ *such that*

$$\mathbb{P}_{\mathbf{x}}\big[\mathbf{V}^{(v[1],n)}_{\mathrm{Po}} \neq \widehat{\mathbf{V}}^{(v[1],n)}_{\mathrm{Po}} \big| \Xi_{\mathbf{x}}\big] \leqslant \frac{C\log\log n}{n}\zeta_0 \quad \textit{for all } n \geqslant 3.$$

Proof. We use the monotone coupling to couple $V_{j\to v[1]}$ and $\widehat{V}_{j\to v[1]}$ for each j, with the convention that the value zero is a Poisson variable with mean zero. Let $\nu_j := \lambda_j^{[1]} \wedge \widehat{P}_{j\to v[1]}$ and

$$V'_{j\to v[1]} \sim \mathrm{Po}(\nu_j), \quad V''_{j\to v[1]} \sim \mathrm{Po}(|\lambda_j^{[1]} - \widehat{P}_{j\to v[1]}|), \quad V'''_{j\to v[1]} = V'_{j\to v[1]} + V''_{j\to v[1]},$$

where $V'_{j\to v[1]}$ is conditionally independent of $V''_{j\to v[1]}$. Define $V_{j\to v[1]} = V'_{j\to v[1]}$ and $\widehat{V}_{j\to v[1]} = V'''_{j\to v[1]}$ if $\lambda_j^{[1]} \leqslant \widehat{P}_{j\to v[1]}$; and vice versa if $\lambda_j^{[1]} > \widehat{P}_{j\to v[1]}$. Then, a moment of thought shows that

$$\mathbb{P}_{\mathbf{x}}\left[\widehat{V}_{j\to v[1]} \neq V_{j\to v[1]} \big| \Xi_{\mathbf{x}}\right] = \mathbb{P}[V''_{j\to v[1]} \geqslant 1 \big| \Xi_{\mathbf{x}}] \leqslant |\lambda_j^{[1]} - \widehat{P}_{j\to v[1]}|,$$

where the last inequality follows from $1 - e^{-x} \leqslant x$. Once again by Lemma 23,

$$\mathbb{P}_{\mathbf{x}}\big[\mathbf{V}_{\mathrm{Po}}^{(v[1],n)} \neq \widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[1],n)} \big| \Xi_{\mathbf{x}}\big] \leqslant \sum_{j=v[1]+1}^{n} |\lambda_j^{[1]} - \widehat{P}_{j\to v[1]}|. \tag{11.9}$$

We now bound the sum in (11.9) on the event $\mathcal{H}_{1,0}$. First, we swap $(a_0^{(\mathrm{i})})^{-1/\mu}$ in $\lambda_j^{[1]}$ for $(v[1]/n)^{-\chi/\mu}$ at a small cost, and then proceed to bound the difference between $\lambda_j^{[1]}$ and $\widehat{P}_{j\to v[1]}$. Recalling that $a_0^{(\mathrm{i})} = U_0^{\chi}$ and $v[1] = \lceil nU_0 \rceil$, we have

$$(a_0^{(\mathrm{i})})^{-1/\mu} - \left(\frac{v[1]}{n}\right)^{-\chi/\mu} \leqslant U_0^{-\chi/\mu}\left[1 - \left(1 + \frac{1}{n}\right)^{-\chi/\mu}\right].$$

Since $\chi/\mu = 1 - \chi$ and $(1 + n^{-1})^{\chi - 1} = \sum_{j\geqslant 0} \binom{\chi-1}{j}(1/n)^j$, there is a constant C_μ such that

$$(a_0^{(\mathrm{i})})^{-1/\mu} - (v[1]/n)^{-\chi/\mu} \leqslant C_\mu n^{-1} U_0^{-\chi/\mu}. \tag{11.10}$$

Because $U_0 \geqslant (\log\log n)^{-1}$ on the event $\mathcal{H}_{1,0}$, it follows that

$$\mathbb{1}[\mathcal{H}_{1,0}]\left\{(a_0^{(\mathrm{i})})^{-1/\mu} - (v[1]/n)^{-\chi/\mu}\right\} \leqslant C_\mu n^{-1}(\log\log n)^{1-\chi} =: \eta_n. \tag{11.11}$$

For $j = v[1] + 2, ..., n$, we use (11.11) to compute

$$\begin{aligned}\lambda_j^{[1]} &= \int_{((j-1)/n)^\chi}^{(j/n)^\chi} \frac{\zeta_0}{\mu(a_0^{(\mathrm{i})})^{1/\mu}} y^{1/\mu - 1} dy = \zeta_0 (a_0^{(\mathrm{i})})^{-1/\mu}\left[\left(\frac{j}{n}\right)^{1-\chi} - \left(\frac{j-1}{n}\right)^{1-\chi}\right] \\ &\leqslant \zeta_0 \left(\frac{j}{n}\right)^{1-\chi}\left[1 - \left(1 - \frac{1}{j}\right)^{1-\chi}\right]\left[\left(\frac{v[1]}{n}\right)^{\chi-1} + \eta_n\right].\end{aligned}$$

Since $(1 - 1/j)^{1-\chi} = \sum_{h\geqslant 0}\binom{1-\chi}{h}(-1/j)^h = 1 - \{(\mu+1)j\}^{-1} + O(j^{-2})$,

$$\lambda_j^{[1]} \leqslant \zeta_0\left[\frac{1}{(\mu+1)j} + O(j^{-2})\right]\left[\left(\frac{j}{v[1]}\right)^{1-\chi} + \eta_n\left(\frac{j}{n}\right)^{1-\chi}\right];$$

and expanding the terms above we obtain

$$\lambda_j^{[1]} \leqslant \widehat{P}_{j\to v[1]} + \frac{\eta_n \zeta_0}{n^{1-\chi}} \frac{1}{j^{\chi}(\mu+1)} + C\zeta_0 j^{-1-\chi}[\eta_n n^{\chi-1} + (v[1])^{\chi-1}],$$

where $C := C(\mu)$ is a constant. Below, we allow the constants $C := C(\mu)$ and $C' := C'(\mu)$ to vary from line to line. Repeating the calculation above for a lower bound on $\lambda_k^{[1]}$, we deduce that on the event $\mathcal{H}_{1,0}$,

$$|\lambda_j^{[1]} - \widehat{P}_{j\to v[1]}| \leqslant \frac{C\zeta_0\eta_n}{n^{1-\chi}} j^{-\chi} + C'\zeta_0(v[1])^{\chi-1} j^{-1-\chi}. \tag{11.12}$$

On the event $\mathcal{H}_{1,0}$, $Cn^{\chi-1}\eta_n\zeta_0 \sum_{j=v[1]+2}^{n} j^{-\chi} \leqslant C(\mu+1)\zeta_0\eta_n$ because

$$\mathbb{1}[\mathcal{H}_{1,0}] \sum_{j=v[1]+2}^{n} j^{-\chi} \leqslant \sum_{j=\lceil n(\log\log n)^{-1}\rceil+2}^{n} j^{-\chi} \leqslant \int_{n(\log\log n)^{-1}+1}^{n} y^{-\chi} dy \leqslant (\mu+1)n^{1-\chi};$$

and similarly, $C'\zeta_0(v[1])^{\chi-1}\sum_{j=v[1]+1}^{n} j^{-1-\chi} \leqslant C'\chi^{-1}\zeta_0 n^{-1}(\log\log n)$. Hence,

$$\mathbb{1}[\mathcal{H}_{1,0}] \sum_{j=v[1]+2}^{n} |\lambda_j^{[1]} - \widehat{P}_{j\to v[1]}| \leqslant \frac{C\log\log n}{n}\zeta_0.$$

Finally, we can use (11.11) and a similar calculation to show that

$$\mathbb{1}[\mathcal{H}_{1,0}]|\lambda_{v[1]+1}^{[1]} - \widehat{P}_{v[1]+1\to v[1]}| \leqslant \frac{C'\log\log n}{n}\zeta_0.$$

Applying the last two displays to (11.9) gives the desired result. □

We now use Lemma 25, 26 and 27 to prove Lemma 9.

