On a preferential attachment and generalized Pólya's urn model

Andrea Collevecchio*
[collevec@unive.it]
Dept. of Management
Università Ca' Foscari Venezia

Codina Cotar
[cotar@ma.tum.de]
Zentrum Mathematik
Technische Universitaet Muenchen

Marco LiCalzi
[licalzi@unive.it]
Dept. of Management
Università Ca' Foscari Venezia

(April 3, 2012)

Abstract. We study a general preferential attachment and Pólya's urn model. At each step a new vertex is introduced, which can be connected to at most one existing vertex. If it is disconnected, it becomes a pioneer vertex. Given that it is not disconnected, it joins an existing pioneer vertex with probability proportional to a function of the degree of that vertex. This function is allowed to be vertex-dependent, and is called the reinforcement function. We prove that there can be at most three phases in this model, depending on the behavior of the reinforcement function. Consider the set whose elements are the vertices with cardinality tending a.s. to infinity. We prove that this set either is empty, or it has exactly one element, or it contains all the pioneer vertices. Moreover, we describe the phase transition in the case where the reinforcement function is the same for all vertices. Our results are general, and in particular we are not assuming monotonicity of the reinforcement function.

Finally, consider the regime where exactly one vertex has a degree diverging to infinity. We give a lower bound for the probability that a given vertex ends up being the leading one, i.e. its degree diverges to infinity. Our proofs rely on a generalization of the Rubin construction given for edge-reinforced random walks, and on a Brownian motion embedding.

AMS 2010 subject classification: Primary: 05C80, 90B15; Secondary: 60C05

Keywords: preferential attachment, reinforcement processes, species sampling sequence, Pólya's urn process

1 Introduction

1.1 Setting and motivation

We study the following model. Given finitely many classes (or groups) each containing a given initial number of members, new members arrive one at a time. For each new member arriving at time n, with probability $s_n \geq 0$ we create a new class in which we place the member; with probability $1 - s_n$, we place the member in an existing class. We assume that each existing class attracts new members with probability proportional to a certain positive function of the cardinality of the group, called the *reinforcement* or *weight scheme* f. If the groups are allowed to have different reinforcement schemes, then we show that looking at the asymptotics as time tends to infinity we have exactly three different regimes: one group is infinite and all the others are finite; all groups are infinite; all groups are finite. Our main result, Theorem 1.4, shows that in the first regime the process will eventually create a unique infinite group: this happens when each group is reinforced quite a bit, but not too much with respect to the other groups. In the second regime, the cardinality of each group goes to infinity. Finally, in the last regime, all the groups will be finite; what happens is that the process creates various peaks: in the beginning one group dominates the others, but sooner or later another group will start dominating, and this change happens infinitely many times. In this way, no group dominates definitively the other groups. This is a kind of 'there is always a faster gun' principle.

Our model is a generalization of two models from two different classes: one model from the class of *preferential attachment models*, as introduced in [13] and in [17], and one model from the class of *reinforcement processes*, as introduced in [5].

The first main model we are generalizing was introduced and studied independently in [13] and in [17], and later studied in more detail in [21] and [26]. This model is part of the class of *preferential attachment models*, which are models of growing networks, and which were first proposed in the highly-influential papers [1] and [2]. In [1] new vertices arrive at the network one at a time and send a fixed number m of edges to already existing vertices; the probability that a new vertex is linked to a given existing vertex is proportional to the in-degree of the respective existing vertex. Here, the in-degree of a vertex is the number of children of that vertex.

The model studied in [13], [17], [21] and [26] is as follows: consider a model of an evolving network in which new vertices arrive one at a time, each connecting by an edge to a previously existing vertex with a probability proportional to a function f of the existing vertex's in-degree. This function f is called *attachment rule*, or *weight function*, and it determines the existence of two main different regimes. The first regime corresponds to $f(j) = j + 1$ and it was studied in [1], [2] and [26]; the second regime corresponds to for $\gamma < 1$ and it was studied in [26]. The third regime corresponds to $f(j) = (j + 1)^\gamma$ for $\gamma > 1$, and it was studied in [21]. In the first two regimes, it is shown that the degrees of all vertices grow to infinity; in the third regime there is a second phase as one vertex eventually dominates all other vertices.

In the first regime, the so-called Pòlya urn, the urn process is exchangeable and is the only case where exchangeability appears (see [15]). (For more results on preferential attachment models, see the survey [3].)

Preferential attachment models have been motivated by real-life problems, especially in regards to network and internet applications. One important example of growing networks is the World Wide Web, in which the more popular a page (or vertex) is the more hits it receives; a similar principle applies to social interaction or to citation networks. Another example is the one of users of a software program who can report bugs on a website. Bugs with the highest number of requests get priority to be fixed. If the user cannot find an existing report of the bug, they can create a new report. However, it could be that there are duplicate reports, in which case the number of requests is split between the reports, making it less likely that the bug the user found will get fixed. Since bugs that have more requests appear higher up the search results, the user is more likely to add a request to an existing report than to a new one.

This can be explained by the fact that such networks are built dynamically and that new vertices prefer to attach themselves to existing popular vertices with high in-degree rather than to existing unpopular vertices with low in-degree.

The second main model we are generalizing is studied in [5], [20] and [28]. It is known as the *generalized Pólya's urn process*, it belongs to the class of *reinforcement processes*, and can be described as follows. Given finitely many bins each containing one ball, new balls arrive one at a time. For each new ball, with probability $p \geq 0$ we create a new bin in which we place the ball; with probability $1-p$, we place the ball in an existing bin. The probability that the ball is placed in an already existing bin is proportional to $f(j) = j^\gamma$, where j is the number of balls in that bin. The case with $p = 0$ and $\gamma = 1$ is the well-known *Pólya's urn problem.* For $p = 0$ and $\gamma > 0$ no new bins are created, and the process is called a *finite Pólya process with exponent* γ. If $p > 0$ then the process is called an *infinite Pólya process.* Similarly to the preferential attachment models, for generalized Pólya urn processes with $f(j) = j^\gamma$, it is known that for $\gamma \leq 1$ the number of balls in all bins eventually grows to infinity, whereas for $\gamma > 1$ one bin eventually comes to dominate all other bins. (A detailed review of a number of other interesting results on Pólya's urn processes and on reinforcement processes in general is provided in the survey [22].)

The generalized Pólya's urn process has applications to many areas. We briefly mention one such application to biology; for an extensive overview of other applications of generalized Pólya urn processes to reinforced random walks, statistics, computer science, clinical trials, biology, pshychology and economics, see for example Chapter 4 in [28].

The generalized Pólya's urn process with $p = 0$ is used in [9] and [27] to study a real-life application; the reinforcement scheme used in these papers is set to $f(j) = j^\gamma$, with $\gamma > 0$, and real-life data are compared against different values of γ and initial configurations. More precisely, the authors study a colony of ants, which explores a chemically unmarked territory randomly, starting from the nest. The exploration is done on a fixed number k of paths of

various lengths. Each ant passes along one of the k paths leaves a pheromone mark and in this way infuences the following ant's decision in choosing a particular path. This decision is also influenced by whether the paths of various lengths are discovered at the same time, or whether they are discovered at different times. In the real-life experiment it is noticed in the case of paths of equal lengths that, after initial fluctuations, one of the paths becomes more or less completely preferred to the others.

We will show in our paper that the above two models, belonging to these two different areas, are in fact closely related because they are both special cases of our much more general model. The first of our results, Theorem 1.1, proved for our general model, unifies the two above-described phase transition results for a very general class of weight functions f; the result holds in particular both for preferential attachment processes and for generelized Pólya's urn processes. It is worth noting that our condition on the weight function is much weaker than all previously-proved results for the models we generalize. Moreover, in our main result, Theorem 1.4, we show, under no assumptions on the weight function, that we can have only three possible phases; in the third phase, all groups (respectively vertices, bins) stay finite as time tends to infinity. To the best of our knowledge, this is the first time when a third regime as described in our Theorem 1.4, has been proved for any model of preferential attachment or Pólya's urn type. In the case of weight functions f which give rise to the second phase, we devise in our Theorem 1.6, and respectively in Corollary 1.7, a test for obtaining an upper bound, and respectively a lower bound, on the probability that a given group ends up being dominant.

The motivation for our model comes from the class of *species sampling sequences*, to which class our model belongs. Species sampling sequences are models for exchangeble sequences (X_n) with a prediction rule, that is, a formula for the conditional distribution of X_{n+1} given $X_1, X_1, \ldots, X_n$ for $n = 1, 2, \ldots, n$. More precisely, given the first n terms of the sequence (X_n), X_{n+1} equals the i-th distinct value observed so far with probability $p_{n+1,i}$, for $i = 1, 2 \ldots n$, and otherwise X_{n+1} is a new value with distribution ν for some probability measure ν. Species sampling sequences were first introduced and studied in [23], [24] and are now used extensively in Bayesian nonparametric statistics. (See for example [14], [16] or [18] for more on species sampling sequences or for their applications to statistics.)

We next introduce precisely our model.

1.2 The main model

We consider the following model where at each step a new vertex and at most one new edge appear according to the following rules. The probability that the new vertex is disconnected is positive and may change in time. When a vertex is disconnected from the existing ones, it becomes a *pioneer vertex*. We label the pioneer vertices in order of appearance. Given that the new vertex is connected to an existing one, the latter is chosen with probability proportional to a reinforcement scheme of its degree. The graph formed with this procedure

is the union of trees. Each tree has a pioneer vertex as a root. The tree with root i observed at time n, is called the i-th group (or i-th component) by time n.

More formally, fix a collection of positive functions $f_k\colon \mathbb{N} \to \mathbb{R}^+$ with $f_k(0) = 0$ and $f_k(i) > 0$ for all $i, k \geq 1$, and a sequence $\{s_n\}$ which takes values in $[0, 1]$. Set $A_1(1) = 1$ and $A_j(1) = 0$ for all $j \geq 2$. Set $L_1 = 1$. We define the random variables $A_i(n+1)$ and L_{n+1} recursively as follows

$$\begin{aligned}
\mathbb{P}\big(A_k(n+1) = A_k(n) + 1 \mid L_n,\, \{A_j(n), \text{with } j \in \mathbb{N}\}\big) &= \big(1 - s_n\big)\frac{f_k(A_k(n))}{\sum_{s=1}^{L_n} f_s(A_s(n))}, \quad \text{for } i \leq L_n, \\
\mathbb{P}(A_{L_n+1}(n+1) = 1 \mid L_n,\, \{A_j(n), \text{with } j \in \mathbb{N}\}\big) &= s_n,
\end{aligned}$$

while $A_j(n+1) = 0$ for all $j > L_n + 1$. Moreover,

$$L_{n+1} \stackrel{def}{=} \max\{j \geq 1\colon A_j(n+1) \geq 1\}.$$

Notice that $A_i(n+1) - A_i(n) \in \{0, 1\}$ and exactly for one index i this difference equals 1. The random variable $A_i(n)$ is the cardinality of the i-th group by time n. We call the process $\{A_i(n), i, n \geq 1\}$ a *generalized attachment model* whose parameters are the sequence $\{s_n\}$ and the reinforcement functions $\{f_k\}$, abbreviated with GAM($\{f_k\}, \{s_n\}$). We emphasize the fact that we do not make any assumptions on the update functions $\{f_k\}$, other than positiveness, and $\{s_n\}$ is allowed to be random. As shown in Theorem 1.4, some of our strongest results hold for a group-dependent deterministic reinforcement scheme $\{f_j\}$, that is, where each group j follows its own reinforcement scheme f_j, independently of the other groups. From the point of view of applications, this allows one to take into account the case where different groups have different update schemes, which is what would be expected in many real-life situations. We use the symbol GAM($f, \{s_n\}$) to denote a generalized attachment model where the update functions f_k are equal to the positive function f for each $k \geq 1$.

We briefly discuss next the link of our work to the recent literature. The two main models that we generalize were studied in detail in the particular case with reinforcement scheme proportional to $f(j) = j^\gamma$, where $\gamma > 0$.

Let us look first at the literature on preferential attachment models connected to our generalized attachment model. The preferential attachment model studied in [13], [17], [21] and [26] is just GAM($f, \{s_n\}$) for the particular case of $s_n = m(n)c/\big(\sum_{s=1}^{L_n} f(A_s(n)) + m(n)c\big)$, where we denoted by $m(n) \leq n$ the number of groups (respectively vertices) with no children at time n, and where $c > 0$. Then in the growing network, $A_j(n) \geq 1$ represents the in-degree at time n of existing vertex j with strictly positive in-degree, that is, vertex j has $A_j(n)$ children. With probability $f(A_j(n))/\big(\sum_{s=1}^{L_n} f(A_s(n) + m(n)c\big)$, a new arriving vertex attaches to an existing vertex j with strictly positive in-degree $A_j(n)$; with probability $m(n)c/\big(\sum_{s=1}^{L_n} f(A_s(n)) + m(n)c\big)$, a new arriving vertex attaches to one of the existing $m(n)$

vertices with 0 in-degree. For the case where the reinforcement function f is linear or super-linear, $\{s_n\}$ is bounded away from 1, so we can apply our results to the case of preferential attachment models.

In [21] the authors look at the preferential attachment model with reinforcement scheme $f(j) = (j+1)^\gamma, \gamma > 1$, for which they prove a similar result to our Theorem 1.1 by using the original Rubin construction. In [4] and [19], respectively in [26], the authors give the limiting degree distribution for a wide range of linear, respectively sub-linear, weight functions.

A different preferential attachment model was studied in [10] and [11]. In this model a new vertex arrives at each step and attaches to every existing vertex independently with a probability proportional to a concave weight function f of the existing vertex's degree. In [10] the authors prove in Theorem 1.5 the same type of phase transition as in our Theorem 1.1, and they study the degree distribution. In [11] they study the existence of a giant component, that is, of a connected component containing a positive fraction of all vertices.

