Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 195 (2021) 445–452

Procedia Computer Science

www.elsevier.com/locate/procedia

XI Latin and American Algorithms, Graphs and Optimization Symposium

Survival for a Galton-Watson tree with cousin mergers

Laura Eslava[a,1,2], Sarah Penington[b,3], Fiona Skerman[c,4]

[a] *Instituto de Investigaciones en Matematicas Aplicadas y en Sistemas. UNAM. Mexico City, Mexico*
[b] *Department of Mathematical Sciences. University of Bath. Bath, UK*
[c] *Mathematics Department. Uppsala University. Uppsala, Sweden*

Abstract

We introduce a generalization of Galton-Watson trees where, individuals have independently a number of Poi$(1 + p)$ offspring and, at each generation, pairs of cousins merge independently with probability q. If $q = 0$ we recover a usual Galton-Watson tree and the survival threshold for the process has $p > 0$. Our main thoerem gives sufficient conditions on p and q for extinction and survival of the Galton-Watson trees with cousin mergers.

In the setting $q > 0$, the Markovian property of regular Galton-Watson trees is lost and so the analysis of the model becomes more involved. In particular, the main obstacle are the intergeneration dependencies since the genealogy of the individuals, having possibly more than one parent, is no longer represented by a tree but by a graph.

Keywords: Branching processes; Survival threshold; Non-Markovian processes.

1. Introduction

Galton-Watson trees are one of the simplest models of branching processes [1]; these were first introduced independently by Bienaymé, and Galton and Watson in the 19th century [2, 4] and represent the genealogy of an asexual population. The process starts with a single individual which has a random number of offspring; each of the children reproduce independently with the same distribution as their parent, and so on. As long as there are individuals to reproduce the process continues; if it continues indefinitely, we say that the process survives. One of the key questions in the area of branching processes is to determine when there is a positive probability of survival.

More precisely, consider a discrete random variable ξ, and a collection $(X_i^{(n)}; i, n \geq 1)$ of independent copies of ξ. A Galton-Watson process with offspring distribution ξ is described by the sequence $(X_n)_{n\geq 0}$ where $X_0 = 1$, and for

[1] This work has been supported by PAPIIT TA100820

[2] Email: laura@sigma.iimas.unam.mx
[3] Email: sp2355@bath.ac.uk
[4] Email: fiona.skerman@math.uu.se

10.1016/j.procs.2021.11.054

$n \geq 1$

$$X_{n+1} = \sum_{i=1}^{X_n} X_i^{(n)}; \tag{1}$$

for $i \leq X_n$, we say that the i-th individual of generation n has $X_i^{(n)}$ children. The event of survival is defined as $\{X_n \geq 1 \text{ for all } n \geq 0\}$, and there is a threshold for the survival probability which, surprisingly, only depends on the expected value of ξ. Namely, if $\mathbb{E}(\xi) \leq 1$ then the probability of survival is zero and if $\mathbb{E}(\xi) > 1$, the probability of survival is positive [1, Chapter 1]. In particular, the survival of the process depends on the expected growth of each generation and not on the genealogy of the individuals within the process.

In this paper, we introduce a simple modification of the construction of Galton-Watson trees, which as we will see in Theorem 1.1, changes the necessary and sufficient conditions for survival. At each generation, we will allow for certain identification of individuals. Loosely speaking, for $p, q \in (0, 1]$, we define $\mathcal{B}(p, q)$ to be the process where individuals independently have a Poi$(1 + p)$ number of offspring and, at each generation, pairs of cousins merge independently with probability q; if $q = 0$ we recover a Galton-Watson tree. A precise description of the construction of $\mathcal{B}(p, q)$ is given in the next section.

The aim is to initiate a line of research of these graph processes (the identification means that the genealogy of the individuals is no longer represented by a tree but by a graph); in particular we study the phases of almost sure extinction and positive survival probability.

1.1. The construction

For $p, q \in (0, 1]$, we define a branching process with merging denoted by $\mathcal{B}(p, q) := \cup_{n\geq 0} G_n$. Each graph G_n represents the first n generations and each vertex corresponds to an individual.