Proof of Lemma 9. Applying Lemma 25, 26 and 27 to Lemma 24, we can couple $\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)}$ and $\mathbf{V}_{\mathrm{Po}}^{(v[1],n)}$ so that on the event $\bigcap_{i=1}^{3} F_{1,i} \cap \mathcal{H}_{1,0}$, there are positive constants $C := C(x_1,\mu)$ and $c := c(\mu)$ such that

$$\mathbb{P}_{\mathbf{x}}[\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)} \neq \mathbf{V}_{\mathrm{Po}}^{(v[1],n)}|\Xi_{\mathbf{x}}] \leqslant \frac{C\zeta_0(\log\log n)^{1-\chi}}{n^{\gamma}} + \frac{\zeta_0^2(\log\log n)^{2-\chi}}{(\mu+1)n} + \frac{c\zeta_0\log\log n}{n}$$

for all $n \geqslant 3$. Taking expectation with respect to ζ_0 on $(0,\infty)$ proves the lemma, since $\mathbb{E}_{\mathbf{x}}[\zeta_0|U_0] = x_{v[1]}$ and $\mathbb{E}_{\mathbf{x}}[\zeta_0^2|U_0] = x_{v[1]}(x_{v[1]}+1)$; and on the event $\mathcal{H}_{1,0}$, $v[1] > 1$ and

$x_{v[1]} \leqslant \kappa$. □

11.2. Proof of Lemma 10

Proof. Recall that $a_0^{(\mathrm{i})} = U_0^{\chi}$, $\zeta_0 \sim \mathrm{Gamma}(x_{v[1]}, 1)$ and $R_0^{(\mathrm{i})} \sim \mathrm{Po}(\zeta_0\{(a_0^{(\mathrm{i})})^{-1/\mu} - 1\})$, where U_0 is a standard uniform variable and $R_0^{(\mathrm{i})}$ is as in (4.1). On the event $\mathcal{H}_{1,0} = \{U_0 > (\log\log n)^{-1}\}$, $R_0^{(\mathrm{i})}$ is stochastically dominated by $\xi \sim \mathrm{Po}(\zeta_0(\log\log n)^{1-\chi})$. Hence,

$$\mathbb{P}_{\mathbf{x}}\Bigg[\bigcap_{i=0}^{2}\mathcal{H}_{1,i}\cap\mathcal{H}_{1,3}^{c}\Bigg] \leqslant \mathbb{P}_{\mathbf{x}}[\mathcal{H}_{1,0}\cap\{\xi \geqslant (\log n)^{1/r}\}]$$
$$\leqslant \mathbb{E}_{\mathbf{x}}\big[\mathbb{1}[\mathcal{H}_{1,0}]\mathbb{P}_{\mathbf{x}}\big[\xi \geqslant (\log n)^{1/r}|\zeta_0, U_0\big]\big]. \tag{11.13}$$

To apply Chebyshev's inequality, let $(\tilde{\xi})_k := \tilde{\xi}(\tilde{\xi}-1)\cdots(\tilde{\xi}-k+1)$ for any non-negative integer k. By [19, equation (6.10), p. 262], we have $\tilde{\xi}^p = \sum_{j=0}^{p}\left\{{p \atop j}\right\}(\tilde{\xi})_k$, where $\left\{{p \atop j}\right\}$ is the Stirling number of the second kind (with $\left\{{0 \atop 0}\right\} = 1$ and $\left\{{p \atop 0}\right\} = 0$ for positive integer p). If $\tilde{\xi} \sim \mathrm{Po}(\theta)$, then $\mathbb{E}[(\tilde{\xi})_k] = \theta^k$. So for such $\tilde{\xi}$ and $\theta \geqslant 1$, taking expectation on both sides of the identity gives

$$\mathbb{E}\tilde{\xi}^p = \sum_{k=0}^{p}\left\{{p \atop k}\right\}\mathbb{E}[(\tilde{\xi})_k] \leqslant C_p\theta^p,$$

where C_p is the sum of the Stirling numbers. By Chebyshev's inequality and the moment bound above,

$$\mathbb{P}_{\mathbf{x}}[\xi \geqslant (\log n)^{1/r}|\zeta_0, U_0] \leqslant (\log n)^{-p/r}\mathbb{E}_{\mathbf{x}}[\xi|\zeta_0, U_0] = C_p\zeta_0^p(\log n)^{-p/r}(\log\log n)^{p(1-\chi)}.$$

The lemma follows from applying the above to (11.13); noting that on the event $\mathcal{H}_{1,0}$, there is a positive constant $C := C(p)$ such that $\mathbb{E}_{\mathbf{x}}[\zeta_0^p|U_0] = \prod_{\ell=0}^{p-1}(x_{v[1]} + \ell) \leqslant C\kappa^p$, because $v[1] \geqslant 2$ and $x_i \leqslant \kappa$ for $i \geqslant 2$. □

11.3. Proof of Lemma 13

We first recall the notation in Section 6 that frequently appears in the proof below. The random variable $\tau[q]$ is the time we probe the type L child in $\partial\mathcal{B}_q := B_q(G_n, k_0^{(\mathrm{u})}) \setminus B_{q-1}(G_n, k_0^{(\mathrm{u})})$, and $L[q] = (0, 1, ..., 1)$ with $|L[q]| = q + 1$, so that $v[\tau[q]] = k_{L[q]}^{(\mathrm{u})}$, with

$k^{(\mathrm{u})}_{L[q]}$ as in the beginning of Section 3.2. Moreover, $\zeta_q := \mathcal{Z}_{v[\tau[q]]}[\tau[q]]$ is the gamma variable in (3.7) with $t = \tau[q]$, $M_{L[q]}$ and $\widetilde{M}_{L[q]}$ are as in (6.6) and (6.8).

To prepare for the intermediate coupling steps, we construct a Bernoulli sequence and a Poisson sequence using the means $\widehat{P}_{j\to v[\tau[q]]}$ in (6.12) and in Table 1.