We turn now to the literature on the generalized Pólya's urn model. This model corresponds to $\mathrm{GAM}(f, \{s_n\})$ in the particular case with $s_n \equiv p$ for all $n \geq n_0$, for some fixed $n_0 \in \mathbb{N}$. In [5] the authors consider both the generalized Pólya's urn model with $p = 0$, when the number of bins is fixed, and with $p > 0$, and they prove by combinatorics techniques a similar result to our Theorem 1.1 for the case of power functions. The case with $p > 0$ and $\gamma < 1$ is studied in [5] under two additional assumptions involving the power function $f(j) = j^\gamma$, assumptions whose validity is left as an open problem in that paper.

The generalized Pólya's urn model with $p = 0$ was also the main object of study in [20] and [28]. In [20] the author studied the case of two fixed bins under a number of technical assumptions on the function f, which exclude for example the (super)-exponential functions, and which assumptions are stated in Section 4 of that paper. Theorem 3.3.1 in [28] proves a result similar to our Theorem 1.1 i) for the case of a fixed number m of bins and under the assumption of monotonicity on the super-linear function f.

Last, we provide below a definition of species sampling sequences and why $\mathrm{GAM}(f, \{s_n\})$ is such a sequence. Consider a Polish space $\mathcal{X}$ and let $\mu(\cdot)$ be a diffuse probability measure on $\mathcal{X}$, i.e. $\mu(\{x\}) = 0$, for all $x \in \mathcal{X}$. Denote with $\mathbb{1}_A$ the indicator function of the event A. A sequence of random variables X_n, with $n \geq 1$, on $\mathcal{X}$ which has the following distribution

$$\mathbb{P}(X_{n+1} \in B \mid X_1, \ldots, X_n) = \sum_{i=1}^{n} p_{n+1,i} \mathbb{1}_{\{X_i \in B\}} + r_{n+1}\mu(B), \tag{1.1}$$

is called a *species sampling sequence* whenever $r_n + \sum_i p_{n,i} = 1$, $r_n, p_{n,i} \geq 0$, and $r_n, p_{n,i}$ are $\mathcal{F}_{n-1}$ measurable, where $\mathcal{F}_n = \sigma\{X_1, X_2, \ldots, X_n\}$. It corresponds to $\mathrm{GAM}(f, \{s_n\})$ for the case with $p_{n,i} = f_i(A_i(n)) / \sum_{s=1}^{L_n} f_s(A_s(n))$ and $r_n = s_n$ for all $n \geq 1$. In particular, the Blackwell-MacQueen urn scheme, also known as Chinese restaurant process, is a species sampling sequence with the choice $s_{n+1} = p_{n+1,i} = 1/(1+n)$; it corresponds to $\mathrm{GAM}(f, \{s_n\})$ with $f(j) = 1/j$ and $s_n = r_n = 1/n$ for all $n \geq 1$.

In this paper we give a complete characterization of the existing phases for a very general class of update functions, for the case $f_j \equiv f$ and for all non-negative random sequences $(s_n)_{n\in\mathbb{N}}$, with $s_n \leq p < 1$ for all $n \geq 1$. In particular we do not assume any monotonicity on f, and our only assumption on f is for Theorem 1.1 i), and it controls the oscillation of the reinforcement function. Moreover, we prove in Theorem 1.4 that for any group-dependent deterministic reinforcement scheme $\{f_j\}$, where $\{f_j\}$ are only assumed to be positive, we can only have three possible phases. We prove the existence of a third phase, by an example. We emphasize the fact that exactly three phases are admitted for this model.

1.3 Results

The following are our main results

Theorem 1.1 *Consider a GAM($f, \{s_n\}$). Suppose that $s_n \leq p$, for some $p < 1$ and all $n \geq 1$.*

i) If

$$\sum_{n=1}^{\infty} \frac{1}{f(n)} < \infty, \quad \text{and} \quad \sum_{n=1}^{\infty} \exp\Big\{ -c\big(\min_{j\geq n} f(j)\big)\Big\} < \infty \ \ \forall c > 0,$$

then there will be, a.s., exactly one group whose cardinality tends to infinity, all the other groups being finite.

ii) The cardinality of each (created) group tends to infinity a.s. if and only if

$$\sum_{n=1}^{\infty} \frac{1}{f(n)} = \infty.$$

Remark 1.2 *If we remove the hypothesis that s_n is bounded away from one, and suppose that $\sum_{n=1}^{\infty}(1-s_n) < \infty$, then by Borel Cantelli's Lemma there exists a random time N such that for any time $n \geq N$ a new group is formed. Hence the cardinality of each group will remain finite, and only finitely many groups will end up having a cardinality larger than 1. We do not study the case of $\limsup_{n\to\infty} s_n = 1$ and $\sum_{n=1}^{\infty}(1-s_n) = \infty$.*

The following result is a corollary to the proof of Theorem 1.1 i). It generalizes the results contained in [21] about the degree of vertices in the preferential attachment model.

Corollary 1.3 *If the hypothesis of Theorem 1.1 i) hold, then $\lim_{n\to\infty} A_i(n) > 1$ for only finitely many i.*

The following theorem establishes that GAM($\{f_j\}, \{s_n\}$) can have only three possible phases. The Theorem holds true if the f_j are random functions independent of s_n satisfying the conditions of the Theorem almost surely.

Theorem 1.4 *Consider a GAM($\{f_j\},\{s_n\}$). Suppose that $s_n \le p < 1$, for some $p < 1$ and all $n \ge 1$.*

i) If

$$\sum_{n=1}^{\infty} \frac{1}{f_j(n)} < \infty, \qquad \textit{for at least one created group } j \in \mathbb{N},$$

then there will be, a.s., **at most** *one group whose cardinality tends to infinity, all the other groups being finite.*

ii) If

$$\sum_{n=1}^{\infty} \frac{1}{f_j(n)} = \infty, \qquad \textit{for all created groups } j \in \mathbb{N},$$

then either the cardinality of each (created) group tends to ∞, a.s., or each of them will be eventually finite, a.s..

We show in Example 4.1 that for the collection of update functions $f_j(n) = \mathrm{e}^{(j^3+n)}$, the cardinality of each group remains finite, a.s.. The third phase seems to arise only when for fixed n, $j \to f_j(n)$ is an unbounded sequence.

The previous two theorems rely on a novel modification of a well-known tool used in reinforced random walk processes, the Rubin construction, which embeds GAM($\{f_j\},\{s_n\}$). We believe that such a generalized Rubin construction as introduced in our paper could have wider applicability to other preferential attachment models.

In the second part of the paper we are going to estimate the probability that a given group is the leading one. Our first result concerns a *reinforced urn model* . Consider an urn with k white balls and 1 red ball and with reinforcement scheme f. Then if we pick a ball at random, it is white with probability $f(k)/(f(k)+f(1))$, and red with probability $f(1)/(f(k)+f(1))$. Suppose that by the time of the n-th extraction we picked j white balls and $n-j$ red ones. The probability to pick a white ball becomes $f(k+j)/(f(k+j)+f(n+1-j))$. We call the urn with these initial conditions and dynamics a reinforced urn model with parameters k and f (abbreviated RUM(k,f)). Denote by $\mathbb{P}^{(k)}$ the probability measure referring to RUM(k, f). We have the following estimate.

Theorem 1.5 *Fix any $k \ge 1$ and consider a RUM(k,f) with $\sum_{j=1}^{\infty} 1/f(j) < \infty$. We have*

$$\mathbb{P}^{(k)}\big(\textit{only a finite number of white balls are picked}\big) \le \frac{1}{2}\prod_{\ell=1}^{k-1} \frac{f(\ell)F_k}{1+f(\ell)F_k}, \tag{1.2}$$

where $F_k \stackrel{def}{=} \sum_{j=k+1}^{\infty} 1/f(j)$.

The above theorem sheds deeper insight on the evolution of RUM(k, f) and on Theorem 1.1 i): it shows that the leading side in the beginning has a great probability to stay the dominant side. As an example of the power of our bound, take $f(j) = j^2$. In this case, a simple computation gives that

$$\mathbb{P}^{(k)}\big(\text{only a finite number of white balls are picked}\big) \leq \frac{1}{2}\exp\big(-(k-1)+\frac{\pi}{2\sqrt{k+2}}\big).$$

Hence for large initial weights k the white has an overwhelming chance to be the one with cardinality tending to infinity. The estimate in (1.2) improves Theorem 3.6.2 in [28]. Theorem 1.5 should be also compared with Theorem 3 in [20], which is proved under the technical assumptions on the update function f stated in Section 4 of that paper. Note also that the bound above is an improvement on the upper bound which could be obtained in (1.2) by means of a similar reasoning to the one in Proposition 2.1 and Proposition 3.1 from [6]. The cause for this is that the lower/upper bounds in [6] are rough for large initial weights. This is one main reason why the methods there only work for finite graphs and not also for infinite graphs. Our proof is based on an embedding of RUM(k, f) into Brownian motion, and gives robust estimates for all initial weights.

Next we turn again to GAM$(f, \{s_n\})$. Suppose that $\sum_{j=1}^{\infty} 1/f(j) < \infty$. Theorem 1.1 i) guarantees the existence of a unique group whose cardinality goes to infinity. We call this the leading group. Denote by **Lead** the label of the leading group. In other words, **Lead** $= j$ if and only if the leading group is the j-th one. Our goal is to test if a given group, which has a certain advantage on the others, is the leader. We start by giving an upper bound for the tail of **Lead**.

We give the following construction of GAM(f, p). Suppose we have two sequences of random variables, b_n and $t(n)$, satisfying the following. The variables b_n are i.i.d. Bernoulli with mean p, while the variables $t(n)$ are described recursively. We define $A_1^*(1) = 1$, and $A_i^*(1) = 0$ for all $i \geq 2$. Moreover, set $L_1^* = 1$. Denote by $\mathcal{F}_n$ the σ-algebra generated by $\big\{\big(b_i, t(i)\big), \text{ with } i \leq n\big\}$. Suppose we defined $A_i^*(n)$, which is $\mathcal{F}_{n-1}$-measurable. The random variable $t(n)$ can be chosen to have the following distribution

$$\mathbb{P}(t(n) = k \mid \mathcal{F}_{n-1}) = \frac{f_k(A_k^*(n))}{\sum_{s=1}^{L_n^*} f_s(A_s^*(n))}.$$

Moreover, we can choose $t(n)$ to be independent of b_i with $i \geq n+1$. Denote by $L_n^* \stackrel{def}{=} \max\{j \geq 1\colon A_j^*(n) \geq 1\}$. We define

$$A_j^*(n+1) \stackrel{def}{=} \begin{cases} 0 & \forall\, j > L_n^* + 1, \\ 1 & \text{if } b_n = 1, \\ A_j^*(n) + \mathbb{1}_{\{t(n)=j\}} & \text{if } b_n = 0. \end{cases}$$

Finally, let $L_{n+1}^* \stackrel{def}{=} \max\{j \geq 1\colon A_j^*(n+1) \geq 1\}$. We have that $\{A_i^*(n),\ i, n \in \mathbb{N}\}$ is distributed like the process $\{A_i(n),\ i, n \in \mathbb{N}\}$ described in section 1.2. At time n, b_n will

determine if the new vertex is disconnected, and $t(n)$ will determine to which of the existing vertices the new arrived will adhere **if** it is not disconnected. Notice that $t(n)$ is defined also in the case that $b_n = 1$, i.e., in the case that the new vertex is disconnected. We denote by $\xi_1 = 0$ and $\xi_i \stackrel{def}{=} \inf\{n > \xi_{i-1} \colon b_n = 1\}$. In words, ξ_i is the time when the i-th group is formed. We say that the i-th group is generated by the u-th group if $t(\xi_i) = u$, i.e. if we flipped the value of b_{ξ_i} into 0, then the new arrival would have joined the group u. In this case we say that u is the parent of i. Notice that there exists exactly one parent for each integer different from one. We build a random tree $\mathcal{G}$, whose root is one, joining each integer to its parent. We say that a vertex is at level n if its distance from the root is n. Denote by g_n the vertices at level n. Let $G_n = \cup_{j \geq n} g_j$. We have

Theorem 1.6 *Suppose that the assumptions of Theorem 1.1 i) hold. Then*

$$\mathbb{P}(\mathbf{Lead} \in G_n) \leq \inf_{r,M \geq 1} \Big[m^n \mathrm{e}^{-c_n(r,M)n} + r^{-n} + C_1 \exp\big\{ - MC_2 \big\} \Big],$$

where the sequence $c_n(r, M) \to \infty$ as $n \to \infty$, for fixed value of $r \geq 1$ and $M \geq 1$, and $m, C_1, C_2 > 0$. The quantities C_1, C_2 and m are computable. The functions $c_n(r, M)$ are computable for fixed values of r and M.

The following result is a direct consequence of Theorems 1.6 and 1.5.

Corollary 1.7 *Suppose that the assumption of Theorem 1.1 i) hold. Then*

$$\mathbb{P}(\mathbf{Lead} = 1) \geq 1 - \Big(\sum_{k=1}^{\infty} \frac{1}{2} \prod_{\ell=1}^{k-1} \frac{f(\ell) F_k}{1 + f(\ell) F_k} \Big) - \inf_{r,M \geq 1} \Big[m^2 \mathrm{e}^{-2c_2(r,M)} + r^{-2} - C_1 \exp\big\{ - MC_2 \big\} \Big]$$

where the quantities $c_n(r, M)$, m and C_1 and C_2 are the same as Theorem 1.6.

The rest of the paper is structured as follows: in Section 2 we introduce our generalized Rubin construction and give the proof of Theorem 1.1 (i), in Section 3 we give the proof of Theorem 1.1 (ii). In Section 4 we prove our main result Theorem 1.4 and present an example where the third phase occurs, in Section 5 we introduce our Brownian motion embedding and provide the proof of Theorem 1.5. In Section 6 we give the proofs of Theorem 1.6 and of Corollary 1.7. Finally, in the Appendix we give a brief introduction to the Rubin construction, as introduced in [7].

2 Proof of Theorem 1.1 i)

We introduce a modified version of the Rubin construction which fits our model. For a detailed explanation of the original Rubin construction, see for example [6] and [7].

