

FURTHER PROPERTIES OF A RANDOM GRAPH MODEL DRIVEN BY TIME-DEPENDENT BRANCHING DYNAMICS

Tamás F. Móri and Sándor Rokob

(Budapest, Hungary)

Communicated by Imre Kátai

(Received March 5, 2018; accepted May 9, 2018)

Abstract. We continue the analysis of a recently introduced increasing graph process. This model is motivated by pairwise collaborations, and it is driven by time-dependent branching dynamics. Our model is different from most of the similar ones in the sense that edges, not vertices, are subject to branching. In this note, we consider a slightly generalized version of the original model, and compare the stable age distributions of the edges for physical, and for biological age. The former is simply the time elapsed from birth, while the latter is measured by the number of offspring, and the death rate is connected with it. Somewhat surprising, we find that the tail behaviour of the two distributions is completely different, although the expected value of the biological age is a linear function of the physical age.

1. Introduction

In [3] we introduced and analyzed a random graph process. It was motivated by the following words: “Consider an increasing group of individuals who

Key words and phrases: Evolving random graph, Crump–Mode–Jagers process, stable age distribution.

2010 Mathematics Subject Classification: 05C80, 60J80.

This work was supported by the Hungarian National Research, Development and Innovation Office NKFIH, grant number K 125569 (T.F. Móri), and the ÚNKP-17-2 New National Excellence Program of the Ministry of Human Capacities (S. Rokob).

are organized by pairwise collaborations. A successful collaboration attracts newcomers, who start collaborating with one or both participants. However, the new connections can weaken and exhaust the attracting pair's collaboration, which eventually ceases. Representing collaborations with edges, we get an evolving graph process."

The mathematical model starts with a single edge. Edges are to produce new edges till death. For every edge, the birth events form a homogeneous Poisson process of unit density. At every birth event a new vertex appears, which gets connected to one or both endpoints of the parent edge. Thus the offspring size is either 2, it happens with probability p , or 1, with probability $q = 1 - p$. The *physical age* of an edge is the time elapsed from its birth, and its *biological age* is defined as the number of its offspring up to the moment. The life length of an edge is connected with its biological age: the hazard rate at physical age t and biological age $\xi(t)$ is equal to $b + c\xi(t)$, where b and c are positive constants. The hazard rate $r(t)$ is the (conditional) rate of immediate death at time t provided the edge is still alive. The distribution of a random life span λ is completely determined by its hazard rate function $r(t)$ through the formula

$$P(\lambda > t) = \exp\left(-\int_0^t r(s) ds\right).$$

At death the edge is deleted (but not its endpoints). The life histories of different edges are assumed independent. Thus the edge process is a well known mathematical object: a so called general time-dependent branching process or *Crump-Mode-Jagers* (CMJ) process.

In [3] we dealt with some asymptotic properties of this random graph process. This time we are interested in the so called stable age distribution: this is the limit as $t \rightarrow \infty$ of the proportion of edges with physical age exceeding s , $0 < s < \infty$. We also compute the same quantity for biological age. We will show that the tail behaviour of the two age distributions is significantly different. We will prove it in a generalized model, where the random number ε of offspring at a birth event can take on more than two values: it can be an arbitrary bounded positive integer valued random variable, with a sufficiently light tail.

Generalized model. The offspring distribution is

$$P(\varepsilon = i) = p_i, \quad i = 1, 2, \dots$$

Suppose that the tail of the distribution of ε is lighter than exponential: $\lim_{k \rightarrow \infty} \frac{1}{k} \log(1/p_k) = +\infty$, or equivalently, the generating function $g(z) = E(z^\varepsilon)$ is finite on the whole real line.

It is plausible to assume that the newborn edges share one vertex with their parent, but as long as we do not want to investigate the properties of vertices we do not need to fix how many new edges should be added to the graph, or how the subgraph they form should look like. However, the analysis of the relationship between the physical and the biological ages can provide information on the graph structure too, for example, on the clustering properties. The higher is the biological age of an edge, the larger are the degrees of its endpoints. Short edge life spans lead to smaller vertex degrees, etc. We are planning to return to these issues in a forthcoming paper.