Definition 14. Given $v[\tau[q]]$, $a^{(\mathrm{i})}_{L[q]}$ and ζ_q, let $\widehat{Y}_{j\to v[\tau[q]]}$, $\widetilde{M}_{L[q]} \leqslant j \leqslant n$, be conditionally independent Bernoulli variables, each with parameter $\widehat{P}_{j\to v[\tau[q]]}$ given in (6.12) and in Table 1. Define this Bernoulli sequence by the random vector

$$\widehat{\mathbf{Y}}^{(v[\tau[q]],n)}_{\mathrm{Be}} := \left(\widehat{Y}_{\widetilde{M}_{L[q]}\to v[\tau[q]]}, \widehat{Y}_{(\widetilde{M}_{L[q]}+1)\to v[\tau[q]]}, \dots, \widehat{Y}_{n\to v[\tau[q]]}\right).$$

Definition 15. Given $v[\tau[q]]$, $a^{(\mathrm{i})}_{L[q]}$ and ζ_q, let $\widehat{V}_{j\to v[\tau[q]]}$, $\widetilde{M}_{L[q]} \leqslant j \leqslant n$, be conditionally independent Poisson random variables, each with parameter $\widehat{P}_{j\to v[\tau[q]]}$ given in (6.12) and in Table 1. Define this Poisson sequence by the random vector

$$\widehat{\mathbf{V}}^{(v[\tau[q]],n)}_{\mathrm{Po}} := \left(\widehat{V}_{\widetilde{M}_{L[q]}\to v[\tau[q]]}, \widehat{V}_{(\widetilde{M}_{L[q]}+1)\to v[\tau[q]]}, \dots, \widehat{V}_{n\to v[\tau[q]]}\right).$$

As we assume that the graphs are already coupled such that $(B_q(G_n, v[1]), v[1]) \cong (B_q(\mathcal{T}_{\mathbf{x},n}, 0), 0)$, it is enough to condition on the following collection of random variables in the sequel,

$$\Xi_{\mathbf{x}} := \Big((k^{(\mathrm{u})}_{\bar{w}}, k^{(\mathrm{i})}_{\bar{w}}, a^{(\mathrm{i})}_{\bar{w}})_{k_{\bar{w}}\in\mathcal{A}_{\tau[q]-1}\cup\mathcal{P}_{\tau[q]-1}}, (R^{(\mathrm{u})}_{\bar{w}}, R^{(\mathrm{i})}_{\bar{w}})_{k_{\bar{w}}\in\mathcal{P}_{\tau[q]-1}}, (\mathcal{Z}_j[\tau[q]], \widetilde{\mathcal{Z}}_j[\tau[q]])_{j\in[n]\setminus(\{1\}\cup\mathcal{P}_{\tau[q]-1})}\Big), \tag{11.14}$$

where $R^{(\mathrm{u})}_{\bar{w}}$ and $R^{(\mathrm{i})}_{\bar{w}}$ are as in (3.4) and (4.1), and $\mathcal{Z}_i[\tau[q]]$ and $\widetilde{\mathcal{Z}}_i[\tau[q]]$ are the gamma variables in (3.7) and (3.8). Note also $\mathcal{P}_{\tau[q]-1} = V(B_{q-1}(G_n, v[1]))$ and $\mathcal{A}_{\tau[q]-1} = \partial\mathcal{B}_q$. Given $l \geqslant 1$, we define

$$J := \bigcap_{i=1}^{3} F_{\tau[q],i} \cap \left(\bigcap_{i=1}^{3} \mathcal{H}_{q,i}\right) \cap \{|\partial\mathcal{B}_q| = l\}, \tag{11.15}$$

where $\mathcal{H}_{q,i}$ and $F_{\tau[q],i}$, $i = 1, 2, 3$ are as in (6.1) and (6.11), and we omit l in the notation for simplicity. To prove Lemma 13, observe that on the event $\mathcal{H}_{q,1} \cap \mathcal{H}_{q,2}$,

$$v[\tau[q]] \geqslant n(\log\log n)^{-(q+1)} - C_q n^{1-\beta_q/\chi} \quad \text{and} \quad M_{L[q]} \geqslant n(\log\log n)^{-(q+1)}, \quad (11.16)$$

where $C_q := C_q(x_1, \mu)$ is the positive constant in $\mathcal{H}_{q,2}$.

We are now ready to prove the lemma. As in the case of Lemma 9, we first couple the pairs

$$\Big(\mathbf{Y}_{\mathrm{Be}}^{(v[\tau[q]],n)}, \widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[\tau[q]],n)}\Big), \quad \Big(\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[\tau[q]],n)}, \widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[\tau[q]],n)}\Big) \quad \text{and} \quad \Big(\widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[\tau[q]],n)}, \mathbf{V}_{\mathrm{Po}}^{(v[\tau[q]],n)}\Big)$$

on the event J, and then apply Lemma 24.

Lemma 28. *Let $\mathbf{Y}_{\mathrm{Be}}^{(v[\tau[q]],n)}$ and $\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[\tau[q]],n)}$ be as in (6.7) and Definition 14, and $\Xi_{\mathbf{x}}$ and J be as in (11.14) and (11.15). We can couple the random vectors so that on the event J, there are positive constants $C := C(x_1, \mu, q)$ and $C' := C'(\mu, q)$ such that*

$$\mathbb{P}_{\mathbf{x}}\big[\mathbf{Y}_{\mathrm{Be}}^{(v[\tau[q]],n)} \neq \widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[\tau[q]],n)} \big| \Xi_{\mathbf{x}}\big] \leqslant (\log\log n)^{(1-\chi)(q+1)} \Big\{ C\zeta_q n^{-\gamma} + \frac{C'\zeta_q (\log n)^{q/r}}{(\mu+1)n^{1-\chi}} \Big\}$$

for all $n \geqslant 3$.

Proof. We prove the cases $M_{L[q]} \leqslant v[\tau[q]]$ and $M_{L[q]} \geqslant v[\tau[q]] + 1$ separately. Below we only consider the case $M_{L[q]} \leqslant v[\tau[q]]$ in detail, as the argument for $M_{L[q]} \geqslant v[\tau[q]] + 1$ is the same.

Since $P_{j\to v[\tau[q]]} = \widehat{P}_{j\to v[\tau[q]]} = 0$ for $M_{L[q]} \leqslant j \leqslant v[\tau[q]]$, $Y_{j\to v[\tau[q]]} = \widehat{Y}_{j\to v[\tau[q]]} = 0$; whereas for $v[\tau[q]] + 1 \leqslant j \leqslant n$, we couple $Y_{j\to v[\tau[q]]}$ and $\widehat{Y}_{j\to v[\tau[q]]}$ as follows. Firstly, we use a similar calculation as for (11.6) to show that there is a positive constant $C := C(x_1, \mu, q)$ such that on the event J,

$$(1 - Cn^{-\gamma})\widehat{P}_{j\to v[\tau[q]]} \leqslant P_{j\to v[\tau[q]]} \leqslant (1 + Cn^{-\gamma})\widehat{P}_{j\to v[\tau[q]]}$$

for $j \in \mathcal{N}_{\tau[q]-1} \cap \{v[\tau[q]] + 1, ..., n\}$. For such j, we couple $Y_{j\to v[\tau[q]]}$ and $\widehat{Y}_{j\to v[\tau[q]]}$ as in Lemma 25, such that on the event J,

$$\mathbb{P}_{\mathbf{x}}\big[Y_{j\to v[\tau[q]]} \neq \widehat{Y}_{j\to v[\tau[q]]} \big| \Xi_{\mathbf{x}}\big] \leqslant Cn^{-\gamma}\widehat{P}_{j\to v[\tau[q]]}.$$