Fix a parameter $p < 1$. We first focus on the case $s_n \equiv p < 1$, i.e. GAM(f, p), then we extend to the more general case $s_n \leq p$ using a coupling. For any set $A \subset \mathbb{R}^+$, let

$$A[n] = \inf\{x \colon \#(A \cap [0, x]) \geq n + 1\},$$

where the infimum of an empty set is ∞. In words, $A[n]$ is the $n+1$-th element of A, ordered from the smallest to the largest. For example, if $A = \{2, 8, 6, 9\}$, then $A[0] = 2$ and $A[1] = 6$, $A[5] = \infty$. Notice also that for the example $A = \{1/j \colon j \geq 1\}$, is not possible to identify the $n + 1$-th element. In fact, in this case, we have that $A[n] = 0$ for all $n \geq 0$.

Notice that $A[n]$ is always a non-decreasing sequence, hence $\lim_{n\to\infty} A[n]$ exists, possibly infinite. For each $i \in \mathbb{N}$, let $\{W_n^{(i)}, n \geq 1\}$ be a sequence of independent exponential(1) random variables, with $n \in \mathbb{N}$. Moreover let $\{R_n^{(i)}, n \geq 1\}$ be a sequence of i.i.d. Bernoulli such that $\mathbb{P}(R_n^{(i)} = 1) = p$. We are going to use these sequences to generate a GAM(f, p). The Bernoullis will be used to create new groups, while the exponentials play a central role in the allocation of new individuals into existing groups. We are assuming that all the variables involved are independent of each other. Set $N_m(1) = 1$, for all $m \geq 1$. Then, for $n \geq 2$, let

$$\begin{aligned} N_m(n) &\stackrel{def}{=} 1 + \#\{j \colon j \leq n-1 \text{ such that } R_j^{(m)} = 0\}, \\ \Xi_1 &\stackrel{def}{=} \{0\} \cup \Big\{ \sum_{i=1}^{n} \frac{W_i^{(1)}}{f\big(N_1(i)\big)} \colon n \geq 1 \Big\} \subset \mathbb{R}^+. \end{aligned}$$

In words, for each $m \geq 1$, the processes $\mathbf{N}_m \stackrel{def}{=} \{N_m(n), n \geq 1\}$ are independent processes with the property that $N_m(n) - 1$ are distributed like binomial with parameters $n - 1$ and $1 - p$, while Ξ_1 is a random subset of $\mathbb{R}^+$ composed by 0 and all the partial sums of the sequence $\big\{W_i^{(1)}/f\big(N_1(i)\big)$, with $i \geq 1\big\}$. To each element Ξ_1 we associate a corresponding Bernoulli as follows. Let $g_1 \colon \Xi_1 \to \{0, 1\}$ be a random function defined by $g_1(\Xi_1[n]) \stackrel{def}{=} R_n^{(1)}$. The elements in Ξ_1 with corresponding Bernoulli equal to one, are used to generate new groups for GAM(f, p). The other ones will *potentially* belong to the first group and will be labelled one. We will clarify the last sentence at the end of the construction. Define

$$\widetilde{\Xi}_1 \stackrel{def}{=} \{0\} \cup \Big\{ \sum_{i=1}^{n} \frac{W_i^{(1)}}{f\big(N_1(i)\big)} \colon n \geq 1 \text{ and } R_n^{(1)} = 0 \Big\},$$

i.e., $\widetilde{\Xi}_1$ is composed of $\{0\}$ and all the points in $\Xi_1 \setminus \{0\}$ with Bernoulli equal to 0. These are the points which do not generate other groups. We label the points in $\widetilde{\Xi}_1$ with 1. Set $\tau_1 = 0$ and define

$$\tau_2 \stackrel{def}{=} \inf\{n \geq 1 \colon R_n^{(1)} = 1\}.$$

The random variable τ_2 is the time when the second group is formed. Given τ_2, let

$$\Xi_2 \stackrel{def}{=} \Xi_1 \cup \Big\{ \Xi_1[\tau_2] + \sum_{i=1}^{n} \frac{W_i^{(2)}}{f(N_2(i))} \colon n \geq 0 \Big\},$$

and

$$\widetilde{\Xi}_2 \stackrel{def}{=} \Big\{\Xi_1[\tau_2] + \sum_{i=1}^{n} \frac{W_i^{(2)}}{f\big(N_2(i)\big)} : \text{ either } n = 0 \text{ or both } n \geq 1 \text{ and } R_n^{(2)} = 0\Big\}.$$

We label the elements of $\widetilde{\Xi}_2$ using 2. Define the function $g_2 \colon \Xi_2 \to \{0,1\}$ as follows. If $\Xi_2[n] = \Xi_1[j]$, for some $j \in \mathbb{N}$, then $g_2(\Xi_2[n]) = R_j^{(1)}$. The latter is well-defined because all the elements of Ξ_1 are a.s. distinct. If $\Xi_2[n] = (\Xi_2 \setminus \Xi_1)[j]$ for some $j \in \mathbb{N}$, then $g_2(\Xi_2[n]) = R_j^{(2)}$. Notice that $\widetilde{\Xi}_1$ and $\widetilde{\Xi}_2$ are disjoint and their union is a proper subset of Ξ_2. Let us describe in words the variables defined so far. The reinforcement plays no role up to time τ_2. The latter random variable is geometrically distributed with mean $1/p$. At time τ_2, the first group has cardinality τ_2, because we count also the point 0, and a second group is formed. The random point $\Xi_2[\tau_2]$ is labelled 2, in fact it belongs to $\widetilde{\Xi}_2$, and it is the smallest point belonging to this random set. The next point on the line, i.e. $\Xi_2[\tau_2+1]$ can have label 1, 2 or no label at this stage. If the latter happens, we label this point with 3. If it belongs to $\widetilde{\Xi}_1$, respectively $\widetilde{\Xi}_2$, its label will be 1, respectively 2. Notice that by the definition of these sets, if $\Xi_2[\tau_2+1] \in \widetilde{\Xi}_1 \cup \widetilde{\Xi}_2$ then $g_2\big(\Xi_2[\tau_2+1]\big)$ must be equal to zero. On the other hand, in the case that $g_2\big(\Xi_2[\tau_2+1]\big) = 1$ then a new group is formed, which is labelled 3. The probability that this happens is p. Next we want to compute the probability that $\Xi_2[\tau_2+1]$ has label 1. We have the following equality

$$\{\Xi_2[\tau_2+1] \in \widetilde{\Xi}_1\} = \Big\{\Xi_1[\tau_2+1] - \Xi_1[\tau_2] < \widetilde{\Xi}_2[1] - \Xi_1[\tau_2]\Big\} \cap \Big\{g_2\big(\Xi_1[\tau_2+1]\big) = 0\Big\}. \quad (2.1)$$

Note that $\Xi_2[\tau_2] = \Xi_1[\tau_2]$. Given τ_2, the two events appearing on the right-hand side of (2.1) are independent, because the first one depends on the exponentials while the second is determined by the Bernoullis. The probability of the second event, conditionally on τ_2, is $1-p$. If the random variable $\Xi_2[\tau_2+1]$ was labelled 1, then it would belong to $\widetilde{\Xi}_1$ and would be equal to

$$\widetilde{\Xi}_1[\tau_2+1] = \sum_{i=1}^{\tau_2+1} W_i^{(1)}/f\big(N_1(i)\big) = \Xi_1[\tau_2] + W_{\tau_2+1}^{(1)}/f\big(N_1(\tau_2+1)\big).$$

If $\Xi_2[\tau_2+1]$ was labelled 2, then it would belong to $\widetilde{\Xi}_2$ and would be equal to $\widetilde{\Xi}_2[1] = \Xi_2[\tau_2] + W_1^{(2)}/f\big(1\big)$. Hence

$$\begin{aligned}\Xi_2[\tau_2+1] &= \Xi_2[\tau_2] + \min\Big(\frac{W_{\tau_2+1}^{(1)}}{f\big(N_1(\tau_2+1)\big)}, \frac{W_1^{(2)}}{f\big(1\big)}\Big)\\ &= \Xi_2[\tau_2] + \min\Big(\frac{W_{\tau_2+1}^{(1)}}{f\big(\tau_2\big)}, \frac{W_1^{(2)}}{f\big(1\big)}\Big),\end{aligned}$$

where we used $N_1(\tau_2+1) = \tau_2$. This last equality comes from the fact that among $R_i^{(1)}$, with $i \leq \tau_2$, the only Bernoulli taking value one is $R_{\tau_2}^{(1)}$. As $N_1(\tau_2+1)$ equals one plus the number

of zeroes among the first τ_2 Bernoulli, it is equal to τ_2. The first event on the right-hand side of (2.1) can be rewritten as

$$\Big\{\frac{W^{(1)}_{\tau_2+1}}{f\big(\tau_2\big)} < \frac{W^{(2)}_1}{f\big(1\big)}\Big\}. \tag{2.2}$$

Given τ_2, the random variable $W^{(1)}_{\tau_2}/f\big(\tau_2\big)$ is exponentially distributed with mean $1/f(\tau_2)$. By a simple integration, we can argue that the probability that, among two independent exponentials, a given one is the smallest is equal to its parameter divided by the sum of the parameters. Hence the probability of the event in (2.2), conditionally on τ_2, is $f(\tau_2)/\big(f(\tau_2)+f(1)\big)$. The probability of the event described in (2.1), conditionally on τ_2, is

$$(1-p)\frac{f(\tau_2)}{f(\tau_2)+f(1)}.$$

We infer that the conditional probability that $\Xi_2[\tau_2+1]$ is labelled 2 is $(1-p)f(1)/(f(\tau_2)+f(1))$. This is consistent with what happens in GAM(f,p).

Suppose we defined $\big(\tau_2, \Xi_1, \widetilde{\Xi}_1, g_1, \ldots \tau_{m-1}, \Xi_{m-1}, \widetilde{\Xi}_{m-1}, g_{m-1}\big)$. Define

$$\tau_m \stackrel{def}{=} \inf\{n > \tau_{m-1} \colon g_{m-1}\big(\Xi_{m-1}[n]\big) = 1\},$$

i.e. the time when the m-th group is formed. Given τ_m let

$$\Xi_m \stackrel{def}{=} \Xi_{m-1} \cup \Big\{\Xi_{m-1}[\tau_m] + \sum_{i=1}^{n} \frac{W^{(m-1)}_i}{f(N_m(i))} \colon n \geq 1\Big\}, \tag{2.3}$$

and

$$\widetilde{\Xi}_m \stackrel{def}{=} \Big\{\Xi_{m-1}[\tau_m] + \sum_{i=1}^{n} \frac{W^{(m-1)}_i}{f(N_m(i))} \colon \text{ either } n=0 \text{ or both } n \geq 1 \text{ and } R^{(m)}_n = 0\Big\}. \tag{2.4}$$

The elements of $\widetilde{\Xi}_m$ are labelled m. Moreover let g_m be defined as follows. If there exists j such that $\Xi_m[n] = \Xi_{m-1}[j]$ then $g_m\big(\Xi_m[n]\big) = g_{m-1}\big(\Xi_{m-1}[j]\big)$. If $\Xi_m[n] = (\Xi_m \setminus \Xi_{m-1})[j]$ for some j, then set $g_m\big(\Xi_m[n]\big) = R^{(m)}_j$.

Denote by $\Xi \stackrel{def}{=} \bigcup_{s=1}^{\infty} \Xi_s$. Each point $x \in \Xi$ belongs, a.s., to exactly one $\widetilde{\Xi}_s$ for some $s \geq 1$, i.e. $\Xi \stackrel{def}{=} \bigcup_{s=1}^{\infty} \widetilde{\Xi}_s$. In our construction, we label the point x with s if and only if $x \in \widetilde{\Xi}_s$. Define the random function $g \colon \Xi \to \{0,1\}$ as follows. If $\Xi[n] = \widetilde{\Xi}_j[s]$ for some (a.s. unique) pair $(j,s) \in \mathbb{N}^2$, then $g\big(\Xi[n]\big) = R^{(j)}_s$. Notice that Ξ can be used to generate a generalized attachment model, as follows. Denote by

$$\widetilde{A}_i(n) = \{j \colon j \leq n,\ \Xi[j] \text{ has label } i\}.$$

Then $\{\widetilde{A}_i(n), \text{ with } i,n \geq 1\}$ is distributed like the process $\{A_i(n), \text{ with } i,n \geq 1\}$ introduced in subsection 1.2. To see this, suppose that in the set $\{\Xi[i], \text{ with } i \leq n\}$ there are exactly ℓ_i

points labelled i, with $\sum_{i=1}^m \ell_i = n$ for some $m \in \mathbb{N}$ satisfying also $\ell_i \geq 1$ for all $i \in \{1, \dots, m\}$. Given this, the probability that $\Xi[n+1]$ is labelled $m+1$, i.e., the probability that $g(\Xi[n+1])$ equals one, is exactly p. Given that $\Xi[n+1]$ is not labelled $m+1$, then the probability that it is labelled j, with $j \leq m$, is exactly

$$\frac{f(\ell_j)}{\sum_{i=1}^m f(\ell_i)}, \tag{2.5}$$

where we used the memoryless property of the exponential random variables. In fact, using this property, given that $\Xi[n+1]$ is not labelled $m+1$, the random variable $\Xi[n+1] - \Xi[n]$ is distributed like the minimum of m exponentials with parameters $f(\ell_s)$, for $1 \leq s \leq m$. The probability that the j-th exponential is the minimum is given exactly by (2.5) through a simple integration. Summarizing, given that in the set $\{\Xi[i], \text{ with } i \leq n\}$ there are exactly ℓ_i points labelled i, with $\sum_{i=1}^m \ell_i = n$ and $\sum_{i=1}^{m-1} \ell_i < n$ for some $m \in \mathbb{N}$, the probability that $\Xi[n+1]$ is labelled j, with $j \leq m$, is

$$(1-p)\frac{f(\ell_j)}{\sum_{i=1}^m f(\ell_i)}.$$

Define

$$x_m^* \stackrel{def}{=} \Xi_{m-1}[\tau_m] + \sum_{i=1}^{\infty} \frac{W_i^{(m)}}{f(N_m(i))}, \tag{2.6}$$

and for any integer $j \geq 1$, set

$$\Xi_j^* \stackrel{def}{=} \Big\{\Xi[\tau_j] + \sum_{s=1}^{n} W_s^{(j)}/f\big(N_j(s)\big) : n \geq 0\Big\}. \tag{2.7}$$

In the next result, we prove that x_m^* is a.s. finite, for any $m \geq 1$. This, together with (2.3) and (2.4), implies that x_m^* is an accumulation point for Ξ_m and $\widetilde{\Xi}_m$. We say that a vertex u is generated by j if $\Xi[\tau_u] \in \Xi_j^*$. Notice that each vertex (different from 1) is generated by exactly one other vertex. Our proof of Lemma 2.1 relies on the construction of a random tree $\mathcal{T}$, built by connecting each vertex to its parent. Notice that this random tree shares the same distribution with $\mathcal{G}$, introduced before Theorem 1.6. Suppose that $\tau_u = t$. If we switched $g(\Xi[t])$ from 1 to 0, we would have that $\Xi[t]$ would have been a point of $\widetilde{\Xi}_j$, hence it would have had label j. Fix $j, n \in \mathbb{N}$. Notice that even if the Bernoulli associated to the point $\Xi_j^*[n]$ equals 1, this point might not be able to generate a child in $\mathcal{T}$ using the exponentials and Bernoulli that have been defined so far. This is the case if $\#\big(\Xi \cap [0, \Xi_j^*[n]]\big) = \infty$, when infinitely many vertices have already been generated by the time we reach $\Xi_j^*[n]$ and all the $(W_n^{(i)}, R_n^{(i)})$ have already been used. This is going to be an important point in the proof of Lemma 2.2.

