



Mathematisches
Forschungsinstitut
Oberwolfach

Member of the



Oberwolfach Preprints

OWP 2022 - 18

ADRIAN GONZÁLEZ CASANOVA, CHARLINE SMADI AND
ANTON WAKOLBINGER

Quasi-Equilibria and Click Times for a Variant of
Muller's Ratchet

Mathematisches Forschungsinstitut Oberwolfach gGmbH
Oberwolfach Preprints (OWP) ISSN 1864-7596

Oberwolfach Preprints (OWP)

The MFO publishes a preprint series **Oberwolfach Preprints (OWP)**, ISSN 1864-7596, which mainly contains research results related to a longer stay in Oberwolfach, as a documentation of the research work done at the MFO. In particular, this concerns the Oberwolfach Research Fellows program (and the former Research in Pairs program) and the Oberwolfach Leibniz Fellows (OWLF), but this can also include an Oberwolfach Lecture, for example.

A preprint can have a size from 1 - 200 pages, and the MFO will publish it in electronic and printed form. Every OWRP group or Oberwolfach Leibniz Fellow may receive on request 20 free hard copies (DIN A4, black and white copy) by surface mail.

The full copyright is left to the authors. With the submission of a manuscript, the authors warrant that they are the creators of the work, including all graphics. The authors grant the MFO a perpetual, irrevocable, non-exclusive right to publish it on the MFO's institutional repository. Since the right is non-exclusive, the MFO enables parallel or later publications, e.g. on the researcher's personal website, in arXiv or in a journal. Whether the other journals also accept preprints or postprints can be checked, for example, via the Sherpa Romeo service.

In case of interest, please send a **pdf file** of your preprint by email to owrf@mfo.de. The file should be sent to the MFO within 12 months after your stay at the MFO.

The preprint (and a published paper) should contain an acknowledgement like: *This research was supported through the program "Oberwolfach Research Fellows" (resp. "Oberwolfach Leibniz Fellows") by the Mathematisches Forschungsinstitut Oberwolfach in [year].*

There are no requirements for the format of the preprint, except that the paper size (or format) should be DIN A4, "letter" or "article".

On the front page of the hard copies, which contains the logo of the MFO, title and authors, we shall add a running number (20XX - XX). Additionally, each preprint will get a Digital Object Identifier (DOI).

We cordially invite the researchers within the OWRP and OWLF program to make use of this offer and would like to thank you in advance for your cooperation.

Imprint:

Mathematisches Forschungsinstitut Oberwolfach gGmbH (MFO)
Schwarzwaldstrasse 9-11
77709 Oberwolfach-Walke
Germany

Tel +49 7834 979 50
Fax +49 7834 979 55
Email admin@mfo.de
URL www.mfo.de

The Oberwolfach Preprints (OWP, ISSN 1864-7596) are published by the MFO.
Copyright of the content is held by the authors.

DOI 10.14760/OWP-2022-18

QUASI-EQUILIBRIA AND CLICK TIMES FOR A VARIANT OF MULLER'S RATCHET

ADRIÁN GONZÁLEZ CASANOVA, CHARLINE SMADI, AND ANTON WAKOLBINGER

ABSTRACT. Consider a population of N individuals, each of them carrying a type in \mathbb{N}_0 . The population evolves according to a Moran dynamics with selection and mutation, where an individual of type k has the same selective advantage over all individuals with type $k' > k$, and type k mutates to type $k + 1$ at a constant rate. This model is thus a variation of the classical Muller's ratchet: there the selective advantage is proportional to $k' - k$. For a regime of selection strength and mutation rates which is between the regimes of weak and strong selection/mutation, we obtain the asymptotic rate of the *click times* of the ratchet (i.e. the times at which the hitherto minimal ('best') type in the population is lost), and reveal the quasi-stationary type frequency profile between clicks. The large population limit of this profile is characterized as the normalized attractor of a "dual" hierarchical multitype logistic system, and also via the distribution of the final minimal displacement in a branching random walk with one-sided steps. An important role in the proofs is played by a graphical representation of the model, both forward and backward in time, and a central tool is the ancestral selection graph decorated by mutations.

CONTENTS

1. Introduction	2
2. Model and main results	2
3. Graphical representation of the model	5
4. Potential ancestral paths and their loads	6
5. Percolation of loads along the Ancestral Selection Graph	9
6. A hierarchy of logistic competitions	10
7. Towards the quasi-stationary type frequency profile via time-reversal	13
8. Quick merging along the Ancestral Selection Graph	14
9. Click times on the Ancestral Selection Graph. 1st proof of Theorem 2.3a)	17
10. Click rates: Proof of Theorem 2.2	21
11. First passage percolation in Poisson-decorated Yule trees	22
12. First passage percolation within the ASG. 2nd proof of Theorem 2.3a)	26
13. The quasi-stationary type frequency profile: Proof of Theorem 2.3b)-d)	29
14. Proof of Lemmata 6.3, 6.4 and 6.5	30
Acknowledgement	36
References	36

2010 *Mathematics Subject Classification*. Primary 60K35; secondary 92D15.

Key words and phrases. Mullers ratchet, click rate, quasi-equilibrium, tournament selection, Harris graphical representation, duality, ancestral selection graph, first passage percolation.

1. INTRODUCTION

A well-known model of population genetics named *Muller's ratchet* (cf. [18, 22, 8, 20, 17] and references therein) considers, in its bare bones version, the interplay between selection and stepwise (slightly) deleterious mutation, for a population of constant size $N \gg 1$. Here is a brief description of the model and of the main questions we will address in this paper. For more details we refer to Sections 2 and 3.

Each individual carries, as its current *type*, a number κ of deleterious mutations. The number of mutations along each lineage increases by 1 at a rate m_N , and as soon as an individual reproduces, its current type is inherited to its 'daughter'. Reproduction happens according to a Moran dynamics with selection, where the fitness difference between two individuals of type κ and κ' is $\frac{s_N}{N}\Phi(\kappa' - \kappa)$ for a selection parameter s_N and a non-decreasing antisymmetric function $\Phi : \mathbb{Z} \rightarrow \mathbb{R}$. (For the classical variant of Muller's ratchet, Φ is the identity function on \mathbb{Z} .) The individual-based dynamics in this model, which we briefly call the Φ -*ratchet*, arises as an independent superposition of the following three ingredients:

Moran resampling. For each pair of individuals (i, j) , irrespective of their types, j is replaced by a newborn daughter of individual i at rate $\frac{1}{2N}$.

Selective reproduction. For each pair of individuals (i, j) for which i currently has type κ and j currently has type $\kappa' > \kappa$, individual j is replaced by a newborn daughter of individual i at rate $\frac{s_N}{N}\Phi(\kappa' - \kappa)$.

Stepwise mutation For each individual, its type is increased by 1 at rate m_N .

This dynamics leads to a sequence of times at which the currently lowest (and thus selectively 'best') type in the population disappears. These times will be referred to as 'click times of the ratchet'. In certain regimes of the parameters s_N and m_N , the click times happen rarely as N becomes large, and a quasi-stationary type profile builds up in between the click times, fastly compared to the duration from click to click.

The following questions thus call for an answer:

- A. What is the rate of the ratchet?
- B. What is the quasi-stationary type profile?

For the classical variant of Muller's ratchet, an asymptotic analysis of these problems beyond existence results is a notoriously difficult - and still unsolved - task, see e.g. [15].

We propose a variant of the fitness function Φ which leads to a model that turns out to be tractable by modern probabilistic techniques, allowing for quantitative results for the rate of the ratchet and the quasi-stationary type profile. This specific choice, denoted by φ , is

$$\varphi(\kappa' - \kappa) := \mathbf{1}_{\{\kappa' - \kappa > 0\}} - \mathbf{1}_{\{\kappa' - \kappa < 0\}}. \quad (1)$$

The essential difference to the classical variant of Muller's ratchet thus is that this selective advantage does not depend on the *value* of the difference of κ' and κ , but only on the *sign* of this difference. This corresponds to *binary tournament selection* [3]: the effect of selection may be imagined as due to pairwise *fights* between randomly chosen individuals, where the individual of 'better' type outcompetes that of worse type.

2. MODEL AND MAIN RESULTS

We now give a definition of the jump rates of the type frequencies of the φ -ratchet with selection strength s_N and mutation rate m_N . For fixed $N \in \mathbb{N}$ and for $\kappa \in \mathbb{N}_0$, let $\xi_\kappa(t) = \xi_\kappa^{(N)}(t)$ be the proportion (or *frequency*) of type κ -individuals at time t . (Here and below we

will sometimes suppress the index N .) Denoting by $(e_i, i \in \mathbb{N}_0)$ the canonical basis of $\mathbb{N}_0^{\mathbb{N}_0}$, the process $(\xi_\kappa, \kappa \in \mathbb{N}_0)$ jumps with the following increments:

- Mutation: for $\kappa \in \mathbb{N}_0$,

$$(e_{\kappa+1} - e_\kappa)/N \quad \text{is added at rate} \quad m_N N \xi_\kappa \quad (2)$$

- Selection: for $\kappa < \kappa'$

$$(e_\kappa - e_{\kappa'})/N \quad \text{is added at rate} \quad s_N N \xi_\kappa \xi_{\kappa'} \quad (3)$$

- Coalescence: for $\kappa \neq \kappa'$

$$(e_\kappa - e_{\kappa'})/N \quad \text{is added at rate} \quad \frac{N}{2} \xi_\kappa \xi_{\kappa'}. \quad (4)$$

Intuitively spoken, the type κ -subpopulation is 'fed through mutations' from the type $(\kappa - 1)$ -subpopulation, is 'selectively attacked' by the type $(< \kappa)$ -subpopulation and 'selectively attacks' the type $(> \kappa)$ -subpopulation. Consequently, these rates imply that for any κ , $(\xi_0, \dots, \xi_\kappa)$ is an autonomous process.

Definition 2.1 (Click times). *a) The best type at time t is defined as*

$$K_N^*(t) := \min \left\{ \kappa \in \mathbb{N}_0 : \xi_\kappa^{(N)}(t) > 0 \right\}. \quad (5)$$

b) The ℓ -th click time $\gamma_\ell^{(N)}$ is the ℓ -th jump time of K_N^ , $\ell = 1, 2, \dots$*

To obtain quantitative results for the click rates and the quasi-stationary type profile, we will throughout the paper consider the case of *moderate* selection and mutation

$$s_N = \frac{\alpha}{f(N)}, \quad m_N = \frac{\mu}{f(N)}, \quad (6)$$

where $\mu < \alpha$, $f(N) \rightarrow \infty$ and $f(N) = o\left(\frac{N}{\log \log N}\right)$ as $N \rightarrow \infty$. In particular, this implies that $m_N \rightarrow 0$, $Nm_N \rightarrow \infty$, and m_N and s_N are of the same order.

Theorem 2.2 (Asymptotic rate of clicks). *Assume that all individuals at time 0 are of type 0, i.e. $\xi_\kappa^{(N)}(0) = \delta_{0\kappa}$, $\kappa \in \mathbb{N}_0$. Then there exists a sequence (θ_N) with*

$$\ln \theta_N \sim 2(\alpha - \mu + \mu \ln(\mu/\alpha)) \frac{N}{f(N)} \quad \text{as } N \rightarrow \infty, \quad (7)$$

such that the sequence of rescaled click times

$$\frac{\gamma_\ell^{(N)}}{f(N)\theta_N}, \quad \ell \in \mathbb{N},$$

converges in distribution as $N \rightarrow \infty$ to a rate 1 Poisson point process.

In particular, for the case of *nearly strong selection* $s_N = 1/l(N)$, where $l(N)$ is any slowly varying function that converges to infinity with N , Theorem 2.2 says that the expected time between clicks is only slightly smaller than exponential in N . In contrast to this, for *nearly weak selection* with $(\ln \ln N)/N \ll s_N \ll (\ln N)/N$, Theorem 2.2 says that the timescale $f(N)\theta_N$ of clicks is asymptotically only slightly larger than the evolutionary timescale N .