For $j \in (\mathcal{A}_{\tau[q]-1} \cup \mathcal{P}_{\tau[q]-1}) \cap \{v[\tau[q]]+1, ..., n\}$, we have

$$\mathbb{P}_{\mathbf{x}}\big[Y_{j\to v[\tau[q]]} \neq \widehat{Y}_{j\to v[\tau[q]]}\big|\Xi_{\mathbf{x}}\big] \leqslant \widehat{P}_{j\to v[\tau[q]]};$$

noting that $Y_{j\to v[\tau[q]]} = 0$ for such j. Following from Lemma 23,

$$\begin{aligned}
&\mathbb{P}_{\mathbf{x}}\big[\mathbf{Y}^{(v[\tau[q]],n)}_{\mathrm{Be}} \neq \widehat{\mathbf{Y}}^{(v[\tau[q]],n)}_{\mathrm{Be}}\big|\Xi_{\mathbf{x}}\big] \\
&\quad \leqslant Cn^{-\gamma} \sum_{\substack{j=v[\tau[q]]+1;\\ j\in\mathcal{N}_{\tau[q]-1}}}^{n} \widehat{P}_{j\to v[\tau[q]]} + \sum_{\substack{j=v[\tau[q]]+1;\\ j\in\mathcal{A}_{\tau[q]-1}\cup\mathcal{P}_{\tau[q]-1}}}^{n} \widehat{P}_{j\to v[\tau[q]]}. \qquad (11.17)
\end{aligned}$$

For the first sum in (11.17), we use $n(\log\log n)^{-(q+1)} \leqslant M_{L[q]} \leqslant v[\tau[q]]$ to obtain

$$\begin{aligned}
Cn^{-\gamma} \sum_{\substack{j=v[\tau[q]]+1;\\ j\in\mathcal{N}_{\tau[q]-1}}}^{n} \widehat{P}_{j\to v[\tau[q]]} &\leqslant \frac{Cn^{-\gamma}\zeta_q}{(\mu+1)\{v[\tau[q]]\}^{1-\chi}} \sum_{j=v[\tau[q]]+1}^{n} j^{-\chi} \\
&\leqslant \frac{Cn^{-\gamma}\zeta_q}{(\mu+1)\{v[\tau[q]]\}^{1-\chi}} \int_{v[\tau[q]]}^{n} y^{-\chi}dy \\
&\leqslant C\zeta_q n^{-\gamma}(\log\log n)^{(1-\chi)(q+1)}. \qquad (11.18)
\end{aligned}$$

To bound the second sum in (11.17), note that by a calculation similar to (6.13),

$$|\mathcal{A}_{\tau[q]-1}| + |\mathcal{P}_{\tau[q]-1}| = |V(B_q(G_n, v[1]))| \leqslant 1 + q + q^2(\log n)^{q/r}$$

on the event $\mathcal{H}_{q,3}$, where $v[1]$ is the uniformly chosen vertex. So combining the last display, the lower bound on $v[\tau[q]]$ and $j^{-\chi} \leqslant 1$ for $j \geqslant 1$, we get that on the event J, there is a constant $C' := C'(\mu, q)$ such that

$$\begin{aligned}
\sum_{j\in\mathcal{A}_{\tau[q]-1}\cup\mathcal{P}_{\tau[q]-1}} \widehat{P}_{j\to v[\tau[q]]} &= \frac{\zeta_q}{(\mu+1)[v[\tau[q]]]^{1-\chi}} \sum_{j\in\mathcal{A}_{\tau[q]-1}\cup\mathcal{P}_{\tau[q]-1}} j^{-\chi} \\
&\leqslant \frac{C'\zeta_q(\log\log n)^{(q+1)(1-\chi)}}{n^{1-\chi}}(\log n)^{q/r}. \qquad (11.19)
\end{aligned}$$

Applying (11.18) and (11.19) to (11.17) proves the lemma for $M_{L[q]} \leqslant v[\tau[q]]$. □

Lemma 29. *Let $\widehat{\mathbf{Y}}^{(v[\tau[q]],n)}_{\mathrm{Be}}$ and $\widehat{\mathbf{V}}^{(v[\tau[q]],n)}_{\mathrm{Po}}$ be as in Definition 14 and 15, and $\Xi_{\mathbf{x}}$ and J be as in (11.14) and (11.15). We can couple the random vectors so that on the event*

J, there is a positive constant $C := C(x_1, \mu, q)$ such that

$$\mathbb{P}_{\mathbf{x}}\big[\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[\tau[q]],n)} \neq \widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[\tau[q]],n)}\big|\Xi_{\mathbf{x}}\big] \leqslant \frac{C\zeta_q^2(\log\log n)^{(2-\chi)(q+1)}}{n} \quad \textit{for all } n \geqslant 3.$$

Proof. Since the argument for the case $M_{L[q]} \geqslant v[\tau[q]]+1$ is similar, we only prove the case $M_{L[q]} \leqslant v[\tau[q]]$ in detail. For $M_{L[q]} \leqslant j \leqslant v[\tau[q]]$, we have $\widehat{Y}_{j\to v[\tau[q]]} = \widehat{V}_{j\to v[\tau[q]]} = 0$ because $P_{j\to v[\tau[q]]} = \widehat{P}_{j\to v[\tau[q]]} = 0$. Hence, we only need to couple $\widehat{Y}_{j\to v[\tau[q]]}$ and $\widehat{V}_{j\to v[\tau[q]]}$ for $v[\tau[q]]+1 \leqslant j \leqslant n$. By the standard Poisson-Bernoulli coupling [24, equation (1.11), p. 5], Lemma 23 and the inequality $(v[\tau[q]]/j)^\chi \leqslant 1$,

$$\mathbb{P}_{\mathbf{x}}\big[\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[\tau[q]],n)} \neq \widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[\tau[q]],n)}\big|\Xi_{\mathbf{x}}\big] \leqslant \sum_{j=v[\tau[q]]+1}^{n} \widehat{P}_{j\to v[\tau[q]]}^2 \leqslant \frac{\zeta_q}{(\mu+1)v[\tau[q]]}\sum_{j=v[\tau[q]]+1}^{n}\widehat{P}_{j\to v[\tau[q]]},$$

and bounding the sum using (11.18) proves the lemma for $M_{L[q]} \leqslant v[\tau[q]]$. □

Lemma 30. *Let $\mathbf{V}_{\mathrm{Po}}^{(v[\tau[q]],n)}$ and $\widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[\tau[q]],n)}$ be as in Definition 10 and 15, and $\Xi_{\mathbf{x}}$ and J be as in (11.14) and (11.15). We can couple the random vectors so that on the event J, there is a positive constant $C := C(x_1, \mu, q)$ such that*

$$\mathbb{P}_{\mathbf{x}}\big[\mathbf{V}_{\mathrm{Po}}^{(v[\tau[q]],n)} \neq \widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[\tau[q]],n)}\big|\Xi_{\mathbf{x}}\big] \leqslant C\zeta_q n^{-\beta_q}(\log\log n)^{q+1} \quad \textit{for all } n \geqslant 3.$$