Lemma 2.1 *The random variables x_m^*, with $m \geq 1$, are almost surely finite.*

Proof. Fix $m \geq 1$. Set $Z_m(0) = 0$ and let $Z_m(i) = \inf\{n\colon N_m(n) = i\}$. Then $Z_m(i) - Z_m(i-1)$, with $i \geq 1$ are geometric$(1-p)$ and are independent of the $W_i^{(m)}$, with $i \geq 1$. If $Z_m(i) \leq k < Z_m(i+1)$, then $f\big(N_m(k)\big) = f(i)$. Hence

$$\begin{aligned} x_m^* &= \Xi[\tau_m] + \sum_{i=1}^{\infty} \sum_{j=Z_m(i)}^{Z_m(i+1)-1} \frac{W_j^{(m)}}{f(N_m(j))} \\ &= \Xi[\tau_m] + \sum_{i=1}^{\infty} \frac{1}{f(i)} \sum_{j=Z_m(i)}^{Z_m(i+1)-1} W_j^{(m)} \end{aligned}$$

As the series in the latter expression is composed by non-negative random variables, it is a.s. finite if its mean is finite. Its mean is exactly

$$\frac{1}{1-p} \sum_{i=1}^{\infty} \frac{1}{f(i)} < \infty. \tag{2.8}$$

To see this, notice that $Z_m(i),\ i \geq 1$, is independent of $W_j^{(m)}$, $j \geq 1$, which implies

$$\mathbb{E}[\sum_{j=Z_m(i)}^{Z_m(i+1)-1} W_j^{(m)}] = \frac{1}{1-p}.$$

Moreover, we have that $\Xi[\tau_m]$ is stochastically smaller than

$$\frac{1}{\min_i f(i)} \sum_{s=1}^{\tau_m} W_1^{(s)}. \tag{2.9}$$

This is because $\Xi[n] - \Xi[n-1]$ is stochastically smaller than an exponential random variable whose mean is smaller than $1/(\min_i f(i))$. Moreover, the random variable τ_m is negative binomial with parameters m and p. This can be checked by induction, in fact τ_1 is geometrically distributed with mean $1/p$. Suppose this is true for τ_{m-1}, then we have to wait an independent geometric(p) to create the next group. Combining this fact with (2.9) we have that $\Xi[\tau_m] < \infty$ a.s.. This, together with (2.8) implies the Lemma. □

In the next result we establish the link between the behavior of the generalized attachment model and the quantity $\inf_i x_i^*$.

Lemma 2.2 *The infimum* $\inf_i x_i^*$*, is a.s. attained, i.e. it is actually a minimum. The minimiser is a.s. unique. Moreover*

$$\lim_{n\to\infty} \Xi[n] = \inf_i x_i^*, \qquad a.s.. \tag{2.10}$$

Proof. We select a random subtree of $\mathcal{T}$, denoted by $\mathcal{T}_1$, as follows. The root of this tree is 1 (i.e., it is identified with the first group). Given that a vertex j belongs to $\mathcal{T}_1$, its offspring will be those vertices u such that

$$\Xi[\tau_u] \in \Xi_j^*, \qquad \text{and } \Xi_u^*[1] < x_j^*. \tag{2.11}$$

Recall that $\Xi[\tau_u] = \Xi_{u-1}[\tau_u] = \Xi_u^*[0]$. We are going to prove the following statement.

$$\text{For any fixed } M\text{, only finitely many of the vertices } u \text{ of } \mathcal{T}_1 \text{ satisfy } \widetilde{\Xi}_u[1] > M. \tag{2.12}$$

Before we prove (2.12) we argue that this statement would imply the Lemma. We need only consider the vertices of $\mathcal{T}_1$. In fact, if j is not a vertex of $\mathcal{T}_1$ then there exists a vertex u such that $\Xi_j[1] > x_u^*$, which implies that $x_j^* > x_u^*$. Hence $x_j^* \neq \inf_i x_i^*$.

If (2.12) holds, then for any M there are only finitely many vertices u in $\mathcal{T}_1$ such that $x_u^* < M$. Hence, as each x_u^* is a.s. finite, we have that $\inf_i x_i^*$ is actually a minimum. Next we prove that the minimizer is a.s. unique. To prove this last statement, we prove that for each $i > j$, we have that x_i^* and x_j^* are a.s. different. To see this, notice that $x_i^* - \Xi[\tau_i]$ only depends on $\{W_n^{(i)}, R_n^{(i)}, \text{ with } n \geq 1\}$. Hence $x_i^* - \Xi[\tau_i]$ is independent of $x_j^* - \Xi[\tau_i]$ which is determined by a disjoint collection of exponentials and Bernoullis. The probability that $x_i^* - \Xi[\tau_i]$ and $x_j^* - \Xi[\tau_i]$ are equal is 0, as they are continuous independent random variables. This is exactly the probability that $x_i^* = x_j^*$. As the set of x_i^*, $i \geq 1$, is countable, x_i^* are all, a.s., distinct.

Next we show that (2.12) implies (2.10). As already mentioned, the sequence $\Xi[n]$ is a.s. non-decreasing, i.e. $\Xi[n+1] \geq \Xi[n]$, a.s.. Hence $\lim_{n\to\infty} \Xi[n]$ a.s. exists. Notice that for each i, x_i^* is the limit of an increasing sequence taking values in Ξ. To see this, notice that

$$\Xi[\tau_i] + \sum_{j=1}^{n} \frac{W_j^{(i)}}{f(N_i(j))} < x_i^*, \qquad \forall n \geq 1,$$

by the definition of x_i^*. Hence infinitely many points labelled i are smaller than x_i^*, yielding $\#\big(\Xi \cap [0, x_i^*]\big) = \infty$. This implies that $\lim_{n\to\infty} \Xi[n] \leq x_i^*$ for each $i \geq 1$, i.e. $\lim_{n\to\infty} \Xi[n] \leq \inf_i x_i^*$. Now we turn to the proof of the other inequality which implies (2.10). Fix $\varepsilon > 0$. It is sufficient to prove that (2.12) implies

$$\#(u\colon \widetilde{\Xi}_u[1] \leq \inf_i x_i^* - \varepsilon) < \infty. \tag{2.13}$$

In fact, if (2.13) holds, only finitely many u satisfy $\#\big(\widetilde{\Xi}_u \cap \big[0, \inf_i x_i^* - \varepsilon\big]\big) > 1$. Denote the set of labels of these groups by B. For each element u of the finite set B, there are only finitely many points of Ξ_u^* which are smaller than $\inf_i x_i^* - \varepsilon$, for otherwise we would have $x_u^* \leq \inf_i x_i^* - \varepsilon$ which would yield a contradiction. Hence, for each element u of B, the set $\Xi_u^* \cap \big[\inf_i x_i^* - \varepsilon\big]$ is finite. For each $j \notin B$ we have that there exists a $u \in B$ such that $\widetilde{\Xi}_j[0] \in \Xi_u^*$. This implies that

$$\Xi \cap \big[0, \inf_i x_i^* - \varepsilon\big] = \bigcup_{u\in B} \Xi_u^* \cap [\inf_i x_i^* - \varepsilon],$$

and the latter is a finite set. Notice that $\inf_i x_i^* - \varepsilon$ could take a negative value, but this is not a problem for our reasoning, as then the set appearing in (2.13) would be empty, and there would be nothing to prove. Next, we prove that (2.12) implies (2.13). Fix a vertex u of $\mathcal{T}_1$. Denote by $u(0) = 1, u(1), u(2), \ldots, u(n)$ the ancestors of u in $\mathcal{T}_1$, i.e. the vertices lying on the unique self-avoiding path connecting u to the root 1. Notice that we do not consider u ancestor of itself. If u satisfies $\widetilde{\Xi}_u[1] \leq \inf_i x_i^* - \varepsilon$, then u belongs to $\mathcal{T}_1$. In fact, $\widetilde{\Xi}_u[1] \geq \Xi^*_{u(i)}[1]$ while $\inf_i x_i^* - \varepsilon < x^*_{u(i-1)}$, for all $i \leq n+1$. Hence, $\Xi^*_{u(i)}[1] \leq x^*_{u(i-1)}$, where we set $u(n+1) = u$. As we are assuming that (2.12) holds, the random tree $\mathcal{T}_1$ has only finitely many vertices j satisfying $\widetilde{\Xi}_j[1] > M$. As $\inf_i x_i^* < \infty$, a.s., we have that (2.13) holds.

Next, we are going to prove (2.12). For any vertex j in $\mathcal{T}_1$, denote by σ_j the number of its offspring. Notice that the σ_j are neither independent nor identically distributed and $\mathcal{T}_1$ is not Galton-Watson tree. To see this, fix $j, n \geq 1$. If there is an infinite number of elements of Ξ to the left of $\Xi^*_j[n]$, i.e.

$$\#\big(\Xi \cap \big[0, \Xi^*_j[n]\big]\big) = \infty, \tag{2.14}$$

then already infinitely many groups have been created. Hence $\Xi^*_u[0] < \Xi^*_j[n]$ for all $u \in \mathbb{N}$. This implies that $\Xi^*_j[n]$ cannot generate any new group in $\mathcal{T}_1$ using the exponentials and Bernoullis defined so far, because they have already been used. To overcome this problem, we create a new tree, larger than $\mathcal{T}_1$, by introducing new random variables which allow also the observations $\Xi^*_j[n]$ satisfying (2.14) to create a new group. To this end, we should attach to each of these observations a new sequence of independent exponentials and independent Bernoullis. For example, if $\Xi^*_j[n]$ satisfies (2.14), and the associated Bernoulli equals one, a new group, that we label ν, is created (notice that we cannot use any of the integers as a label, because they are already all taken). In this case, we set $\widetilde{\Xi}_\nu[0] = \Xi^*_j[n]$. We denote the associated sequence of i.i.d. exponentials with mean 1 by $W_n^{(\nu)}$, and let $R_j^{(\nu)}$ be the Bernoulli associated to group ν. We define Ξ^*_ν and $\widetilde{\Xi}_\nu$ using these random variables, as we did in (2.7) and (2.4). If the group ν satisfies the second requirement in (2.11), i.e. $W_1^{(\nu)}/f(1) < x_j^* - \Xi^*_\nu[0]$, then ν belongs to the new tree $\mathcal{T}_2$ that we are going to define. But then, we would have to allow that ν is able to generate groups as well, in the same fashion. This approach would require that we introduce new sequences of exponentials and Bernoullis, and the notation would be quite awkward. Hence we prefer a different approach. Before we proceed in a formal description of $\mathcal{T}_2$, notice that for this tree the number of offspring per vertex are independent and identically distributed. In fact, on the set $\{\nu$ is generated by $j\}$, we have that $\widetilde{\Xi}_\nu[0] \neq \Xi^*_j[0]$, because the latter belongs to $\widetilde{\Xi}_j$. This implies that $\widetilde{\Xi}_\nu[0] \geq \Xi^*_j[1]$. Hence $x_j^* - \widetilde{\Xi}_\nu[0]$ is independent of $\Xi^*_j[1]$. In fact, the former is determined by the exponentials and Bernoulli attached to j excluding $W_1^{(j)}$, while $\Xi^*_j[1]$ is determined by $W_1^{(j)}$ and exponentials and Bernoullis attached to vertices different from j. Moreover, analyzing the event $\{W_1^{(\nu)}/f(1) < x_j^* - \Xi^*_\nu[0]\}$, one can easily argue that it does not depend on the exponentials and Bernoullis attached to vertices different from u and j. Summarizing, the number of offspring of j in this new tree depends only

- on the exponentials attached to j, with the exception of $W_1^{(j)}$,

- on the Bernoullis attached to j,
- and on $W_1^{(\nu)}$, if $\nu = \Xi_j^*[n]$ for some n and $R_n^{(j)} = 1$.

This implies that the number of offspring per vertex are i.i.d..

