Internet Mathematics Vol. 2, No. 2: 121-163

Connectivity Transitions in Networks with Super-Linear Preferential Attachment

Roberto Oliveira and Joel Spencer

Abstract. We analyze an evolving network model of Krapivsky and Redner in which new nodes arrive sequentially, each connecting to a previously existing node b with probability proportional to the pth power of the in-degree of b. We restrict to the super-linear case $p > 1$. When $1 + \frac{1}{k} < p < 1 + \frac{1}{k-1}$, the structure of the final countable tree is determined. There is a finite tree T with distinguished v (which has a limiting distribution) on which is "glued" a specific infinite tree; v has an infinite number of children, an infinite number of which have $k - 1$ children, and there are only a finite number of nodes (possibly only v) with k or more children. Our basic technique is to embed the discrete process in a continuous time process using exponential random variables, a technique that has previously been employed in the study of balls-in-bins processes with feedback.

1. Introduction

In some important examples of growing networks, such as the World Wide Web or the scientific citation network, one can interpret the fact that a given node has high in-degree as indicative that the node is "popular." For instance, popular papers are the ones more often cited by other works, and popular web pages receive more links than less popular ones. A consequence of differences in popularity is that a node with high in-degree has more propensity to receive further edges as the network evolves than an unpopular node with low in-degree. In other words, the more popular a node is, the more visible it is to the community

1542-7951/05 $0.50 per page

that creates the network and/or interacts through it, and high visibility makes future increases in popularity more likely.

Barabási and Albert [Barabási and Albert 99] incorporated this so-called *preferential attachment* phenomenon into a generative model for these and other networks. In this model, nodes arrive at the network one at a time and direct a fixed number m of edges to previously existing nodes that are chosen with probabilities proportional to their in-degrees. It is quite remarkable that this simple model already replicates many nontrivial features of the above networks, such as power-law degree distributions, small diameter, and high resistance to random failures, as argued nonrigorously by physicists (see [Albert et al. 99, Albert and Barabási 02] and references therein) and later proven rigorously by mathematicians [Bollobás et al. 00, Bollobás and Riordan 03b, Bollobás and Riordan 04, Bollobás and Riordan 03a].

The success of the Barabási-Albert model has also inspired many different variants. The models in [Cooper and Frieze 03, Bollobás et al. 03] permit that the power-law exponent of the degree distribution be adjusted to fit real-world data. Other models [Bianconi and Barabási 01] feature a preferential attachment that is dictated both by node fitness and popularity. This work is dedicated to yet another kind of variant of the model of [Barabási and Albert 99], one in which the *strength* of preferential attachment can be varied.

This model was proposed and studied by Krapivsky and Redner [Krapivsky and Redner 01] and independently by Drinea, Mitzenmacher, and Enachescu [Drinea et al. 01]. It differs from the Barabási-Albert network in that each incoming node chooses a preexisting vertex to which to link with probability proportional to a fixed function f (the *attachment kernel*) of the degree of that vertex.[1] While the Barabási-Albert model is recovered by setting $f(x) = x$, we will be mostly concerned with kernels of the form $f(x) \sim x^p$ with $p > 1$ thought of as a tunable parameter; this is referred to in [Krapivsky and Redner 01] as the *super-linear case*. One of the many remarkable nonrigorous results about this so-called GN (Growing Network) model is that it undergoes an infinite sequence of *connectivity transitions* at $p = p_k \equiv 1+1/k$, $k = 1, 2, 3, \ldots$. By this it is meant that for $p > p_k$ the GN process has only finitely many vertices that receive more than k links, whereas for $p \leq p_k$ the number of such vertices is infinite. Another way of stating this property is the following: the smallest integer k for which $p > p_k = 1 + 1/k$ is also the smallest number k for which only finitely many nodes ever reach in-degree k.

[1] As in the original Barabási-Albert model, one could also consider a similar model in which each incoming node creates a fixed number m of new edges, but we will only consider the case $m = 1$ in this paper.

The connectivity transitions are both mathematically intriguing and physically interesting. The fact that the $p > 1$, $p = 1$, and (conjecturally) $p < 1$ cases of the model are very different leads the authors of [Dorogovtsev and Mendes 02] to suggest that so-called *self-organized criticality* is at work in networks with power-law degree distributions (the $p = 1$ case). It was also noted elsewhere [Albert and Barabási 02] that the condensation regime of the fitness model of Bianconi and Barabási [Bianconi and Barabási 01] has qualitatively similar behavior to the super-linear GN; [Albert and Barabási 02] even suggests that a direct connection between the two models could exist. There is also some modeling interest in the connectivity transitions, since networks in which preferential attachment is very strong (conceivably even some parts of the World Wide Web) should exhibit behavior that is qualitatively similar to the GN model in the super-linear regime.

Despite the striking characteristics, we do not know of any rigorous work on the GN model to the present date. A modified model was addressed in independent work by Chung, Handjani, and Jungreis [Chung et al. 03]. In their process, an attachment kernel is still present, but at each time step either a new vertex and a new edge are added with probability $0 < q < 1$ or only a new edge is added with probability $1-q$. This modified model does exhibit connectivity transitions in the sense of [Krapivsky and Redner 01], but it is not clear how to deduce the analogous results for the original GN model from the techniques in [Chung et al. 03].

In this paper we attempt to give a rigorous description of the super-linear GN process in the large-time limit. Our rigorous results imply the existence of connectivity transitions, but they also go beyond that. The first result that we prove is the following.

Theorem 1.1. *Let $\{T_m\}_{m\geq 1}$ be the* GN *process with attachment kernel $f(x) = (x+1)^p$ (defined in Section 3). Also let T_∞ be the increasing limit of the $\{T_m\}_{m\geq 1}$ process, and assume that $p > p_k = 1 + 1/k$. Then with probability 1 all but finitely many nodes of T_∞ have less than k descendants (see the definition in Section 2.2).*

A vertex of T_∞ with in-degree larger than or equal to k necessarily has at least k descendants. For this reason, Theorem 1.1 implies that, for $p > p_k$, only finitely many vertices of T_∞ have in-degree $\geq k$. As a result, the number of vertices in T_m with in-degree bigger than k is bounded as $m \to +\infty$. This shows that Theorem 1.1 implies the nonrigorous "$p > p_k$" result of [Krapivsky and Redner 01] and is in fact stronger than it. Similarly, Theorem 1.2 implies the $p \leq p_k$ case of Krapivsky and Redner's result.

Theorem 1.2. *Let f, $p > 1$, and T_∞ be as in Theorem 1.1, and let $k = k_p$ be the smallest positive integer for which $p > p_k = 1 + 1/k$. Consider the construction* $\textsc{Glue}(T, v, k)$ *defined in Section 2.2. Then the set of values (up to isomorphism) that T_∞ attains with positive probability is precisely the set of all trees that can be obtained by choosing a finite rooted tree T and a distinguished vertex $v \in T$ and setting* $T_\infty = \textsc{Glue}(T, v, k)$.

Theorem 1.2 completely describes (up to isomorphisms) the limit set of the GN process in the large-time limit. In particular, it also implies that if $p \leq p_k$, the number of vertices of in-degree $\geq k$ in T_m diverges as $m \to +\infty$. This differs from the original claim in [Krapivsky and Redner 01], in which the authors argue that the expectation of the number of vertices of degree $\ell \geq k$ diverges at a certain rate. While we have nothing to say about this rate, Theorem 1.2 is stronger than the claim of [Krapivsky and Redner 01] in that divergence of the expected number is implied but almost sure divergence is not implied. Moreover, our description of the structure of T_∞ is new. Finally, we note that there is nothing special about the choice of $f(x) = (x+1)^p$ as our super-linear kernel. In fact, the proof of both theorems will make it clear that it suffices to assume that $f(x) > 0$ for all x and that $f(x) = \Theta(x^p)$ for $x \gg 1$, with only minor modifications in our arguments.

We now briefly outline our proof techniques. On a high level, we rely strongly on the similarity pointed out by Drinea, Frieze, and Mitzenmacher [Drinea et al. 02] between the GN process and *balls-in-bins models with feedback*. The latter model describes the evolution of a system with a fixed number of bins at which balls are thrown. A ball arrives at each discrete time step and chooses a bin to go into with probability proportional to a fixed function f (that we call the feedback function) of the number of balls currently in that bin. This model can also be viewed as a static variant of the GN process in which new edges are repeatedly added but without the creation of any new nodes/bins. This analogy permits that a certain technique applied to the study of balls-in-bins problems [Khanin and Khanin 01, Mitzenmacher et al. 04, Spencer and Wormald 04] is adapted to the GN process. It consists of building a continuous-time process out of exponential random variables and of showing that it embeds the original discrete-time process. For this reason we call this construction the *exponential embedding*. The GN version of the exponential embedding is essential to the construction and analysis of the infinite tree limit T_∞, and we view it as an important part of our paper's contribution.

The remainder of the paper is organized as follows. In Section 2 we introduce our notation and review a few basic concepts. We formally define the GN process in Section 3, starting with its original definition in [Krapivsky and Redner 01,

Drinea et al. 01] and then describing a useful labeled version of it. Section 4 introduces the exponential embedding technique. We begin with a review of the simpler balls-in-bins case, then we move on to the construction of the embedding of the GN process for general attachment kernels. We then employ the embedding to show that so-called "explosive kernels" give rise to GN processes for which $T_m \to T_\infty$ in finite time under the exponential embedding. This section ends with some lemmas on sums of exponential random variables that will be useful later on. Theorem 1.1 is proven in the subsequent Section 5. The section starts with weaker results that intuitively pave the way for the actual proof of the theorem, which relies on a careful consideration of the time of the birth of the kth descendant of a given node in the exponential embedding setting. In Section 6 we prove Theorem 1.2, relying on Theorem 1.1 and on the techniques developed in the previous sections. We discuss some consequences of our main theorems and some related open questions in the Conclusion (Section 7). The appendix contains the proofs of some technical results.

2. Preliminaries

2.1. Probabilistic Ingredients

We briefly remind the reader of some basic probabilistic concepts and tools, while also fixing some notation.

2.1.1. Distributions. We say that two random variables X, Y taking values on the same set U *have the same distribution* (or *are identical in law*) if for all measurable subsets $A \subseteq U$ $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$. This will be symbolically represented by $X =^d Y$.

2.1.2. The exponential distribution. A random variable X is said to be *exponentially-distributed with rate* $\lambda > 0$ if X almost surely takes values on the positive reals and

$$\mathbb{P}(X > t) = e^{-\lambda t},\ t \geq 0.$$

We denote this property by $X =^d \exp(\lambda)$. The shorthand $\exp(\lambda)$ will also denote a generic exponentially-distributed random variable with rate λ. We list below some elementary but extremely useful properties of those random variables.

1. *Lack of memory.* Let $X =^d \exp(\lambda)$ and $Z \geq 0$ be independent from X. The distribution of $X - Z$ conditioned on $X > Z$ is still equal to $\exp(\lambda)$.

2. *Minimum property.* Let $\{X_i =^d \exp(\lambda_i)\}_{i=1}^m$ be independent. Then $X_{\min} \equiv \min_{1\le i\le m} X_i =^d \exp(\lambda_1 + \lambda_2 + \ldots \lambda_m)$. Moreover, for all $1 \le i \le m$,

$$\mathbb{P}(X_i = X_{\min}) = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \ldots \lambda_m} .$$

3. *Multiplication property.* If $X =^d \exp(\lambda)$ and $\eta > 0$ is a fixed number, $\eta X =^d \exp(\lambda/\eta)$.

2.1.3. The Borel-Cantelli Lemma. Let $\{A_n\}_{n\in N}$ be a sequence of events in some fixed probability space, with N a countable set. The event "A_n *infinitely often* $(n \in N)$" (or "A_n *i.o.* $(n \in N)$") contains all outcomes that belong to an infinite number of the events A_n. The *Borel-Cantelli Lemma* states that

$$\sum_{n\in N} \mathbb{P}(A_n) < +\infty \Rightarrow \mathbb{P}(A_n \text{ i.o. } (n \in N)) = 0$$

and

$$\sum_{n\in N} \mathbb{P}(A_n) = +\infty \text{ and } \{A_n\}_{n\in N} \text{ independent } \Rightarrow \mathbb{P}(A_n \text{ i.o. } (n \in N)) = 1 .$$

2.1.4. Discrete-time Markov chains. A (discrete-time) Markov chain on the countable set Ω is specified by *transition probabilities* $\Pi : \Omega \times \Omega \to [0, 1]$ and a *initial condition* $X_0 \in \Omega$ (possibly nondeterministic). The recipe

$$\mathbb{P}(\forall 0 \le i \le t \;\; X_i = \omega_i) = \mathbb{P}(X_0 = \omega_0) \prod_{i=1}^{t} \Pi(\omega_{i-1}, \omega_i)$$

defines the distribution of a sequence $\{X_i\}_{i=0}^{+\infty}$ of Ω-valued random variables.

2.2. Tree Terminology

2.2.1. Trees. All trees are rooted and have their edges directed towards the root. No loops or parallel edges are allowed. Given vertices a, b in a tree T, the existence of the oriented edge (a, b) will be indicated by saying that *a is a child of b*, or that *b is a's parent*, or that *a links to b*. With this terminology, the (in-)degree $\mathrm{d}_T(b)$ of b in T is the number of its children. If r is a node of T, the subtree T_r of T rooted at r is the tree with root r, together with r's children, the children of those children, and so on. The nodes in $T_r \backslash \{r\}$ are referred to as the descendants of r, and r is said to be *k-fertile* in T if it has k or more descendants. Theorem 1.1 consists of showing that for $p > p_k$, only finitely many nodes in T_∞ are k-fertile.

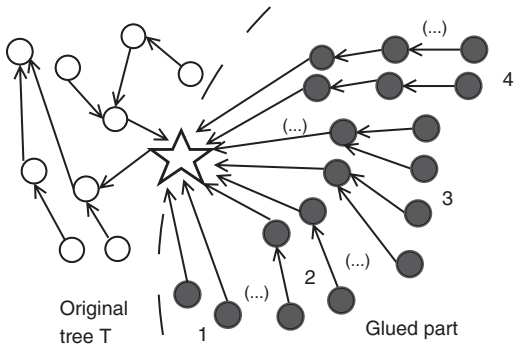

Figure 1. An example of $\textsc{Glue}(\mathrm{T}, v, k)$ for $k = 3$.

2.2.2. The Glue construction. Given a finite (rooted, oriented) tree T, a distinguished node v of T and an integer $k \geq 1$, we define $\textsc{Glue}(\mathrm{T}, v, k)$ as follows. For each finite (rooted, oriented) tree S on k or less nodes, take countably many copies $\{S_i\}_{i\geq 1}$ of S. $\textsc{Glue}(\mathrm{T}, v, k)$ is the union of T with all the trees S_i as above, with the addition of edges from the root of each one of the S_i to v.

