THE NUMBER OF DESCENDANTS IN A PREFERENTIAL ATTACHMENT GRAPH

SVANTE JANSON, TIFFANY Y. Y. LO

Abstract. We study the number $X^{(n)}$ of vertices that can be reached from the last added vertex n via a directed path (the descendants) in the standard preferential attachment graph. In this model, vertices are sequentially added, each born with outdegree $m \geqslant 2$; the endpoint of each outgoing edge is chosen among previously added vertices with probability proportional to the current degree of the vertex plus some number ρ.

We show that $X^{(n)}/n^{\nu}$ converges in distribution as $n \to \infty$, where ν depends on both m and ρ, and the limiting distribution is given by a product of a constant factor and the $(1-\nu)$-th power of a Gamma$(m/(m-1), 1)$ variable. The proof uses a Pólya urn representation of preferential attachment graphs, and the arguments of Janson (2024) where the same problem was studied in uniform attachment graphs. Further results, including convergence of all moments and analogues for the version with possible self-loops are provided.

1. Introduction

Preferential attachment models have emerged as a popular class of random graphs since it was proposed in [2] as an explanation for the power-law degree sequences observed in real-world networks. There are several versions of these models, differing in minor details, see e.g. [11]; we will use the version defined below, which is the sequential model in [3]. In this version, self-loops are not allowed but multiple edges are possible. The graph is often treated as undirected, but we regard it as directed, with all edges directed from the younger vertex (with larger label) to the older vertex (with smaller label).

Definition 1.1 (Preferential attachment graph)**.** Fix an integer $m \geqslant 2$ and a real number $\rho > -m$, and let $(G_n)_{n\geqslant 1}$ be the sequence of random graphs that are generated as follows; G_n has n vertices with labels in $[n] := \{1, \dots, n\}$. The initial graph G_1 consists of a single vertex (labelled 1) with no edges. Given G_{n-1}, we construct G_n from G_{n-1} by adding the new vertex with label n, and sequentially attaching m edges between vertex n and at most m vertices in G_{n-1} as follows. Let $d_j(n)$ be the degree of vertex j in G_n. If $n \geqslant 2$, each outgoing edge of vertex n is attached to vertex $j \in [n-1]$ with probability proportional to $\rho +$ the current degree of vertex j. (In particular, if $n = 2$, we add m edges from vertex 2 to vertex 1.) This means that the first outgoing edge of vertex n is attached to vertex $j \in [n-1]$ with probability

$$\frac{d_j(n-1)+\rho}{2m(n-2)+(n-1)\rho}; \tag{1.1}$$

noting that $\sum_{k=1}^{n-1} d_k(n-1) = 2m(n-2)$ and $d_j(n-1)+\rho \geqslant m+\rho > 0$. Furthermore, given that the first $1 \leqslant k \leqslant m-1$ outgoing edges of vertex n have been added to the graph, the $(k+1)$th edge of vertex n is attached to vertex $j \in [n-1]$ with probability

$$\frac{d_j(n-1)+\sum_{\ell=1}^{k} \mathbf{1}[n \overset{\ell}{\to} j]+\rho}{2m(n-2)+k+(n-1)\rho}, \tag{1.2}$$

Date: 18 December, 2024; typo corrected 14 January, 2025.

Supported by the Knut and Alice Wallenberg Foundation, Ragnar Söderberg Foundation, the Swedish Research Council (VR), and Sverker Lerheden Foundation.

where $n \overset{\ell}{\to} j$ is shorthand for the event that the ℓ-th outgoing edge of vertex n is attached to vertex j. The resulting graph G_n is a preferential attachment graph with n vertices with parameters m and ρ, and we denote its law by $\mathrm{PA}(n, m, \rho)$.

The formulation of the sequential model in [3] is somewhat different, but is easily seen to be equivalent. Note also that [3] assume (in our notation) $\rho \geqslant 0$, but in the formulation above, only $\rho > -m$ is needed. The definition above is valid also for $m = 1$ (in which case the graph is a tree), but we do not consider this case in the present paper; see Remark 1.5 below for a further discussion.

Since [6] proved that the degree sequence of a certain class of preferential attachment models indeed has a power-law behaviour, many other properties of the model above and its variants have been investigated over the last two decades. These results include for example, vertex degrees, distance and local weak convergences; and we refer to the books [11, 12] for a comprehensive overview.

In this paper, we study the number of vertices that can be reached from the lastly added vertex n via a directed path in the preferential attachment graph. We refer to these vertices (including vertex n) as the *descendants* of n and their count as $X^{(n)}$, even though all of them (apart from vertex n itself) are added to G_n before n. The problem was first considered in [15, Exercise 7.2.2.3-371 and 372] for a uniform attachment graph, where each vertex has $m \geqslant 2$ outgoing edges and the endpoints of these edges are chosen uniformly among the existing vertices. ([15] uses drawing without replacement, thus avoiding multiple edges, but as shown in [13], this makes no difference asymptotically.) This uniform attachment version is studied in [13], where it is shown that as $n \to \infty$, if $\nu = (m-1)/m$, then $X^{(n)}/n^{\nu}$ converges in distribution, and the limiting distribution is given by a product of a constant factor and the $(1-\nu)$-th power of a $\mathrm{Gamma}(m/(m-1), 1)$ variable. The main result of the present paper is that for the preferential attachment graph defined above, $X^{(n)}$ behaves similarly, but with a different exponent ν which furthermore depends on both m and ρ.

As in previous works such as [3, 18, 21], the analysis in this work is hinged on a connection between Pólya urns and the preferential attachment mechanism. We use, in particular, the Pólya urn representation of [3] that was originally devised to study the local weak limit of preferential attachment graphs. As we show later, this representation result enables us to adapt the framework of [13] to study the problem in the preferential attachment setting.

We state our main results in the next subsection.

1.1. **Main results.** The parameters $m \geqslant 2$ and $\rho > -m$ are fixed throughout the paper. We define

$$\nu := \frac{(m-1)(m+\rho)}{m(m+\rho+1)} \in (0, 1). \tag{1.3}$$

The proofs of the results below are developed in Sections 2–11, and as by-products of the proofs, we also prove some results on the structure of the subgraph of descendants of n. In Section 12 we show that the following results hold also for a preferential attachment model with possible self-loops.

Theorem 1.2. *As $n \to \infty$,*

$$n^{-\nu} X \overset{\mathrm{d}}{\longrightarrow} \frac{\Gamma\big(\frac{(m-1)(m+\rho)}{m(m+\rho+1)}\big)\Gamma\big(\frac{m+\rho}{m(m+\rho+1)}+1\big)}{\Gamma\big(\frac{m+\rho}{m+\rho+1}\big)} \Big(\frac{(m+\rho+1)(m-1)}{2m+\rho}\xi_1\Big)^{1-\nu}, \tag{1.4}$$

where $\xi_1 \in \mathrm{Gamma}(m/(m-1), 1)$.

Theorem 1.3. *All moments converge in* (1.4)*. In other words, for any* $p > 0$*, as* $n \to \infty$*,*

$$\mathbb{E}[X^p]/n^{p\nu} \to \left(\frac{\Gamma\left(\frac{(m-1)(m+\rho)}{m(m+\rho+1)}\right)\Gamma\left(\frac{m+\rho}{m(m+\rho+1)}+1\right)}{\Gamma\left(\frac{m+\rho}{m+\rho+1}\right)} \left(\frac{(m+\rho+1)(m-1)}{2m+\rho}\right)^{1-\nu} \right)^p \cdot \frac{\Gamma\left(p(1-\nu)+\frac{m}{m-1}\right)}{\Gamma\left(\frac{m}{m-1}\right)}. \tag{1.5}$$

Remark 1.4. In the special case $\rho = 0$, (1.3) and (1.4) simplify to $\nu = (m-1)/(m+1)$ and

$$n^{-\nu} X \xrightarrow{\mathrm{d}} \frac{1}{m+1} \frac{\Gamma\left(\frac{m-1}{m+1}\right)\Gamma\left(\frac{1}{m+1}\right)}{\Gamma\left(\frac{m}{m+1}\right)} \left(\frac{m^2-1}{2m}\xi_1\right)^{2/(m+1)}. \tag{1.6}$$

If we specialize further to the case $m = 2$ and $\rho = 0$, we get $\nu = 1/3$, and (1.4) simplifies further to

$$n^{-1/3} X \xrightarrow{\mathrm{d}} \frac{\Gamma\left(\frac{1}{3}\right)^2}{2^{4/3}3^{1/3}\Gamma\left(\frac{2}{3}\right)} \xi_1^{2/3} = \frac{3^{1/6}\Gamma\left(\frac{1}{3}\right)^3}{2^{7/3}\pi} \xi_1^{2/3}, \tag{1.7}$$

with $\xi_1 \in \mathrm{Gamma}(2,1)$ and

$$\frac{\Gamma\left(\frac{1}{3}\right)^2}{2^{4/3}3^{1/3}\Gamma\left(\frac{2}{3}\right)} \doteq 1.45833. \tag{1.8}$$

In this case, (1.5) yields, for example,

$$\mathbb{E}[X]/n^{1/3} \to \frac{\Gamma\left(\frac{1}{3}\right)^2}{2^{4/3}3^{1/3}\Gamma\left(\frac{2}{3}\right)} \Gamma\left(2+\tfrac{2}{3}\right) = \frac{5\,\Gamma\left(\frac{1}{3}\right)^2}{2^{1/3}3^{7/3}} \doteq 2.19416. \tag{1.9}$$

Remark 1.5. Definition 1.1 is valid also for $m = 1$, and then defines a random tree; such preferential attachment trees have been studied by many authors. In this case, $X^{(n)}$ equals 1 + the depth of vertex n, and it is known that $X^{(n)}$ grows like $\log n$, in contrast to the case $m \geqslant 2$ studied in the present paper, where we show that $X^{(n)}$ grows as a power of n. More precisely, as $n \to \infty$,

$$X^{(n)}/\log n \xrightarrow{\mathrm{p}} \frac{1+\rho}{2+\rho}, \tag{1.10}$$

and precise results are known on the exact distribution, Poisson approximation, and a central limit theorem, see [8], [20, Theorem 6], and [16, Theorem 3]. (Papers on preferential attachment trees usually use a slightly different definition, where the attachment probabilities depend on the outdegree rather than the degree as in (1.1); apart from a shift in the parameter ρ, this makes a difference only at the root. This minor difference ought not to affect asymptotic result; for $X^{(n)}$ this follows rigorously by the bijection in [16] which yields both exact and asymptotic results, and in particular (1.10), by straightforward calculations for both versions of the definition.)

We mention also an open problem, which we have not studied, where the same methods might be useful.

Problem 1.6. Study the asymptotic behaviour of $\max\{X^{(n+1)}, \dots, X^{(n+i)}\}$ for a fixed $i \geqslant 2$, in both uniform and preferential attachment graphs. Perhaps also do the same for $i = i(n)$ growing with n at some rate.

1.2. **Notation.** As above, $k \overset{\ell}{\to} i$ (where $1 \leqslant i < k \leqslant n$ and $\ell \in [m]$) denotes that in G_n the ℓ-th outgoing edge of vertex k is attached to vertex i. We say that vertex i is a *child* of vertex k if there is such an edge.

As usual, empty sums are 0, and empty products are 1.

Convergence in distribution, in probability, and a.s. (almost surely) are denoted by $\overset{\mathrm{d}}{\longrightarrow}$, $\overset{\mathrm{p}}{\longrightarrow}$, and $\overset{\mathrm{a.s.}}{\longrightarrow}$, respectively. Equality in distribution is denoted by $\overset{\mathrm{d}}{=}$, and w.h.p. (with high probability) is short for "with probability tending one as $n \to \infty$".

We frequently use two standard probability distributions. The Gamma(a, b) distribution, with $a, b > 0$, has density $\Gamma(a)^{-1} b^{-a} x^{a-1} e^{-x/b}$ on $(0, \infty)$. The Beta(a, b) distribution, with $a, b > 0$, has density $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$ on $(0, 1)$.

Most quantities defined below depend on n. We sometimes indicate this by a superscript ${}^{(n)}$, but usually we omit this to simplify the notation. We may in proofs sometimes tacitly assume that n is large enough.

$C[a, b]$, $C[0, \infty)$ and $C(0, \infty)$ denote the spaces of continuous functions on the indicated intervals, equipped with the topology of uniform convergence on compact subsets. These spaces are complete separable metric spaces. Note that a sequence of random functions in $C[0, \infty)$ or $C(0, \infty)$ converges (a.s., in probability, or in distribution) if and only if it converges in the same sense in $C[a, b]$ for each compact interval $[a, b]$ in $[0, \infty)$ or $(0, \infty)$, respectively. (For $C[0, \infty)$ it is obviously equivalent to consider intervals $[0, b]$ only.) The case $C[0, \infty)$ is treated in detail in [22]; the case $C(0, \infty)$ is similar.

C denotes positive constants (not depending on n) that may vary from one occasion to another. The constants may depend on the parameters m and ρ; we indicate dependence on other parameters (if any) by writing e.g. C_a.

2. Pólya urn representation

We shall use a celebrated result of [3], which states that the dynamics of the preferential attachment graph can be encoded in a collection of classical Pólya urns; see also [12, Chapter 5] for more details. In a classical Pólya urn with initially a red balls and b black balls, a ball is randomly sampled from the urn at each step, and is then returned to the urn with another ball of the same colour. (The "numbers" of balls are not necessarily integers; any positive real numbers are allowed.) In the preferential attachment graph, for each $i \geqslant 2$, the weight of vertex i, defined as the degree + ρ, and the total weight of the first $i-1$ vertices evolve like the numbers of red and black balls in a classical Pólya urn. The initial numbers of red and black balls are $a = m + \rho$ and $b = (2i-3)m + (i-1)\rho$, which are the weights of vertex i and the first $i-1$ vertices before the edges of vertex $i+1$ are added to the graph. When one of the first i vertices is chosen as a recipient of a newly added edge, the number of red balls in the urn increases by one if vertex i is the recipient; otherwise we add a new black ball to the urn. It is well-known, for example as a consequence of exchangeability and de Finetti's theorem, that the proportion of red balls a.s. converges to a random number $\beta \in \mathrm{Beta}(a, b)$, and that conditioned on β, the indicators that a red ball is chosen at each step are distributed as conditionally independent Bernoulli variables with parameter β. Consequently, by conditioning on suitable beta variables, the preferential attachment graph can instead be generated using independent steps.

The model and the theorem below are easy variations of their counterparts in [3, Section 2.2]. The only difference is that ρ is allowed to be negative here.

Definition 2.1 (Pólya urn representation, [3])**.** Given the integer $m \geqslant 2$ and the real number $\rho > -m$, let $(B_j)_{j=1}^{\infty}$ be independent random variables such that $B_1 = 1$ and

$$B_j \in \mathrm{Beta}(m+\rho, (2j-3)m + (j-1)\rho), \qquad j \geqslant 2. \tag{2.1}$$

Given $(B_j)_{j=1}^{\infty}$, construct for each $n \geqslant 1$ a (directed) graph G_n on n vertices (labelled by $[n]$) such that each vertex $2 \leqslant k \leqslant n$ has m outgoing edges, and the recipient of each outgoing edge of k is $i \in [k-1]$ with probability

$$B_i \prod_{j=i+1}^{k-1} (1 - B_j), \tag{2.2}$$

with the endpoints of all edges in G_n chosen (conditionally) independently. The law of G_n is denoted by $\mathrm{PU}(n, m, \rho)$, where PU is short for Pólya Urn.

Remark 2.2. The probabilities (2.2) can be interpreted as follows, which will be useful below: Given $(B_j)_{j=1}^{\infty}$, each edge from vertex k tries to land at $k-1, k-2, \dots$ successively; at each vertex j it stops with probability B_j, and otherwise it continues to the next vertex. (All random choices are independent, given $(B_j)_{j=1}^{\infty}$.)

Remark 2.3. The construction in [3] is actually formulated in the following somewhat different way, which obviously is equivalent; we will use this version too below. Define

$$S_{n,j} = \prod_{i=j+1}^{n-1} (1 - B_i) \qquad \text{for } 0 \leqslant j \leqslant n-1. \tag{2.3}$$

(In particular, $S_{n,0} = 0$ and $S_{n,n-1} = 1$.) Conditioned on $(B_j)_{j=2}^{n-1}$, let $(U_{k,\ell})_{k=2,\ell=1}^{n,m}$ be independent random variables with

$$U_{k,\ell} \in \mathsf{U}[0, S_{n,k-1}). \tag{2.4}$$

For each vertex $2 \leqslant k \leqslant n$, add the m outgoing edges such that

$$k \overset{\ell}{\to} i \iff U_{k,\ell} \in [S_{n,i-1}, S_{n,i}), \qquad \ell \in [m],\ i \in [k-1]. \tag{2.5}$$

Note also that a natural way to achieve (2.4) is to let $(\widetilde{U}_{k,\ell})_{k=2,\ell=1}^{n-1,m}$ be independent $\mathsf{U}[0,1]$ variables, independent of $(B_i)_{i=2}^{n-1}$, and set

$$U_{k,\ell} := S_{n,k-1} \widetilde{U}_{k,\ell}. \tag{2.6}$$

Theorem 2.4 ([3], Theorem 2.1)**.** *For all integers $n \geqslant 2$, $m \geqslant 2$ and real $\rho > -m$, $\mathrm{PA}(n, m, \rho) = \mathrm{PU}(n, m, \rho)$.*

In view of this theorem, it is enough to consider the Pólya urn representation instead of the preferential attachment graph. We shall do so in the subsequent analysis and always have $G_n \in \mathrm{PU}(n, m, \rho)$.

Remark 2.5. The uniform directed acyclic graph studied in [13], where each new edge from k is attached uniformly to a vertex in $[k-1]$, can be seen as the limit as $\rho \to \infty$ of the construction above; it can be constructed by the same procedure, except that we let $B_j := 1/j$ (deterministically). This may help seeing the similarities and differences in the arguments below and in [13]. Not surprisingly, formally taking the limit $\rho \to \infty$ in (1.4) yields the main result of [13].

Remark 2.6. Unless we say otherwise, we use the same sequence $(B_i)_{i=1}^{\infty}$ for every n. (But see Section 8 for an exception.)

3. Preliminaries

For convenience, we define the positive constants

$$\theta := 2m + \rho, \tag{3.1}$$

$$\chi := \frac{m+\rho}{2m+\rho} = \frac{m+\rho}{\theta}, \tag{3.2}$$

noting that if $\rho = 0$, then $\chi = 1/2$ for any m.

Recall that if $B \in \mathrm{Beta}(a, b)$, it follows by evaluating a beta integral that the moments are given by

$$\mathbb{E}\, B^s = \frac{\Gamma(a+b)\Gamma(a+s)}{\Gamma(a)\Gamma(a+b+s)} = \frac{\Gamma(a+s)/\Gamma(a)}{\Gamma(a+b+s)/\Gamma(a+b)}, \qquad s > 0. \tag{3.3}$$

Recall also that for any fixed real (or complex) a and b, and $x > 0$ (with $x + a \notin \{0, -1, -2, \dots\}$,

$$\frac{\Gamma(x+a)}{\Gamma(x+b)} = x^{a-b}\bigl(1 + O(x^{-1})\bigr), \tag{3.4}$$

which follows readily from Stirling's formula; see also [19, 5.11.13].

Similarly to the definition of $S_{n,j}$ in (2.3), we also define

$$\Phi_k = \prod_{j=1}^{k} (1 + (m-1)B_j), \quad \text{for } k \geqslant 0. \tag{3.5}$$

We collect here some simple results for these variables that will be used later.

Lemma 3.1. *For $2 \leqslant i < \infty$, let B_i be as in (2.1); and for $1 \leqslant i < \infty$, let Φ_i be as in (3.5). We then have for $2 \leqslant i < \infty$,*

$$\mathbb{E}(B_i) = \frac{m+\rho}{\theta i - 2m} = \frac{\chi}{i - 2m/\theta} = \frac{\chi}{i} + O\Bigl(\frac{1}{i^2}\Bigr), \tag{3.6}$$

$$\mathbb{E}(B_i^2) = \frac{(m+\rho+1)(m+\rho)}{(\theta i - 2m + 1)(\theta i - 2m)} = O\Bigl(\frac{1}{i^2}\Bigr). \tag{3.7}$$

Furthermore, for $2 \leqslant j \leqslant k < \infty$,

$$\begin{aligned}\prod_{i=j}^{k} \mathbb{E}(1 + (m-1)B_i) &= \frac{\Gamma\bigl(k + 1 + [(m-1)(m+\rho) - 2m]/\theta\bigr)\Gamma(j - 2m/\theta)}{\Gamma(k+1-2m/\theta)\Gamma\bigl(j + [(m-1)(m+\rho) - 2m]/\theta\bigr)} \\ &= \Bigl(\frac{k}{j}\Bigr)^{(m-1)\chi}\bigl(1 + O(j^{-1})\bigr)\end{aligned} \tag{3.8}$$

and

$$\mathbb{E}\,\Phi_k = \frac{m \cdot \Gamma\bigl(2 - 2m/\theta\bigr)}{\Gamma\bigl(2 + [(m-1)(m+\rho) - 2m]/\theta\bigr)} k^{(m-1)\chi}\bigl(1 + O(k^{-1})\bigr). \tag{3.9}$$

Finally, there is a positive constant C such that, for $2 \leqslant j \leqslant k < \infty$,

$$\prod_{i=j}^{k} \mathbb{E}(1 + (m-1)B_i)^2 \leqslant C\Bigl(\frac{k}{j}\Bigr)^{2(m-1)\chi} \tag{3.10}$$

and, for $2 \leqslant k < \infty$,

$$\mathbb{E}(\Phi_k^2) \leqslant Ck^{2(m-1)\chi}, \quad \mathbb{E}(\Phi_k^{-1}) \leqslant Ck^{-(m-1)\chi}, \quad \mathbb{E}(\Phi_k^{-2}) \leqslant Ck^{-2(m-1)\chi}. \tag{3.11}$$

Remark 3.2. If $m = 2$ and $\rho = 0$, then $\theta = 4$ and so in (3.9),

$$\frac{m \cdot \Gamma\bigl(2 - 2m/\theta\bigr)}{\Gamma\bigl(2 + [(m-1)(m+\rho) - 2m]/\theta\bigr)} = \frac{2}{\Gamma(3/2)} = \frac{4}{\sqrt{\pi}}. \tag{3.12}$$

Proof. The equalities in (3.6) and (3.7) follow from (3.3), recalling (3.1)–(3.2). For (3.8), we use (3.6) to obtain

$$\prod_{i=j}^{k} \mathbb{E}(1+(m-1)B_i) = \prod_{i=j}^{k} \frac{i+[(m-1)(m+\rho)-2m]/\theta}{i-2m/\theta}$$
$$= \frac{\Gamma\bigl(k+1+[(m-1)(m+\rho)-2m]/\theta\bigr)\Gamma(j-2m/\theta)}{\Gamma(k+1-2m/\theta)\Gamma\bigl(j+[(m-1)(m+\rho)-2m]/\theta\bigr)}, \tag{3.13}$$

and so (3.8) follows from (3.4) and (3.2). The formula (3.9) follows similarly by taking $j=2$ in (3.13) and using $B_1=1$.

To prove (3.10), we write for $i \geqslant 2$, using (3.6)–(3.7),

$$\begin{aligned}\mathbb{E}(1+(m-1)B_i)^2 &= 1+2(m-1)\,\mathbb{E}\,B_i+(m-1)^2\,\mathbb{E}\,B_i^2 \\ &= 1+2(m-1)\frac{\chi}{i}+O\bigl(i^{-2}\bigr) \\ &=: 1+y_i.\end{aligned}$$

Taking logarithms,

$$\log\Bigl(\prod_{i=j}^{k}(1+y_i)\Bigr) = \sum_{i=j}^{k}\log(1+y_i) \leqslant \sum_{i=j}^{k} y_i = 2(m-1)\chi\sum_{i=j}^{k}\bigl(i^{-1}+O(i^{-2})\bigr)$$
$$= 2(m-1)\chi\log\Bigl(\frac{k}{j}\Bigr)+O(j^{-1}). \tag{3.14}$$

This implies the inequality in (3.10). The bound on $\mathbb{E}(\Phi_k^2)$ in (3.11) follows from the definition in (3.5) and applying (3.10) with $j=2$. The upper bound on $\mathbb{E}(\Phi_k^{-2})$ in (3.11) can be proved similarly, where we can use $(1+x)^{-2} \leqslant 1-2x+3x^2$ for $x \geqslant 0$, and thus

$$\begin{aligned}\mathbb{E}\bigl[(1+(m-1)B_i)^{-2}\bigr] &\leqslant 1-2(m-1)\,\mathbb{E}\,B_i+3(m-1)^2\,\mathbb{E}\,B_i^2 \\ &= 1-2(m-1)\chi i^{-1}+O(i^{-2}),\end{aligned} \tag{3.15}$$

together with $\log(1-x) \leqslant -x$. Finally, by the Cauchy–Schwarz inequality and the just proven $\mathbb{E}(\Phi_k^{-2}) \leqslant Ck^{-2(m-1)\chi}$,

$$\mathbb{E}(\Phi_k^{-1}) \leqslant \sqrt{\mathbb{E}(\Phi_k^{-2})} \leqslant Ck^{-(m-1)\chi}, \tag{3.16}$$

which completes the proof of all three inequalities claimed in (3.11). □

3.1. **An infinite product.**

Lemma 3.3. *The infinite product*

$$\beta := \prod_{k=1}^{\infty}\frac{1+(m-1)B_k}{\mathbb{E}(1+(m-1)B_k)} = \lim_{k\to\infty}\frac{\Phi_k}{\mathbb{E}\,\Phi_k} \tag{3.17}$$

exists a.s. and in L^p for every $p<\infty$. Furthermore, $\mathbb{E}\,\beta = 1$ and $\beta>0$ a.s.