There are many papers on random graphs and networks with heuristic computations instead of exact, mathematically rigorous proofs. Such heuristics often lead to (but not prove) correct results, but there exist interesting counterexamples, too. A favourite method for getting rid of complicated mathematical inferences is to replace random quantities with their expectations. In our case, the expectation of the biological age is a linear function of the physical age, but plugging a linear transform into the stable age distribution yields a completely false approximation to the asymptotic distribution of the biological age.

2. Preliminaries

In this section we briefly summarize the notions, notations and theorems we will use from the general theory of CMJ processes, together with some results proved in [3].

Let us recall the notations of [3]. For a generic edge let $\pi(t)$ denote the Poisson process of birth events. The random times of successive birth events are denoted by τ_1, τ_2, \dots . The number of progenies at the i th birth is ε_i , and the cumulative sum $S_i = \varepsilon_1 + \dots + \varepsilon_i$ gives the biological age of the parent after τ_i . The life span of the edge is denoted by λ , with distribution function $L(t) = P(\lambda \leq t)$. The stopped compound Poisson process $\xi(t) = S_{\pi(t \wedge \lambda)}$ is called the reproduction process. By Wald's identity we have

$$\mu(t) := E\xi(t) = E(\varepsilon)E(\lambda \wedge t) = E(\varepsilon_1) \int_0^t [1 - L(s)] ds.$$

The history of every edge e is characterized by the triple

$$(\lambda_e, \pi_e(\cdot), \xi_e(\cdot))$$

which are iid copies of $(\lambda, \pi(\cdot), \xi(\cdot))$ introduced above. If edge e was born at time σ_e , then at time t its biological age is $\xi_e(t - \sigma_e)$ (defined as zero for $t < \sigma_e$), and the edge is deleted at time $\sigma_e + \lambda_e$.

CMJ processes are often analyzed by the help of *random characteristics*. A random characteristic is a stochastic process $\phi(\cdot)$, which is connected with the history of an edge. It is supposed to change exclusively during the lifetime: $\phi(t) = 0$ for $t \leq 0$ and $\phi(t) = \phi(\lambda)$ for $t \geq \lambda$, and it doesn't need to be independent of $\lambda, \pi(\cdot)$, and $\xi(\cdot)$. The processes $\phi_e(\cdot)$ belonging to different edges are iid. The stochastic process

$$Z^\phi(t) = \sum_e \phi_e(t - \sigma_e),$$

is called *the time-dependent branching process counted by random characteristic*. For example, $\phi(t) = I(0 \leq t < \lambda)$ produces $Z^\phi(t) = Z(t)$, the number of edges at time t , while in the original model of [3], where every birth event brings exactly one new vertex to the graph, the number $V(t)$ of vertices is by 2 greater than $Z^\phi(t)$ for the random characteristic $\phi(t) = \pi(t \wedge \lambda)$.

For certain values of the parameters b, c, p_1, p_2, \dots it can happen that the process dies out: eventually all edges get deleted. From the classical theory of discrete Galton–Watson branching processes we know that the probability of extinction is equal to 1 if the mean number of offspring of an edge during its lifetime, $E\xi(\infty)$, is less than or equal to 1 (subcritical and critical cases, respectively). In the supercritical case, i.e., when every edge produces more than 1 new edge on the average, the probability of extinction is strictly less than 1. In this paper, we only deal with supercritical processes, on the event of non-extinction.

A general treatment of CMJ processes is found in the classical book of Jagers [2], or in the more recent monograph [1]. A brief summary is presented in [3, Section 3]. Here we only refer to a deep result of Nerman [4].