Write $I_n := G_n \setminus G_{n-1}$ for the set of vertices corresponding to individuals in the n^{th} generation. The process starts with G_0 consisting of a single vertex; let $I_0 := G_0$. For $n \in \mathbb{N}$, we construct the n^{th} generation, I_n, and the graph G_n as follows. Each individual $u \in I_{n-1}$ has ξ_u offspring with distribution Poi$(1 + p)$ independently from the rest of the individuals. Form a new graph $\tilde{G}_n$ by adding new vertices to G_{n-1} for each of these offspring, with edges joining them to their parents and denote the set of these new vertices by $\tilde{I}_n$. We define equivalence classes of vertices in $\tilde{I}_n$ as follows. Let $(B_{i,j};\ 1 \leq i < j \leq |\tilde{I}_n|)$ be independent Bernoulli random variables with mean q. For each $1 \leq i < j \leq |\tilde{I}_n|$, if the graph distance between the i^{th} and j^{th} individuals in $\tilde{I}_n$ is exactly 4 and $B_{i,j} = 1$ then identify the two individuals. Finally, form the graph G_n and the set of vertices I_n from $\tilde{G}_n$ and $\tilde{I}_n$ by merging each class of identified vertices into a single vertex, with an edge set given by the union of their edge sets, replacing multi-edges with single edges. See Figure 1 for a representation of a merger. Observe that two individuals in $\tilde{I}_n$ which have graph distance exactly four in $\tilde{G}_n$ share a common grandparent but not a common parent in $\tilde{G}_n$ and so we shall sometimes refer to them as cousins.

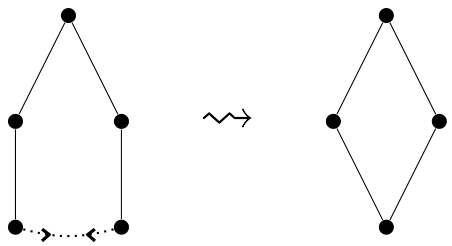

Fig. 1. Merging of cousins. On the left, $\tilde{G}_n$ has a pair of cousins that will merge. On the right, the pair of cousins form a vertex with two parents in G_n.

1.2. Statement of results

Let Z_n be the number of vertices corresponding to individuals in the n^{th} generation. Our main theorem, Theorem 1.1, gives sufficient conditions for extinction and survival of the process.

Theorem 1.1. *There exists $C > 0$ such that the branching process $\mathcal{B}(p, q)$ with $0 < p \leq 1/(2C)$ exhibits (at least) two phases:*

i) If $q < 2p(1 - Cp)$ then the process survives with positive probability;
ii) If $q > 2p(1 + Cp)$ then the process dies out almost surely.

We conjecture that there is a critical threshold for survival of the process; that is, for fixed p, small enough, there is $q_c = q_c(p)$ such that $\mathcal{B}(p, q)$ dies out almost surely if $q > q_c$, and it survives with positive probability if $q < q_c$. This would be an immediate consequence of establishing monotonicity for the survival probability for $\mathcal{B}(p, q)$ as q varies.

We note that the construction of $\mathcal{B}(p, q)$ implies that a simple recurrence between Z_n and Z_{n-1} such as in (1) is not possible since the identification of pairs of individuals now depends on their ancestry. In other words, the sequence $(Z_n)_{n\geq 0}$ is no longer a Markovian process. Despite the subtle dynamics between the expected (initial) offsprings and their identification by pairs of cousins, we are able to obtain sufficiently good estimates on the expected growth per generation.

Our next theorem states, broadly, that Z_n grows, in expectation, by a factor of $(1 + p - \frac{q}{2})$ up to first order terms.

Theorem 1.2. *The generations of $\mathcal{B}(p, q)$ satisfy for $n \geq 2$ and $m \geq 3$,*

$$\mathbb{E}(Z_n) \geq (1 + p)\mathbb{E}(Z_{n-1}) - \frac{q}{2}(1 + p)^4\mathbb{E}(Z_{n-2}), \tag{2}$$

$$\mathbb{E}(Z_m) \leq (1 + p - \frac{q}{2})\mathbb{E}(Z_{m-1}) + 224q^2 \max_{k=2,3} \mathbb{E}(Z_{m-k}). \tag{3}$$

As our result concerns the first order term of the (conjectured) survival threshold, we do not attempt to optimize the constant C in Theorem 1.1. In particular, the large constant in the upper bound of (3) is due to the loose control on the number of multiple mergers caused by the complexity of the genealogy of individuals. As we will see in Lemma 2.1, we circumvent such complexity by restricting the study of mergers to individuals with no mergers up to two generations in their ancestry.