We have also, as $k\to\infty$,

$$k^{-(m-1)\chi}\Phi_k \xrightarrow{\text{a.s.}} \tilde{\beta} := \frac{m\cdot\Gamma\bigl(2-2m/\theta\bigr)}{\Gamma\bigl(2+[(m-1)(m+\rho)-2m]/\theta\bigr)}\beta. \tag{3.18}$$

Proof. Define for $k \geqslant 1$

$$\widetilde{M}_k := \frac{\Phi_k}{\mathbb{E}(\Phi_k)} = \prod_{i=1}^{k}\frac{1+(m-1)B_i}{\mathbb{E}(1+(m-1)B_i)}. \tag{3.19}$$

This is a product of independent random variables with mean 1, and thus a martingale. For every fixed integer $r>1$, we have by the binomial theorem, $|B_k|\leqslant 1$, and (3.6)–(3.7),

$$\begin{aligned}\mathbb{E}(1+(m-1)B_k)^r &= \sum_{j=0}^{r}\binom{r}{j}(m-1)^j\,\mathbb{E}\,B_k^j = 1+r(m-1)\,\mathbb{E}\,B_k + O(\mathbb{E}\,B_k^2)\\ &= 1+r(m-1)\,\mathbb{E}\,B_k+O(k^{-2})\\ &= (1+(m-1)\,\mathbb{E}\,B_k)^r+O(k^{-2}).\end{aligned}\tag{3.20}$$

Hence, for every $k\geqslant 1$,

$$\mathbb{E}\,\widetilde{M}_k^r = \prod_{i=1}^{k}\frac{\mathbb{E}(1+(m-1)B_i)^r}{(\mathbb{E}(1+(m-1)B_i))^r} = \prod_{i=1}^{k}\bigl(1+O(i^{-2})\bigr)\leqslant C_r \tag{3.21}$$

and thus the martingale $\widetilde{M}_k$ is L^r-bounded; consequently it converges in L^r, and thus in L^p for all real $0<p\leqslant r$. Since r is arbitrary, this holds for all $p>0$.

In particular, $\widetilde{M}_k\to\beta$ in L^1, which shows that $\mathbb{E}\,\beta=\lim_{k\to\infty}\mathbb{E}\,\widetilde{M}_k=1$.

The event $\{\beta=0\}$ is independent of any finite number of $B_1,B_2,\dots$, and is thus a tail event. The Kolmogorov zero-one law, see e.g. [10, Theorem 1.5.1], thus shows that $\mathbb{P}(\beta=0)=0$ or 1, but $\mathbb{P}(\beta=0)=1$ is impossible since $\mathbb{E}\,\beta=1$. Hence, $\beta>0$ a.s.

Finally, (3.18) follows by (3.17) and (3.9). □

3.2. **Estimates for $S_{n,k}$.** Below, let ψ_n be a positive function such that $\psi_n\leqslant n-1$ and $\psi_n\to\infty$ as $n\to\infty$ (we later choose $\psi_n=n/\log n$). The next lemma shows that w.h.p., for all $k\geqslant\psi_n$, the random variables $S_{n,k}$ are close enough to the constants $(k/n)^\chi$.

Lemma 3.4. *Let $S_{n,k}$ be as in (2.3) and ψ_n be as above. Define $\delta_n=\psi_n^{-\varepsilon}$ for some $\varepsilon\in(0,1/2)$. Then, there is a positive constant C such that*

$$\mathbb{P}\left[\max_{\lceil\psi_n\rceil\leqslant k<n}\left|S_{n,k}-\left(\frac{k}{n}\right)^{\chi}\right|\geqslant 2\delta_n\right]\leqslant C{\psi_n}^{2\varepsilon-1}. \tag{3.22}$$

The proof of Lemma 3.4 is based on a standard martingale argument that is similar to [17] (see also [3]), but we present it here for completeness. To prepare for the main proof, we start by estimating $\mathbb{E}(S_{n,k})$.

Lemma 3.5. *Let $S_{n,k}$ be as in (2.3). For every $1\leqslant k\leqslant n-1$, we have*

$$\mathbb{E}(S_{n,k})=\frac{\Gamma\bigl(n-(3m+\rho)/\theta\bigr)}{\Gamma\bigl(n-2m/\theta\bigr)}\frac{\Gamma\bigl(k+1-2m/\theta\bigr)}{\Gamma\bigl(k+1-(3m+\rho)/\theta\bigr)}. \tag{3.23}$$

Proof. Recalling that $(B_j)_{j=2}^{n-1}$ are independent, we obtain from (3.6)

$$\begin{aligned}\mathbb{E}\,S_{n,k} &= \prod_{j=k+1}^{n-1}\mathbb{E}(1-B_i)=\prod_{j=k+1}^{n-1}\frac{\theta j-3m-\rho}{\theta j-2m}=\prod_{j=k+1}^{n-1}\frac{j-(3m+\rho)/\theta}{j-2m/\theta}\\ &= \frac{\Gamma\bigl(n-(3m+\rho)/\theta\bigr)}{\Gamma\bigl(n-2m/\theta\bigr)}\frac{\Gamma\bigl(k+1-2m/\theta\bigr)}{\Gamma\bigl(k+1-(3m+\rho)/\theta\bigr)},\end{aligned}\tag{3.24}$$

as claimed in the lemma. □

Lemma 3.6. *Let $S_{n,k}$ be as in (2.3). Then, there is a positive constant C such that, for $1\leqslant k\leqslant n-1$,*

$$\left|\mathbb{E}(S_{n,k})-\left(\frac{k}{n}\right)^{\chi}\right|\leqslant\frac{C}{n^\chi k^{1-\chi}}. \tag{3.25}$$

Proof. By (3.23) and (3.4), we have, recalling (3.2),

$$\mathbb{E}(S_{n,k}) = n^{-\chi}\big(1+O(n^{-1})\big)k^{\chi}\big(1+O(k^{-1})\big) = \Big(\frac{k}{n}\Big)^{\chi}\big(1+O(k^{-1})\big), \tag{3.26}$$

which yields (3.25). □

Proof of Lemma 3.4. By Lemma 3.6, for n large enough and any $k \in [\psi_n, n]$,

$$\left|\mathbb{E}(S_{n,k}) - \Big(\frac{k}{n}\Big)^{\chi}\right| \leqslant \frac{C}{n^{\chi}k^{1-\chi}} \leqslant \frac{C}{k} \leqslant \frac{C}{\psi_n} < \delta_n. \tag{3.27}$$

Hence,

$$\begin{aligned}\mathbb{P}\left[\max_{\lceil\psi_n\rceil\leqslant k<n}\left|S_{n,k} - \Big(\frac{k}{n}\Big)^{\chi}\right| \geqslant 2\delta_n\right] &\leqslant \mathbb{P}\left[\max_{\lceil\psi_n\rceil\leqslant k<n}\big|S_{n,k} - \mathbb{E}(S_{n,k})\big| \geqslant \delta_n\right]\\ &\leqslant \mathbb{P}\left[\max_{\lceil\psi_n\rceil\leqslant k<n}\big|S_{n,k}/\,\mathbb{E}(S_{n,k}) - 1\big| \geqslant \delta_n\right],\end{aligned} \tag{3.28}$$

noting that $\mathbb{E}\,S_{n,k} \leqslant 1$ for $k \geqslant 1$. To bound the right-hand side of (3.28), we first observe that for $k \geqslant 0$,

$$\widehat{M}_k := \prod_{j=n-k}^{n-1} \frac{1-B_j}{\mathbb{E}(1-B_j)} = \frac{S_{n,n-1-k}}{\mathbb{E}\,S_{n,n-1-k}} \tag{3.29}$$

is a martingale with respect to the σ-algebras generated by $(B_j)_{j=n-k}^{n-1}$, with $\mathbb{E}\,\widehat{M}_k = 1$. Now, by Doob's inequality for the submartingale $(\widehat{M}_k - 1)^2$, see e.g. [10, Theorem 10.9.1],

$$\begin{aligned}\mathbb{P}\left[\max_{\lceil\psi_n\rceil\leqslant k<n}\big|S_{n,k}/\,\mathbb{E}(S_{n,k}) - 1\big| \geqslant \delta_n\right] &= \mathbb{P}\left[\max_{0\leqslant k\leqslant n-1-\lceil\psi_n\rceil}|\widehat{M}_k - 1| \geqslant \delta_n\right]\\ &\leqslant \delta_n^{-2}\mathrm{Var}\big(\widehat{M}_{n-1-\lceil\psi_n\rceil}\big).\end{aligned} \tag{3.30}$$

Using $\mathbb{E}\,\widehat{M}_{n-1-\lceil\psi_n\rceil} = 1$ and the independence of the beta variables, we have

$$\mathrm{Var}\big(\widehat{M}_{n-1-\lceil\psi_n\rceil}\big) = \mathbb{E}\left(\widehat{M}^2_{n-1-\lceil\psi_n\rceil}\right) - 1 = \prod_{k=\lceil\psi_n\rceil+1}^{n-1} \frac{\mathbb{E}[(1-B_k)^2]}{(\mathbb{E}[1-B_k])^2} - 1, \tag{3.31}$$

and by (2.1), (3.3), and simplifying, we get

$$\begin{aligned}\mathrm{Var}\big(\widehat{M}_{n-1-\lceil\psi_n\rceil}\big) &= \prod_{k=\lceil\psi_n\rceil+1}^{n-1}\Big(\frac{\theta k - 3m - \rho + 1}{\theta k - 3m - \rho}\cdot\frac{\theta k - 2m}{\theta k - 2m + 1}\Big) - 1\\ &\leqslant \prod_{k=\lceil\psi_n\rceil+1}^{n-1}(1 + Ck^{-2}) - 1 \leqslant \frac{C}{\psi_n}.\end{aligned} \tag{3.32}$$

Applying (3.32) and $\delta_n = \psi_n^{-\varepsilon}$ to (3.28) and (3.30) yields

$$\mathbb{P}\left[\max_{\lceil\psi_n\rceil\leqslant k<n}\left|S_{n,k} - \Big(\frac{k}{n}\Big)^{\chi}\right| \geqslant 2\delta_n\right] \leqslant \delta_n^{-2}\frac{C}{\psi_n} = C\psi_n^{2\varepsilon-1},$$

hence proving the lemma. □

3.3. **Asymptotics of two sums.** Fix a sequence $\lambda_n \to \infty$ (we will later choose $\lambda_n = n^{\nu}$) and define, for $y \geqslant 0$ and $0 \leqslant k \leqslant \ell < \infty$,

$$H^y_{k,\ell} := \sum_{i=k+1}^{\ell}\big[(1-B_i)^{\lambda_n y} - 1 + \lambda_n y B_i\big], \qquad I^y_{k,\ell} := \sum_{i=k+1}^{\ell}\big[1 - (1-B_i)^{\lambda_n y}\big] \tag{3.33}$$

and, for $y \geqslant 0$ and $0 \leqslant s \leqslant t < \infty$,

$$\widehat{H}^y_{s,t} := \lambda_n^{-1}H^y_{\lfloor s\lambda_n\rfloor,\lfloor t\lambda_n\rfloor}, \qquad \widehat{I}^y_{s,t} := \lambda_n^{-1}I^y_{\lfloor s\lambda_n\rfloor,\lfloor t\lambda_n\rfloor}. \tag{3.34}$$

Lemma 3.7. *Let $0 < s \leqslant t < \infty$. Then, for every $y \geqslant 0$, as $n \to \infty$,*

$$\widehat{H}^y_{s,t} \overset{\mathrm{p}}{\longrightarrow} \int_s^t \Bigl(\frac{(\theta u)^{m+\rho}}{(\theta u + y)^{m+\rho}} - 1 + \frac{\chi y}{u}\Bigr) \,\mathrm{d}u = \int_s^t \Bigl(\frac{1}{(1 + y/(\theta u))^{m+\rho}} - 1 + \frac{\chi y}{u}\Bigr) \,\mathrm{d}u \tag{3.35}$$

and

$$\widehat{I}^y_{s,t} \overset{\mathrm{p}}{\longrightarrow} \int_s^t \Bigl(1 - \frac{(\theta u)^{m+\rho}}{(\theta u + y)^{m+\rho}}\Bigr) \,\mathrm{d}u = \int_s^t \Bigl(1 - \frac{1}{(1 + y/(\theta u))^{m+\rho}}\Bigr) \,\mathrm{d}u. \tag{3.36}$$

Proof. Denote the summand in (3.33) by ΔH_i. Then $-1 \leqslant \Delta H_i \leqslant \lambda_n y B_i$, and thus, using (3.7), for $\lfloor s\lambda_n \rfloor < i \leqslant \lfloor \lambda_n t \rfloor$,

$$\operatorname{Var}(\Delta H_i) \leqslant \mathbb{E}(\Delta H_i)^2 \leqslant C + C\lambda_n^2 \,\mathbb{E}\, B_i^2 \leqslant C_s. \tag{3.37}$$

The summands ΔH_i are independent, and thus (3.33)–(3.34) and (3.37) yield

$$\operatorname{Var}(\widehat{H}^y_{s,t}) = \lambda_n^{-2} \sum_{i=\lfloor s\lambda_n \rfloor + 1}^{\lfloor t\lambda_n \rfloor} \operatorname{Var}\bigl(\Delta H_i\bigr) \leqslant C_{s,t}\lambda_n^{-1} = o(1). \tag{3.38}$$

Hence, it suffices to show that the expectation $\mathbb{E}\, \widehat{H}^y_{s,t}$ converges to the limit in (3.35). We have, applying (3.3) to $1 - B_i \in \operatorname{Beta}(\theta i - 3m - \rho, m + \rho)$ and using (3.4), uniformly for $s\lambda_n < i \leqslant t\lambda_n$,

$$\begin{aligned}
\mathbb{E}(1 - B_i)^{\lambda_n y} &= \frac{\Gamma(\theta i - 2m)\Gamma(\theta i - 3m - \rho + \lambda_n y)}{\Gamma(\theta i - 3m - \rho)\Gamma(\theta i - 2m + \lambda_n y)} = \Bigl(\frac{\theta i}{\theta i + \lambda_n y}\Bigr)^{m+\rho} + o(1) \\
&= \Bigl(\frac{\theta i/\lambda_n}{\theta i/\lambda_n + y}\Bigr)^{m+\rho} + o(1).
\end{aligned} \tag{3.39}$$

Hence, using also (3.6), if $i = u\lambda_n$ with $u \in (s, t]$,

$$\begin{aligned}
\mathbb{E}[\Delta H_i] &= \mathbb{E}(1 - B_i)^{\lambda_n y} - 1 + \lambda_n y \,\mathbb{E}\, B_i \\
&= \Bigl(\frac{\theta i/\lambda_n}{\theta i/\lambda_n + y}\Bigr)^{m+\rho} - 1 + \frac{\chi}{i}\lambda_n y + o(1).
\end{aligned} \tag{3.40}$$

It follows that $\mathbb{E}\, \widehat{H}^y_{s,t}$ is $o(1)$ plus a Riemann sum of the integral in (3.35).

The proof of (3.36) is similar, where we now replace ΔH_i with

$$\Delta I_i = 1 - (1 - B_i)^{\lambda_n y} \tag{3.41}$$

and, for $s\lambda_n < i \leqslant t\lambda_n$, use the estimates $0 \leqslant \Delta I_i \leqslant 1$ and thus, using (3.39),

$$\operatorname{Var}(\Delta I_i) \leqslant \mathbb{E}(\Delta I_i)^2 \leqslant 1, \tag{3.42}$$

$$\mathbb{E}[\Delta I_i] = 1 - \mathbb{E}(1 - B_i)^{\lambda_n y} = 1 - \Bigl(\frac{\theta i/\lambda_n}{\theta i/\lambda_n + y}\Bigr)^{m+\rho} + o(1) \tag{3.43}$$

to proceed. □

4. Basic analysis

In this and subsequent sections, we follow the framework (and hence the notation) in [13]. To concentrate on the important aspects of the proof, we assume that $m = 2$ and $\rho = 0$; note that then $\chi = \frac{1}{2}$ and $\theta = 2m = 4$. The minor modifications for the general case are discussed in Section 10.

4.1. **The stochastic recursion.** Let D_n be the subgraph in G_n, consisting of vertex n, all vertices that can be reached from n via a directed path, and all the edges between them. We think of the vertices and edges in D_n as coloured red. We use the following stochastic recursion to construct D_n. It is similar to the recursion used in [13], with differences that stem from the difference of the models.

(1) Sample the beta variables $(B_j)_{j=2}^{n-1}$ defined in (2.1).

(2) Declare vertex n to be red and all others black. Initiate the recursion by setting $k := n$.

(3) If vertex k is red, choose the recipients of the two outgoing edges from vertex k according to the construction given in Definition 2.1. After sampling the recipients, declare them as red.
If vertex k is black, delete k and do nothing else.

(4) If $k = 2$ then stop; otherwise let $k := k - 1$ and repeat from (3).

For integers $0 \leqslant k \leqslant n-1$, let Y_k be the number of edges in D_n that start from $\{k+1, \dots, n\}$ and end in $\{1, \dots, k\}$. Define Z_k as the number of edges in Y_k that end in k. Note that we have the boundary conditions $Y_{n-1} = 2$ and $Y_0 = 0$; as well as $Z_1 = Y_1$ and $Z_0 = 0$.

For $1 \leqslant k \leqslant n-1$, denote the indicator that at least one edge of Y_k ends at k as

$$J_k = \mathbf{1}[Z_k \geqslant 1], \tag{4.1}$$

which is the same as the indicator that k is red. Thus summing J_k over $k \in [n-1]$ gives the number $X^{(n)}$ of red vertices. For $2 \leqslant k \leqslant n-1$, the number of edges that start at k is $2J_k$, and we thus have

$$Y_{k-1} = Y_k - Z_k + 2 \cdot J_k = Y_k - Z_k + 2 \cdot \mathbf{1}[Z_k \geqslant 1]. \tag{4.2}$$

As in [13], we use a modified version of the procedure above, where we use the construction in Remark 2.2. In (3) above, we thus do not choose the recipients of the outgoing edges, we just note that they have endpoints in $[k-1]$. We then at the next vertex toss a coin for each edge with unassigned endpoint to decide whether it ends there or not. This yields the following equivalent version of the construction.

(1) Sample the beta variables $(B_j)_{j=2}^{n-1}$ defined in (2.1).

(2) Declare vertex n to be red and all others black. Initiate the recursion by setting $k := n$.

(3) If vertex k is red, add two outgoing edges from vertex k, with as yet undetermined endpoints in $[k-1]$; mark these edges *incomplete*.

(4) Let $k := k - 1$.

(5) For each incomplete edge, toss a coin with heads probability B_k, independently given B_k. If the outcome is heads, the edge ends at k and is marked complete; furthermore, vertex k is coloured red. Otherwise do nothing (so the edge is still incomplete).

(6) If $k = 1$ then stop; otherwise repeat from (3).

Let $\mathcal{F}_k$ be the σ-field generated by all beta variables $(B_j)_{j=2}^{n-1}$ and the coin tosses at vertices $n-1, \dots, k+1$. Then $\mathcal{F}_1, \dots, \mathcal{F}_{n-1}$ forms a decreasing sequence of σ-fields, and $Y_{n-1}, \dots, Y_k$ are measurable with respect to $\mathcal{F}_k$. Moreover, conditioned on $\mathcal{F}_k$, we have

$$Z_k \mid \mathcal{F}_k \in \mathrm{Bin}(Y_k, B_k) \quad \text{for } 1 \leqslant k \leqslant n-1. \tag{4.3}$$

Thus, in view of the recursion (4.2), we have for $2 \leqslant k \leqslant n-1$,

$$\begin{aligned}\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) &= Y_k - \mathbb{E}(Z_k \mid \mathcal{F}_k) + 2 \cdot \mathbb{P}(Z_k \geqslant 1 \mid \mathcal{F}_k) \\ &= Y_k - B_k Y_k + 2\bigl(1-(1-B_k)^{Y_k}\bigr). \end{aligned} \tag{4.4}$$

By Markov's inequality, we also have for $2 \leqslant k \leqslant n-1$,

$$\begin{aligned}\mathbb{E}[Y_{k-1} \mid \mathcal{F}_k] \leqslant Y_k - \mathbb{E}[Z_k \mid \mathcal{F}_k] + 2 \cdot \mathbb{E}[Z_k \mid \mathcal{F}_k] = Y_k + \mathbb{E}[Z_k \mid \mathcal{F}_k] \\ =(1+B_k)Y_k. \end{aligned} \tag{4.5}$$

Define, recalling (3.5),

$$W_k = \Phi_k Y_k \quad \text{for } 0 \leqslant k \leqslant n-1, \tag{4.6}$$

noting that $W_0 = 2Y_0 = 0$. Using (4.5) and (3.5), we find for $2 \leqslant k \leqslant n-1$,

$$\mathbb{E}(W_{k-1} \mid \mathcal{F}_k) = \Phi_{k-1}\,\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) \leqslant \Phi_{k-1}(1+B_k)Y_k = \Phi_k Y_k = W_k; \tag{4.7}$$

and so $W_0, \dots, W_{n-1}$ is a reverse supermartingale. The initial value is

$$W_{n-1} = \Phi_{n-1}Y_{n-1} = 2\Phi_{n-1}. \tag{4.8}$$

By Doob's decomposition,

$$W_k = M_k - A_k, \quad 0 \leqslant k \leqslant n-1, \tag{4.9}$$

where

$$M_k := 2\Phi_{n-1} + \sum_{j=k+1}^{n-1} \bigl(W_{j-1} - \mathbb{E}(W_{j-1} \mid \mathcal{F}_j)\bigr) \tag{4.10}$$

is a reverse martingale and

$$A_k := \sum_{j=k+1}^{n-1} \bigl(W_j - \mathbb{E}(W_{j-1} \mid \mathcal{F}_j)\bigr) \tag{4.11}$$

is positive and reverse increasing. To see these properties of A_k, we note $A_{n-1} = 0$ and by (4.7),

$$A_{k-1} - A_k = W_k - \mathbb{E}(W_{k-1} \mid \mathcal{F}_k) \geqslant 0 \quad \text{for } 1 \leqslant k \leqslant n-1. \tag{4.12}$$

Hence, for $0 \leqslant k \leqslant n-1$, we have $0 \leqslant W_k \leqslant M_k$.

From the exact formula (4.4),

$$\mathbb{E}(W_{k-1} \mid \mathcal{F}_k) = \Phi_{k-1}\,\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) = \Phi_{k-1}(1-B_k)Y_k + 2\Phi_{k-1}\bigl(1-(1-B_k)^{Y_k}\bigr), \tag{4.13}$$

and so (4.12) can be written as

$$\begin{aligned}A_{k-1} - A_k &= 2B_k\Phi_{k-1}Y_k - 2\Phi_{k-1}\bigl(1-(1-B_k)^{Y_k}\bigr) \\ &= 2\Phi_{k-1}\bigl((1-B_k)^{Y_k} - 1 + B_kY_k\bigr). \end{aligned} \tag{4.14}$$

Following the steps in [13, equation (2.16)] for evaluating $\operatorname{Var}(Y_{k-1} \mid \mathcal{F}_k)$ in the uniform attachment case, here we have, for $1 \leqslant k \leqslant n-1$,

$$\begin{aligned}&\operatorname{Var}(Y_{k-1} \mid \mathcal{F}_k) \\ &\quad= \operatorname{Var}(Z_k - 2\cdot \mathbf{1}[Z_k \geqslant 1] \mid \mathcal{F}_k) \leqslant 2\operatorname{Var}(Z_k \mid \mathcal{F}_k) + 2\operatorname{Var}(2\cdot\mathbf{1}[Z_k \geqslant 1] \mid \mathcal{F}_k) \\ &\quad\leqslant 2B_kY_k + 8\mathbb{P}(Z_k \geqslant 1 \mid \mathcal{F}_k) \leqslant 2B_kY_k + 8\,\mathbb{E}(Z_k \mid \mathcal{F}_k) = 10B_kY_k. \end{aligned} \tag{4.15}$$

Thus,

$$\operatorname{Var}(W_{k-1} \mid \mathcal{F}_k) = \Phi_{k-1}^2 \operatorname{Var}(Y_{k-1} \mid \mathcal{F}_k) \leqslant 10\Phi_{k-1}^2 B_kY_k. \tag{4.16}$$

Let $\mathbf{B}$ be the σ-field generated by the beta variables $(B_j)_{j=2}^{n-1}$, and let $\mathbb{E}_{\mathbf{B}}$ and $\operatorname{Var}_{\mathbf{B}}$ denote conditional expectation and variance with respect to $\mathbf{B}$. Note that $M_0, \dots, M_{n-1}$

is a reverse martingale also conditioned on $\mathbf{B}$, since $\mathbf{B} = \mathcal{F}_{n-1} \subseteq \mathcal{F}_k$ for every k. In particular,

$$\mathbb{E}_{\mathbf{B}} W_k \leqslant \mathbb{E}_{\mathbf{B}} M_k = M_{n-1} = 2\Phi_{n-1} \quad \text{for } 0 \leqslant k \leqslant n-1. \tag{4.17}$$

Hence, by applying (4.10), the reverse martingale property, (4.16), (3.5), (4.6), and then (4.17), for $0 \leqslant k \leqslant n-1$,

$$\begin{aligned}
\operatorname{Var}_{\mathbf{B}}(M_k) = \mathbb{E}_{\mathbf{B}}(M_k - 2\Phi_{n-1})^2 &= \sum_{j=k+1}^{n-1} \mathbb{E}_{\mathbf{B}} \operatorname{Var}(W_{j-1} \mid \mathcal{F}_j) \\
&\leqslant 10 \sum_{j=k+1}^{n-1} \Phi_{j-1}^2 B_j \, \mathbb{E}_{\mathbf{B}}(Y_j) = 10 \sum_{j=k+1}^{n-1} \frac{B_j}{1+B_j} \Phi_{j-1} \, \mathbb{E}_{\mathbf{B}}(W_j) \\
&\leqslant 20 \sum_{j=k+1}^{n-1} \frac{B_j}{1+B_j} \Phi_{j-1}\Phi_{n-1} = 20 \sum_{j=k+1}^{n-1} \Phi_{j-1}^2 B_j \prod_{i=j+1}^{n-1} (1+B_i).
\end{aligned} \tag{4.18}$$

4.2. **Some estimates.** Below we provide several estimates for W_k, M_k, Z_k and A_k that we need later. The results are analogous to [13, Lemmas 2.1–2.3]. Recall that $\chi = 1/2$ for $m = 2$ and $\rho = 0$.