Now we are ready to give a formal construction of $\mathcal{T}_2$. Suppose that to each $x \in \Xi_1^*$ we associate an extra exponential random variable Θ_x. Let

$$\eta_1 \stackrel{def}{=} \#\{x \in \Xi_1^* \colon \Theta_x/f(1) < x_1^* - x \text{ and } g(x) = 1\}.$$

The previous random variable counts also the n satysfying $\#\big(\Xi \cap [0, \Xi_1^*[n]]\big) = \infty$, hence η_1 is stochastically larger than σ_i for any i. Then the Galton-Watson tree $\mathcal{T}_2$ whose offspring distribution is the same as the one of η_1 is stochastically larger than $\mathcal{T}_1$. We assume that $\mathcal{T}_2$ is built on the same probability space of $\mathcal{T}_1$. In other words, we can assume, and we will, that $\mathcal{T}_1$ is a subtree of $\mathcal{T}_2$. Next we prove that the average number of offspring is bounded by a finite constant m. Define

$$\Omega_{k,j} \stackrel{def}{=} \Big\{W_1^{(k)}/f(1) < \sum_{s=k+1}^{\infty} W_s^{(j)}/f\big(N_j(s)\big)\Big\}. \tag{2.15}$$

Notice that we should have used different exponentials(1) instead of $W_1^{(k)}$, but the two share the same distribution and are independent of the right-hand side, and this notation is easier to handle. Of course, we are allowed to do that because we are interested only in estimating the probability of this event.

We have that

$$\mathbb{E}[\eta_j] \le \mathbb{E}[\sum_{k=1}^{\infty} \mathbb{1}_{\Omega_{k,j}}] = \sum_{k=1}^{\infty} \mathbb{P}(\Omega_{k,j}). \tag{2.16}$$

In order to prove (2.16), notice that on the left-hand side we count the number of elements in Ξ_u^* with Bernoulli equal to 1, and which satisfy an extra condition. The right-hand side counts only those vertices which satisfy the extra condition. Hence we only need to prove that $\mathbb{P}(\Omega_{k,j})$ is summable. Notice that $\mathbb{P}(\Omega_{k,j})$ is independent of j.

Set $a_k \stackrel{def}{=} \min_{s \ge k} f(s)$ and define

$$\Gamma \stackrel{def}{=} \{k \in \mathbb{N} \colon (1-p)^2 a_{k+1} < f(1) \text{ or } \sum_{s=k+1}^{\infty} 1/f(s) > 0.05(1/f(1))\}.$$

The set Γ is finite. This is implied by the facts that $a_k \to \infty$ and $\sum_{s=k+1}^{\infty} 1/f(s) \to 0$. For $k \in \Gamma$ we use the trivial bound $\mathbb{P}(\Omega_{k,j}) \le 1$. In the following, we fix $k \notin \Gamma$, i.e. we assume that k satisfies $(1-p)^2 a_{k+1} \ge f(1)$ and $\sum_{s=k+1}^{\infty} 1/f(s) \le 0.05(1/f(1))$. We prove the following bound,

$$\mathbb{P}(\Omega_{k,j}) \le \mathrm{e}^{-C_1 a_{k+1}}, \qquad \text{for } k \notin \Gamma, \tag{2.17}$$

where $C_1 = 0.15(1-p)^2(1/f(1))$. Denote by $Y_k = \sum_{s=k+1}^{\infty} W_s^{(j)}/f\big(N_j(s)\big)$, and $Z = W_1^{(k)}/f(1)$. Set $\theta = (1-p)^2 a_{k+1}$. We have

$$\begin{aligned} E[\mathrm{e}^{\theta Y_k}] &= \prod_{s=k+1}^{\infty} \sum_{j=1}^{\infty} \Big(\frac{f(s)}{f(s)-\theta}\Big)^j p^{j-1}(1-p) \\ &= \prod_{s=k+1}^{\infty} \Big(\frac{f(s)}{f(s)-\theta}\Big)(1-p)\frac{1}{1-\big(pf(s)/(f(s)-\theta)\big)} \\ &= \prod_{s=k+1}^{\infty} \frac{f(s)(1-p)}{f(s)(1-p)-\theta} \\ &= \prod_{s=k+1}^{\infty} \frac{1}{1-\big(\theta/(f(s)(1-p))\big)} \\ &\le \exp\{3\theta \sum_{s=k+1}^{\infty} 1/f(s)\}, \end{aligned} \tag{2.18}$$

where we used that $\theta/((1-p)f(s)) \le (1-p) < 1$ for $s \ge k$ and our choice of θ, and the inequality $1-x \ge \mathrm{e}^{-3x}$, for $x \in (0,1)$. Moreover

$$\mathbb{E}[\mathrm{e}^{-\theta Z}] = \frac{f(1)}{\theta + f(1)} \le \mathrm{e}^{-0.3\theta/f(1)},$$

where we use that $k \notin \Gamma$ (hence $\theta/f(1) < 1$), and the inequality $1+x \ge \mathrm{e}^{0.3x}$, still valid for $x \in (0,1)$. This implies that

$$\begin{aligned} \mathbb{P}(\Omega_{k,j}) = \mathbb{P}\big(Y_k - Z > 0\big) &= \mathbb{P}(\mathrm{e}^{\theta Y_k}\mathrm{e}^{-\theta Z} > 1) \\ &\le \exp\Big\{-\theta\big(0.3/f(1) - 3\sum_{j=k+1}^{\infty} 1/f(j)\big)\Big\} \\ &\le \exp\Big\{-\theta\big(0.15/f(1)\big)\Big\}, \end{aligned} \tag{2.19}$$

where in the last step we used that $\sum_{s=k+1}^{\infty} 1/f(s) \le 0.05(1/f(1))$ which holds because we picked $k \notin \Gamma$. The last inequality in (2.19) proves (2.17) for our choice of θ.

Hence

$$\mathbb{E}[\eta_j] \le \mathbb{E}[\sum_{k=1}^{\infty} \mathbb{1}_{\Omega_{k,j}}] \le \sum_{k\in\Gamma} 1 + \sum_{k\notin\Gamma} \mathrm{e}^{-C_1 a_{k+1}} \stackrel{def}{=} m < \infty, \tag{2.20}$$

where, for the finitess of m we used the second assumption in Theorem 1.1 i), and the fact that Γ is a finite set.

For each vertex u in $\mathcal{T}_2$, recall that we denote by $\Xi_u^*[0]$ the time when this vertex was generated and by $\Xi_u^*[n] = \Xi_u^*[0] + \sum_{j=1}^{n} W_j^{(u)}/f(N_u(j))$. This is consistent with our definition given in (2.7), but now it is defined for indices which are not necessarily integers. Next we

prove that each vertex u at level $n+1$ has a probability to satisfy $\widetilde{\Xi}_u[1] < M$ which decreases faster than e^{-cn} for any $c > 0$. For any vertex $u \in \mathcal{T}_2$ we denote by $|u|$ its distance from the root of the tree. Recall that the set of vertices at distance k from the root is called level k. Fix a large parameter M. A vertex u of $\mathcal{T}_2$ is **good** if the element which generates u is smaller than M. A path is a (possibly finite) sequence of vertices $u(i), i \geq 1$, such that $u(i+1)$ is generated by $u(i)$. We say that a path connects vertex a to level $n+1$ if the first element of the path is a and the last lies at level $n+1$. We build the following random path $\mathbf{u}$. We start from $1 = u(0)$ and if this vertex has at least one offspring in $\mathcal{T}_2$, we choose one at random assigning the same probability to each offspring. We denote its label as $u(1)$. If $u(1)$ has at least one offspring, we choose one of them at random and denote its label by $u(2)$. We follow this procedure until we either reach level $n+1$ or find a vertex with no offspring. The event $\{$the path $\mathbf{u}$ connects 1 to a vertex at level $n+1\}$ equals the event that each of the $u(i)$ has at least one offspring. Hence

$$\big\{\text{the path } \mathbf{u} \text{ connects 1 to a vertex at level } n+1\big\} = \bigcap_{i=0}^{n} \big\{\eta_{u(i)} \geq 1\big\}.$$

Notice that each event $\big\{\eta_{u(i)} \geq 1\big\}$ is independent of $\Xi^*_{u(i-1)}[1]$ and is independent of each $W_k^{(\ell)}$ with $\ell < u(i-1)$, and $k \geq 1$. Moreover the events $\big\{\eta_{u(i)} \geq 1\big\}$ are independent. Define

$$\Psi(n,k) \stackrel{def}{=} \Big\{ \sum_{i=1}^{n} \mathbb{1}_{\{\eta_{u(i)} \leq k\}} \geq 0.5n \Big\}.$$

Fix $k \geq 1$. We have

$$\begin{aligned}
&\mathbb{P}(u(n+1) \text{ is good } \mid \text{the path } \mathbf{u} \text{ connects 1 to a vertex at level } n+1) \\
&\leq \mathbb{P}\Big(\sum_{i=1}^{n} \Xi^*_{u(i)}[1] - \Xi^*_{u(i)}[0] \leq M \;\Big|\; \text{the path } \mathbf{u} \text{ connects 1 to a vertex at level } n+1 \Big) \\
&= \mathbb{P}\Big(\sum_{i=1}^{n} \Xi^*_{u(i)}[1] - \Xi^*_{u(i)}[0] \leq M \mid \bigcap_{i=1}^{n} \big\{\eta_{u(i)} \geq 1\big\} \Big) \\
&\leq \mathbb{P}\Big(\sum_{i=1}^{n} \Big(\Xi^*_{u(i)}[1] - \Xi^*_{u(i)}[0]\Big) \mathbb{1}_{\{\eta_{u(i)} \leq k\}} \leq M \mid \bigcap_{i=1}^{n} \big\{\eta_{u(i)} \geq 1\big\} \Big) \\
&\leq \mathbb{P}\Big(\sum_{i=1}^{n} \Big(\Xi^*_{u(i)}[1] - \Xi^*_{u(i)}[0]\Big) \mathbb{1}_{\{\eta_{u(i)} \leq k\}} \leq M \mid \bigcap_{i=1}^{n} \big\{\eta_{u(i)} \geq 1\big\} \cap \Psi(n,k) \Big) + \\
&\qquad + \mathbb{P}\Big(\Psi^c(n,k) \mid \bigcap_{i=1}^{n} \big\{\eta_{u(i)} \geq 1\big\} \Big).
\end{aligned} \tag{2.21}$$

In the last step we used that for any triplet of events A, B, C we have

$$\mathbb{P}(A \mid B) \leq \mathbb{P}(A \mid B \cap C) + \mathbb{P}(C^c \mid B).$$

Next, we bound the last probability in (2.21).

$$\begin{aligned}
&\mathbb{P}\Big(\Gamma^c(n,k) \mid \bigcap_{i=1}^n \{\eta_{u(i)} \geq 1\}\Big) \\
&= \mathbb{P}\Big(\sum_{i=1}^n \mathbb{1}\{\eta_{u(i)} \leq k\} \leq 0.5n \mid \bigcap_{i=1}^n \{\eta_{u(i)} \geq 1\}\Big) \\
&= \mathbb{P}\Big(\sum_{i=1}^n \mathbb{1}\{\eta_{u(i)} > k\} \geq 0.5n \mid \bigcap_{i=1}^n \{\eta_{u(i)} \geq 1\}\Big).
\end{aligned} \tag{2.22}$$

Let $\xi_{u(i)}$, $i \leq n$, be i.i.d. random variables taking values in $\mathbb{N}$, with distribution

$$\mathbb{P}(\xi_{u(i)} \geq k) = \mathbb{P}(\eta_{u(i)} \geq k \mid \eta_{u(i)} \geq 1) = \frac{\mathbb{P}(\eta_{u(i)} \geq k)}{\mathbb{P}(\eta_{u(i)} \geq 1)} \stackrel{def}{=} q_k, \qquad \text{for } k \geq 1.$$

The sequence q_k is independent of $u(i)$ because the random variables η_i are i.i.d.. Moreover, as the $\eta_{u(i)}$ are independent, $\sum_{i=1}^n \mathbb{1}\{\eta_{u(i)} > k\}$ is, conditionally on $\cap_{i=1}^n \{\eta_{u(i)} \geq 1\}$, binomially distributed with mean nq_k. If X is a binomial with parameters (n,q) then

$$\mathbb{P}(X \geq 0.5n) \leq \exp\Big\{-\big(\frac{1}{2q}\ln(\frac{1}{2q}) + \frac{1}{2(1-q)}\ln\frac{1}{2(1-q)}\big)n\Big\}, \tag{2.23}$$

by a simple exponential bound (see, for example [8] pages 27 and 35).