1.3. Outline of the proof

The key to the survival result is to prove that with positive probability, after a constant number of generations the process contains individuals that *renew* the dynamics on the genealogies. This means that we can observe an underlying Galton-Watson process occurring with 'long-range' generations. We make precise this idea in Section 3 and provide the required bounds to have positive probability of having at least one *renewed vertex* every so often. The proof of Theorem 1.1 is based on Theorem 1.2, Proposition 3.2 and Lemma 3.1.

The proof of Theorem 1.2 boils down to a careful inclusion-exclusion process that starts with the assumption that the generations of the process grow, roughly, by a factor $(1 + p)$.

In Section 2 we present almost sure upper and lower bounds for Z_n in terms of new random variables; see Lemma 2.1. The required upper bounds are straightforward computations from a double-counting argument summarized in Proposition 2.2 while the lower bound in Proposition 2.4 involves the analysis of mergers among vertices with no mergers in the two previous generations of their ancestry.

2. Merger stage and proof of Theorem 1.2

For distinct vertices $i, j \in I_n$ (alternatively, $i, j \in \tilde{I}_n$) we write $i \overset{s}{\sim} j$ if these vertices share a common ancestor in I_{n-1} (that is, i and j are siblings) and we write $i \overset{c}{\sim} j$ if they share a common ancestor in I_{n-2} but none in I_{n-1} (that is, i and j are cousins). Write $J_n^{(1)}$ for the set of individuals in I_n which have exactly one parent, and write $\tilde{J}_n^{(2)}$ for the set of offspring in $\tilde{I}_n$ of $J_{n-1}^{(1)}$; that is, that have exactly one grandparent.

Let us consider now the mergers stage. Let $Y_n := |\tilde{I}_n|$; define M_n and M'_n as the number of pairwise mergers in $\tilde{I}_n$ and $\tilde{J}_n^{(2)}$, respectively; also let L'_n be the number of triplets (u, v, w) where vertices are in $\tilde{J}_n^{(2)}$ such that $u \overset{c}{\sim} v$, $v \overset{c}{\sim} w$ are distinct pairs of cousins that merge.

Lemma 2.1. *For all $n \geq 2$,*

$$Y_n - M_n \leq Z_n \leq Y_n - M'_n + L'_n.$$

Proof. Let $\sigma = \{v_1, \dots, v_t\} \subset \tilde{I}_n$ be an equivalence class of offspring that merges into a single individual $v \in I_n$; we say that $t - 1$ of such vertices are *eliminated*. Write m_σ for the number of pairwise mergers within vertices in σ and write ℓ_σ for the number of triplets (v, u, w) of distinct vertices in σ such that $u \overset{c}{\sim} v$, $v \overset{c}{\sim} w$ and both pairs merge. The difference $Y_n - Z_n$ counts the total number of eliminated vertices across all equivalence classes in $\tilde{I}_n$.

First, observe that within each $\sigma = \{v_1, \dots, v_t\}$ there must have occurred at least $t - 1$ mergers; that is, $t - 1 \leq m_\sigma$. Summing over all equivalence classes σ it follows that $Y_n - Z_n \leq M_n$; and so the lower bound on Z_n follows.

Next, let $t' = |\sigma \cap \tilde{J}^{(2)}|$ and define m'_σ and ℓ'_σ as above with the restriction that vertices are contained in $\sigma \cap \tilde{J}^{(2)}$. If $t' = 0$, then $m'_\sigma - \ell'_\sigma = 0 \leq t - 1$. Assume $t' \geq 1$. Consider the graph where the set of vertices is $\sigma \cap \tilde{J}_n^{(2)}$ and the edges are pairs of merged vertices. It is a straightforward fact that a graph with t vertices, m edges and ℓ directed paths of length two satisfies $m - \ell < t$. Thus, $m'_\sigma - \ell'_\sigma \leq t' - 1 \leq t - 1$ for any σ. Summing over all equivalence classes σ yields $M'_n - L'_n \leq Y_n - Z_n$, establishing the desired upper bound on Z_n. □