Lemma 4.1. *For $1 \leqslant k \leqslant n-1$, we have*

$$\mathbb{E}_{\mathbf{B}} W_k^2 \leqslant \mathbb{E}_{\mathbf{B}} M_k^2 \leqslant 20 \sum_{j=k+1}^{n-1} \Phi_{j-1}^2 B_j \prod_{i=j+1}^{n-1} (1+B_i) + 4\Phi_{n-1}^2. \tag{4.19}$$

Furthermore,

$$\mathbb{E}_{\mathbf{B}} \max_{0\leqslant k\leqslant n-1} W_k^2 \leqslant \mathbb{E}_{\mathbf{B}} \max_{0\leqslant k\leqslant n-1} M_k^2 \leqslant 4\, \mathbb{E}_{\mathbf{B}} M_0^2 \tag{4.20}$$

and there is a positive constant C such that

$$\mathbb{E} \max_{0\leqslant k\leqslant n-1} W_k^2 \leqslant \mathbb{E} \max_{0\leqslant k\leqslant n-1} M_k^2 \leqslant Cn. \tag{4.21}$$

Proof. The first inequalities in (4.19) and (4.20) follow from $0 \leqslant W_k \leqslant M_k$. For the second inequality in (4.19), note that

$$\mathbb{E}_{\mathbf{B}} M_k^2 = \operatorname{Var}_{\mathbf{B}}(M_k) + (\mathbb{E}_{\mathbf{B}} M_k)^2, \tag{4.22}$$

and the inequality follows from this by the inequality (4.18) and the equalities in (4.17).

The second inequality in (4.20) follows from Doob's inequality. By (4.20), (4.21) follows from showing $\mathbb{E} M_0^2 \leqslant Cn$. For every $0 \leqslant k \leqslant n-1$, we use (4.18) and the independence of the beta variables to obtain

$$\begin{aligned}
\mathbb{E}\big(\operatorname{Var}_{\mathbf{B}}(M_k)\big) &\leqslant 20\, \mathbb{E} \sum_{j=k+1}^{n-1} \Phi_{j-1}^2 B_j \prod_{i=j+1}^{n-1} (1+B_i) \\
&= 20 \sum_{j=k+1}^{n-1} \mathbb{E}\, \Phi_{j-1}^2 \, \mathbb{E}\, B_j \prod_{i=j+1}^{n-1} \mathbb{E}(1+B_i).
\end{aligned} \tag{4.23}$$

So applying (3.6), (3.8) and (3.11), we get

$$\mathbb{E}\big(\operatorname{Var}_{\mathbf{B}}(M_k)\big) \leqslant C \sum_{j=k+1}^{n-1} \frac{j}{j} \cdot \Big(\frac{n}{j}\Big)^{1/2} \leqslant Cn^{1/2} \sum_{j=k+1}^{n-1} \frac{1}{j^{1/2}} \leqslant Cn. \tag{4.24}$$

Using the equalities in (4.17) and the estimate in (3.11), we also have

$$\mathbb{E}\big[(\mathbb{E}_{\mathbf{B}} M_k)^2\big] = \mathbb{E}\big(M_{n-1}^2\big) = 4\, \mathbb{E}(\Phi_{n-1}^2) \leqslant Cn \quad \text{for } 0 \leqslant k \leqslant n-1. \tag{4.25}$$

Applying (4.24) and (4.25) to (4.22), with $k=0$, we thus have

$$\mathbb{E}\, M_0^2 \leqslant \mathbb{E}\big[\operatorname{Var}_{\mathbf{B}}(M_0)\big] + \mathbb{E}\big[(\mathbb{E}_{\mathbf{B}}\, M_0)^2\big] \leqslant Cn, \tag{4.26}$$

which together with (4.20) implies (4.21). □

Lemma 4.2. *There is a positive constant C such that, for $1 \leqslant k \leqslant n-1$,*

$$\mathbb{P}(Z_k \geqslant 1) \leqslant C\frac{n^{1/2}}{k^{3/2}}; \tag{4.27}$$

$$\mathbb{P}(Z_k \geqslant 2) \leqslant C\frac{n}{k^3}. \tag{4.28}$$

Proof. We start by proving (4.27). Firstly, it follows from (4.3) and (4.6) that

$$\mathbb{E}(Z_k \mid \mathcal{F}_k) = Y_k B_k = \Phi_k^{-1} B_k W_k, \tag{4.29}$$

which, along with (4.17), imply that

$$\mathbb{E}_{\mathbf{B}}(Z_k) = \Phi_k^{-1} B_k\, \mathbb{E}_{\mathbf{B}}(W_k) \leqslant 2\Phi_k^{-1} B_k \Phi_{n-1} = 2B_k \prod_{i=k+1}^{n-1}(1+B_i). \tag{4.30}$$

Using the independence of $(B_k)_{k=2}^{n-1}$, (3.6) and (3.8), we therefore have

$$\mathbb{E}(Z_k) \leqslant 2\,\mathbb{E}(B_k) \prod_{i=k+1}^{n-1} \mathbb{E}(1+B_i) \leqslant \frac{C}{k}\Big(\frac{n}{k}\Big)^{1/2} = C\frac{n^{1/2}}{k^{3/2}}, \tag{4.31}$$

and (4.27) follows from Markov's inequality.

The proof for (4.28) is similar, this time we observe that by Markov's inequality,

$$\mathbb{P}(Z_k \geqslant 2 \mid \mathcal{F}_k) \leqslant \mathbb{E}\bigg(\binom{Z_k}{2} \,\bigg|\, \mathcal{F}_k\bigg) = \binom{Y_k}{2} B_k^2 \leqslant B_k^2 Y_k^2 = B_k^2 \Phi_k^{-2} W_k^2. \tag{4.32}$$

By (4.19) of Lemma 4.1 and (3.5), we have

$$\begin{aligned}
&\mathbb{E}\left(B_k^2 \Phi_k^{-2} W_k^2\right) = \mathbb{E}\left(B_k^2 \Phi_k^{-2}\, \mathbb{E}_{\mathbf{B}}(W_k^2)\right)\\
&\leqslant 20\,\mathbb{E}\bigg(B_k^2\Phi_k^{-2} \sum_{j=k+1}^{n-1} \Phi_{j-1}^2 B_j \prod_{i=j+1}^{n-1}(1+B_i)\bigg) + 4\,\mathbb{E}\big(B_k^2 \Phi_k^{-2}\Phi_{n-1}^2\big)\\
&= 20\,\mathbb{E}\bigg(B_k^2 \sum_{j=k+1}^{n-1} B_j \prod_{i=k+1}^{j-1}(1+B_i)^2 \prod_{l=j+1}^{n-1}(1+B_l)\bigg) + 4\,\mathbb{E}\Big(B_k^2 \prod_{i=k+1}^{n-1}(1+B_i)^2\Big)\\
&= 20\,\mathbb{E}(B_k^2) \sum_{j=k+1}^{n-1} \mathbb{E}(B_j) \prod_{i=k+1}^{j-1} \mathbb{E}(1+B_i)^2 \prod_{l=j+1}^{n-1} \mathbb{E}(1+B_l) + 4\,\mathbb{E}(B_k^2) \prod_{i=k+1}^{n-1} \mathbb{E}(1+B_i)^2
\end{aligned} \tag{4.33}$$

Applying the estimates in (3.7), (3.8) and (3.10) to (4.33), we find

$$\mathbb{E}\left(B_k^2 \Phi_k^{-2} W_k^2\right) \leqslant \frac{C}{k^2} \sum_{j=k+1}^{n-1} \frac{1}{j}\cdot\frac{j}{k}\cdot\Big(\frac{n}{j}\Big)^{1/2} + \frac{Cn}{k^3} \leqslant \frac{Cn}{k^3}. \tag{4.34}$$

Taking the expectation in (4.32) and plugging in (4.34) yields (4.28). □

Lemma 4.3. *For $1 \leqslant k \leqslant n-1$,*

$$A_{k-1} - A_k \leqslant (W_k B_k)^2 \Phi_k^{-1} \tag{4.35}$$

and there is a positive constant C such that

$$\mathbb{E}\, A_k \leqslant \frac{Cn}{k^{3/2}}. \tag{4.36}$$

Proof. By (4.14), Taylor's formula, the increasing property of Φ_k, and (4.6),

$$\begin{aligned} A_{k-1} - A_k = 2\Phi_{k-1}\big((1-B_k)^{Y_k} - 1 + B_k Y_k\big) &\leqslant 2\Phi_{k-1}\binom{Y_k}{2}B_k^2 \\ &\leqslant Y_k^2\Phi_{k-1}B_k^2 \leqslant Y_k^2\Phi_k B_k^2 = W_k^2\Phi_k^{-1}B_k^2, \end{aligned} \tag{4.37}$$

yielding (4.35). To prove (4.36), we note that by a telescoping argument, (4.35) implies

$$A_k \leqslant \sum_{i=k+1}^{n-1} \big(W_i B_i\big)^2 \Phi_i^{-1} \tag{4.38}$$

Furthermore, (4.19) of Lemma 4.1 and (3.5) together yield

$$\begin{aligned} &\mathbb{E}_{\mathbf{B}}\big(W_i^2 B_i^2 \Phi_i^{-1}\big) \\ &\leqslant 20 B_i^2 \Phi_i^{-1} \sum_{j=i+1}^{n-1} \Phi_{j-1}^2 B_j \prod_{h=j+1}^{n-1}(1+B_h) + 4B_i^2\Phi_i^{-1}\Phi_{n-1}^2 \\ &= 20 B_i^2 \Phi_i \sum_{j=i+1}^{n-1} B_j \prod_{l=i+1}^{j-1}(1+B_l)^2 \prod_{h=j+1}^{n-1}(1+B_h) + 4B_i^2\Phi_i \prod_{h=i+1}^{n-1}(1+B_h)^2 \\ &\leqslant 40 B_i^2 \Phi_{i-1} \sum_{j=i+1}^{n-1} B_j \prod_{l=i+1}^{j-1}(1+B_l)^2 \prod_{h=j+1}^{n-1}(1+B_h) + 8B_i^2\Phi_{i-1} \prod_{h=i+1}^{n-1}(1+B_h)^2. \end{aligned} \tag{4.39}$$

Taking expectation and again using the independence of $(B_k)_{k=2}^{n-1}$, we have

$$\begin{aligned} \mathbb{E}\big(W_i^2 B_i^2 \Phi_i^{-1}\big) \leqslant 40\,\mathbb{E}(B_i^2)\,\mathbb{E}(\Phi_{i-1}) \sum_{j=i+1}^{n-1} \mathbb{E}(B_j) \prod_{l=i+1}^{j-1} \mathbb{E}(1+B_l)^2 \prod_{h=j+1}^{n-1} \mathbb{E}(1+B_h) \\ + 8\,\mathbb{E}(B_i^2)\,\mathbb{E}(\Phi_{i-1}) \prod_{h=i+1}^{n-1} \mathbb{E}(1+B_h)^2. \end{aligned} \tag{4.40}$$

Applying (3.6), (3.7), (3.8), (3.9) and (3.10) to the last display gives

$$\mathbb{E}\big(W_i^2 B_i^2 \Phi_i^{-1}\big) \leqslant \frac{C}{i^{3/2}} \sum_{j=i+1}^{n-1} \frac{1}{2j-2} \cdot \frac{j}{i} \cdot \Big(\frac{n}{j}\Big)^{1/2} + \frac{Cn}{i^{5/2}} \leqslant \frac{Cn}{i^{5/2}}. \tag{4.41}$$

Thus, in view of (4.41) and (4.38), we deduce that for $1 \leqslant k \leqslant n-1$,

$$\mathbb{E}(A_k) \leqslant \sum_{i=k+1}^{n-1} \frac{Cn}{i^{5/2}} \leqslant \frac{Cn}{k^{3/2}}, \tag{4.42}$$

hence proving (4.36). ☐

5. The early part and a Yule process

We continue to study the case $m = 2$ and $\rho = 0$, and recall that then $\chi = 1/2$. We show that the early part of the growth of D_n can be closely coupled to the same time-changed Yule process as in [13], and use this coupling to study Y_k and W_k. We start by presenting its construction and key features, following the description in [13].

Let $\mathcal{Y}$ be a Yule process starting with two particles, and let $\mathcal{Y}_t$ be the number of (living) particles at time t (thus $\mathcal{Y}_0 = 2$). Note that $\mathcal{Y}_t$ has the same distribution as the sum of two copies of the standard Yule process, which starts with a single particle. (See e.g. [1, Section III.5] for definition and some basic properties.) To better compare the process to D_n, it is convenient too to view the Yule process $\mathcal{Y}$ as a tree, where the root $\gamma_0 := 0$ marks the beginning of the process and the vertex γ_i marks the time of the i-th particle split in the process. Note also these split times are a.s. distinct. In this way, each particle can be represented by an edge from its time of birth to its time of death.

The time-changed Yule tree $\widehat{\mathcal{Y}}$ appearing in [13] is obtained by applying the mapping $t \to e^{-t}$, so that the vertices in $\widehat{\mathcal{Y}}$ have labels $e^{-\gamma_i} \in (0,1]$. Hence, the root in $\widehat{\mathcal{Y}}$ has label 1, and a particle in the original Yule process that is born at time γ_i and has lifetime $\tau \in \mathrm{Exp}(1)$ is now represented by an edge from $x = e^{-\gamma_i}$ to $e^{-(\gamma_i+\tau)} = xe^{-\tau} = xU$, where $U := e^{-\tau} \in \mathsf{U}(0,1)$. In light of this, as well as that $e^{-\gamma_0} = 1$ and all lifetimes in the original Yule process are independent and have the Exp(1) distribution, any vertex in $\widehat{\mathcal{Y}}$ that is d generations away from the root therefore take labels of the form $\widehat{U}_1 \cdots \widehat{U}_d$, where $\widehat{U}_1, \dots, \widehat{U}_d \in \mathsf{U}(0,1)$ and are independent.

Let $\widehat{D}_n$ be the random red graph D_n with label k replaced with $(k/n)^\chi$, so that the labels now take values in $(0,1]$. We regard D_n as rooted at n; thus the root of $\widehat{D}_n$ has label 1. We shall compare the time-changed Yule tree $\widehat{\mathcal{Y}}$ to $\widehat{D}_n$, considering only vertices with large enough labels. In preparation, let

$$n_1 = n_1^{(n)} := \lfloor n/\log n \rfloor. \tag{5.1}$$

We will use the construction of G_n in (2.3)–(2.6), using the variables $S_{n,k}$ defined there; in particular, recall that $\widetilde{U}_{k,\ell}$ are independent $\mathsf{U}(0,1)$ random variables.

Lemma 5.1. *Let $\kappa_n := \log n/n^{1/3}$. With probability at least $1-C\log n/n^{1/3}$, the following hold:*

(i) *For every path in D_n between vertex n (the root) and a vertex $k > n_1$ consisting of $d+1 \geqslant 2$ red vertices $n = v_0 > v_1 > \cdots > v_{d-1} > v_d = k$ such that $v_i \xrightarrow{\ell_i} v_{i+1}$ for $0 \leqslant i < d$, we have*

$$\bigl|\widetilde{U}_{v_0,\ell_0} \cdots \widetilde{U}_{v_{d-1},\ell_{d-1}} - \bigl(\tfrac{k}{n}\bigr)^\chi\bigr| \leqslant 3d\kappa_n. \tag{5.2}$$

(ii) *For every such path in D_n between vertex n and a vertex $k \leqslant n_1$, we have*

$$\widetilde{U}_{v_0,\ell_0} \cdots \widetilde{U}_{v_{d-1},\ell_{d-1}} \leqslant \bigl(\tfrac{n_1}{n}\bigr)^\chi + 3d\kappa_n. \tag{5.3}$$

Proof. We may assume that n is large enough such that $n_1^{-1} < \kappa_n$, since the result is trivial for small n by choosing C large enough.

Firstly it follows from (2.5) and (2.6) that if $k \xrightarrow{\ell} i$, then

$$S_{n,i-1} \leqslant \widetilde{U}_{k,\ell} S_{n,k-1} < S_{n,i}. \tag{5.4}$$

Secondly, again assuming that n is large, we take $\varepsilon = 1/3$ and $\psi_n = n_1$ in Lemma 3.4 and find that there is a positive constant C such that with probability at least $1-C\log n/n^{1/3}$,

$$\max_{n_1 \leqslant j \leqslant n} \bigl|S_{n,j} - \bigl(\tfrac{j}{n}\bigr)^\chi\bigr| \leqslant 2n_1^{-1/3} \leqslant 3\log^{1/3} n/n^{1/3} \leqslant \kappa_n. \tag{5.5}$$

We assume in the rest of the proof that (5.5) holds, and show first that (5.2) follows by induction on d. Note first that if $j > n_1$, then (5.5) implies

$$\begin{aligned} S_{n,j-1} &\geqslant \bigl(\tfrac{j-1}{n}\bigr)^\chi - \kappa_n = \bigl(\tfrac{j}{n}\bigr)^\chi\bigl(1-\tfrac{1}{j}\bigr)^\chi - \kappa_n \geqslant \bigl(\tfrac{j}{n}\bigr)^\chi\bigl(1-\tfrac{1}{j}\bigr) - \kappa_n \geqslant \bigl(\tfrac{j}{n}\bigr)^\chi - \tfrac{1}{j} - \kappa_n \\ &\geqslant \bigl(\tfrac{j}{n}\bigr)^\chi - 2\kappa_n. \end{aligned} \tag{5.6}$$

For the base case $d=1$ we have by (5.4), (5.5), (5.6), and recalling $S_{n,n-1} = 1$,

$$\widetilde{U}_{n,\ell_0} \leqslant S_{n,k} \leqslant \bigl(\tfrac{k}{n}\bigr)^\chi + \kappa_n \tag{5.7}$$

and

$$\widetilde{U}_{n,\ell_0} \geqslant S_{n,k-1} \geqslant \bigl(\tfrac{k}{n}\bigr)^\chi - 2\kappa_n, \tag{5.8}$$

which show (5.2) in this case.

For $d \geqslant 2$, we use induction and find, using the induction hypothesis and (5.4)–(5.6),

$$\begin{aligned} \widetilde{U}_{v_0,\ell_0}\cdots\widetilde{U}_{v_{d-1},\ell_{d-1}} &\leqslant \Bigl(\bigl(\tfrac{v_{d-1}}{n}\bigr)^{\chi}+3(d-1)\kappa_n\Bigr)\widetilde{U}_{v_{d-1},\ell_{d-1}} \\ &\leqslant \bigl(S_{n,v_{d-1}-1}+2\kappa_n+(3d-3)\kappa_n\bigr)\widetilde{U}_{v_{d-1},\ell_{d-1}} \\ &\leqslant S_{n,v_{d-1}-1}\widetilde{U}_{v_{d-1},\ell_{d-1}}+(3d-1)\kappa_n \\ &< S_{n,k}+(3d-1)\kappa_n \\ &\leqslant \bigl(\tfrac{k}{n}\bigr)^{\chi}+3d\kappa_n \end{aligned} \tag{5.9}$$

and similarly, using also $S_{n,v_{d-1}} \geqslant S_{n,v_{d-1}-1}$,

$$\begin{aligned} \widetilde{U}_{v_0,\ell_0}\cdots\widetilde{U}_{v_{d-1},\ell_{d-1}} &\geqslant \Bigl(\bigl(\tfrac{v_{d-1}}{n}\bigr)^{\chi}-3(d-1)\kappa_n\Bigr)\widetilde{U}_{v_{d-1},\ell_{d-1}} \\ &\geqslant \bigl(S_{n,v_{d-1}}-\kappa_n-(3d-3)\kappa_n\bigr)\widetilde{U}_{v_{d-1},\ell_{d-1}} \\ &\geqslant S_{n,v_{d-1}-1}\widetilde{U}_{v_{d-1},\ell_{d-1}}-(3d-2)\kappa_n \\ &\geqslant S_{n,k-1}-(3d-2)\kappa_n \\ &\geqslant \bigl(\tfrac{k}{n}\bigr)^{\chi}-3d\kappa_n. \end{aligned} \tag{5.10}$$

These inequalities prove (5.2), which completes the proof of (i).

To prove (ii), assume first $v_{d-1} > n_1 \geqslant k$. Then, using (5.2), the first lines of (5.9) still hold and yield

$$\widetilde{U}_{v_0,\ell_0}\cdots\widetilde{U}_{v_{d-1},\ell_{d-1}} < S_{n,k}+(3d-1)\kappa_n. \tag{5.11}$$

Furthermore, by $n_1 \geqslant k$ and (5.5),

$$S_{n,k} \leqslant S_{n,n_1} \leqslant \bigl(\tfrac{n_1}{n}\bigr)^{\chi}+\kappa_n, \tag{5.12}$$

and (5.3) follows by (5.11) and (5.12). Finally, in the remaining case $v_{d-1} \leqslant n_1$, we use the trivial $\widetilde{U}_{v_0,\ell_0}\cdots\widetilde{U}_{v_{d-1},\ell_{d-1}} \leqslant \widetilde{U}_{v_0,\ell_0}\cdots\widetilde{U}_{v_{d-2},\ell_{d-2}}$ and induction on d. □

Recall that a vertex in $\widehat{\mathcal{Y}}$ that is d generations away from the root has label of the form $\widehat{U}_1\cdots\widehat{U}_d$, where $\widehat{U}_i \in \mathsf{U}[0,1]$ and are independent. In view of (5.2), we couple $\widehat{\mathcal{Y}}$ and $\widehat{D}_n$ by generating them together as follows, where we also construct a mapping Ψ of the vertices of $\widehat{\mathcal{Y}}$ to the vertices of $\widehat{D}_n$. In the construction below, $\widehat{\mathcal{Y}}$ and $\widehat{D}_n$ will be finite subsets of the final Yule tree and digraph, and Ψ maps the current $\widehat{\mathcal{Y}}$ onto the current $\widehat{D}_n$. Recall that $\widehat{\mathcal{Y}}$ and $\widehat{D}_n$ determine $\mathcal{Y}$ and D_n by (deterministic) relabelling.

(1) Sample the beta variables $(B_j)_{j=2}^{n-1}$ defined in (2.1). This defines also $S_{n,j}$ by (2.3).

(2) Start the construction by letting $\widehat{\mathcal{Y}}$ and $\widehat{D}_n$ just consist of their roots, both labelled 1. Let Ψ map the root of $\widehat{\mathcal{Y}}$ to the root of $\widehat{D}_n$.

(3) Let x be the vertex in the constructed part of $\widehat{\mathcal{Y}}$ that has the largest label among all vertices that do not yet have children assigned. Give x children $xU'_{x,1}$ and $xU'_{x,2}$ (which are added to the current $\widehat{\mathcal{Y}}$), where $U'_{x,\ell}$ are independent $\mathsf{U}[0,1]$ variables that are independent of all other variables.

The vertex x is mapped to some vertex $\Psi(x) = (k/n)^{\chi}$ in $\widehat{D}_n$, which thus corresponds to vertex k in D_n. There are three cases:

(a) If $k > 1$ and $\Psi(x)$ has not yet got any children, define $\widetilde{U}_{k,\ell} := U'_{x,\ell}$ for $\ell = 1, 2$. This defines by (2.5)–(2.6) the edges from k and thus the children of k in D_n; if these children are k_1 and k_2, the corresponding children in $\widehat{D}_n$ are $(k_1/n)^{\chi}$ and $(k_2/n)^{\chi}$; we add them to $\widehat{D}_n$ and we define $\Psi(x_\ell) := (k_\ell/n)^{\chi}$, thus mapping the children of x in $\widehat{\mathcal{Y}}$ to the children of $\Psi(x)$ in $\widehat{D}_n$.

(b) If $k > 1$ and $\Psi(x)$ already has children (because it equals $\Psi(y)$ for some $y > x$), then we just extend Ψ by mapping the children of x to the children of $\Psi(x)$ (in any order).