Theorem 2.3 (Quasi-stationary type frequency profile). a) *Again assume that all individuals at time 0 are of type 0. Introduce the empirical type profile at time t (seen from the currently best type)*

$$X_k^{(N)}(t) := \xi_{K_N^*(t)+k}^{(N)}(t), \quad k \in \mathbb{N}_0. \quad (8)$$

Let (t_N) be a deterministic sequence of times such that

$$\frac{t_N}{f(N) \ln N} \rightarrow \infty \quad \text{as } N \rightarrow \infty. \quad (9)$$

Then for all $k \in \mathbb{N}_0$

$$X_k^{(N)}(t_N) \rightarrow p_k \quad \text{in probability as } N \rightarrow \infty,$$

where $(p_k)_{k \in \mathbb{N}_0}$ is a sequence of probability weights given by the recursion

$$p_0 = 1 - \frac{\mu}{\alpha} \quad \text{and} \quad p_k^2 - p_k \left(1 - \frac{\mu}{\alpha} - 2 \sum_{k'=0}^{k-1} p_{k'} \right) = \frac{\mu}{\alpha} p_{k-1}, \quad k \geq 1. \quad (10)$$

b) *The recursion (10) is equivalent to the (mutation-selection equilibrium) system*

$$\alpha p_k \left(\sum_{k' \in \mathbb{N}_0} p_{k'} (\mathbf{1}_{\{k' > k\}} - \mathbf{1}_{\{k' < k\}}) \right) = \mu (p_k - p_{k-1}), \quad k \geq 0, \quad (11)$$

with the boundary conditions $p_{-1} = 0$, $p_0 > 0$, $\sum_{k \in \mathbb{N}_0} p_k = 1$.

c) *Let $(p_k)_{k \in \mathbb{N}_0}$ be the probability weights given by (10). Then*

- (i) *for $0 < \frac{\mu}{\alpha} < \frac{2}{3}$, $k \mapsto p_k$ is strictly monotonically decreasing,*
- (ii) *for $\frac{\mu}{\alpha} = \frac{2}{3}$, $k \mapsto p_k$ is monotonically decreasing with $p_0 = p_1 > p_2 > \dots$,*
- (iii) *for $\frac{2}{3} < \frac{\mu}{\alpha} < 1$, $(p_k)_{k \in \mathbb{N}_0}$ is unimodal in the sense that there exist $k_1 \leq k_2$ with $k_2 - k_1 \leq 1$ for which $p_0 < p_1 < \dots < p_{k_1} = p_{k_2}$ and $p_{k_2} > p_{k_2+1} > \dots$.*

In any case, $p_k \sim C \cdot \left(\frac{\mu}{\mu+\alpha}\right)^k$ as $k \rightarrow \infty$ for some constant C depending on $\frac{\mu}{\alpha}$.

d) *Two alternative probabilistic descriptions of $(p_k)_{k \in \mathbb{N}_0}$ given by (10) are as follows:*

- *Consider a Yule tree with splitting rate α whose branches are decorated by a rate μ Poisson point process. Then p_k is the probability that there is an infinite lineage carrying exactly k points but no infinite lineage with less than k points.*
- *Consider a branching random walk on \mathbb{N}_0 starting with one individual at the origin with binary branching at rate α (and no death) and with migration of individuals from k to $k+1$ at rate μ . Then, as $t \rightarrow \infty$, the minimal position of the individuals alive at time t converges in law to a random variable with distribution $(p_k)_{k \in \mathbb{N}_0}$.*

Remark 2.4. a) Eq. (11) characterizes the type frequency profile $(p_k)_{k \in \mathbb{N}_0}$ as the fixed point of a deterministic mutation-selection equilibrium, with the out-flux due to mutation on its right hand side and the in-flux due to selection on its left hand side. The latter can be written as $p_k \alpha \sum_{k' \in \mathbb{N}_0} \varphi(k' - k) p_{k'}$ with φ as in (1). If φ would be replaced by $\Phi(k' - k) := k' - k$, then (cf. [12]) the solution of (11) would be the Poisson weights with parameter μ/α . An essential advantage of the form (1) of the fitness function is that it opens the way to a mathematically tractable analysis of the probabilistic system via a dual process within a graphical representation. This graphical representation together with the analysis of the dual

process will be the main tool in our paper. The quasi-stationary center of attraction of the dual process turns out to be proportional to $(p_k)_{k \in \mathbb{N}_0}$ given by (10), see Remark 6.1 e).

b) The graphical representation allows for a clear interpretation of the 'first passage percolation'-description of the type frequency profile given in part d) of Theorem 2.3. For large population size N and on a suitably short timescale, the *decorated Ancestral Selection Graph* (see Definitions 4.2 and 4.6) of a typical individual (sampled from the population) looks like a Yule tree with mutations. It merges quickly with the best class within the *equilibrium Ancestral Selection Graph* [21] between clicks, and the type of the sampled individual is (asymptotically as $N \rightarrow \infty$) determined by the minimal load of mutations collected along the lineages of the Yule tree. All this will be made precise in the sequel.

3. GRAPHICAL REPRESENTATION OF THE MODEL

The type frequency process $\xi^{(N)}$ of the φ -ratchet, which was introduced at the beginning of Section 2, can be constructed (in a similar way as in [11, 7, 10]) on top of a Moran graph with selection parameter s_N , with mutations added by means of an independent Poisson process.

Definition 3.1 (Graphical elements). *For fixed $N \in \mathbb{N}$, we consider three independent Poisson point processes, $\mathcal{C}^{(N)}$, $\mathcal{S}^{(N)}$ and $\mathcal{M}^{(N)}$. The processes $\mathcal{C}^{(N)}$ and $\mathcal{S}^{(N)}$ are supported by $\{(i, j) : i, j \in [N], i \neq j\} \times \mathbb{R}$ and have on each component $\{(i, j)\} \times \mathbb{R}$ the constant intensity $\frac{1}{2N}$ and $\frac{s_N}{N}$, respectively. The third process, $\mathcal{M}^{(N)}$, is a Poisson point process on $G^{(N)} := [N] \times \mathbb{R}$ with constant intensity m_N on each component $\{i\} \times \mathbb{R}$.*

Remark 3.2. When there is no risk of confusion, we will suppress the index N and write $G, \mathcal{C}, \mathcal{S}, \mathcal{M}$. We will speak of the points $(i, t) \in G$, $i \in [N]$, as *the individuals living at time t* . Each point $(i, j, t) \in \mathcal{C} \cup \mathcal{S}$ can be visualized as an arrow pointing from line i to line j at time t . At an $(i, j, t) \in \mathcal{C}$, the individual (i, t) , irrespective of its type, bears a daughter (j, t) who replaces the individual $(j, t-)$. At an $(i, j, t) \in \mathcal{S}$, the same happens, but only provided the individual (i, t) carries less mutations than the individual $(j, t-)$. The process \mathcal{M} describes the mutations occurring along the lines; each point of \mathcal{M} increases the mutational load along the lineage by 1. This is made precise in Definition 3.3, and illustrated in Figure 1.

Definition 3.3 (Type configurations and their transport). *A type configuration is an element of $\mathbb{N}_0^{[N]}$, thus assigning a (nonnegative integer) type to each $i \in [N]$. The process $(\mathcal{C}, \mathcal{S}, \mathcal{M})$ gives rise to a transport of type configurations $(\eta(i, t))_{i \in [N]}$ as $t \in \mathbb{R}$ increases. Specifically, think of an initial type configuration $(\eta(i, s))_{i \in [N]}$ being given at some time s . At times $t > s$, the three Poisson point processes \mathcal{C}, \mathcal{S} and \mathcal{M} act as follows:*

- if the point (i, j, t) belongs to \mathcal{C} , then

$$\eta(j, t) = \eta(i, t-) \quad (= \eta(i, t) \text{ a.s.})$$

- if the point (i, j, t) belongs to \mathcal{S} , then

$$\eta(j, t) = \begin{cases} \eta(j, t-) & \text{if } \eta(j, t-) \leq \eta(i, t-) \\ \eta(i, t-) & \text{if } \eta(i, t-) < \eta(j, t-) \end{cases}$$

- if the point (i, t) belongs to \mathcal{M} , then

$$\eta(i, t) = \eta(i, t-) + 1.$$

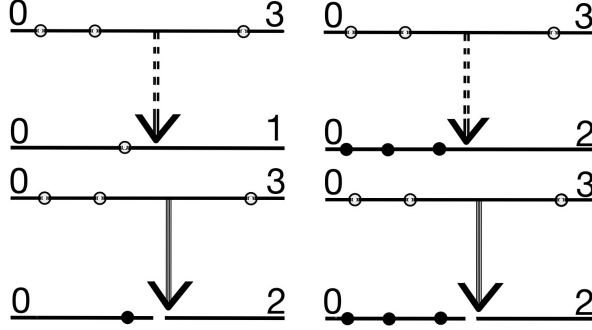


FIGURE 1. Graphical elements and their impact on the transport of types. Time is running from left to right, and in each of the four panels two levels ($i = 1, 2$) are considered, with the initial type configuration $(0, 0)$. Selective and neutral arrows are drawn with dashed and solid shafts, respectively. Mutations are drawn as circles. Because of the rules described in Definition 3.3, some of the mutations do not have an effect on the outcome of the types at the final time; these mutations are represented as filled black circles.

In the next section and thereafter we will use the Poisson point processes $(\mathcal{C}, \mathcal{S}, \mathcal{M})$ also for a transport (of potential ancestral paths and mutational loads) *backwards in time*. In order to clearly distinguish between forward and backward concepts, we define two filtrations generated by $(\mathcal{C}, \mathcal{S}, \mathcal{M})$.

Definition 3.4 (Forward and backward filtrations). *For $t \in \mathbb{R}$ let $\mathcal{C}_{\leq t}$ and $\mathcal{S}_{\leq t}$ be the restrictions of \mathcal{C} and \mathcal{S} to $\bigcup_{i,j \in [N], i \neq j} \{(i, j)\} \times (-\infty, t]$, and let $\mathcal{M}_{\leq t}$ be the restriction of \mathcal{M} to*

$\bigcup_{i \in [N]} \{i\} \times (-\infty, t]$. Likewise, define $\mathcal{C}_{\geq t}$, $\mathcal{S}_{\geq t}$, $\mathcal{M}_{\geq t}$, replacing $(-\infty, t]$ by $[t, \infty)$. Let \mathcal{F}_t and

\mathcal{P}_t be the σ -algebras generated by $\mathcal{C}_{\leq t}, \mathcal{S}_{\leq t}, \mathcal{M}_{\leq t}$ and $\mathcal{C}_{\geq t}, \mathcal{S}_{\geq t}, \mathcal{M}_{\geq t}$, respectively. The forward filtration is $\mathcal{F} := (\mathcal{F}_t)_{t \geq 0}$ and the backward filtration is $\mathcal{P} := (\mathcal{P}_t)_{t \in \mathbb{R}}$; note that \mathcal{P}_t increases as t decreases. All these objects are understood for fixed population size N ; sometimes we will write $\mathcal{P}^{(N)}$ and $\mathcal{F}^{(N)}$ instead of \mathcal{P} and \mathcal{F} , to make the dependence on N explicit.

Remark 3.5. With $s := 0$ and $\eta(i, 0) := 0$, $i \in [N]$, and with $\eta(\cdot, t)$, $t \geq 0$, constructed according to Definition 3.3, the process of type frequency evolutions that figures in Theorems 2.2 and 2.3 can now be represented as the \mathcal{F} -adapted process

$$\xi_{\kappa}(t) := \frac{1}{N} \# \{i \in [N] : \eta(i, t) = \kappa\}, \quad \kappa \in \mathbb{N}_0, \quad t \geq 0. \quad (12)$$

Indeed it is straightforward that this process has the jump rates given in (2), (3), (4). In terms of η , the best type at time t (defined in (5)) has the representation

$$K_N^*(t) = \min\{\eta(j, t) : j \in [N]\}. \quad (13)$$

4. POTENTIAL ANCESTRAL PATHS AND THEIR LOADS

While the graphical representation given in the previous section was a *forward in time* construction, we now take a *backward in time* point of view. This is based on the concept of *potential ancestral lineages* which goes back to pioneering work of Krone and Neuhauser [14, 19]. The key idea is to construct in a first stage an untyped version of the (potential) genealogy

backwards in time and decide in a second stage forwards in time which lineages become “real”. Specifically, a “selective arrow” $(i, j, t) \in \mathcal{S}$ introduces the two potential parents $(i, t-)$ and $(j, t-)$ of the individual (j, t) . Thus, a potential ancestral lineage backwards in time should jump from (j, t) to $(i, t-)$ as soon as it encounters the head j of a “neutral arrow” $(i, j, t) \in \mathcal{C}$, and should branch into two selective lineages as soon as it encounters the head j of a “selective arrow” $(i, j, t) \in \mathcal{S}$. We will formalize this by the concept of (*potential ancestral*) *paths*.

Definition 4.1 (Paths and potential ancestors). *Let $(i, s), (j, t) \in G$ with $s < t$. A (potential ancestral) path connecting (i, s) and (j, t) is a subset of G of the form*

$$(\{i_1\} \times [s_0, s_1]) \cup (\{i_2\} \times [s_1, s_2]) \cup \dots \cup (\{i_n\} \times [s_{n-1}, s_n]),$$

with the following properties

- a) $s = s_0 \leq s_1 < \dots < s_{n-1} \leq s_n = t$,
- b) $i = i_1, j = i_n$,
- c) $(i_g, i_{g+1}, s_g) \in \mathcal{C} \cup \mathcal{S}$ for $g = 1, \dots, n-1$,
- d) $\mathcal{C}(([N] \setminus \{i_g\}) \times \{i_g\} \times (s_{g-1}, s_g)) = 0$ for $g = 1, \dots, n-1$.

We write $(i, s) \prec (j, t)$ if there is a path connecting (i, s) and (j, t) . In this case we say that (i, s) is a potential ancestor of (j, t) .

In words, the conditions mean that jumps between different levels $h, h' \in [N]$ may only occur at time points of either neutral or selective arrows, and that none of the time intervals (s_{g-1}, s_g) , $g = 1, \dots, n$, may be hit by a neutral arrow whose arrow-head is at i_g .

As a consequence of this definition we observe (see Figure 2 for an illustration):

- If the point (i, j, t) belongs to \mathcal{C} , the point (j, t) is disconnected with $(j, t-)$ and connected with $(i, t-)$.
- If the point (i, j, t) belongs to \mathcal{S} , the point (j, t) is connected both with $(i, t-)$ and $(j, t-)$.

Definition 4.2 (Ancestral selection graph (ASG)). *For $t \in \mathbb{R}$ and $J_t \subset [N] \times \{t\}$ we define, suppressing the index N ,*

$$\mathcal{A}_s^{J_t} := \{(i, s) : (i, s) \prec v \text{ for some } v \in J_t\} \quad \text{and} \quad \mathcal{A}^{J_t} := \bigcup_{s \leq t} \mathcal{A}_s^{J_t}.$$

Thinking of \mathcal{A}^{J_t} as a union of paths jointly with the graphical elements from \mathcal{C} and \mathcal{S} by which it was induced, we call \mathcal{A}^{J_t} the ASG back from J_t .

For a singleton $J_t = \{(j, t)\}$ we write $\mathcal{A}_s^{j,t}$ instead of $\mathcal{A}_s^{\{(j,t)\}}$, and for $J_t = [N] \times \{t\}$ we briefly write \mathcal{A}_s^t instead of $\mathcal{A}_s^{[N] \times \{t\}}$, and \mathcal{A}^t instead of $\mathcal{A}^{[N] \times \{t\}}$.