As a simple example suppose T consists of a single node (the root v) and $k = 2$. Then in $\textsc{Glue}(\mathrm{T}, v, k)$ the root has a countably infinite number of children. Infinitely many of these children are childless and infinitely many of these children have precisely one child and none of them have more than one child. Further, all grandchildren of the root are childless.

A more complex example of $\textsc{Glue}(\mathrm{T}, v, k)$, now with $k = 3$, is portrayed in Figure 1. The starred node is v, and the finite tree T lies to the left of the dashed line. The countably many copies of the four rooted trees on three or less vertices (numbered 1, 2, 3, and 4 in the figure) appear to the right of the line and are all connected to v by their roots.

Our task in proving Theorem 1.2 will be to show that with probability 1 there exist T and v as above with $\mathrm{T}_\infty = \textsc{Glue}(\mathrm{T}, v, k)$ and that all such $\textsc{Glue}(\mathrm{T}, v, k)$ occur as values of T_∞ with some positive probability.

2.3. Labeled Trees and Parent-Closed Sets

2.3.1. Labels. It will be convenient for us to label the vertices of trees. For our purposes, a label is a (possibly empty) sequence of elements of the set $\mathbb{N} = \{1, 2, 3, \dots\}$ of positive integers. The empty sequence is denoted by ϵ, and all other sequences $a = \{a_i\}_{i=1}^m \subset \mathbb{N}$ (with $m \geq 1$) will be represented by $a = a_1 a_2 \dots a_m$. Moreover, we call $a_1 \dots a_{m-1}$ (the sequence a without its last element) the *parent sequence* of a. The set of all labels will be denoted by $\mathbb{N}^*$.

2.3.2. Labeling trees. A labeling of a finite tree T is an assignment of labels to the vertices of T that obeys two rules:

- the label of the root of T is the empty sequence ϵ;
- if vertex v has degree d and is labeled by the sequence $v_1 \dots v_m$, its children will receive labels $v_1 \dots v_m i$, $1 \leq i \leq d$.

The second rule implies that the label of a vertex v's parent in T is the parent sequence of the label of v.

2.3.3. Parent-closed subsets. A subset $A \subset \mathbb{N}^*$ is said to be *parent-closed* if it is nonempty and, for all nonempty sequences $a \in A$, the parent sequence of a is also in A. Any parent-closed A corresponds to a finite tree with vertex set A and edges from each $a \in A \backslash \{\epsilon\}$ to a's parent. Conversely, given a tree T, the labeling procedure above provides a proper set $A = A(T)$ that corresponds to tree T. This set $A(T)$ is not uniquely defined, but this will not keep us from representing finite trees by finite parent-closed $A \subset \mathbb{N}^*$ in what follows. For this reason, we will often apply tree terminology to parent-closed $A \subset \mathbb{N}^*$, speaking, for instance, of the degree $\mathrm{d}_A(a)$ of an element $a \in A$. We also observe that the potential descendants of $a = a_1 \dots a_m \in \mathbb{N}^*$ are obtained by adjoining the terms of another sequence $b = b_1 \dots b_n$ to a, thus forming the *concatenation* $ab \equiv a_1 \dots a_m b_1 \dots b_n$. Finally, we define for convenience

$$\mathcal{E}^{\mathrm{fin}} \equiv \{A \subset \mathbb{N}^* \ : \ A \text{ parent-closed and finite}\}.$$

3. Definition of the GN Process

3.1. The Standard Definition

The $\mathsf{GN}[f]$ process is defined in terms of an *attachment kernel*, that is, a function $f : \mathbb{N} \cup \{0\} \to \mathbb{R}^+$. The process evolves in discrete time $m = 0, 1, 2, 3, \dots$; its state at time $m \geq 0$ is a (rooted, oriented) tree $\mathrm{T}_m = (V_m, E_m)$ with vertex set V_m and edge set E_m. Initially, T_0 contains a single *root* node and no edges. At each time $m > 0$, the tree is updated by the addition of a new node ($V_m = V_{m-1} \cup \{v_m\}$) and a new edge ($E_m = E_{m-1} \cup \{v_m w_m\}$), where w_m is chosen according to the following probability distribution:

$$\forall w \in V_{m-1} \ \mathbb{P}\left(w_m = w \mid \mathrm{T}_{m-1}\right) = \frac{f(\mathrm{d}_{\mathrm{T}_{m-1}}(w))}{\sum_{v \in V_{m-1}} f(\mathrm{d}_{\mathrm{T}_{m-1}}(v))} \ .$$

These definitions already specify the process completely as a finite-tree-valued Markov chain.

3.2. The Labeled Process

In the labeled GN[f] process, we start by labeling the root (and unique element) of T_0 by the empty sequence ϵ. At subsequent times $m \geq 1$, assume that the incoming node v_m links to a node w_m, which is labeled by the sequence $a_1 \dots a_n$, and that v_m is the ℓth node to link to w_m. Then, the label of v_m is defined to be $a_1 \dots a_n \ell$, i.e., the sequence corresponding to v_m's parent w_m, with a new number ℓ added to it.

This recursive labeling obeys the definition of a labeling of a tree given in Section 2.3 and provides an alternative description of the process as a Markov chain on $\mathcal{E}^{\mathrm{fin}}$, as defined in Section 2.3. The transition probabilities of the GN process on $\mathcal{E}^{\mathrm{fin}}$ are

$$\begin{aligned} \Pi(A,B) &= \frac{f(\mathrm{d}_A(a))}{\sum_{b\in A} f(\mathrm{d}_A(b))}, \text{ if } \exists a \in A : B = A \cup \{a(\mathrm{d}_A(a)+1)\} \quad (3.1) \\ &= 0 \text{ otherwise,} \end{aligned}$$

and its initial state is $\mathrm{T}_0 = \{\epsilon\}$. We note in passing that the limit $\mathrm{T}_\infty = \bigcup_{m\geq 0} \mathrm{T}_m$ of the GN process takes values in the uncountable set

$$\mathcal{E} \equiv \{A \subset \mathbb{N}^* : A \text{ parent-closed and nonempty}\};$$

$\mathcal{E}$ a closed subset of the topological space $2^{\mathbb{N}^*}$ (with the product topology). We will refrain from explicitly considering measurability questions related to T_∞ and $\mathcal{E}$ in what follows, since all such problems can be addressed in a rather straightforward manner.

4. Exponential Embedding

Our aim in the present section is to present the special construction of the labeled GN process that we alluded to in Section 1. We will show how one can explicitly embed the process in continuous time by employing sequences of independent exponential random variables. Although perhaps complicated at first sight, this embedding will prove to be fundamental to our analysis, with the independence of the involved random variables playing a key role in most of our computations.

4.1. The Balls-in-Bins Case

Davis [Davis 90] applied the elementary properties of exponential random variables to the study of reinforced random walks in a very interesting way. His method was later adapted by Khanin and Khanin [Khanin and Khanin 01] to the

balls-in-bins setting. We present this latter use of exponential random variables (which was also rediscovered by Spencer and Wormald [Spencer and Wormald 04]) as a preparation for the more difficult GN case.

Consider independent random variables $\{X_j, Y_j =^d \exp(f(j))\}_{j\in\mathbb{N}\cup\{0\}}$, and define for $t \geq 0$

$$\begin{aligned} N(t) &\equiv \sup\{n \in \mathbb{N}\cup\{0\} \mid \sum_{i=0}^{n-1} X_i \leq t\}, \\ M(t) &\equiv \sup\{m \in \mathbb{N}\cup\{0\} \mid \sum_{j=0}^{m-1} Y_j \leq t\}. \end{aligned}$$

We interpret the times $\sum_{i=0}^{n-1} X_i$ and $\sum_{j=0}^{m-1} Y_j$ as the times when $N(\cdot)$ and $M(\cdot)$ receive their nth and mth "hits," respectively. We now fix some $t \geq 0$ and $n, m \in \mathbb{N}\cup\{0\}$ and define the event

$$A_{n,m}^t \equiv \{N(t) = n, M(t) = m\} = \left\{ \sum_{i=0}^{n-1} X_i \leq t < \sum_{i=0}^{n} X_i, \sum_{j=0}^{m-1} Y_j \leq t < \sum_{j=0}^{m} Y_j \right\}.$$

What is the probability that the $N(\cdot)$ process is the first one to receive a hit after time t, conditioned on $A_{n,m}^t$? This probability can be written as

$$\mathbb{P}\left(\sum_{i=0}^{n} X_i < \sum_{j=0}^{m} Y_j \;\middle|\; A_{n,m}^t \right).$$

If we further condition on $\sum_{i=0}^{n-1} X_i = s_1 \leq t$ and $\sum_{j=0}^{m-1} Y_j = s_2 \leq t$, we can write this probability as

$$\mathbb{P}\left(X_n + s_1 < Y_m + s_2 \mid X_n \geq t - s_1, Y_m \geq t - s_2\right).$$

The lack-of-memory property of exponentials implies that under the conditioning event above $X_n - t + s_1 =^d \exp(f(n))$ and $Y_m - t + s_2 =^d \exp(f(m))$. The minimum property then implies

$$\begin{aligned} &\mathbb{P}\left(X_n + s_1 < Y_m + s_2 \mid X_n \geq t - s_1, Y_m \geq t - s_2\right) = \\ &\qquad = \mathbb{P}\left(\exp(f(n)) + t < \exp(f(m)) + t\right) = \frac{f(n)}{f(n) + f(m)}. \end{aligned}$$

Since this holds for all $0 \leq s_1$ and $s_2 \leq t$, we have in fact proven that

$$\mathbb{P}\left(\sum_{i=0}^{n} X_i < \sum_{j=0}^{m} Y_j \;\middle|\; A_{n,m}^t \right) = \frac{f(n)}{f(n) + f(m)}.$$

We thus arrive at a surprising conclusion.

Fact 4.1. (Exponential embedding for balls-in-bins.) [Davis 90, Spencer and Wormald 04] *Consider the balls-in-bins process [Drinea et al. 02] with two bins and feedback function f, i.e., the discrete Markov Chain that evolves from state $(n, m) \in (\mathbb{N} \cup \{0\})^2$ to state $(n + 1, m)$ with probability $\frac{f(n)}{f(n)+f(m)}$ and from (n, m) to $(n, m + 1)$ with probability $\frac{f(m)}{f(n)+f(m)}$. It then holds that the joint hit counts of the $(N(\cdot), M(\cdot))$ processes up to the (possibly finite) time when either one becomes infinite is identical in law to the balls-in-bins process with feedback function f started from $(0, 0)$. That is, the balls-in-bins process is embedded in the continuous time $(N(\cdot), M(\cdot))$ process, with X_j (respectively Y_j) parameterizing the time between the arrivals of the jth and $(j + 1)$th balls at the first (respectively second) bin.*

Many nontrivial results that do not have direct combinatorial proofs can be deduced from the construction in Fact 4.1. This method seems to be especially powerful in the case when either N or M reaches an infinite value in finite time. The reader is directed to [Khanin and Khanin 01, Mitzenmacher et al. 04, Spencer and Wormald 04] for many examples of applications of the exponential embedding. We will now show how we can adapt this technique to our present context.

4.2. Exponential Embedding of the GN Process

As pointed out in Section 1, a balls-in-bins process with feedback function f is very similar to a GN process with attachment kernel f to which only new edges (and no new vertices) are added. Conversely, one may think of a GN process as a balls-in-bins process in which each new ball also creates a corresponding bin. This analogy was exploited in [Chung et al. 03], in which a variant of the original GN process was modeled as an "infinite Pólya Urn process" for the purposes of studying the degree sequence. We take this analogy further by adapting the exponential embedding technique to the labeled GN process as defined in Section 3.2.

Our construction starts from an independent sequence $\{X(a, j) =^d \exp(f(j)) \mid a \in \mathbb{N}^*, j \in \mathbb{N} \cup \{0\}\}$ of random variables. The random variable $X(a, 0)$ shall correspond to the age of vertex a at the time its first child a_1 is born. For $j \geq 1$, $X(a, j)$ shall parameterize the time between the births of the jth and $(j + 1)$th children of the a. Therefore, the sequence $\{X(a, j)\}_{a,j}$ plays a role that is similar to that of the X_i and Y_j in Section 4.1. There is, however, one important difference: whereas balls-in-bins processes always have a fixed number

of bins at which balls/hits arrive, the number of "bins" in the GN process grows. That is, the potential vertices $a \in \mathbb{N}^*$ of the trees $\{\mathrm{T}_m\}_{m\geq 0}$ do not all come into existence at the same time; they are rather *born* at appropriate times. We therefore introduce a notion of *birth time*, which is defined recursively as follows:

- the birth time of the empty string $a = \epsilon$ is $\mathcal{B}(\epsilon) = 0$;
- let $a_1, a_2, \ldots, a_n \in \mathbb{N}$, and consider the sequence $a = a_1 \ldots a_n$. The birth time of a is the birth time of the parent sequence $b = a_1 \ldots a_{n-1}$ plus the time until a_n's birth at b. More precisely,

$$\mathcal{B}(a) = \mathcal{B}(a_1 \ldots a_n) = \mathcal{B}(a_1 \ldots a_{n-1}) + \sum_{j=0}^{a_n - 1} X(a_1 \ldots a_{n-1}, j).$$

An equivalent form of the definition of $\mathcal{B}(a)$ is

$$\mathcal{B}(a) = \mathcal{B}(a_1 \ldots a_n) = \sum_{i=0}^{n-1} \sum_{j=0}^{a_{i+1}-1} X(a_1 \ldots a_i, j). \tag{4.1}$$

Our continuous time process is defined by setting

$$\mathcal{W}(t) \equiv \{a \in \mathbb{N}^* \ : \ \mathcal{B}(a) \leq t\}, \quad t \in \mathbb{R}.$$

$\mathcal{W}(\cdot)$ always takes values in the set $\mathcal{E}$ of parent-closed subsets of $\mathbb{N}^*$ (defined in Section 2.3). This is because the definition of birth time implies that the birth time of $a_1 \ldots a_{n-1}$ is always smaller than or equal to that of $a_1 \ldots a_n$.

Let us now specialize to the case where $f : \mathbb{N} \cup \{0\} \to \mathbb{R}$ is given by $f(x) = (x+1)^p$ for some constant $p > 1$. Such attachment kernels satisfy the *explosion condition*

$$\sum_{n\geq 0} \frac{1}{f(n)} < +\infty. \tag{4.2}$$

The condition implies that the expectation of

$$\mathcal{P}(a) \equiv \sum_{j=0}^{+\infty} X(a, j) = \sup_{k\in\mathbb{N}} \mathcal{B}(ak) - \mathcal{B}(a), \quad a \in \mathbb{N}^*, \tag{4.3}$$

is finite. Therefore, all the random variables defined in (4.3) are almost surely finite.