(c) If $k = 1$, so x maps to $v = (1/n)^\chi$ (which has no children in the final $\widehat{D}_n$), we extend Ψ by mapping also the children of x to v.

(4) Repeat from (3) (*ad infinitum*).

It is easy to see that running this "algorithm" an infinite number of iterations yields $\widehat{\mathcal{Y}}$ and $\widehat{D}_n$ with the right distributions, together with a map Ψ of the vertices of $\widehat{\mathcal{Y}}$ onto the vertices of $\widehat{D}_n$ such that every path from the root in $\widehat{D}_n$ is the image of a path from the root in $\widehat{\mathcal{Y}}$. Ψ is obviously not injective since $\widehat{\mathcal{Y}}$ is an infinite tree. Nevertheless, we show that restricted to rather large labels, the mapping Ψ is w.h.p. a bijection which perturbs the vertex labels with small errors.

Theorem 5.2. *Let $n_1 := \lfloor n/\log n\rfloor$. We may w.h.p. couple the $\widehat{D}_n$ and the time-changed Yule tree $\widehat{\mathcal{Y}}$, such that considering only vertices with labels in $((n_1/n)^\chi, 1]$ and edges with the starting points in this set, there is a bijection between these sets of vertices in the two models which displaces each label by at most $3\log^2 n/n^{1/3}$, and a corresponding bijection between the edges (preserving the incidence relations). In particular, w.h.p.*

$$Y^{(n)}_{n_1} = \widehat{\mathcal{Y}}_{(n_1/n)^\chi}, \tag{5.13}$$

where $\widehat{\mathcal{Y}}_x = \mathcal{Y}_{-\log x}$ is the number of edges in $\widehat{\mathcal{Y}}$ alive at time x.

Proof. The proof is similar to that of [13, Theorem 3.1], but with several technical complications. We use the coupling constructed before the theorem. Let $\delta_n := 3\log^2 n/n^{1/3} = (\log n)\kappa_n$, with κ_n as in Lemma 5.1.

Step 1. We first note that if some vertex in $\widehat{D}_n$ with label in $[(n_1/n)^\chi, 1]$ is the image of two or more vertices in $\widehat{\mathcal{Y}}$, then the corresponding vertex $k \geqslant n_1$ in D_n can be reached from n by at least two different paths in D_n, and if we let k be maximal with this property, then its indegree $Z_k \geqslant 2$. Consequently, the probability that this happens is at most, using (4.28) of Lemma 4.2,

$$\sum_{k=n_1}^{n-1} \mathbb{P}(Z_k \geqslant 2) \leqslant Cn \sum_{k=n_1}^{n-1} k^{-3} = O(n/n_1^2) = O(\log^2 n/n) = o(1). \tag{5.14}$$

Hence, w.h.p. the mapping Ψ from $\widehat{\mathcal{Y}}$ to $\widehat{D}_n$ is injective at every vertex in $\widehat{D}_n \cap [(n_1/n)^\chi, 1]$. We may thus in the sequel assume that this injectivity holds. Note that this implies that in the construction of the mapping Ψ above, we have $\widetilde{U}_{k,\ell} = U'_{x,\ell}$ for every $k \geqslant n_1$ and vertex $x \in \widehat{\mathcal{Y}}$ such that $\Psi(x) = (k/n)^\chi$.

Step 2. As in [13], $\widehat{\mathcal{Y}}_x = \mathcal{Y}_{-\log x}$ for every $x \in (0, 1]$, so by standard properties of the Yule process (see e.g. [13, Section 3])

$$\mathbb{E}\,\widehat{\mathcal{Y}}_x = \mathbb{E}\,\mathcal{Y}_{-\log x} = 2e^{-\log x} = 2/x. \tag{5.15}$$

In $\widehat{\mathcal{Y}}$, there are $\widehat{\mathcal{Y}}_x - 1$ vertices with labels in $[x, 1]$. Thus, (5.15) implies that w.h.p. there are less than $\lfloor\log n\rfloor$ such vertices for $x = (n_1/n)^\chi \sim \log^{-\chi} n$. It follows trivially that w.h.p., in $\widehat{\mathcal{Y}}$ the number of generations from the root to any point in $[(n_1/n)^\chi, 1]$ is less than $\lfloor\log n\rfloor$. Hence, we may assume this property too.

Step 3. The expected number of vertices in $\widehat{\mathcal{Y}}$ that are within δ_n from $(n_1/n)^\chi$ is, by (5.15),

$$\mathbb{E}\big(\widehat{\mathcal{Y}}_{(n_1/n)^\chi-\delta_n} - \widehat{\mathcal{Y}}_{(n_1/n)^\chi+\delta_n}\big) = \frac{2}{(n_1/n)^\chi - \delta_n} - \frac{2}{(n_1/n)^\chi + \delta_n}$$

$$= O\Big(\frac{\delta_n}{(n_1/n)^{2\chi}}\Big) = O\big(\delta_n \log^{2\chi} n\big) = o(1). \tag{5.16}$$

Hence, w.h.p. there are no vertices x in $\widehat{\mathcal{Y}}$ with $|x-(n_1/n)^\chi| \leqslant \delta_n$. We may in the sequel assume this.

Step 4. Consequently, w.h.p. the properties in Steps 1–3 hold, and also the conclusions (i) and (ii) of Lemma 5.1. We assume this for the rest of the proof.

Suppose that $k > n_1$ is a vertex of D_n, and let $v = (k/n)^\chi$ be the corresponding vertex of $\widehat{D}_n$. By Step 1, $v = \Psi(x)$ for a unique vertex $x \in \widehat{\mathcal{Y}}$. If $x \in ((n_1/n)^\chi, 1]$, then the number of generations from the root to x in $\widehat{\mathcal{Y}}$ is at most $\log n$ by Step 2. The number of generations to $\Psi(x)$ in $\widehat{D}_n$ is the same, and it follows from (5.2) and the equality $\widetilde{U}_{k,\ell} = U'_{x,\ell}$ in Step 1 that, denoting the path from n to k as in Lemma 5.1,

$$|x-\Psi(x)| = \big|\widetilde{U}_{v_0,\ell_0}\cdots\widetilde{U}_{v_{d-1},\ell_{d-1}} - \big(\tfrac{k}{n}\big)^\chi\big| \leqslant 3(\log n)\kappa_n = 3\log^2 n/n^{1/3} = \delta_n. \tag{5.17}$$

It remains to show only that no vertex x in $\widehat{\mathcal{Y}}$ is pushed over the boundary $(n_1/n)^\chi$ (in any direction) by Ψ.

Step 5. Suppose that there exists a vertex $x \leqslant (n_1/n)^\chi$ in $\widehat{\mathcal{Y}}$ such that $\Psi(x) > (n_1/n)^\chi$. Let $y > x$ be the parent of x in $\widehat{\mathcal{Y}}$, so that $\Psi(y) > \Psi(x)$. Assume also $y > (n_1/n)^\chi$. By Step 2, it follows that the number of generations from the root to y is less than $\lfloor \log n\rfloor$, and thus the number of generations to x is at most $\lfloor \log n\rfloor$. Consequently, (5.2) shows, similarly to (5.17), that

$$|x-\Psi(x)| \leqslant 3(\log n)\kappa_n = \delta_n \tag{5.18}$$

and hence

$$x \geqslant \Psi(x) - \delta_n \geqslant (n_1/n)^\chi - \delta_n. \tag{5.19}$$

However, we have also $x \leqslant (n_1/n)^\chi$, so $|x-(n_1/n)^\chi| \leqslant \delta_n$, and by Step 3, there is no such vertex x in $\widehat{\mathcal{Y}}$. If $y \leqslant (n_1/n)^\chi$, we may instead replace x by y (and repeat this if necessary) until we encounter a vertex x with parent y such that $x \leqslant (n_1/n)^\chi$, $\Psi(x) > (n_1/n)^\chi$ and $y > (n_1/n)^\chi$. However, we have shown that such a pair cannot exist.

Step 6. Suppose that there exists a vertex $x > (n_1/n)^\chi$ in $\widehat{\mathcal{Y}}$ such that $\Psi(x) \leqslant (n_1/n)^\chi$. By Step 2, it follows that the number of generations from the root to x is less than $\log n$. This time (5.3) applies and shows that

$$x \leqslant \big(\tfrac{n_1}{n}\big)^\chi + 3(\log n)\kappa_n = \big(\tfrac{n_1}{n}\big)^\chi + \delta_n. \tag{5.20}$$

However, by Step 3 again, there is no such vertex x in $\widehat{\mathcal{Y}}$.

The various claims proved in the steps above show that with the coupling and mapping Ψ constructed before the theorem, w.h.p. Ψ yields a bijection with the stated properties. In particular, w.h.p. Y_{n_1} equals the number of edges in $\widehat{\mathcal{Y}}$ that are alive at $(n_1/n)^\chi$, meaning that they start in $J := [(n_1/n)^\chi, 1]$ and end outside J. (Note that a.s. $\widehat{\mathcal{Y}}$ has no point exactly at $(n_1/n)^\chi$, so it does not matter whether we include that point in J or not.) Finally, note that for any $x \in (0,1)$, the number of edges of $\widehat{\mathcal{Y}}$ that are alive at x equals the number of edges in $\mathcal{Y}$ that are alive at $-\log x$, which equals the number $\mathcal{Y}_{-\log x}$ of particles at $-\log x$, since each edge represents the lifeline of one particle. □

We define

$$\Xi = \Xi^{(n)} := \frac{W^{(n)}_{n_1}}{n^\chi}. \tag{5.21}$$

This random variable plays the same role as Ξ in [13], but note the different scaling. Recall also $\tilde\beta$ defined in (3.17)–(3.18).

Lemma 5.3. *As $n \to \infty$,*

$$\Xi^{(n)} \xrightarrow{\mathrm{d}} \tilde{\beta} \cdot \xi, \tag{5.22}$$

where $\tilde{\beta}$ is given by (3.18), and $\xi \in \mathrm{Gamma}(2,1)$ is independent of $\tilde{\beta}$.

Proof. As in [13, Lemma 3.2], we first generate the Yule process $\mathcal{Y}$, and then for each n separately, we construct D_n by the construction given before Theorem 5.2. This yields for each n the coupling in Theorem 5.2. In particular, (5.13) holds w.h.p. By standard properties of the Yule process (see e.g. [1, Section III.5 and Problem III.2]),

$$x\widehat{\mathcal{Y}}_x = x\mathcal{Y}_{-\log x} \xrightarrow{\mathrm{a.s.}} \xi \quad \text{as } x \to 0, \tag{5.23}$$

with $\xi \in \mathrm{Gamma}(2,1)$. Therefore, (5.13) and (5.23) together imply that

$$\Big(\frac{n_1}{n}\Big)^{\chi} Y^{(n)}_{n_1} \xrightarrow{\mathrm{p}} \xi. \tag{5.24}$$

Hence, using also (5.21), (4.6), and (3.18),

$$\Xi^{(n)} = n^{-\chi}\Phi_{n_1}Y_{n_1} = n_1^{-\chi}\Phi_{n_1}\Big(\frac{n_1}{n}\Big)^{\chi}Y_{n_1} \xrightarrow{\mathrm{p}} \tilde{\beta}\xi. \tag{5.25}$$

Finally, $Y^{(n)}_{n_1}$ is a function of $(B_i)_{i>n_1}$ and the coin tosses made for $k > n_1$ in the construction. Hence, for any fixed K, $Y^{(n)}_{n_1}$ is independent of $(B_i)_{i=1}^K$ for large enough n. Consequently, (5.24) implies that ξ is independent of $(B_i)_{i=1}^K$ for every $K < \infty$, and thus ξ and β are independent. ☐

6. The flat middle part

Let n_2 be any sequence of integers satisfying $n^{1/3} \ll n_2 \leqslant n_1$. We show that similar to the case of uniform attachment, the variable W_k does not fluctuate much in the range $n_1 \geqslant k \geqslant n_2$. Below we give analogues of [13, Lemmas 4.1–4.2 and Theorem 4.3], where we recall the definitions of W_k, M_k, A_k and $\Xi^{(n)}$ in (4.6), (4.9) and (5.21). The results and proofs are again similar, but with a different scaling.

Lemma 6.1. *As $n \to \infty$,*

$$\max_{n_2 \leqslant k \leqslant n-1}\Big|\frac{A_k}{n^{1/2}}\Big| = \frac{A_{n_2}}{n^{1/2}} \xrightarrow{\mathrm{p}} 0. \tag{6.1}$$

Proof. The first equality in (6.1) follows from the fact that A_k is reverse increasing. By (4.36) in Lemma 4.3 and the choice of n_2,

$$\mathbb{E}\,\frac{A_{n_2}}{n^{1/2}} \leqslant C\frac{n}{n^{1/2}n_2^{3/2}} = o(1), \tag{6.2}$$

implying the convergence in probability in (6.1). ☐

Lemma 6.2. *As $n \to \infty$,*

$$\max_{0 \leqslant k \leqslant n_1}\Big|\frac{M_k}{n^{1/2}} - \Xi^{(n)}\Big| \xrightarrow{\mathrm{p}} 0. \tag{6.3}$$

Proof. Recall that M_k is a reverse martingale. Hence we obtain by Doob's inequality, (4.10), (4.16), (4.6), (3.5), and (4.17),

$$\begin{aligned}\mathbb{E}\max_{0 \leqslant k \leqslant n_1}|M_k - M_{n_1}|^2 &\leqslant 4\,\mathbb{E}\,|M_0 - M_{n_1}|^2 = 4\sum_{i=1}^{n_1}\mathbb{E}\,\mathrm{Var}(W_{i-1} \mid \mathcal{F}_i) \\ &\leqslant 40\sum_{i=1}^{n_1}\mathbb{E}\big(\Phi_{i-1}^2 B_i Y_i\big) = 40\sum_{i=1}^{n_1}\mathbb{E}\big(\Phi_{i-1}W_iB_i(1+B_i)^{-1}\big)\end{aligned}$$

$$\begin{aligned}
&= 40 \sum_{i=1}^{n_1} \mathbb{E}\bigl(\Phi_{i-1} B_i (1+B_i)^{-1} \, \mathbb{E}_{\mathbf{B}}(W_i)\bigr) \\
&\leqslant 80 \sum_{i=1}^{n_1} \mathbb{E}\bigl(\Phi_{i-1} B_i (1+B_i)^{-1} \Phi_{n-1}\bigr) \\
&= 80 \sum_{i=1}^{n_1} \mathbb{E}\Bigl(\Phi_{i-1}^2 B_i \prod_{j=i+1}^{n-1} (1+B_j)\Bigr). \qquad (6.4)
\end{aligned}$$

Using the independence of $(B_i)_{i=2}^{n-1}$, (3.6), (3.8) and (3.11), it follows from (6.4) that

$$\begin{aligned}
\mathbb{E}\bigl(\max_{0\leqslant k\leqslant n_1} |M_k - M_{n_1}|\bigr)^2 = \mathbb{E} \max_{0\leqslant k\leqslant n_1} |M_k - M_{n_1}|^2 &\leqslant C \sum_{i=1}^{n_1} \frac{i}{i} \Bigl(\frac{n}{i}\Bigr)^{1/2} \\
&\leqslant C(nn_1)^{1/2} = o(n). \qquad (6.5)
\end{aligned}$$

Thus, by (5.21), (4.9), the triangle inequality, Lemma 6.1, and (6.5),

$$\max_{0\leqslant k\leqslant n_1} \left| \frac{M_k}{n^{1/2}} - \Xi^{(n)} \right| \leqslant \max_{0\leqslant k\leqslant n_1} \left| \frac{M_k}{n^{1/2}} - \frac{M_{n_1}}{n^{1/2}} \right| + \frac{A_{n_1}}{n^{1/2}} \overset{\mathrm{p}}{\longrightarrow} 0, \qquad (6.6)$$

as required. □

Theorem 6.3. *As $n \to \infty$,*

$$\max_{n_2\leqslant k\leqslant n_1} \left| \frac{W_k}{n^{1/2}} - \Xi^{(n)} \right| \overset{\mathrm{p}}{\longrightarrow} 0. \qquad (6.7)$$

Proof. The result is a direct consequence of (4.9), the triangle inequality, and Lemmas 6.1 and 6.2. □

7. The final part: tightness

Most vertices in D_n turn out to have labels of the order $n^{1/3}$. To study this region in detail, we extend the processes W_k, M_k and A_k to real arguments $t \in [0, n-1]$ by linear interpolation. We for convenience extend them further to $t \in [0,\infty)$ by defining them to be constant on $[n-1,\infty)$.

The aim of this and the next section is to prove convergence of $W_{tn^{1/3}}$ and $Y_{tn^{1/3}}$ as $n \to \infty$, after suitable rescaling. A key ingredient is the tightness of the random function

$$\widehat{A}_t^{(n)} := n^{-1/2} A_{tn^{1/3}}^{(n)}, \qquad t \geqslant 0. \qquad (7.1)$$

Recall that $C[a, b]$ is the space of continuous functions on $[a, b]$.

Lemma 7.1. *Let $0 < a < b < \infty$. Then the stochastic processes $\widehat{A}_t^{(n)}$, $n \geqslant 1$, are tight in $C[a, b]$.*

The proof of Lemma 7.1 is more complicated than for the corresponding Lemma 5.2 in [13], and we show first two other lemmas. We begin by stating a simple general lemma on tightness in the space $C[a, b]$. (See e.g. [4] for a background.)

Lemma 7.2. *Let $-\infty < a < b < \infty$. Let $(X_n(t))_{n=1}^\infty$ and $(Y_n(t))_{n=1}^\infty$ be two sequences of random continuous functions on $[a, b]$. Suppose that there exists a sequence $(Z_n)_{n=1}^\infty$ of random variables such that for every n and $s, t \in [a, b]$, we have*

$$|X_n(t) - X_n(s)| \leqslant Z_n |Y_n(t) - Y_n(s)|. \qquad (7.2)$$

If the sequences $(X_n(a))_{n=1}^\infty$ and $(Z_n)_{n=1}^\infty$ are tight, and $(Y_n(t))_{n=1}^\infty$ is tight in $C[a, b]$, then the sequence $(X_n(t))_{n=1}^\infty$ is tight in $C[a, b]$.

Proof. We may for convenience assume $[a,b]=[0,1]$. We define for any function f on $[0,1]$ its modulus of continuity

$$\omega(f;\delta) := \sup_{s,t\in[0,1];\ |s-t|<\delta} |f(s)-f(t)|, \qquad \delta>0. \tag{7.3}$$

Then [4, Theorem 8.2] says that a sequence $(X_n(t))_{n=1}^\infty$ in $C[0,1]$ is tight if and only if

(i) the sequence $(X_n(0))_n$ is tight, and

(ii) for each positive ε and η, there exists $\delta>0$ such that, for every n,

$$\mathbb{P}\bigl(\omega(X_n;\delta)\geqslant\varepsilon\bigr)\leqslant\eta. \tag{7.4}$$

We have already assumed (i). Moreover, the assumption (7.2) implies

$$\omega(X_n;\delta)\leqslant Z_n\omega(Y_n;\delta) \tag{7.5}$$

for every δ. Let $\varepsilon,\eta>0$ be given. Since the sequence $(Z_n)_{n=1}^\infty$ is tight, there exists a number $K>0$ such that $\mathbb{P}(|Z_n|>K)<\eta/2$ for every n. Hence, (7.5) implies

$$\begin{aligned}\mathbb{P}\bigl(\omega(X_n;\delta)\geqslant\varepsilon\bigr)&\leqslant\mathbb{P}(|Z_n|>K)+\mathbb{P}\bigl(K\omega(Y_n;\delta)\geqslant\varepsilon\bigr)\\&\leqslant\eta/2+\mathbb{P}\bigl(\omega(Y_n;\delta)\geqslant\varepsilon/K\bigr).\end{aligned} \tag{7.6}$$

Since $(Y_n(t))_{n=1}^\infty$ is tight, conditions (i)–(ii) hold for $Y_n(t)$; in particular, there exists $\delta>0$ such that $\mathbb{P}\bigl(\omega(Y_n;\delta)\geqslant\varepsilon/K\bigr)\leqslant\eta/2$ for every n. Then (7.6) shows that (ii) holds. Consequently, $(X_n(t))_{n=1}^\infty$ is tight. ☐

Recall that Lemma 4.3 shows that

$$0\leqslant A_{k-1}-A_k\leqslant W_k^2\Phi_k^{-1}B_k^2\leqslant M_k^2\Phi_k^{-1}B_k^2, \qquad 1\leqslant k\leqslant n-1. \tag{7.7}$$

We define the simpler

$$V_k:=\sum_{j=1}^k B_j^2, \quad T_k:=\sum_{j=1}^k B_j, \qquad 0\leqslant k\leqslant n-1; \tag{7.8}$$

we extend also V_k and T_k by linear interpolation to real arguments, and define

$$\widehat{V}_t^{(n)}:=n^{1/3}V^{(n)}_{tn^{1/3}}, \quad \widehat{T}_t^{(n)}:=T^{(n)}_{tn^{1/3}}, \qquad t\geqslant 0. \tag{7.9}$$

The proof of Lemma 7.1 only uses $\widehat{V}_t^{(n)}$, but $\widehat{T}_t^{(n)}$ is needed later when we prove (1.7).

Lemma 7.3. *Let $0<a<b<\infty$. Then the stochastic processes $\widehat{V}_t^{(n)}-\widehat{V}_a^{(n)}$ and $\widehat{T}_t^{(n)}-\widehat{T}_a^{(n)}$, $n\geqslant 1$, are tight in $C[a,b]$.*

Proof. We start with $\widehat{V}_t^{(n)}-\widehat{V}_a^{(n)}$, $n\geqslant1$. If $1\leqslant k\leqslant\ell\leqslant n-1$, then

$$\mathbb{E}\,|V_\ell-V_k|^2=\mathbb{E}\Bigl(\sum_{i,j=k+1}^{\ell}B_i^2B_j^2\Bigr)=\sum_{i,j=k+1}^{\ell}\mathbb{E}\bigl[B_i^2B_j^2\bigr]. \tag{7.10}$$

If $i\neq j$, then B_i and B_j are independent, and thus, by (3.7), $\mathbb{E}\bigl[B_i^2B_j^2\bigr]=\mathbb{E}[B_i^2]\,\mathbb{E}[B_j^2]=O\bigl(i^{-2}j^{-2}\bigr)$. On the other hand, if $i=j$, then, by (3.3), recalling that $B_i\in\mathrm{Beta}(2,4i-6)$,

$$\mathbb{E}\,B_i^4=\frac{2\cdot3\cdot4\cdot5}{(4i-4)(4i-3)(4i-2)(4i-1)}=O\bigl(i^{-4}\bigr). \tag{7.11}$$

Consequently, (7.10) yields

$$\mathbb{E}\,|V_\ell-V_k|^2\leqslant\sum_{i,j=k+1}^{\ell}Ci^{-2}j^{-2}\leqslant C(\ell-k)^2k^{-4}. \tag{7.12}$$

This trivially holds for $\ell>n-1$ too, since then $V_\ell=V_{n-1}$ by definition. Furthermore, writing (7.12) as $\|V_\ell-V_k\|_{L^2}\leqslant C(\ell-k)k^{-2}$, it follows by Minkowski's inequality that we can interpolate between integer arguments and conclude that (7.12) holds for all real k and ℓ with $1 \leqslant k \leqslant \ell$.

Let s and t be real numbers with $0 < s \leqslant t$. Then (7.9) and (7.12) yield, with $k := sn^{1/3}$ and $\ell := tn^{1/3}$,

$$\mathbb{E}\bigl|\widehat{V}_t^{(n)} - \widehat{V}_s^{(n)}\bigr|^2 = n^{2/3}\,\mathbb{E}\,|V_\ell - V_k|^2 \leqslant Cn^{2/3}|\ell - k|^2 k^{-4} = C|t-s|^2 s^{-4}. \tag{7.13}$$

For the restriction to $[a,b]$ we thus have

$$\mathbb{E}\bigl|\widehat{V}_t^{(n)} - \widehat{V}_s^{(n)}\bigr|^2 \leqslant C_a|t-s|^2, \qquad a \leqslant s \leqslant t \leqslant b, \tag{7.14}$$

which shows the tightness of $\widehat{V}_t^{(n)} - \widehat{V}_a^{(n)}$ by [4, Theorem 12.3].

Tightness of $\widehat{T}_t^{(n)} - \widehat{T}_a^{(n)}$ can be shown by using

$$\mathbb{E}\,|T_k - T_\ell|^2 = \sum_{i,j=k+1}^{\ell} \mathbb{E}(B_iB_j) \leqslant C(\ell-k)^2k^{-2}, \quad 1 \leqslant k \leqslant \ell \leqslant n-1, \tag{7.15}$$

instead of (7.12) and proceeding similarly. ☐

Proof of Lemma 7.1. Note first that (4.36) implies that

$$\mathbb{E}\,\widehat{A}_a^{(n)} \leqslant n^{-1/2}\frac{Cn}{(an^{1/3})^{3/2}} = Ca^{-3/2}, \tag{7.16}$$

and thus the sequence $\widehat{A}_a^{(n)}$ is tight.