Definition 4.3 (Load and \mathcal{M} -distance). (i) *The load of a path is the number of points of \mathcal{M} carried by the path.*

(ii) *The \mathcal{M} -distance $d_{\mathcal{M}}((i, s), (j, t))$ of two points $(i, s), (j, t) \in G$ with $s < t$ is the minimal load of all paths connecting them, with the convention that the minimum over an empty set is infinity. We say that (i, s) is a load k potential ancestor of (j, t) if $d_{\mathcal{M}}((i, s), (j, t)) = k$.*

(iii) *For $s < t$ and $I_s \subset [N] \times \{s\}, J_t \subset [N] \times \{t\}$ we put*

$$d_{\mathcal{M}}(I_s, J_t) := \min\{d_{\mathcal{M}}(v, w) : v \in I_s, w \in J_t\}.$$

Remark 4.4. For three points $(i, s), (j, t), (g, u) \in G$ with $s < t < u$ one may have $(i, s) \prec (j, t)$ and $(i, s) \prec (g, u)$ but $(j, t) \not\prec (g, u)$. This shows that in general $d_{\mathcal{M}}$ violates the triangle

inequality. If, however, the three points satisfy $(i, s) \prec (j, t) \prec (g, u)$, then one has

$$d_{\mathcal{M}}((i, s), (g, u)) \leq d_{\mathcal{M}}((i, s), (j, t)) + d_{\mathcal{M}}((j, t), (g, u)).$$

Indeed, in this case the concatenation of a path of minimal load connecting (i, s) and (j, t) with a path of minimal load connecting (j, t) and (g, u) is a path connecting (i, s) and (g, u) ; hence $d_{\mathcal{M}}$ follows the claimed “restricted” triangle inequality.

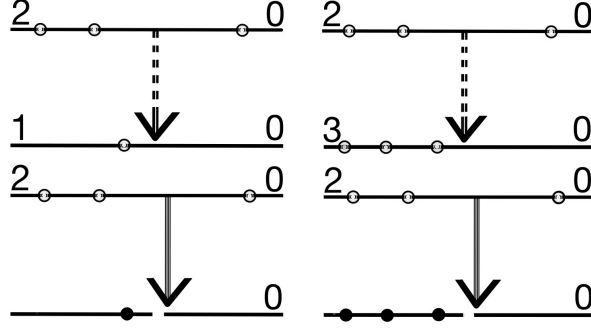


FIGURE 2. This figure contains the same graphical elements than Figure 1, but now the paths are followed backwards. Let us think of the left hand side of each of the four panels corresponding to time s and the right hand side corresponding to $t > s$. At time s , the \mathcal{M} -distance between the set $\{1, 2\} \times \{t\}$ and its potential ancestors at time s is annotated. A comparison with Figure 1 shows differences and similarities between the backward and forward transport.

The next remark is immediate from the graphical construction.

Remark 4.5 (Flow of type configurations). Let $\eta = \eta^{(N)}$ be as specified in Definition 3.3. Then for any $j \in [N]$, $0 \leq s < t$, one has a.s.

$$\eta(j, t) = \min_{i \in [N]} \{\eta(i, s) + d_{\mathcal{M}}((i, s), (j, t))\}.$$

In particular, if $\eta(i, 0) = 0$ for all $i \in [N]$, then

$$\eta(j, t) = d_{\mathcal{M}}([N] \times \{0\}, (j, t)) = d_{\mathcal{M}}(\mathcal{A}_0^{j,t}, (j, t)) \quad (14)$$

and

$$K_N^*(t) = \min\{\eta(j, t) : j \in [N]\} = d_{\mathcal{M}}([N] \times \{0\}, [N] \times \{t\}). \quad (15)$$

Definition 4.6. a) (Load k potential ancestors) Let $k \in \mathbb{N}_0$ and $s \leq t \in \mathbb{R}$. For $J_t \subset [N] \times \{t\}$ we define $\mathcal{A}_s^{J_t}(k)$ as the set of individuals (i, s) which are load k potential ancestors of some individual in J_t (cf. Definition 4.3). Taking the union over $s \in (\infty, t]$ we define $\mathcal{A}^{J_t}(k)$ as the set of all individuals which are load k potential ancestors of some individual in J_t .

b) Minimum load potential ancestors. We define the set of minimum load potential ancestors at time s of the population J_t as

$$\bar{\mathcal{A}}_s^{J_t} := \mathcal{A}_s^{J_t}(\underline{k}) \quad (16)$$

where

$$\underline{k} := \underline{k}(s, t) := \min\{k \in \mathbb{N}_0 : \mathcal{A}_s^{J_t}(k) \neq \emptyset\}.$$

To ease notation we write $\bar{\mathcal{A}}_s^t$ instead of $\bar{\mathcal{A}}_s^{[N] \times \{t\}}$, and $\bar{\mathcal{A}}_s^{i,t}$ instead of $\bar{\mathcal{A}}_s^{\{(i,t)\}}$.

c) The definitions in a) and b) extend directly from deterministic t and J_t to a \mathcal{P} -stopping time T and a \mathcal{P}_T -measurable random set $\mathcal{J}_T \subset [N] \times \{T\}$.

5. PERCOLATION OF LOADS ALONG THE ANCESTRAL SELECTION GRAPH

In this section we fix a population size $N \in \mathbb{N}$ which we suppress in the notation. For fixed $t \in \mathbb{R}$ and $J_t \subset [N] \times \{t\}$, as $s \leq t$ decreases, the evolution of the set-valued processes $\mathcal{A}_s^{J_t}(k)$, $k \in \mathbb{N}_0$, is driven by the Poisson point processes $(\mathcal{C}, \mathcal{S}, \mathcal{M})$, now in a \mathcal{P} -adapted manner. In view of Definitions 4.1 and 4.2, the actions of $(\mathcal{C}, \mathcal{S}, \mathcal{M})$ on the sets $\mathcal{A}_s^{J_t}(k)$ are as follows (note the analogy and the differences to Definition 3.3 for the transport of type configurations which there was forward in time):

- *Coalescences:* Let $(i, j, s) \in \mathcal{C}$; $k, k' \in \mathbb{N}_0$. If $(i, s) \in \mathcal{A}_s^{J_t}(k)$ and $(j, s) \in \mathcal{A}_s^{J_t}(k')$, then for $k \leq k'$

$$\mathcal{A}_{s-}^{J_t}(k') = \mathcal{A}_s^{J_t}(k') \setminus \{(j, s)\} \quad \text{and} \quad \mathcal{A}_{s-}^{J_t}(k) = \mathcal{A}_s^{J_t}(k)$$

whereas for $k > k'$

$$\mathcal{A}_{s-}^{J_t}(k') = (\mathcal{A}_s^{J_t}(k') \setminus \{(j, s)\}) \cup \{(i, s)\} \quad \text{and} \quad \mathcal{A}_{s-}^{J_t}(k) = \mathcal{A}_s^{J_t}(k) \setminus \{(i, s)\}.$$

- *Selective branching:* Let $(i, j, s) \in \mathcal{S}$; $k \in \mathbb{N}_0$. If $(i, s) \in \mathcal{A}_s^{J_t}(k)$ and $(j, s) \notin \mathcal{A}_s^{J_t}$, then

$$\mathcal{A}_{s-}^{J_t}(k) = \mathcal{A}_s^{J_t}(k) \cup \{(j, s)\}.$$

- *Selective competition:* Let $(i, j, s) \in \mathcal{S}$; $k, k' \in \mathbb{N}_0$. If $(i, s) \in \mathcal{A}_s^{J_t}(k)$ and $(j, s) \in \mathcal{A}_s^{J_t}(k')$, then for $k \leq k'$

$$\mathcal{A}_{s-}^{J_t}(k') = \mathcal{A}_s^{J_t}(k') \setminus \{(j, s)\} \quad \text{and} \quad \mathcal{A}_{s-}^{J_t}(k) = \mathcal{A}_s^{J_t}(k) \cup \{(j, s)\}$$

whereas for $k > k'$

$$\mathcal{A}_{s-}^{J_t}(k') = \mathcal{A}_s^{J_t}(k') \cup \{(i, s)\} \quad \text{and} \quad \mathcal{A}_{s-}^{J_t}(k) = \mathcal{A}_s^{J_t}(k) \setminus \{(i, s)\}.$$

- *Mutation:* Let $(i, s) \in \mathcal{M}$; $k \in \mathbb{N}_0$. If $(i, s) \in \mathcal{A}_s^{J_t}(k)$ then

$$\mathcal{A}_{s-}^{J_t}(k) = \mathcal{A}_s^{J_t}(k) \setminus \{(i, s)\} \quad \text{and} \quad \mathcal{A}_{s-}^{J_t}(k+1) = \mathcal{A}_s^{J_t}(k+1) \cup \{(i, s)\}.$$

Due to the symmetry properties of the dynamics (backwards in time) that is induced by the just described transitions, we may focus our attention on the configuration of cardinalities of the sets $\mathcal{A}_s^{J_t}(k)$, and define

$$A_s^{J_t}(k) := \#\mathcal{A}_s^{J_t}(k), \quad k \in \mathbb{N}_0.$$

For $J_t = [N] \times \{t\}$ we write $A_s^t(k)$ instead of $A_s^{[N] \times \{t\}}(k)$, and $A^t(k)$ instead of $A^{[N] \times \{t\}}(k)$.

The following lemma is immediate from the above described actions of the Poisson point processes $(\mathcal{C}, \mathcal{S}, \mathcal{M})$ on the sets $\mathcal{A}_s^{J_t}(k)$.

Lemma 5.1. *For all $t \in \mathbb{R}$ and all sets $J_t \subset [N] \times \{t\}$, the process*

$$(A_{t-r}^{J_t}(0), A_{t-r}^{J_t}(1), \dots, A_{t-r}^{J_t}(k), \dots)_{r \geq 0}$$

is Markovian when randomized over $(\mathcal{C}, \mathcal{S}, \mathcal{M})$. Its state space is the set

$$\mathcal{Z}_N := \{z = (z_0, z_1, \dots) : z_k \in \mathbb{N}_0, z_0 + z_1 + \dots \leq N\}. \quad (17)$$

Its jump rates from $z \in \mathcal{Z}_N$ are (with e_k as in Section 2 and s_N, m_N as in (6))

- *Coalescences:* for any $k \in \mathbb{N}$,

$$z \rightarrow z - e_k \quad \text{with rate} \quad \frac{1}{2N} z_k (z_k - 1) + \frac{1}{N} z_k \sum_{0 \leq k' < k} z_{k'}. \quad (18)$$

- *Selective branching:* for any $k \in \mathbb{N}_0$,

$$z \rightarrow z + e_k \quad \text{with rate} \quad \frac{s_N}{N} z_k \left(N - \sum_{0 \leq k' < \infty} z_{k'} \right). \quad (19)$$

- *Selective competition:* for any pair of integers (k, k') such that $0 \leq k < k'$,

$$z \rightarrow z + e_k - e_{k'} \quad \text{with rate} \quad \frac{s_N}{N} z_k z_{k'}. \quad (20)$$

- *Mutation:* for any $k \in \mathbb{N}_0$,

$$z \rightarrow z + e_{k+1} - e_k \quad \text{with rate} \quad m_N z_k$$

and for any $k \in \mathbb{N}$,

$$z \rightarrow z + e_k - e_{k-1} \quad \text{with rate} \quad m_N z_{k-1}. \quad (21)$$

Remark 5.2. In Lemma 5.1 we may replace the deterministic time t by a \mathcal{P} -stopping time T , and the deterministic set J_t by a \mathcal{P}_T -measurable random set $\mathcal{J}_T \subset [N] \times \{T\}$, with the backward filtration \mathcal{P} specified in Definition 3.4.

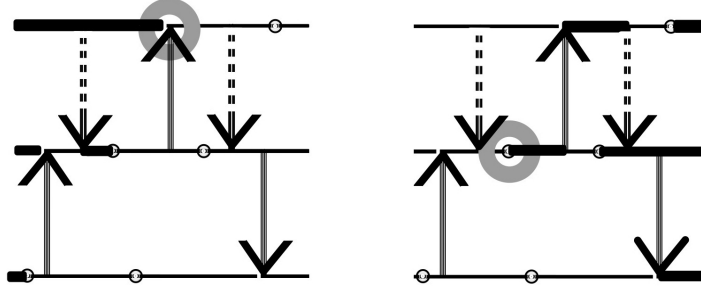


FIGURE 3. Key to the analysis of the click times of the ratchet are the instances at which (seen backward in time) the paths with minimal load are lost. In this figure we observe how the click times forward and backward are different, but strongly related to each other and also close in time. Both panels contain the same graphical elements, with the left panel showing the forward transport and the right panel showing the backward transport of \mathcal{M} -distances. In each case, the points with \mathcal{M} -distance 0 from the left respectively from the right boundary are shown by thick lines, and *clicks* are indicated by a large circle.

6. A HIERARCHY OF LOGISTIC COMPETITIONS

Throughout this section we consider, for any given $N \in \mathbb{N}$, a Markovian jump process $Z^{(N)} := (Z^{(N)}(r))_{r \geq 0}$, whose state space is \mathcal{Z}_N defined in (17) and whose jump rates are given by (18) to (21). When there is no risk of confusion, we suppress the superscript N and write e.g. Z_0 instead of $Z^{(N)}$.

