Elect. Comm. in Probab. **15** (2010), 475–488

ELECTRONIC COMMUNICATIONS in PROBABILITY

THE ALDOUS-SHIELDS MODEL REVISITED WITH APPLICATION TO CELLULAR AGEING

KATHARINA BEST
Albert-Ludwigs Universität, Abteilung Mathematische Stochastik, Eckerstr. 1, D–79104 Freiburg
email: katharina.best@stochastik.uni-freiburg.de

PETER PFAFFELHUBER
Albert-Ludwigs Universität, Abteilung Mathematische Stochastik, Eckerstr. 1, D–79104 Freiburg
email: p.p@stochastik.uni-freiburg.de

Submitted April 9, 2010, *accepted in final form* September 30, 2010

AMS 2000 Subject classification: 60K35, 92D20, 60J85, 05C05
Keywords: Random tree, cellular senescence, telomere, Hayflick limit

Abstract
In Aldous and Shields (1988) a model for a rooted, growing random binary tree with edge lengths 1 was presented. For some $c > 0$, an external vertex splits at rate c^{-i} (and becomes internal) if its distance from the root (depth) is i. We reanalyse the tree profile for $c > 1$, i.e. the numbers of external vertices in depth $i = 1, 2, \ldots$. Our main result are concrete formulas for the expectation and covariance-structure of the profile. In addition, we present the application of the model to cellular ageing. Here, we say that nodes in depth $h + 1$ are senescent, i.e. do not split. We obtain a limit result for the proportion of non-senesced vertices for large h.

1 Introduction

Trees arise in several applied sciences: In linguistics and biology, trees describe the relationship of items (languages, species) and in computer science, trees are used as data structures, e.g. for sorting. Randomizing the input leads to random trees, which are subject of a large body of research. For applications in biology, see e.g. [6, 14]. Here, important examples are trees arising from branching processes (e.g. Yule trees). In computer science, prominent examples are search trees; see e.g. [18, 10].

In this note, we are concerned with an application of random trees to cellular biology. In the 1960s it was known that eukaryotic cells have a limited replication capacity ([16]). The number of generations until cells do not proliferate any more is today known as the *Hayflick limit* and the phenomenon that cells lose their ability to proliferate is called *cellular senescence*. The molecular bases for cellular senescence were uncovered starting in the 1970s. During each round of replication, the *telomeres* (which are the end part of each chromosome) are shortened due to physical

constraints of the DNA copying mechanism [20]. In humans, these telomeres are a multiple (i.e. more than 1000-fold) repetition of the base pairs TTAGGG and up to 200 bases are lost in each replication round [17]. Most importantly, telomeres have a stabilizing effect on the DNA. The *DNA repair mechanism* of a cell must be able to distinguish between usual DNA breaks (which it is assumed to repair) and the telomeres (which it is assumed to ignore). Hence, when telomeres become shorter this stabilizing effect seizes and ageing occurs. In human cells, telomeres shrink from 15 kilobases at birth to less than 5 kilobases during a lifetime [24]. However, the enzyme *telomerase* can decrease the loss of telomeres during replication. This enzyme is active in stem cells and cancer cells, which both are cell types with an (almost) unbounded replication potential. The deeper understanding of the role of telomeres and telomerase is an active field of research because of the medical implications for ageing and cancer. In particular, it was awarded the Nobel prize in medicine in 2009 [25].

We study the model of random trees introduced in Aldous and Shields in [1] (hereafter referred to as [AS]) and extend it for an application to cellular ageing. Given some $c > 0$ and a full binary tree $\mathbb{T}$, the model introduced in [AS] describes the evolution of the vertices of the tree. Here, we distinguish *internal* and *external* vertices. At $t = 0$, the root is the only external vertex (and there are no internal vertices). An external vertex $u \in \mathbb{T}$ in depth $|u|$ becomes internal at rate $c^{-|u|}$. At the time it becomes internal, the two daughter vertices in depth $|u| + 1$ become external. We present our results on the Aldous-Shields model in Theorem 1.
For our application to cellular senescence, we will analyze a relative of the Aldous-Shields model for $c > 1$. Here, a critical depth h is fixed, and only external vertices in depth at most h can become internal. External vertices in depth $h + 1$ never become internal. Hereby, external vertices can be thought of as cells. The depth of a vertex is the number of generations from the first cell. Vertices in depth at most h represent proliferating cells, because they are able to produce offspring (i.e. daughter cells). Vertices in depth $h + 1$ represent senescent cells. This model has two features, which appear to be realistic in cellular senescence. First, the rate of cell proliferation decreases with the generation of a cell, parameterized by $c > 1$. Second, cells which have already split too often lose their ability to proliferate at all. For this model, we obtain a limit result for the frequency of proliferating cells in Theorem 2.

The paper is organized as follows: In Section 2, we state our results on the Aldous-Shields model. The application to cellular senescence is carried out in Section 3, where we also give an overview of other models for cellular senescence in the literature. Section 4 contains the proofs for our results on the Aldous-Shields model (Theorem 1), and in Section 5, we give proofs for the results on the model of cellular ageing (Theorem 2).

2 Model and results

We start by introducing some notation. Let $\mathbb{T}$ be the complete binary tree, given through

$$\mathbb{T} = \bigcup_{n=0}^{\infty} \mathbb{T}_n$$

and

$$\mathbb{T}_0 = \{\emptyset\}, \qquad \mathbb{T}_n = \{0,1\}^n \text{ for } n = 1, 2, \ldots$$

We refer to elements in $\mathbb{T}$ by *vertices* and identify $u \in \mathbb{T}_n$ by a word of length n over the alphabet $\{0,1\}$, whose ith letter is u_i, $n \geq 1$. The vertex $\emptyset$ is the root of the tree and vertex $u \in \mathbb{T}$ has two daughter vertices, $u0$ and $u1$. (We make the convention that $\emptyset 0 := 0, \emptyset 1 := 1$.) For $u \in \mathbb{T}$ we set $|u| = n$ iff $u \in \mathbb{T}_n$.
We say that u is an ancestor of v if $|u| < |v|$ and there are $i_1, ..., i_{|v|-|u|} \in \{0,1\}$ with $v = ui_1 \cdots i_{|v|-|u|}$. This induces a partial order on $\mathbb{T}$ if we write $u \prec v$ iff u is ancestor of v.