Let

$$\mathfrak{M}_n := n^{-1/2}\max_{k\geqslant 1} M_k \qquad \text{and} \qquad \widehat{\Psi}_n := n^{1/6}\Phi^{-1}_{\lfloor an^{1/3}\rfloor}. \tag{7.17}$$

Then Lemma 4.1 shows that $\mathbb{E}\,\mathfrak{M}_n^2 \leqslant C$, and (3.11) yields $\mathbb{E}\,\widehat{\Psi}_n = O(1)$; hence the sequences $(\mathfrak{M}_n)_{n=1}^\infty$ and $(\widehat{\Psi}_n)_{n=1}^\infty$ are tight.

By (7.7) and (7.8), we have for any integer k with $k \geqslant an^{1/3}$, since Φ_k is increasing by the definition (3.5), using (7.17),

$$\begin{aligned}|A_k - A_{k-1}| &\leqslant M_k^2\Phi_k^{-1}(V_k - V_{k-1}) \leqslant n\mathfrak{M}_n^2\Phi^{-1}_{\lfloor an^{1/3}\rfloor}(V_k - V_{k-1})\\ &= n^{5/6}\mathfrak{M}_n^2\widehat{\Psi}_n(V_k - V_{k-1}).\end{aligned} \tag{7.18}$$

Thus, if $\lfloor an^{1/3}\rfloor \leqslant k \leqslant \ell$,

$$|A_\ell - A_k| \leqslant n^{5/6}\mathfrak{M}_n^2\widehat{\Psi}_n(V_\ell - V_k), \tag{7.19}$$

We can interpolate between integer arguments and conclude that (7.19) holds for all real k and ℓ with $\lfloor an^{1/3}\rfloor \leqslant k \leqslant \ell$. By (7.1) and (7.9), this shows that if $a \leqslant s \leqslant t$, then

$$\bigl|\widehat{A}_t^{(n)} - \widehat{A}_s^{(n)}\bigr| \leqslant \mathfrak{M}_n^2\widehat{\Psi}_n\bigl(\widehat{V}_t^{(n)} - \widehat{V}_s^{(n)}\bigr). \tag{7.20}$$

The result now follows from Lemmas 7.2 and 7.3, taking $X_n(t) := \widehat{A}_t^{(n)}$, $Y_n(t) := \widehat{V}_t^{(n)} - \widehat{V}_a^{(n)}$, and $Z_n := \mathfrak{M}_n^2\widehat{\Psi}_n$, noting that $(Z_n)_{n=1}^\infty$ is tight since $(\mathfrak{M}_n)_{n=1}^\infty$ and $(\widehat{\Psi}_n)_{n=1}^\infty$ are. ☐

8. The final part: convergence

Recall the definition of the spaces $C(0,\infty)$ and $C[0,\infty)$ in Section 1.2.

Theorem 8.1. *As $n \to \infty$ we have*

$$\frac{W_{tn^{1/3}}}{n^{1/2}} \overset{\mathrm{d}}{\longrightarrow} 4\tilde{\beta}\bigl((t^{9/2} + \tfrac{3}{4}\xi t^3)^{1/3} - t^{3/2}\bigr) \qquad \text{in } C[0,\infty), \tag{8.1}$$

and

$$\frac{Y_{tn^{1/3}}}{n^{1/3}} \xrightarrow{\mathrm{d}} 4\bigl(\bigl(t^3+\tfrac{3}{4}\xi t^{3/2}\bigr)^{1/3}-t\bigr) \qquad \text{in } C(0,\infty), \tag{8.2}$$

with $\tilde\beta$ *as in* (3.18) *and* $\xi \in \mathrm{Gamma}(2,1)$ *independent of* $\tilde\beta$*.*

Remark 8.2. We believe that (8.2) holds also in $C[0,\infty)$, but we see no simple proof so we leave this as an open problem.

Proof. The proof is similar to that of [13, Theorem 5.3], but with several technical complications.

Step 1. Convergence in $C(0,\infty)$ *for a subsequence.* As in [13], Lemma 7.1 implies that by considering a subsequence, we may assume that the processes $\widehat{A}^{(n)}_t$ converge in distribution in every space $C[a,b]$ with $0<a<b$ to some continuous stochastic process $\mathcal{A}(t)$ on $(0,\infty)$; in other words, as $n\to\infty$,

$$\widehat{A}^{(n)}_t \to \mathcal{A}(t) \tag{8.3}$$

holds in distribution in the space $C(0,\infty)$. Furthermore, also as in [13], we may by the Skorohod coupling theorem [14, Theorem 4.30], assume that this convergence holds a.s.; in other words (as $n\to\infty$ along the subsequence) a.s. (8.3) holds uniformly on every interval $[a,b]$. We use such a coupling until further notice. Note that this means that we consider all random variables as defined separately for each n (with some unknown coupling); in particular, this means that we have potentially a different sequence $B^{(n)}_i$ ($i\geqslant 1$) for each n, and thus different limits $\beta^{(n)}$ and $\tilde\beta^{(n)}$. (The variables $B^{(n)}_i$ for a fixed n are independent, but we do not know how they are coupled for different n.) Hence, we cannot directly use the a.s. convergence results in Lemma 3.3. Instead we note that (3.18) (which holds for each n, with the distributions of the variables the same for all n) implies, for any coupling,

$$\sup_{k\geqslant \log n} \bigl|k^{-1/2}\Phi^{(n)}_k - \tilde\beta^{(n)}\bigr| \xrightarrow{\mathrm{p}} 0. \tag{8.4}$$

Furthermore, trivially (since the distributions are the same)

$$\tilde\beta^{(n)} \xrightarrow{\mathrm{d}} \tilde\beta \tag{8.5}$$

for some random variable $\tilde\beta$ with, by Lemma 3.3, $\tilde\beta>0$ a.s.; note also that (8.5) holds jointly with (5.22), since this is true for the coupling used in the proof of Lemma 5.3, where B_i does not depend on n and thus trivially $\tilde\beta^{(n)}\to\tilde\beta$ holds together with (5.25).

We may select the subsequence above such that (8.3) (in distribution), (8.5) and (5.22) hold jointly (with some joint distribution of the limits). We may then assume that (8.3), (8.4), (8.5), (5.22), (6.3), and (6.7) all hold a.s., by redoing the application of the Skorohod coupling theorem and including all these limits. It then follows from (4.9), (7.1), (6.3), (8.3), and (5.22) that a.s.

$$n^{-1/2}W_{tn^{1/3}} = n^{-1/2}M_{tn^{1/3}} - \widehat{A}^{(n)}_t = \Xi^{(n)} - \mathcal{A}(t) + o(1) \to \mathcal{B}(t) := \tilde\beta\xi - \mathcal{A}(t) \tag{8.6}$$

uniformly on every compact interval in $(0,\infty)$. In other words, (8.6) holds a.s. in $C(0,\infty)$. From (8.6) we obtain by (4.6), (8.4), and (8.5), letting $k:=\lfloor tn^{1/3}\rfloor$, a.s.

$$Y_{\lfloor tn^{1/3}\rfloor} = \frac{W_k}{\Phi_k} = \frac{W_k}{k^{1/2}}\bigl(\tilde\beta^{(n)}+o(1)\bigr)^{-1} = \frac{n^{1/2}}{k^{1/2}}\bigl(\mathcal{B}(t)+o(1)\bigr)\bigl(\tilde\beta+o(1)\bigr)^{-1} \tag{8.7}$$

and thus

$$n^{-1/3}Y_{\lfloor tn^{1/3}\rfloor} \to t^{-1/2}\tilde\beta^{-1}\mathcal{B}(t), \tag{8.8}$$

again uniformly on every compact interval in $(0,\infty)$, and thus in $C(0,\infty)$.

Step 2. Identifying the limit. Fix $0 < s < t$ and let $k := \lfloor sn^{1/3} \rfloor$ and $\ell := \lfloor tn^{1/3} \rfloor$. It follows from (8.3) that $\widehat{A}_s^{(n)} - \widehat{A}_{n^{-1/3}k}^{(n)} \to 0$ and $\widehat{A}_t^{(n)} - \widehat{A}_{n^{-1/3}\ell}^{(n)} \to 0$ a.s. Hence, (7.1) and (4.14) imply that

$$\begin{aligned}\widehat{A}_s^{(n)} - \widehat{A}_t^{(n)} &= \sum_{i=k+1}^{\ell} n^{-1/2}(A_i - A_{i-1}) + o(1) \\ &= n^{-1/2} \sum_{i=k+1}^{\ell} 2\Phi_{i-1}\bigl[(1-B_i)^{Y_i} - 1 + Y_i B_i\bigr] + o(1). \end{aligned} \tag{8.9}$$

For any $B \in (0,1)$, with $x := -\log(1-B)$, and any $y \geqslant 1$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}y}\bigl((1-B)^y - 1 + yB\bigr) &= \frac{\mathrm{d}}{\mathrm{d}y}\bigl(e^{-xy} - 1 + y(1-e^{-x})\bigr) = -xe^{-xy} + 1 - e^{-x} \\ &\geqslant 1 - (1+x)e^{-x} > 0. \end{aligned} \tag{8.10}$$

Hence $(1-B)^y - 1 + yB$ is an increasing function of $y \geqslant 1$. Let $\varepsilon > 0$, and let

$$y_+ := \max\{u^{-1/2}\tilde{\beta}^{-1}\mathcal{B}(u) : u \in [s,t]\} + \varepsilon. \tag{8.11}$$

Then (8.8) implies that, for large enough n, $Y_i \leqslant y_+ n^{1/3}$ when $k \leqslant i \leqslant \ell$, and hence (8.9) implies, noting that Φ_i is increasing in i,

$$\widehat{A}_s^{(n)} - \widehat{A}_t^{(n)} \leqslant 2n^{-1/2}\Phi_\ell \sum_{i=k+1}^{\ell} \bigl[(1-B_i)^{n^{1/3}y_+} - 1 + n^{1/3}y_+ B_i\bigr] + o(1). \tag{8.12}$$

Using the notation (3.33)–(3.34), we thus have by (3.35) and (3.18)

$$\begin{aligned}\widehat{A}_s^{(n)} - \widehat{A}_t^{(n)} &\leqslant 2n^{-1/6}\Phi_\ell \widehat{H}_{s,t}^{y_+} + o(1) \\ &= 2t^{1/2}\tilde{\beta}\int_s^t \Bigl(\frac{1}{(1+y_+/(4u))^2} - 1 + \frac{y_+}{2u}\Bigr)\,\mathrm{d}u + o_{\mathrm{p}}(1). \end{aligned} \tag{8.13}$$

Similarly, defining

$$y_- := \min\{u^{-1/2}\tilde{\beta}^{-1}\mathcal{B}(u) : u \in [s,t]\} - \varepsilon \tag{8.14}$$

(adjusted to 0 if (8.14) becomes negative), we obtain a lower bound

$$\widehat{A}_s^{(n)} - \widehat{A}_t^{(n)} \geqslant 2s^{1/2}\tilde{\beta}\int_s^t \Bigl(\frac{1}{(1+y_-/4u)^2} - 1 + \frac{y_-}{2u}\Bigr)\,\mathrm{d}u + o_{\mathrm{p}}(1). \tag{8.15}$$

We now subdivide $[s,t]$ into a large number N of small subintervals of equal length and apply (8.13) and (8.15) for each subinterval $[s_i, t_i]$. Since $u^{-1/2}\tilde{\beta}^{-1}\mathcal{B}(u)$ is continuous, we may choose N such that with probability $> 1-\varepsilon$, for each subinterval, the corresponding values of y_+ and y_- differ by at most 3ε, and also that $t_i/s_i < 1+\varepsilon$. We may then, for each subinterval, replace y_+ and y_- in (8.13) and (8.15) by $u^{-1/2}\tilde{\beta}^{-1}\mathcal{B}(u)$ with a small error, and by summing over all subintervals it finally follows by letting $\varepsilon \to 0$ (we omit the routine details) that

$$\widehat{A}_s^{(n)} - \widehat{A}_t^{(n)} = 2\tilde{\beta}\int_s^t u^{1/2}\Bigl(\bigl(1 + \tfrac14 u^{-3/2}\tilde{\beta}^{-1}\mathcal{B}(u)\bigr)^{-2} - 1 + \tfrac12 u^{-3/2}\tilde{\beta}^{-1}\mathcal{B}(u)\Bigr)\,\mathrm{d}u + o_{\mathrm{p}}(1). \tag{8.16}$$

Since we also assume (8.3), the left-hand side converges a.s. to $\mathcal{A}(s) - \mathcal{A}(t)$, and thus we have, a.s.,

$$\mathcal{A}(s) - \mathcal{A}(t) = 2\tilde{\beta}\int_s^t u^{1/2}\Bigl(\bigl(1 + \tfrac14 u^{-3/2}\tilde{\beta}^{-1}\mathcal{B}(u)\bigr)^{-2} - 1 + \tfrac12 u^{-3/2}\tilde{\beta}^{-1}\mathcal{B}(u)\Bigr)\,\mathrm{d}u. \tag{8.17}$$

This holds a.s. simultaneously for every pair of rational s and t, and thus by continuity a.s. for every real s and t with $0 < s \leqslant t$. Consequently, $\mathcal{A}(t)$ is a.s. continuously differentiable on $(0,\infty)$, with

$$\mathcal{A}'(t) = -2\tilde{\beta}t^{1/2}\Big(\big(1+\tfrac14 t^{-3/2}\tilde{\beta}^{-1}\mathcal{B}(t)\big)^{-2} - 1 + \tfrac12 t^{-3/2}\tilde{\beta}^{-1}\mathcal{B}(t)\Big). \tag{8.18}$$

By the definition of $\mathcal{B}(t)$ in (8.6), this implies that also $\mathcal{B}(t)$ is continuously differentiable and

$$\mathcal{B}'(t) = -\mathcal{A}'(t) = 2\tilde{\beta}t^{1/2}\Big(\big(1+\tfrac14 t^{-3/2}\tilde{\beta}^{-1}\mathcal{B}(t)\big)^{-2} - 1 + \tfrac12 t^{-3/2}\tilde{\beta}^{-1}\mathcal{B}(t)\Big). \tag{8.19}$$

We may simplify a little by defining

$$\widetilde{\mathcal{B}}(t) := \tilde{\beta}^{-1}\mathcal{B}(t). \tag{8.20}$$

Then (8.19) becomes

$$\widetilde{\mathcal{B}}'(t) = 2t^{1/2}\Big(\big(1+\tfrac14 t^{-3/2}\widetilde{\mathcal{B}}(t)\big)^{-2} - 1 + \tfrac12 t^{-3/2}\widetilde{\mathcal{B}}(t)\Big). \tag{8.21}$$

By definition, $\widehat{A}^{(n)}_t$ is decreasing on $[0,\infty)$, and thus (8.3) shows that $\mathcal{A}(t)$ is decreasing, and thus $\mathcal{B}(t)$ is increasing by (8.6). (This also follows from (8.21), since the right-hand side is positive.) Furthermore, (8.3), (7.1), (4.36), and Fatou's inequality yield, for every $t > 0$,

$$\mathbb{E}\,\mathcal{A}(t) \leqslant \liminf_{n\to\infty} \mathbb{E}\,\widehat{A}^{(n)}_t \leqslant \liminf_{n\to\infty} n^{-1/2}\frac{Cn}{(tn^{1/3})^{3/2}} = \frac{C}{t^{3/2}}. \tag{8.22}$$

Hence, by dominated convergence,

$$\mathbb{E}\lim_{t\to\infty}\mathcal{A}(t) = \lim_{t\to\infty}\mathbb{E}\,\mathcal{A}(t) = 0. \tag{8.23}$$

Consequently, a.s. $\mathcal{A}(t)\to 0$ as $t\to\infty$, and thus by (8.20) and (8.6)

$$\widetilde{\mathcal{B}}(t) \nearrow \xi, \qquad \text{as } t\to\infty. \tag{8.24}$$

We show in Appendix A below, see in particular (A.14) and (A.16), that the differential equation (8.21) has a unique solution satisfying the boundary condition (8.24), viz.

$$\widetilde{\mathcal{B}}(t) = 4t^{3/2}\big(\big(1+\tfrac34\xi t^{-3/2}\big)^{1/3} - 1\big), \qquad t > 0. \tag{8.25}$$

(It can easily be verified by differentiation that this is a solution; Appendix A shows how the solution may be found, and that it is unique.) Hence, by (8.20),

$$\begin{aligned}\mathcal{B}(t) = \tilde{\beta}\widetilde{\mathcal{B}}(t) &= 4\tilde{\beta}t^{3/2}\big(\big(1+\tfrac34\xi t^{-3/2}\big)^{1/3} - 1\big)\\ &= 4\tilde{\beta}\big(\big(t^{9/2}+\tfrac34\xi t^{3}\big)^{1/3} - t^{3/2}\big).\end{aligned} \tag{8.26}$$

Step 3. Convergence in $C[0,\infty)$. Note that (8.26) shows that $\mathcal{B}(t)$ extends to a continuous function on $[0,\infty)$ with $\mathcal{B}(0) = 0$; hence it follows from (8.6) that also $\mathcal{A}(t)$ extends to a continuous function on $[0,\infty)$ with $\mathcal{A}(0) = \tilde{\beta}\xi$. Using $A^{(n)}_0 = M^{(n)}_0 - W^{(n)}_0 \leqslant M^{(n)}_0$, and the assumed a.s. versions of (6.3) and (5.22), we have that, a.s.,

$$\limsup_{n\to\infty} n^{-1/2}A^{(n)}_0 \leqslant \limsup_{n\to\infty} n^{-1/2}M^{(n)}_0 = \tilde{\beta}\xi = \mathcal{A}(0). \tag{8.27}$$

By the reverse increasing property of $A^{(n)}_k$, we also have, for every $t > 0$,

$$\liminf_{n\to\infty} n^{-1/2}A^{(n)}_0 \geqslant \liminf_{n\to\infty} n^{-1/2}A^{(n)}_{tn^{1/3}} = \mathcal{A}(t). \tag{8.28}$$

Sending $t \searrow 0$ and thus $\mathcal{A}(t)\nearrow\mathcal{A}(0)$ then yields

$$\liminf_{n\to\infty} n^{-1/2}A^{(n)}_0 \geqslant \mathcal{A}(0). \tag{8.29}$$

It follows that, $\widehat{A}_0^{(n)} \to \mathcal{A}(0)$ a.s., and thus (8.3) holds a.s. for every fixed $t \geqslant 0$. Since $\widehat{A}_t^{(n)}$ and $\mathcal{A}(t)$ are decreasing in t, and $\mathcal{A}(t)$ is continuous, this implies that (8.3) holds a.s. uniformly for every interval $[0,b]$ with $0 < b < \infty$. It then follows that also (8.6) holds a.s. uniformly on every compact interval in $[0,\infty)$; in other words, (8.6) holds a.s. in $C[0,\infty)$, with $\mathcal{B}(t)$ as in (8.26).

Step 4. Conclusion. We have so far considered a subsequence, and a special coupling, and have then shown (8.6) in $C[0,\infty)$ and (8.8) in $C(0,\infty)$, which by (8.26) yields (8.1) and (8.2). Since (8.1) and (8.2) use convergence in distribution, they hold in general along the subsequence, also without the special coupling used in the proof. Moreover, the limits in (8.1) and (8.2) do not depend on the subsequence, and the proof shows that every subsequence has a subsequence such that the limits in distribution (8.1) and (8.2) hold. As is well known, this implies that the full sequences converge in distribution, see e.g. [10, Section 5.7]). ☐

Remark 8.3. The argument above using possibly different $B_i^{(n)}$ is rather technical. A more elegant, and perhaps more intuitive version, of the argument would be to assume that B_i is the same for all n, and then condition on $(B_i)_{i=1}^{\infty}$ before applying the Skorohod coupling theorem. However, while intuitively clear, this seems technically more difficult to justify, and it seems to require that we prove that earlier convergence results hold also conditioned on $(B_i)_{i=1}^{\infty}$, a.s. We therefore prefer the somewhat clumsy argument above.

9. The number of descendants

Let $X = X^{(n)}$ be the number of red vertices in the preferential attachment graph. Vertex n is red by definition, and $J_k = \mathbf{1}[Z_k \geqslant 1]$ is the indicator that takes value 1 if vertex k is red for $k < n$; thus

$$X = 1 + \sum_{k=1}^{n-1} J_k, \tag{9.1}$$

noting that J_k is $\mathcal{F}_{k-1}$-measurable. We now set out to prove (1.7) (a special case of Theorem 1.2 with $m = 2$ and $\rho = 0$), which we for convenience restate below as a separate theorem.

Theorem 9.1. *As $n \to \infty$,*

$$\frac{X^{(n)}}{n^{1/3}} \overset{\mathrm{d}}{\longrightarrow} 2^{-4/3}3^{-1/3}\frac{\Gamma(\frac{1}{3})^2}{\Gamma(\frac{2}{3})}\xi^{2/3}, \tag{9.2}$$

where $\xi \in \mathrm{Gamma}(2,1)$.

Remark 9.2. Unlike in (8.1), $\tilde{\beta}$ does not appear in the distributional limit of $X^{(n)}/n^{1/3}$. This is because β in (3.17) is essentially determined by B_k corresponding to the first few vertices; and in the red subgraph D_n, the number of these vertices is insignificant, since most vertices have labels of the order $n^{1/3}$.

As in [13], we use the Doob decomposition

$$X = 1 + L_0 + P_0, \tag{9.3}$$

where for $0 \leqslant k \leqslant n-1$,

$$L_k := \sum_{i=k+1}^{n-1} \bigl(J_i - \mathbb{E}(J_i \mid \mathcal{F}_i)\bigr) \tag{9.4}$$

is a reverse martingale, and by (4.3),

$$P_k := \sum_{i=k+1}^{n-1} \mathbb{E}(J_i \mid \mathcal{F}_i) = \sum_{i=k+1}^{n-1} \mathbb{P}(Z_i \geqslant 1 \mid \mathcal{F}_i) = \sum_{i=k+1}^{n-1} \Bigl(1-(1-B_i)^{Y_i}\Bigr) \tag{9.5}$$

is positive and reverse increasing. Furthermore, by Markov's inequality and (4.3),

$$P_{k-1} - P_k = \mathbb{P}(Z_k \geqslant 1 \mid \mathcal{F}_k) \leqslant B_k Y_k, \quad 1 \leqslant k \leqslant n-1. \tag{9.6}$$

By (9.5) and Lemma 4.2, for $1 \leqslant k \leqslant n-1$,

$$\mathbb{E}\, P_k = \sum_{i=k+1}^{n-1} \mathbb{P}(Z_i \geqslant 1) \leqslant \sum_{i=k+1}^{n-1} \frac{Cn^{1/2}}{i^{3/2}} \leqslant \frac{Cn^{1/2}}{k^{1/2}}. \tag{9.7}$$

Using also the crude bound $0 \leqslant J_i \leqslant 1$, we have $P_0 - P_k \leqslant k$. Choosing $k = \lfloor n^{1/3} \rfloor$ and applying (9.7) thus yield

$$\mathbb{E}\, P_0 \leqslant \mathbb{E}\, P_{\lfloor n^{1/3} \rfloor} + \lfloor n^{1/3} \rfloor \leqslant Cn^{1/3}. \tag{9.8}$$

Moreover, it follows from the reverse martingale property of L_k, $\operatorname{Var}(J_i \mid \mathcal{F}_i) \leqslant \mathbb{E}(J_i \mid \mathcal{F}_i)$ and (9.8) that

$$\mathbb{E}\, L_0^2 = \sum_{i=1}^{n-1} \mathbb{E}[\operatorname{Var}(J_i \mid \mathcal{F}_i)] \leqslant \sum_{i=1}^{n-1} \mathbb{E}\, J_i = \mathbb{E}\, P_0 \leqslant Cn^{1/3}. \tag{9.9}$$

This in turn implies that as $n \to \infty$,

$$\frac{L_0}{n^{1/3}} \xrightarrow{p} 0, \tag{9.10}$$

and thus by (9.3), it is enough to show that as $n \to \infty$, $n^{-1/3} P_0$ converges in distribution to the right-hand side of (9.2). To this end, we extend also P_k to real arguments by linear interpolation and define

$$\widehat{P}_t^{(n)} = n^{-1/3} P_{tn^{1/3}}^{(n)}, \qquad t \geqslant 0. \tag{9.11}$$

Lemma 9.3. *Let $0 < a < b < \infty$. Then the stochastic processes $\widehat{P}_t^{(n)}$, $n \geqslant 1$, are tight in $C[a, b]$.*

Proof. The proof is very similar to that of Lemma 7.1. First, by (9.11) and (9.8),

$$\mathbb{E}\, \widehat{P}_a^{(n)} = n^{-1/3}\, \mathbb{E}\, P_{an^{1/3}}^{(n)} \leqslant n^{-1/3}\, \mathbb{E}\, P_0^{(n)} \leqslant C, \tag{9.12}$$

and thus the sequence $(\widehat{P}_a^{(n)})_{n=1}^{\infty}$ is tight.