Remark 6.1. a) The processes Z_0 and (Z_0, Z_1) are Markovian. Specifically, the process Z_0 jumps with the following rates:

$$\begin{aligned} z_0 \rightarrow z_0 + 1 & \quad \text{at rate} \quad s_N z_0 \left(1 - \frac{z_0}{N}\right) =: z_0 b_0(z_0, N), \\ z_0 \rightarrow z_0 - 1 & \quad \text{at rate} \quad z_0 \left(m_N + \frac{z_0 - 1}{2N}\right) =: z_0 d_0(z_0, N). \end{aligned} \quad (22)$$

b) We may consider the process Z_1 as living in an environment given by Z_0 : Given that the current state of Z_0 is $z_0 \in \mathbb{N}_0$, the process Z_1 jumps with the rates

$$\begin{aligned} z_1 \rightarrow z_1 + 1 & \quad \text{at rate} \quad s_N z_1 \left(1 - \frac{z_0 + z_1}{N}\right) + m_N z_0, \\ z_1 \rightarrow z_1 - 1 & \quad \text{at rate} \quad \frac{z_1(z_1 - 1)}{2N} + z_1 \left(m_N + \frac{z_0}{N}\right) + s_N \frac{z_0 z_1}{N}. \end{aligned}$$

c) For each $k \in \mathbb{N}$ the process (Z_0, \dots, Z_{k-1}) is Markovian, and given that the current state of $(Z_0 + \dots + Z_{k-1}, Z_{k-1})$ is (\tilde{z}, z_{k-1}) , the jump rates of Z_k from a state z_k are

$$\begin{aligned} z_k \rightarrow z_k + 1 & \quad \text{at rate} \quad s_N z_k \left(1 - \frac{\tilde{z} + z_k}{N}\right) + m_N z_{k-1}, \\ z_k \rightarrow z_k - 1 & \quad \text{at rate} \quad \frac{z_k(z_k - 1)}{2N} + \frac{z_k \tilde{z}}{N} + s_N \frac{z_k \tilde{z}}{N} + m_N z_k. \end{aligned} \quad (23)$$

d) An inspection of the rates in (23) and an application of a dynamical law of large numbers from [9] shows that if $Z_0^{(N)}(0)$ is of order $N/f(N)$, the process $(f(N)Z^{(N)}(f(N)r)/N)_{r \geq 0}$ is on each time interval $[0, r_0]$ for large N close (uniformly in $r \in [0, r_0]$) to the solution of the dynamical system

$$\frac{dn_k(t)}{dt} = \mu n_{k-1}(t) + n_k(t) \left(\alpha - \mu - \frac{n_k(t)}{2} - \sum_{i=0}^{k-1} n_i(t) \right), \quad k \geq 0 \quad (24)$$

with $n_{-1} := 0$. Without going into all details here, let us mention that two steps are needed to prove this convergence. First, we consider a modified version of the process $Z^{(N)}$, namely $\tilde{Z}^{(N)}$, where the rates in (19) and (20) are replaced by $s_N z_k$ and 0 respectively. Choosing as the mass rescaling parameter the carrying capacity $N/f(N)$, we can directly apply Theorem 11.2.1 in [9] to the process $(\tilde{Z}^{(N)}(f(N)r))_{r \geq 0}$. Then applying Lemma C.1 in [6] as in the proof of Lemma 6.3, we obtain that the sum of the components of the process $(\tilde{Z}^{(N)}(f(N)r))_{r \geq 0}$ does not reach a size of order N within a time of order $\ln N$ with a probability close to 1 for large N . The modification of the jump rates is thus negligible on a time scale of order 1, and the claimed convergence holds for the process $(Z^{(N)}(f(N)r))_{r \geq 0}$.

e) The system (24) has a unique attracting equilibrium $(\bar{n}_k)_{k \in \mathbb{N}_0}$ which follows the recursion

$$\bar{n}_0 := 2(\alpha - \mu) \quad \text{and} \quad \mu \bar{n}_{k-1} + \bar{n}_k \left(\alpha - \mu - \frac{\bar{n}_k}{2} - \sum_{i=0}^{k-1} \bar{n}_i \right) = 0, \quad k \geq 1. \quad (25)$$

The process $(Z_0^{(N)}, \dots, Z_k^{(N)})$ has a *quasi-stationary equilibrium* (see e.g. [5, Example 7]) which we denote by $\nu_N^{(k)}$. We thus obtain that for any $r \geq 0$,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\nu_N^{(k)}} \left[\frac{f(N)Z_k^{(N)}(r)}{N} \mid Z_0^{(N)}(r) \geq 1 \right] = \bar{n}_k.$$

Summing over k in (25) and defining $\bar{n} := \sum_{k=0}^{\infty} \bar{n}_k$, we get

$$0 = \alpha \bar{n} - \frac{1}{2} \left(\sum_{k=0}^{\infty} \bar{n}_k^2 + 2 \sum_{0 \leq i < k < \infty} \bar{n}_k \bar{n}_i \right) = \alpha \bar{n} - \frac{1}{2} \bar{n}^2.$$

This yields

$$\bar{n} = \sum_{k=0}^{\infty} \bar{n}_k = 2\alpha. \quad (26)$$

Remark 6.2. Let $(\bar{n}_k)_{k \in \mathbb{N}_0}$ be defined by the recursion (25). In view of (25) and (26) it is clear that $\frac{\bar{n}_k}{2\alpha}$, $k = 0, 1, \dots$, is a sequence of probability weights which satisfies the recursion (10) and thus coincides with the probability weights p_k , $k \in \mathbb{N}_0$, that are defined in Theorem 2.3a).

The next lemma roughly says that for any $k \in \mathbb{N}_0$ the process $Z_k^{(N)}$ with high probability grows quickly to a size of order $N/f(N)$ and stays there at least for a time of order $f(N) \ln N$, provided only that for some $\ell \leq k$ the initial size of $Z_\ell^{(N)}$ is not too small. In view of Lemma 5.1, the quantity $N/f(N)$ thus characterizes the typical size of the ASG on the $f(N) \ln N$ - timescale.

Lemma 6.3. *Let $(\bar{n}_g, g \in \mathbb{N}_0)$ be given by the recursion (25). Let $(Z_0, Z_1, \dots) = (Z_0^{(N)}, Z_1^{(N)}, \dots)$ be a process with jump rates given by (18) to (21), and let $R > 0$. Then for any $k \in \mathbb{N}_0$ and $\varepsilon > 0$, there exist finite constants C_k and $C_k(\varepsilon)$ such that*

$$\liminf_{N \rightarrow \infty} \mathbb{P} \left(\frac{f(N)}{N} Z_k(r) \in [\bar{n}_k - C_k \varepsilon, \bar{n}_k + C_k \varepsilon] \text{ for all } r \text{ such that } \frac{r}{f(N) \ln N} \in [C_k(\varepsilon), C_k(\varepsilon) + R] \mid \exists \ell \leq k \text{ such that } Z_\ell(0) \geq 1/\varepsilon \right) = 1 - \delta(\varepsilon), \quad (27)$$

where $\delta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

The proof of this lemma (as well as the proofs of the other lemmata of this section) will be given in Section 14.

For the process $Z_0^{(N)}$ with jump rates (22), we introduce the first time at which $Z_0^{(N)}$ hits the trap 0:

$$H_0^{(N)} := \inf\{r \geq 0, Z_0^{(N)}(r) = 0\}. \quad (28)$$

Lemma 6.4. *Let $\nu_N := \nu_N^{(0)}$ denote the quasi-stationary distribution of $Z_0^{(N)}$. For every N there exists θ_N such that*

$$\mathbb{P}_{\nu_N} \left(H_0^{(N)} > r \right) = e^{-r/f(N)\theta_N}, \quad r > 0, \quad (29)$$

where the sequence (θ_N) satisfies (7).

Lemma 6.5. *a) There exists $C > 0$ such that for $\varepsilon > 0$*

$$\liminf_{N \rightarrow \infty} \mathbb{P} \left(Z_1^{(N)}(H_0^{(N)}) \geq \frac{C}{\varepsilon^2} \mid Z_0^{(N)}(0) \geq 1/\varepsilon \right) = 1 - \delta(\varepsilon),$$

where $\delta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

b) With (θ_N) as in Lemma 6.4, let (r_N) be a sequence with $\frac{r_N}{f(N) \ln N} \rightarrow \infty$ as $N \rightarrow \infty$, and define $\underline{k}_N := \min\{k \in \mathbb{N} : Z^{(N)}(r_N) > 0\}$. Then

$$\liminf_{N \rightarrow \infty} \mathbb{P} \left(Z_{\underline{k}_N}^{(N)}(r_N) \geq \frac{1}{\varepsilon} \mid Z_0^{(N)}(0) \geq 1/\varepsilon \right) = 1 - \delta(\varepsilon),$$

where $\delta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

These two lemmas are key for obtaining the renewal structure of the dynamics of the potential ancestors with minimal load. They imply in particular that when the set of potential ancestors with the currently minimal load gets extinct, the number of minimum load potential ancestors that “come next” is large enough for reaching a size of order $N/f(N)$ given by the quasi-stationary distribution ν_N . As we will see in Section 10, this will ensure, using duality, that the succession of several clicks (in the sense of Definition 2.1) within a time frame of order smaller than $f(N)\theta_N$ is not likely.

7. TOWARDS THE QUASI-STATIONARY TYPE FREQUENCY PROFILE VIA TIME-REVERSAL

A principal tool in this section will be the *equilibrium ASG* $\mathcal{A}^{N,\text{eq}}$. Its definition relies on the observation that for each fixed $N \in \mathbb{N}$ and $t > 0$ there exists some $u_0 > t$ such that

$$\mathcal{A}_{[0,t]}^{N,\text{eq}} := \mathcal{A}^{u_0} \cap ([N] \times [0, t]) \stackrel{\text{a.s.}}{=} \mathcal{A}^u \cap ([N] \times [0, t]) \quad \text{for all } u \geq u_0.$$

In accordance with (14) we observe that for $(j, t) \in \mathcal{A}_t^{N,\text{eq}}$,

$$\eta(j, t) = d_{\mathcal{M}}(\mathcal{A}_0^{N,\text{eq}}, (j, t)).$$

We define for $v > 0$ and $(i, 0) \in \mathcal{A}_0^{N,\text{eq}}$,

$$\zeta_t(i, 0) := d_{\mathcal{M}}((i, 0), \mathcal{A}_t^{N,\text{eq}}).$$

It was proved in [21, Lemma 2.3] that $\mathcal{A}^{N,\text{eq}}$ is time-reversible in the sense that for all $t > 0$

$$R_0^t \mathcal{A}_{[0,t]}^{N,\text{eq}} \stackrel{d}{=} \mathcal{A}_{[0,t]}^{N,\text{eq}}$$

where R_0^t is the bijection on $[N] \times [0, t]$ that takes (i, s) into $(i, t-s)$; note that this time-reversal takes coalescences into branchings and vice versa. Since time-homogeneous Poisson point processes are time-reversible, this equality in distribution carries over to the \mathcal{M} -decorated version of $\mathcal{A}_{[0,t]}^{N,\text{eq}}$ as well. As a consequence we have the following equality in distribution for the configuration of \mathcal{M} -distances observed backwards and forwards in $\mathcal{A}_{[0,t]}^{N,\text{eq}}$:

$$(\zeta_t(i, 0))_{(i,0) \in \mathcal{A}_0^{N,\text{eq}}} \stackrel{d}{=} (\eta(j, t))_{(j,t) \in \mathcal{A}_t^{N,\text{eq}}} \quad (30)$$

Next we define

$$\tilde{Z}_k^N(t) := \# \left\{ (i, 0) \in \mathcal{A}_0^{N,\text{eq}} : \zeta_t(i, 0) = k \right\}, \quad k \in \mathbb{N}_0.$$

It is clear from the definition of $Z_k^{(N)}$ at the beginning of Section 6 and from Lemma 5.1 that for all fixed $t > 0$

$$\left(Z_k^{(N)}(t) \right)_{k \in \mathbb{N}_0} \stackrel{d}{=} \left(\tilde{Z}_k^N(t) \right)_{k \in \mathbb{N}_0}. \quad (31)$$

For $r > 0$ we define

$$\bar{k}(r) := \bar{k}_N(r) := \min \left\{ k : Z_k^{(N)}(r) > 0 \right\} \quad \text{and} \quad \tilde{k}(r) := \tilde{k}_N(r) := \min \left\{ k : \tilde{Z}_k^N(r) > 0 \right\}. \quad (32)$$

Let t_N be a sequence in \mathbb{R}_+ with $\frac{t_N}{f(N) \ln N} \rightarrow \infty$ and $t_N = O(f(N)\theta_N)$. The three lemmata at the end of Section 6 together with Remark 6.2 show that for all $k \in \mathbb{N}_0$, as $N \rightarrow \infty$,

$$\frac{f(N)}{2\alpha N} Z_{k+\bar{k}(t_N)}^{(N)}(t_N) \rightarrow p_k \quad \text{in distribution.}$$

In view of (31) this implies that for all $k \in \mathbb{N}_0$, as $N \rightarrow \infty$,

$$\frac{f(N)}{2\alpha N} \tilde{Z}_{k+\tilde{k}(t_N)}^{(N)}(t_N) \rightarrow p_k \quad \text{in distribution.} \quad (33)$$

In analogy to (15) we define

$$K_N^{**}(t) := \min \{ \eta(j, t) : (j, t) \in \mathcal{A}_t^{N, \text{eq}} \} \left(= d_{\mathcal{M}}(\mathcal{A}_0^{N, \text{eq}}, \mathcal{A}_t^{N, \text{eq}}) = \min \{ \zeta_t(i, 0) : (i, 0) \in \mathcal{A}_0^{N, \text{eq}} \} \right).$$

The distributional identity (30) implies that for all $k \in \mathbb{N}_0$ and $t > 0$

$$\# \left\{ j \in [N] : (j, t) \in \mathcal{A}_t^{N, \text{eq}}, \eta(j, t) = K_N^{**}(t) + k \right\} \stackrel{d}{=} \tilde{Z}_{k+\tilde{k}(t)}^{(N)}(t).$$

Together with (33) this shows that for all $k \in \mathbb{N}_0$, as $N \rightarrow \infty$,

$$\frac{f(N)}{2\alpha N} \# \left\{ j \in [N] : (j, t_N) \in \mathcal{A}_{t_N}^{N, \text{eq}}, \eta(j, t_N) = K_N^{**}(t_N) + k \right\} \rightarrow p_k \quad \text{in distribution.} \quad (34)$$

Remark 7.1. The convergence assertion (34) comes close to a proof of Theorem 2.3a). Indeed, since $\mathcal{A}_{t_N}^{N, \text{eq}}$ is \mathcal{P}_{t_N} -measurable, the individuals in that set can be seen as a random sample of size close to $2\alpha N/f(N) \gg 1$ taken at time t_N . What is still missing to complete the proof of Theorem 2.3a) is to check that $\mathcal{A}_{t_N}^{N, \text{eq}}$ contains with high probability as $N \rightarrow \infty$ an individual that is of best type among all the individuals in $[N] \times \{t_N\}$, or in other words to prove that

$$\mathbb{P}(K_N^*(t_N) = K_N^{**}(t_N)) \rightarrow 1 \quad \text{as } N \rightarrow \infty. \quad (35)$$

This convergence will be proved in Lemma 9.5 and Remark 9, which thus completes our first proof of Theorem 2.3a).