In our model, the processes $Z^\phi(t)$ grow exponentially fast on the event of non-extinction. The exponent of growth is the so called *Malthusian parameter*, denoted by α . It is the only positive solution of the equation

$$(2.1) \quad \int_0^\infty e^{-\alpha t} \mu(dt) = 1.$$

Theorem 2.1. [4, Theorem 6.3] *Suppose $1 < E\xi(\infty) < \infty$. Let ϕ and ψ be random characteristics, both satisfying the following conditions (only formulated for ϕ).*

- (i) $\phi(t) \geq 0$,

- (ii) *the trajectories of ϕ belong to the Skorohod D -space, that is, they do not have discontinuities of the second kind,*
- (iii) $E[\sup_t \phi(t)] < \infty$,

Then

$$\lim_{t \rightarrow \infty} \frac{Z^\phi(t)}{Z^\psi(t)} = \frac{\int_0^\infty e^{-\alpha t} E\phi(t) dt}{\int_0^\infty e^{-\alpha t} E\psi(t) dt}$$

almost everywhere on the event of non-extinction.

3. Stable age distribution

Firstly, let us deal with the distribution of the biological age.

Let $g(z) = E(z^\varepsilon)$ and $G_t(z)$, $0 \leq z \leq 1$ denote the probability generating functions of the offspring size and the empirical distribution of biological age at time t , respectively. Consider the uniformly bounded random characteristic $\phi(t) = z^{\xi(t)} I(0 \leq t < \lambda)$. Clearly, $G_t(z) = Z^\phi(t)/Z(t)$. In order to apply Theorem 2.1 we have to compute $E\phi(t)$.

Theorem 3.1.

$$\begin{aligned} E\phi(t) &= \exp\left(- (1+b)t + t \int_0^1 g(e^{-cts} z) ds\right) = \\ &= \exp\left(- (1+b)t + \frac{1}{c} \int_{e^{-ct}}^1 \frac{g(vz)}{v} dv\right). \end{aligned}$$

Before proving this theorem, let us list a few of its consequences.

By setting $z = 1$ we get the survival function of the lifetime λ .

Corollary 3.1.

$$1 - L(t) = P(\lambda > t) = \exp\left(- (1+b)t + \frac{1}{c} \int_{e^{-ct}}^1 \frac{g(v)}{v} dv\right).$$

In our model $\mu(\infty) = E\varepsilon \int_0^\infty [1 - L(t)] dt$. Let us substitute $u = e^{-ct}$ in the integral, then we obtain the following condition of supercriticality.

Corollary 3.2. *Our process is supercritical if and only if*

$$\frac{E\varepsilon}{c} \int_0^1 u^{\frac{1+b}{c}-1} \exp\left(\frac{1}{c} \int_u^1 \frac{g(v)}{v} dv\right) du > 1.$$

The same substitution in the integral involved in the Malthusian equation (2.1) yields the following expression.

Corollary 3.3. *The Malthusian parameter α is determined by the equation*

$$\frac{E\varepsilon}{c} \int_0^1 u^{\frac{\alpha+1+b}{c}-1} \exp\left(\frac{1}{c} \int_u^1 \frac{g(v)}{v} dv\right) du = 1.$$

Proof of Theorem 3.1. The proof follows the lines of the proof of [3, Theorem 4.1]. There we have seen that

$$P(\lambda > t \mid \pi(t) = k, \tau_1, \dots, \tau_k, \varepsilon_1, \dots, \varepsilon_k) = \exp\left(c \sum_{i=1}^k \varepsilon_i \tau_i - (b + cS_k)t\right),$$

hence

$$\begin{aligned} E(\phi(t) \mid \pi(t) = k, \tau_1, \dots, \tau_k, \varepsilon_1, \dots, \varepsilon_k) &= \\ &= z^{S_k} P(\lambda > t \mid \pi(t) = k, \tau_1, \dots, \tau_k, \varepsilon_1, \dots, \varepsilon_k) = \\ &= z^{S_k} \exp\left(c \sum_{i=1}^k \varepsilon_i \tau_i - (b + cS_k)t\right), \end{aligned}$$