Let $\mathfrak{M}_n$ and $\widehat{\Psi}_n$ be as in (7.17), T_k and $\widehat{T}_k^{(n)}$ be as in (7.8) and (7.9). From (9.6) and (4.6), $W_k \leqslant M_k$, the increasing property of Φ_k and also (7.8), for any integer $k \geqslant an^{1/3}$,

$$|P_k - P_{k-1}| \leqslant B_k Y_k = B_k \Phi_k^{-1} W_k \leqslant B_k \Phi_{\lfloor an^{1/3} \rfloor}^{-1} \max_{k \geqslant 1} M_k = n^{1/3} \mathfrak{M}_n \widehat{\Psi}_n (T_k - T_{k-1}). \tag{9.13}$$

Extending (9.13) to real arguments and using (9.11), we thus have

$$|\widehat{P}_t^{(n)} - \widehat{P}_s^{(n)}| \leqslant \mathfrak{M}_n \widehat{\Psi}_n (\widehat{T}_t^{(n)} - \widehat{T}_s^{(n)}) \quad \text{if } a \leqslant s \leqslant t. \tag{9.14}$$

The result then follows from Lemmas 7.2 and 7.3, this time taking $X_n(t) := \widehat{P}_t^{(n)}$, $Y_n(t) := \widehat{T}_t^{(n)} - \widehat{T}_a^{(n)}$ and $Z_n := \mathfrak{M}_n \widehat{\Psi}_n$; tightness of $\mathfrak{M}_n \widehat{\Psi}_n$ follows from that of $(\mathfrak{M}_n)_0^{\infty}$ and $(\widehat{\Psi}_n)_0^{\infty}$. ∎

Proof of Theorem 9.1. In view of Lemma 9.3, by considering a subsequence, we may assume that the processes $\widehat{P}_t^{(n)}$ converge in distribution in every space $C[a, b]$ for $0 < a < b < \infty$, and thus in $C(0, \infty)$, to some stochastic process $\mathcal{P}(t)$ on $(0, \infty)$. Again using the Skorohod coupling theorem, we can assume that all a.s. convergence results in the proof of Theorem 8.1 hold and also

$$\widehat{P}_t^{(n)} \to \mathcal{P}(t) \tag{9.15}$$

a.s. uniformly on every interval $[a, b]$. From (8.8),

$$n^{-1/3} Y_{\lfloor tn^{1/3} \rfloor} \to t^{-1/2} \tilde{\beta}^{-1} \mathcal{B}(t) = t^{-1/2} \widetilde{\mathcal{B}}(t) \tag{9.16}$$

a.s. uniformly on each compact interval in $(0, \infty)$.

Let s, t be real numbers with $0 < s < t$, and let $k = \lfloor sn^{1/3} \rfloor$ and $\ell = \lfloor tn^{1/3} \rfloor$. By the same argument leading to (8.9), now using (9.15) and (9.5), we have

$$\widehat{P}_s^{(n)} - \widehat{P}_t^{(n)} = n^{-1/3} \sum_{i=k+1}^{\ell} \bigl(1 - (1 - B_i)^{Y_i}\bigr) + o(1) \tag{9.17}$$

Let y_+ and y_- be as in (8.11) and (8.14). By (3.36) of Lemma 3.7, we can therefore conclude that

$$\widehat{P}_s^{(n)} - \widehat{P}_t^{(n)} \leqslant \int_s^t \Bigl(1 - \frac{1}{(1 + y_+/(4u))^2}\Bigr)\,\mathrm{d}u + o_{\mathrm{p}}(1) \tag{9.18}$$

and

$$\widehat{P}_s^{(n)} - \widehat{P}_t^{(n)} \geqslant \int_s^t \Bigl(1 - \frac{1}{(1 + y_-/(4u))^2}\Bigr)\,\mathrm{d}u + o_{\mathrm{p}}(1). \tag{9.19}$$

The sandwich argument in the proof of (8.16), the bounds in (9.18) and (9.19), together with (8.25), imply that for $0 < s < t < \infty$,

$$\begin{aligned} \widehat{P}_s^{(n)} - \widehat{P}_t^{(n)} &= \int_s^t \Bigl(1 - \frac{1}{(1 + u^{-1/2}\widetilde{\mathcal{B}}(u)/(4u))^2}\Bigr)\,\mathrm{d}u + o_{\mathrm{p}}(1) \\ &= \int_s^t \Bigl(1 - \frac{1}{\bigl(1 + \frac{3}{4}\xi u^{-3/2}\bigr)^{2/3}}\Bigr)\,\mathrm{d}u + o_{\mathrm{p}}(1). \end{aligned} \tag{9.20}$$

In light of (9.15), we thus have a.s.,

$$\mathcal{P}(t) - \mathcal{P}(s) = \int_s^t \Bigl(1 - \frac{1}{\bigl(1 + \frac{3}{4}\xi u^{-3/2}\bigr)^{2/3}}\Bigr)\,\mathrm{d}u. \tag{9.21}$$

Furthermore,

$$\frac{\mathrm{d}}{\mathrm{d}s} \widehat{P}_s^{(n)} = -\mathbb{E}(J_k \mid \mathcal{F}_k) = (1 - B_k)^{Y_k} - 1, \tag{9.22}$$

where $k = \lceil sn^{1/3} \rceil$. Thus $\bigl|\frac{\mathrm{d}}{\mathrm{d}s} \widehat{P}_s^{(n)}\bigr| \leqslant 1$, which in turn implies that

$$\bigl|\widehat{P}_0^{(n)} - \widehat{P}_s^{(n)}\bigr| \leqslant s. \tag{9.23}$$

For $t \geqslant 1$, it follows from the reverse increasing property of P_k and (9.7) that

$$\mathbb{E}\,\widehat{P}_t^{(n)} = n^{-1/3}\,\mathbb{E}\,P_{tn^{1/3}}^{(n)} \leqslant n^{-1/3}\,\mathbb{E}\,P_{\lfloor tn^{1/3} \rfloor}^{(n)} \leqslant \frac{Cn^{1/6}}{\lfloor tn^{1/3} \rfloor^{1/2}} \leqslant Ct^{-1/2}. \tag{9.24}$$

Sending $s \to 0$ and $t \to \infty$, we deduce from (9.23) and (9.24) that $\widehat{P}_0^{(n)} - (\widehat{P}_s^{(n)} - \widehat{P}_t^{(n)}) \xrightarrow{\mathrm{p}} 0$, uniformly in n. Combining this and (9.21) with a standard argument gives

$$\widehat{P}_0^{(n)} \xrightarrow{\mathrm{p}} \int_0^\infty \Bigl(1 - \frac{1}{\bigl(1 + \frac{3}{4}\xi u^{-3/2}\bigr)^{2/3}}\Bigr)\,\mathrm{d}u. \tag{9.25}$$

The change of variable $x = 3\xi u^{-3/2}/4$ yields, using Lemma B.1,

$$\begin{aligned} \int_0^\infty \Bigl(1 - \frac{1}{\bigl(1 + \frac{3}{4}\xi u^{-3/2}\bigr)^{2/3}}\Bigr)\,\mathrm{d}u &= \frac{2}{3}\Bigl(\frac{3\xi}{4}\Bigr)^{2/3} \int_0^\infty \Bigl(1 - \frac{1}{(1+x)^{2/3}}\Bigr) x^{-5/3}\,\mathrm{d}x \\ &= \frac{2}{3}\Bigl(\frac{3}{4}\Bigr)^{2/3} \xi^{2/3} \frac{-\Gamma(-\frac{2}{3})\Gamma(\frac{4}{3})}{\Gamma(\frac{2}{3})} \end{aligned}$$

$$= 2^{-4/3}3^{-1/3}\frac{\Gamma(\frac{1}{3})^2}{\Gamma(\frac{2}{3})}\xi^{2/3}. \tag{9.26}$$

The result thus follows from (9.3), (9.10), (9.11), (9.25), and (9.26). □

10. The general case

Here we consider the general case. The argument is similar to the case where $m = 2$ and $\rho = 0$, so we give only the main changes here.

10.1. **The stochastic recursions and new estimates.** Define $Y_k, J_k, Z_k, \mathcal{F}_k$ as in Section 4, but now with the boundary condition $Y_{n-1} = m$. We use the stochastic recursions in Section 4.1 to obtain the subgraph D_n, where we now sample m outgoing edges instead of two. The recursion in (4.2) now becomes

$$Y_{k-1} = Y_k - Z_k + m \cdot J_k = Y_k - Z_k + m \cdot \mathbf{1}[Z_k \geqslant 1], \qquad 2 \leqslant k \leqslant n-1. \tag{10.1}$$

As (4.3) still holds, we have

$$\begin{aligned}\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) &= Y_k - \mathbb{E}(Z_k \mid \mathcal{F}_k) + m\mathbb{P}(Z_k \geqslant 1 \mid \mathcal{F}_k)\\ &= Y_k - B_kY_k + m\bigl(1-(1-B_k)^{Y_k}\bigr),\end{aligned} \tag{10.2}$$

and, again by Markov's inequality,

$$\mathbb{E}(Y_{k-1} \mid \mathcal{F}_k) \leqslant Y_k - B_kY_k + mB_kY_k = (1+(m-1)B_k)Y_k. \tag{10.3}$$

Thus, with Φ_k as in (3.5), we can define

$$W_k := \Phi_k Y_k, \qquad 0 \leqslant k \leqslant n-1. \tag{10.4}$$

It follows from (10.3) and (10.4) that $W_0, \dots W_{n-1}$ is a reverse supermartingale with

$$W_{n-1} = \Phi_{n-1}Y_{n-1} = m\Phi_{n-1.} \tag{10.5}$$

We again consider the Doob decomposition $W_k = M_k - A_k$, with

$$M_k := m\Phi_{n-1} + \sum_{j=k+1}^{n-1}(W_{j-1} - \mathbb{E}(W_{j-1} \mid \mathcal{F}_j)), \tag{10.6}$$

and A_k as in (4.11). Analogous to (4.14) and (4.16),

$$A_{k-1} - A_k = m\Phi_{k-1}\bigl((1-B_k)^{Y_k} - 1 + B_kY_k\bigr) \tag{10.7}$$

and

$$\operatorname{Var}(W_{k-1} \mid \mathcal{F}_k) \leqslant C\Phi_{k-1}^2 B_k Y_k. \tag{10.8}$$

Using (10.8) and arguing as in (4.18), we obtain

$$\operatorname{Var}_{\mathbf{B}}(M_k) \leqslant C\sum_{j=k+1}^{n-1}\Phi_{j-1}^2 B_j \prod_{i=j+1}^{n-1}\bigl(1+(m-1)B_i\bigr). \tag{10.9}$$

By the same proofs as for Lemmas 4.1–4.3, again using (3.6), (3.7), (3.8), (3.9) and (3.11), we get

$$\mathbb{E}\max_{0\leqslant k\leqslant n-1} W_k^2 \leqslant \mathbb{E}\max_{0\leqslant k\leqslant n-1} M_k^2 \leqslant Cn^{2(m-1)\chi}; \tag{10.10}$$

$$\mathbb{P}(Z_k \geqslant 1) \leqslant \frac{Cn^{(m-1)\chi}}{k^{1+(m-1)\chi}}, \qquad \mathbb{P}(Z_k \geqslant 2) \leqslant \frac{Cn^{2(m-1)\chi}}{k^{2+2(m-1)\chi}}; \tag{10.11}$$

$$A_{k-1} - A_k \leqslant C\Phi_k^{-1}W_k^2B_k^2, \qquad \mathbb{E}A_k \leqslant \frac{Cn^{2(m-1)\chi}}{k^{1+(m-1)\chi}}. \tag{10.12}$$

10.2. **The branching process.** As in Section 5, we can couple the early part of D_n to a suitable time-changed branching process. Let $\mathcal{Y}$ be a branching process that starts with m particles at time 0, and each particle has an independent Exp(1) lifetime, before splitting into m new particles. Let $\widehat{\mathcal{Y}}$ be the time-changed counterpart of $\mathcal{Y}$, again by the mapping $t \mapsto e^{-t}$; thus $\widehat{\mathcal{Y}}_x = \mathcal{Y}_{-\log x}$ is the number of particles in $\widehat{\mathcal{Y}}$ alive at time x. By standard properties of branching processes (see [13, Section 8] and e.g. [1, Chapter III])

$$\mathbb{E}\,\mathcal{Y}_t = me^{(m-1)t} \tag{10.13}$$

and as $t \to \infty$ and thus $x = e^{-t} \to 0$,

$$x^{m-1}\widehat{\mathcal{Y}}_x = e^{-(m-1)t}\mathcal{Y}_t \xrightarrow{\text{a.s.}} \xi \in \operatorname{Gamma}\Bigl(\frac{m}{m-1}, m-1\Bigr). \tag{10.14}$$

The statements of Lemma 5.1 and Theorem 5.2 hold with the same n_1 and κ_n, but χ is now as in (3.2), and $\delta_n = 3\log^m n/n^{1/3}$. The analogue of Lemma 5.1 can be proved using entirely the same argument, but several straightforward modifications are needed to obtain the analogue of Theorem 5.2. For instance, in Step 1, we use (10.11) and $n_1 = \lfloor n/\log n \rfloor$ to show that

$$\sum_{k=n_1}^{n-1} \mathbb{P}(Z_k \geqslant 2) = O\bigl(n^{-1}(\log n)^{1+2(m-1)\chi}\bigr) = o(1). \tag{10.15}$$

In Step 2, it follows from (10.13) that

$$\mathbb{E}\,\widehat{\mathcal{Y}}_x = \mathbb{E}\,\mathcal{Y}_{-\log x} = \frac{m}{x^{m-1}}, \quad 0 < x \leqslant 1, \tag{10.16}$$

and so for $x = (n_1/n)^\chi \sim \log^\chi n$, w.h.p. there are at most $\log^{m-1} n$ generations from the root 1 to any point in $[(n_1/n)^\chi, 1]$. Adjusting the remaining steps accordingly then yield the desired conclusion.

Redefine

$$\Xi^{(n)} := \frac{W_{n_1}^{(n)}}{n^{(m-1)\chi}}. \tag{10.17}$$

It follows from (10.17), (4.6). (3.5), (3.18), (5.13), and (10.14), that the statement of Lemma 5.3 holds, with $\tilde{\beta}$ as in (3.18), and $\xi \in \operatorname{Gamma}(m/(m-1), m-1)$, independent of $\tilde{\beta}$.

10.3. **The flat middle part.** We first note that ν defined in (1.3) also satisfies, by (3.2) and a simple calculation,

$$\nu = \frac{(m-1)\chi}{1+(m-1)\chi} \tag{10.18}$$

and thus

$$(1-\nu)(m-1)\chi = \nu. \tag{10.19}$$

We now choose $n^\nu \ll n_2 \leqslant n_1 := \lfloor n/\log n \rfloor$. Then, as in Section 6, (10.12) and (10.18) yield

$$\mathbb{E} \max_{n_2 \leqslant k \leqslant n-1} \left| \frac{A_k}{n^{(m-1)\chi}} \right| = \frac{\mathbb{E}\,A_{n_2}}{n^{(m-1)\chi}} \leqslant C\frac{n^{(m-1)\chi}}{n_2^{1+(m-1)\chi}} = C\Bigl(\frac{n^\nu}{n_2}\Bigr)^{1+(m-1)\chi} = o(1). \tag{10.20}$$

Hence Lemma 6.1 holds, with denominators $n^{(m-1)\chi}$. Similarly, the proofs in Section 6, now using (10.9), (3.8), (3.11), and (10.17), yield the conclusion that as $n \to \infty$,

$$\max_{0 \leqslant k \leqslant n_1} \left| \frac{M_k}{n^{(m-1)\chi}} - \Xi^{(n)} \right| \xrightarrow{\text{p}} 0; \tag{10.21}$$

$$\max_{n_2 \leqslant k \leqslant n_1} \left| \frac{W_k}{n^{(m-1)\chi}} - \Xi^{(n)} \right| \xrightarrow{\text{p}} 0. \tag{10.22}$$

10.4. **The final part.** Let V_k and T_k be as in (7.8). As before, we extend W_k, M_k, A_k, V_k, T_k to real arguments by linear interpolation. Now let, for $t \geqslant 0$,

$$\widehat{A}_t^{(n)} := n^{-(m-1)\chi} A_{tn^\nu}^{(n)}, \tag{10.23}$$

$$\widehat{V}_t^{(n)} := n^{\nu} V_{tn^\nu}^{(n)}, \tag{10.24}$$

$$\widehat{T}_t^{(n)} := T_{tn^\nu}^{(n)}. \tag{10.25}$$

Note that Lemma 7.3 holds in this more general setting (with the exponent 1/3 replaced by ν in the proof). Moreover, fix $a > 0$ and define also

$$\mathfrak{M}_n := n^{-(m-1)\chi} \max_{k\geqslant 1} M_k \qquad \text{and} \qquad \widehat{\Psi}_n := n^{\nu(m-1)\chi} \Phi_{\lfloor an^\nu \rfloor}^{-1}. \tag{10.26}$$

In view of (10.12), we have, for all real $\lfloor an^\nu \rfloor \leqslant k \leqslant \ell$,

$$|A_\ell - A_k| \leqslant C n^{(m-1)\chi(2-\nu)} \mathfrak{M}_n^2 \widehat{\Psi}_n (V_\ell - V_k). \tag{10.27}$$

Since $(2-\nu)(m-1)\chi - (m-1)\chi = \nu$ by (10.19), it follows from (10.27) that if $a \leqslant s \leqslant t$, then

$$|\widehat{A}_t^{(n)} - \widehat{A}_s^{(n)}| \leqslant C \mathfrak{M}_n^2 \widehat{\Psi}_n (\widehat{V}_t^{(n)} - \widehat{V}_s^{(n)}). \tag{10.28}$$

By (10.23), (10.12), and (10.18), $\mathbb{E}\, \widehat{A}_a^{(n)} \leqslant C_a$, which implies that the sequence $(\widehat{A}_a^{(n)})_{n=1}^\infty$ is tight. From (10.10) and (3.11), $\mathbb{E}\, \mathfrak{M}_n^2 \leqslant C$ and $\mathbb{E}\, \widehat{\Psi}_n = O(1)$, implying that $(\mathfrak{M}_n)_{n=1}^\infty$ and $(\widehat{\Psi}_n)_{n=1}^\infty$ are tight also. Following the proof of Lemma 7.1, with the ingredients above, we then conclude that the stochastic processes $\widehat{A}_t^{(n)}$, $n \geqslant 1$, are tight in $C[a, b]$ for $0 < a < b < \infty$.

Therefore, arguing as in the beginning of the proof of Theorem 8.1, we may assume (by considering a subsequence and a special coupling) that (8.3) holds a.s. in $C(0, \infty)$ together with (8.5), (5.22), (10.21), (10.22), and, instead of (8.4),

$$\sup_{k \geqslant \log n} \left| k^{-(m-1)\chi} \Phi_k^{(n)} - \tilde{\beta}^{(n)} \right| \overset{\mathrm{p}}{\longrightarrow} 0. \tag{10.29}$$

Then, analogously to (8.6) and (8.8),

$$n^{-(m-1)\chi} W_{tn^\nu} \to \mathcal{B}(t) := \tilde{\beta}\xi - \mathcal{A}(t) \tag{10.30}$$

and, again using (10.19),

$$n^{-\nu} Y_{\lfloor tn^\nu \rfloor} \to t^{-(m-1)\chi} \tilde{\beta}^{-1} \mathcal{B}(t) \tag{10.31}$$

a.s. uniformly on every compact interval in $(0, \infty)$.

Let $k := \lfloor sn^\nu \rfloor$ and $\ell := \lfloor tn^\nu \rfloor$ for some $0 < s < t$. Similarly to (8.9), we have by (8.3) and (10.7)

$$\widehat{A}_s^{(n)} - \widehat{A}_t^{(n)} = m n^{-(m-1)\chi} \sum_{i=k+1}^{\ell} \Phi_{i-1} [(1 - B_i)^{Y_i} - 1 + Y_i B_i] + o(1). \tag{10.32}$$

Following (8.10)–(8.16) with minor adjustments (in particular, choosing $\lambda_n = n^\nu$ in Lemma 3.7 and using again (10.19)) leads to, a.s. for every real $0 < s \leqslant t$,

$$\mathcal{A}(s) - \mathcal{A}(t) = m\tilde{\beta} \int_s^t u^{(m-1)\chi} \Big(\big(1 + \tfrac{1}{\theta} u^{-(1+(m-1)\chi)} \tilde{\beta}^{-1} \mathcal{B}(u)\big)^{-(m+\rho)} - 1 + \chi u^{-(1+(m-1)\chi)} \tilde{\beta}^{-1} \mathcal{B}(u) \Big) \,\mathrm{d}u. \tag{10.33}$$

Let, for convenience, recalling (10.18),

$$\alpha := 1 + (m-1)\chi = \frac{1}{1-\nu}. \tag{10.34}$$

Then, by (10.33), a.s. $\mathcal{A}(t)$ is differentiable on $(0,\infty)$ and

$$\mathcal{A}'(t) = -m\tilde\beta t^{(m-1)\chi}\Bigl(\bigl(1+\tfrac{1}{\theta}t^{-\alpha}\tilde\beta^{-1}\mathcal{B}(t)\bigr)^{-(m+\rho)} - 1 + \chi t^{-\alpha}\tilde\beta^{-1}\mathcal{B}(t)\Bigr), \tag{10.35}$$

and $\mathcal{B}'(t) = -\mathcal{A}'(t)$ by (10.30). Define again

$$\widetilde{\mathcal{B}}(t) = \tilde\beta^{-1}\mathcal{B}(t), \tag{10.36}$$

so that

$$\widetilde{\mathcal{B}}'(t) = mt^{(m-1)\chi}\Bigl(\bigl(1+\tfrac{1}{\theta}t^{-\alpha}\widetilde{\mathcal{B}}(t)\bigr)^{-(m+\rho)} - 1 + \chi t^{-\alpha}\widetilde{\mathcal{B}}(t)\Bigr). \tag{10.37}$$

Moreover, $\mathbb{E}\,\mathcal{A}(t) \leqslant Ct^{-\alpha}$ for $t \geqslant 1$, say, by (10.23), (10.12), and Fatou's lemma, and dominated convergence further implies that $\mathbb{E}\lim_{t\to\infty}\mathcal{A}(t) = 0$. Hence $\mathcal{A}(t) \to 0$ a.s. as $t \to \infty$, and thus we have we have from (10.30) and (10.36) that

$$\widetilde{\mathcal{B}}(t) \nearrow \xi \qquad \text{as } t\to\infty. \tag{10.38}$$

As shown in detail in Appendix A, see (A.14) and (A.16), the unique solution to (10.37) satisfying (10.38) is given by

$$\widetilde{\mathcal{B}}(t) = \theta t^{\alpha}\Bigl(\bigl(1+\tfrac{m+\rho+1}{\theta}\xi t^{-\alpha}\bigr)^{\frac{1}{m+\rho+1}} - 1\Bigr). \tag{10.39}$$

Hence, by (10.36) and (10.39),

$$\mathcal{B}(t) = \tilde\beta\widetilde{\mathcal{B}}(t) = \theta\tilde\beta t^{\alpha}\Bigl(\bigl(1+\tfrac{m+\rho+1}{\theta}\xi t^{-\alpha}\bigr)^{\frac{1}{m+\rho+1}} - 1\Bigr). \tag{10.40}$$

Now, proceeding as in the remaining steps of the proof of Theorem 8.1 yields the conclusion that as $n\to\infty$,

$$n^{-(m-1)\chi}W_{tn^{\nu}} \xrightarrow{\mathrm{d}} \theta\tilde\beta t^{\alpha}\Bigl(\bigl(1+\tfrac{m+\rho+1}{\theta}\xi t^{-\alpha}\bigr)^{\frac{1}{m+\rho+1}} - 1\Bigr) \qquad \text{in } C[0,\infty), \tag{10.41}$$

and

$$n^{-\nu}Y_{tn^{\nu}} \xrightarrow{\mathrm{d}} \theta t\Bigl(\bigl(1+\tfrac{m+\rho+1}{\theta}\xi t^{-\alpha}\bigr)^{\frac{1}{m+\rho+1}} - 1\Bigr) \qquad \text{in } C(0,\infty). \tag{10.42}$$

10.5. **The number of descendants.** As in Section 9, let $X = X^{(n)}$ be the number of red vertices, and define L_k and P_k as in (9.4) and (9.5). As in (9.7), it follows from (10.11) that

$$\mathbb{E}\,P_k \leqslant C\Bigl(\frac{n}{k}\Bigr)^{(m-1)\chi}, \qquad 1\leqslant k\leqslant n-1. \tag{10.43}$$

Furthermore, arguing as in (9.8) with the cutoff n^{ν} yields, recalling (10.19),

$$\mathbb{E}\,P_0 \leqslant Cn^{\nu}. \tag{10.44}$$

The argument for (9.9) now yields

$$\mathbb{E}\,L_0^2 \leqslant Cn^{\nu}, \tag{10.45}$$

which implies that

$$n^{-\nu}L_0 \xrightarrow{\mathrm{p}} 0 \qquad \text{as } n\to\infty. \tag{10.46}$$

As before, we extend P_k to real arguments by linear interpolation, but now let

$$\widehat{P}_t^{(n)} = n^{-\nu}P_{tn^{\nu}}^{(n)}, \qquad t\geqslant 0. \tag{10.47}$$

The same proof as for Lemma 9.3 then shows that for $0<a<b<\infty$, the sequences $\widehat{P}_t^{(n)}$, $n\geqslant 1$ are tight in $C[a,b]$. Proceeding as in the proof of Theorem 9.1, where we use the Skorohod coupling theorem again, we get

$$\begin{aligned}\widehat{P}_0^{(n)} &\xrightarrow{\mathrm{p}} \int_0^{\infty}\Bigl(1-\bigl(1+\theta^{-1}u^{-\alpha}\widetilde{\mathcal{B}}(u)\bigr)^{-(m+\rho)}\Bigr)\,\mathrm{d}u.\\ &= \int_0^{\infty}\Bigl(1-\bigl(1+\tfrac{m+\rho+1}{\theta}\xi u^{-\alpha}\bigr)^{-\frac{m+\rho}{m+\rho+1}}\Bigr)\,\mathrm{d}u.\end{aligned} \tag{10.48}$$

By the change of variable $v = \theta^{-1}(m+\rho+1)\xi u^{-\alpha}$,

$$\int_0^\infty \Bigl(1 - \bigl(1+\tfrac{m+\rho+1}{\theta}\xi u^{-\alpha}\bigr)^{-\frac{m+\rho}{m+\rho+1}}\Bigr)\,\mathrm{d}u = \frac{1}{\alpha}\Bigl(\frac{m+\rho+1}{\theta}\xi\Bigr)^{1/\alpha}\int_0^\infty \Bigl(1-(1+v)^{-\frac{m+\rho}{m+\rho+1}}\Bigr)v^{-(1+1/\alpha)}\,\mathrm{d}v. \tag{10.49}$$

We take $a = -1/\alpha$ and $b = (m+\rho)/(m+\rho+1)$ in Lemma B.1, and note that $1/\alpha = 1-\nu$ by (10.34), and thus by (1.3),

$$b - a = \frac{m+\rho}{m+\rho+1} + \frac{1}{\alpha} = \frac{m+\rho}{m+\rho+1} - \nu + 1 = \frac{m+\rho}{m(m+\rho+1)} + 1. \tag{10.50}$$

Hence, (10.49) and Lemma B.1 yield

$$\begin{aligned}\int_0^\infty \Bigl(1 - \bigl(1+\tfrac{m+\rho+1}{\theta}\xi u^{-\alpha}\bigr)^{-\frac{m+\rho}{m+\rho+1}}\Bigr)\,\mathrm{d}u &= -\frac{1}{\alpha}\cdot\frac{\Gamma(-\frac{1}{\alpha})\Gamma\bigl(\frac{m+\rho}{m(m+\rho+1)}+1\bigr)}{\Gamma\bigl(\frac{m+\rho}{m+\rho+1}\bigr)}\Bigl(\frac{m+\rho+1}{\theta}\xi\Bigr)^{1/\alpha}\\ &= \frac{\Gamma(1-\frac{1}{\alpha})\Gamma\bigl(\frac{m+\rho}{m(m+\rho+1)}+1\bigr)}{\Gamma\bigl(\frac{m+\rho}{m+\rho+1}\bigr)}\Bigl(\frac{m+\rho+1}{\theta}\xi\Bigr)^{1/\alpha}.\end{aligned} \tag{10.51}$$

Finally, (9.3), (10.46), (10.47), (10.48), and (10.51) together imply that as $n\to\infty$,

$$n^{-\nu}X \xrightarrow{\mathrm{d}} \frac{\Gamma(1-\frac{1}{\alpha})\Gamma\bigl(\frac{m+\rho}{m(m+\rho+1)}+1\bigr)}{\Gamma\bigl(\frac{m+\rho}{m+\rho+1}\bigr)}\Bigl(\frac{m+\rho+1}{\theta}\xi\Bigr)^{1/\alpha}. \tag{10.52}$$

We here note that, by (10.34) and (1.3),

$$1-\frac{1}{\alpha} = \nu = \frac{(m-1)(m+\rho)}{m(m+\rho+1)}. \tag{10.53}$$

We write also $\xi = (m-1)\xi_1$, with $\xi_1 \in \mathrm{Gamma}(m/(m-1),1)$, and recall that $\theta = 2m+\rho$. Hence, (10.52) can be written as (1.4). □

11. Moment convergence

In this section, we prove Theorem 1.3 on moment convergence; we use the standard method of proving uniform moment estimates and thus uniform integrability. This time we choose to treat general m and ρ from the beginning.