Proof. Separately considering the cases $M_{L[q]} \leqslant v[\tau[q]]$ and $M_{L[q]} \geqslant v[\tau[q]]+1$, we apply the monotone coupling in Lemma 27 to each case. Below, we let $\lambda_j^{[\tau[q]]}$ and $\widehat{P}_{j\to v[\tau[q]]}$ be as in (6.9) and (6.12) for any j. By comparing the Poisson means in Table 1, and applying Lemma 23, we have

$$\begin{aligned}&\mathbb{P}_{\mathbf{x}}\big[\mathbf{V}_{\mathrm{Po}}^{(v[\tau[q]],n)} \neq \widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[\tau[q]],n)}\big|\Xi_{\mathbf{x}}\big]\\ &\leqslant \begin{cases}\sum_{j=M_{L[q]}}^{v[\tau[q]]}\lambda_j^{[\tau[q]]} + \sum_{j=v[\tau[q]]+1}^{n}\big|\lambda_j^{[\tau[q]]} - \widehat{P}_{j\to v[\tau[q]]}\big|, & M_{L[q]} \leqslant v[\tau[q]];\\ \sum_{j=v[\tau[q]]+1}^{n}\big|\lambda_j^{[\tau[q]]} - \widehat{P}_{j\to v[\tau[q]]}\big|, & M_{L[q]} = v[\tau[q]]+1;\\ \sum_{j=v[\tau[q]]+1}^{M_{L[q]}}\widehat{P}_{j\to v[\tau[q]]} + \sum_{j=M_{L[q]}}^{n}\big|\lambda_j^{[\tau[q]]} - \widehat{P}_{j\to v[\tau[q]]}\big|, & M_{L[q]} \geqslant v[\tau[q]]+2.\end{cases}\end{aligned} \tag{11.20}$$

We first handle the sum of absolute mean differences in all three cases in (11.20). When

$M_{L[q]} \leqslant v[\tau[q]]+1$, clearly

$$\sum_{j=v[\tau[q]]+1}^{n} \left|\lambda_j^{[\tau[q]]} - \widehat{P}_{j\to v[\tau[q]]}\right| \leqslant \sum_{j=M_{L[q]}}^{n} \left|\lambda_j^{[\tau[q]]} - \widehat{P}_{j\to v[\tau[q]]}\right|,$$

so it is enough to consider the sum from $M_{L[q]}$ to n in all three cases. On the event $\mathcal{H}_{q,2}$, $|a^{(\mathrm{i})}_{L(q)} - (v[\tau[q]]/n)^{\chi}| \leqslant C_q n^{-\beta_q}$, so a computation similar to that of (11.11) gives

$$\left|(a^{(\mathrm{i})}_{L[q]})^{-\frac{1}{\mu}} - (v[\tau[q]]/n)^{-\frac{\chi}{\mu}}\right| \leqslant c_\mu n^{-\beta_q}(\log\log n)^{1+q} =: \eta_n, \tag{11.21}$$

where $c_\mu := c_\mu(q, x_1, \mu)$. Using (11.21) and the second inequality in (11.16), we can repeat the calculation of (11.12) to obtain

$$\left|\lambda_j^{[\tau[q]]} - \widehat{P}_{j\to v[\tau[q]]}\right| \leqslant \frac{C\zeta_q\eta_n}{n^{1-\chi}} j^{-\chi} + C'\zeta_q(\log\log n)^{(1-\chi)(q+1)} j^{-2}, \tag{11.22}$$

where $C := C(x_1,\mu)$ and $C' := C'(\mu)$. Again using the second inequality in (11.16), a straightforward computation shows that on the event J, $\sum_{j=M_{L[q]}}^{n} j^{-\chi} \leqslant (\mu+1)n^{1-\chi}$ and $\sum_{j=M_{L[q]}}^{n} j^{-2} \leqslant 2n^{-1}(\log\log n)^{q+1}$. Hence, summing (11.22) over j, we get that there is a constant $C := C(x_1,\mu,q)$ such that

$$\sum_{j=M_{L[q]}}^{n} \left|\lambda_j^{[\tau[q]]} - \widehat{P}_{j\to v[\tau[q]]}\right| \leqslant C\zeta_q n^{-\beta_q}(\log\log n)^{q+1}. \tag{11.23}$$

Next, we bound the remaining sums appearing in (11.20). When $M_{L[q]} \leqslant v[\tau[q]]$,

$$\sum_{j=M_{L[q]}}^{v[\tau[q]]} \lambda_j^{[\tau[q]]} = \int_{a^{(\mathrm{i})}_{L[q]}}^{\left(\frac{v[\tau[q]]}{n}\right)^{\chi}} \frac{\zeta_q}{\mu(a^{(\mathrm{i})}_{L[q]})^{1/\mu}} y^{\frac{1}{\mu}-1} dy = \zeta_q \left[\left(\frac{v[\tau[q]]}{n}\right)^{\frac{\chi}{\mu}} (a^{(\mathrm{i})}_{L[q]})^{-\frac{1}{\mu}} - 1\right].$$

Choose n large enough such that $C_q n^{-\beta_q}(\log\log n)^{\chi(q+1)} < 1$, where C_q is the constant in $\mathcal{H}_{q,2}$. Using $(v[\tau[q]]/n)^{\chi} \leqslant a^{(\mathrm{i})}_{L[q]} + C_q n^{-\beta_q}$ and the first inequality in (11.16), a little calculation shows that on the event J,

$$\begin{aligned}\left(\frac{v[\tau[q]]}{n}\right)^{\frac{\chi}{\mu}} (a^{(\mathrm{i})}_{L[q]})^{-\frac{1}{\mu}} - 1 &\leqslant (a^{(\mathrm{i})}_{L[q]})^{-\frac{1}{\mu}} \left\{a^{(\mathrm{i})}_{L[q]} + C_q n^{-\beta_q}\right\}^{\frac{1}{\mu}} - 1 \\ &\leqslant \{1 + C_q n^{-\beta_q}(\log\log n)^{(q+1)\chi}\}^{\frac{1}{\mu}} - 1\end{aligned}$$

$$\leqslant Cn^{-\beta_q}(\log\log n)^{(q+1)\chi}$$

for some constant $C := C(x_1, \mu, q)$. Hence, on the same event,

$$\sum_{j=M_{L[q]}}^{v[\tau[q]]} \lambda_j^{[\tau[q]]} \leqslant C\zeta_q n^{-\beta_q}(\log\log n)^{(q+1)\chi}. \tag{11.24}$$

When $M_{L[q]} \geqslant v[\tau[q]]+2$, we need an upper bound on $M_{L[q]}$ to control $\sum_{j=v[\tau[q]]+1}^{M_{L[q]}} \widehat{P}_{j\to v[\tau[q]]}$. Since $((M_{L[q]}-1)/n)^\chi \leqslant a_{L[q]}^{(\mathrm{i})}$ by definition of $M_{L[q]}$, we have $M_{L[q]} \leqslant n(a_{L[q]}^{(\mathrm{i})})^{1/\chi}+1$. Let

$$b_n := \lfloor n(a_{L[q]}^{(\mathrm{i})})^{1/\chi} - C_q n^{1-\beta_q/\chi} \rfloor \quad \text{and} \quad c_n := n(a_{L[q]}^{(\mathrm{i})})^{1/\chi}+1,$$

so that $v[\tau[q]] \geqslant b_n$ and $M_{L[q]} \leqslant c_n$ on the event J. Pick n large enough so that $e_n := (\log\log n)^{q+1}(C_q n^{-\beta_q/\chi} + n^{-1}) < 1$. On the event J, we use $(a_{L[q]}^{(\mathrm{i})})^{-1/\chi} \leqslant (\log\log n)^{q+1}$ and the first inequality in (11.16) to obtain