Given $\pi(t) = k$, the conditional joint distribution of the birth times τ_1, \dots, τ_k coincides with that of $(tU_1^{(k)}, \dots, tU_k^{(k)})$, where $(U_1^{(k)}, \dots, U_k^{(k)})$ is an ordered sample of size k from the uniform distribution $U(0, 1)$ (and independent of the litter sizes $\varepsilon_1, \dots, \varepsilon_k$). Thus,

$$\begin{aligned} E(\phi(t) \mid \pi(t) = k, \varepsilon_1, \dots, \varepsilon_k) &= \\ &= z^{S_k} E\left(\exp\left(-bt + ct \sum_{i=1}^k \varepsilon_i (U_i^{(k)} - 1)\right) \middle| \varepsilon_1, \dots, \varepsilon_k\right) = \\ &= z^{S_k} E\left(\exp\left(-bt + ct \sum_{i=1}^k \varepsilon_i (U_i - 1)\right) \middle| \varepsilon_1, \dots, \varepsilon_k\right) \end{aligned}$$

by the interchangeability of $\varepsilon_1, \dots, \varepsilon_k$, where U_1, \dots, U_k are iid random variables with uniform distribution $U(0, 1)$, and they are independent of the ε 's.

Thus, by independence,

$$\begin{aligned} E(\phi(t) | \pi(t) = k) &= e^{-bt} E \left(\prod_{i=1}^k \left(e^{ct(U_i-1)} z \right)^{\varepsilon_i} \right) = \\ &= e^{-bt} \left[E \left(e^{ct(U_1-1)} z \right)^{\varepsilon_1} \right]^k. \end{aligned}$$

By the law of total expectation we have

$$E(\phi(t) | \pi(t) = k) = e^{-bt} \left[E g \left(e^{ct(U_1-1)} z \right) \right]^k = e^{-bt} \left[\int_0^1 g(e^{-cts} z) ds \right]^k,$$

therefore

$$\begin{aligned} E\phi(t) &= \sum_{k=0}^{\infty} P(\pi(t) = k) E(\phi(t) | \pi(t) = k) = \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{-t} \cdot e^{-bt} \left[\int_0^1 g(e^{-cts} z) ds \right]^k = \\ &= \exp \left(- (1+b)t + t \int_0^1 g(e^{-cts} z) ds \right), \end{aligned}$$

as claimed. The second line can be obtained by substituting $v = e^{-cts}$ in the inner integral. \blacksquare

Theorem 3.2. *The probability generating function $G_t(z)$ of the empirical distribution of the biological age converges to*

$$\begin{aligned} \frac{E\varepsilon}{c} \int_0^1 u^{\frac{\alpha+b+1}{c}-1} \exp \left(\frac{1}{c} \int_u^1 \frac{g(vz)}{v} dv \right) du = \\ = \frac{E\varepsilon}{c} \int_0^1 u^{\frac{\alpha+b+1}{c}-1} \exp \left(\sum_{i=1}^{\infty} \frac{p_i}{ci} (1-u^i) z^i \right) du, \quad 0 \leq z \leq 1. \end{aligned}$$

as $t \rightarrow \infty$, almost everywhere on the event of non-extinction.

Proof. As remarked before, $G_t(z) = Z^\phi(t)/Z^\psi(t)$ with random characteristics $\phi(t) = z^{\xi(t)} I(0 \leq t < \lambda)$ and $\psi(t) = I(0 \leq t < \lambda)$. By Theorem 2.1 it converges

almost everywhere on the event of non-extinction to

$$\frac{\int_0^{\infty} e^{-\alpha t} E\phi(t) dt}{\int_0^{\infty} e^{-\alpha t} E\psi(t) dt}.$$

In the denominator

$$(3.1) \quad \int_0^{\infty} e^{-\alpha t} E\psi(t) dt = \int_0^{\infty} e^{-\alpha t} [1 - L(t)] dt = \frac{1}{E\varepsilon} \int_0^{\infty} e^{-\alpha t} \mu(dt) = \frac{1}{E\varepsilon},$$

by the Malthusian equation (2.1).