8. QUICK MERGING ALONG THE ANCESTRAL SELECTION GRAPH

The main result of this section, which will be a key ingredient in the proofs of Proposition 9.2 and Theorem 2.2 as well as in our second proof of Theorem 2.3a), is an upper estimate for the time it takes for the merging of the sets of load k potential ancestors of two \mathcal{P}_T -measurable random sets \mathcal{J}_T^1 and \mathcal{J}_T^2 of $[N] \times \{T\}$, where T is a \mathcal{P} -stopping time. Roughly stated, this result (stated as Proposition 8.1) says that this merging happens with high probability as $N \rightarrow \infty$ within a time frame of order $f(N) \ln N$, provided only that the sets \mathcal{J}_T^1 and \mathcal{J}_T^2 are sufficiently large. With reference to Definition 4.6, we define the (*random*) *merging time* of the two load k ASG's $\mathcal{A}^{\mathcal{J}_T^1}(k)$ and $\mathcal{A}^{\mathcal{J}_T^2}(k)$ as

$$\mathcal{C}_k^{\mathcal{J}_T^1, \mathcal{J}_T^2} := \sup \left\{ s \leq T : \mathcal{A}_s^{\mathcal{J}_T^1}(k) = \mathcal{A}_s^{\mathcal{J}_T^2}(k) \right\}. \quad (36)$$

Proposition 8.1. *Let T be a \mathcal{P} -stopping time and let $\mathcal{J}_T^1, \mathcal{J}_T^2$ be \mathcal{P}_T -measurable random subsets of $[N] \times \{T\}$. Then, for any $k \geq 0$ and $\varepsilon > 0$, there exists a finite constant $C(\varepsilon)$ s.t.*

$$\liminf_{N \rightarrow \infty} \mathbb{P} \left(\mathcal{C}_k^{\mathcal{J}_T^1, \mathcal{J}_T^2} \geq T - C(\varepsilon) f(N) \ln N \mid \#\mathcal{J}_T^1 \geq 1/\varepsilon \text{ and } \#\mathcal{J}_T^2 \geq 1/\varepsilon \right) \geq 1 - \delta(\varepsilon) \quad (37)$$

with $\delta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

Proof. The strategy of the proof consists in showing by induction that for all $k \geq 0$ the sets

$$\mathcal{A}^{\mathcal{J}_T^1}(0) \cup \dots \cup \mathcal{A}^{\mathcal{J}_T^1}(k) \quad \text{and} \quad \mathcal{A}^{\mathcal{J}_T^2}(0) \cup \dots \cup \mathcal{A}^{\mathcal{J}_T^2}(k)$$

merge within a time of order $f(N) \ln N$. Let us begin with the case $k = 0$. For abbreviation we will write $s := T - r$, $r \geq 0$, and $\mathcal{A}_s^i := \mathcal{A}_s^{\mathcal{J}_T^i}(0)$, $i = 1, 2$

We will study the dynamics of the set-valued process

$$\mathcal{A}_{T-r}^1 \Delta \mathcal{A}_{T-r}^2 = (\mathcal{A}_{T-r}^1 \cup \mathcal{A}_{T-r}^2) \setminus (\mathcal{A}_{T-r}^1 \cap \mathcal{A}_{T-r}^2), \quad r \geq 0,$$

and of its cardinality $\#(\mathcal{A}_{T-r}^1 \Delta \mathcal{A}_{T-r}^2) = \#(\mathcal{A}_{T-r}^1 \cup \mathcal{A}_{T-r}^2) - \#(\mathcal{A}_{T-r}^1 \cap \mathcal{A}_{T-r}^2)$ as r increases. Four possible types of elements of the processes $(\mathcal{C}, \mathcal{S}, \mathcal{M})$ may have an impact on $\mathcal{A}_s^1 \Delta \mathcal{A}_s^2$:

- $(i, j, s) \in \mathcal{S}$ with $(j, s) \in \mathcal{A}_s^1 \Delta \mathcal{A}_s^2$ and $(i, s) \notin \mathcal{A}_s^1 \cup \mathcal{A}_s^2$; then

$$\mathcal{A}_{s-}^1 \Delta \mathcal{A}_{s-}^2 = \mathcal{A}_s^1 \Delta \mathcal{A}_s^2 \cup \{(j, s)\}.$$

The rate of this event is

$$q_1 := s_N \#(\mathcal{A}_s^1 \Delta \mathcal{A}_s^2) \left(1 - \frac{\#(\mathcal{A}_s^1 \cup \mathcal{A}_s^2)}{N} \right) \quad (38)$$

- $(i, j, s) \in \mathcal{C}$ with $(j, s) \in \mathcal{A}_s^1 \Delta \mathcal{A}_s^2$ and $(i, s) \in \mathcal{A}_s^1 \cap \mathcal{A}_s^2$; then

$$\mathcal{A}_{s-}^1 \Delta \mathcal{A}_{s-}^2 = \mathcal{A}_s^1 \Delta \mathcal{A}_s^2 \setminus \{(j, s)\}.$$

The rate of this event is

$$q_2 := \frac{1}{N} \#(\mathcal{A}_s^1 \Delta \mathcal{A}_s^2) \#(\mathcal{A}_s^1 \cap \mathcal{A}_s^2)$$

- $(i, j, s) \in \mathcal{C}$ with (j, s) and $(i, s) \in \mathcal{A}_s^1 \Delta \mathcal{A}_s^2$; then

$$\mathcal{A}_{s-}^1 \Delta \mathcal{A}_{s-}^2 = \mathcal{A}_s^1 \Delta \mathcal{A}_s^2 \setminus \{(j, s)\}.$$

The rate of this event is

$$q_3 := \frac{1}{2N} \#(\mathcal{A}_s^1 \Delta \mathcal{A}_s^2) (\#(\mathcal{A}_s^1 \Delta \mathcal{A}_s^2) - 1) \quad (39)$$

- $(i, s) \in \mathcal{M}$ with $(i, s) \in \mathcal{A}_s^1 \Delta \mathcal{A}_s^2$; then

$$\mathcal{A}_{s-}^1 \Delta \mathcal{A}_{s-}^2 = \mathcal{A}_s^1 \Delta \mathcal{A}_s^2 \setminus \{(i, s)\}.$$

The rate of this event is

$$q_4 = m_N \#(\mathcal{A}_s^1 \Delta \mathcal{A}_s^2). \quad (40)$$

The sum of q_2, q_3 and q_4 equals

$$\begin{aligned} q_2 + q_3 + q_4 &= \#(\mathcal{A}_s^1 \Delta \mathcal{A}_s^2) \left(m_N + \frac{\#(\mathcal{A}_s^1 \cup \mathcal{A}_s^2) + \#(\mathcal{A}_s^1 \cap \mathcal{A}_s^2) - 1}{2N} \right) \\ &= \#(\mathcal{A}_s^1 \Delta \mathcal{A}_s^2) \left(m_N + \frac{\#\mathcal{A}_s^1 + \#\mathcal{A}_s^2 - 1}{2N} \right). \end{aligned}$$

From Lemma 5.1, Remark 5.2 and Lemma 6.3 we know that for any $\varepsilon > 0$ there exists a constant $C(\varepsilon)$ such that if

$$\#\mathcal{A}_T^i \geq 1/\varepsilon, \quad i = 1, 2,$$

then for any $R > 0$, with a probability close to 1 for ε small enough and N large enough,

$$\frac{2N}{f(N)}(\alpha - \mu - 2\varepsilon) \leq \#\mathcal{A}_{T-r}^1, \quad \#\mathcal{A}_{T-r}^2 \leq \frac{2N}{f(N)}(\alpha + \mu - 2\varepsilon)$$

for $C(\varepsilon)f(N) \ln N \leq r \leq (C(\varepsilon) + R)f(N) \ln N$. In such a time window, we have

$$q_1 \leq \frac{\alpha}{f(N)} (\#\mathcal{A}_s^1 \Delta \mathcal{A}_s^2)$$

and

$$q_2 + q_3 + q_4 \geq \#\mathcal{A}_s^1 \Delta \mathcal{A}_s^2 \left(m_N + \frac{\#\mathcal{A}_s^1 + \#\mathcal{A}_s^2 - 1}{2N} \right) \geq \#\mathcal{A}_s^1 \Delta \mathcal{A}_s^2 \frac{\alpha + (\alpha - \mu - 5\varepsilon)}{f(N)}$$

for N large enough. The process $(\#\mathcal{A}_{T-r}^1 \Delta \mathcal{A}_{T-r}^2)_{r \geq 0}$ is thus stochastically dominated by a branching process with individual birth rate $\alpha/f(N)$ and death rate

$$(\alpha + (\alpha - \mu - 5\varepsilon)) / f(N).$$

The extinction time of such a process, with an initial state smaller than N , is smaller than

$$2 \frac{f(N) \ln N}{\alpha - \mu - 5\varepsilon}$$

with a probability converging to 1 when N goes to infinity (see e.g. [4] Lemma A.1). This concludes the proof of the proposition for the case $k = 0$.

Assume now that the sets

$$\mathcal{A}^{\mathcal{J}_T^1}(0) \cup \dots \cup \mathcal{A}^{\mathcal{J}_T^1}(k-1) \quad \text{and} \quad \mathcal{A}^{\mathcal{J}_T^2}(0) \cup \dots \cup \mathcal{A}^{\mathcal{J}_T^2}(k-1)$$

merge at time $T_{k-1} = T - R_{k-1}$, where $R_{k-1} = O(f(N) \ln N)$. From Lemma 6.3 we know that there exists $R < \infty$ such that for any $K < \infty$, with a probability close to one the size of this union is close to

$$\frac{N}{f(N)} \sum_{l=0}^{k-1} \bar{n}_l$$

during the time interval $[T_{k-1} + Rf(N) \ln N, T_{k-1} + (R + K)f(N) \ln N]$, and remains to be so during any time frame of order $f(N) \ln N$. We also know that the sizes of $\mathcal{A}^{\mathcal{J}_T^1}(k)$ and $\mathcal{A}^{\mathcal{J}_T^2}(k)$ are close to $N\bar{n}_k/f(N)$ during the same time frame. Let us again use the abbreviations $s := T - r$ and \mathcal{A}_s^1 and \mathcal{A}_s^2 , now for

$$\mathcal{A}_s^i := \mathcal{A}_s^{\mathcal{J}_T^i}(0) \cup \dots \cup \mathcal{A}_s^{\mathcal{J}_T^i}(k), \quad i = 1, 2.$$

By definition of T_{k-1} we have the equality

$$\# \left(\mathcal{A}_{T_{k-1}-r}^1 \Delta \mathcal{A}_{T_{k-1}-r}^2 \right) = \# \left(\mathcal{A}_{T_{k-1}-r}^{\mathcal{J}_T^1}(k) \Delta \mathcal{A}_{T_{k-1}-r}^{\mathcal{J}_T^2}(k) \right), \quad r \geq 0. \quad (41)$$

Another crucial observation is that the upward and downward jump rates of the process

$$\# \left(\mathcal{A}_{T_{k-1}-r}^1 \Delta \mathcal{A}_{T_{k-1}-r}^2 \right)_{r \geq 0}$$

are the same as those of the process

$$\# \left(\mathcal{A}_{T-r}^{\mathcal{J}_T^1} \triangle \mathcal{A}_{T-r}^{\mathcal{J}_T^2} \right)_{r \geq 0},$$

resulting from (38) – (40). (In particular, for $s \leq T_{k-1}$, the mutational events only affect the set $\mathcal{A}_s^{\mathcal{J}_T^1}(k) \triangle \mathcal{A}_s^{\mathcal{J}_T^2}(k)$, whose cardinality by (41) equals that of $\mathcal{A}_s^1 \triangle \mathcal{A}_s^2$.) The rest of the proof now follows that same lines as in the case $k = 0$. \square

Similarly as in (36), we define the (*random*) *merging time* of the ASG's $\mathcal{A}^{\mathcal{J}_T^1}$ and $\mathcal{A}^{\mathcal{J}_T^2}$ as

$$\mathcal{C}^{\mathcal{J}_T^1, \mathcal{J}_T^2} := \max \left\{ s \leq T : \mathcal{A}_s^{\mathcal{J}_T^1} = \mathcal{A}_s^{\mathcal{J}_T^2} \right\}.$$

Since in the special case $m_N = 0$ the load zero ASG $\mathcal{A}_s^{\mathcal{J}_T}(0)$ equals the ‘untyped’ ASG $\mathcal{A}_s^{\mathcal{J}}$, we immediately obtain the following corollary by putting $\mu = 0$ and $k = 0$ in Proposition 8.1:

Corollary 8.2. *Let $T, \mathcal{J}_T^1, \mathcal{J}_T^2$ be as in Proposition 8.1. Then for any $\varepsilon > 0$, there exists a finite constant $C(\varepsilon)$ such that (37) also holds for $\mathcal{C}^{\mathcal{J}_T^1, \mathcal{J}_T^2}$ in place of $\mathcal{C}_k^{\mathcal{J}_T^1, \mathcal{J}_T^2}$.*

9. CLICK TIMES ON THE ANCESTRAL SELECTION GRAPH. 1ST PROOF OF THEOREM 2.3A)

Definition 9.1. *For $N \in \mathbb{N}$, $t \in \mathbb{R}$ and $\ell = 0, 1, \dots$ we define (again partially suppressing N in the notation) the click times on the ASG back from $[N] \times \{u\}$ as follows*

$$\hat{T}_\ell^{N,u} := \min \{ s \leq u : d_{\mathcal{M}}(\mathcal{A}_s^u, [N] \times \{u\}) = \ell \}.$$

We thus get a point process

$$\mathcal{T}^{N,u} := \left\{ \hat{T}_\ell^{N,u}, \ell \in \mathbb{N}_0 \right\}.$$

The proof of the following proposition will be based on results in Sections 5, 6 and 8. Roughly stated, Proposition 9.2 says that the process of click times on the ASG, back from times that are large on the $f(N)\theta_N$ -scale, converges on that scale locally around time 0 to a standard Poisson process. This result is key for the proof of Theorem 2.2. Indeed, in Section 10 we will argue that the process of (forward) click times figuring in Theorem 2.2, which are represented as the jump times of (15), is locally on the $f(N)\theta_N$ -scale with high probability (as $N \rightarrow \infty$) close to the process figuring in Proposition 9.2. This latter process, however, can be read off from the ASG together with the points of \mathcal{M} . See Figure 3 and also Figure 4 for illustrations.