Definition 2.1 (Aldous-Shields model)**.** *Fix $c > 0$. The (time-continuous)* Aldous-Shields model *with parameter c is a a Markov jump process $\mathscr{Y} = (Y(t))_{t\geq 0}$, $Y(t) = (Y_u(t))_{u\in\mathbb{T}}$ with state space $\{0,1\}^{\mathbb{T}}$, starting in $Y(0) = (\mathbb{1}_{u=\emptyset})_{u\in\mathbb{T}}$. Given $Y(t) = y \in \{0,1\}^{\mathbb{T}}$ and $u \in \mathbb{T}$ with $y_u = 1$, it jumps to $(\widetilde{y}_v)_{v\in\mathbb{T}}$, given by*

$$\widetilde{y}_v = \begin{cases} 0, & v = u, \\ 1, & v = u0 \text{ or } v = u1, \\ y_v, & \text{else,} \end{cases}$$

at rate $c^{-|u|}$. In this case, we say that vertex u splits.

Remark 2.2 (Internal and external vertices)**.** Let $\mathscr{Y} = (Y(t))_{t\geq 0}$ be the Aldous-Shields model and $Y = Y(t)$ for some $t \geq 0$. It is important to note that the dynamics is such that any path $\emptyset, i_1, i_1 i_2, ... \in \mathbb{T}$ with $i_1, i_2, ... \in \{0,1\}$, starting at the root, has exactly one element u with $Y_u = 1$. In particular, the sets

$$\{u : \text{ there is } v \text{ with } u \prec v \text{ and } Y_v = 1\}$$

of internal vertices and

$$\{u : Y_u = 1\}$$

of external vertices are disjoint.

Definition 2.3 (Profile)**.** *Let $\mathscr{Y} = (Y(t))_{t\geq 0}$ be the Aldous-Shields model and $Y = Y(t)$ for some $t \geq 0$. We define*

$$X_n := \sum_{u\in\mathbb{T}_n} Y_u, \qquad \widehat{X}_n := 2^{-n} X_n, \tag{2.1}$$

the total number of external vertices *and the* relative proportion of external vertices *in depth n, respectively. The vector $(X_n)_{n=0,1,2,...}$ is also called the* profile *and*

$$X = \sum_{n=0}^{\infty} X_n$$

is the total number of external vertices.

Remark 2.4 (Dependence on c)**.** The behaviour of the Aldous-Shields model strongly depends on c: A larger c implies the profile being more concentrated around certain depths. The reason is that the difference in the splitting rate between external vertices at different depths is increasing in c. Therefore, external vertices in smaller depth can easily catch up with external vertices in larger depth for large c. See Figure 1 for an illustration.
Two values of c are of particular importance in applications from computer science: for $c = 1$, and if $X = n$, the set of external vertices is a binary search tree with n external vertices. For $c = 2$ and $X = n$, the set of external vertices is a digital search tree with n external vertices; see e.g. [10].

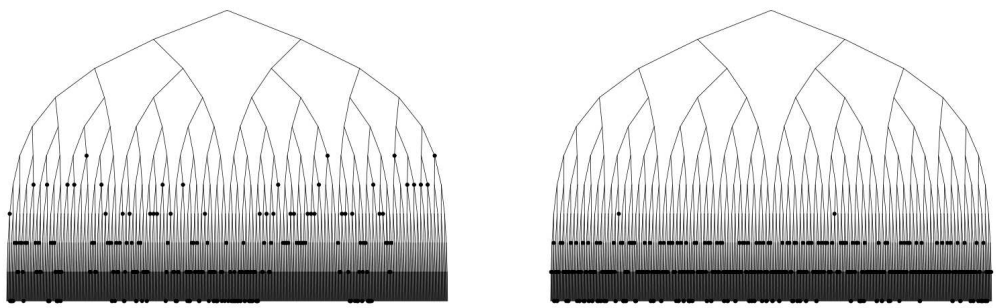

Figure 1: Two realizations of the Aldous-Shields model at the time when 500 vertices are external for $c = 1.05$ (A) and $c = 3$ (B). Only depths $0, \ldots, 10$ are drawn and the external vertices are marked.

Remark 2.5 (Relative frequencies)**.** We observe that

$$\sum_{n=0}^{\infty} \widehat{X}_n(t) = 1 \tag{2.2}$$

for all t, almost surely. To see this, note that $\widehat{X}_0(0) = 1$ and $\widehat{X}_n(0) = 0$ for $n > 0$, i.e. (2.2) holds at $t = 0$. Additionally, assume that some u with $|u| = n$ splits at time t. Then, we have that $\widehat{X}_n(t) - \widehat{X}_n(t-) = -2^{-n}$ and $\widehat{X}_{n+1}(t) - \widehat{X}_{n+1}(t-) = 2 \cdot 2^{-(n+1)} = 2^{-n}$. In particular, every split leaves $\sum_{n=0}^{\infty} \widehat{X}_n$ unchanged which shows (2.2).

Remark 2.6 (Notation)**.** In our results, we will give asymptotics of moments of $X_{n+i}(tc^n)$ for large n. Generally, for two sequences $(x_n)_{n=1,2,\ldots}$ and $(y_n)_{n=1,2,\ldots}$, which may depend on other parameters, we write

$$x_n \overset{n\to\infty}{\approx} y_n \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \frac{x_n}{y_n} = 1.$$

In our exposition, several products appear. Throughout, we take the convention that an empty product equals 1.