Coming to the numerator, by Theorem 3.1 we have

$$\begin{aligned} \int_0^{\infty} e^{-\alpha t} E\phi(t) dt &= \int_0^{\infty} \exp\left(-(\alpha + 1 + b)t + t \int_0^1 g(e^{-cts} z) ds\right) dt \\ &= \int_0^{\infty} \exp\left(-(\alpha + 1 + b)t + \frac{1}{c} \int_{e^{-ct}}^1 \frac{g(vz)}{v} dt\right) dt, \end{aligned}$$

and the proof can be completed by substituting $u = e^{-ct}$, and computing the inner integral. \blacksquare

It is easy to obtain the individual probabilities from the generating function of the limit distribution. All we have to do is expanding the integrand into Taylor series by the powers of z , then integration and summation can be interchanged.

Corollary 3.4. *Let $X_t(k)$ denote the (random) proportion of edges with biological age k at time t , $k = 0, 1, \dots$. Then, almost everywhere on the event of non-extinction,*

$$\begin{aligned} \lim_{t \rightarrow \infty} X_t(k) &= \text{coeff} \left(\lim_{t \rightarrow \infty} G_t(z), z^k \right) = \\ &= \frac{E\varepsilon}{c} \sum \frac{1}{i_1! \dots i_k!} \int_0^1 u^{\frac{\alpha+b+1}{c}-1} \prod_{j=1}^k \left[\frac{(1-u^j)p_j}{cj} \right]^{i_j} du, \end{aligned}$$

where the sum is taken over all sequences i_1, i_2, \dots, i_k of non-negative integers such that $i_1 + 2i_2 + \dots + ki_k = k$ is satisfied.

Let r_k denote the limit above. Next, we will show that the tail probability $r_k + r_{k+1} + \dots$ decreases faster than exponentially as $k \rightarrow \infty$.

Theorem 3.3.

$$r_k + r_{k+1} + \dots \leq \frac{E\varepsilon}{\alpha + 1 + b} \left(\frac{e}{z_k} \right)^k,$$

where $z_k = g^{-1}(ck) \rightarrow \infty$ as $k \rightarrow \infty$.

Proof. For every positive z we obviously have

$$X_t(k) + X_t(k+1) + \dots \leq z^{-k} G_t(z).$$

This is just the Markov inequality for the empirical distribution of the biological age. By Theorem 3.2, taking limits as $t \rightarrow \infty$ on both sides yields

$$\begin{aligned} r_k + r_{k+1} + \dots &\leq \frac{E\varepsilon}{cz^k} \int_0^1 u^{\frac{\alpha+b+1}{c}-1} \exp\left(\frac{1}{c} \int_u^1 \frac{g(vz)}{v} dv\right) du \leq \\ &\leq \frac{E\varepsilon}{cz^k} \int_0^1 u^{\frac{\alpha+b+1}{c}-1} \exp\left(\frac{1}{c} \int_0^1 \frac{g(vz)}{v} dv\right) du = \\ &= \frac{E\varepsilon}{\alpha + 1 + b} \exp\left(\frac{1}{c} \int_0^z \frac{g(v)}{v} dv - k \log z\right). \end{aligned}$$

Let us minimize the exponent. By differentiating and finding the zero of the derivative we get

$$\frac{g(z)}{cz} - \frac{k}{z} = 0,$$

that is, $z = z_k = g^{-1}(ck)$. Since $g(v)/v$ is an increasing function of z , it follows that

$$\frac{1}{c} \int_0^z \frac{g(v)}{v} dv \leq \frac{g(z)}{c} = k,$$

completing the proof. ■

Remark. If the offspring distribution is bounded: $\varepsilon \leq m$, then $g(z) \leq z^m$, hence $z_k \geq (ck)^{1/m}$, and we obtain the following upper bound.

$$r_k + r_{k+1} + \dots \leq \frac{E\varepsilon}{\alpha + 1 + b} \cdot \frac{e^k}{(ck)^{k/m}}.$$

Next, let us turn to the physical age.