Definition 4.2. For an element $a \in \mathbb{N}^*$, the random variable $\mathcal{P}(a)$ defined in (4.3) is the *explosion time of* a. The infimum of $\mathcal{B}(a) + \mathcal{P}(a)$ over all $a \in \mathbb{N}^*$ is the *tree explosion time*, or the explosion time of the $\mathcal{W}(\cdot)$ process, and is denoted by $\mathcal{S}$:

$$\mathcal{S} \equiv \inf_{a\in\mathbb{N}} \mathcal{B}(a) + \mathcal{P}(a). \tag{4.4}$$

The intuition behind the definition of $\mathcal{S}$ is that it is the first time when some node in the $\mathcal{W}(\cdot)$ process has an infinite number of children. In fact, we *claim* that

Claim 4.3. *The following events hold with probability 1. The birth times $\mathcal{B}(a)$ that are smaller than $\mathcal{S}$ are pairwise distinct and can be well ordered with order type ω. Let $0 = \mathcal{B}(\epsilon) = \mathcal{B}_0 < \mathcal{B}_1 < \ldots < \mathcal{B}_n < \ldots$ denote their ordered sequence, then $\mathcal{B}_n \nearrow \mathcal{S}$ as $n \to +\infty$. Moreover, there exists a unique $v \in \mathbb{N}^*$ that has infinite degree in $\mathcal{W}(\mathcal{S})$; this v satisfies $\mathcal{B}(v) + \mathcal{P}(v) = \mathcal{S}$ and $\mathcal{B}(w) + \mathcal{P}(w) > \mathcal{S}$ for all $w \neq v$.*

A direct consequence of Claim 4.3 is Theorem 4.4 below.

Theorem 4.4. *Let $\{\mathcal{B}_n\}_{n\geq 0}$ be as in Claim 4.3. Then, the sequence $\{T_n \equiv \mathcal{W}(\mathcal{B}_n)\}_{n\geq 0}$ is identical in law to the labeled* GN *process. Moreover, $T_n \to T_\infty \equiv \mathcal{W}(\mathcal{S})$ as $n \to +\infty$.*

However, to prove Claim 4.3, we will need some elements of the proof of Theorem 1.1. This could potentially result in a problem: using Theorem 1.1 to prove Claim 4.3, then employing the claim to prove Theorem 4.4, and finally using this theorem in the proof of Theorem 1.1 would not be acceptable. Instead, we circumvent this difficulty as follows.

1. In the beginning of the next section, we state Lemma 5.1, which is the same as Theorem 1.1 but with $\mathcal{W}(\mathcal{S})$ replacing T_∞ in the statement.

2. Claim 4.3 is then proven, assuming the lemma.

3. The remainder of the section proves Lemma 5.1, *without assuming* Claim 4.3 or Theorem 4.4 in any way.

4. The argument below shows that Claim 4.3 implies Theorem 4.4, which directly implies that Lemma 5.1 can be strengthened to Theorem 1.1.

Irrespective of formal proofs, the reader should keep in mind that $\mathcal{W}(\mathcal{S})$ represents the tree T_∞ in the statements of Theorem 1.1 and Theorem 1.2. Vertices $a \in \mathbb{N}^*$ whose birth times satisfy $\mathcal{B}(a) > \mathcal{S}$ are not really "born" in T_∞ but rather constitute a *fictitious continuation* of T_∞ in which new vertices continue to arrive even though infinitely many vertices have already appeared. We will use this continuation to our advantage in many of the proofs below.

Proof of Theorem 4.4. Assuming Claim 4.3, it suffices to show that for all $A \in \mathcal{E}^{\text{fin}}$ and all $a \in A$,

$$\begin{aligned}
&\mathbb{P}\,(\text{first birth of } \mathcal{W}(\cdot) \text{ after time } t \text{ is at } a \mid \mathcal{W}(t) = A)\\
&= \mathbb{P}\left(\mathcal{B}(a\mathrm{d}_A(a)) + X(a, \mathrm{d}_A(a)) = \min_{b\in A} \mathcal{B}(b\mathrm{d}_A(b)) + X(b, \mathrm{d}_A(b)) \mid \mathcal{W}(t) = A\right)\\
&= \Pi(A, A \cup \{a(\mathrm{d}_A(a) + 1)\}).
\end{aligned} \tag{4.5}$$

To prove this, we first observe that the conditioning event is

$$\{\mathcal{W}(t) = A\} = \left\{\forall c \in A \;\; \mathcal{B}(c) + \sum_{j=0}^{\mathrm{d}_A(c)-1} X(c, j) \le t < \mathcal{B}(c) + \sum_{j=0}^{\mathrm{d}_A(c)} X(c, j)\right\}.$$

We proceed as in the previous section and condition on the values $X(b, j) = x(b, j) \ge 0$ for $b \in A$ and $0 \le j \le \mathrm{d}_A(b) - 1$. We want this event to be a subset of $\{W(t) = A\}$, so we require that the birth times of all $b \in A$ are at most t; that is, we must have

$$\forall b_1 \dots b_r \in A \;\; y(b_1 \dots b_r) \equiv \sum_{i=0}^{r-1} \sum_{j=0}^{b_{i+1}-1} x(b_1 \dots b_i, j) \le t. \tag{4.6}$$

Under this more stringent conditioning, the probability that we wish to compute is

$$\mathbb{P}\left(y(a\mathrm{d}_A(a)) + X(a, \mathrm{d}_A(a)) = \min_{b\in A} y(b\mathrm{d}_A(b)) + X(b, \mathrm{d}_A(b)) \,\middle|\, B\right), \tag{4.7}$$

where

$$B \equiv \{\forall b \in A \;\; X(b, \mathrm{d}_A(b)) > t - y(b, \mathrm{d}_A(b))\}.$$

The exponential random variables in (4.7) are all independent. Moreover, by the lack-of-memory property, $X(b, \mathrm{d}_A(b)) + y(b, \mathrm{d}_A(b)) - t$ conditioned on $X(b, \mathrm{d}_A(b)) > t - y(b, \mathrm{d}_A(b))$ is distributed as $\exp(f(\mathrm{d}_A(b)))$. It follows that

$$\begin{aligned}
&\mathbb{P}\left(y(a\mathrm{d}_A(a)) + X(a, \mathrm{d}_A(a)) = \min_{b\in A} y(b\mathrm{d}_A(b)) + X(b, \mathrm{d}_A(b)) \mid B\right)\\
&\qquad = \mathbb{P}\left(\exp\big(f(\mathrm{d}_A(a))\big) - t = \min_{b\in A} \exp\big(f(\mathrm{d}_A(b))\big) - t\right),
\end{aligned} \tag{4.8}$$

where all the experimental functions are independent. From the minimum property, this last probability is

$$\frac{f(\mathrm{d}_A(a))}{\sum_{b\in A} f(\mathrm{d}_A(b))}, \tag{4.9}$$

and this holds irrespective of the values $\{x(b,j)\}$, as long as (4.6) is satisfied. As a result, (4.5) holds. □

Remark 4.5. The proof of Theorem 4.4 makes it clear that the $\mathcal{W}(\cdot)$ process is a continuous-time Markov Chain on $\mathcal{E}^{\text{fin}}$ up to time $\mathcal{S}$. A consequence of this is the following. Let $t \geq 0$ be given, and let E be an event for $\mathcal{W}(\cdot)$ that is entirely defined in terms of $\{\mathcal{W}(s)\}_{0\leq s\leq t}$. Assume that inside the event E, $\mathcal{W}(t) = A \in \mathcal{E}^{\text{fin}}$. Finally, let F be an event defined entirely in terms of $\{\mathcal{W}(s)\}_{s\geq t}$. Then,

$$\mathbb{P}\left(F \mid E\right) = \mathbb{P}\left(F \mid \mathcal{W}(t) = A\right).$$

We will employ this Remark in the proofs of Lemma 6.1 and Lemma 6.2.

4.3. Three Useful Lemmas

Before we move on to prove the main theorems in the paper, we collect three lemmas (proven in the appendix) that will be useful in dealing with sums of independent exponential random variables. The present lemmas provide estimates of several probabilities that are intimately related to the presence of nodes with k descendants in the final tree T_∞. All of them are key ingredients of the proofs of Theorem 1.1 and Theorem 1.2. We assume that $f(x) = (x+1)^p$ with $p > 1$ in all statements.

Lemma 4.6. (A large-deviations bound.) *There exist constants $C, n_0 > 0$ depending only on f such that for all $n \geq n_0$, all independent sequences of random variables $\{X_j =^d \exp(f(j))\}_{j\geq n}$, and all $\delta > 0$,*

$$\begin{aligned}
\mathbb{P}\left(\sum_{j\geq n} X_j > \mathbb{E}\left[\sum_{j\geq n} X_j\right] + \delta\right) &\leq Ce^{-\delta n^{p-\frac{1}{2}}} \\
\mathbb{P}\left(\sum_{j\geq n} X_j < \mathbb{E}\left[\sum_{j\geq n} X_j\right] - \delta\right) &\leq Ce^{-\delta n^{p-\frac{1}{2}}}.
\end{aligned}$$

Lemma 4.7. *Let $Y = Y_1 + \cdots + Y_k$ be a sum of k independent random variables for which*

$$\forall 1 \leq i \leq k \;\; \mathbb{P}\left(Y_i \leq \epsilon\right) = \Theta(\epsilon) \text{ as } \epsilon \searrow 0.$$

Then, there exist constants C and n_0 depending only on k, f, and the distributions Y_i such that for all independent sequences of random variables $\{X_j =^d$

$\exp(f(j))\}_{j\geq n}$ *that are independent of* Y *and all* $n \geq n_0$,

$$\frac{1}{Cn^{k(p-1)}} \leq \mathbb{P}\left(Y \leq \sum_{j>n} X_j\right) \leq \frac{C}{n^{k(p-1)}}.$$

Lemma 4.8. *Let* $Z_1, \ldots, Z_k$ *be independent exponentials with mean one and let* $Z \equiv Z_1 + \ldots + Z_k$. *Then, for all* $\lambda > 0$,

$$\mathbb{P}(Z \leq \lambda) = e^{-\lambda} \sum_{j\geq k} \frac{\lambda^j}{j!} \leq \frac{\lambda^k}{k!}.$$

5. Finitely Many k-Fertile Vertices

In this section we prove the first of our main results about the GN process, Theorem 1.1. As noted in the previous section, Claim 4.3—which has not been proven yet—is necessary for the connection between the exponential process and the GN process. Proving the claim will require a preliminary form of Theorem 1.1 that we shall present. We assume throughout the section that $f(x) = (x+1)^p$ for some $p > p_k = 1 + 1/k$.

Recall that a node is *k-fertile* if it has k or more descendants in the corresponding tree.

Lemma 5.1. *Consider the* $\mathcal{W}(\cdot)$ *process defined in Section 4, and assume that its attachment kernel is* $f(x) = (x+1)^p$, $p > p_k$. *Then, for all* $T > 0$,

$$\mathbb{E}\left[\#\{a \in \mathcal{W}(\mathcal{S}) : a \text{ is } k\text{-fertile}\} \mid \mathcal{P}(\epsilon) \leq T\right] < +\infty, \tag{5.1}$$

and therefore

$$\mathbb{P}\left(\#\{a \in \mathcal{W}(\mathcal{S}) : a \text{ is } k\text{-fertile}\} < +\infty \mid \mathcal{P}(\epsilon) \leq T\right) = 1.$$

Since $\mathcal{P}(\epsilon) < +\infty$ *almost surely, this implies that* $\mathcal{W}(\mathcal{S})$ *almost surely has only finitely many k-fertile vertices.*

As noted in Section 4, we will use Lemma 5.1 to prove Claim 4.3, and this in turn will imply that Theorem 4.4 holds. This last theorem and Lemma 5.1 directly imply Theorem 1.1. Therefore, most of the present section will be devoted to proving Lemma 5.1.

This section is organized as follows. In Section 5.1 we show how Lemma 5.1 implies Claim 4.3. Having settled that matter, we move on to proving Lemma 5.1. Our proof will consist of bounding the probabilities of the form

$$\mathbb{P}\left(a \text{ is } k\text{-fertile} \mid \mathcal{P}(\epsilon) \leq T\right)$$

and then showing that their sum is finite. We illustrate our techniques for doing so in Section 5.2, where we show a partial result in the direction of Lemma 5.1. We then show in Section 5.3 that the time at which a given $a \in \mathbb{N}^*$ becomes k-fertile in the $\mathcal{W}(\cdot)$ process can be bounded in terms of a sum of k exponential random variables (Lemma 5.6). This permits an improved bound on the probability of k-fertility (Section 5.4), which is then applied to prove Lemma 5.1 in Section 5.5.

5.1. Lemma 5.1 Implies Claim 4.3

Proof of Claim 4.3. The following lemma is a well-known combinatorial result.

Lemma 5.2. (Konig's Infinity Lemma.) *Let T be an infinite rooted tree in which every vertex has finite degree. Then T contains an infinite path starting from the root.*

We will use Konig's Infinity Lemma and Lemma 5.1 to prove a series of almost-sure statements that imply the claim.

All birth times are almost surely distinct. This occurs because, for all distinct $a, b \in \mathbb{N}^*$, the difference $\mathcal{B}(a) - \mathcal{B}(b)$ is a sum of terms of the form $\pm X(c, j)$ for some $(c, j) \in \mathbb{N}^* \times \mathbb{N} \cup \{0\}$. Each such term has a smooth distribution with no point masses, and all terms are independent, hence $\mathcal{B}(a) - \mathcal{B}(b) \neq 0$ with probability 1.

There almost surely exists at least one vertex $v \in \mathbb{N}^*$ with infinite degree in $\mathcal{W}(\mathcal{S})$. Suppose that this is not the case. Since $\mathcal{W}(\mathcal{S})$ is infinite, Konig's Infinity Lemma would imply that there is an infinite path starting from the root in $\mathcal{W}(\mathcal{S})$. But all the infinitely many vertices on such path would have $\geq k$ descendants, for any $k \in \mathbb{N}$. However, $p > 1$ implies that $p > p_k = 1 + 1/k$ for *some* $k \in \mathbb{N}$, and Lemma 5.1 then implies that only finitely many vertices in $\mathcal{W}(\mathcal{S})$ can be k-fertile, a contradiction.

There almost surely exists a unique vertex v for which $\mathcal{S} = \mathcal{B}(v) + \mathcal{P}(v)$. With probability 1, there is a vertex v of $\mathcal{W}(\mathcal{S})$ with infinite degree. Since the degree of v is infinite in $\mathcal{W}(\mathcal{S})$, all the children of v must have been born before time $\mathcal{S}$:

$$\forall n \in \mathbb{N},\ \mathcal{B}(vn) \leq \mathcal{S}.$$

As $n \to +\infty$, $\mathcal{B}(vn) \to \mathcal{B}(v)+\mathcal{P}(v)$, thus $\mathcal{B}(v)+\mathcal{P}(v) \leq \mathcal{S}$. Then, by the definition

of $\mathcal{S}$, $\mathcal{S} = \mathcal{B}(v) + \mathcal{P}(v)$. Thus, there *exists* a v as claimed. For uniqueness, one can show that $\mathcal{B}(a) + \mathcal{P}(a) \neq \mathcal{B}(b) + \mathcal{P}(b)$ for all distinct $a, b \in \mathbb{N}^*$.