Theorem 1 (Moments of the profile and their limits)**.** *Let $c > 1$. Define for $k \in \mathbb{Z}_+$*

$$a_k = (-1)^k \frac{c^k}{(c-1)\cdots(c^k-1)}, \qquad b_k := \frac{c^1 \cdots c^k}{(c-1)\cdots(c^k-1)} \tag{2.3}$$

and $a_0 = b_0 = 1$. For $k < 0$, we set $a_k = b_k = 0$. Moreover,

$$b_\infty := \Big(\prod_{l=1}^{\infty}(1 - c^{-l})\Big)^{-1} \in (1; \infty).$$

Then, for $t > 0$ and $i > -n$,

$$\mathbb{E}[X_{n+i}(tc^n)] = 2^{n+i} \cdot \sum_{k=0}^{n} a_k b_{n-k} e^{-c^{-i+k}t} \overset{n\to\infty}{\approx} b_\infty \cdot 2^{n+i} \cdot \sum_{k=0}^{\infty} a_k e^{-c^{-i+k}t}. \tag{2.4}$$

Moreover, for $i, i' \in \mathbb{Z}$, $i \le i'$,

$$\mathbb{COV}[X_{n+i}(tc^n), X_{n+i'}(tc^n)] \overset{n\to\infty}{\approx} a_{i,i'} \cdot \begin{cases} \dfrac{2}{2-c^2}\left(\dfrac{2}{c}\right)^{2n+i+i'}, & c < \sqrt{2} \\ 2^n n \sqrt{2}^{i+i'}, & c = \sqrt{2}, \\ \dfrac{c^4}{2(c^2-1)(c^2-2)} 2^{n+i'} c^{i-i'}, & c > \sqrt{2}. \end{cases} \tag{2.5}$$

where

$$a_{i,i'} := b_\infty^2 \sum_{k,k'=0}^{\infty} a_k a_{k'} e^{-c^{-i+k}-c^{-i'+k'}t} c^{k+k'}. \tag{2.6}$$

Remark 2.7 (Convergence and covariance)**.** 1. It is immediate from the Theorem and the definition of $\widehat{X}_n$ that $\lim_{n\to\infty} \mathbb{V}[\widehat{X}_{n+i}(tc^n)] = 0$ and hence

$$\lim_{n\to\infty} \widehat{X}_{n+i}(tc^n) = b_\infty \cdot \sum_{k=0}^{\infty} a_k e^{-c^{-i+k}t} \tag{2.7}$$

in probability, for all $t > 0$ and all $c > 1$.

2. The covariances given in (2.5) show a phase transition at $c = \sqrt{2}$. Such a phase transition is already known from results by [AS] and [9]. However, these papers do not give explicit formulas for the covariance structure.

3. Using (4.4), (4.6) and (4.10) it is also possible to obtain exact results for the covariance on the left hand side of (2.5).

Remark 2.8 (Connection to results from Aldous and Shields (1988))**.** In [AS], the evolution of the vector $(\widehat{X}_n(t))_{n=0,1,2,\ldots}$ is studied. In their Theorems, a law of large numbers for $X_n(t)$ to some deterministic limit $x_n(t)$ is stated and proved using martingale methods. Their result implies that (2.7) holds even almost surely on compact time intervals. In particular, they claim that the limit $x_i(t)$ of $\widehat{X}_{n+i}(tc^n)$ must satisfy $x_i(t) = x_{i+1}(ct)$, which clearly holds for the right hand side of (2.7). In addition, they show that a suitably rescaled process, $2^{n/2}(\widehat{X}_{n+i}(t) - x_i(t))_{t\ge 0}$, converges weakly to a diffusion for $c > \sqrt{2}$. The rescaling factor $2^{n/2}$ can also be recognized in the above Theorem. Moreover, (2.5) shows that a convergence of $c^n(\widehat{X}_{n+i}(t) - x_i(t))_{t\ge 0}$ to a diffusion can be conjectured for $1 < c < \sqrt{2}$.

Remark 2.9 (Comparison to work of Dean and Majumdar (2006))**.** In [9], the total number X of external vertices was studied in the context of the Aldous-Shields model on an m-ary tree. In the binary case, a functional equation (their equation (2)) for the Laplace transform of $X(t)$ was shown to hold true. This equation uses the following fact: Given that T is the random time of the first split in the model, it is clear that T is mean one exponential and, in addition,

$$X(t) \overset{d}{=} \mathbb{1}_{T\ge t} + \mathbb{1}_{T\le t}\left(X'\left(\tfrac{t-T}{c}\right) + X''\left(\tfrac{t-T}{c}\right)\right)$$

where X' and X'' are independent of T and of each other and distributed like X. From their identity on Laplace transforms, [9] show the phase transition for the variance of the number of occupied vertices at $c = \sqrt{2}$, which is also seen from Theorem 1.

3 Application: cellular ageing

The first mathematical model for cellular senescence was given in [17]. It takes several biological facts into account. When DNA is copied, the double helix is unfolded and both strands of DNA are copied. Only in one of the two strands there are physical constraints by which the end of a chromosome cannot be perfectly copied. This shortening of telomeres is independent for all chromosomes. In [17], it is assumed that (i) the telomeres are of fixed length, (ii) they decrease by a fixed length at each proliferation event for one of the daughter cells and (iii) proliferation occurs along a full binary tree. If the telomere length of a single chromosome falls below a threshold, a cell cannot replicate any more and becomes senescent. This threshold takes the Hayflick limit into account, which states that a cell line can only live for a limited number of generations before it becomes senescent.
The model in [17] was extended in several directions. A stochastic amount of loss of telomeres was studied in [2]. In [3] and [19], the binary tree of proliferating cells considered in [17] was replaced by the tree of a branching process. In particular, [19] included cell death into their model, with different death rates above and below a critical threshold of telomere length. Age structure of cells (i.e. structure which phase of the cell cycle the cell is in) is considered in [11, 12]. Moreover, [4] extends the model of [17] by making telomerase activity (which is present in stem cells and cancer cells) explicit.

The idea to use the Aldous-Shields model for cellular ageing was influenced by the following recent results:

1. In [7], a model is proposed which distinguishes two states of telomeres: *capped* and *uncapped*. Only in the capped state, proliferation of the cell is still possible. In somatic cells, an uncapped telomere cannot be transformed to the capped state any more leading to senesced cells; see [22]. In stem and tumor cells, telomerase is (among other things) responsible for transitions from uncapped back to capped telomeres. Following [23], the transition rate of the uncapped to the capped state in stem cell decreases with shorter telomeres.