It is well known [4, Corollary 6.4] that the proportion of edges at time t that are older than s converges almost everywhere on the event of non-extinction to

$$(3.2) \quad 1 - F(s) := \frac{\int_0^\infty e^{-\alpha t} [1 - L(t)] dt}{\int_0^\infty e^{-\alpha t} [1 - L(t)] dt},$$

as it can be seen by considering the fraction $Z^\phi(t)/Z(t)$ with random characteristic $\phi(t) = I(s < t < \lambda)$.

Theorem 3.4.

$$1 - F(s) \sim \frac{E\varepsilon}{\alpha + 1 + b} e^{-(\alpha+1+b)s+M} \quad \text{as } s \rightarrow \infty,$$

where

$$M = \frac{1}{c} \int_0^1 \frac{g(v)}{v} dv = \sum_{i=1}^{\infty} \frac{p_i}{ci}.$$

Proof. By (3.1) and Corollary 3.1,

$$\begin{aligned} 1 - F(s) &= E\varepsilon \int_s^\infty e^{-\alpha t} [1 - L(t)] dt = \\ &= E\varepsilon \int_s^\infty \exp\left(-(\alpha + 1 + b)t + \frac{1}{c} \int_{e^{-ct}}^1 \frac{g(v)}{v} dv\right) dt. \end{aligned}$$

Substitution $u = e^{-ct}$ yields

$$1 - F(s) = \frac{E\varepsilon}{c} \int_0^{e^{-cs}} u^{\frac{\alpha+1+b}{c}-1} \exp\left(\frac{1}{c} \int_u^1 \frac{g(v)}{v} dv\right) du.$$

As s tends to infinity, the domain of integration shrinks onto the origin, therefore the exponential term becomes more and more constant. It immediately follows that

$$(3.3) \quad 1 - G(s) \sim \frac{E\varepsilon}{\alpha + 1 + b} \exp\left(-(\alpha + 1 + b)s + M\right),$$

indeed. ■

Thus the tail of the physical age distribution is exponentially decreasing. Now, although the expected value of the biological age is $E\varepsilon t$, the asymptotic proportion of vertices with biological age s cannot be approximated by

$1 - F(s/E\varepsilon)$: the asymptotic behaviour of the two functions is completely different.

The much lighter tail of the stable biological age distribution is similar in flavour to the counter-intuitive fact found in [3]. There it was proved that

$$E(\xi(t) \mid \lambda > t) = \frac{1}{c}E(1 - e^{-ct\varepsilon}) = \frac{1}{c}E(1 - g(e^{-ct})) < \frac{1}{c}.$$

and the proof did not depend on the offspring distribution, thus the result is also valid in our generalized model. This is surprising, because the biological age $\xi(t)$ grows linearly before it is stopped at λ , hence one may think that if the physical age of an edge is large then so is its biological age, at least doesn't remain bounded in mean. Both discrepancies might be explained by the observation that reaching an extremely high physical age would point to the fact that the biological age has not increased as fast as usually.

References

- [1] **Haccou, P., P. Jagers and V.A. Vatunin**, *Branching Processes: Variation, Growth, and Extinction of Populations*, Cambridge University Press, 2005.
- [2] **Jagers, P.**, *Branching Processes with Biological Applications*, Wiley, London, 1975.
- [3] **Móri, T.F. and S. Rokob**, A random graph model driven by time-dependent branching dynamics. *Annales Univ. Sci. Budapest., Sect. Comp.*, **46** (2017), 191–213.
- [4] **Nerman, O.**, On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrscheinlichkeit.*, **57** (1981), 365–395.

T.F. Móri and S. Rokob

Department of Probability Theory and Statistics

Faculty of Science

ELTE Eötvös Loránd University

Budapest

Hungary

mori@math.elte.hu

rodnasbokor@gmail.com