Fix $r > 1$. We can choose K_r^* such that $q_{K_r^*} < 1/2$ and

$$\mathbb{P}\Big(\sum_{i=1}^n \mathbb{1}\{\xi_{u(i)} > K_r^*\} \geq 0.5n\Big) \leq \exp\Big\{-\big(\frac{1}{2q_{K_r^*}}\ln(\frac{1}{2q_{K_r^*}}) + \frac{1}{2(1-q_{K_r^*})}\ln\frac{1}{2(1-q_{K_r^*})}\big)n\Big\} \leq \frac{1}{(rm)^n},$$

where m has been defined in (2.20). We can choose such K_r^* because $\lim_{k\to\infty}(1/(2q_k))\ln(1/(2q_k)) = \infty$. Notice that for any $k \geq K_r^*$, we have $q_k \leq q_{K_r^*} < 1/2$. Moreover if $k \geq K_r^*$ then

$$\exp\Big\{-\big(\frac{1}{2q_k}\ln(\frac{1}{2q_k}) + \frac{1}{2(1-q_k)}\ln\frac{1}{2(1-q_k)}\big)n\Big\} \leq \frac{1}{(rm)^n}.$$

This fact is due to the monotonicity of q_k and the convexity of the function $2x\ln(2x) + 2(1-x)\ln 2(1-x)$, for $x \in (0,1)$, and the fact that this function attains its minimum at $1/2$. Next, let (e_i) be a sequence of i.i.d. random variables with distribution

$$\mathbb{P}(e_i \leq x) = \mathbb{P}\Big(W_1^{(2)}/f(1) \leq x \;\Big|\; W_1^{(2)} \leq \sum_{t=K_r^*+1}^{\infty} W_t^{(1)}/f(N_1(t))\Big).$$

In words, e_i is distributed like an exponential with mean $1/f(1)$ conditioned to be smaller than $Y_{K_r^*}$. We claim that the first probability in the last equation of (2.21) is smaller or equal to

$$\mathbb{P}\big(\sum_{j=1}^{\lfloor 0.5n \rfloor} e_i \leq M\big). \tag{2.24}$$

To see this, notice that by a simple exchangeability argument we have that

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{i=1}^{n}\Big(\Xi^*_{u(i)}[1]-\Xi^*_{u(i)}[0]\Big)\mathbb{1}_{\{\eta_{u(i)}\le K^*_r\}}\le M \;\Big|\; \bigcap_{i=1}^{n}\{\eta_{u(i)}\ge 1\}\cap\Psi(n,K^*_r)\Big)\\
&\quad=\mathbb{P}\Big(\sum_{i=1}^{n}\Big(\Xi^*_{u(i)}[1]-\Xi^*_{u(i)}[0]\Big)\mathbb{1}_{\{\eta_{u(i)}\le K^*_r\}}\le M \;\Big|\; \bigcap_{i=1}^{\lfloor 0.5n\rfloor}\{1\le\eta_{u(i)}\le K^*_r\}\Big)\\
&\quad\le\mathbb{P}\Big(\sum_{i=1}^{\lfloor 0.5n\rfloor}\Big(\Xi^*_{u(i)}[1]-\Xi^*_{u(i)}[0]\Big)\mathbb{1}_{\{\eta_{u(i)}\le K^*_r\}}\le M \;\Big|\; \bigcap_{i=1}^{\lfloor 0.5n\rfloor}\{1\le\eta_{u(i)}\le K^*_r\}\Big)
\end{aligned}$$

Again, notice that the events $\{1\le\eta_{u(i)}\le K^*_r\}$, with $i\le n+1$, are independent. Given $\{1\le\eta_{u(i)}\le K^*_r\}$, the random variable $\Xi^*_{u(i)}[1]-\Xi^*_{u(i)}[0]$ is stochastically larger than e_i, as $\sum_{s=k}^{\infty}W^{(1)}_s/f(N_1(s))$ is a.s. decreasing in k. This proves the relationship between (2.24) and the first probability in the last equation of (2.21). Next a simple exponential bound, which uses the fact that e_i are independent, yields

$$\begin{aligned}
\mathbb{P}\Big(\sum_{j=1}^{\lfloor 0.5n\rfloor}e_i\le M\Big)&=\mathbb{P}\Big(\frac{1}{0.5n}\sum_{j=1}^{\lfloor 0.5n\rfloor}e_i\le\frac{1}{0.5M}\Big)\\
&=\mathbb{P}\Big(\exp\Big\{-\theta\frac{1}{0.5n}\sum_{j=1}^{\lfloor 0.5n\rfloor}e_i\Big\}\ge\exp\{-\theta\frac{1}{0.5M}\}\Big)\\
&\le\exp\{-c_n(r,M)n\},
\end{aligned}$$

where $c_n(r,M)\to\infty$ as $n\to\infty$. For each n, $c_n(r,M)$ is the Fenchel-Legendre transform (i.e. we minimize the exponent on θ) of e_i in the point $\frac{1}{0.5M}$. Hence, the number of good vertices in $\mathcal{T}_1$ at level n is smaller or equal to

$$m^n\Big(\exp\Big\{-c_n(r,M)n\Big\}+\frac{1}{(rm)^n}\Big). \tag{2.25}$$

Hence only finitely many vertices in $\mathcal{T}_2$ are good. This implies that only finitely many vertices in $\mathcal{T}_1$ are good, and this, in turn, implies (2.12). □

Proof of Theorem 1.1 i). First suppose that $s_n\equiv p<1$. The minimum of $\inf_i x^*_i$ is a.s. unique, and we denote it by J^*. By Lemma 2.2 $\lim_{n\to\infty}\Xi[n]=x^*_{J^*}$, hence the cardinality of group J^* tends to infinity, while the cardinality of each of the other groups is finite.

Now we reason for general $s_n\le p$, using a simple coupling. Let $\{S_i, i\ge 1\}$ be a sequence of independent Bernoullis with $\mathbb{P}(S_i=1)=s_i/p=1-\mathbb{P}(S_i=0)$. We use these random variables to relabel the points in Ξ as follows. If $S_1=0$ then we set $\Theta_1=\Xi_2[\tau_2]\cup\Xi\setminus\widetilde{\Xi}_2$. If $S_1=1$ then $\Theta_1=\Xi$. Define $\widetilde{\tau}_3\stackrel{def}{=}\inf\{n>\tau_2\colon g(\Theta_1[n])=1\}$. Suppose we have defined Θ_{m-1} and $\widetilde{\tau}_i$, for $i\le m$. On the event $\{\sum_{i=1}^{m-1}S_i=k\}$, if $S_m=0$, respectively $S_m=1$, set $\Theta_m=\Xi_{k+1}[\widetilde{\tau}_{k+1}]\cup\Theta_{m-1}\setminus\widetilde{\Xi}_{k+1}$, respectively $\Theta_m=\Theta_{m-1}$. We set $\widetilde{\tau}_{m+1}\stackrel{def}{=}\inf\{n>$

$\widetilde{\tau}_m\colon g(\Theta_m[n]) = 1\}$. Let $\Theta = \bigcap_i \Theta_i$. The process $\Theta[n]$, with $n \geq 1$ is a GAM$(f, \{s_n\})$. Let $\kappa(n) = \sum_{j=1}^{n} S_j$. Denote by $h(i) = \inf\{n\colon \kappa(n) = i\}$. This implies that the i-th group in Θ is the $h(i)$-th group in Ξ. Let $U_i \stackrel{def}{=} \widetilde{\Xi}_{h(i)}$, and

$$u_i^* \stackrel{def}{=} x_{h(i)}^*. \tag{2.26}$$

This implies that $\inf_j\{u_j^*\colon j \geq 1\}$ is actually a minimum and has a unique minimizer. Following the same reasoning given in the previous paragraph we conclude that the only group whose cardinality grows to infinity is K_r^*. □

Proof of Corollary 1.3. We first assume that $s_n \equiv p$. For any i, denote by $E(i)$ the set of groups which are generated by i. In virtue of (2.20), we have that

$$V(u) \stackrel{def}{=} \{\widetilde{\Xi}_i[1] < x_u^* \quad \text{for only finitely many } i \in E(u)\} \quad \text{holds a.s..} \tag{2.27}$$

Notice that for u which is not a vertex of $\mathcal{T}_1$ we have that $\widetilde{\Xi}_u[1] > \inf_i x_i^* = \lim_{n\to\infty} \Xi[n]$. Hence, we do not have to consider such u. Recall the definition of G_N given before Theorem 1.6. As for each N, there are only finitely many good vertices in $\mathcal{T}_1$, we get $\lim_{N\to\infty} \mathbb{P}(\mathbf{Lead} \in G_N) = 0$. Combining the latter limit with (2.27) we have that

$$\begin{aligned}
&\mathbb{P}(\lim_{n\to\infty} A_u(n) > 1 \text{ for only finitely many } u)\\
&\qquad = \mathbb{P}(\widetilde{\Xi}_u[1] < \inf_i x_i^* \text{ for only finitely many } u)\\
&\qquad = \lim_{N\to\infty} \mathbb{P}(\{\widetilde{\Xi}_u[1] < \inf_i x_i^* \text{ for only finitely many } u\} \cap \{\mathbf{Lead} \notin G_N\})\\
&\qquad \geq \lim_{N\to\infty} \mathbb{P}\Big(\bigcap_{u\in\mathcal{T}_1\colon u\notin G_N} V(u) \cap \{\mathbf{Lead} \notin G_N\}\Big)\\
&\qquad = \lim_{N\to\infty} \mathbb{P}(\mathbf{Lead} \notin G_N) = 1.
\end{aligned}$$

For the general case $s_n \leq p$, apply the same coupling we used at the end of the previous proof. □

3 Proof of Theorem 1.1 ii)

We first deal with the case $s_n = p$. Repeat the construction given in the proof of Theorem 1.1 i), under the hypothesis of Theorem 1.1 ii). Recall the definition of Ξ_u^*, $\widetilde{\Xi}_u$ and x_u^*. Recall also the definition of $\mathcal{T}$. The random variables x_u^* , for $u \geq 1$, are a.s. infinite, because the infinite sum of independent exponentials is finite if and only if its mean is finite. We prove next that for any fixed $M > 0$,

$$\liminf_{u\to\infty} \widetilde{\Xi}_u[1] > M, \qquad \text{a.s..} \tag{3.1}$$

Fix a vertex u_n of $\mathcal{T}$, and denote by u_i, with $i \leq n-1$ its ancestors: i.e. $\Xi_{u_j}[\tau_{u_j}] \in \Xi^*_{u_{j-1}}$, for all $j \leq n$. Then $\Xi_{u_n}[1]$ is stochastically larger than a sum of $n-1$ i.i.d. exponentials with parameter $f(1)$. Hence $\lim_{n\to\infty} \Xi_{u_n}[1] = \infty$, a.s.. Now notice that $\Xi^*_s \cap [0, M]$ is a.s. finite for each $s \geq 1$. Hence, as u grows to infinity, the number of its ancestors grows to infinity, proving (3.1). Since it is easy to adapt the above reasoning to the case $s_n \leq p$, we will leave this task to the reader.

4 Proof of Theorem 1.4

Proof of Theorem 1.4 i). We first analyze the case $s_n \equiv p$. We build a similar construction as the one given in the proof of Theorem 1.1, the only difference being that we place f_i instead of f. We leave to the reader to check that this construction embeds our GAM($\{f_j\}, p$). In this setting

$$x^*_i \stackrel{def}{=} \Xi_i[\tau_i] + \sum_{j=1}^{\infty} \frac{W^{(i)}_j}{f_i\big(N_i(j)\big)}.$$

Notice that x^*_i is a.s. finite if and only if $\sum_{s=1}^{\infty} 1/f_i(s)$ is finite. Hence, we do not exclude that $x^*_i = \infty$, a.s., but we know that

$$\text{there exists at least one } j \text{ for which } x^*_j < \infty, \text{ a.s..} \tag{4.1}$$

Denote by y^* the smallest accumulation point of Ξ. This minimum accumulation point exists because the set of accumulation points of Ξ is closed, and the set Ξ is a subset of $\mathbb{R}^+$. Moreover y^* is a.s. finite because of (4.1). If $y^* < x^*_i$ for all $i \geq 1$, then

$$\lim_{n\to\infty} \Xi[n] < x^*_i, \qquad \text{for all } i \geq 1. \tag{4.2}$$

We need to prove (4.2) only for the case $x^*_i < \infty$, because for the other cases the result is implied by the fact that y^* is an accumulation point which is a.s. finite and $\#\Xi \cap [0, y^* + \varepsilon] = \infty$. Assume that x^*_i is a.s. finite and notice that for fixed i, as $y^* < x^*_i$, then $\delta_i \stackrel{def}{=} (x^*_i + y^*)/2 < x^*_i$. As y^* is an accumulation point for Ξ, then $\#\Xi \cap [0, \delta_i]$ is a.s. infinite. In words there are infinitely many points of Ξ smaller than δ_i. Hence

$$\lim_{n\to\infty} \Xi[n] \leq \delta_i < x^*_i. \tag{4.3}$$

The inequality in (4.3) holds for each i, yielding (4.2). Moreover, (4.3) implies that each group will end up having finite cardinality. This is because $\#\widetilde{\Xi}_i \cap [0, \delta_i]$, as δ_i is strictly less than x^*_i which is the only accumulation point for $\widetilde{\Xi}_i$. The latter statement is a direct consequence of the definitions of x^*_i and $\widetilde{\Xi}_i$.

On the other hand, if $y^* = x^*_i$ for some i, then using again that all the x^*_j which are finite are also a.s. distinct, we have that $\lim_{n\to\infty} \Xi[n] = x^*_i$. To prove the latter inequality, suppose

it is not true, i.e. $\lim_{n\to\infty}\Xi[n] < x_i^*$. Then there would be an accumulation point smaller than y^*, which would yield a contradiction.

Next we analyse the general case, i.e. $s_n \le p$, for some $p < 1$ and all $n \ge 1$. The problem here is that the reinforcement function is group dependent. In the special case $s_n \equiv p$ we had that the first point labelled i was $\Xi_i^*[0]$. We need to translate the points labelled i in the new construction for the general case. Denote by $v(i)$ the time when the i-th group is created and denote by Υ_{i-1} the union of the points labelled j, with $j \le i-1$. We have that the first point labelled i is exactly $\Upsilon_{i-1}[v(i)]$. Set

$$\begin{aligned} U_i^* &\stackrel{def}{=} \Big\{\Xi_i^*[n] - \Xi_i^*[0] + \Upsilon_{i-1}[v(i)] \colon n \ge 0\Big\}, \\ u_i^* &\stackrel{def}{=} x_i^* - \Xi_i^*[0] + \Upsilon_{i-1}[v(i)]. \end{aligned} \tag{4.4}$$

Hence $\Upsilon_i = \Upsilon_{i-1} \cup U_i^*$. Moreover, let $\Upsilon = \cup_{i=1}^{\infty}\Upsilon_i$. It is easy to check that Υ embeds $\mathrm{GAM}(\{f_i\},\{s_n\})$. We prove the theorem on the event $\{u_j^*$ is a.s. finite for at least one created group $j\}$. Repeating the argument we gave for the case $s_n \equiv p$, we see that either all the groups remains finite or there exists exactly one dominating the others. □

Proof of Theorem 1.4 ii). First assume that $s_n \equiv p$, for some $p < 1$. Under the assumptions of this part of the Theorem, we have that each $x_j^* = \infty$, a.s.. Hence $\inf_j x_j^* = \infty$. By our construction, either $\lim_{n\to\infty}\Xi[n] = \infty$, in which case the cardinality of each group is a.s. diverging to infinity, or $\lim_{n\to\infty}\Xi[n] = \Gamma < \infty$, a.s., in which case $\#\widetilde{\Xi}_u \cap [0,\Gamma] < \infty$, a.s.. In words, in the latter case, the cardinality of each group will eventually remain finite, for otherwise $x_j^* \le \Gamma < \infty$ for some j, and this would give a contradiction.