With probability 1, $\mathcal{W}(t)$ *is finite for all* $t < s$. Suppose that this is not the case. For all $a \in \mathbb{N}^*$,

$$\mathcal{B}(a) + \mathcal{P}(a) = \lim_{d \to +\infty} \mathcal{B}(ad) \geq \mathcal{S} > t,$$

which implies that for all a there is an integer $d_a \geq 0$ such that $\mathcal{B}(ad_a) > t$. Therefore, any a has finite degree $\leq d_a - 1$ in $\mathcal{W}(t)$. By Konig's Infinity Lemma, $\mathcal{W}(t)$ must then have an infinite path from $\epsilon, a_1, a_1a_2, a_1a_2a_3, \ldots$. But all nodes along this path have infinitely many descendants in $\mathcal{W}(t)$, and hence also in $\mathcal{W}(\mathcal{S})$, which was shown above to have probability 0. The contradiction implies the assertion.

The set of birth times before $\mathcal{S}$ *can be well ordered.* This is a consequence of the previous assertion.

With probability 1, *when the descendants of* v *are removed from* $\mathcal{W}(\mathcal{S})$, *the result is a finite tree.* Again, the key property here is that all $a \in \mathbb{N}^* \backslash \{v\}$ have finite degree. So, if $\mathcal{W}(\mathcal{S})$ without the descendants of v would be infinite, Konig's Infinity Lemma would imply the existence of an infinite path in $\mathcal{W}(\mathcal{S})$, which would imply that all nodes along the path have infinitely many descendants. Since this is impossible, the assertion must be true.

The ordered birth times $\mathcal{B}_0 = 0 \leq \mathcal{B}_1 \leq \mathcal{B}_2 \leq \ldots$ *are almost surely distinct and converge almost surely to* $\mathcal{S}$. That they are distinct follows from the first assertion. Since they form an increasing sequence bounded by $\mathcal{S} < +\infty$, they converge to some finite limit. But the birth times $\{\mathcal{B}(vm)\}_{m=1}^{+\infty}$ (with v as in the previous paragraph) form a subsequence of $\{\mathcal{B}_n\}_{n \in \mathbb{N} \cup \{0\}}$ that converges to $\mathcal{B}(v) + \mathcal{P}(v) = \mathcal{S}$, so the $\{\mathcal{B}_n\}_n$ sequence converges to $\mathcal{S}$ as well.

The series of assertions implies the claim. □

5.2. Two Instructive Examples

Having shown that Lemma 5.1 implies Claim 4.3, we now turn to the proof of the lemma. Recall that the goal of the lemma is to prove that only finitely many vertices have k or more descendants in $\mathcal{W}(\mathcal{S})$. For the sake of the reader, however, we first consider two special classes of $a \in \mathbb{N}^*$ and prove that only finitely many nodes in each class have large degree. While the corresponding general result combines ingredients of the two special cases, we believe that our techniques become much clearer if introduced separately.

To state the present results, we need two definitions. Fix a number $L > 0$, and call $a \in \mathbb{N}^* \backslash \{\epsilon\}$ *L-moderate* if all numbers in the sequence a are smaller than

or equal to L. If on the other hand all numbers in a are bigger than L, call it L-*extreme*. Our two simple lemmas are presented below.

Lemma 5.3. *For all integers $L > 0$ and all $T \geq 0$, the expected number of L-moderate, 1-fertile vertices in $\mathcal{W}(\mathcal{S})$ conditioned on $\mathcal{P}(\epsilon) = T$ is finite.*

Lemma 5.4. *There exists a constant $L_0 > 0$ defined only in terms of p such that for all integers $L \geq L_0$, the expected number of L-extreme vertices in $\mathcal{W}(\mathcal{S})$ that have at least k children is finite.*

Proof of Lemma 5.3. For any $a = a_1 \dots a_m$, the time for the birth of the first child of a is

$$\begin{aligned}\mathcal{B}(a1) &= \mathcal{B}(a) + X(a,0) \\ &= \sum_{j=0}^{m-1} \sum_{i=0}^{a_{j+1}-1} X(a_1 \dots a_j, i) + X(a_1 \dots a_m, 0) \geq \sum_{j=1}^{m} X(a_1 \dots a_j, 0). \quad (5.2)\end{aligned}$$

Notice that this lower bound on $\mathcal{B}(a1)$ is actually *independent* of $\mathcal{P}(\epsilon)$, which is at least as big as the tree explosion time $\mathcal{S}$. As a result,

$$\begin{aligned}\mathbb{P}\left(a \text{ is 1-fertile} \mid \mathcal{P}(\epsilon) = T\right) &= \mathbb{P}\left(\mathcal{B}(a1) \leq \mathcal{S} \mid \mathcal{P}(\epsilon) = T\right) \\ &\leq \mathbb{P}\left(\sum_{j=1}^{m} X(a_1 \dots a_j, 0) \leq \mathcal{P}(\epsilon) \mid \mathcal{P}(\epsilon) = T\right) \\ &= \mathbb{P}\left(\sum_{j=1}^{m} X(a_1 \dots a_j, 0) \leq T\right).\end{aligned}$$

We now apply Lemma 4.8 with $\lambda = T$ and $Z = \sum_{j=1}^{m} X(a_1 \dots a_j, 0)$ to deduce

$$\mathbb{P}\left(a \text{ is 1-fertile } \mid \mathcal{P}(\epsilon) = T\right) \leq \mathbb{P}\left(\sum_{i=1}^{m} X(a_1 \dots a_i, 1) \leq T\right) \leq \frac{T^m}{m!}.$$

There are L^m L-moderate a of length m, and this implies that

$$\sum_{a \ L\text{-moderate}} \mathbb{P}\left(a \text{ is 1-fertile} \mid \mathcal{P}(\epsilon) = T\right) \leq \sum_{m=1}^{+\infty} \frac{T^m L^m}{m!} = e^{TL} - 1 < +\infty.$$

This finishes the proof. □

Proof of Lemma 5.4. We assume $L \geq k+1, n_0$, where n_0 comes from Lemma 4.7. Fix an L-extreme $a = a_1 \dots a_m$ with all $a_i \in \mathbb{N}$, and let $a_{m+1} = k$. The event

$$\{a \text{ has at least } k \text{ children}\} = \{\mathcal{B}(ak) \leq \mathcal{S}\} \tag{5.3}$$

is contained in the event

$$H_{a,i} \equiv \left\{ \sum_{j=0}^{k-1} X(a_1 \dots a_i, j) \leq \sum_{s=a_i}^{+\infty} X(a_1 \dots a_{i-1}, s) \right\}, \tag{5.4}$$

for each $1 \leq i \leq m$. This is true because for all $1 \leq i \leq m$

$$\mathcal{B}(a_1 \dots a_i) + \sum_{j=0}^{k-1} X(a_1 \dots a_i, j) = \mathcal{B}(a_1 \dots a_i k)$$

by properties of birth times

$$\leq \mathcal{B}(a_1 \dots a_i a_{i+1})$$

since $a_i \geq L \geq k$ for $i < m$, and $a_{m+1} = k$

$$\leq \mathcal{B}(ak)$$

since ak is either a descendant of $a_1 \dots a_{i+1}$ or equal to ak and

$$\mathcal{S} \leq \mathcal{B}(a_1 \dots a_{i-1}) + \mathcal{P}(a_1 \dots a_{i-1})$$

by definition of $\mathcal{S}$

$$= \mathcal{B}(a_1 \dots a_{i-1}) + \sum_{j=0}^{+\infty} X(a_1 \dots a_{i-1}, j)$$

$$= \mathcal{B}(a_1 \dots a_i) + \sum_{j=a_i}^{+\infty} X(a_1 \dots a_{i-1}, j).$$

Therefore,

$$\begin{aligned}
& \{\text{a has at least } k \text{ children}\} \\
= \; & \{\mathcal{B}(ak) \leq \mathcal{S}\} \\
\subseteq \; & \left\{ \mathcal{B}(a_1 \dots a_i) + \sum_{j=0}^{k-1} X(a_1 \dots a_i, j) \leq \mathcal{B}(a_1 \dots a_i) + \sum_{j=a_i}^{+\infty} X(a_1 \dots a_{i-1}, j) \right\} \\
= \; & \left\{ \sum_{j=0}^{k-1} X(a_1 \dots a_i, j) \leq \sum_{j=a_i}^{+\infty} X(a_1 \dots a_{i-1}, j) \right\} = H_{a,i}.
\end{aligned}$$

Now note that all the events $\{H_{a,i}\}_{1\leq i\leq m}$ are in fact *independent*. In fact, for any $1 \leq i \leq n-1$, $H_{a,i}$ depends only on the random variables $X(a_1 \dots a_{i-1}, j)$ with $j \geq a_i \geq L \geq k+1$ and $X(a_1 \dots a_i, \ell)$ with $0 \leq \ell \leq k$. Therefore, the choice of L implies that no random variable can appear in the definitions of two different $H_{a,i}$. Thus,

$$\mathbb{P}\left(a \text{ has at least } k \text{ children}\right) \leq \prod_{i=1}^{m} \mathbb{P}\left(H_{a,i}\right).$$

Now notice that

$$\mathbb{P}\left(H_{a,i}\right) = \mathbb{P}\left(\sum_{\ell=1}^{k} Y_\ell \leq \sum_{j\geq a_i} X(a_1 \dots a_{i-1}, j)\right), \tag{5.5}$$

with $Y_\ell = X(a_1 \dots a_i, \ell - 1)$. It is straightforward to check that the assumptions of Lemma 4.7 hold (since we know $a_i \geq L \geq n_0$) and that as a result

$$\mathbb{P}\left(H_{a,i}\right) \leq \frac{C}{a_i^{(p-1)k}}, \tag{5.6}$$

where C depends only on p, as the distributions of the Y_ℓ are determined by p. It follows that

$$\begin{aligned}\sum_{a \text{ } L\text{-ext.}} \mathbb{P}\left(a \text{ has } k \text{ children}\right) &\leq \sum_{m=1}^{+\infty} \sum_{a_1,\dots,a_m > L} \prod_{i=1}^{m} \frac{C}{a_i^{k(p-1)}} \\ &= \sum_{m=1}^{+\infty} \left(\frac{C}{L^{k(p-1)-1}}\right)^m.\end{aligned} \tag{5.7}$$

Noting that $p > p_k \Rightarrow k(p-1) > 1$, we can now take $L \geq L_0 \equiv (2C)^{\frac{1}{k(p-1)-1}}$ to have a finite sum. □

Remark 5.5. One can show by the same proof technique used above that for all fixed $v \in \mathbb{N}^*$ and all fixed $k \in \mathbb{N}$,

$$\mathbb{P}\left(vi \text{ has } k \text{ children before } v \text{ explodes}\right) = O\left(i^{-(p-1)k}\right) \text{ as } i \to +\infty. \tag{5.8}$$

To prove this, note that the event in (5.8) is

$$\{\mathcal{B}(vik) \leq \mathcal{B}(v) + \mathcal{P}(v)\} = \left\{\sum_{j=0}^{k-1} X(vi, k) \leq \sum_{j\geq i} X(v, j)\right\},$$

because $\mathcal{B}(vik) = \mathcal{B}(vi) + \sum_{j=0}^{k-1} X(vi,k)$ and $\mathcal{B}(v) + \mathcal{P}(v) = \mathcal{B}(vi) + \sum_{j \geq i} X(v,j)$. Then, apply Lemma 4.7, as in the previous proof.

Similarly, one can show that, for all $t \geq 0$, all $v, w \in \mathbb{N}^*$, and all finite trees T_n such that v has n children in T_n,

$$\mathbb{P}\left(w \text{ has } k \text{ children after time } t \text{ and before } v \text{ explodes} \mid \mathcal{W}(t) = \mathrm{T}_n\right) = O\left(n^{-(p-1)k}\right) \text{ as } n \to +\infty.$$

We will employ this remark in the proof of Lemma 6.2.

5.3. Subtrees and the Time until k Descendants Are Born

There are two reasons why Lemma 5.3 and Lemma 5.4 do not imply Lemma 5.1. First, there are $a \in \mathbb{N}^*$ that are neither L-moderate nor L-extreme. Second, the two lemmas only bound the probability of a certain node having degree $\geq k$, which is different from k-fertility for all $k \geq 2$. The next lemma deals with the latter difficulty. Fix some $a \in \mathbb{N}^*$, and let $\mathcal{W}_a(t) \equiv \{c \in \mathbb{N}^* \,:\, \mathcal{B}(ac) - \mathcal{B}(a) \leq t\}$, for $t \in \mathbb{R}$, be the subtree of $\mathcal{W}(t + \mathcal{B}(a))$ rooted at a. Clearly, $\mathcal{W}_a(\cdot)$ and $\mathcal{W}(\cdot) = \mathcal{W}_\epsilon(\cdot)$ have the same distribution. Moreover, a is k-fertile if and only if the size of $\mathcal{W}_a(\mathcal{S} - \mathcal{B}(a))$ is at least $k+1$ (i.e., $\mathcal{W}_a(\mathcal{S} - \mathcal{B}(a))$ has at least k vertices other than the root). Lemma 5.6 provides tools for the analysis of the k-fertility event.