2. In data, it has been observed that proliferating cells can behave differently. Motivated by data from [5, 8, 15], it is argued in [21] that the rate of proliferation decreases for shorter telomeres. Their model produces a Gompertzian growth model which is known to fit to empirical data for somatic and tumor cells.

In stem cells, the decreasing rate for an uncapped telomere to re-enter the capped state for shorter telomeres follows the behaviour of the Aldous-Shields model as shown in [23] remarkably well: cells with a long replicative history proliferate slower. While [21] uses a linear decrease in replication rate, depending on the telomere length, the Aldous-Shields model uses a geometric decay of the proliferation rate.
Note that short telomeres can be seen as a form of *damage*. In [13], models for cellular damage were introduced. In their model, cells pass on damage to the daughter cells. This model, as well as the Aldous-Shields model are among the analytically tractable ones.

We state our model of cellular ageing:

Definition 3.1. *Fix $h \in \mathbb{N}$, $r > 0$ and $c > 1$ and let $\mathbb{T}^h := \bigcup_{n=0}^{h+1} \mathbb{T}_n$. The process $\mathscr{Z} = (Z(t))_{t \geq 0}$, where $Z(t) = (Z_u(t))_{u \in \mathbb{T}^h}$ is a Markov jump process with state space $\{0,1\}^{\mathbb{T}^h}$, starting in $Z(0) = (\mathbb{1}_{u=\emptyset})_{u \in \mathbb{T}^h}$.*

Given $Z(t) = z \in \{0,1\}^{\mathbb{T}^h}$ and $u \in \mathbb{T}^h \setminus \mathbb{T}_{h+1}$ with $z_u = 1$, it jumps to $(\widetilde{z}_v)_{v\in\mathbb{T}}$, given by

$$\widetilde{z}_v = \begin{cases} 0, & v = u, \\ 1, & v = u0 \text{ or } v = u1, \\ z_v, & \text{else}, \end{cases}$$

at rate $rc^{-|u|}$. (Note that vertices $u \in \mathbb{T}_{h+1}$ do not split.)

Informally, every external vertex u in this process represents a cell. If $|u| = n$, we say that the cell is in generation n. The process starts with a single mother cell. It proliferates at rate r. All cells up to generation h from the mother cell follow the usual dynamics of the Aldous-Shields model (with time rescaled by a factor of r), such that cells in generation n proliferate at rate $r \cdot c^{-n}$. If a cell is in generation $h+1$ from the mother cell, its telomeres have reached the Hayflick limit and the cell is not able to proliferate any more.

Definition 3.2 (Relative frequency of proliferating cells)**.** *In applications, the relative frequency of proliferating cells,*

$$L(t) := \frac{Z^p(t)}{Z^p(t) + Z^s(t)} \tag{3.1}$$

with

$$Z^p(t) := \sum_{u\in\mathbb{T}^h\setminus\mathbb{T}_{h+1}} Z_u(t), \qquad Z^s(t) := \sum_{u\in\mathbb{T}_{h+1}} Z_u(t),$$

is of particular importance. Here, $Z^p(t)$ and $Z^s(t)$ is the number of proliferating and senescent cells at time t, respectively.

Theorem 2 (Frequency of replicating cells)**.** *For $\mathscr{Z}$ and L as in Definition 3.1 and 3.2,*

$$\lim_{h\to\infty} L(tc^h) = \frac{\sum_{i=0}^{\infty}\sum_{k=0}^{\infty} 2^{-i} a_k e^{-c^{i+k}t/r}}{\sum_{i=0}^{\infty}\sum_{k=0}^{\infty} a_k (2^{-i} e^{-c^{i+k}t/r} + 2e^{-c^{-i-1+k}t/r})} \tag{3.2}$$

in probability, for all $t > 0$.

Remark 3.3 (Simulations)**.** In our model for cellular senescence, Theorem 2 describes the decrease in the frequency of proliferating cells. This frequency has been measured empirically; see e.g. [4, Figure 5] and [3, Figure 2]. As can be seen from the Theorem, every c gives a specific curve of decrease; see also Figure 2. These curves can be fit to data in order to estimate c. As the figure shows, the limiting result of Theorem 2 already gives a good fit for simulations which use $h = 20$.

4 Proof of Theorem 1

4.1 Preliminaries

The key ingredient in the proof is the quantity

$$y_n(t) := \mathbb{E}[Y_{0_n}(t)] \text{ with } 0_n := \underbrace{0\cdots0}_{n \text{ times}}. \tag{4.1}$$

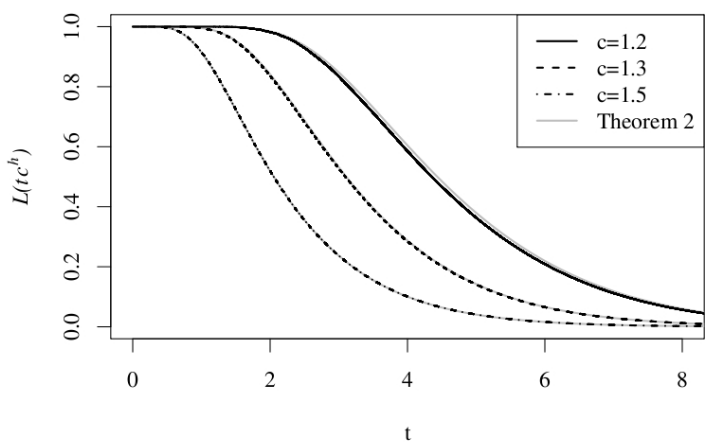

Figure 2: The decrease of the frequency of proliferating cells strongly depends on c. The figure shows simulations for $c = 1.2, 1.3, 1.5$ and $h = 20$. The grey lines show the limit result (3.2) from Theorem 2 with $r = 1$.