For general $s_n \le p$, we have that $u_i^* = \infty$, where the u_i^* are the random variables defined in (4.4). Reasoning as in the previous paragraph we get the result for this more general case. □

4.1 An example when the third phase occurs

Next we show an example where a third phase occurs, i.e.

$$\lim_{n\to\infty} A_i(n) < \infty, \qquad \text{a.s. for each } i \ge 1. \tag{4.5}$$

In this example we pick $f_j(n) = \mathrm{e}^{(j^3+n)}$ and $s_n \equiv p \in (0,1)$. Notice that $\tau_{i+1} - \tau_i$, with $i \ge 1$, is an i.i.d. sequence of geometrically distributed random variables, with mean $1/p$. Hence, by a standard exponential bound, we have

$$\mathbb{P}\Big(\tau_n > \big((1/p) - \varepsilon\big)n\Big) = \mathbb{P}\Big(\sum_{i=1}^{n}(\tau_i - \tau_{i-1}) > \big((1/p) - \varepsilon\big)n\Big) \le \mathrm{e}^{-Cn}.$$

This implies that

$$\sum_{n=1}^{\infty}\mathbb{P}(\tau_n > n^2) < \infty. \tag{4.6}$$

Next, we use this fact to prove that

$$\text{for each } j \geq 1 \text{ there exists an } s > j \text{ such that } x_s^* < x_j^*. \tag{4.7}$$

The latter implies that $\inf_j x_j^*$ is not attained. As this infimum is an accumulation point for Ξ, this would imply that the smallest accumulation point of Ξ is smaller than x_j^*, for all $j \geq 1$. Hence, (4.5) would hold.

Next we turn to the proof of (4.7). Fix $j \in \mathbb{N}$. As $\Xi_j \subset \Xi$, we have that $\Xi[\tau_u] \leq \Xi_j[\tau_u]$. Hence,

$$\begin{aligned}
&\mathbb{P}\Big(x_u^* > x_j^*\Big) \\
&= \mathbb{P}\Big(\Xi[\tau_u] + \sum_{\ell=1}^{\infty} W_\ell^{(u)}/f_u(N_u(\ell)) > x_j^*\Big) \\
&\leq \mathbb{P}\Big(\Xi_j[\tau_u] + \sum_{\ell=1}^{\infty} W_\ell^{(u)}/f_u(N_u(\ell)) > x_j^*\Big) \\
&\leq \mathbb{P}\Big(\sum_{\ell=1}^{\infty} W_\ell^{(u)}/f_u(N_u(\ell)) > \sum_{\ell=u^2}^{\infty} W_\ell^{(j)}/f_j(N_j(\ell)) \;\Big|\; \tau_u < u^2\Big) + \mathbb{P}\big(\tau_u \geq u^2\big)
\end{aligned} \tag{4.8}$$

The last inequality in (4.8) is justified as follows. For any pair of events A and B we have that

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \cap B^c) \leq \mathbb{P}(A \mid B) + \mathbb{P}(B^c).$$

Notice that $\{\tau_u < u^2\}$ is measurable with respect to the σ-algebra

$$\sigma\Big\{R_i^{(t)} \colon t < u \text{ and } i < u^2\Big\}.$$

In words, if we know the first $u^2 - 1$ observations of each $\widetilde{\Xi}_t$, with $t < u$, and the associated Bernoullis, we know if the event $\{\tau_u < u^2\}$ holds. Hence the latter event is independent of the pair

$$\Big(\sum_{\ell=1}^{\infty} W_\ell^{(u)}/f_u(N_u(\ell)), \sum_{\ell=u^2}^{\infty} W_\ell^{(j)}/f_j(N_j(\ell)\Big).$$

Hence the last expression in (4.8) equals

$$\mathbb{P}\Big(\sum_{\ell=1}^{\infty} W_\ell^{(u)}/f_u(N_u(\ell)) > \sum_{\ell=u^2}^{\infty} W_\ell^{(j)}/f_j(N_j(\ell))\Big) + \mathbb{P}\big(\tau_u > u^2\big).$$

The last expression is summable. To see this, in virtue of (4.6), we just need to prove that the first term is summable. Then our argument follows from an application of the first Borel-Cantelli lemma. In fact, the summability implies that $\{x_u^* < x_j^*\}$ for infinitely many u. Set

$\gamma_{u,j} = (1/j)\mathrm{e}^{-j^3-u^2}$, and recall that j is fixed. For any pair of random variables X and Y and any constant a, we have that

$$\begin{aligned}\mathbb{P}(X > Y) = \mathbb{P}(X > Y, X > a) + \mathbb{P}(X > Y, X < a) &\le \mathbb{P}(X > Y, X > a) + \mathbb{P}(Y < a)\\ &\le \mathbb{P}(X > a) + \mathbb{P}(Y < a).\end{aligned}$$

We apply this fact to obtain

$$\begin{aligned}&\mathbb{P}\Big(\sum_{\ell=1}^{\infty} W_\ell^{(u)}/f_u(N_u(\ell)) > \sum_{\ell=u^2}^{\infty} W_\ell^{(j)}/f_j(N_j(\ell))\Big)\\ &\quad\le \mathbb{P}\Big(\sum_{\ell=1}^{\infty} W_\ell^{(u)}/f_u(N_u(\ell)) > \gamma_{u,j}\Big) + \mathbb{P}\Big(\sum_{\ell=u^2}^{\infty} W_\ell^{(j)}/f_j(N_j(\ell)) < \gamma_{u,j}\Big).\end{aligned}$$

Notice that

$$\sum_{n=1}^{\infty} 1/f_u(n) = \sum_{n=1}^{\infty} \mathrm{e}^{-u^3-n} = \mathrm{e}^{-u^3}\sum_{n=1}^{\infty} \mathrm{e}^{-n} = C_1\mathrm{e}^{-u^3},$$

while, by a similar reasoning, $\sum_{n=u^2} 1/f_j(n) \sim C_2\mathrm{e}^{-u^2-j^3}$. Next we reason like in the proof of (2.17), with $\min_{n\ge k} \mathrm{e}^{n+u^3} = \mathrm{e}^{+k+u^3}$. We have

$$\mathbb{P}\Big(\sum_{\ell=1}^{\infty} W_\ell^{(u)}/f_u(N_u(\ell)) > \gamma_{u,j}\Big) \le \exp\{-(1-p)^2(\mathrm{e}^{1+u^3})(\gamma_{u,j} - C_1\mathrm{e}^{-u^3})\},$$

and the right hand-side is summable in u for fixed j. This is because for fixed j, $\gamma_{u,j} > C_1\mathrm{e}^{-u^3}$ for all sufficiently large u. In a similar way, using Chebyshev inequality after applying the function $\mathrm{e}^{\theta x}$ to both sides and choosing $\theta = (1-p)^2\mathrm{e}^{u^2+j^3}$, we obtain

$$\mathbb{P}\Big(\sum_{\ell=u^2}^{\infty} W_\ell^{(j)}/f_j(N_j(\ell)) < \gamma_{u,j}\Big) \le \exp\{-(\mathrm{e}^{u^2+j^3})(C_2\mathrm{e}^{-u^2-j^3} - \gamma_{u,j}\}.$$

The last expression is summable in u, because, for fixed j, $C_2\mathrm{e}^{-u^2-j^3}$ is larger than $\gamma_{u,j}$ for all sufficiently large u.

5 Brownian motion embedding

Suppose that the positive function f satisfies the condition $\sum_{j=1}^{\infty} 1/f(j) < \infty$. Consider an urn with k white balls and 1 red one. We pick a ball at random, and it is white with probability $f(k)/(f(k)+f(1))$. Suppose that by the time of the n-th extraction we picked j white balls and $n-j$ red ones. The probability to pick a white ball at the next stage becomes $f(k+j)/(f(k+j)+f(n+1-j))$. Let

$$D \stackrel{def}{=} \{\text{only a finite number of white balls are picked}\},$$

Denote by $\mathbb{P}^{(k)}$ the probability measures referring to the urn with initial conditions and dynamics described above.

Let $F \stackrel{def}{=} \sum_{j=1}^{\infty} 1/f(j)$ and recall that $F_k \stackrel{def}{=} \sum_{j=k}^{\infty} 1/f(j)$. Let the process $\mathbf{B} := \{B_t,\, t \geq 0\}$ be a standard Brownian motion, which starts from the point $F - F_k = \sum_{i=1}^{k-1} 1/f(i)$. Denote by $\mathbb{Q}^{(k)}$ the measure associated with this Brownian motion. We use this process to generate the urn sequence described at the beginning of this section, as follows. Set $m_0 = 0$ and let

$$m_1 \stackrel{def}{=} \inf\Big\{t \geq 0 :\ B_t \text{ hits either } 1/f(k) \text{ or } -1/f(1)\Big\}. \tag{5.1}$$

If $B_{m_1} - B_{m_0} > 0$ then set $z_1 = 1$, otherwise set $z_1 = 0$.

Suppose we defined m_n and $z_1, z_2, \ldots, z_n$. Set $\phi(n) = \sum_{i=1}^{n} z_i$. On the event $\phi(n) = s$, we define

$$m_{n+1} = \inf\left\{t \geq m_n :\ B_t - B_{m_n} \text{ hits either } \frac{1}{f(s+k)} \text{ or } -\frac{1}{f(n-s+1)}\right\}.$$

Set

$$z_{n+1} \stackrel{def}{=} \begin{cases} 1 & \text{if } B_{m_{n+1}} - B_{m_n} = 1/f(k+s) \\ \\ 0 & \text{if } B_{m_{n+1}} - B_{m_n} = -1/f(n-s+1) \end{cases}.$$

By the ruin problem for Brownian motion, we have that

$$\mathbf{P}\left(z_{n+1} = 0 \mid \phi(n) = s\right) = \frac{1/f(s+k)}{(1/f(s+k)) + (1/f(n-s+1))} = \frac{f(n-s+1)}{f(s+k) + f(n-s+1)},$$

which is exactly the urn transition probability.

In this way we embedded the urn into Brownian motion. In fact, the process $\phi(n)$, with $n \geq 1$, is distributed like the number of white balls withdrawn from the urn associated to the reinforcement scheme described at the beginning of this section. Notice that

$$B_{m_n} = \sum_{j=1}^{k+\phi(n)} (1/f(j)) - \sum_{s=1}^{n-\phi(n)} (1/f(s)), \qquad \text{with } n \geq 0. \tag{5.2}$$

Define

$$S \stackrel{def}{=} \lim_{n\to\infty} m_n. \tag{5.3}$$

This limit exists because the sequence of stopping times $\{m_n\}$ is increasing. For this reason S is itself a stopping time. Define

$$D_1 \stackrel{def}{=} \Big\{\exists n \geq 1\colon B_S = \sum_{j=1}^{k+n}(1/f(j)) - \sum_{j=1}^{\infty}(1/f(j))\Big\} = \big\{B_S < 0\big\},$$

$$D_2 \stackrel{def}{=} \Big\{\exists n \geq 1\colon B_S = \sum_{j=1}^{\infty}(1/f(j)) - \sum_{j=1}^{n}(1/f(j))\Big\} = \big\{B_S > 0\big\}.$$

Moreover, in virtue of Theorem 7.1 we have that exactly one of the collection of events $\big\{\{z_i = 0\}, i \geq 1\big\}$ and $\big\{\{z_i = 1\}, i \geq 1\big\}$ holds finitely many times, a.s.. This implies that the event $D_1 \cup D_2$ holds $\mathbb{Q}^{(k)}$-a.s.. By our embedding, we have that

$$\mathbb{Q}^{(k)}(D_1) = \mathbb{P}^{(k)}(D),$$

where D was defined at the beginning of this section.

Proof of Theorem 1.5. In order to prove our result we only need to prove the following

$$\mathbb{Q}^{(k)}(D_1) \leq \frac{1}{2} \prod_{s=1}^{k-1} \frac{f(s)F_k}{1 + f(s)F_k}.$$

Let

$$T \stackrel{def}{=} \inf\Big\{n \geq 1 \colon \phi(n) = \frac{n-k}{2}\Big\}. \tag{5.4}$$

This stopping time can be infinite with positive probability. Notice that on $\{T < \infty\}$, by (5.2), we have that the urn generated by the Brownian motion contains, at time T, an equal number of white and red balls, and $B_{m_T} = 0$. Viceversa, if we let

$$H \stackrel{def}{=} \inf\{t \geq 0 \colon B_t = 0\},$$

then we have that

$$\{H < S\} = \{T < \infty\}. \tag{5.5}$$

To prove (5.5), notice that for $k \in \mathbb{N}$, with $k > 0$, the random sequence

$$n \to \sum_{j=1}^{k+\phi(n)} \frac{1}{f(j)} - \sum_{j=1}^{n-\phi(n)} \frac{1}{f(j)},$$

cannot switch sign without becoming 0. So if $B_{m_j} > 0$ and $B_{m_t} < 0$, for some $j < t$, then there exists an s, with $j < s < t$, such that $B_{m_s} = 0$. In this case, by time s we have a tie. We use this fact throughout the proof.

Recall that under $\mathbb{Q}^{(k)}$ the Brownian motion $\mathbf{B}$ starts from $F - F_k$. For $j \leq k$, let

$$H_j \stackrel{def}{=} \inf\{t \geq 0 \colon B_t = F_j - F_k\}.$$

Notice that $F_j - F_k \geq 0$ for $j \leq k$. Moreover, by time H_j, with $j \leq k-1$, on the event $\{H_j < S\}$, at least j red balls have been extracted. To see this, we first focus on H_1, and prove that by this time, on the event $\{H_1 < S\}$, at least one red ball has been picked. Suppose that this is not true, i.e. suppose that we picked 0 red balls by time H_1. The reader can check from our embedding that this implies that

$$\min_{0 \leq t \leq S} B_t > F - F_k - 1/f_1 = F_1 - F_k.$$

This would imply that $H_1 > S$ contradicting our hypothesis. By reiterating the same reasoning we get that the statement holds true for any $j \leq k$.