Lemma 5.6. *For a fixed $a \in \mathbb{N}^*$, let $\mathcal{T}_0(a)$ be the time of the first birth of a node other than the root in the $\mathcal{W}_a(\cdot)$ process. Moreover, for $i \in \mathbb{N}$, let $\mathcal{T}_i(a)$ be the time elapsed between the ith and $(i+1)$th births in $\mathcal{W}_a(\cdot)$ (again excluding the birth time of the root). Then, there exists a sequence of random variables $\{\mathcal{R}_j(a)\}_{j=0}^{+\infty}$ such that*

1. *$\{\mathcal{R}_j(a)\}_{j=0}^{+\infty}$ is a sequence of independent random variables;*
2. *the sequence $\{\mathcal{R}_j(a)\}_{j=0}^{+\infty}$ is a deterministic function of the random variables $\{X(ac,i) \mid c \in \mathbb{N}^*, i \in \mathbb{N}\}$;*
3. *for each $j \in \mathbb{N} \cup \{0\}$, $\mathcal{R}_j(a) =^d \exp((j+1)f(j))$;*
4. *$\mathcal{R}_0(a) = \mathcal{T}_0(a)$ and, for all $j \in \mathbb{N}$, $\mathcal{R}_j(a) \leq \mathcal{T}_j(a)$.*

Proof. It suffices to consider the case $a = \epsilon$. For convenience, we introduce the notation

$$\Sigma(c) \equiv \begin{cases} \sum_{i=1}^m c_i & c = c_1 \ldots c_m \in \mathbb{N}^* \backslash \{\epsilon\} \\ 0 & c = \epsilon \end{cases}.$$

We prove inductively that the random variables $\{\mathcal{R}_j(\epsilon)\}_{j=0}^r$ can be defined as above so that, for all $j \in \mathbb{N} \cup \{0\}$, $\mathcal{R}_j(\epsilon)$ is completely defined by the values of $X(c,r)$ for $c \in \mathbb{N}^*$, $0 \leq \Sigma(c) + j \leq r$. For $r = 0$, this is easy: just set $\mathcal{R}_0(\epsilon) = \mathcal{T}_1(\epsilon) = X(\epsilon, 0)$. Now assume inductively that $\mathcal{R}_j(\epsilon)$ has been defined for all $0 \leq j \leq r = n-1$. To prove that the same is possible for $r = n$, condition on a particular value

$$\mathcal{W}_\epsilon \left(\sum_{j=0}^{n-1} \mathcal{T}_j(\epsilon) \right) = A \in \mathcal{E}, |A| < +\infty. \tag{5.9}$$

The sum $\sum_{j=1}^{n-1} \mathcal{T}_j(\epsilon)$ is exactly the birth time of the nth descendant of the root in $\mathcal{W}_\epsilon(\cdot)$ (for ϵ is born at time 0), hence $|A| = n+1$. We also notice that $\Sigma(c) + \mathrm{d}_A(c) \leq n$ for all $c \in A$. Indeed, the sequence

$$b = \begin{cases} c\mathrm{d}_A(c) & \text{if } \mathrm{d}_A(c) > 0 \\ c & \text{if } \mathrm{d}_A(c) = 0 \end{cases}$$

is an element of A with $\Sigma(b) = \Sigma(c) + \mathrm{d}_A(c)$, and it is a simple fact (whose proof we omit) that $\Sigma(b) \leq |A| - 1$ for *any* $b \in A \in \mathcal{E}^{\mathrm{fin}}$.

Conditioned on the event in (5.9), the random variable $\mathcal{T}_n(a)$ has exponential distribution with rate $\sum_{c \in A} f(\mathrm{d}_A(c))$, which is bounded by $|A| f(n) \leq (n+1) f(n)$ by the previous remarks. Therefore,

$$\mathcal{R}_n(a) \equiv \frac{\sum_{c \in A} f(\mathrm{d}_A(c))}{(n+1) f(n)} \mathcal{T}_n(a) \leq \mathcal{T}_{n+1}(a) \quad \left(\text{where } A = \mathcal{W}_\epsilon \left(\sum_{j=1}^{n-1} \mathcal{T}_j(\epsilon) \right) \right)$$

is exponential with rate $(n+1)f(n)$ irrespective of A, by the multiplication property of exponentials (see Section 2.1). Because $\mathcal{W}_a(\sum_{j=1}^{n-1} \mathcal{T}_j(a))$ and $\mathcal{T}_n(a)$ are completely defined by the random variables $\{X(ac, j) \ : \ c \in \mathbb{N}^*, \Sigma(c) + j \leq n+1\}$, the same is true of $\mathcal{R}_n(a)$. This finishes the proof. □

5.4. A General Bound on the Probability of k-Fertility

Lemma 5.6 is now used to prove a stronger form of the bounds in Section 5.2 that applies to *all* $a \in \mathbb{N}^*$ (not just L-moderate or L-large sequences). To present this bound, we need a definition. For a fixed $L > 0$ and a sequence $a = a_1 a_2 \ldots a_m \in \mathbb{N}^*$ of length m, the set of *small indices in* a is $\mathsf{sm}_L(a) = \{1 \leq i \leq m \ : \ a_i \leq L\}$, and the set of *large indices in* a is $\mathsf{lg}_L(a) = \{1 \leq i \leq m \ : \ a_i > L\}$.

Lemma 5.7. *There exist constants $C, L_0 > 0$ depending only on k and p such that for any $T > 0$, $L \geq L_0$, and $a = a_1 \dots a_m \in \mathbb{N}^*$,*

$$\mathbb{P}\left(a \text{ is } k\text{-fertile and } \mathcal{P}(\epsilon) \leq T\right) \leq \frac{\max\{T, C\}^m}{|\mathsf{sm}_L(a)|!} \prod_{j \in \mathsf{lg}_L(a)} \frac{1}{a_j^{(p-1)k}}. \tag{5.10}$$

Proof. For most of the proof, we will only assume that $L_0 \geq k$; more conditions on L_0 will be imposed later. Set $a_{m+1} \equiv k$, and for each $i \in \mathsf{lg}_L(a)$, define I_i to be the smallest $j \in \mathsf{lg}_L(a) \cup \{m+1\}$ satisfying $j > i$; notice that the choice of L_0 implies $a_{I_i} > L_0 \geq k$ whenever $I_i < m+1$. Employing the random variables $\{\mathcal{R}_j(a)\}_{j=0}^{k-1}$ whose existence Lemma 5.6 guarantees, we deduce that

$$\begin{aligned} \{a \text{ is } k\text{-fertile}\} &\subset \{\mathcal{B}(a) + \sum_{n=0}^{k-1} \mathcal{R}_n(a) \leq \mathcal{S}\} \\ &= \{\mathcal{B}(a) + X(a,0) + \sum_{n=1}^{k-1} \mathcal{R}_n(a) \leq \mathcal{S}\}. \end{aligned} \tag{5.11}$$

In what follows, we will bound the probability on the right-hand side, noting that the $\mathcal{R}_j(a)$ and $X(b,i)$ that appear in the definitions below are all independent because of Lemma 5.6. Consider the following events:

$$F_a^T \equiv \left\{\mathcal{B}(a) + X(a,0) + \sum_{n=1}^{k-1} \mathcal{R}_n(a) \leq \mathcal{S} \text{ and } \mathcal{P}(\epsilon) \leq T\right\}, \tag{5.12}$$

$$G_a^T \equiv \left\{\sum_{i \in \mathsf{sm}_L(a)} X(a_1 \dots a_i, 0) \leq T\right\}, \tag{5.13}$$

$$H_{a,i} \equiv \left\{\begin{pmatrix} X(a_1 \dots a_i, 0) \\ +\sum_{j=1}^{k-1} X(a_1 \dots a_{I_i - 1}, j) \end{pmatrix} \leq \sum_{j \geq a_i} X(a_1 \dots a_{i-1}, j)\right\}, \tag{5.14}$$

$$i \in \mathsf{lg}_L(a), I_i \neq m+1,$$

$$H_{a,i} \equiv \left\{X(a_1 \dots a_i, 0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a) \leq \sum_{j \geq a_i} X(a_1 \dots a_{i-1}, j)\right\}, \tag{5.15}$$

$$i \in \mathsf{lg}_L(a), I_i = m+1.$$

The first event is the one whose probability we want to bound. The second event is similar to the one in the proof of Lemma 5.3, whereas the remaining events are reminiscent of those in the proof of Lemma 5.4. We now *claim* the following.

Claim 5.8. *It holds that*

$$F_a^T \subseteq G_a^T \cap \bigcap_{i \in \mathsf{lg}_L(a)} H_{a,i}. \tag{5.16}$$

Moreover, the events on the right-hand side of (5.16) are independent.

Claim 5.8 is proven at the end of the current proof, but we now present the following concrete example of its application to illustrate our argument. Assume that $k = 2$, $L = L_0 = 3$, and $a = a_1 a_2 \dots a_6 = 142461$, in which case $\mathsf{sm}_L(a) = \{1, 3, 6\}$ and $\mathsf{lg}_L(a) = \{2, 4, 5\}$. Figure 2 represents some of the random variables involved in (5.16) by rectangles. The first six columns of rectangles stand for random variables of the form $X(b, j)$ for $b = \epsilon$ (the empty string), 1, 14, ..., 14246 and $j = 0, 1, \dots, 6$, while the last column represents the random variables $\mathcal{R}_0(a) = X(a, 0)$ and $\mathcal{R}_1(a)$. The rectangles that lie completely below the dashed line correspond to the random variables that appear in

$$\begin{aligned} \mathcal{B}(a) \;=\; & X(\epsilon, 0) + \sum_{j_1=0}^{3} X(1, j_1) + \sum_{j_2=0}^{1} X(14, j_2) \\ & + \sum_{j_3=0}^{3} X(142, j_3) + \sum_{j_4=0}^{5} X(1424, j_4) + X(14246, 0). \end{aligned} \tag{5.17}$$

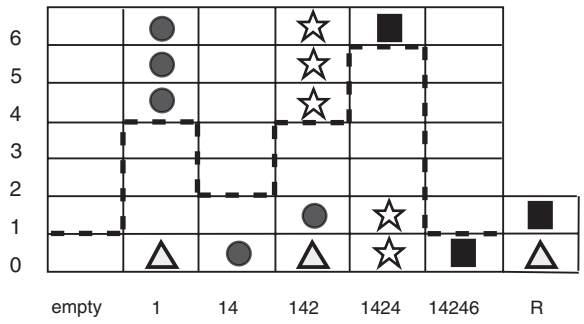

Figure 2. Illustration of Claim 5.8 for $a = 142461$.

Moreover,

$$I_i = \begin{cases} 4, & i = 2 \\ 5, & i = 4 \\ 7, & i = 5. \end{cases}$$

By checking the definitions of G_a^T and $H_{a,i}$, one can check that the following statements hold:

1. $G_a^T \supseteq F_a^T$, since in the event F_a^T the explosion time $\mathcal{S}$ is at most T, and the sum (5.17) defining $\mathcal{B}(a)+X(a,0)$ contains the terms of $\sum_{i\in \mathsf{sm}_L(a)} X(a_1,\ldots a_i,0)$. Moreover, the random variables appearing in G_a^T correspond to the rectangles marked with triangles in Figure 2.

2. $H_{a,2} \supseteq F_a^T$. In order for F_a^T to happen, $a_1a_2\ldots a_{I_2} = 1424$ must be born before $a_1 = 1$ explodes. In particular, using the critical fact that $k = 2 < L = 3$, so that $a_{I_2} > k$, $1422 = a_1a_2\ldots a_{I_2-1}k$ must be born before node $a_1 = 1$ explodes. Since a_1 explodes at time $\mathcal{B}(a_1)+\mathcal{P}(a_1) = \mathcal{B}(1)+X(1,0)+X(1,1)+\ldots$, 1422 is born at a time that is larger than $\mathcal{B}(14)+X(14,0)+X(142,1)$ and $\mathcal{B}(14) = \mathcal{B}(1)+X(1,0)+\cdots+X(1,3)$, it follows that

$$\begin{aligned} F_a^T &\subset \left\{ \begin{pmatrix} \mathcal{B}(1)+X(1,0)+ \\ +\cdots+X(1,3)+ \\ X(14,0)+X(142,1) \end{pmatrix} \le \mathcal{B}(1)+X(1,0)+X(1,1)+\ldots \right\} \\ &\subset \left\{ X(14,0)+X(142,1) \le \sum_{j\ge 4} X(1,j) \right\} = H_{a,2}, \end{aligned}$$

   so $H_{a,2} \supseteq F_a^T$ is indeed true. Moreover, one can check that the random variables appearing in the definition of $H_{a,2}$ are precisely the ones marked with circles in Figure 2.

3. Similarly, one can show that $H_{a,4} \supseteq F_a^T$ (respectively, $H_{a,5} \supseteq F_a^T$) and that the random variables marked with stars (respectively squares) are precisely the ones appearing in the definition of $H_{a,4}$ (respectively $H_{a,5}$).

Items 1, 2, and 3 imply not only the validity of (5.16) but also that *no random variable of the form $X(\cdot,\cdot)$ or $\mathcal{R}_{(\cdot)}(a)$ appears in the definition of more than one of the events in (5.13)–(5.15).* Since those random variables are also *independent*, we have proven that G_a^T, $H_{a,2}$, $H_{a,4}$, and $H_{a,5}$ are independent events, which implies the claim in this special case. The proof of Claim 5.8 for general L, k, and a is entirely analogous to the argument sketched above.

We continue with the proof of Lemma 5.7, noting that Claim 5.8 implies

$$\mathbb{P}\left(F_a^T\right) \leq \mathbb{P}\left(G_a^T\right) \times \prod_{i \in \mathsf{lg}_L(a)} \mathbb{P}\left(H_{a,i}\right). \tag{5.18}$$

The remainder of our proof consists of bounding the probabilities on the right-hand side of (5.18), which is done in roughly the same way as in Lemmas 5.3 and 5.4. The probability of G_a^T is bounded using Lemma 4.8 with the Z_i corresponding to the $X(a_1 \dots a_i, 0)$ for $i \in \mathsf{sm}_L(a)$ and $\lambda = T$:

$$\mathbb{P}\left(G_a^T\right) \leq \frac{T^{|\mathsf{sm}_L(a)|}}{|\mathsf{sm}_L(a)|!}. \tag{5.19}$$

Now fix some $i \in \mathsf{lg}_L(a)$ with $I_i \neq m+1$. We apply Lemma 4.7 with $Y_1 = X(a_1 \dots a_i, 0)$, $Y_\ell = X(a_1 \dots a_{I_i - 1}, \ell - 1)$ for $2 \leq \ell \leq k$, and

$$\{X_j\}_{j \geq n} = \{X(a_1 \dots a_{i-1}, j)\}_{j \geq a_i}.$$

In the present case, the distributions of the Y_ℓ are all defined in terms of f and k. Therefore, there exist C and n_0 depending only on k and f such that if $a_i \geq n_0$,

$$\mathbb{P}\left(H_{a,i}\right) \leq \frac{C}{a_i^{(p-1)k}}. \tag{5.20}$$

For $i \in \mathsf{lg}_L(a)$ with $I_i = m+1$, a similar reasoning with $Y_\ell = \mathcal{R}_\ell(a)$ for $2 \leq \ell \leq k$ implies that for (possibly enlarged) C and n_0 depending only on k and f, and all $a_i \geq n_0$, (5.20) still holds. So, if we take $L_0 \geq n_0$, we can plug (5.19) and (5.20) into (5.18) for any a, which finishes the proof. □

To conclude, we now prove Claim 5.8.