Note that the dynamics of $(Y_{0_n})_{n=0,1,2,\dots}$ (taking values in $\{0,1\}^{\{0,1,2,\dots\}}$ with only one entry being 1) is autonomous. It is given by the following rule: if $Y_{0_n}(t) = 1$, then, at rate c^{-n} a transition occurs to the configuration $Y_{0_n}(t) = 0, Y_{0_{n+1}}(t) = 1$. From the dynamics of the Aldous-Shields model, it is clear that $\underline{y}(t) = (y_n(t))_{n=0,1,2,\dots}$ follows the differential equations

$$\dot{y}_n = c^{-(n-1)} y_{n-1} - c^{-n} y_n, \qquad n \geq 0 \tag{4.2}$$

with $y_{-1} = 0$; compare (2.1) in [AS]. We rewrite the equation to obtain

$$\dot{\underline{y}} = C\underline{y}$$

with

$$C = \begin{pmatrix} -1 & & & & & & \\ 1 & -c^{-1} & & & & & \\ & c^{-1} & -c^{-2} & & & & \\ & & c^{-2} & -c^{-3} & & & \\ & & & c^{-3} & -c^{-4} & & \\ & & & & c^{-4} & -c^{-5} & \\ & & & & & \cdots & \cdots \end{pmatrix}$$

Our first Lemma provides essential facts about the matrix C. Recall a_k and b_k from (2.3).

Lemma 4.1. *Let $A = (a_{ij})_{i,j=0,1,2,\dots}$ and $B = (b_{ij})_{i,j=0,1,2,\dots}$ be given by*

$$a_{ij} := a_{i-j}, \qquad b_{ij} := b_{i-j}.$$

The matrix A contains the eigenvectors of C to the eigenvalues $\lambda_i = -c^{-i}, i = 0, 1, 2, \dots$ and A and B are inverse to each other.

Proof. To see that A contains the eigenvectors of C, note that for $i \geq j$

$$(CA)_{ij} = c^{-i+1}a_{i-1-j} - c^{-i}a_{i-j} = (c^{-i+1}\frac{1-c^{i-j}}{c} - c^{-i})\,a_{i-j} = -c^{-j}a_{i-j}.$$

To see that A and B are inverse to each other, it follows from the definition of a_k and b_k that

$$(BA)_{ii} = a_0 \cdot b_0 = 1 \qquad \text{and} \qquad (BA)_{ij} = 0 \text{ for } i < j.$$

For $i > j$, we set $n := i - j > 0$ and obtain

$$\begin{aligned}
(BA)_{ij} &= \sum_{k=j}^{i} b_{i-k}a_{k-j} = \sum_{k=j}^{n+j} b_{j+n-k}a_{k-j} = \sum_{k=0}^{n} b_{n-k}a_k \\
&= \sum_{k=0}^{n}(-1)^k \frac{c\cdots c^{n-k}\cdot c^k}{(c-1)\cdots(c^{n-k}-1)\cdot(c-1)\cdots(c^k-1)} \\
&= \frac{c^n}{(c-1)\cdots(c^n-1)}\sum_{k=0}^{n}(-1)^k\frac{c^{1+\cdots+(n-k-1)}}{(c-1)\cdots(c^{n-k}-1)}(c^{k+1}-1)\cdots(c^n-1) \\
&= \frac{(-c)^n}{(c-1)\cdots(c^n-1)}\sum_{k=0}^{n}(-1)^k\frac{c^{\binom{k}{2}}}{(c-1)\cdots(c^k-1)}(c^{n-k+1}-1)\cdots(c^n-1),
\end{aligned}$$

where we have reversed the order of the summands in the last equality. We rewrite the product $(c^l-1)\cdots(c^{l'}-1) = \prod_{j=l}^{l'}(c^j-1) = 1$ for $l > l'$ and claim that for all $n > 0$

$$\sum_{k=0}^{n}(-1)^k\frac{c^{\binom{k}{2}}}{(c-1)\cdots(c^k-1)}(c^{n-k+1}-1)\cdots(c^n-1) = 0 \tag{4.3}$$

which implies that A and B are inverse to each other. We use induction and note that the assertion is clear for $n = 1$. Given it is true for n, we have

$$\begin{aligned}
&\sum_{k=0}^{n+1}(-1)^k\frac{c^{\binom{k}{2}}}{(c-1)\cdots(c^k-1)}(c^{n-k+2}-1)\cdots(c^{n+1}-1) \\
&\quad = 1 + \sum_{k=1}^{n+1}(-1)^k\frac{c^{\binom{k}{2}}}{(c-1)\cdots(c^k-1)}(c^{n-k+2}-1)\cdots(c^n-1)[(c^{n+1}-c^{n+1-k})+(c^{n-k+1}-1)] \\
&\quad = \sum_{k=1}^{n+1}(-1)^k\frac{c^{\binom{k}{2}}}{(c-1)\cdots(c^k-1)}c^{n+1-k}(c^k-1)(c^{n-k+2}-1)\cdots(c^n-1) \\
&\quad = -\sum_{k=1}^{n+1}(-1)^{k-1}\frac{c^{\binom{k-1}{2}}c^k}{(c-1)\cdots(c^{k-1}-1)}c^{n+1}c^{-k}(c^{n-k+2}-1)\cdots(c^n-1) \\
&\quad = -c^{n+1}\sum_{k=1}^{n+1}(-1)^{k-1}\frac{c^{\binom{k-1}{2}}}{(c-1)\cdots(c^{k-1}-1)}(c^{n-(k-1)+1}-1)\cdots(c^n-1) = 0
\end{aligned}$$

where we have used the induction hypothesis in the second and in the last equality. Hence, we have shown (4.3) and the proof is complete. □

4.2 First order structure; proof of (2.4)

By linearity, we can now explicitly solve (4.2) using Lemma 4.1. Since A contains the eigenvalues of C and A is inverse to B, we immediately write, using $\underline{y}(t) := (y_0(t), y_1(t), y_2(t), \dots)$ and $D := \mathrm{diag}(-1, -c^{-1}, -c^{-2}, \dots)$

$$\begin{aligned}\big(2^{-n}\mathbb{E}[X_n(t)]\big)_{n=0,1,2,\dots} &= y(t) = e^{Ct}\underline{y}(0)^\dagger = Ae^{Dt}B\underline{y}(0)^\dagger = Ae^{Dt}(b_0, b_1, \dots)^\dagger \\ &= \Big(\sum_{k=0}^{n} a_{n-k} b_k e^{-c^{-k}t}\Big)_{n=0,1,2,\dots}\end{aligned}$$

because $z(0) = (1, 0, 0, \dots)$ as the process starts with $(Y(0)) = (\mathbb{1}_{u=\emptyset})_{u\in\mathbb{T}}$. We have shown the first part of (2.4) and in order to prove the second part, fix i, k, t and note that $n \mapsto |a_k| b_{n+i-k} e^{-c^{-i+k}t}$ is increasing with a summable limit. Hence, by dominated convergence,

$$y_{n+i}(tc^n) = \sum_{k=0}^{\infty} a_k b_{n+i-k} e^{-c^{-n-i+k}tc^n} \overset{n\to\infty}{\approx} b_\infty \sum_{k=0}^{\infty} a_k e^{-c^{-i+k}t}.$$