By a double-counting argument, the number of ordered pairs of cousins in $\tilde{I}_n$ is the same as summing up, for each individual in I_{n-2}, the number of ordered pairs of grandchildren that do not share a common parent. This motivates the following notation, together with the associated bound in Proposition 2.2. Let $j_1, \dots, j_k$ be non-negative integers. For a vertex $u \in I_n$, let $N^{(j_1,\dots,j_k)}(u)$ count the number of distinct vertices $v_1, \dots, v_k$ with $v_i \in I_{n+j_i}$ and whose most recent pairwise ancestor is u. Also, let

$$N_n^{(j_1,\dots,j_k)} := \sum_{u \in I_n} N^{(j_1,\dots,j_k)}(u).$$

Proposition 2.2. *Let $j_1, \dots, j_k$ be positive integers. For any $n \geq 0$,*

$$\mathbb{E}(N_n^{(j_1,\dots,j_k)}) \leq (1 + p)^{\sum j_i} \mathbb{E}(Z_n).$$

Proof. By linearity of the expectation, it suffices to prove the uniform bound

$$\mathbb{E}(N^{(j_1,\dots,j_k)}(u)) \leq (1 + p)^{\sum j_i}; \tag{4}$$

for any $n \geq 0$ and $u \in I_n$. Without loss of generality, we can assume that $j_1 \geq j_2 \geq \cdots \geq j_k \geq 1$.

We will prove (4) by induction on j_1. Let us first consider the case $N^{(1,\dots,1)}(u)$ counts the number of k-tuples of distinct children of u. The number of offspring ξ_u may only decrease after the merger stage due to multiple mergers; that is, ξ_u is stochastically dominated by a $\mathrm{Poi}(1 + p)$ random variable. We then have uniformly for u,

$$\mathbb{E}(N^{(1,\dots,1)}(u)) \leq \mathbb{E}(\xi_u(\xi_u - 1) \cdots (\xi_u - k + 1)) \leq (1 + p)^k.$$

For the induction step, suppose that $j_1 \geq 2$ and let $\ell = \max\{i : j_i = j_1\}$. Set $j'_i = j_i - \mathbf{1}_{(i \leq \ell)} \geq 1$ for $1 \leq i \leq k$. Denote by $\mathrm{mrca}(u, w)$ to the most recent common ancestor of u and w and let

$$\mathcal{N}^{(j_1,\dots,j_k)}(u) := \{(v_1, \dots, v_k) \in \prod_{i=1}^{k} I_{n+j_i} : v_i \neq v_j, \ \mathrm{mrca}(v_i, v_j) = u \text{ for all } i \neq j\},$$

and similarly for $\mathcal{N}^{(j'_1,\dots,j'_k)}(u)$. Then for each $(v_1, \dots, v_k) \in \mathcal{N}^{(j_1,\dots,j_k)}(u)$ there is at least one $(w_1, \dots, w_k) \in \mathcal{N}^{(j'_1,\dots,j'_k)}(u)$; observe that there are multiple choices when v_i has more than one parent for some $i \leq \ell$. It follows that

$$\begin{aligned}\mathbb{E}(N^{(j_1,\dots,j_k)}(u)|G_{n+j_1-1}) &\leq \sum_{(w_1,\dots,w_k) \in \mathcal{N}^{(j'_1,\dots,j'_k)}(u)} \prod_{i=1}^{\ell} \mathbb{E}(\xi_{w_i}|G_{n+j_1-1}) \\ &\leq N^{(j'_1,\dots,j'_k)}(u)(1 + p)^\ell;\end{aligned}$$

which, by taking expectations, establishes the induction step for (4). □

Lemma 2.3. *For $n \geq 2$,*

$$\mathbb{E}(M_n) \leq \frac{q}{2}(1+p)^4 \mathbb{E}(Z_{n-2}), \tag{5}$$

$$\mathbb{E}(L'_n) \leq 96q^2 \mathbb{E}(Z_{n-2}). \tag{6}$$

Proof. For $\mathbb{E}(M_n)$, observe that the number of (unordered) pairs of cousins is counted by $\frac{N^{(2,2)}_{n-2}}{2}$ and recall that mergers occur with probability q. By Proposition 2.2 we have

$$\mathbb{E}(M_n) = q\mathbb{E}\left(\frac{N^{(2,2)}_{n-2}}{2}\right) \leq \frac{q}{2}(1+p)^4 \mathbb{E}(Z_{n-2}).$$

For $\mathbb{E}(L'_n)$, recall that we consider triplets (u, v, w) with vertices in $\tilde{J}^{(2)}_n$ such that $u \overset{c}{\sim} v$ and $v \overset{c}{\sim} w$. Since mergers are independent, each triplet of distinct vertices counts towards L'_n with probability q^2; so it remains to provide an overcount for the number of such triples (u, v, w). We divide them according to whether u and w share a common parent.