For later reference we will consider a sequence (u_N) of time points with the property

$$\frac{u_N}{\theta_N f(N)} \rightarrow \infty \quad \text{as } N \rightarrow \infty. \quad (42)$$

Proposition 9.2. *Let (u_N) obey (42). Let $(T_g^{N,u_N})_{1 \leq g \leq g_N}$ be the points contained in the set $\mathcal{T}^{N,u_N} \cap [0, u_N]$ ordered such that*

$$0 < T_1^{N,u_N} < \dots < T_{g_N}^{N,u_N} \leq u_N.$$

Putting $T_0^{N,u_N} := 0$, we have for $n \in \mathbb{N}$ the following convergence in distribution as $N \rightarrow \infty$:

$$\left(\frac{T_g^{N,u_N} - T_{g-1}^{N,u_N}}{f(N)\theta_N} \right)_{1 \leq g \leq n} \rightarrow (\mathcal{W}_g)_{1 \leq g \leq n},$$

where $(W_g)_{g \in \mathbb{N}}$ is a sequence of i.i.d. standard exponential random variables. Consequently, the sequence of processes \mathcal{N}^N defined by

$$\mathcal{N}_t^N := \sum_{g \geq 1} \mathbf{1}_{\{T_g^{N,tN} \leq f(N)\theta_N t\}}, \quad t \geq 0$$

converges as $N \rightarrow \infty$ to a standard Poisson process.

A large part of the remainder of this section is devoted to the proof of Proposition 9.2. For any $N \in \mathbb{N}$ and any $\mathcal{P}^{(N)}$ -stopping time T we define the $\mathcal{P}^{(N)}$ -stopping time $S^{N,T}$

$$S^{N,T} := \sup \{s \leq T : A_s^T(0) = 0\}. \quad (43)$$

In words, among all times at which all the potential ancestral paths of the population that lives at time T carry at least one mutation, the time $S^{N,T}$ is the one which is closest to T . Let us also note that for fixed N the distribution of $T - S^{N,T}$ does not depend on the choice of the $\mathcal{P}^{(N)}$ -stopping time T , cf. Lemma 5.1 and Remark 5.2. A key step in the proof of Proposition 9.2 is provided by

Lemma 9.3. *For any sequence of $\mathcal{P}^{(N)}$ -stopping times T_N the sequence*

$$\frac{T_N - S^{N,T_N}}{f(N)\theta_N}$$

converges in law to an exponential random variable with rate parameter 1.

Proof. The process $Z_0^{(N)}(r) := A_{T_N - r}^{T_N}(0)$, $r \geq 0$, has the jump rates (22) and starts in N . Lemma 6.3 shows that the quasi-equilibrium of $Z_0^{(N)}$ builds up within a time of order $f(N) \ln N$ when started in $Z_0^{(N)}(0) = N$. But notice that the assumption $f(N) = o(N/\ln \ln N)$ yields

$$\frac{f(N)\theta_N}{f(N)\ln N} \gg \frac{e^{\ln \ln N}}{\ln N} \gg 1.$$

The asymptotic exponentiality of $T_N - S^{N,T_N}$ with the claimed time scaling thus follows from Lemma 6.4. \square

Let us now consider a sequence of (deterministic) times u_N as in Proposition 9.2 and recall the definition of $S^{N,T}$ in (43). For each fixed $N \in \mathbb{N}$ define recursively

$$\begin{aligned} S_0^{N,u_N} &:= u_N, \\ S_\ell^{N,u_N} &:= S^{N,S_{\ell-1}^{N,u_N}}, \quad \ell = 1, 2, \dots \end{aligned} \quad (44)$$

The following corollary is now immediate from Lemma 9.3.

Corollary 9.4. *Let S_ℓ^{N,u_N} be defined by (44).*

a) The sequences

$$\left(\frac{S_{\ell-1}^{N,u_N} - S_\ell^{N,u_N}}{f(N)\theta_N} \right)_{\ell \geq 1}$$

converge as $N \rightarrow \infty$ in the sense of finite dimensional distributions to a sequence of i.i.d. standard exponential random variables.

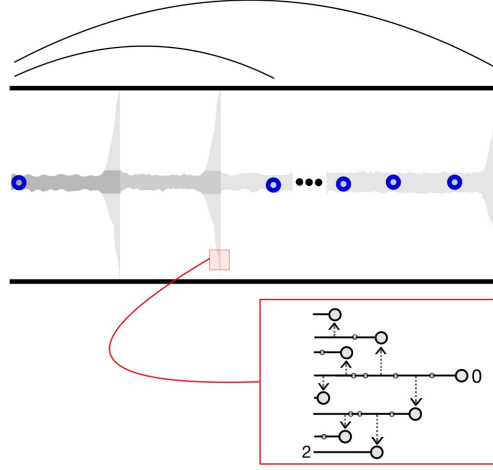


FIGURE 4. This cartoon reflects two main insights that help to prove Theorems 2.2 and 2.3. First, ASGs started from the entire population merge “quickly” (i.e. on the scale $f(N) \ln N$) into a “backbone” ASG (coloured in grey). Along the latter, clicks (depicted by thick circles) happen asymptotically at scale $\theta_N f_N$, and these can be coupled locally on that scale with the clicks forward in time. Second, most of the variability on the number of mutations among contemporaneous individuals comes from mutations acquired in their recent past. These recent mutations can be studied via duality by means of a Yule process approximation on the scale $f(N) \ln N$, leading to the quasi-stationary type frequency profile (p_k) .

b) Let C be an arbitrary positive constant. The sequence of point processes

$$\mathcal{S}^{N, u_N} := \left\{ \frac{S_\ell^{N, u_N}}{f(N)\theta_N} : \ell \in \mathbb{N}_0 \right\}, \quad N = 1, 2, \dots$$

converges, when restricted to $[0, C]$ in distribution to a standard Poisson point process restricted to $[0, C]$.

Proof of Proposition 9.2. From Definition 9.1 we recall the point process \mathcal{T}^{N, u_N} of click times on the ASG back from $[N] \times \{u_N\}$. The strategy of the proof will be to compare this process “locally on the $f(N)\theta_N$ -timescale” to the process \mathcal{S}^{N, u_N} which on that scale according to Corollary 9.4 is approximately Poisson.

To this purpose we define for each time $N \in \mathbb{N}$ and each time point $s > 0$

$$S^{N, u_N}(s) := \min(\mathcal{S}^{N, u_N} \cap [s, \infty)), \quad \hat{T}^{N, u_N}(s) := \min(\mathcal{T}^{N, u_N} \cap [s, \infty)),$$

Let $\bar{\mathcal{A}}_s^t$ be the set of minimum load potential ancestors at time s of the population at some (deterministic or random) time t , as specified in Definition 4.6. For abbreviation we put

$$\mathcal{B}_s^N := \bar{\mathcal{A}}_s^{S^{N, u_N}(s)}.$$

For any fixed $C > 0$ we abbreviate $s_N := Cf(N)\theta_N$. We will use the following properties:

- The process $(\#\bar{\mathcal{A}}_{u_N-r}^{u_N}, r \geq 0)$ follows the dynamics of $Z_{\bar{k}} = (Z_{\bar{k}(r)}(r))_{r \geq 0}$, where \bar{k} has been defined in (32). According to Lemmas 6.4 and 6.5

$$\liminf_{N \rightarrow \infty} \mathbb{P} \left(\#\bar{\mathcal{A}}_{s_N}^{u_N} \geq \frac{1}{\varepsilon} \right) = 1 + \delta(\varepsilon).$$

- A similar reasoning yields

$$\liminf_{N \rightarrow \infty} \mathbb{P} \left(\#\mathcal{B}_{s_N}^N \geq \frac{1}{\varepsilon} \right) = 1 + \delta(\varepsilon).$$

- Take a sequence v_N such that $f(N) \ln N \ll v_N \ll f(N)\theta_N$. Then according to part *a*) of Corollary 9.4,

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left([s_N - v_N, s_N] \cap f(N)\theta_N \mathcal{S}^{N, u_N} = \emptyset \right) = 1.$$

- Finally, according to Proposition 8.1 (on the quick merging of load zero ASG's),

$$\liminf_{N \rightarrow \infty} \mathbb{P} \left(\mathcal{C}_0^{\bar{\mathcal{A}}_{s_N}^{u_N}, \mathcal{B}_{s_N}^N} \geq s_N - v_N \mid \#\bar{\mathcal{A}}_{s_N}^{u_N} \geq 1/\varepsilon \text{ and } \#\mathcal{B}_{s_N}^N \geq 1/\varepsilon \right) \geq 1 - \delta(\varepsilon).$$

From these facts we deduce that

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\hat{T}^{N, u_N}(s_N) = S^{N, u_N}(s_N) \right) = 1.$$

We proceed in a similar way to cover the timeframe $[0, s_N]$, which contains a random number of points of \mathcal{S}^{N, u_N} that has a finite expectation. We thus add a sum of errors that converges to 0 as $N \rightarrow \infty$, which allows us to conclude the proof. \square

The following lemma says, roughly spoken, that at generic, suitably large times t_N the minimum load ASG (which is a backward in time construction) with high probability not only is appreciably large but also contains an individual whose type is best among the total population at time t_N . This lemma is a building block in the second proof of Theorem 2.3a) that will be carried out in Section 12. It will also help completing the first proof of this theorem, see the end of the current section.

Lemma 9.5. *Let (t_N) be as in (9) and (u_N) be such that $(u_N - t_N)/(\theta_N f(N)) \rightarrow \infty$ as $N \rightarrow \infty$. Let $\bar{\mathcal{A}}_{t_N}^{u_N}$ be the minimum load ASG at time t_N of the total population at time u_N , as specified in Definition 4.6. Then*

$$\mathbb{P} \left(\#\bar{\mathcal{A}}_{t_N}^{u_N} > \sqrt{N/f(N)}; \exists v \in \bar{\mathcal{A}}_{t_N}^{u_N} : \eta(v) = K_N^*(t_N) \right) \rightarrow 1 \quad \text{as } N \rightarrow \infty. \quad (45)$$

Proof. Let λ_N be such that $f(N) \ln N \ll \lambda_N \ll f(N)\theta_N \wedge t_N$. From Lemma 5.1 we know that the law of the process $(\#\bar{\mathcal{A}}_{u_N-r}^{u_N}, r \geq 0)$ is the same as the law of the process $(Z_{\bar{k}(r)}^{(N)}(r), r \geq 0)$ studied in Section 6 (defined in (32)), with initial state $(N, 0, \dots, 0, \dots)$. Thus from Lemma 6.4 we obtain that the distribution of $\#\bar{\mathcal{A}}_{u_N-s}^{u_N}$ at a given time $s \gg f(N)\theta_N$ is of order $N/f(N)$ with a probability close to one. We may thus apply Corollary 8.2 with $\mathcal{J}_{t_N}^1 := \bar{\mathcal{A}}_{t_N}^{u_N}$ and $\mathcal{J}_{t_N}^2 := [N] \times \{t_N\}$ to obtain

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\mathcal{C}^{\bar{\mathcal{A}}_{t_N}^{u_N}, [N] \times \{t_N\}} \geq t_N - \lambda_N \right) = 1. \quad (46)$$

Take an individual (i, t_N) belonging to the best class at time t_N , that is to say

$$\eta(i, t_N) = K_N^*(t_N) =: g^*. \quad (47)$$

On the event $\mathcal{E}_N^{(1)} := \{\mathcal{E}^{\bar{\mathcal{A}}_{t_N}^{u_N}, [N] \times \{t_N\}} \geq t_N - \lambda_N\}$ we have

$$\mathcal{A}_{t_N - \lambda_N}^{i, t_N} \subset \mathcal{A}_{t_N - \lambda_N}^{\bar{\mathcal{A}}_{t_N}^{u_N}} =: \mathcal{A}_N,$$

and consequently also

$$\eta(i, t_N) \geq \min \left\{ \eta(v) : v \in \mathcal{A}_{t_N - \lambda_N}^{i, t_N} \right\} \geq \min \left\{ \eta(v) : v \in \mathcal{A}_N \right\} = d_{\mathcal{M}} \left(\mathcal{A}_0^{\mathcal{A}_N}, \mathcal{A}_N \right).$$

Now consider the event that there is no click on the ASG between times $t_N - \lambda_N$ and t_N , i.e.

$$\mathcal{E}_N^{(2)} := \left\{ T_{g^*}^{N, u_N} \leq t_N - \lambda_N \leq t_N < T_{g^*+1}^{N, u_N} \right\}.$$

On this event we have

$$d_{\mathcal{M}} \left(\mathcal{A}_0^{\mathcal{A}_N}, \mathcal{A}_N \right) = d_{\mathcal{M}} \left(\mathcal{A}_0^{\bar{\mathcal{A}}_{t_N}^{u_N}}, \bar{\mathcal{A}}_{t_N}^{u_N} \right) = \min \left\{ \eta(v) : v \in \bar{\mathcal{A}}_{t_N}^{u_N} \right\} \geq K_N^*(t_N), \quad (48)$$

where the last equality and the last inequality hold by definition. The chain of (in-)equalities (47)–(48) shows that on the event $\mathcal{E}_N^{(1)} \cap \mathcal{E}_N^{(2)}$

$$\min \left\{ \eta(v) : v \in \bar{\mathcal{A}}_{t_N}^{u_N} \right\} = K_N^*(t_N).$$

Proposition 9.2 and (46) ensure that $\mathbb{P}(\mathcal{E}_N^{(1)} \cap \mathcal{E}_N^{(2)}) \rightarrow 1$ as $N \rightarrow \infty$, which ends the proof. \square

Completion of the first proof of Theorem 2.3a). Without loss of generality we may choose u_N in Lemma 9.5 so large that $\mathcal{A}_{t_N}^{u_N} = \mathcal{A}_{t_N}^{\text{eq}}$. In this way we obtain (35) as a consequence of (45); hence, as announced in Remark 7.1, Lemma 9.5 also concludes our first proof of Theorem 2.3a). \square

10. CLICK RATES: PROOF OF THEOREM 2.2

The next lemma relates the click times of the ratchet, defined as the jump times of the process K_N^* given by (15), to the times T_g^{N, u_N} obtained from the point process \mathcal{T}^{N, u_N} of backward click times, see Definition 9.1 and Proposition 9.2. As will become clear from the following proof, each time T_g^{N, u_N} with high probability 'announces' a click time of the ratchet, with waiting times whose lengths on the $f(N)\theta_N$ -scale tend to zero in probability as $N \rightarrow \infty$.