$$\begin{aligned}
\sum_{j=v[\tau[q]]+1}^{M_{L[q]}} \widehat{P}_{j\to v[\tau[q]]} \quad &\leqslant \frac{\zeta_q}{(\mu+1)(v[\tau[q]])^{1-\chi}} \int_{b_n}^{c_n} y^{-\chi} dy \\
\leqslant \frac{\zeta_q n^{1-\chi}(a_{L[q]}^{(\mathrm{i})})^{\frac{1}{\mu}}}{(v[\tau[q]])^{1-\chi}} &\left\{ \left[1+n^{-1}(a_{L[q]}^{(\mathrm{i})})^{-\frac{1}{\chi}}\right]^{1-\chi} - \left[1 - C_q(a_{L[q]}^{(\mathrm{i})})^{-\frac{1}{\chi}} n^{-\frac{\beta_q}{\chi}} - n^{-1}(a_{L[q]}^{(\mathrm{i})})^{-\frac{1}{\chi}}\right]^{1-\chi} \right\} \\
\leqslant C\zeta_q(\log\log n)^{q+1} &\left[\{1+n^{-1}(\log\log n))^{q+1}\}^{1-\chi} - \{1-e_n\}^{1-\chi}\right] \\
\leqslant C'\zeta_q n^{-\frac{\beta_q}{\chi}}(\log\log n)^{2(q+1)} &
\end{aligned} \tag{11.25}$$

for some constants $C := C(x_1, \mu, q)$ and $C' := C'(x_1, \mu, q)$. The proof follows from applying (11.23), (11.24) and (11.25) to (11.20). □

Next, we apply Lemma 28, 29 and 30 to prove Lemma 13.

Proof of Lemma 13. By applying Lemma 28, 29 and 30 to Lemma 24, we can couple $\mathbf{Y}_{\mathrm{Be}}^{(v[\tau[q]],n)}$ and $\mathbf{V}_{\mathrm{Po}}^{(v[\tau[q]],n)}$ so that on the event J, there are positive constants $C := C(x_1, \mu, q)$, $C' := C'(x_1, \mu)$, $C'' := C''(x_1, \mu, q)$ and $C''' := C'''(x_1, \mu, q)$ such that

$$\begin{aligned}
&\mathbb{P}_{\mathbf{x}}\big[\mathbf{Y}_{\mathrm{Be}}^{(v[\tau[q]],n)} \neq \mathbf{V}_{\mathrm{Po}}^{(v[\tau[q]],n)} \big| \Xi_{\mathbf{x}}\big] \\
&\quad \leqslant \frac{C\zeta_q(\log\log n)^{(1-\chi)(q+1)}}{n^\gamma} + \frac{C'\zeta_q(\log\log n)^{(q+1)(1-\chi)}}{(\mu+1)n^{1-\chi}}(\log n)^{q/r}
\end{aligned}$$

$$+\frac{C''\zeta_q^2(\log\log n)^{(2-\chi)(q+1)}}{(\mu+1)n}+\frac{C'''\zeta_q(\log\log n)^{q+1}}{n^{\beta_q}} \tag{11.26}$$

for all $n \geqslant 3$. The lemma follows from taking expectation with respect to $\Xi_{\mathbf{x}}$, since on the event J (in particular $\mathcal{H}_{q,1} \cap \mathcal{H}_{q,2}$), $\mathbb{E}_{\mathbf{x}}\{\zeta_q|v[\tau[q]]\} \leqslant \kappa+1$ and $\mathbb{E}_{\mathbf{x}}\{\zeta_q^2|v[\tau[q]]\} \leqslant (\kappa+2)(\kappa+1)$. □

11.4. Proof of Theorem 2

The proof is similar to that of Theorem 1 for $r=1$; but to improve the convergence rate, the key is to choose a suitable threshold n^{ψ}, and then construct a coupling for each fixed vertex $\lceil n^{\psi}\rceil \leqslant j \leqslant n$ in the $(\mathbf{x},n)$-Pólya urn tree (Definition 4). The rate follows from randomising over the uniformly chosen vertex $v[1]$, and taking expectation with respect to the fitness sequence $\mathbf{X}$.

Proof. Given $p>4$, choose α such that $1/2+1/p<\alpha<3/4$, and let $A_{\alpha,n}$ be as in (2.1). In preparation for the coupling, define $\psi > \max\{1-(1-\alpha)/8, \chi\}$ and $\gamma' < \min\{\psi, \chi(3-4\alpha)/4\}$. Let $\Xi_{\mathbf{x}} := ((\mathcal{Z}_j[1], \widetilde{\mathcal{Z}}_j[1]), 2 \leqslant j \leqslant n)$, $(S_{n,j}[1], 1 \leqslant j \leqslant n)$ and the events $F_{1,j}$, $j=1,2,3$ be as in (3.7), (3.8), (3.10) and (5.5). Furthermore, let $U_0 \sim \mathrm{U}[0,1]$, $a_0 = U_0^{\chi}$ and $v[1] = \lceil nU_0\rceil$, with the random vectors $\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)}$, $\mathbf{V}_{\mathrm{Po}}^{(v[1],n)}$, $\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(v[1],n)}$ and $\widehat{\mathbf{V}}_{\mathrm{Po}}^{(v[1],n)}$ as in Definition 7, 9, 12 and 13.

Assume that the realisation of the fitness sequence $\mathbf{x}$ is such that $A_{\alpha,n}$ holds (denoted abusively as $\mathbf{x} \in A_{\alpha,n}$), and $U_0 \in ((j-1)/n, j/n]$ for some positive integer $\lceil n^{\psi}\rceil \leqslant j \leqslant n$, so that the uniformly chosen vertex is $v[1]=j$. We start by coupling $\mathbf{Y}_{\mathrm{Be}}^{(j,n)}$ and $\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(j,n)}$. We first show that, on the event $\bigcap_{i=1}^3 F_{1,i}$, there is a positive constant $C := C(x_1,\mu)$ such that

$$(1-Cn^{-\gamma'})\widehat{P}_{h\to j} \leqslant P_{h\to j} \leqslant (1+Cn^{-\gamma'})\widehat{P}_{h\to j}, \tag{11.27}$$

where $P_{h\to j}$ and $\widehat{P}_{h\to j}$ are as in (3.11) and (5.6), with $v[1]$ replaced by j. Note that we need $\psi > \chi$ in order to apply Lemma 2 and 3, but to see why $\psi > 1-(1-\alpha)/8$ is needed, we sketch the calculation of the upper bound in (11.27). Arguing the same way as for (11.6), there is a constant $C' := C'(x_1,\mu)$ such that

$$\frac{S_{n,j}[1]}{S_{h-1,n}[1]} \leqslant \left(\frac{j}{h}\right)^{\chi} + C'n^{(1-\psi-(1-\alpha)/4)\chi} \quad \text{for } j < h \leqslant n;$$

noting that $C'n^{(1-\psi-(1-\alpha)/4)\chi} \to 0$ as $n \to \infty$ due to our choice of ψ. Letting $\varepsilon = n^{-\gamma'}$ in (2.3) of Lemma 3 and $\phi_n := C'n^{(1-\psi-(1-\alpha)/4)\chi}$, we can repeat the same calculation as for (11.7) to deduce that

$$\begin{aligned} P_{h\to j} &\leqslant \widehat{P}_{h\to j}\Big\{1 + n^{-\gamma'} + \phi_n\Big(\frac{h}{j}\Big)^{\chi} + \phi_n n^{-\gamma'}\Big(\frac{h}{j}\Big)^{\chi}\Big\} \\ &\leqslant \widehat{P}_{h\to j}\Big\{1 + n^{-\gamma'} + \phi_n n^{(1-\psi)\chi} + \phi_n n^{-\gamma'+(1-\psi)\chi}\Big\}. \end{aligned}$$