4.3 Second order structure; proof of (2.5)

Now we come to the second order structure. Similar to the definition of y_n in (4.1), we set for $n \le n'$

$$y_{n,n',m}(t) := \mathbb{COV}[Y_{0_n}(t), Y_{0_m 1_{n'-m}}(t)] \text{ with } 0_m 1_{n'-m} := \underbrace{0\cdots0}_{m \text{ times}}\ \underbrace{1\cdots1}_{n'-m \text{ times}}.$$

In order to see the connection of $y_{n,n',m}(t)$ and $\mathbb{COV}[X_n(t), X_{n'}(t)]$, we define the depth of the most recent common ancestor of u, u' as

$$M_{u,u'} := \sup\{|v| : v \preceq u, v \preceq u'\}, \qquad u, u' \in \mathbb{T},$$

where $v \preceq u$ if $v \prec u$ or $v = u$. Let U, U' be two random variables, where U is uniformly distributed on $\mathbb{T}_n$ and U' uniformly distributed on $\mathbb{T}_{n'}$, independent of all the rest. The distribution of $M_{U,U'}$ is given by (recall $n \le n'$)

$$\mathbb{P}[M_{U,U'} = m] = 2^{-((m+1)\wedge n)}, \qquad m = 0, \dots, n.$$

We write

$$\begin{aligned}\mathbb{COV}[X_n(t), X_{n'}(t)] &= \sum_{u\in\mathbb{T}_n}\sum_{u'\in\mathbb{T}_{n'}} \big(\mathbb{E}[Y_u(t)Y_{u'}(t)] - \mathbb{E}[Y_u(t)]\cdot\mathbb{E}[Y_{u'}(t)]\big) \\ &= 2^{n+n'}\mathbb{COV}[Y_U(t), Y_{U'}(t)] \\ &= 2^{n+n'}\big(\mathbb{E}\big[\mathbb{COV}[Y_U(t), Y_{U'}(t)|M_{U,U'}]\big] + \mathbb{COV}\big[\mathbb{E}[Y_U|M_{U,U'}], \mathbb{E}[Y_{U'}|M_{U,U'}]\big]\big) \qquad (4.4) \\ &= 2^{n+n'}\mathbb{E}\big[\mathbb{COV}[Y_U(t), Y_{U'}(t)|M_{U,U'}]\big] \\ &= 2^{n+n'}\sum_{m=0}^{n} 2^{-((m+1)\wedge n)} y_{n,n',m}(t).\end{aligned}$$

The second to last equality holds as $\mathbb{E}[Y_U|M_{U,U'}] = \mathbb{E}[Y_U]$ and $\mathbb{E}[Y_{U'}|M_{U,U'}] = \mathbb{E}[Y_{U'}]$. In order to use the last expression, note that for $y_{n,n',m}$ the vertices 0_n and $0_m 1_{n'-m}$ have the vertex 0_m as

their most recent common ancestor and so $M_{0_n,0_m1_{n'-m}} = m$. We let T_m be the last time when Y_{0_m} is external, equivalently the time when $Y_{0_{m+1}}$ becomes external, i.e. T_m is the sum of exponentials with parameters $1, c^{-1}, c^{-2}, ..., c^{-m}$. Hence,

$$\mathbb{E}[e^{-\lambda T_m}] = \prod_{l=1}^{m} \frac{c^{-l}}{c^{-l}+\lambda} = \prod_{l=1}^{m} \frac{1}{1+\lambda c^l}. \tag{4.5}$$

Using the last equations we now prove (2.5) in three steps. First, we give a representation of $y_{n,n',m}$ in terms of a functional of T_m. Second, we derive the asymptotics of $y_{n+i,n+i',m}(tc^n)$ for large n using this representation. Last, we plug this asymptotics into (4.4).

Step 1 (Exact representation of $y_{n,n',m}(t)$)**.** In this step we show that for $m < n \le n'$

$$\begin{aligned} y_{n,n',m}(t) = \sum_{k=0}^{n-m}\sum_{k'=0}^{n'-m} & a_k a_{k'} b_{n-m-k} b_{n'-m-k'} \\ & \cdot \mathbb{COV}[e^{-c^{-n+k}(t-T_m)}\mathbb{1}_{t\ge T_m}, e^{-c^{-n'+k'}(t-T_m)}\mathbb{1}_{t\ge T_m}] \end{aligned} \tag{4.6}$$

and for $m = n \le n'$

$$y_{n,n',m}(t) = \delta_{n,n'} y_n(t) - y_n(t) y_{n'}(t), \tag{4.7}$$

where $\delta_{n,n'}$ is Kronecker's δ.