Proof of Claim 5.8. We first show that each of the events (re)defined as

$$G_a^T \equiv \left\{ \sum_{i \in \mathsf{sm}_L(a)} X(a_1 \dots a_i, 0) \leq T \right\}, \tag{5.21}$$

$$H_{a,i} \equiv \left\{ \begin{pmatrix} X(a_1 \dots a_i, 0) \\ +\sum_{j=1}^{k-1} X(a_1 \dots a_{I_i - 1}, j) \end{pmatrix} \leq \sum_{j \geq a_i} X(a_1 \dots a_{i-1}, j) \right\}, \tag{5.22}$$

$$i \in \mathsf{lg}_L(a), I_i \neq m+1, \text{ and}$$

$$H_{a,i} \equiv \left\{ X(a_1 \dots a_i, 0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a) \leq \sum_{j \geq a_i} X(a_1 \dots a_{i-1}, j) \right\}, \tag{5.23}$$

$$i \in \mathsf{lg}_L(a), I_i = m + 1,$$

contains

$$F_a^T = \left\{ \mathcal{B}(a) + X(a,0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a) \leq \mathcal{S} \text{ and } \mathcal{P}(\epsilon) \leq T \right\}. \tag{5.24}$$

First containment: $G_a^T \supset F_a^T$. On the one hand, all terms appearing in the sum

$$\sum_{i \in \mathsf{sm}_L(a)} X(a_1 \dots a_i, 0)$$

also appear in the sum defining $\mathcal{B}(a) = \mathcal{B}(a_1 \dots a_m)$ (see (4.1)), so that

$$\sum_{i \in \mathsf{sm}_L(a)} X(a_1 \dots a_i, 0) \leq \mathcal{B}(a) \leq \mathcal{B}(a) + X(a,0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a). \tag{5.25}$$

On the other hand, by the definition of $\mathcal{S}$ in (4.4),

$$\mathcal{S} \leq \mathcal{B}(\epsilon) + \mathcal{P}(\epsilon) = \mathcal{P}(\epsilon) \tag{5.26}$$

since $\mathcal{B}(\epsilon) = 0$. Therefore,

$$F_a^T \text{ occurs } \Rightarrow \sum_{i \in \mathsf{sm}_L(a)} X(a_1 \dots a_i, 0) \leq \mathcal{P}(\epsilon) \leq T \Rightarrow G_a^T \text{ occurs.}$$

Second containment: $H_{a,i} \supset F_a^T$ *if* $\boldsymbol{i \in \mathsf{lg}_L(a)}$ *and* $\boldsymbol{I_i < m + 1}$. Consider the sum

$$X(a_1 \dots a_i, 0) + \sum_{j=1}^{k-1} X(a_1 \dots a_{I_i - 1}, j).$$

In the present case, $I_i \in \mathsf{lg}_L(a)$. Our choice of $L \geq k$ is now used, for it implies that $a_{I_i} \geq L \geq k$, and hence

$$X(a_1 \dots a_i, 0) + \sum_{j=1}^{k-1} X(a_1 \dots a_{I_i - 1}, j) \leq X(a_1 \dots a_i, 0) + \sum_{j=1}^{a_{I_i} - 1} X(a_1 \dots a_{I_i - 1}, j).$$

The terms in this sum each appear once in

$$\sum_{t=i}^{m-1} \sum_{j=0}^{a_{t+1}-1} X(a_1 \dots a_t, j) + X(a,0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a)$$

$$= \mathcal{B}(a) - \mathcal{B}(a_1 \dots a_i) + X(a,0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a), \tag{5.27}$$

and it follows that

$$X(a_1 \dots a_i, 0) + \sum_{j=1}^{k-1} X(a_1 \dots a_{I_i - 1}, j) \leq \mathcal{B}(a) - \mathcal{B}(a_1 \dots a_i) + X(a, 0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a). \quad (5.28)$$

Therefore,

$$\begin{aligned} F_a^T \text{ occurs } &\Rightarrow \mathcal{B}(a) + X(a, 0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a) \leq \mathcal{S} \\ &\Rightarrow X(a_1 \dots a_i, 0) + \sum_{j=1}^{k-1} X(a_1 \dots a_{I_i - 1}, j) \leq \mathcal{S} - \mathcal{B}(a_1 \dots a_i). \end{aligned}$$

But, it is always true that

$$\mathcal{S} \leq \mathcal{B}(a_1 \dots a_{i-1}) + \mathcal{P}(a_1 \dots a_{i-1})$$

and

$$\mathcal{B}(a_1 \dots a_{i-1}) + \mathcal{P}(a_1 \dots a_{i-1}) - \mathcal{B}(a_1 \dots a_i) = \sum_{j \geq a_i} X(a_1 \dots a_{i-1}, j).$$

Hence,

$$\begin{aligned} F_a^T \text{ occurs } &\Rightarrow X(a_1 \dots a_i, 0) + \sum_{j=1}^{k-1} X(a_1 \dots a_{I_i - 1}, j) \leq \sum_{j \geq a_i} X(a_1 \dots a_{i-1}, j) \\ &\Rightarrow H_{a,i} \text{ occurs.} \end{aligned}$$

Third containment: $H_{a,i} \subset F_a^T$ *if* $i \in \mathsf{lg}_L(a)$ *and* $I_i = m + 1$. In this case, the terms of the sum

$$X(a_1 \dots a_i, 0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a)$$

are all contained in

$$\sum_{t=i}^{m-1} \sum_{j=0}^{a_{t+1}-1} X(a_1 \dots a_t, j) + X(a, 0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a)$$

$$= \mathcal{B}(a) - \mathcal{B}(a_1 \dots a_i) + X(a, 0) + \sum_{j=1}^{k-1} \mathcal{R}_j(a). \quad (5.29)$$

The rest of the proof proceeds exactly as in the case of the second containment.

We now show that the events in (5.21) to (5.23) are *independent*. This is proven by showing that no term $X(b, r)$ appears in the definition of more than one of those events. We will analyze three different cases.

1. Comparing G_a^T to the remaining events. G_a^T is entirely defined in terms of $X(a_1 \dots a_t, 0)$ for $t \in \mathsf{sm}_L(a)$. The only terms of the form $X(b, 0)$ appearing in the definition of the events $H_{a,i}$ have $b = a_1 \dots a_i$ for $i \in \mathsf{lg}_L(a)$. This implies that no random variable appears in the definition of both G_a^T and $H_{a,i}$, for all $i \in \mathsf{lg}_L(a)$.

2. Comparing $H_{a,i}$ to $H_{a,\ell}$ for $i < I_i < \ell$, $i, \ell \in \mathsf{lg}_L(a)$. The definition of $H_{a,i}$ only involves random variables of the form $X(a_1 \dots a_t, j)$ for some $j \in \mathbb{N} \cup \{0\}$ and $t \leq I_i - 1 < \ell - 1$, whereas the definition of $H_{a,\ell}$ involves $X(a_1 \dots a_s, j)$ for $s \geq \ell - 1$. Therefore, the ranges of the indices t and s will never overlap in this case.

3. Comparing $H_{a,i}$ to $H_{a,\ell}$ for $i < I_i = \ell$, $i, \ell \in \mathsf{lg}_L(a)$. By the same argument and with the same notation as above, the only possibility for trouble is when $t = I_i - 1 = \ell - 1 = s$. This is precisely where the assumption that $L \geq k$ comes in. The event $H_{a,i}$ involves random variables of the form

$$\{X(a_1 \dots a_{I_i - 1}, j) : 1 \leq j \leq k - 1\}, \quad (5.30)$$

whereas the event $H_{a,\ell}$ uses the random variables

$$\{X(a_1 \dots a_{\ell-1}, j) : j \geq a_\ell\}. \quad (5.31)$$

Since $\ell \in \mathsf{lg}_L(a)$, $a_\ell > L \geq k$, the ranges of j in the two formulae above do not overlap, and we are done.

□

5.5. Proof of Lemma 5.1

Having proven Lemma 5.7, we now come to the end of the proof of Lemma 5.1.

Proof of Lemma 5.1. Our aim is to show that, for any $T > 0$,

$$\sum_{a \in \mathbb{N}^*} \mathbb{P}\left(a \text{ is } k\text{-fertile} \mid \mathcal{P}(\epsilon) \leq T\right) < +\infty. \tag{5.32}$$

To this end, we employ Lemma 5.7 and prove instead that, for some fixed number $L \geq L_0$ depending only on f, k, and T,

$$\sum_{a \in \mathbb{N}^*} \frac{\max\{C,T\}^m}{|\mathsf{sm}_L(a)|!} \prod_{j \in \mathsf{lg}_L(a)} \frac{1}{a_j^{(p-1)k}} < +\infty. \tag{5.33}$$

We will eventually choose some L such that

$$\sum_{|a|=m} \frac{\max\{C,T\}^m}{|\mathsf{sm}_L(a)|!} \prod_{j \in \mathsf{lg}_L(a)} \frac{1}{a_j^{(p-1)k}} = 2^{-\Omega(m)} \text{ as } m \to +\infty, \tag{5.34}$$

which clearly implies (5.33). Fix some m and a subset $S \subseteq \{1, \ldots, m\}$ of size $|S| = s$. The sum of these quantities over all a of length $|a| = m$ with $\mathsf{sm}_L(a) = S$ is

$$L^s \frac{\max\{T,C\}^m}{s!} \prod_{i \in \{1,\ldots,m\} \setminus S} \sum_{a_i > L} \frac{1}{a_i^{(p-1)k}},$$

because there are L^s ways of choosing the a_j with $j \in \mathsf{sm}_L(a)$. Now note that S can be chosen in $\binom{m}{s}$ for any $0 \leq s \leq m$, and therefore

$$\begin{aligned}
&\sum_{\substack{|a|=m \\ |\mathsf{sm}_L(a)|=s}} \frac{\max\{C,T\}^m}{|\mathsf{sm}_L(a)|!} \prod_{j \in \mathsf{lg}_L(a)} \frac{1}{a_j^{(p-1)k}} \\
&\qquad = \binom{m}{s} L^s \frac{\max\{C,T\}^m}{s!} \prod_{j=1}^{m-s} \sum_{a_j > L} \frac{1}{a_j^{(p-1)k}} \\
&\qquad \leq \binom{m}{s} L^s \frac{\hat{T}^m}{s!} \left(\int_L^{+\infty} \frac{dx}{x^{k(p-1)}} \right)^{m-s} \\
&\qquad \leq \binom{m}{s} (L)^{s-\alpha(m-s)} \frac{\hat{T}^m}{s!},
\end{aligned}$$

where $\alpha \equiv k(p-1) - 1$ and

$$\hat{T} \equiv \max\{C,T\} \times \max\left\{1, \frac{1}{\alpha}\right\}.$$

Here, we make critical use of the condition $p > p_k = 1 + 1/k$: under this assumption, $\alpha > 0$. Summing over s, we discover that

$$\sum_{|a|=m} \frac{\max\{C,T\}^m}{|\mathsf{sm}_L(a)|!} \prod_{j\in \mathsf{lg}_L(a)} \frac{1}{a_j^{(p-1)k}} \le \sum_{s=0}^{m} \binom{m}{s} \frac{\hat{T}^m L^{s-(m-s)\alpha}}{s!}. \tag{5.35}$$

To bound this last sum, we split it into two parts, corresponding to $s \le \alpha m/2(1+\alpha)$ and $s > \alpha m/2(1+\alpha)$. For the first part, we forget the $s!$ term and bound $s-(m-s)\alpha \le -\alpha m/2$; for the second, we simply bound $s! \ge \lceil \alpha m/2(1+\alpha)\rceil!$ and $s-(m-s)\alpha \le m$:

$$\begin{aligned}
\sum_{s\le \frac{\alpha m}{2(1+\alpha)}} \binom{m}{s} \frac{\hat{T}^m L^{s-(m-s)\alpha}}{s!} &\le \left(\frac{\hat{T}}{L^{\frac{\alpha}{2}}}\right)^m \sum_{s\le \frac{\alpha m}{2(1+\alpha)}} \binom{m}{s}, \\
\sum_{s> \frac{\alpha m}{2(1+\alpha)}} \binom{m}{s} \frac{\hat{T}^m L^{s-(m-s)\alpha}}{s!} &\le \frac{\left(\hat{T}L\right)^m}{\left\lceil \frac{\alpha m}{2(1+\alpha)}\right\rceil!} \sum_{s> \frac{\alpha m}{2(1+\alpha)}} \binom{m}{s}.
\end{aligned}$$

It follows that for $L \ge (4\hat{T})^{2/\alpha}$, which only depends on f, p, and T,

$$\begin{aligned}
\sum_{|a|=m} \binom{m}{s} \frac{\hat{T}^m L^{s-(m-s)\alpha}}{s!} &\le 2^m \left(\frac{1}{4^m} + \frac{\left(\hat{T}^{1+2/\alpha}\right)^m}{\left\lceil \frac{\alpha m}{2(1+\alpha)}\right\rceil!}\right) &(5.36)\\
&= 2^{-\Omega(m)} \text{ as } m \to +\infty. &(5.37)
\end{aligned}$$

This proves (5.34) and finishes the proof. □

6. The Structure of the Infinite Tree

Now that the proof of Theorem 1.1 is complete, we proceed to prove Theorem 1.2. We will assume throughout the section that $f(x) = (x+1)^p$ (with $p > 1$) and that $k = k_p$ is as in the statement of the theorem. As in the previous section, it is convenient to break the proof down into steps.

Lemma 6.1. *If S is a rooted tree with $|S| = \ell + 1$ vertices, then for all $a \in \mathbb{N}^*$*

$$\begin{aligned}
&\mathbb{P}\left(\#\{n \in \mathbb{N} : \mathcal{W}_{an}(\mathcal{P}(a) + \mathcal{B}(a) - \mathcal{B}(an)) \text{ is isomorphic to } S\} = +\infty\right) \\
&\qquad = \begin{cases} 1 & p \le p_\ell, \\ 0 & p > p_\ell. \end{cases} \qquad (6.1)
\end{aligned}$$

Lemma 6.2. *Let $\tilde{T}$ be any finite tree and v be a vertex of $\tilde{T}$. There is a positive probability that all of the following events hold:*

1. *the labeled* GN *process reaches state $\tilde{T}$,*
2. *v is the unique vertex present in $\tilde{T}$ to have any children after state $\tilde{T}$ is reached, and*
3. *all nodes that are born after state $\tilde{T}$ is reached are ℓ-fertile for some $\ell < k$.*

As we shall see below, these lemmas permit that Theorem 1.2 is easily proven.

Proof of Theorem 1.2. By Claim 4.3, there almost surely exists a unique node $v \in \mathbb{N}^*$ with $\mathcal{B}(v) + \mathcal{P}(v) = \mathcal{S}$, and all other nodes have finitely many descendants in T_∞. Moreover, since $p > 1 + 1/k_p$, one can apply Theorem 1.1 and deduce that with probability 1 there are only finitely many children vn of v that are k-fertile.