It is easy to check that $\phi_n n^{(1-\psi)\chi} \to 0$ as $n \to \infty$, if $\psi > 1-(1-\alpha)/8$. This implies the upper bound in (11.27). The lower bound can be deduced in a similar way. Therefore, using the same argument as for Lemma 25, on the event $\bigcap_{i=1}^{3} F_{1,i}$,

$$\mathbb{P}_{\mathbf{x}}\big[\mathbf{Y}_{\mathrm{Be}}^{(j,n)} \neq \widehat{\mathbf{Y}}_{\mathrm{Be}}^{(j,n)} \big| \Xi_{\mathbf{x}}\big] = \frac{Cn^{-\gamma'}\mathcal{Z}_j[1]}{(\mu+1)j^{1-\chi}} \sum_{h=j+1}^{n} h^{-\chi} \leqslant \frac{Cn^{1-\chi-\gamma'}\mathcal{Z}_j[1]}{j^{1-\chi}}. \tag{11.28}$$

The coupling of $\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(j,n)}$ and $\widehat{\mathbf{V}}_{\mathrm{Po}}^{(j,n)}$ is entirely similar to Lemma 26, yielding

$$\mathbb{P}_{\mathbf{x}}[\widehat{\mathbf{Y}}_{\mathrm{Be}}^{(j,n)} \neq \widehat{\mathbf{V}}_{\mathrm{Po}}^{(j,n)} | \Xi_{\mathbf{x}}] \leqslant \frac{\mathcal{Z}_j^2[1] n^{1-\chi}}{(\mu+1)j^{2-\chi}}. \tag{11.29}$$

Next, we couple $\widehat{\mathbf{V}}_{\mathrm{Po}}^{(j,n)}$ and $\mathbf{V}_{\mathrm{Po}}^{(j,n)}$ as in the proof of Lemma 27. To bound the sum of the absolute differences of the Poisson means, we apply (11.10) to obtain $(a_0^{(\mathrm{i})})^{-1/\mu} - (j/n)^{-\chi/\mu} \leqslant c''n^{-\chi}j^{\chi-1}$, where $c'' := c''(\mu)$. Using the bound to continue as for (11.12), and then adding the absolute differences over $j+1,...,n$, there are constants $c' := c'(\mu)$ and $c'' := c''(\mu)$ such that

$$\mathbb{P}_{\mathbf{x}}\big[\mathbf{V}_{\mathrm{Po}}^{(j,n)} \neq \widehat{\mathbf{V}}_{\mathrm{Po}}^{(j,n)} \big| \Xi_{\mathbf{x}}\big] \leqslant \mathcal{Z}_j[1]\{c'n^{-\chi^2}j^{\chi-1} + c''j^{-2}\}. \tag{11.30}$$

Applying the arguments for proving Lemma 9, this time with the bounds (11.28), (11.29) and (11.30), it follows that

$$\begin{aligned} &\mathbb{P}_{\mathbf{x}}\Big[\bigcap_{i=1}^{3} F_{1,i} \cap \{\mathbf{Y}_{\mathrm{Be}}^{(j,n)} \neq \mathbf{V}_{\mathrm{Po}}^{(j,n)}\}\Big] \\ &\qquad \leqslant \frac{cn^{1-\chi-\gamma'}x_j}{j^{1-\chi}} + \frac{x_j(x_j+1)n^{1-\chi}}{(\mu+1)j^{2-\chi}} + \frac{c'x_j n^{-\chi^2}}{j^{1-\chi}} + \frac{c''x_j}{j^2}. \end{aligned} \tag{11.31}$$

To conclude the proof using (11.31), we note that

$$\mathbb{P}\big[\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)} \neq \mathbf{V}_{\mathrm{Po}}^{(v[1],n)}\big] \leqslant \mathbb{E}\big[\mathbb{1}[A_{\alpha,n}]\mathbb{P}_{\mathbf{x}}\big[\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)} \neq \mathbf{V}_{\mathrm{Po}}^{(v[1],n)}\big]\big] + \mathbb{P}[A_{\alpha,n}^c], \tag{11.32}$$

where for the first term, we write

$$\begin{aligned}
\mathbb{P}_{\mathbf{x}}\big[\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)} \neq \mathbf{V}_{\mathrm{Po}}^{(v[1],n)}\big] &= \frac{1}{n}\sum_{j=1}^{n} \mathbb{P}_{\mathbf{x}}\big[\mathbf{Y}_{\mathrm{Be}}^{(j,n)} \neq \mathbf{V}_{\mathrm{Po}}^{(j,n)}\big] \\
&\leqslant \frac{\lceil n^\psi \rceil - 1}{n} + \frac{1}{n}\sum_{j=\lceil n^\psi \rceil}^{n} \mathbb{P}_{\mathbf{x}}\left[\mathbf{Y}_{\mathrm{Be}}^{(j,n)} \neq \mathbf{V}_{\mathrm{Po}}^{(j,n)}\right] \\
&\leqslant n^{\psi-1} + \frac{1}{n}\sum_{j=\lceil n^\psi \rceil}^{n} \mathbb{P}_{\mathbf{x}}\left[\bigcap_{i=1}^{3} F_{1,i} \cap \{\mathbf{Y}_{\mathrm{Be}}^{(j,n)} \neq \mathbf{V}_{\mathrm{Po}}^{(j,n)}\}\right] + \mathbb{P}_{\mathbf{x}}\left[\left(\bigcap_{i=1}^{3} F_{1,i}\right)^c\right] \\
&\leqslant n^{\psi-1} + \frac{1}{n}\sum_{j=\lceil n^\psi \rceil}^{n} \mathbb{P}_{\mathbf{x}}\left[\bigcap_{i=1}^{3} F_{1,i} \cap \{\mathbf{Y}_{\mathrm{Be}}^{(j,n)} \neq \mathbf{V}_{\mathrm{Po}}^{(j,n)}\}\right] + \sum_{i=1}^{3}\mathbb{P}_{\mathbf{x}}[F_{1,i}^c].
\end{aligned} \tag{11.33}$$