First, consider a child-aunt pair $(v, u') \in I_n \times I_{n-1}$ and a pair of offsprings (u, w) of u'; then the triplet (u, v, w) may be considered (if the vertices of the triplet are in $\tilde{J}^{(2)}_n$). The pairs (v, u') are counted by $N^{(2,1)}_{n-2}$ and for each u' the expected number of pairs (u, w) is $(1+p)^2$. On the other hand, if u and w do not share a common parent, then the triplet (u, v, w) has a common grandparent in I_{n-2} but distinct parents; this is overcounted by $N^{(2,2,2)}_{n-2}$. Placing these observations together with the bounds from Proposition 2.2 we have

$$\begin{aligned}\mathbb{E}(L'_n) &\leq q^2(1+p)^2 \mathbb{E}(N^{(2,1)}_{n-2}) + q^2 \mathbb{E}(N^{(2,2,2)}_{n-2}) \\ &\leq q^2(1+p)^5 \mathbb{E}(Z_{n-2}) + q^2(1+p)^6 \mathbb{E}(Z_{n-2});\end{aligned}$$

the bound in (6) is obtained using $1 + p \leq 2$ on the last estimate above. □

Proposition 2.4. *For $n \geq 3$,*

$$\mathbb{E}(M'_n) \geq \frac{q}{2}(1+p)^4 \mathbb{E}(Z_{n-2}) - 128q^2 \mathbb{E}(Z_{n-3}).$$

Proof. Recall that M'_n counts the number of pairwise mergers between $w, w' \in \tilde{J}^{(2)}_n$ with $w \overset{c}{\sim} w'$.

Consider $w, w' \in J^{(2)}_n$. By definition, these vertices have exactly one parent, say $v, v' \in I_{n-1}$ respectively, and one grandparent $u \in I_{n-2}$. That is; ordered pairs (w, w') map to (v, v', u) where $v \overset{s}{\sim} v'$ and u is the unique parent of both v and v'. The number of such triples (v, v', u) is counted by $S_{n-1} := |\mathcal{S}_{n-1}|$ where

$$\mathcal{S}_n := \{(v, v') \subset J^{(1)}_n : v \overset{s}{\sim} v'\}.$$

We first obtain a lower bound on $\mathbb{E}(S_n)$. The pairs of ordered siblings $\tilde{u}_1, \tilde{u}_2$ in $\tilde{I}_n$ are simply counted by $\sum_{u \in I_{n-1}} \xi_u(\xi_u - 1)$, which has mean $(1+p)^2 \mathbb{E}(Z_{n-1})$. Now, a pair of ordered siblings $\tilde{u}_1, \tilde{u}_2$ in $\tilde{I}_n$ contributes to S_n unless one of the siblings merges with a cousin. In this case, there is $\tilde{v} \in \tilde{I}_n$ and $w \in I_{n-2}$ which is grandparent of $\tilde{u}_1, \tilde{u}_2$ and $\tilde{v}$, and $\tilde{u}_1 \overset{c}{\sim} \tilde{v}$. So the number of ordered pairs $(\tilde{u}_1, \tilde{u}_2)$ that do not contribute to S_n is at most twice the number of 4-tuples (u_1, u_2, v, w) such that all vertices are ancestors of w, $\tilde{u}_1 \overset{s}{\sim} \tilde{u}_2$, $\tilde{u}_1 \overset{c}{\sim} \tilde{v}$ and the last pair merges. Using Proposition 2.2 we have that the expected number of such 4-tuples is

$$q\mathbb{E}\Big(\sum_{v, v' \in I_{n-1}, v \overset{s}{\sim} v'} \xi_{v'} \xi_v(\xi_v - 1)\Big) \leq q(1+p)^3 \mathbb{E}(N^{(1,1)}_{n-1}) \leq q(1+p)^5 \mathbb{E}(Z_{n-1}).$$