If we remove all *other* children of v (i.e., those that have $\leq k-1$ descendants, which must be infinitely many) and their descendants from T_∞, we obtain a finite tree T. We *claim* that in fact $\mathrm{T}_\infty = \textsc{Glue}(\mathrm{T}, v, k)$. Consider some (rooted, oriented) tree S with $|S| \leq k$. By Lemma 6.1, there almost surely exist infinitely many $n \in \mathbb{N}$ such that $\mathcal{W}_{vn}(\mathcal{P}(v) - \sum_{j=0}^{n-1} X(v,j))$ is isomorphic to S, and because $\mathcal{S} = \mathcal{B}(v) + \mathcal{P}(v) = \mathcal{B}(vn) + (\mathcal{P}(v) - \sum_{j=0}^{n-1} X(v,j))$, this implies that $\mathcal{W}_{vn}(\mathcal{S} - \mathcal{B}(vn))$ is isomorphic to S for infinitely many n. But $\mathcal{W}_{vn}(\mathcal{S} - \mathcal{B}(vn))$ is the subtree of $\mathrm{T}_\infty = \mathcal{W}(\mathcal{S})$ rooted at (and oriented towards) vn, hence with probability 1 there are infinitely many $n \in \mathbb{N}$ such that the subtree of T_∞ rooted at vn is isomorphic to S. This is true for any S of size $\leq k$, so all such trees must appear infinitely often, and this finishes the proof of the claim.

We have shown that T_∞ is always isomorphic to some $\textsc{Glue}(\mathrm{T}, v, k)$. Moreover, Lemma 6.2 says that *any* $\textsc{Glue}(\mathrm{T}, v, k)$ has a positive probability of being the value of T_∞. This finishes the proof. □

We now proceed to prove Lemmas 6.1 and 6.2.

6.1. Proof of Lemma 6.1

Proof of Lemma 6.1. The $p > p_\ell$ case is implied by Theorem 1.1, so we focus on $p \leq p_\ell$, using "$\approx$" to denote a rooted oriented tree isomorphism. We will prove the theorem only for the case $a = \epsilon$. This entails no loss of generality because the joint distribution $\mathcal{W}_a(\cdot), \mathcal{P}(a)$, and $\{\mathcal{B}(an) - \mathcal{B}(a)\}_{n\in\mathbb{N}}$ does not depend on the choice of $a \in \mathbb{N}^*$.

Define the sequence of events

$$B_n \equiv \{\mathcal{W}_n(\mathcal{P}(\epsilon) - \mathcal{B}(n)) \approx S\}, \quad n \in \mathbb{N}. \tag{6.2}$$

Our goal is to show that

$$\mathbb{P}\left(B_n \text{ infinitely often}\right) = 1.$$

If the events B_n were independent, we could apply the Borel-Cantelli Lemma for independent events to prove this statement. Since independence is lacking, we will substitute the events B_n by a sequence of independent events A_n such that

$$\mathbb{P}\left(A_n \text{ i.o. but not } B_n \text{ i.o.}\right) = 0, \tag{6.3}$$

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(A_n\right) = +\infty. \tag{6.4}$$

Because the sequence $\{A_n\}$ consists of independent events, (6.4) implies A_n infinitely often almost surely, which implies (via (6.3)) B_n infinitely often almost surely. Therefore, (6.3) and (6.4) imply the lemma.

We define the sequence A_n as follows

$$A_n \equiv \left\{\forall t \in \left[\frac{1}{2(p-1)n^{p-1}}, \frac{3}{(p-1)n^{p-1}}\right] \;\; \mathcal{W}_n(t) \approx S\right\}, \;\; n \in \mathbb{N}. \tag{6.5}$$

The independence of those events is a consequence of the independence of the processes $\{\mathcal{W}_n(\cdot)\}_{n\in\mathbb{N}}$. Moreover,

$$\mathbb{P}\left(A_n \text{ i.o. but not } B_n \text{ i.o.}\right) \leq \mathbb{P}\left(\mathcal{P}(\epsilon) - \mathcal{B}(n) \notin \left[\frac{1}{2(p-1)n^{p-1}}, \frac{3}{2(p-1)n^{p-1}}\right] \text{ i.o.}\right). \tag{6.6}$$

We *claim* that the event on the RHS of (6.6) has probability 0. To see this, note that

$$\mathcal{P}(\epsilon) - \mathcal{B}(n) = \sum_{j\geq n} X(\epsilon, j)$$

is a sum of independent, rate-$f(j)$ exponentials and that

$$\mathbb{E}\left[\mathcal{P}(\epsilon) - \mathcal{B}(n)\right] = \sum_{j\geq n} \frac{1}{f(j)}.$$

As a result, direct use of Lemma 4.6 and the estimate

$$S_1(n) = \sum_{j=n}^{+\infty} \frac{1}{(j+1)^p} \sim \frac{1}{(p-1)n^{p-1}}, \quad n \gg 1,$$

implies

$$\sum_{j\geq 1} \mathbb{P}\left(\mathcal{P}(\epsilon) - \mathcal{B}(n) \notin \left[\frac{1}{2(p-1)n^{p-1}}, \frac{3}{2(p-1)n^{p-1}}\right]\right) < +\infty.$$

Therefore, the Borel-Cantelli Lemma implies that

$$\mathbb{P}\left(\mathcal{P}(\epsilon) - \mathcal{B}(n) \notin \left[\frac{1}{2(p-1)n^{p-1}}, \frac{3}{2(p-1)n^{p-1}}\right] \text{ i.o.}\right) = 0, \tag{6.7}$$

thereby proving the claim and (via (6.6)) proving (6.3).

It remains to prove (6.4). For this purpose, we will only need a very rough lower bound on the probability of A_n. Consider a *labeling* of the elements of S. That is, pick a finite parent-closed subset of $\mathbb{N}^*$, i.e., an element $\hat{S} \in \mathcal{E}^{\text{fin}}$, that corresponds to a labeling of the vertex set of S as defined in Section 3.2. We assume that $\hat{S}$ is ordered

$$\hat{S} = \{s^{(0)} = \epsilon, s^{(1)}, \dots s^{(\ell)}\} \tag{6.8}$$

in a way such that, for all $1 \leq i \leq \ell$, there is an index $p_i < i$ such that $s^{(p_i)}$ is the parent sequence of $s^{(i)}$. We also define the subsets

$$\hat{S}(i) \equiv \{s^{(0)}, \dots, s^{(i)}\}, \;\; 1 \leq i \leq \ell.$$

The ordering property implies that $\hat{S}(i)$ is also a parent-closed subset of $\mathbb{N}^*$. Now define (for $1 \leq i \leq \ell$, where applicable):

$$\begin{aligned}
t_n &= \frac{1}{2(p-1)n^{p-1}}, & (6.9)\\
T_n &= \frac{3}{2(p-1)n^{p-1}}, & (6.10)\\
C_n(i) &= \left\{s^{(i)} \text{ is the only vertex born in } \{\mathcal{W}(t)\}_{t\in\left[\frac{i-1}{\ell}t_n, \frac{i}{\ell}t_n\right]}\right\}, & (6.11)\\
D_n &= \left\{\text{no vertex is born in } \{\mathcal{W}(t)\}_{t\in[t_n,T_n]}\right\}. & (6.12)
\end{aligned}$$

Clearly,

$$\mathbb{P}\left(A_n\right) \geq \mathbb{P}\left(C_n(1)\right) \times \left\{\prod_{i=2}^{\ell} \mathbb{P}\left(C_n(i) \middle| \bigcap_{j=1}^{i-1} C_n(j)\right)\right\} \times \mathbb{P}\left(D_n \middle| \bigcap_{r=1}^{\ell} C_n(r)\right). \tag{6.13}$$

(In fact, A_n is defined in terms of $\mathcal{W}_n(\cdot)$ rather than $\mathcal{W}(\cdot)$, but in terms of evaluating the probabilities, that does not make any difference since these two processes have the same distribution.) We will lower bound the probabilities on the RHS of the above inequality.

Probability of $C_n(1)$. The probability of $C_n(1)$ is the probability that the birth time of $s^{(1)}$ is $\mathcal{B}(s^{(1)}) \leq t_n/\ell$ and that no other birth occurs in the time interval

$[\mathcal{B}(s^{(1)}), t_n/\ell]$. Conditioning on a value $0 \leq \mathcal{B}(s^{(1)}) = t \leq t_n/\ell$, the time of the next birth in $\mathcal{W}(\cdot)$ is

$$\min\{X(s^{(0)}, 0), X(s^{(1)}, 1)\} =^d \exp(f(0) + f(1)).$$

Hence,

$$\begin{aligned}\mathbb{P}\left(C_n(1) \mid \mathcal{B}(s^{(1)}) = u\right) &= \mathbb{P}\left(\exp(f(0) + f(1)) \geq \frac{t_n}{\ell} - u\right) \\ &\geq e^{-(f(0)+f(1))\frac{t_n}{\ell} - u} \geq e^{-(2f(2))(\frac{t_n}{\ell})}.\end{aligned} \tag{6.14}$$

Moreover,

$$\mathbb{P}\left(\mathcal{B}(s^{(0)}) \leq \frac{t_n}{\ell}\right) = 1 - e^{-f(0)\frac{t_n}{\ell}}.$$

Since $t_n \to 0$ as $n \to +\infty$, it follows that there exist constants $C_1, n_1 > 0$ such that for all $n \geq n_1$

$$\mathbb{P}\left(\mathcal{B}(s^{(0)}) \leq \frac{t_n}{\ell}\right) \geq C_1 t_n.$$

We conclude that

$$\mathbb{P}\left(C_n(1)\right) \geq C_1 t_n \, e^{-2f(2)\frac{t_n}{\ell}}. \tag{6.15}$$

Probability of $C_n(i)$, $2 \leq i \leq \ell$. In this part, we will make use of the Markov property of the continuous-time process (see Remark 4.5). Notice that the conditioned event is defined entirely in terms of $\{\mathcal{W}(s)\}_{0 \leq s \leq (i-1)t_n/\ell}$, whereas $C_n(i)$ is defined entirely in terms of $\{\mathcal{W}(s)\}_{s \geq (i-1)t_n/\ell}$. Moreover, it is also true that inside the event $\cap_{j=1}^{i-1} C_n(j)$

$$\mathcal{W}\left(\frac{i-1}{\ell} t_n\right) = \hat{S}(i-1). \tag{6.16}$$

Therefore, we can apply Remark 4.5 to deduce

$$\mathbb{P}\left(C_n(i) \,\middle|\, \bigcap_{j=1}^{i-1} C_n(j)\right) = \mathbb{P}\left(C_n(i) \,\middle|\, \mathcal{W}\left(\frac{i-1}{\ell} t_n\right) = \hat{S}(i-1)\right). \tag{6.17}$$

For $C_n(i)$ to happen, two conditions must be satisfied:

1. $\mathcal{B}(s^{(i)}) - (i-1)t_n/\ell \leq t_n \ell$. That is, $s^{(i)}$ must be born in the interval

$$[(i-1)t_n/\ell, it_n/\ell].$$

2. No other birth happens in the interval $[(i-1)t_n/\ell, it_n/\ell]$.

Choose a value $0 \leq u \leq t_n/\ell$. We will now bound

$$\mathbb{P}\left(C_n(i) \,\middle|\, \mathcal{W}\left(\frac{(i-1)t_n}{\ell}\right) = \hat{S}(i-1),\ \mathcal{B}(s^{(i)}) = u + \frac{(i-1)t_n}{\ell}\right). \tag{6.18}$$

In this case, note that the rate at which the first birth of a node $a \neq s^{(i)}$ happens in $\hat{S}(i-1)$ is[2]

$$R_{i-1} \equiv \sum_{0 \leq j < i,\ j \neq p_i} f(\mathrm{d}_{\hat{S}(i-1)}(s^{(j)})) \leq (i+1) f(i+1), \tag{6.19}$$

the inequality being justified by the fact that the cardinality of $\hat{S}(i-1)$ is i. The rate of births *after* time $(i-1)t_n/\ell + u$ under the conditioning of (6.18) is

$$T_i \equiv \sum_{0 \leq j \leq i} f(\mathrm{d}_{\hat{S}(i)}(s^{(j)})) \leq (i+1) f(i+1). \tag{6.20}$$

Under the conditioning in (6.18), $C_n(i)$ holds iff no $a \neq s^{(i)}$ is born in the time interval $[(i-1)t_n/\ell, (i-1)t_n/\ell + u]$ and no births happen in $[(i-1)t_n/\ell + u, it_n/\ell]$. By the Markov property of $\mathcal{W}(\cdot)$, these events in different time intervals are independent given $\mathcal{W}((i-1)t_n/\ell + u) = \hat{S}(i)$. Therefore, we can write

$$\begin{aligned}
&\mathbb{P}\left(C_n(i) \,\middle|\, \mathcal{W}\left(\frac{(i-1)t_n}{\ell}\right) = \hat{S}(i-1),\ \mathcal{B}(s^{(i)}) = u + \frac{(i-1)t_n}{\ell}\right) \\
&\quad = \mathbb{P}\left(\exp(R_{i-1}) > u\right) \times \mathbb{P}\left(\exp(T_i) > t_n/\ell - u\right) \\
&\quad = e^{-R_{i-1}u} e^{-T_i(\frac{t_n}{\ell} - u)} \geq e^{-(i+1)f(i+1)\frac{t_n}{\ell}}.
\end{aligned} \tag{6.21}$$

As a result,

$$\begin{aligned}
&\mathbb{P}\left(C_n(i) \,\middle|\, \mathcal{W}\left(\frac{(i-1)t_n}{\ell}\right) = \hat{S}(i-1)\right) \\
&\geq e^{-(i+1)f(i+1)\frac{t_n}{\ell}} \mathbb{P}\left(\mathcal{B}(s^{(i)}) - \frac{(i-1)t_n}{\ell} \leq \frac{t_n}{\ell} \,\middle|\, \mathcal{W}\left(\frac{(i-1)t_n}{\ell}\right) = \hat{S}(i-1)\right).
\end{aligned} \tag{6.22}$$

Now notice that conditioned on $\mathcal{W}((i-1)t_n/\ell)$,

$$\mathcal{B}(s^{(i)}) - \frac{(i-1)t_n}{\ell} =^d \exp(f(\mathrm{d}_{\hat{S}(i-1)}(s^{(p_i)}))).$$

Hence,

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{B}(s^{(i)}) - \frac{(i-1)t_n}{\ell} \leq \frac{t_n}{\ell} \,\middle|\, \mathcal{W}\left(\frac{(i-1)t_n}{\ell}\right) = \hat{S}(i-1)\right) \\
&\qquad = 1 - e^{-f(\mathrm{d}_{\hat{S}(i-1)}(s^{(p_i)}))\frac{t_n}{\ell}} \geq 1 - e^{-f(\ell)\frac{t_n}{\ell}}.
\end{aligned} \tag{6.23}$$

[2]This is the rate *until some birth happens*, whether it is the birth of a or of some $s^{(i)} \neq a$.