We consider first the reverse martingale M_k, recalling that $M_k \geqslant W_k \geqslant 0$. We denote the maximal function by

$$M_* := \max_{n-1\geqslant k\geqslant 0} M_k, \tag{11.1}$$

and define the martingale differences, for $n-1\geqslant k\geqslant 1$, recalling (10.6), (10.4), (10.1), and that Y_k is $\mathcal{F}_k$-measurable,

$$\begin{aligned}\Delta M_k := M_{k-1} - M_k &= W_{k-1} - \mathbb{E}\bigl(W_{k-1} \mid \mathcal{F}_k\bigr)\\ &= \Phi_{k-1}\bigl(Y_{k-1} - \mathbb{E}(Y_{k-1}\mid\mathcal{F}_k)\bigr)\\ &= -\Phi_{k-1}\bigl(Z_k - \mathbb{E}(Z_k\mid\mathcal{F}_k)\bigr) + m\Phi_{k-1}\bigl(J_k - \mathbb{E}(J_k\mid\mathcal{F}_k)\bigr).\end{aligned} \tag{11.2}$$

We define also the conditional square function

$$s(M) := \Bigl(\sum_{i=1}^{n-1}\mathbb{E}\bigl((\Delta M_i)^2 \mid \mathcal{F}_i\bigr)\Bigr)^{1/2}. \tag{11.3}$$

Let for convenience

$$\varkappa := (m-1)\chi. \tag{11.4}$$

(Thus, in the case $m = 2$, $\rho = 0$, we have $\varkappa = \chi = \frac{1}{2}$.) We use also the standard notation, for any random variable $\mathcal{X}$,

$$\|\mathcal{X}\|_p := \bigl(\mathbb{E}[|\mathcal{X}|^p]\bigr)^{1/p}. \tag{11.5}$$

Note that for any $p > 0$, (2.1), (3.3), and (3.4) yield, cf. (3.6)–(3.7),

$$\mathbb{E}[B_k^p] \leqslant C_p k^{-p}. \tag{11.6}$$

Lemma 11.1. *For every $p > 0$,*

$$\mathbb{E}[M_*^p] \leqslant C_p n^{p\varkappa}. \tag{11.7}$$

Proof. We assume in the proof for simplicity that $p \geqslant 2$ is an integer; the general case follows by Lyapunov's inequality.

We use as in [13] one of Burkholder's martingale inequalities [7, Theorem 21.1], [10, Corollary 10.9.1] on the reverse martingale $M_k - M_{n-1} = M_k - m\Phi_{n-1}$, which yields

$$\begin{aligned}\mathbb{E}[M_*^p] &\leqslant C_p\,\mathbb{E}[\Phi_{n-1}^p] + C_p\,\mathbb{E}\bigl[\bigl(\max_k |M_k - M_{n-1}|\bigr)^p\bigr]\\ &\leqslant C_p\,\mathbb{E}[\Phi_{n-1}^p] + C_p\,\mathbb{E}[s(M)^p] + C_p\,\mathbb{E}\bigl[\max_k |\Delta M_k|^p\bigr]\\ &\leqslant C_p\,\mathbb{E}[\Phi_{n-1}^p] + C_p\,\mathbb{E}\bigl[s(M)^p\bigr] + C_p\sum_{k=1}^{n-1}\mathbb{E}\bigl[|\Delta M_k|^p\bigr]. \end{aligned}\tag{11.8}$$

We estimate the three terms on the right-hand side separately.

First, we have by the independence of B_i, (3.20), and (3.6), similarly to (3.10)–(3.11),

$$\begin{aligned}\mathbb{E}[\Phi_k^p] &= \prod_{i=1}^k \mathbb{E}\,(1 + (m-1)B_i)^p = \prod_{i=1}^k \Bigl(1 + p(m-1)\frac{\chi}{i} + O\bigl(i^{-2}\bigr)\Bigr)\\ &= \exp\Bigl(\sum_{i=1}^k \Bigl(\frac{p\varkappa}{i} + O\bigl(i^{-2}\bigr)\Bigr)\Bigr) = \exp\Bigl(p\varkappa \log k + O(1)\Bigr)\\ &\leqslant C_p k^{p\varkappa}. \end{aligned}\tag{11.9}$$

Next, by (11.2), (10.8), and (10.4),

$$\begin{aligned}\mathbb{E}\bigl[(\Delta M_k)^2 \mid \mathcal{F}_k\bigr] &= \operatorname{Var}\bigl[W_{k-1} \mid \mathcal{F}_k\bigr] \leqslant C\Phi_{k-1}^2 B_k Y_k \leqslant C\Phi_{k-1}B_k W_k\\ &\leqslant C\Phi_{k-1}B_k M_k \leqslant C\Phi_{k-1}B_k M_*. \end{aligned}\tag{11.10}$$

Note that $\Phi_k - \Phi_{k-1} = (1 + (m-1)B_k - 1)\Phi_{k-1} = (m-1)B_k\Phi_{k-1}$. Hence, (11.3) and (11.10) yield

$$s(M)^2 \leqslant C\sum_{k=1}^{n-1}(\Phi_k - \Phi_{k-1})M_* \leqslant C\Phi_{n-1}M_*. \tag{11.11}$$

Hölder's inequality (or Cauchy–Schwarz's) and (11.9) thus yield

$$\mathbb{E}[s(M)^p] \leqslant C_p\,\mathbb{E}\bigl[\Phi_{n-1}^{p/2}M_*^{p/2}\bigr] \leqslant C_p\Bigl(\mathbb{E}[\Phi_{n-1}^p]\,\mathbb{E}[M_*^p]\Bigr)^{1/2} \leqslant C_p n^{p\varkappa/2}\|M_*\|_p^{p/2}. \tag{11.12}$$

For the final term in (11.8), we use the decomposition of ΔM_k in (11.2) and treat the two terms on the last line there separately. We use as in [13, (7.9)] the well-known general estimate for a binomial random variable $\zeta \in \operatorname{Bin}(N, q)$:

$$\mathbb{E}\,|\zeta - \mathbb{E}\,\zeta|^p \leqslant C_p(Nq)^{p/2} + C_p Nq. \tag{11.13}$$

Conditioned on $\mathcal{F}_k$, we have $Z_k \in \operatorname{Bin}(Y_k, B_k)$ by (4.3), and thus (11.13) yields

$$\mathbb{E}\bigl(\bigl|Z_k - \mathbb{E}(Z_k \mid \mathcal{F}_k)\bigr|^p \mid \mathcal{F}_k\bigr) \leqslant C_p(Y_k B_k)^{p/2} + C_p Y_k B_k. \tag{11.14}$$

Similarly, since $J_k = \mathbf{1}[Z_k \geqslant 1]$ has a conditional Bernoulli distribution,

$$\begin{aligned}\mathbb{E}\big(\big|J_k - \mathbb{E}(J_k \mid \mathcal{F}_k)\big|^p \mid \mathcal{F}_k\big) \leqslant C_p \,\mathbb{E}\big(|J_k|^p \mid \mathcal{F}_k\big) = C_p \,\mathbb{E}\big(J_k \mid \mathcal{F}_k\big) \\ \leqslant C_p \,\mathbb{E}\big(Z_k \mid \mathcal{F}_k\big) = C_p Y_k B_k. \end{aligned} \tag{11.15}$$

Hence, (11.2), (11.14), and (11.15) yield,

$$\begin{aligned}\mathbb{E}\big[|\Delta M_k|^p \mid \mathcal{F}_k\big] &\leqslant C_p \Phi_{k-1}^p \big[\mathbb{E}\big(\big|Z_k - \mathbb{E}(Z_k \mid \mathcal{F}_k)\big|^p \mid \mathcal{F}_k\big) + \mathbb{E}\big(\big|J_k - \mathbb{E}(J_k \mid \mathcal{F}_k)\big|^p \mid \mathcal{F}_k\big)\big] \\ &\leqslant C_p \Phi_{k-1}^p Y_k^{p/2} B_k^{p/2} + C_p \Phi_{k-1}^p Y_k B_k \\ &\leqslant C_p \Phi_{k-1}^{p/2} W_k^{p/2} B_k^{p/2} + C_p \Phi_{k-1}^{p-1} W_k B_k \\ &\leqslant C_p \Phi_{k-1}^{p/2} M_*^{p/2} B_k^{p/2} + C_p \Phi_{k-1}^{p-1} M_* B_k. \end{aligned} \tag{11.16}$$

Hence, using Hölder's inequality, the independence of Φ_{k-1} and B_k, (11.9), and (11.6),

$$\begin{aligned}\mathbb{E}\big[|\Delta M_k|^p\big] &\leqslant C_p \,\mathbb{E}\big[\Phi_{k-1}^{p/2} B_k^{p/2} M_*^{p/2}\big] + C_p \,\mathbb{E}\big[\Phi_{k-1}^{p-1} B_k M_*\big] \\ &\leqslant C_p \Big(\mathbb{E}\big[\Phi_{k-1}^p B_k^p\big]\,\mathbb{E}\big[M_*^p\big]\Big)^{1/2} + C_p \Big(\mathbb{E}\big[\Phi_{k-1}^{2p-2} B_k^2\big]\,\mathbb{E}\big[M_*^2\big]\Big)^{1/2} \\ &\leqslant C_p k^{p\varkappa/2-p/2} \|M_*\|_p^{p/2} + C_p k^{(p-1)\varkappa-1} \|M_*\|_2 \\ &\leqslant C_p k^{p\varkappa/2-1} \|M_*\|_p^{p/2} + C_p k^{(p-1)\varkappa-1} \|M_*\|_p. \end{aligned} \tag{11.17}$$

Consequently,

$$\sum_{k=1}^{n-1} \mathbb{E}\big[|\Delta M_k|^p\big] \leqslant C_p n^{p\varkappa/2} \|M_*\|_p^{p/2} + C_p n^{(p-1)\varkappa} \|M_*\|_p. \tag{11.18}$$

Finally, (11.8) yields, collecting the estimates (11.9), (11.12), and (11.18),

$$\mathbb{E}[M_*^p] \leqslant C_p n^{p\varkappa} + C_p n^{p\varkappa/2} \|M_*\|_p^{p/2} + C_p n^{(p-1)\varkappa} \|M_*\|_p. \tag{11.19}$$

It follows trivially from the definitions that for every n, M_* is deterministically bounded by some constant (depending on n), and thus $\|M_*\|_p < \infty$. Let $x := \|M_*\|_p / n^{\varkappa} \in (0, \infty)$; then (11.19) can be written as

$$x^p \leqslant C_p + C_p x^{p/2} + C_p x. \tag{11.20}$$

Since $p > 1$, it follows that $x \leqslant C_p$, which is the same as (11.7). Alternatively, we can proceed as in [13] to consider only $p = 2^j$, with j being positive integers. The conclusion (11.7) then follows from an induction over j, (11.19) and the base case ($p = 2$) proved in (10.10). □

We use the decomposition $X = 1 + L_0 + P_0$ in (9.3), and estimate the terms L_0 and P_0 separately.

Lemma 11.2. *For every $p > 0$,*

$$\mathbb{E}[P_0^p] \leqslant C_p n^{p\nu}. \tag{11.21}$$

Proof. We may by Lyapunov's inequality assume that $p \geqslant 1$ is an integer. By (9.5) and (4.3),

$$P_k = \sum_{i=k+1}^{n-1} \mathbb{E}(J_i \mid \mathcal{F}_i) \leqslant \sum_{i=k+1}^{n-1} Y_i B_i = \sum_{i=k+1}^{n-1} \Phi_i^{-1} W_i B_i \leqslant M_* \sum_{i=k+1}^{n-1} \Phi_{i-1}^{-1} B_i. \tag{11.22}$$

Hence, by Hölder's and Minkowski's inequalities,

$$\|P_k\|_p \leqslant \|M_*\|_{2p} \left\| \sum_{i=k+1}^{n-1} \Phi_{i-1}^{-1} B_i \right\|_{2p} \leqslant \|M_*\|_{2p} \sum_{i=k+1}^{n-1} \big\|\Phi_{i-1}^{-1} B_i\big\|_{2p}. \tag{11.23}$$

We have $(1+x)^{-p} \leqslant 1 - px + C_p x^2$ for all $x \geqslant 0$, and thus by (3.6)–(3.7) and (11.4), generalizing (3.15),

$$\mathbb{E}\big[(1+(m-1)B_i)^{-p}\big] \leqslant 1 - p(m-1)\,\mathbb{E}[B_i] + C_p\,\mathbb{E}[B_i^2] = 1 - p\varkappa i^{-1} + O(i^{-2}). \tag{11.24}$$

Hence, by the same argument as for (11.9), for any integers $p \geqslant 1$ and $k \geqslant 1$,

$$\begin{aligned}\mathbb{E}[\Phi_k^{-p}] &= \prod_{i=1}^{k} \mathbb{E}\big[(1+(m-1)B_i)^{-p}\big] = \prod_{i=1}^{k}\Big(1 - \frac{p\varkappa}{i} + O\big(i^{-2}\big)\Big) \\ &= \exp\Big(-\sum_{i=1}^{k}\Big(\frac{p\varkappa}{i} + O\big(i^{-2}\big)\Big)\Big) = \exp\Big(-p\varkappa \log k + O(1)\Big) \\ &\leqslant C_p k^{-p\varkappa}.\end{aligned} \tag{11.25}$$

In other words, $\|\Phi_k^{-1}\|_p \leqslant C_p k^{-\varkappa}$. Furthermore, $\|B_k\|_p \leqslant C_p k^{-1}$ by (11.6). Since Φ_{i-1} and B_i are independent, it follows that, for $i \geqslant 2$,

$$\big\|\Phi_{i-1}^{-1} B_i\big\|_p = \|\Phi_{i-1}^{-1}\|_p \|B_i\|_p \leqslant C_p i^{-\varkappa-1}. \tag{11.26}$$

We may here replace p by $2p$, and it follows from (11.23) and (11.7) that, for $k \geqslant 1$,

$$\|P_k\|_p \leqslant \|M_*\|_{2p} \sum_{i=k+1}^{n-1} \big\|\Phi_{i-1}^{-1} B_i\big\|_{2p} \leqslant C_p n^{\varkappa} \sum_{i=k+1}^{n-1} i^{-\varkappa-1} \leqslant C_p n^{\varkappa} k^{-\varkappa}. \tag{11.27}$$

Furthermore, as in Section 9, we have $P_0 - P_k \leqslant k$ for any $k \geqslant 0$, and thus Minkowski's inequality and (11.27) yield, choosing $k := \lfloor n^\nu \rfloor$ and noting that (10.18) and (11.4) imply $\varkappa(1-\nu) = \nu$,

$$\|P_0\|_p \leqslant \|P_k\|_p + k \leqslant C_p n^{\varkappa - \varkappa\nu} + n^\nu \leqslant C_p n^\nu, \tag{11.28}$$

which completes the proof. □

Lemma 11.3. *For every $p > 0$,*

$$\mathbb{E}[|L_0|^p] \leqslant C_p n^{p\nu/2}. \tag{11.29}$$

Proof. Recall that $(L_k)_{k=0}^{n-1}$ is a reverse martingale. By (9.4), its conditional square function is given by

$$s(L)^2 := \sum_{i=1}^{n-1} \mathbb{E}\big[\big(J_i - \mathbb{E}(J_i \mid \mathcal{F}_i)\big)^2 \mid \mathcal{F}_i\big] = \sum_{i=1}^{n-1} \operatorname{Var}[J_i \mid \mathcal{F}_i] \leqslant \sum_{i=1}^{n-1} \mathbb{E}[J_i \mid \mathcal{F}_i] = P_0, \tag{11.30}$$

where the inequality follows because J_i has a conditional Bernoulli distribution. Furthermore, again using (9.4), the martingale differences $\Delta L_k := L_{k-1} - L_k$ are bounded by

$$|\Delta L_k| = \big|J_k - \mathbb{E}(J_k \mid \mathcal{F}_k)\big| \leqslant 1. \tag{11.31}$$

Hence, Burkholder's inequality yields, similarly to (11.8), using also Lemma 11.2,

$$\mathbb{E}[L_0^p] \leqslant C_p\,\mathbb{E}[s(L)^p] + C_p\,\mathbb{E}\big[\max_k |\Delta L_k|^p\big] \leqslant C_p\,\mathbb{E}[P_0^{p/2}] + C_p \leqslant C_p n^{p\nu/2}, \tag{11.32}$$

which completes the proof. □

Proof of Theorem 1.3. It follows from (9.3) and Lemmas 11.2 and 11.3 that, for any $p > 0$,

$$\mathbb{E}[X^p] \leqslant C_p + C_p\,\mathbb{E}[L_0^p] + C_p\,\mathbb{E}[P_0^p] \leqslant C_p n^{p\nu}. \tag{11.33}$$

In other words, $\mathbb{E}[(X^{(n)}/n^\nu)^p] \leqslant C_p$ for every $p > 0$. By a standard argument, see e.g. [10, Theorems 5.4.2 and 5.5.9], this implies uniform integrability of the sequence $|X^{(n)}/n^\nu|^p$

for every $p > 0$ and thus the convergence in distribution in (1.4) implies convergence of all moments.

Since $\xi_1 \in \operatorname{Gamma}\bigl(\frac{m}{m-1}, 1\bigr)$,

$$\mathbb{E}\bigl[\xi_1^{p(1-\nu)}\bigr] = \frac{\Gamma(p(1-\nu)+\frac{m}{m-1})}{\Gamma(\frac{m}{m-1})}, \tag{11.34}$$

and thus the explicit formula (1.5) follows. □

12. The model with self-loops

In this section, we consider a variation of the preferential attachment graph in Definition 1.1, where self-loops are possible. We use the version in [11, Section 8.2] (see also [5, 6]) and start with a single vertex 1 with m self-loops. For $n \geqslant 2$, each outgoing edge of vertex n is now attached to a vertex $j \in [n]$, again with probability proportional to ρ + the current degree of vertex j, where we define the current degree of vertex n when we add the $(k+1)$th edge from it to be k + 1 + the number of loops attached to n so far. (We thus count all outgoing edges up to the $(k+1)$th; a loop contributes 2 to the degree.) Hence, recalling that $d_j(n)$ is the degree of vertex j in G_n, when adding vertex $n \geqslant 2$ to G_{n-1}, the $(k+1)$-th outgoing edge of vertex n attaches to vertex $j \in [n]$ with probability

$$\begin{cases} \frac{d_j(n-1)+\sum_{\ell=1}^k \mathbf{1}[n\overset{\ell}{\to}j]+\rho}{2(n-1)m+2k+1+n\rho}, & j<n, \\ \frac{k+1+\sum_{\ell=1}^k \mathbf{1}[n\overset{\ell}{\to}j]+\rho}{2(n-1)m+2k+1+n\rho}, & j=n. \end{cases} \tag{12.1}$$

Remark 12.1. The details of the model can be modified without affecting the following asymptotic result, with only straightforward changes to its proof. For example, we may again start with m edges between vertices 1 and 2, and thus no loops there, or we may include all m outgoing edges in the weight of vertex n when we add edges from it. We leave the details to the reader.

Theorem 12.2. *Let $X^{(n)}$ be the number of descendants of vertex n in the model above. Then, the statements of Theorems 1.2 and 1.3 hold.*

The proof of Theorem 12.2 is largely similar to those of Theorems 1.2 and 1.3 so we only sketch the main differences here.

First, let N_i be the number of self-loops at vertex i. When we add the m edges from a new vertex i, the weight of vertex i and the total weight of the first $i-1$ vertices evolve like a Pólya urn $\mathcal{U}_i'$ with initially $1+\rho$ red and $(2i-2)m+(i-1)\rho$ black balls, where we add 2 new balls at each draw: 2 red balls when a red ball is drawn, and one ball of each colour when a black ball is drawn; N_i is the number of times a red ball is drawn. Note that this urn does not depend on what has happened when the edges from earlier vertices were added, and in particular not on $N_1,\dots,N_{i-1}$. Consequently, the random numbers $(N_i)_{i=1}^\infty$ are independent. Furthermore, if we condition on the entire sequence $(N_i)_{i=1}^\infty$, then the non-loop edges are added from each new vertex $n \geqslant 2$ to $[n-1]$ by the same random procedure as in Definition 1.1, except that now we add $m - N_n$ new edges from n, and that the degrees of the vertices include also any existing loops. This means that after we have added vertex $j \geqslant 2$, the weight of vertex j and the total weight of the first $j-1$ vertices evolve like a standard Pólya urn $\mathcal{U}_j''$ with initially $m+N_j+\rho$ red and $(2j-1)m-N_j+(j-1)\rho$ black balls, after each draw adding one ball of the same colour as the drawn ball. As a consequence, the proportion of red balls converges a.s. to a random number B_j with the (conditional) beta distribution

$$B_j \mid (N_i)_{i=1}^\infty \in \operatorname{Beta}(m+N_j+\rho, (2j-1)m-N_j+(j-1)\rho), \qquad j \geqslant 2. \tag{12.2}$$

Moreover, conditioned on $(N_i)_{i=1}^{\infty}$, we can again construct the preferential attachment graph by the Pólya urn representation in Definition 2.1–Remark 2.3, using (conditionally) independent B_j with the distributions (12.2). (As before, we also let $B_1 := 1$.) In particular, note that since $(N_i)_{i=2}^{\infty}$ are independent, the random variables $(B_i)_{i=2}^{\infty}$ are independent. and so are the pairs of random variables (N_i, B_i), $i \geqslant 2$.