When $\mathbf{x} \in A_{\alpha,n}$, we can bound $\mathbb{P}_{\mathbf{x}}[F_{1,i}^c]$ using Lemma 2 and (2.4) and (2.5) of Lemma 3. Note that the bound in (2.5) is of order $n^{-\chi}$, due to our moment assumption on the fitness sequence. The probability $\mathbb{P}[A_{\alpha,n}^c]$ can be bounded by Lemma 1. Applying these bounds, (11.33) and (11.31) to (11.32), and then taking expectation with respect to $\mathbf{X}$, $\mathbb{P}\left[\mathbf{Y}_{\mathrm{Be}}^{(v[1],n)} \neq \mathbf{V}_{\mathrm{Po}}^{(v[1],n)}\right]$ is at most

$$C'n^{-b} + \sum_{j=\lceil n^\psi \rceil}^{n} \left\{ \frac{c\mu n^{-\chi-\gamma'}}{j^{1-\chi}} + \frac{(\mathbb{E}X_2^2+\mu)n^{-\chi}}{(\mu+1)j^{2-\chi}} + \frac{c'\mu n^{-1-\chi^2}}{j^{1-\chi}} + \frac{c''\mu n^{-1}}{j^2} \right\},$$

where $b = \min\{\chi[p(\alpha-1/2)-1], \chi(1-\alpha)/2, \chi(3-4\alpha)-4\gamma', 1-\psi\}$ and $C' := C'(x_1,\mu,p)$. By an integral comparison, the sum above is bounded by $C''n^{-\min\{\gamma',\psi(1-\chi)+\chi,1+\chi^2-\chi\}}$ for some $C'' := C''(x_1,\mu)$. Choosing $d = \min\{b,\gamma',\psi(1-\chi)+\chi, 1+\chi^2-\chi\}$ and $C = 2\max\{C',C''\}$ in the statement of the theorem, and noting that $\chi(1-\alpha)/2 < \chi(1/4-1/(2p))$ concludes the proof. □

Acknowledgements

The author is grateful to the referees and Nathan Ross for their careful reading of the manuscript and numerous helpful suggestions; and to Tejas Iyer for pointing out the connection to the work [18]. This research is supported by an Australian Government Research Training Program scholarship, and partially by ACEMS and the David Lachlan Hay Memorial Fund.

References

[1] Aldous, D. and Steele, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures.* vol. 110 of *Encyclopaedia Math. Sci.* Springer, Berlin pp. 1–72.

[2] Barabási, A. L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286,** 509–512.

[3] Benjamini, I. and Schramm, O. (2001). Recurrence of distributional limits of finite planar graphs. *Electronic Journal of Probability* **6,** 1–13.

[4] Berger, N., Borgs, C., Chayes, J. T. and Saberi, A. (2014). Asymptotic behavior and distributional limits of preferential attachment graphs. *Annals of Probability* **42,** 1–40.

[5] Bhamidi, S. (2007). Universal techniques to analyze preferential attachment trees: Global and local analysis. In preparation.

[6] Bloem-Reddy, B. and Orbanz, P. (2017). Preferential attachment and vertex arrival times. Preprint `https://arxiv.org/pdf/1710.02159.pdf`.

[7] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms* **18,** 279–290.

[8] Borgs, C., Chayes, J., Daskalakis, C. and Roch, S. (2007). First to market is not everything: an analysis of preferential attachment with fitness. In *Proceedings of the 39th annual ACM symposium on Theory of computing.* pp. 135–144.

[9] Borovkov, K. A. and Vatutin, V. A. (2006). On the asymptotic behaviour of random recursive trees in random environments. *Advances in Applied Probability* **38,** 1047–1070.

[10] Bubeck, S., Mossel, E. and Rácz, M. (2015). On the influence of the seed graph in the preferential attachment model. *IEEE Transactions on Network Science and Engineering* **2,** 30–39.

[11] Dereich, S. (2016). Preferential attachment with fitness: unfolding the condensate. *Electronic Journal of Probability* **21,** 1–38.

[12] Dereich, S. and Ortgiese, M. (2014). Robust analysis of preferential attachment models with fitness. *Combinatorics, Probability and Computing* **23,** 386–411.

[13] Dharmadhikari, S. W. and Jogdeo, K. (1969). Bounds on moments of certain random variables. *Annals of Mathematical Statistics* **40,** 1506–1509.

[14] Durrett, R. (2019). *Probability: Theory and Examples* 5 ed. No. 31 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[15] Ergün, G. and Rodgers, G. J. (2002). Growing random networks with fitness. *Physica A: Statistical Mechanics and its Applications* **303,** 261–272.

[16] Garavaglia, A. (2019). Preferential attachment models for dynamic networks. *PhD thesis.* Technische Universiteit Eindhoven.

[17] Garavaglia, A., Hazra, R. S., van der Hofstad, R. and Ray, R. (2022). Universality of the local limit of preferential attachment models. Preprint `https://arxiv.org/abs/2212.05551`.

[18] Garavaglia, A., van der Hofstad, R. and Litvak, N. (2020). Local weak convergence for pagerank. *Annals of Applied Probability* **30,** 40–79.

[19] Graham, R. L., Knuth, D. E. and Patashnik, O. (1994). *Concrete mathematics: A Foundation for Computer Science* 2 ed. Addision-Wesley Longman Publishing Co.

[20] Holmgren, C. and Janson, S. (2017). Fringe trees, crump–mode–jagers branching processes and m-ary search trees. *Probability Surveys* **14,** 53–154.

[21] Iyer, T. (2023). Degree distributions in recursive trees with fitnesses. *Advances in Applied Probability* **55,** 407–443.

[22] Jameson, G. J. O. (2013). Inequalites for gamma function ratios. *American Mathematical Monthly* **120,** 936–940.

[23] Karp, R. M. (1990). The transitive closure of a random digraph. *Random Structures & Algorithms* **1,** 73–93.

[24] Lindvall, T. (2002). *Lectures on the coupling method.* Courier Corporation.

[25] Lo, T. Y. Y. (2023). Weak local limit of preferential attachment random trees with additive fitness. To appear in Advances in Applied Probability.

[26] Lodewijks, B. and Ortgiese, M. (2020). A phase transition for preferential attachment models with additive fitness. *Electronic Journal of Probability.* **25,** 1–54.

[27] Lukacs, E. (1955). A characterization of the gamma distribution. *Annals of Mathematical Statistics* **26,** 319–324.

[28] Mahmoud, H. (2008). *Pólya urn models.* Chapman and Hall.

[29] Móri, T. F. (2005). The maximum degree of the barabási-albert random tree. *Combinatorics, Probability and Computing* **14,** 339–348.

[30] Pain, M. and Sénizergues, D. (2022). Correction terms for the height of weighted recursive trees. *Annals of Applied Probability* **32,** 3027–3059.

[31] Peköz, E. A., Röllin, A. and Ross, N. (2013). Total variation error bounds for geometric approximation. *Bernoulli* **19,** 610–632.

[32] Peköz, E. A., Röllin, A. and Ross, N. (2019). Pólya urns with immigration at random times. *Bernoulli* **25,** 189–220.

[33] Peköz, E. A., Röllin, A. and Ross, N. (2016). Generalized gamma approximation with rates for urns, walks and trees. *Annals of Probability* **44,** 1776–1816.

[34] Peköz, E. A., Röllin, A. and Ross, N. (2017). Joint degree distributions of preferential attachment random graphs. *Advances in Applied Probability* **49,** 368–387.

[35] Petrov, V. V. (1975). *Sums of independent random variables.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. Springer Verlag.

[36] Ross, N. (2013). Power laws in preferential attachment graphs and stein's method for the negative binomial distribution. *Advances in Applied Probability* **45,** 876–893.

[37] Rudas, A., Tóth, B. and Valkó, B. (2007). Random trees and general branching processes. *Random Structures & Algorithms* **31,** 186–202.

[38] Sénizergues, D. (2021). Geometry of weighted recursive and affine preferential attachment trees. *Electronic Journal of Probability* **26,** 1–56.

[39] van der Hofstad, R. (2017). *Random graphs and complex networks* vol. 1 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press.

[40] van der Hofstad, R. (2021). Random graphs and complex networks, vol. 2. Book draft of 6 December 2021: `https://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf`.