Lemma 10.1. *For any $t \geq 0$,*

$$\mathbb{P} \left(K_N^*(f(N)\theta_N t) = \sum_{g \geq 1} \mathbf{1}_{\{T_g^{N, u_N} \leq f(N)\theta_N t\}} \right) \rightarrow 1 \quad \text{as } N \rightarrow \infty. \quad (49)$$

Proof. Since $K_N^*(0) = 0$ and $T_g^{N, u_N} > 0$ a.s. for $g > 0$, we need only to consider the case $t > 0$. For abbreviation we put $t_N := f(N)\theta_N t$. Let us recall the definition (given in formula (16)) of the set $\bar{\mathcal{A}}_s^{u_N}$ of minimal load potential ancestors at time s of the total population at time u_N . Let g^* be such that

$$T_{g^*}^{N, u_N} \leq t_N < T_{g^*+1}^{N, u_N}$$

and take $(i, t_N) \in \bar{\mathcal{A}}_{t_N}^{u_N}$. Then by definition

$$K_N^*(t_N) \leq d_{\mathcal{M}} \left(\mathcal{A}_0^{i, t_N}, (i, t_N) \right) = g^*.$$

Now, notice that we have the following event inclusion:

$$\{K_N^*(t_N) < g^*\} \subset \left\{ \mathcal{A}_{T_{g^*}^{N,u_N}}^{u_N} \subsetneq \mathcal{A}_{T_{g^*}^{t_N,u_N}} \right\}.$$

But according to Proposition 9.2,

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left(\exists g \in \mathbb{N}, \frac{T_g^{N,u_N}}{f(N)\theta_N} \in [t - \varepsilon, t + \varepsilon] \right) = \delta(\varepsilon),$$

and according to Corollary 8.2, the merging time of \mathcal{A}^{u_N} and \mathcal{A}^{t_N} back from time t_N is $O(f(N) \ln N)$ with a probability close to one for large N . As by assumption $f(N) \ln N \ll f(N)\theta_N$, this concludes the proof of (49), and shows the lemma. \square

Proposition 9.2 then implies that, as $N \rightarrow \infty$, the sequence of processes $(K_N^*(f(N)\theta_N t))_{t \geq 0}$ converges in distribution to a rate 1 Poisson process. This is the assertion of Theorem (2.2).

11. FIRST PASSAGE PERCOLATION IN POISSON-DECORATED YULE TREES

In this section we consider a Yule tree \mathcal{Y} with splitting rate α , and regard \mathcal{Y} as the union of the (infinitely many) *lineages* \mathfrak{l} leading from the root to ∞ . Given \mathcal{Y} , let Π be a Poisson process on \mathcal{Y} whose intensity is μ times the length measure on \mathcal{Y} . (In Section 12 we will prove that these *Poisson-decorated Yule trees* indeed appear in the ASG as $N \rightarrow \infty$, see Figure 4 for an illustration). Again we assume $\mu < \alpha$ and define the *minimal Π -load in \mathcal{Y}* as

$$L := \min\{\Pi(\mathfrak{l}) : \mathfrak{l} \text{ is a lineage of } \mathcal{Y}\}. \quad (50)$$

In this section we will use the abbreviation

$$q := \frac{\mu}{\alpha + \mu}, \quad \rho := \frac{\mu}{\alpha} = \frac{q}{1 - q}. \quad (51)$$

Proposition 11.1. *a) L defined by (50) is an \mathbb{N}_0 -valued random variable satisfying*

$$\mathbb{P}(L > \ell) = \underbrace{\mathfrak{G} \circ \dots \circ \mathfrak{G}}_{\ell \text{ times}}(\rho), \quad \ell \in \mathbb{N}_0, \quad (52)$$

where ρ is as in (51) and

$$\mathfrak{G}(s) := \frac{1}{2} \left(1 + \rho - \sqrt{(1 + \rho)^2 - 4\rho s} \right), \quad 0 \leq s \leq 1. \quad (53)$$

b) The distribution weights

$$\pi_k := \mathbb{P}(L = k), \quad k \in \mathbb{N}_0,$$

satisfy the recursion (10), and thus are equal to the probability weights p_k , $k \in \mathbb{N}_0$ appearing in Theorem 2.3.

Proof. a) Consider a binary branching Galton-Watson tree \mathcal{G} with offspring distribution

$$\mathbb{P}(\text{no child}) = q = 1 - \mathbb{P}(\text{two children}).$$

Let \mathfrak{m}_1 be the number of leaves of \mathcal{G} . We put

$$g(s) := \mathbb{E}[s^{\mathfrak{m}_1} I_{\{\#\mathcal{G} < \infty\}}] = \begin{cases} \mathbb{E}[s^{\mathfrak{m}_1}], & 0 \leq s < 1 \\ \mathbb{P}(\#\mathcal{G} < \infty), & s = 1. \end{cases} \quad (54)$$

A first generation decomposition gives

$$g(s) = qs + (1 - q)g(s)^2, \quad 0 \leq s \leq 1.$$

From the two solutions of this equation only the function \mathfrak{G} given by (53) is admissible, since from (54) we have that $g(1) < 1$. Consequently, we have

$$\mathbb{E}[s^{\mathfrak{m}_1} I_{\{\#\mathcal{G} < \infty\}}] = \mathfrak{G}(s), \quad 0 \leq s \leq 1. \quad (55)$$

In particular, putting $s = 1$, we recover the extinction probability of the underlying Galton-Watson process as

$$\mathbb{P}(\#\mathcal{G} < \infty) = \mathfrak{G}(1) = \rho. \quad (56)$$

The Galton-Watson tree \mathcal{G} can be seen as embedded in the Yule tree \mathcal{Y} : when moving away from the root of \mathcal{Y} , every first encounter with a point of Π stands for a death in \mathcal{G} , while every splitting point of \mathcal{Y} stands for a birth in \mathcal{G} . Hence the event that \mathcal{G} is finite equals the event that there is no lineage \mathfrak{l} in \mathcal{Y} with $\Pi(\mathfrak{l}) = 0$, which in turn equals the event $\{L > 0\}$. In particular we obtain

$$\mathbb{P}(\#\mathcal{G} < \infty) = \mathbb{P}(L > 0).$$

Together with (56) this gives

$$\mathbb{P}(L > 0) = \rho, \quad (57)$$

which is (52) in the special case $\ell = 0$.

Exploring the lineages of \mathcal{Y} beyond the points of Π that are closest to the root of \mathcal{Y} , we encounter a self-similar situation: any such point can be seen as the root of an independent copy of \mathcal{G} , and the event $\{L > 1\}$ equals the event that all of these Galton-Watson trees are finite, which in view of (57), (54) and (55) has probability

$$\mathbb{P}(L > 1) = \mathbb{E}[\rho^{\mathfrak{m}_1}] = \mathfrak{G}(\rho).$$

This is (52) for $\ell = 1$. Proceeding further, $\{L > 2\}$ means that all of the \mathfrak{m}_1 many Poisson points are founders of lineages that carry more than one point of Π . This event has probability

$$\mathbb{P}(L > 2) = \mathbb{E}[(\mathfrak{G}(\rho))^{\mathfrak{m}_1}] = \mathfrak{G}(\mathfrak{G}(\rho)),$$

which is (52) for $\ell = 2$. Part a) of the proposition now follows by induction.

b) Let \mathfrak{e} be the edge that is between the root of \mathcal{Y} and its closest branch point. With q as in (51), the random variable $\Pi(\mathfrak{e})$ satisfies

$$\mathbb{P}(\Pi(\mathfrak{e}) \geq \ell) = q^\ell, \quad \ell \in \mathbb{N}_0. \quad (58)$$

The random variable L satisfies the stochastic fixed point equation

$$L \stackrel{d}{=} M + \min(L_1, L_2) \quad (59)$$

where L_1 and L_2 have the same distribution as L , M has distribution (58) and L, L_1, L_2, M are independent. We thus obtain

$$\mathbb{P}(M + \min(L_1, L_2) = k) = \sum_{i=0}^k \mathbb{P}(M = k - i) \mathbb{P}(\min(L_1, L_2) = i), \quad k \in \mathbb{N}_0. \quad (60)$$

From the independence of L_1, L_2 we have

$$\mathbb{P}(\min(L_1, L_2) = i) = \pi_i^2 + 2\pi_i \sum_{j>i} \pi_j =: w_i.$$

From (58) we have

$$\mathbb{P}(M = k - i) = q^{k-i}(1 - q).$$

Inserting this into (60) and observing (59) we obtain

$$\pi_k = \sum_{i=0}^k q^{k-i}(1 - q)w_i, \quad k \in \mathbb{N}_0.$$

Taking differences yields

$$\pi_k - \pi_{k-1} = (1 - q)w_k - \sum_{i=0}^{k-1} q^{k-1-i}(1 - q)^2 w_i = (1 - q)(w_k - \pi_{k-1}).$$

Observing that $(1 - q)(1 + \rho) = 1$ we arrive at

$$(1 + \rho)(\pi_k - \pi_{k-1}) = \pi_k \left(\pi_k + 2 \sum_{j>k} \pi_j \right) - \pi_{k-1} \quad (61)$$

which is equivalent to

$$\rho(\pi_k - \pi_{k-1}) = \pi_k \left(\pi_k - 1 + 2 \sum_{j>k} \pi_j \right). \quad (62)$$

Since

$$\pi_k - 1 + 2 \sum_{j>k} \pi_j = -\pi_k + 1 - 2 + 2\pi_k + 2 \sum_{j>k} \pi_j = -\pi_k + 1 - 2 \sum_{j<k} \pi_j,$$

we see that (π_k) satisfies the recursion (10). \square

Remark 11.2. a) From (53) we have

$$\mathfrak{G}'(s) = \rho \left((1 + \rho)^2 - 4\rho s \right)^{-1/2}, \quad 0 \leq s \leq 1$$

and hence $\mathfrak{G}'(0) = \frac{\rho}{1+\rho} = q = \frac{\mu}{\mu+\alpha}$. In view of (52) this shows that L has the geometric tails

$$\mathbb{P}(L > \ell) \sim C_{\mu/\alpha} \left(\frac{\mu}{\mu + \alpha} \right)^\ell. \quad (63)$$

b) The setting of Proposition 11.1 gives an instance of Example 40 in [1]: our stochastic fixed point equation(59) corresponds to Eq. (49) in [1] with a geometrically distributed “toll” random variable η . Thus, the results of Proposition 11.1 apply to a specific case of a situation which, according to [1], “does not seem to have been studied generally”. As stated in Theorem 2.3d) and explained in Section 13, this connects to the asymptotic minimum of a branching random walk whose increment distribution is supported on \mathbb{R}_+ . (See [13] and references therein for the asymptotics of minima of random walks with two-sided increment distributions.)

c) Let \mathscr{Y} be the Yule tree described at the beginning of the section. For a node $v \in \mathscr{Y}$, let $\mathbf{a}(v)$ be the path from v to the root, and for $h > 0$, $k \in \mathbb{N}_0$ let $\mathscr{Y}_h(k)$ be the set of nodes of \mathscr{Y} that have height h and obey $\Pi(\mathbf{a}(v)) = k$. Let L be as in (50). The following lemma says that the minimal Π -load of the (infinite) lineages in \mathscr{Y} can with high probability be observed already at a height w_N which is large but of smaller order than $\ln N$ as $N \rightarrow \infty$; moreover at this height there are many nodes of the Yule tree whose ancestral paths collect this load.

Finally, we define, as an analogue to (50), the *minimal Π -load in \mathcal{Y} up to height h* as

$$L_h := \min\{\Pi(\mathbf{a}(v)) : v \in \mathcal{Y}_h\}. \quad (64)$$

Lemma 11.3. *Let $w_N \rightarrow \infty$ with $w_N = o(\ln N)$ as $N \rightarrow \infty$. Then for all $k \in \mathbb{N}_0$,*

$$\mathbb{P}\left(L_{w_N} = k, e^{(\alpha-\mu)w_N/2} \leq \#\mathcal{Y}_{w_N}(k) \leq e^{2(\alpha-\mu)w_N}\right) \rightarrow p_k \text{ as } N \rightarrow \infty. \quad (65)$$

Proof. As described in the proof of Proposition 11.1, an equivalent representation of a binary branching Galton-Watson tree with mutation at rate μ is a sequence of trees of different types killed at rate μ . Descendants of the root are of type 0. Every death of an individual of type 0 (respectively 1, 2, ...) leads to a new binary branching Galton-Watson tree of type 1 (respectively 2, 3, ...). Let us consider the event $\{L_{w_N} = k\}$. This event implies that all the trees of types $l \leq k-1$ are extinct at time w_N . Every such tree is a supercritical tree with birth rate α and death rate μ . Conditioned on extinction, it is thus a subcritical tree with birth rate μ and death rate α . A first step decomposition shows that the expected number of leaves of this subcritical tree is $\frac{\rho}{1-\rho} = \frac{\mu}{\alpha-\mu} < \infty$. In particular, this subcritical tree has a finite mean extinction time. Hence

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\#\mathcal{Y}_{w_N^{1/2}}(l) = 0, 0 \leq l \leq k-1 | L_{w_N} = k\right) = 1.$$

Moreover, by definition, any tree of type k still alive at time w_N is born before the death of the last alive type $k-1$ individual. Let us denote by $\widehat{\mathcal{Y}}$ a binary Galton-Watson tree of type k with birth rate α and death rate μ . On the event of survival (see for instance [2] p.112),

$$\lim_{t \rightarrow \infty} (\ln \widehat{\mathcal{Y}}_t)/t = \alpha - \mu.$$

On the event

$$\mathfrak{E}_{k,N} := \left\{ \#\mathcal{Y}_{w_N^{1/2}}(l) = 0, 0 \leq l \leq k-1 \right\} \cap \{L_{w_N} = k\}$$

we know that

- There is a finite mean number of independent copies of $\widehat{\mathcal{Y}}$ and a positive number of them survive after time w_N which goes to infinity with N .
- These independent copies have a root born between the times 0 and $w_N^{1/2}$.