The distribution of each B_j is thus a mixed beta distribution, but we do not need exact expressions. We will show that all estimates in Section 3 still hold (possibly with different constants C). Note first that in the urn $\mathcal{U}_i'$ used to determine N_i, we make m draws and thus the number of red balls is at most $m+\rho = O(1)$; hence the probability of drawing a red ball is $O(1/i)$ for each draw, and thus

$$\mathbb{P}(N_i > 0) \leqslant \mathbb{E}\, N_i = O(1/i). \tag{12.3}$$

Recall θ and χ in (3.1) and (3.2). Using $0 \leqslant N_i \leqslant m$, (12.2), (12.3), and (3.3), it is easy to show that

$$\mathbb{E}\, B_i = \frac{m+\rho+\mathbb{E}\, N_i}{\theta i} = \frac{\chi}{i} + O(i^{-2}), \tag{12.4}$$

$$\mathbb{E}\, B_i^r \leqslant \prod_{j=0}^{r-1} \frac{2m+\rho+j}{\theta i + j} \leqslant C_r i^{-r}, \quad r \geqslant 2. \tag{12.5}$$

Similarly, we have, by first conditioning on N_i,

$$\mathbb{E}[N_i B_i] = \frac{\mathbb{E}[N_i(m+\rho+N_i)]}{\theta i} \leqslant \frac{\mathbb{E}[N_i(2m+\rho)]}{\theta i} = O(i^{-2}), \tag{12.6}$$

and the bound

$$\mathbb{E}\big[N_i^r B_i^r\big] = O\big(i^{-(r+1)}\big), \qquad \text{for each } r \geqslant 2. \tag{12.7}$$

Define Φ_i and $S_{n,i}$ by (3.5) and (2.3) as before. Then (12.4) and a little calculation using (3.4) shows that, for $2 \leqslant j \leqslant k < \infty$,

$$\begin{aligned}
\prod_{i=j}^{k} \mathbb{E}(1+(m-1)B_i) &= \prod_{i=j}^{k} \frac{i+(m-1)\chi}{i} \cdot \prod_{i=j}^{k} \frac{\mathbb{E}(1+(m-1)B_i)}{1+(m-1)\chi/i} \\
&= \frac{\Gamma\big(k+1+(m-1)\chi\big)\Gamma(j)}{\Gamma\big(j+(m-1)\chi\big)\Gamma(k+1)} \prod_{i=j}^{k} \Big(1+O\big(i^{-2}\big)\Big) \\
&= \Big(\frac{k}{j}\Big)^{(m-1)\chi} \big(1+O(j^{-1})\big),
\end{aligned} \tag{12.8}$$

as in (3.8), and, recalling that B_i are independent and taking $j=1$ in (12.8),

$$\begin{aligned}
\mathbb{E}\, \Phi_k &= \prod_{i=1}^{k} \mathbb{E}(1+(m-1)B_i) = \frac{\Gamma\big(k+1+(m-1)\chi\big)}{\Gamma\big(1+(m-1)\chi\big)\Gamma(k+1)} \prod_{i=1}^{k} \frac{\mathbb{E}(1+(m-1)B_i)}{1+(m-1)\chi/i} \\
&= Qk^{(m-1)\chi}\big(1+O(k^{-1})\big),
\end{aligned} \tag{12.9}$$

where

$$Q := \frac{\prod_{i=1}^{\infty} \frac{\mathbb{E}(1+(m-1)B_i)}{1+(m-1)\chi/i}}{\Gamma\big(1+(m-1)\chi\big)}; \tag{12.10}$$

note that the infinite product in (12.10) converges as a consequence of (12.4).

Using (12.4) and (12.5), the upper bounds (3.10) and (3.11) follow by the same proof as before. The statements in Lemmas 3.3 and 3.4 hold exactly, except for (3.18), which in view of (12.9), is now replaced with

$$\tilde{\beta} := Q\beta. \tag{12.11}$$

Let Y_k, Z_k, J_k, W_k be as in Section 4. To streamline the arguments, from here onwards we concentrate on the $m = 2$, $\rho = 0$ case, and leave the general case (with modifications as in Section 10) to the reader. Once we have sampled the self-loops at every vertex, the stochastic recursions for obtaining D_n are similar to the ones in Section 4.1: we sample $(B_i)_{i=1}^{n-1}$ according to (12.2), and for each red vertex k, we add $2 - N_k$ outgoing edges and proceed as before. The boundary conditions are the same, except now we have $Y_{n-1} = 2 - N_n$. For $2 \leqslant k \leqslant n-1$, the recursion takes the form

$$Y_{k-1} = Y_k - Z_k + (2 - N_k)J_k, \tag{12.12}$$

and because $0 \leqslant N_k \leqslant 2$, we also have

$$Y_{k-1} \leqslant Y_k - Z_k + 2J_k. \tag{12.13}$$

Let $\mathcal{F}_k$ be the σ-algebra generated by $(N_i)_{i=2}^n$, $(B_i)_{i=2}^{n-1}$ and the coin tosses at vertices $n-1, \dots, k+1$ in the stochastic recursion. Note that (4.3) holds, and in view of (12.13), also (4.5)–(4.12) and (4.15)–(4.18) hold, with the number 2 in (4.10) and (4.18) replaced with $2 - N_n$, and with the last equality in (4.17) replaced with $\leqslant$. Instead of (4.14), from (12.12) we have

$$\begin{aligned} A_{k-1} - A_k &= W_k - \mathbb{E}(W_{k-1} \mid \mathcal{F}_k) \\ &= 2\Phi_{k-1}((1-B_k)^{Y_k} - 1 + B_k Y_k) + \Phi_{k-1} N_k \,\mathbb{E}(J_k \mid \mathcal{F}_k). \end{aligned} \tag{12.14}$$

Now, let $\mathbf{B}$ be the σ-field generated by $(B_i)_{i=2}^{n-1}$ and $(N_i)_{i=2}^n$. As the upper bounds in Section 3 still hold, and $(B_i)_{i=2}^{n-1}$ are independent, Lemmas 4.1 and 4.2 hold. The probability that vertex $k \geqslant 2$ is red and has at least one self-loop is

$$\mathbb{P}(Z_k \geqslant 1, N_k \geqslant 1) = \mathbb{E}\big[\mathbf{1}\{N_k \geqslant 1\} \mathbb{P}_{\mathbf{B}}(Z_k \geqslant 1)\big]; \tag{12.15}$$

and so by Markov's inequality and (4.30),

$$\mathbb{P}(Z_k \geqslant 1, N_k \geqslant 1) \leqslant 2\,\mathbb{E}\Big(N_k B_k \prod_{i=k+1}^{n-1} (1 + B_i)\Big). \tag{12.16}$$

By the independence of the pairs (B_i, N_i), (12.6), and (3.8), this yields

$$\mathbb{P}(Z_k \geqslant 1, N_k \geqslant 1) \leqslant 2\,\mathbb{E}(N_k B_k) \prod_{i=k+1}^{n-1} \mathbb{E}(1 + B_i) \leqslant C \frac{n^{1/2}}{k^{5/2}}. \tag{12.17}$$

In view of (12.14) and Markov's inequality, (4.35) in Lemma 4.3 is replaced with

$$A_{k-1} - A_k \leqslant (W_k B_k)^2 \Phi_k^{-1} + \Phi_k N_k B_k Y_k = (W_k B_k)^2 \Phi_k^{-1} + N_k B_k W_k. \tag{12.18}$$

Using (4.17), we have

$$\mathbb{E}_{\mathbf{B}}[N_k B_k W_k] = N_k B_k \,\mathbb{E}_{\mathbf{B}}[W_k] \leqslant 2 N_k B_k \Phi_{n-1}. \tag{12.19}$$

Since the pairs (B_i, N_i) are independent, it follows from (12.19) and (3.5) that

$$\mathbb{E}[N_k B_k W_k] \leqslant 2\,\mathbb{E}\big[N_k B_k (1 + B_k)\big] \prod_{\substack{i=1 \\ i \neq k}}^{n-1} \mathbb{E}(1 + B_i) \leqslant 4\,\mathbb{E}[N_k B_k]\,\mathbb{E}\,\Phi_{n-1}, \tag{12.20}$$

and applying (12.6) and (12.9), we get

$$\mathbb{E}\big[N_k B_k W_k\big] \leqslant C \frac{n^{1/2}}{k^2} \leqslant C \frac{n}{k^{5/2}}. \tag{12.21}$$

With (12.18) and (12.21), we may proceed as in the proof of Lemma 4.3 to show that (4.36) holds.

Lemma 5.1 follows from Lemma 3.4. Thus, the early part of the growth of D_n can be coupled to the same time-changed Yule process $\widehat{\mathcal{Y}}$ with some extra modifications. Recall that $\Psi(x)$ is the mapping of vertex x in $\widehat{\mathcal{Y}}$ to a vertex k in D_n (or vertex $(k/n)^\chi$ in $\widehat{D}_n$).

In Step (1) of the coupling, we sample $(N_i)_{i=1}^n$ and then $(B_i)_{i=1}^{n-1}$ as in (12.4). If Ψ maps x to some k that has at least one self-loop, we extend Ψ in Section 5 by mapping all children of x to k (so all other descendants of x are also mapped to k). To prove that Theorem 5.2 also holds in this case, we need to show that the extended mapping above is w.h.p. injective at every vertex in $\widehat{D}_n \cap [(n_1/n)^\chi, 1]$. By (12.17) and (12.3), the probability that a vertex in $\widehat{D}_n \cap [(n_1/n)^\chi, 1]$ has at least one self-loop is at most

$$\mathbb{P}(N_n \geqslant 1) + \sum_{k=n_1}^{n-1} \mathbb{P}(Z_k \geqslant 1, N_k \geqslant 1) \leqslant \frac{C}{n} + \sum_{k=n_1}^{n-1} \frac{Cn^{1/2}}{k^{5/2}} = O(\log^{3/2} n/n) = o(1). \tag{12.22}$$

The same argument as in Step 1 in the proof of Theorem 5.2 and (12.22) then give the desired claim. The remaining steps of the proof can be applied without any changes.

Lemma 6.1, Lemma 6.2 and Theorem 6.3 hold with the same proofs as before, since we have shown that (4.36) and the various other estimates that we use there still hold.

Let $\widehat{A}_t^{(n)}$ be as in (7.1). When proving tightness of $\widehat{A}_t^{(n)}$ in $C[a,b]$ for $0 < a < b < \infty$, we have to use (12.18) instead of (4.35). Let V_k, T_k, $\widehat{V}_t^{(n)}$, $\widehat{T}_t^{(n)}$, $\mathfrak{M}_n$, and $\widehat{\Psi}_n$ be as in (7.8), (7.9) and (7.17). Using (12.18) we obtain instead of (7.18), using the crude bound $N_k \leqslant m$,

$$\begin{aligned} |A_k - A_{k-1}| &\leqslant M_k^2 \Phi_k^{-1}(V_k - V_{k-1}) + mM_k(T_k - T_{k-1}) \\ &\leqslant n^{5/6}\mathfrak{M}_n^2 \widehat{\Psi}_n (V_k - V_{k-1}) + mn^{1/2}\mathfrak{M}_n(T_k - T_{k-1}) \end{aligned} \tag{12.23}$$

and thus, arguing as for (7.20), for real numbers s, t such that $a \leqslant s \leqslant t$,

$$|\widehat{A}_t^{(n)} - \widehat{A}_s^{(n)}| \leqslant \widehat{\Psi}_n \mathfrak{M}_n^2 (\widehat{V}_t^{(n)} - \widehat{V}_s^{(n)}) + m\mathfrak{M}_n(\widehat{T}_t^{(n)} - \widehat{T}_s^{(n)}). \tag{12.24}$$

We have already shown in Section 7 that the processes $\widehat{V}_t^{(n)} - \widehat{V}_a^{(n)}$ and $\widehat{T}_t^{(n)} - \widehat{T}_a^{(n)}$ are tight in $C[a,b]$ (Lemma 7.3) and that the sequences $(\mathfrak{M}_n)_{n=1}^\infty$ and $(\widehat{\Psi}_n)_{n=1}^\infty$ are tight. Hence, by simple applications of Lemma 7.2, the processes $\widehat{\Psi}_n \mathfrak{M}_n^2 (\widehat{V}_t^{(n)} - \widehat{V}_a^{(n)})$ and $m\mathfrak{M}_n(\widehat{T}_t^{(n)} - \widehat{T}_a^{(n)})$, $n \geqslant 1$, are tight in $C[a,b]$. If $(X_n(t))_{n=1}^\infty$ and $(Y_n(t))_{n=1}^\infty$ are any two sequences of random continuous functions on $[a,b]$ that both are tight in $C[a,b]$, then so is the sequence $((X_n(t)+Y_n(t)))_{n=1}^\infty$. Hence, the sequence $\widehat{\Psi}_n \mathfrak{M}_n^2 (\widehat{V}_t^{(n)} - \widehat{V}_a^{(n)}) + m\mathfrak{M}_n(\widehat{T}_t^{(n)} - \widehat{T}_a^{(n)})$, $n \geqslant 1$, is tight in $C[a,b]$; finally (12.24) and another application of Lemma 7.2 (now with $Z_n = 1$) show that $\widehat{A}_t^{(n)}$, $n \geqslant 1$, are tight in $C[a,b]$, so Lemma 7.1 still holds.

Some minor adjustments are also required to yield the same result as in Theorem 8.1. When applying the Skorohod coupling theorem, $N_i^{(n)}$, $n \geqslant 1$, are potentially different for each n. Let $0 < s < t$ and define $k := \lfloor sn^{1/3} \rfloor$ and $\ell := \lfloor tn^{1/3} \rfloor$. By (7.1) and (12.14),

$$\begin{aligned} \widehat{A}_s^{(n)} - \widehat{A}_t^{(n)} = n^{-1/2} \sum_{i=k+1}^{\ell} 2\Phi_{i-1}\big[(1-B_i)^{Y_i} - 1 + Y_i B_i\big] \\ + n^{-1/2} \sum_{i=k+1}^{\ell} \Phi_{i-1} N_i \, \mathbb{E}(J_i \mid \mathcal{F}_i) + o(1). \end{aligned} \tag{12.25}$$

However, by Markov's inequality and (12.21),

$$\begin{aligned} \mathbb{E} \sum_{i=k+1}^{\ell} \Phi_{i-1} N_i \, \mathbb{E}(J_i \mid \mathcal{F}_i) &\leqslant \mathbb{E} \sum_{i=k+1}^{\ell} \Phi_{i-1} N_i B_i Y_i \leqslant \mathbb{E} \sum_{i=k+1}^{\ell} N_i B_i W_i \\ &\leqslant \sum_{i=k+1}^{\ell} C\frac{n^{1/2}}{i^2} \leqslant C\frac{n^{1/2}}{k} \leqslant C_s n^{1/6}, \end{aligned} \tag{12.26}$$

implying that

$$n^{-1/2} \sum_{i=k+1}^{\ell} \Phi_{i-1} N_i \, \mathbb{E}(J_i \mid \mathcal{F}_i) \xrightarrow{\mathrm{p}} 0. \tag{12.27}$$

The remainder of the proof of Theorem 8.1 is then the same as before.

Proof of Theorem 12.2. With the preparations above, the same argument as in Section 9 yields Theorem 9.1 for this model too; with modifications as in Section 10 we obtain Theorem 1.2. Similarly, the arguments in Section 11 still hold, and thus Theorem 1.3 holds. □

Appendix A. The differential equations in (8.21) and (10.37)

We rewrite the equation in (10.37) as

$$f'(t) = mt^{\alpha-1}\Big(\big(1+\tfrac{1}{\theta}t^{-\alpha}f(t)\big)^{-(m+\rho)} - 1 + \chi t^{-\alpha} f(t)\Big). \tag{A.1}$$

where, as above,

$$\alpha := 1 + (m-1)\chi. \tag{A.2}$$

Note that in the special case $m = 2$ and $\rho = 0$, we have $\chi = 1/2$, $\theta = 4$, and $\alpha = 3/2$, so the above yields the differential equation in (8.21). We define

$$g(t) := \theta^{-1} t^{-\alpha} f(t) \tag{A.3}$$

so that (A.1) simplifies to, recalling $\chi\theta = m + \rho$, see (3.2),

$$g'(t) = -\frac{\alpha}{t} g(t) + \frac{m}{\theta t}\Big((1+g(t))^{-(m+\rho)} - 1 + (m+\rho)g(t)\Big). \tag{A.4}$$

Letting

$$h(x) := g(e^{(m+\rho)x}) \tag{A.5}$$

then yields, using (A.2),

$$\begin{aligned} h'(x) &= -(m+\rho)(1+(m-1)\chi)h(x) + \chi m\big((1+h(x))^{-(m+\rho)} - 1 + (m+\rho)h(x)\big) \\ &= \chi m(1+h(x))^{-(m+\rho)} - \chi m + (m+\rho)(\chi-1)h(x) \\ &= \chi m\big((1+h(x))^{-(m+\rho)} - 1 - h(x)\big), \end{aligned} \tag{A.6}$$

where the last equality follows from $(m+\rho)(1-\chi) = \chi m$, see again (3.2). The autonomous differential equation in (A.6) can be integrated to

$$\frac{1}{\chi m} \int \frac{1}{(1+h)^{-(m+\rho)} - 1 - h} \, \mathrm{d}h = \int 1 \, \mathrm{d}x \tag{A.7}$$

Furthermore, with $v := 1 + h$,

$$\begin{aligned} \frac{1}{\chi m} \int \frac{1}{(1+h)^{-(m+\rho)} - 1 - h} \, \mathrm{d}h &= \frac{1}{\chi m} \int \frac{(1+h)^{m+\rho}}{1 - (1+h)^{m+\rho+1}} \, \mathrm{d}h \\ &= -\frac{1}{\chi m} \int \frac{v^{m+\rho}}{v^{m+\rho+1} - 1} \, \mathrm{d}v. \end{aligned} \tag{A.8}$$

The change of variable $u = v^{m+\rho+1} - 1$ then gives

$$\frac{1}{\chi m} \int \frac{v^{m+\rho}}{v^{m+\rho+1} - 1} \, \mathrm{d}v = \frac{1}{\chi m(m+\rho+1)} \int \frac{1}{u} \, \mathrm{d}u = \frac{1}{\chi m(m+\rho+1)} \log u + C. \tag{A.9}$$

Thus, reverting back to the original variable h, (A.7) is equivalent to

$$-\frac{1}{\chi m(m+\rho+1)} \log\big((1+h)^{m+\rho+1} - 1\big) = x + C, \tag{A.10}$$

which yields the solution

$$h(x) = \bigl(1 + ce^{-\chi m(m+\rho+1)x}\bigr)^{\frac{1}{m+\rho+1}} - 1, \qquad \text{for some } c \in \mathbb{R}. \tag{A.11}$$

From (A.3) and (A.5),

$$f(t) = \theta t^{1+(m-1)\chi} h\bigl(\tfrac{1}{m+\rho} \log t\bigr). \tag{A.12}$$

so plugging in (A.11) into (A.12), and using

$$\alpha = 1 + (m-1)\chi = 1 + \frac{(m-1)(m+\rho)}{2m+\rho} = \frac{\chi m(m+\rho+1)}{m+\rho}, \tag{A.13}$$

we get

$$f(t) = \theta t^{\alpha} \Bigl(\bigl(1 + ct^{-\alpha}\bigr)^{\frac{1}{m+\rho+1}} - 1\Bigr). \tag{A.14}$$

Using L'Hôpital's rule (or a Taylor expansion) and (A.14), we obtain

$$f(\infty) := \lim_{t\to\infty} f(t) = \frac{\theta c}{m+\rho+1} \lim_{t\to\infty} \bigl(1 + ct^{-\alpha}\bigr)^{\frac{1}{m+\rho+1}-1} = \frac{\theta c}{m+\rho+1}. \tag{A.15}$$

Hence, the unique solution f to (A.1) with a given $f(\infty)$ is given by (A.14) with

$$c = \frac{m+\rho+1}{\theta} f(\infty). \tag{A.16}$$

Appendix B. A beta integral

Recall the standard beta integral [19, 5.12.3]

$$\int_0^{\infty} \frac{x^{a-1}}{(1+x)^b} \,\mathrm{d}x = \frac{\Gamma(a)\Gamma(b-a)}{\Gamma(b)} \tag{B.1}$$

when $0 < \Re a < \Re b$. We use the following less well-known extension; it is not new but we give a proof for completeness.

Lemma B.1. *If* $-1 < \Re a < 0$ *and* $\Re b > 0$*, then*

$$\int_0^{\infty} \Bigl(\frac{1}{(1+x)^b} - 1\Bigr) x^{a-1} \,\mathrm{d}x = \frac{\Gamma(a)\Gamma(b-a)}{\Gamma(b)}. \tag{B.2}$$

Proof. We consider a more general integral. Assume first $\Re a > 0$ and $\Re b > 0$, and let $\Re c > \Re a$. Then, by using (B.1) twice,

$$\int_0^{\infty} \Bigl(\frac{1}{(1+x)^{b+c}} - \frac{1}{(1+x)^c}\Bigr) x^{a-1} \,\mathrm{d}x = \frac{\Gamma(a)\Gamma(b+c-a)}{\Gamma(b+c)} - \frac{\Gamma(a)\Gamma(c-a)}{\Gamma(c)}. \tag{B.3}$$

For fixed b and c with $\Re b, \Re c > 0$, the left-hand side converges for $-1 < \Re a < \Re c$, and defines an analytic function of a in this strip. Hence, by analytic continuation, (B.3) holds throughout this range. Similarly, if $\Re a > -1$ and $\Re b > 0$, then the left-hand side of (B.3) is an analytic function of c in the domain $\Re c > \Re a$, and thus (B.3) holds whenever $-1 < \Re a < \Re c$ and $\Re b > 0$.

For $-1 < \Re a < 0$ we thus may take $c = 0$ in (B.3) which yields (B.2). (Recall that $1/\Gamma(0) = 0$.) □

Remark B.2. Note that (B.1) and (B.2) give the same formula, but for different ranges of a. The integrals can be interpreted as the Mellin transforms of $(1+x)^{-b}$ and $(1+x)^{-b} - 1$, respectively, and thus this is an instance of a general phenomenon when considering the Mellin transforms of a function $f(x)$ and of the difference $f(x) - p(x)$ where, for example, $p(x)$ is a finite Taylor polynomial at 0, see [9, p. 19].

References

[1] Krishna B. Athreya and Peter E. Ney: *Branching Processes.* Springer-Verlag, Berlin, 1972.
[2] Albert-László Barabási and Reka Albert: Emergence of scaling in random networks. *Science* **286** (1999), no. 5439, pp. 509–512.
[3] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi: Asymptotic behavior and distributional limits of preferential attachment graphs. *Ann. Probab.* **42** (2014), no. 1, pp. 1–40.
[4] Patrick Billingsley: *Convergence of Probability Measures.* Wiley, New York, 1968.
[5] Béla Bollobás and Oliver Riordan: The diameter of a scale-free random graph. *Combinatorica* **24** (2004), no. 1, pp. 5–34.
[6] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády: The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** (2001), no. 3, pp. 279–290.
[7] Donald L. Burkholder: Distribution function inequalities for martingales. *Ann. Probab.* **1** (1973), no. 1, pp. 19—42.
[8] Robert P. Dobrow and Robert T. Smythe: Poisson approximations for functionals of random trees. *Random Structures Algorithms* **9** (1996), no. 1-2, 79–92.
[9] Philippe Flajolet, Xavier Gourdon, and Philippe Dumas: Mellin transforms and asymptotics: harmonic sums. *Theoret. Comput. Sci.*, **144** (1995), no. 1-2, pp. 3–58.
[10] Allan Gut: *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.
[11] Remco van der Hofstad: *Random Graphs and Complex Networks, Vol. 1*, Cambridge Univ. Press, Cambridge, 2017.
[12] Remco van der Hofstad: *Random Graphs and Complex Networks, Vol. 2*, Cambridge Univ. Press, Cambridge, 2024.
[13] Svante Janson: The number of descendants in a random directed acyclic graph. *Random Structures Algorithms* **64** (2024), no. 3, pp. 768–803.
[14] Olav Kallenberg: *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002.
[15] Donald E. Knuth: *The Art of Computer Programming*, Section 7.2.2.3. Preliminary draft, 29 January 2023. (To appear in Volume 4, Fascicle 7.)
[16] Markus Kuba and Stephan Wagner: On the distribution of depths in increasing trees. *Electron. J. Combin.* **17** (2010), no. 1, Research Paper 137, 9 pp.
[17] Tiffany Y. Y. Lo: Local weak limit of preferential attachment random trees with additive fitness. *Advances in Applied Probability* **56** (2024), no. 3, pp. 785–824.
[18] Tamás F. Móri: The maximum degree of the Bárabási-Albert random tree. *Combinatorics, Probability and Computing* **14** (2005), no. 3, pp. 339–348
[19] *NIST Handbook of Mathematical Functions.* Edited by Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert and Charles W. Clark. Cambridge Univ. Press, 2010.
Also available as *NIST Digital Library of Mathematical Functions*, http://dlmf.nist.gov/
[20] Alois Panholzer and Helmut Prodinger: Level of nodes in increasing trees revisited. *Random Structures Algorithms* **31** (2007), no. 2, 203–226
[21] Erol Peköz, Adrian Röllin and Nathan Ross: Joint degree distributions of preferential attachment random graphs. *Advances in Applied Probability* **49** (2017), no. 2, pp. 368–387
[22] Ward Whitt: Weak convergence of probability measures on the function space $C[0,\infty)$. *Annals of Mathematical Statistics* **41** (1970), no. 3, pp. 939–944.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: svante.janson@math.uu.se

Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden

Email address: tiffany.y.y.lo@math.su.se