Proof. For (4.6), observe that $Y_{0_n}(t)$ and $Y_{0_m1_{n'-m}}(t)$ are independent, given T_m. Moreover, it is clear that $\mathbb{E}[Y_{0_n}(t)|T_m] = y_{n-m}(c^{-m}(t-T_m))\mathbb{1}_{t\ge T_m}$. Hence, by (2.4),

$$\begin{aligned} &\mathbb{COV}[Y_{0_n}(t), Y_{0_m1_{n'-m}}(t)] \\ &= \mathbb{E}[\mathbb{COV}[Y_{0_n}(t), Y_{0_m1_{n'-m}}(t)|T_m]] \\ &\qquad\qquad + \mathbb{COV}[\mathbb{E}[Y_{0_n}(t)|T_m], \mathbb{E}[Y_{0_m1_{n'-m}}(t)|T_m]] \\ &= \mathbb{COV}[y_{n-m}(c^{-m}(t-T_m))\mathbb{1}_{t\ge T_m}, y_{n'-m}(c^{-m}(t-T_m))\mathbb{1}_{t\ge T_m}] \\ &= \sum_{k=0}^{n-m}\sum_{k'=0}^{n'-m} a_k a_{k'} b_{n-m-k} b_{n'-m-k'} \mathbb{COV}[e^{-c^{-n+k}(t-T_m)}\mathbb{1}_{t\ge T_m}, e^{-c^{-n'+k'}(t-T_m)}\mathbb{1}_{t\ge T_m}]. \end{aligned}$$

For (4.7), note that $m = n$ implies that $Y_{0_n}(t)Y_{0_m1_{n'-m}}(t) = \delta_{n,n'} Y_{0_n}(t)$ and the result follows. □

Step 2 (Asymptotics of $y_{n+i,n+i',m}(tc^n)$ for large n)**.** The aim of this step is to establish that for $i, i' \in \mathbb{Z}$ and $m \in \mathbb{Z}_+$,

$$y_{n+i,n+i',m}(tc^n) \overset{n\to\infty}{\approx} b_\infty^2 \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} a_k a_{k'} e^{-t(c^{-i+k}+c^{-i'+k'})} c^{-2n-i-i'+k+k'} \frac{c^{2(m+1)}-1}{c^2-1}. \tag{4.8}$$

Proof. For large n, we have $tc^n > T_m$ and so, by (4.6), (4.7) and dominated convergence,

$$\begin{aligned} y_{n+i,n+i',m}(tc^n) &\overset{n\to\infty}{\approx} b_\infty^2 \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} a_k a_{k'} \mathbb{COV}[e^{-c^{-n-i+k}(tc^n-T_m)}, e^{-c^{-n-i'+k'}(tc^n-T_m)}] \\ &= b_\infty^2 \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} a_k a_{k'} e^{-t(c^{-i+k}+c^{-i'+k'})} \mathbb{COV}[e^{c^{-n-i+k}T_m}, e^{c^{-n-i'+k'}T_m}]. \end{aligned} \tag{4.9}$$

Now, for all $j, j' \in \mathbb{Z}$, using (4.5), for n large enough,

$$\begin{aligned}
&\mathbb{COV}[e^{c^{-n-j}T_m}, e^{c^{-n-j'}T_m}] \\
&\quad = \prod_{l=0}^{m} \frac{1}{1 - c^{-n-j+l} - c^{-n-j'+l}} - \prod_{l=0}^{m} \frac{1}{1 - c^{-n-j+l} - c^{-n-j'+l} + c^{-2n-j-j'+2l}} \\
&\quad \overset{n\to\infty}{\approx} \prod_{l=0}^{m} (1 + c^{-n-j+l} + c^{-n-j'+l} + c^{-2n-2j+2l} + 2c^{-2n-j-j'+2l} + c^{-2n-2j'+2l}) \\
&\qquad - \prod_{l=0}^{m} (1 + c^{-n-j+l} + c^{-n-j'+l} + c^{-2n-2j+2l} + c^{-2n-j-j'+2l} + c^{-2n-2j'+2l}) \\
&\quad \overset{n\to\infty}{\approx} \sum_{l=0}^{m} c^{-2n-j-j'+2l} \\
&\quad = c^{-2n-j-j'} \frac{c^{2(m+1)} - 1}{c^2 - 1}.
\end{aligned} \tag{4.10}$$

Plugging the last expression into (4.9) gives (4.8). □

Step 3 (Combining (4.8) and (4.4))**.** We write immediately, using $j = i - k$ and $j' = i' - k'$,

$$\begin{aligned}
&\mathbb{COV}[X_{n+i}(tc^n), X_{n+i'}(tc^n)] \\
&\quad \overset{n\to\infty}{\approx} b_\infty^2 \left(\frac{2}{c}\right)^{2n+i+i'} \sum_{m=1}^{n+i+1} \sum_{k,k'=0}^{\infty} a_k a_{k'} e^{-t(c^{-i+k}+c^{-i'+k'})} c^{k+k'} 2^{-m} \frac{c^{2m} - 1}{c^2 - 1}.
\end{aligned} \tag{4.11}$$

Noting that $\sum_{k=0}^{\infty} |a_k| c^k < \infty$, we see that for $a_{i,i'}(t)$ given by (2.6)

$$\begin{aligned}
\mathbb{COV}[X_{n+i}(tc^n), X_{n+i'}(tc^n)] &\overset{n\to\infty}{\approx} a_{i,i'}(t) \left(\frac{2}{c}\right)^{2n+i+i'} \sum_{m=1}^{n+i+1} \frac{2^{-m}(c^{2m} - 1)}{c^2 - 1} \\
&\overset{n\to\infty}{\approx} a_{i,i'}(t) \cdot \begin{cases} \dfrac{2}{2 - c^2} \left(\dfrac{2}{c}\right)^{2n+i+i'}, & c < \sqrt{2} \\ 2^n n \sqrt{2}^{i+i'}, & c = \sqrt{2}, \\ \dfrac{c^4}{2(c^2 - 1)(c^2 - 2)} 2^{n+i'} c^{i-i'}, & c > \sqrt{2}. \end{cases}
\end{aligned}$$

This finally shows (2.5) and finishes the proof of Theorem 1.