Putting these two observations together, and using the fact that $1 + p < 2$ we get

$$\mathbb{E}(S_n) \geq (1+p)^2 \mathbb{E}(Z_{n-1}) - 64q \mathbb{E}(Z_{n-2}). \tag{7}$$

Finally, M'_n has the distribution of a sum of Bernoulli random variables with mean q, one variable for each (unordered) pair of cousins in $\tilde{J}^{(2)}_n$. In the following estimate we use the mapping of pairs (w, w') described above. We also use that the offspring variables ξ_v are independent to get,

$$\mathbb{E}(M'_n) = \frac{q}{2}\mathbb{E}\Big(\sum_{\substack{v,v' \in J^{(1)}_{n-1} \\ v \overset{s}{\sim} v'}} \xi_v \xi_{v'}\Big) = \frac{q}{2}(1+p)^2 \mathbb{E}(S_{n-1});$$

we obtain the desired lower bound by applying the lower bound on $\mathbb{E}(S_{n-1})$ given by (7) and using $1 + p < 2$. □

Proof of Theorem 1.2. Lemma 2.1 gives $Y_n - M_n \le Z_n \le Y_n - M'_n + L'_n$, for $n \ge 2$. We will take expectations and use the estimates from Lemma 2.3 and Proposition 2.4; clearly, $\mathbb{E}(Y_n) = (1 + p)\mathbb{E}(Z_{n-1})$. Thus, for $n \ge 2$, (2) holds directly. For $n \ge 3$ we get

$$\begin{aligned}\mathbb{E}(Z_n) &\le (1+p)\mathbb{E}(Z_{n-1}) - \frac{q}{2}(1+p)^4\mathbb{E}(Z_{n-2}) + 224q^2 \max_{k=2,3} \mathbb{E}(Z_{n-k}) \\ &\le (1 + p - \frac{q}{2})\mathbb{E}(Z_{n-1}) + 224q^2 \max_{k=2,3} \mathbb{E}(Z_{n-k});\end{aligned}$$

where we use $(1 + p)^4\mathbb{E}(Z_{n-2}) \ge (1 + p)\mathbb{E}(Z_{n-1}) \ge \mathbb{E}(Z_{n-1})$; completing the proof. □

3. Renewed vertices and proof of Theorem 1.1

An individual v in generation $n \ge 2$ is called a *renewed individual* if there are exactly two individuals in G_n at distance at most two from v; in genealogical terms, a renewed individual has exactly one parent and no siblings (see Figure 2). For $n \ge 2$, let $\mathcal{R}_n := \{v_1, \ldots, v_m\}$ be the set of renewed vertices in generation n and let $R_n := |\mathcal{R}_n|$.

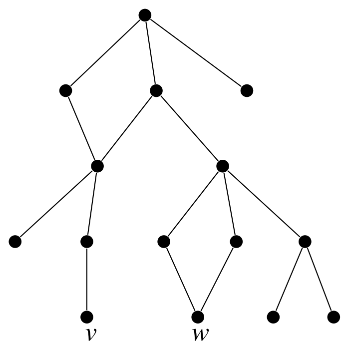

Fig. 2. An instance of $\mathcal{B}(p, q)$ up to the fourth generation. Vertex v is a renewed vertex, while vertex w is a merger of two only-child individuals.

Lemma 3.1. *Let $n \ge 2$. For $v \in \mathcal{R}_n$, let $\mathcal{B}_v$ be the graph process representing v and its descendants in $\mathcal{B}$. Conditional on G_n, the processes $\{\mathcal{B}_v\}_{v\in\mathcal{R}_n}$ are independent copies of $\mathcal{B}$.*

Proof. Suppose individual v in generation $N \ge 2$ is a renewed vertex. Since v had no siblings, the children of v have no cousins and hence cannot merge, playing the same role as vertices in generation 1. Their offspring in generation $N+2$ have v as their unique grandparent; similarly, every descendant of v in subsequent generations, after any mergers occur, will have v as their sole ancestor in generation N. This implies that the law of the descendants of renewed vertices becomes independent and identically distributed as $\mathcal{B}(p, q)$. □

Proposition 3.2. *For $n \ge 2$, if $q \le 2p \le 2^{-5}e^{-1}$,*

$$\mathbb{E}(R_n) \ge e^{-1}\mathbb{E}(Z_{n-1}) - (2e)^{-1}\mathbb{E}(Z_{n-2}).$$