Define

$$M_j \stackrel{def}{=} \Big\{\text{after time } H_{j-1}, \text{ the process } \mathbf{B} \text{ reaches } F_j \text{ before it hits } F_j - F_k\Big\}.$$

On M_j the Brownian motion, after time H_{j-1}, will hit F_j before there is a tie in the urn, because $F_j - F_k \geq 0$, for $j \leq k$. Next we prove that for any $j \in \{1, 2, \ldots, k\}$, if M_j holds then only a finite number of red balls are extracted, i.e. $M_j \subset D_2$. We split this proof into two parts, we first prove that $M_j \cap \{S \leq H_{j-1}\} \subset D_2$ and then $M_j \cap \{S > H_{j-1}\} \subset D_2$. In order to prove the first inclusion, recall that under $\mathbb{Q}^{(k)}$ the Brownian motion starts at $F - F_k$. This implies that if $S \leq H_{j-1}$, then infinitely many balls will be extracted before the Brownian motion hits $F_{j-1} - F_k$. As $F - F_k > F_{j-1} - F_k > 0$, we have that infinitely many balls will be extracted before $\mathbf{B}$ hits 0, i.e. before a tie. This implies that $B_S > 0$, which in turn implies $M_j \cap \{S \leq H_{j-1}\} \subset D_2$.

Next we prove that $M_j \cap \{S > H_{j-1}\} \subset D_2$. On the set $M_j \cap \{S > H_{j-1}\}$, by time H_{j-1} the number of red balls extracted is at least $j - 1$. This implies that

$$B_S \leq \sum_{j=1}^{\infty}(1/f(j)) - \sum_{t=1}^{j-1}(1/f(t)) = F_j, \qquad \forall k \geq n. \tag{5.6}$$

This is a consequence of (5.2) and the fact that $n - \phi(n)$ is a non-decreasing random sequence, and if $n - \phi(n) = j - 1$ for some n then $\lim_{n\to\infty} n - \phi(n) \geq j - 1$. Let

$$V_1 \stackrel{def}{=} \inf\{m_n \colon m_n > H_{j-1}, \text{ and } B_{m_n} - B_{m_{n-1}} > 0\},$$

i.e. the first time after H_{j-1} that a white ball is extracted. The stopping time V_1 could be infinite. Next we prove that on M_j the random time V_1 is a.s. finite. Recall that H_{j-1} is the first time that the process $\mathbf{B}$ hits $F_{j-1} - F_k$, and that $0 < F_{j-1} - F_k < F - F_k$. This implies that by time H_{j-1} the number of white balls generated by the Brownian motion, plus the initial k, overcomes that of the red ones. On M_j, after time H_{j-1}, the process will hit F_j before it hits 0. This implies that $V_1 < \infty$ a.s. on M_j. In fact if no white balls are extracted after time H_{j-1} the process would hit 0 before it hits F_j giving a contradiction. Moreover on M_j, we have that $B_{V_1} > 0$, hence by time V_1 the white balls are still ahead with respect the red ones. We can repeat the same reasoning with

$$V_2 \stackrel{def}{=} \inf\{m_n \colon m_n > V_1, \text{ and } B_{m_n} - B_{m_{n-1}} > 0\},$$

to argue that V_2 is a.s. finite and by time V_2 the white balls are still in advantage. By reiterating this argument we get that only finite many red balls will be extracted, because each V_i occurs before a tie, a.s.. Hence D_2 holds when M_j holds. This implies that $D_2^c \subset M_j^c$ for each $j \in \{1, 2, \ldots, k-1\}$. If $\cap_{j=1}^{k-1} M_j^c$ holds, then either $\{B_S > 0\}$ holds or $\{H < S\}$ holds.

If the latter event holds, independently of the past, the probability that only finitely many white balls are picked is exactly 1/2, by symmetry. Moreover, the events M_j are independent, because they are determined by the behavior of disjoint increments of the Brownian motion. By the standard ruin problem for this process, we have that

$$\mathbb{Q}^{(k)}(M_j) = \frac{1}{1+f(j)F_k}. \tag{5.7}$$

We get

$$\mathbb{Q}^{(k)}(D_1) = \mathbb{Q}^{(k)}(D_2^c) \leq \frac{1}{2}\prod_{j=1}^{s}\Big(1-\frac{1}{1+f(j)F_k}\Big).$$

□

6 Proofs of Theorem 1.6 and Corollary 1.7

Proof of Theorem 1.6. Notice that **Lead** must be a vertex of $\mathcal{T}_1$. Under the assumptions of the Theorem, the probability that $\inf_i x_i^* > M$ is smaller or equal to the probability that $x_1^* > M$. The latter probability is bounded as follows

$$\mathbb{P}(\sum_{j=1}^{\infty} W_j^{(1)}/f\big(N_1(j)\big) > M) \leq \exp\{-(1-p)^2 a_1(M-3F)\}.$$

We set $C_1 = \exp\{3(1-p)^2F\}$ and $C_2 = (1-p)^2 a_1$, where $a_1 = \inf_{k\geq 1} f(k)$. In virtue of (2.25), the probability that all the vertices at level n are good is at least

$$1 - m^n \inf_{r>1} \mathrm{e}^{-c_n(r,M)n} + r^{-n},$$

where $c_n(r,M)$ were introduced at the end of the proof of Lemma 2.2 and m was introduced in (2.20). Recall that g_n is the set of the vertices of $\mathcal{G}$ at level n. Moreover, recall that $G_n = \cup_{j\geq n} g_j$. We have

$$\begin{aligned}\mathbb{P}(\mathbf{Lead} \in G_n) &\leq \mathbb{P}\big(\{\inf_i x_i^* > M\} \cup \{\text{at least one vertex in } g_n \text{ is not good}\}\big)\\ &\leq C_1 \mathrm{e}^{-C_2 M} + m^n \inf_{r>1}(\mathrm{e}^{-c_n(r,M)n} + r^{-n}).\end{aligned}$$

□

Proof of Corollary 1.7. Set $i(1) = \tau_2$ and define recursively $i(n) = \inf\{j > i(n-1)\colon R_j^{(1)} = 1\}$. Notice that $i(k) \geq k$. If a vertex ν of $\mathcal{G}$ belongs to g_1 then we have that $\tau_\nu = i(k)$ for some k. We have

$$\mathbb{P}\big(\mathbf{Lead} = 1\big) \geq 1 - \mathbb{E}[\sum_{j\in g_1} \mathbb{1}_{\{x_j^* < x_1^*\}}] - \mathbb{P}(\mathbf{Lead} \in G_2).$$

We bound the last probability in the previous expression using Theorem 1.6. Order the groups at level one, starting from the smaller. As $i(k) \geq k$, we have that by the time the k-th group at level 1 is created, there are at least k balls in urn 1. Hence, using Theorem 1.5, we get

$$\mathbb{E}[\sum_{j \in g_1} \mathbb{1}_{\{x_j^* < x_1^*\}}] \leq \sum_{k=1}^{\infty} \frac{1}{2} \prod_{\ell=1}^{k-1} \frac{f(\ell)F_k}{1+f(\ell)F_k}.$$

□

7 Appendix

Fix two real numbers r and w, and two sequences of positive real numbers $\{W(k), k \geq w\}$ and $\{R(i), i \geq r\}$. Suppose we have an urn with w (resp r) white (resp. red) balls. If at step $n \geq 0$ there are exactly j white balls, with $n - j \geq 0 \geq w - j$, then the probability to pick a white ball is

$$\frac{W(j)}{W(j) + R(n-j+w)}.$$

If a white (resp. red) ball is picked, at time $n+1$ the composition of the urn becomes $j+1$ (resp. j) white balls and $n-j+w$ (resp. $n-j+w+1$) red ones. Denote by

$$\begin{aligned} A_R^c &\stackrel{def}{=} \{ \text{ the number of red balls in the urn goes to } \infty \text{ as } n \to \infty\}, \\ A_W^c &\stackrel{def}{=} \{ \text{ the number of red balls in the urn goes to } \infty \text{ as } n \to \infty\}. \end{aligned}$$

Let $\mathbb{Q}$ be the measure describing the dynamics of this urn. We have the following result, due to Herman Rubin (see the Appendix in [7])

Theorem 7.1 *[H. Rubin] We have the following 3 cases*

i) If $\sum_{k=w}^{\infty} \big(W(k)\big)^{-1} = \infty$ and $\sum_{k=r}^{\infty} \big(R(k)\big)^{-1} = \infty$, then both the number or red balls and the number of white balls in the urn goes to ∞, a.s., as $n \to \infty$.

ii) If $\sum_{k=w}^{\infty} \big(W(k)\big)^{-1} < \infty$ and $\sum_{k=r}^{\infty} \big(R(k)\big)^{-1} = \infty$, then

$$\mathbb{Q}(A_R) = 1.$$

iii) If $\sum_{k=w}^{\infty} \big(W(k)\big)^{-1} < \infty$ and $\sum_{k=r}^{\infty} \big(R(k)\big)^{-1} < \infty$, then

$$\mathbb{Q}(A_R) + \mathbb{Q}(A_W) = \mathbb{Q}(A_R \cup A_W) = 1,$$

and both $\mathbb{Q}(A_R)$ and $\mathbb{Q}(A_W)$ are strictly positive.

Acknowledgement. We thank two anonymous referees for their suggestions. A.C. and C.C. thank IAS-Technische Universitaet Muenchen for financial support.

References

[1] A.-L. Barabasi and R. Albert (1999), Emergence of Scaling in Random Networks, *Science*, N 286, 509-512.

[2] R. Albert and A.-L. Barabasi (2002),Statistical mechanics of complex networks, *Review of Modern Physics*, N 74, 47-97.

[3] S. Bhamidi (2011), Universal techniques to analyze preferential attachment tree and networks: Global and Local analysis , *Probab. Surveys*, to appear.

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

[5] F. Chung, S.Handjani and D Jungreis (2003), Generalizations of Polya's Urn Problem, *Annals of Combinatorics.*

[6] C. Cotar and V. Limic (2009), Attraction time for strongly reinforced walks, *Annals of Applied Probability* Vol 19, N. 5, 1972-2007.

[7] B. Davis (1990), Reinforced random walk, *Prob. Theory and rel. fields* Vol 84, N 2, 203-229.

[8] A. Dembo and O. Zeitouni (1998), Large deviations techniques and applications, second edition, Springer.

[9] J.-L. Deneubourg, S. Aron, S. Goss and J.M. Pasteels (1990), The self-organising exploratory pattern of the argentine ant, *Journal of insect behavior*, N. 3, 159-168.

[10] S. Dereich and P. Mörters (2009), Random networks with sublinear preferential attachment: Degree evolutions, *Elec. Journ. Probab.*, Vol. 14, N. 43, 1222-1267.

[11] S. Dereich and P. Mörters (2011), Random networks with sublinear preferential attachment: the giant component, preprint.

[12] E. Drinea, A. Frieze and M. Mitzenmacher (2002), Balls and bins models with feedback, *In Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, PA: Society for Industrial and Applied Mathematics, 308-315.

[13] E. Drinea, M. Enachescu and M. Mitzenmacher (2001), Variations on Random Graph Models of the Web,*Harvard Technical Report*, TR-06-01.

[14] B. Hansen and J. Pitman (2000), Prediction rules for exchangeable sequences related to species sampling, *Stats. and Prob Lett.*, N.46, 251-256.

[15] B. M. Hill, D. Lane and W. Sudderth (1987) Exchangeable Urn Processes, *Annals of Probability*, N. 4, 1586-1592

[16] H. Ishwaran and L. F. James (2003), Generalized weighted chinese restaurant processses for species sampling mixture models, *Statistica Sinica*, N 3, 1211-1235.

[17] P. L. Krapivsky and S. L. Redner (2001), Organization of Growing Random Networks, *Physics Reviews E*, N 63, 69-77.

[18] J. Lee, F. A. Quintana, P. Müller and L. Trippa (2011), Defining Predictive Probability Functions for Species Sampling Models, preprint.

[19] T. Móri (2002), On random trees, *Studia Sci. Math. Hungar.*, N 39,143- 155.

[20] R. Oliveira (2009), The Onset of Dominance in Balls-in-Bins Processes with Feedback, *Rand. Struct. and Alg* Vol 34, N 4, 454-477.

[21] R. Oliveira and J. Spencer (2005), Connectivity Transitions in Networks with Super-Linear Preferential Attachment, *Internet Mathematics* Vol 2, Nr 2, 121-163.

[22] R. Pemantle (2007), A survey of random proceses with reinforcement, *Probab. Surv.*, N 4, 1-79.

[23] J. Pitman (1995), Exchangeable and partially exchangeable random partitions, *Probab. Theory Related Fields*, N 102 145-158.

[24] J. Pitman (1996), Some developments of the Blackwell-MacQueen urn scheme, *In Statistics, Probability and Game Theory (Edited by T. S. Ferguson, L. S. Shapley and J. B. MacQueen)*, IMS Lecture Notes-Monograph series, Vol 30, 245-267.

[25] J. Pitman (2006), Combinatorial Stochastic Processes, Springer.

[26] A. Rudas, B. Tóth and B. Valko (2007), Random trees and general branching processes, *Random Structures and Algorithms, N 31, 186-202.*

[27] S. Shah, R. Kothari, Jayadeva, S, Chandra (2010), Trail formation in ants. A generalized Polya urn process, Swarm Intelligence, N 2, 145-171.

[28] T. Zhou (2009) Nonlinear Pòlya Urn Models and Self-Organizing Processes, PhD Thesis University of Pennsylvania. Available at http://www.math.upenn.edu/grad/dissertations/tongzhuThesis.pdf