To state our bound for the probability $C_n(i)$, we note that $t_n \to 0$ as $n \to +\infty$, and therefore there exist constants $C_i, n_i > 0$ depending only on ℓ and f such that for all $n \geq n_i$

$$\mathbb{P}\left(C_n(i) \left| \mathcal{W}\left(\frac{(i-1)t_n}{\ell}\right) = \hat{S}(i-1)\right.\right) \geq e^{-(i+1)f(i+1)\frac{t_n}{\ell}} C_i t_n. \tag{6.24}$$

Probability of D_n. For this bound, we again use the Markov property of $\mathcal{W}(\cdot)$. Notice that whereas D_n is only defined in terms of $\{\mathcal{W}(t)\}_{t \geq t_n}$, the definition of $\cup_{i \leq \ell} C_n(i)$ only depends on $\{\mathcal{W}(t)\}_{0 \leq t \leq t_n}$. Moreover, inside the latter event, $\mathcal{W}(t_n) = \hat{S}$. We can then apply Remark 4.5 to conclude that

$$\mathbb{P}\left(D_n \left| \bigcup_{i=1}^{\ell} C_n(i)\right.\right) = \mathbb{P}\left(D_n \left| \mathcal{W}(t_n) = \hat{S}\right.\right). \tag{6.25}$$

Under this last conditioning, the rate of new births in $\mathcal{W}(t_n)$ is

$$\sum_{s \in S} f(\mathrm{d}_S(s)) \leq (\ell+1) f(\ell+1),$$

and the probability that none of those births occur in $[t_n, T_n]$ is precisely

$$\mathbb{P}\left(D_n \left| \mathcal{W}(t_n) = \hat{S}\right.\right) = e^{-(T_n - t_n)\sum_{s \in S} f(\mathrm{d}_S(s))} \geq e^{-(T_n - t_n)(\ell+1)f(\ell+1)}. \tag{6.26}$$

Wrapping up. To finish this proof, we plug (6.15), (6.24), and (6.26) into (6.13), letting $n \geq \max\{n_i : 1 \leq i \leq \ell\}$ and $C = C_1 C_2 \ldots C_\ell$ to get

$$\begin{aligned} \mathbb{P}(A_n) &\geq C\left(\prod_{i=1}^{\ell} e^{-(i+1)f(i+1)\frac{t_n}{\ell}}\right) \times e^{-(T_n - t_n)(\ell+1)f(\ell+1)} \times t_n^{\ell} \quad (6.27) \\ &\geq C \exp(-(\ell+1)f(\ell+1)T_n) t_n^{\ell}. \quad (6.28) \end{aligned}$$

For ℓ fixed, $n \to +\infty$, we deduce (using the definition of t_n and T_n in (6.9) and (6.10))

$$\mathbb{P}(A_n) = \Omega\left(t_n^{\ell}\right) = \Omega\left(n^{-(p-1)\ell}\right).$$

Now the assumption $p \leq p_\ell$ comes into play, for it implies that $(p-1)\ell \leq 1$. As a result,

$$\sum_{n \geq 1} \mathbb{P}(A_n) = \sum_{n \geq 1} \Omega\left(n^{-(p-1)\ell}\right) = +\infty.$$

This proves (6.4) and finishes the proof. □

6.2. Proof of Lemma 6.2

Proof of Lemma 6.2. Let T have $u+1$ vertices, and let v have r children in T. Let $(\mathrm{T}, v)+n$ denote T with n additional children added to v. Asymptotically in n we consider the probability that T_∞ is not isomorphic to $\textsc{Glue}(\mathrm{T}, v, k)$ conditional on the GN process reaching $(\mathrm{T}, v)+n$. Each $w \in \mathrm{T}$, $w \neq v$, has probability $o(1)$ of having a child before v explodes. Each of the n additional children of v has probability $O(n^{k(1-p)})$ of having k (or more) descendants before v explodes. For $i > n+l$ the ith child of v has k (or more) descendants before v explodes with probability $O(i^{k(1-p)})$. The total probability of any of these events occurring is then bounded from above by $u \cdot o(1) + n \cdot O(n^{k(1-p)}) + \sum_{i>n+l} O(i^{k(1-p)})$, which is $o(1)$ because $k(1-p) < -1$. Therefore, we can find an explicit n so that this probability is less than, say, $\frac{1}{2}$.

With positive (perhaps small) probability the first $n+l$ steps of the GN process yield $(T, v)+n$. Then, with probability at least $\frac{1}{2}$, the final T_∞ is $\textsc{Glue}(\mathrm{T}, v, k)$ as desired. □

7. Conclusion

The two main theorems of this paper completely characterize the limits of the super-linear GN process. Some of their consequences are the fact that the tree T_∞ has finite height (and thus the finite-time GN trees have bounded height) and the fact that the nodes of in-degree $\leq k-1$ (where $p > p_k$) are all but finitely many. However, these characteristics raise many interesting questions about *distributions* of these quantities. For instance, what does the tail of the height distribution of T_∞ look like? We believe that the methods presented in this paper might be sharpened to prove this and other results.

There are many more open questions about the $p < 1$ case of GN. The authors of [Chung et al. 03] have derived some results on their modified model for this range of p under the assumption that certain limits exist. Proving unconditional results of this nature for the GN model remains an important open problem that is also potentially amendable to treatment by our techniques, since the exponential embedding applies to any attachment kernel.

It would also be quite interesting if the exponential embedding could be used to prove known and new properties of related network models, in particular the original Barabási-Albert preferential attachment model. The rigorous version of the process defined in [Bollobás and Riordan 04] is essentially the GN process defined in our paper with attachment kernel $f(x) = x+1$, and it could be the case that the embedding method is a viable technical alternative to the "linearized chord diagrams" of [Bollobás and Riordan 04].

A. Appendix—Proofs of Technical Lemmas

Proof of Lemma 4.6. We will only prove the first inequality, for the proof of the second one is very similar. The technique that we employ is fairly standard and is commonly used in other proofs of Chernoff-type large deviation inequalities [Alon and Spencer 00]. Let $A_n = \sum_{j \geq n} X_j - f(j)^{-1}$. Fix any $0 < s \leq (n+1)^p/2$, and notice that, by the standard Bernstein's trick, the formulae in Section 2.1, the inequality "$1 + x \leq e^x$", and some simple calculations,

$$\begin{aligned}
\mathbb{P}\left(A_n > \delta\right) &= \mathbb{P}\left(e^{s\,A_n} > e^{s\,\delta}\right) \\
&\leq e^{-s\,\delta}\mathbb{E}\left[e^{\sum_{j\geq n} s\left(X_j - \frac{1}{f(j)}\right)}\right] \\
&= e^{-s\,\delta}\prod_{j\geq n}\mathbb{E}\left[e^{s\left(X_j - \frac{1}{f(j)}\right)}\right] \\
&= e^{-s\,\delta}\prod_{j\geq n}\frac{e^{-\frac{s}{f(j)}}}{1-\frac{s}{f(j)}} \\
&= e^{-s\,\delta} \\
&\quad \times \prod_{j\geq n} e^{-\frac{s}{f(j)}}\left(1 + \frac{s}{f(j)} + \frac{s^2}{f(j)^2}\frac{1}{1-\frac{s}{f(j)}}\right) \\
&\leq e^{-s\,\delta}\prod_{j\geq n}\exp\left(2\frac{s^2}{(j+1)^{2p}}\right) \\
&\leq \exp\left(\frac{2s^2}{(2p-1)n^{2p-1}} - s\delta\right).
\end{aligned}$$

To finish the proof, we set $s \equiv n^{p-1/2}$, which is permissible since $n^{p-1/2} \leq (n+1)^p/2$ for all large enough n. □

Proof of Lemma 4.7. To begin with, we note that

$$\forall \epsilon > 0 \prod_{i=1}^{k}\mathbb{P}\left(Y_i \leq \frac{\epsilon}{k}\right) \leq \mathbb{P}\left(Y \leq \epsilon\right) \leq \prod_{i=1}^{k}\mathbb{P}\left(Y_i \leq \epsilon\right)$$

and therefore the assumptions imply the existence of a constant C_0 depending only on the distributions of the Y_i and on k such that

$$\forall \epsilon > 0, \quad \frac{\epsilon^k}{C_0} \leq \mathbb{P}\left(Y \leq \epsilon\right) \leq C_0\epsilon^k.$$

We now use the notation and results in the proof of Lemma 4.6 with $\delta = n^{3/4-p}$.

Then,

$$\forall n \geq n_0,\ \mathbb{P}\left(\left|\sum_{j\geq n} X_j - \mu\right| > \delta\right) \leq 2C_1\, e^{-n^{\frac{1}{4}}}$$

for some constant C_1. Then,

$$\mathbb{P}\left(Y \leq \mu - \delta\right) - \mathbb{P}\left(\left|\sum_{j\geq n} X_j - \mu\right| > \delta\right)$$
$$\leq \mathbb{P}\left(Y \leq \sum_{j\geq n} X_j\right) \leq \mathbb{P}\left(Y \leq \mu + \delta\right) + \mathbb{P}\left(\left|\sum_{j\geq n} X_j - \mu\right| > \delta\right),$$

and by the previous bounds

$$\frac{(\mu-\delta)^k}{C_0} - 2C_1\, e^{-n^{\frac{1}{4}}} \leq \mathbb{P}\left(Y \leq \sum_{j\geq n} X_j\right) \leq C_0(\mu+\delta)^k + 2C_1\, e^{-n^{\frac{1}{4}}}.$$

The result now follows from the fact that, as $n \to +\infty$,

$$\mu \sim \frac{1}{(p-1)n^{p-1}} \sim \mu \pm \delta \gg e^{-n^{\frac{1}{4}}}.$$

□

Proof of Lemma 4.8. $\mathbb{P}\left(Z \leq \lambda\right)$ is equal to the probability that there are at least k arrivals up to time λ in a Poisson process with rate 1. This has a Poisson distribution with rate λ; hence, we have the exact result

$$\mathbb{P}\left(Z \leq \lambda\right) = \sum_{j=k}^{+\infty} e^{-\lambda}\frac{\lambda^j}{j!}.$$

The upper bound follows from

$$\sum_{j\geq k}\frac{\lambda^j}{j!} = \frac{\lambda^k}{k!}\sum_{\ell\geq 0}\frac{\lambda^\ell}{(k+1)(k+2)\ldots(k+\ell)} \leq \frac{\lambda^k e^\lambda}{k!}.$$

□

Acknowledgments. We thank Eleni Drinea and Michael Mitzenmacher for bringing this problem to our attention and for useful discussions. We also thank the anonymous referees for pointing out several typos and making suggestions that greatly improved our presentation. Work for this paper was done while the first author was a PhD student at the Courant Institute, New York University, supported by a fellowship from CNPq, Brazil.

References

[Albert and Barabási 02] Réka Albert and Albert-László Barabási. "Statistical Mechanics of Complex Networks." *Reviews of Modern Physics* 74 (2002), 47–97. Available at `cond-mat/0106096`.

[Albert et al. 99] Réka Albert, Haowoong Jeong, and Albert-László Barabási. "Diameter of the World Wide Web." *Nature* 401 (1999), 130–131.

[Alon and Spencer 00] Noga Alon and Joel Spencer. *The Probabilistic Method*, Second edition. Wiley-Interscience Series in Discrete Mathematics and Optimization. New York: Wiley-Interscience, 2000.

[Barabási and Albert 99] Albert-László Barabási and Réka Albert. "Emergence of Scaling in Random Networks." *Science* 286 (1999), 509–512.

[Bianconi and Barabási 01] Ginestra Bianconi and Albert-László Barabási. "Bose-Einstein Condensation in Complex Networks." *Physical Review Letters* 86 (2001), 5632-5635.

[Bollobás and Riordan 03a] Béla Bollobás and Oliver Riordan. "Mathematical Results on Scale-Free Random Graphs." In *Handbook of Graphs and Networks*, edited by Stefan Burnholdt and Heinz Georg Schuster, pp. 1–34. Berlin: Wiley-VCH, 2003.

[Bollobás and Riordan 03b] Béla Bollobás and Oliver Riordan. "Robustness and Vulnerability of Scale-Free Random Graphs." *Internet Mathematics* 1:1 (2003), 1–35.

[Bollobás and Riordan 04] Béla Bollobás and Oliver Riordan. "The Diameter of a Scale-Free Random Graph." *Combinatorica* 4 (2004), 5–34.

[Bollobás et al. 00] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tárdos. "The Degree Sequence of a Scale-Free Random Graph Process." *Random Structures and Algorithms* 18:3 (2000), 279–290.

[Bollobás et al. 03] Béla Bollobás, Christian Borgs, Jennifer Chayes, and Oliver Riordan. "Directed Scale-Free Graphs." In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 132–139. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2003.

[Chung et al. 03] Fan Chung, Shirin Handjani, and Doug Jungreis. "Generalizations of Pólya's Urn Problem." *Annals of Combinatorics* 7:2 (2003), 141–153.

[Cooper and Frieze 03] Colin Cooper and Alan Frieze. "On a General Model of Web Graphs." *Random Structures and Algorithms* 22 (2003), 311–335.

[Davis 90] Burgess Davis. "Reinforced Random Walk." *Probability Theory and Related Fields* 84:2 (1990), 203–229.

[Dorogovtsev and Mendes 02] Serguei N. Dorogovtsev and José F. F. Mendes. "Evolution of Networks." *Advances in Physics* 51 (2002), 1079–1187. Available at `cond-mat/0106144`.

[Drinea et al. 01] Eleni Drinea, Mihaela Enachescu, and Michael Mitzenmacher. "Variations on Random Graph Models of the Web." Harvard Technical Report TR-06-01, 2001.

[Drinea et al. 02] Eleni Drinea, Alan Frieze, and Michael Mitzenmacher. "Balls in Bins Models with Feedback." In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 308–315. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2002.

[Khanin and Khanin 01] Kostya Khanin and Raya Khanin. "A Probabilistic Model for the Establishment of Neuron Polarity." *Journal of Mathematical Biology* 42:1 (2001), 26–40.

[Krapivsky and Redner 01] P. L. Krapivsky and Sidney L. Redner. "Organization of Growing Random Networks." *Physics Reviews E* 63 (2001), 066123-1–066123-14. Available at `cond-mat/0011094`.

[Mitzenmacher et al. 04] Michael Mitzenmacher, Roberto Oliveira, and Joel Spencer. "A Scaling Result for Explosive Processes." *Electronic Journal of Combinatorics* 11:1 (2004), R31.

[Spencer and Wormald 04] Joel Spencer and Nicholas Wormald. *Explosive Processes.* Manuscript, 2004.

Roberto Oliveira, IBM T. J. Watson Research Laboratory, Yorktown Heights, NY 10598 (riolivei@us.ibm.com)

Joel Spencer, Courant Institute, New York University, New York, NY 10012 (spencer@cims.nyu.edu)

Received March 9, 2004; accepted April 11, 2005.