Proof. A vertex $v \in I_n$ is a renewed vertex if its parent has exactly one offspring and v does not merge with any cousin. Therefore we have that

$$\mathbb{E}(R_n) \ge (1 + p)e^{-(1+p)}\mathbb{E}(Z_{n-1}) - 2\mathbb{E}(M_n);$$

where the first term in the bound is the expected number of parents in I_{n-1} that have a unique offspring and the second term is a loose bound on the number of such vertices lost by a merging event (the worst case being a merger between two only-child cousins). By (5), $2\mathbb{E}(M_n) \leq q(1+p)^4\mathbb{E}(Z_{n-2})$, while the assumptions on p and q gives $q(1+p)^4 \leq q2^4 \leq (2e)^{-1}$. Finally, $(1+p)e^{-(1+p)} \geq e^{-1}$ for all $p \leq 1$, completing the proof. $\square$

Proof of Theorem 1.1. Let $C' = 224$, $C = 3C'$ and $p_0 < \max\{(6C')^{-1}, 2^{-5}e^{-1}\}$. We start with two technical bounds. Straightforward computations give

$$(1+p)^4(1-Cp) < (1+15p)(1-Cp) < 1+(15-C)p < 1;$$

in other words, $(1-Cp) < (1+p)^{-4}$ and so we infer that

$$b := 1 + p - \frac{q}{2}(1+p)^4 > 1. \tag{8}$$

Now, if $p < 1/6C'$ then

$$(1+3C'p)(1-2C'p) = 1 + C'p(1-6C'p) > 1,$$

which implies that whenever $p \leq p_0$, $(1+3C'p) \leq (1-2C'p)^{-1}$ and, in this case

$$a := 1 + p - \frac{q}{2} + C'q^2 < 1. \tag{9}$$

We first consider the case of extinction. Let $q > 2p(1+Cp)$ and a as in (9). Now, for $n \geq 3$, we can relax (3) to

$$\mathbb{E}(Z_n) \leq a \max_{k\in\{1,2,3\}} \mathbb{E}(Z_{n-k}). \tag{10}$$

Next, the assumption on $q < 1$ implies that $a > 0$, observe that (10) applied to $\mathbb{E}(Z_{n+1})$ and $\mathbb{E}(Z_n)$ gives

$$\begin{aligned}
\mathbb{E}(Z_{n+1}) &\leq a \max_{k\in\{1,2,3\}} \mathbb{E}(Z_{n+1-k}) \\
&= a \max\left(\mathbb{E}(Z_n), \max_{l=1,2} \mathbb{E}(Z_{n-l})\right) \\
&\leq a \max_{k\in\{0,1,2\}} \mathbb{E}(Z_{n-k}).
\end{aligned}$$

Similarly, we obtain, $\mathbb{E}(Z_{n+2}) \leq a \max_{k\in\{0,1,2\}} \mathbb{E}(Z_{n-k})$ and so

$$\max_{l\in\{0,1,2\}} \mathbb{E}(Z_{n+l}) \leq a \max_{k\in\{0,1,2\}} \mathbb{E}(Z_{n-k}).$$

An iterative argument yields

$$\max_{l\in\{0,1,2\}} \mathbb{E}(Z_{3m+l}) \leq a^m \max_{k\in\{0,1,2\}} \mathbb{E}(Z_k).$$

Letting $m \to \infty$, since $a \in (0,1)$, we infer that $\mathbb{E}(Z_n) \to 0$ as $n \to \infty$, therefore,

$$\mathbb{P}(Z_n > 0\ \forall n \in \mathbb{N}) \leq \mathbb{P}(Z_n > 0) \leq \mathbb{E}(Z_n) \to 0;$$

that is, extinction occurs with probability 1.

We next consider the case of survival. Let $q < 2p(1-Cp)$ and $b > 1$ as in (8). Observe that $\mathbb{E}(Z_1) = 1 + p > 1 = \mathbb{E}(Z_0)$ and so (2) reduces to $\mathbb{E}(Z_2) \geq b\mathbb{E}(Z_1)$. Then $\mathbb{E}(Z_2) \geq \mathbb{E}(Z_1)$ and an induction argument gives, for any $n \geq 2$, $\mathbb{E}(Z_n) \geq b\mathbb{E}(Z_{n-1}) \geq b^{n-1}(1+p)$; which implies $\mathbb{E}(Z_n) \to \infty$ as $n \to \infty$.