□

5 Proof of Theorem 2

The parameter r is only a rescaling of time. Hence, we can safely assume $r = 1$ in our proof. Let $\mathscr{Y} = (Y(t))_{t \geq 0}$ be the Aldous-Shields model with parameter c and $X_n(t)$ as in (2.1). Defining

$$\widetilde{X}_{h+1}(t) := \sum_{n=h+1}^{\infty} 2^{h+1-n} X_n(t),$$

it is important to note that

$$\widetilde{X}_{h+1}(t) = \#\{u \in \mathbb{T}_{h+1} : \exists v : u \preceq v, Y_v(t) = 1\},$$

almost surely; see also Remark 2.5. This implies that we can couple $\mathscr{Y}$ and $\mathscr{Z}$ in the sense that

$$(X_1(t), ..., X_h(t), \widetilde{X}_{h+1}(t))_{t\geq 0} \stackrel{d}{=} \Big(\sum_{u\in\mathbb{T}_1} Z_u(t), ..., \sum_{u\in\mathbb{T}_{h+1}} Z_u(t) \Big)_{t\geq 0}.$$

By Theorem 1, for $n \leq h$ and $i = 0, 1, 2, ...$

$$Z^p(tc^h) \stackrel{d}{=} \sum_{i=0}^{h} X_i(tc^h) = \sum_{i=0}^{h} X_{h-i}(tc^h) \stackrel{h\to\infty}{\approx} b_\infty \cdot 2^h \sum_{i=0}^{\infty}\sum_{k=0}^{\infty} 2^{-i} a_k e^{-c^{i+k}t},$$

$$Z^s(tc^h) \stackrel{d}{=} \widetilde{X}_{h+1}(tc^h) = \sum_{i=1}^{\infty} 2^{-i+1} X_{h+i}(tc^h) \stackrel{h\to\infty}{\approx} b_\infty \cdot 2^h \sum_{i=1}^{\infty}\sum_{k=0}^{\infty} 2 a_k e^{-c^{-i+k}t}$$

in probability, for all $t > 0$. Using the last two limits in the definition of L in (3.1) gives the result. □

Acknowledgements

This research was supported by the BMBF through FRISYS (Kennzeichen 0313921).

References

[1] D. Aldous and P. Shields. A diffusion limit for a class of randomly-growing binary trees. *Prob. Theo. Rel. Fields*, 79(4):509–542, 1988. MR0966174

[2] T. Antal, K. B. Blagoev, S. A. Trugman, and S. Redner. Aging and immortality in a cell proliferation model. *J. Theor. Biol.*, 248(3):411–417, 2007.

[3] O. Arino, M. Kimmel, and G. F. Webb. Mathematical modeling of the loss of telomere sequences. *J. Theor. Biol.*, 177(1):45–57, 1995.

[4] N. Arkus. A mathematical model of cellular apoptosis and senescence through the dynamics of telomere loss. *J. Theor. Biol.*, 235(1):13–32, 2005. MR2139011

[5] Melissa A. Baxter, R. F. Wynn, S. N. Jowitt, J. Ed Wraith, L. J. Fairbairn, and I. Bellantuono. Study of telomere length reveals rapid aging of human marrow stromal cells following in vitro expansion. *Stem Cells*, 22(5):675–682, 2004.

[6] N. Berestycki. Recent progress in coalescent theory. *Ensaios Matematicos*, 16:1–193, 2009. MR2574323

[7] E. H. Blackburn. Telomere states and cell fates. *Nature*, 408:53–56, 2000.

[8] M. Bonab, K. Alimoghaddam, F. Talebian, S. Ghaffari, A. Ghavamzadeh, and B. Nikbin. Aging of mesenchymal stem cell in vitro. *BMC Cell Biology*, 7(1):14, 2006.

[9] D. S. Dean and S. N. Majumdar. Phase Transition in a Generalized Eden Growth Model on a Tree. *J. Stat. Phys.*, 124(6):1351–1376, 2006. MR2266447

[10] M. Drmota. *Random Trees*. Springer, 2009. MR2484382

[11] J. Dyson, E. Sánchez, R. Villella-Bressan, and G. F. Webb. Stabilization of telomeres in nonlinear models of proliferating cell lines. *J. Theor. Biol.*, 244(3):400–408, 2007. MR2293122

[12] J. Dyson, R. Villella-Bressan, and G. F. Webb. Asynchronous exponential growth in an age structured population of proliferating and quiescent cells. *Math. Biosci.*, 177-178:73–83, 2002. MR1923804

[13] S. N. Evans and D. Steinsaltz. Damage segregation at fissioning may increase growth rates: A superprocess model. *Theo. Pop. Biol.*, 71(4):473–490, 2007.

[14] J. Felsenstein. *Inferring Phylogenies*. Palgrave Macmillan, 2002.

[15] N. Gupta, R. Taneja, A. Pandey, M. Mukesh, H. Singh, and S. C. Gupta. Replicative senescence, telomere shortening and cell proliferation rate in gaddi goat's skin fibroblast cell line. *Cell Biol. Int.*, 31:1257–1264, 2007.

[16] L. Hayflick and P. S. Moorhead. The serial cultivation of human diploid cell strains. *Exp. Cell Res.*, 25:585–621, 1961.

[17] M. Z. Levy, R. C. Allsopp, A. B. Futcher, C. W. Greider, and C. B. Harley. Telomere end-replication problem and cell aging. *J. Mol. Biol.*, 225(4):951–960, 1992.

[18] R. Neininger. Stochastische Analyse von Algorithmen, Fixpunktgleichungen und ideale Metriken, Habilitationsschrift, University of Frankfurt, 2004.

[19] P. Olofsson and M. Kimmel. Stochastic models of telomere shortening. *Math. Biosci.*, 158(1):75–92, 1999. MR1681442

[20] A. M. Olovnikov. Telomeres, telomerase, and aging: origin of the theory. *Exp. Gerontol.*, 31:443–448, 1996.

[21] R. D. Portugal, M. G. Land, and B. F. Svaiter. A computational model for telomere-dependent cell-replicative aging. *BioSystems*, 91(1):262–267, 2008.

[22] C. J. Proctor and T. B. Kirkwood. Modelling cellular senescence as a result of telomere state. *Aging Cell*, 2(3):151–157, 2003.

[23] I. A. Rodriguez-Brenes and C. S. Peskin. Quantitative theory of telomere length regulation and cellular senescence. *Proc. Natl. Acad. Sci.*, 107(12):5387–5392, 2010.

[24] J. W. Shay and W. E. Wright. Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis*, 26:867–874, 2005.

[25] G. Vogel, E. Pennisi, E. Blackburn, C. Greider, and J. Szostak. Physiology Nobel. U.S. Researchers recognized for work on telomeres. *Science*, 326:212–213, 2009.