Moreover, we also have that $\mathbb{E}(Z_n)$ is increasing which, together with Proposition 3.2, implies that there is N large enough that

$$\mathbb{E}(R_N) \geq e^{-1}\mathbb{E}(Z_{N-1}) - (2e)^{-1}\mathbb{E}(Z_{N-2}) \geq (2e)^{-1}\mathbb{E}(Z_{N-1}) > 1.$$

Now, define a Galton-Watson process $(X_n)_{n\geq 0}$ as follows. Let $X_0 = 1$ and for $n \geq 1$, let

$$X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)},$$

where $\{\xi_j^{(n)}, n \in \mathbb{N}, j \in \mathbb{N}\}$ are i.i.d. with $\xi_j^{(n)} \stackrel{d}{=} R_N$. The choice of N implies that X_n is a supercritical branching process, that is $\mathbb{P}(X_n > 0\ \forall n \in \mathbb{N}) > 0$. By Lemma 3.1, we can couple $(Z_n)_{n\geq 0}$ with $(X_n)_{n\geq 0}$ in such a way that $Z_{nN} \geq X_n\ \forall n \in \mathbb{N}$. Finally, this implies that

$$\begin{aligned}\mathbb{P}(Z_n > 0\ \forall n \in \mathbb{N}) &\geq \mathbb{P}(Z_{nN} > 0\ \forall n \in \mathbb{N})\\ &\geq \mathbb{P}(X_n > 0\ \forall n \in \mathbb{N}) > 0;\end{aligned}$$

that is, there is a positive probability of survival. □

4. Conclusion

In this paper we introduce a new model $\mathcal{B}(p, q)$ that generalizes Galton-Watson trees by allowing, at each generation, pairs of cousins to merge into a single individual. We analyse the survival and extinction conditions for the particular case of Poisson offspring distribution, Poi$(1 + p)$, and constant probability of cousin merging q. Our main result gives sufficient conditions on q, in terms of p, to either have extinction occurring with probability 1 or survival with positive probability (Theorem 1.1).

Contrary to the case of Galton-Watson trees, where the structure of the tree is irrelevant to determine the survival of the process, in the case of the Galton-Watson tree with cousin merges, the structure of the graph plays a key role in determining the expected growth of each generation and a inclusion-exclusion argument is applied to bound the generation sizes Z_n (Theorem 1.2).

4.1. Open questions

As a novel graph process, there are plenty of open questions, for example,

i) what is the distribution of individuals with at least three parents?
ii) what is the distribution of the largest cycle of the graph G_n?
iii) when do the individuals that have no mergers in their ancestry form a supercritical process?
iv) establish the monotonicity of the survival probability for $\mathcal{B}(p, q)$
v) establish the existence of a critical $q_c = q_c(p)$ for the survival of the process,
vi) extend the results to a broader class of offspring distributions.
vii) couple the model to graph exploration for random graphs.

We close this section by giving further remarks on the last line of research. Galton-Watson trees and, more generally, branching processes are linked with the graph exploration of connected components in random graphs (see, e.g. [5, Chapter 4]). It would be interesting to explore whether there is a coupling between graph exploration of distint graph classes and variants of the Galton-Watson trees with cousin mergers. In [3], the authors address the necessary adjustments to this first toy model so that the (conjectured) critical threshold for the survival matches the expansion of the critical threshold for percolation in both site and bond percolation of hypercubic graphs of large dimension.

References

[1] Athreya, K. B., Ney, P.E., 1972. “Branching Processes,” Math. Surveys **7**, Springer-Verlag, Berlin.
[2] Bienaymé, I. J., 1845. De la loi de multiplication et de la duree des families, Soc. Philomat. Paris Extraits, Ser. **5**, 37-39.
[3] Eslava, L., Penington, S., Skerman, F. Infanticide and Incest: a branching process with deletions and mergers that matches the threshold for hypercube percolation, In preparation.
[4] Galton, F., Watson, H.W., 1875. On the probability of the extinction of families. Journal of the Royal Anthropological Institute 4, 138–144.
[5] Hofstad, R., 2016. “Random Graphs and Complex Networks,” (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge: Cambridge University Press.