We deduce that

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(e^{(\alpha-\mu)w_N/2} \leq \#\mathcal{Y}_{w_N}(k) \leq e^{2(\alpha-\mu)w_N} | \mathfrak{E}_{k,N}\right) = 1.$$

Finally, notice that from properties of supercritical Galton-Watson processes,

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\#\mathcal{Y}_\infty(k) \geq 1 | \#\mathcal{Y}_{w_N}(k) \geq e^{(\alpha-\mu)w_N/2}\right) = 1.$$

We thus obtain

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(L \leq k | L_{w_N} = k, \#\mathcal{Y}_{w_N}(k) \geq e^{(\alpha-\mu)w_N/2}\right) = 1.$$

But by definition, $L_{w_N} \leq L$, which yields

$$\mathbb{P}\left(L_{w_N} = k, e^{(\alpha-\mu)w_N/2} \leq \#\mathcal{Y}_{w_N}(k) \leq e^{2(\alpha-\mu)w_N}\right) \rightarrow \mathbb{P}(L = k) \text{ as } N \rightarrow \infty.$$

An application of Proposition 11.1 ends the proof. \square

Lemma 11.4. *Let (w_N) be as in Lemma 11.3, and assume that the splitting rate $\alpha_N(h)$ and the decoration rate $\mu_N(h)$ may depend on N and h such that, uniformly in $h \in [0, w_N]$,*

$$\lim_{N \rightarrow \infty} \alpha_N(h) = \alpha \text{ and } \lim_{N \rightarrow \infty} \mu_N(h) = \mu. \quad (66)$$

Let $\mathcal{Y}^{(N)}$ be the corresponding $\Pi^{(N)}$ -decorated Yule tree, grown up to the height w_N . For $0 \leq h \leq w_N$, define in analogy to (64)

$$L_h^{(N)} := \min\{\Pi^{(N)}(\mathbf{a}(v)) : v \in \mathcal{Y}_h^{(N)}\}.$$

Then, for all $k \in \mathbb{N}$,

$$\pi_k^{(N)} := \mathbb{P}(L_{w_N}^{(N)} = k) \rightarrow \pi_k \text{ as } N \rightarrow \infty,$$

where (π_k) is as in Proposition 11.1. Moreover, the analogue of (65) holds for $L^{(N)}$ and $\mathcal{Y}^{(N)}$ instead of L and \mathcal{Y} .

Proof. All the previous quantities (probabilities, mean numbers, growth rates and expected times) are continuous functions of the parameters α and μ . Sandwiching arguments thus allow to extend the proof of the previous lemma. \square

12. FIRST PASSAGE PERCOLATION WITHIN THE ASG. 2ND PROOF OF THEOREM 2.3A)

In this section we will give a proof of Theorem 2.3a) along the program laid out in Figure 4. This program has two parts. The first one says that the \mathcal{M} -decorated ASG's of single individuals look at the timescale $f(N)$ like the Poisson-decorated Yule processes studied in Section 11. We will prove this in Lemma 12.3, together with the fact that on this timescale (and slightly beyond it) the ASG's of finitely many individuals are asymptotically independent. Roughly stated, Lemma 12.3 says that (with w_N as in Lemma 11.3) the minimal-load potential ancestors at time $t_N - w_N f(N)$ of an individual (i, t_N) are numerous and that the minimal load $L_{w_N f(N)}$ acquired over the time span $w_N f(N)$ has asymptotically as $N \rightarrow \infty$ the distribution $(p_k)_{k \in \mathbb{N}_0}$ given by the recursion (10), which we encountered also in Proposition 11.1. The second part of the program announced in Figure 4 can be stated as the fact that a suitably large set of individuals that live at a generically late time s_N , contains with high probability an individual that is of the best type among all individuals living at time s_N (or in other words, has type $K_N^*(s_N)$). This will be proved in Lemma 12.1, and will be applied to the set of minimum-load potential ancestors at time $s_N := t_N - w_N f(N)$ of an individual (i, t_N) , showing that this individual's type $\eta^{(N)}(i, t_N)$ is indeed the sum of the best type $K_N^*(s_N)$ in the total population at time s_N and the minimal load $L_{w_N f(N)}$ acquired by the potential ancestry of (i, t_N) between times s_N and t_N .

Lemma 12.1. *Let (t_N) obey (9). For $\varepsilon > 0$ let $\mathcal{J}_{t_N}^{(N)}$ be a sequence of \mathcal{P}_{t_N} -measurable subsets of $[N] \times \{t_N\}$ with $\lim_{N \rightarrow \infty} \mathbb{P}(\#\mathcal{J}_{t_N}^{(N)} \geq 1/\varepsilon) = 1$. Then*

$$\liminf_{N \rightarrow \infty} \mathbb{P}\left(\min\left\{\eta^{(N)}(v) : v \in \mathcal{J}_{t_N}^{(N)}\right\} = K_N^*(t_N)\right) = 1 - \delta(\varepsilon),$$

with $\delta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

Remark 12.2. Due to the a.s. independence of \mathcal{P}_{t_N} and \mathcal{F}_{t_N} , the type distribution within the set $\mathcal{J}_{t_N}^{(N)}$ is the same as that within a random sample drawn from the population at time t_N . Theorem 2.3a) asserts that the frequency of best type at time t_N is, as $N \rightarrow \infty$, with high probability close to $p_0 = 1 - \mu/\alpha > 0$. Hence Lemma 12.1 is in fact a consequence of

Theorem 2.3a), for which we have provided a first proof in Section 9, using results of Section 7. Our aim here is, however, to give yet another conceptual proof of Theorem 2.3a) which (in contrast to the first proof), does not use the time-reversal arguments of Section 7. Hence our proof of Lemma 12.1, which prepares for the second proof of Theorem 2.3a), will build on Proposition 8.1 (which guarantees the “quick merging” of zero-load ASG’s), combined with Lemma 9.5.

Proof of Lemma 12.1. Let (u_N) be such that $(u_N - t_N)/(\theta_N f(N)) \rightarrow \infty$ as $N \rightarrow \infty$. Recall the notation of the merging time of two sets in (36) and introduce for brevity the notation

$$\mathcal{T}_N := \mathcal{C}_0^{\mathcal{J}_{t_N}^{(N)}, \bar{\mathcal{A}}_{t_N}^{u_N}}$$

for the merging time of zero-load ASG’s of $\mathcal{J}_{t_N}^{(N)}$ and $\bar{\mathcal{A}}_{t_N}^{u_N}$. Since Lemma 9.5 guarantees that $\bar{\mathcal{A}}_{t_N}^{u_N}$ is sufficiently large with high probability, an application of Proposition 8.1 yields

$$\liminf_{N \rightarrow \infty} \mathbb{P}(\mathcal{T}_N \geq s_N - C(\varepsilon)f(N) \ln N) \geq 1 - \delta(\varepsilon).$$

On the event that there is no click on \mathcal{A}^{u_N} between times t_N and \mathcal{T}_N , the minimum load ASG $\bar{\mathcal{A}}^{u_N}$ does not acquire additional mutations between those times, hence we have on that event the equality

$$\mathcal{A}_{\mathcal{T}_N}^{\bar{\mathcal{A}}_{t_N}^{u_N}}(0) = \bar{\mathcal{A}}_{\mathcal{T}_N}^{u_N}.$$

One more application of Lemma 9.5, now to the time \mathcal{T}_N in place of t_N , implies

$$\mathbb{P}\left(\exists v \in \bar{\mathcal{A}}_{\mathcal{T}_N}^{u_N} : \eta^{(N)}(v) = K_N^*(\mathcal{T}_N)\right) \rightarrow 1 \quad \text{as } N \rightarrow \infty.$$

Because of the definition of \mathcal{T}_N , the individual v is a load zero potential ancestor of some $v^* \in \mathcal{J}_{t_N}^{(N)}$. Consequently, with probability tending to 1 as $N \rightarrow \infty$,

$$K_N^*(\mathcal{T}_N) = \eta^{(N)}(v) = \eta^{(N)}(v^*) = K_N^*(t_N),$$

which ends the proof. \square

For the next lemma let us define, as an analogue to (64), the *minimal load of potential ancestors at time $t - r$ of an individual $(i, t) \in G^{(N)}$* as

$$L_r^{i,t} := d_{\mathcal{M}}\left(\mathcal{A}_{t-r}^{i,t}, (i, t)\right). \quad (67)$$

Lemma 12.3. *Let t_N be as in (9) and w_N be as in Lemma 11.3, i.e. $w_N \rightarrow \infty$ and $w_N = o(\ln N)$ as $N \rightarrow \infty$. Choose a sample size $n \in \mathbb{N}$, let $i_1, \dots, i_n \in \mathbb{N}$ be pairwise distinct, and $k_1, \dots, k_n \in \mathbb{N}_0$. Then, for all $\varepsilon > 0$,*

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(L_{f(N)w_N}^{i_\ell, t_N} - k_\ell, A_{t_N - f(N)w_N}^{i_\ell, t_N}(k_\ell) \geq 1/\varepsilon, 1 \leq \ell \leq n\right) = p_{k_1} \cdots p_{k_n}.$$

Proof. First notice that if we take independent Yule trees $(\mathcal{Y}^{(\ell)}, 1 \leq \ell \leq n)$ as in Lemma 11.3 and denote their minimal Π -load accordingly, we get from an application of Lemma 11.3,

$$\mathbb{P}\left(L_{w_N}^{(\ell)} = k_\ell, e^{(\alpha - \mu)w_N/2} \leq \#\mathcal{Y}_{w_N}^{(\ell)}(k_\ell) \leq e^{2(\alpha - \mu)w_N}, 1 \leq \ell \leq n\right) \rightarrow p_{k_1} \cdots p_{k_n} \text{ as } N \rightarrow \infty.$$

According to Lemma 11.4, this result still holds true if the splitting rate $\alpha_N(h)$ and the decoration rate $\mu_N(h)$ of the Yule trees $(\mathcal{Y}^{(\ell)}, 1 \leq \ell \leq n)$ may depend on N and h in such a way that (66) is fulfilled. Hence, to prove the lemma we need to show two properties on the processes $(A_{t_N - r}^{i_\ell, t_N}, r \geq 0)$

- The rates of the processes $(A_{t_N - hf(N)}^{i_\ell, t_N}, h \geq 0)$ indeed follow (66)
- They are asymptotically independent in the following sense

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\bigcap_{1 \leq \ell \leq n} \mathcal{A}_{t_N - r}^{i_\ell, t_N} = \emptyset, \forall r \leq f(N)w_N \right) = 1.$$

Recall the rates of the processes $(A_{t_N - r}^{i_\ell, t_N}, r \geq 0)$ as stated in Lemma 5.1. In particular, they imply that these processes are dominated by a Yule process with birth rate $\alpha/f(N)$. Hence, if we introduce the event

$$\mathcal{E}^{(N, \text{notbig})} := \left\{ \sup_{r \leq f(N)w_N, 1 \leq \ell \leq n} A_{t_N - r}^{i_\ell, t_N} \leq e^{2\alpha w_N} \right\},$$

and apply Lemma A.1 in [6], we get

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\mathcal{E}^{(N, \text{notbig})} \right) = 1. \quad (68)$$

Recall r_3 in (39) which describes the rate at which two currently disjoint ASG's acquire a common point. On the event $\mathcal{E}^{(N, \text{notbig})}$, the mean number of such events during the time $[0, f(N)w_N]$ for two processes among $(\mathcal{A}_{t_N - r}^{i_\ell, t}, 1 \leq \ell \leq n)$ is thus bounded by

$$\frac{1}{2N} (2e^{2\alpha w_N}) (2e^{2\alpha w_N} - 1) w_N,$$

which converges to 0 as $N \rightarrow \infty$. Applying the Markov inequality and (68) ends the proof. \square

Proposition 12.4. *Let the type configurations $\eta^{(N)}(i, t)$, $i \in [N]$, $t \geq 0$, be as in Section 3, with $\eta^{(N)}(i, 0) := 0$, and let $K_N^*(t)$ be as in (15). Let (t_N) be a sequence of time points which obeys (9). Choose a sample size $n \in \mathbb{N}$, let $i_1, \dots, i_n \in \mathbb{N}$ be pairwise distinct, and $k_1, \dots, k_n \in \mathbb{N}_0$. Then, with $(p_k)_{k \in \mathbb{N}_0}$ given by the recursion (10),*

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\eta^{(N)}(i_1, t_N) - K_N^*(t_N) = k_1, \dots, \eta^{(N)}(i_n, t_N) - K_N^*(t_N) = k_n \right) = p_{k_1} \cdots p_{k_n}.$$

Proof. Let w_N be as in Lemma 12.3, and $s_N := t_N - w_N f(N)$. Once again we let (u_N) satisfy $(u_N - t_N)/(\theta_N f(N)) \rightarrow \infty$ as $N \rightarrow \infty$. Applying Lemma 12.1 to the set $\bar{\mathcal{A}}_{s_N}^{i_\ell, t_N}$ of minimum load potential ancestors at time s_N of the individual (i_ℓ, t_N) (see Definition 4.6) yields

$$\mathbb{P} \left(\min\{\eta^{(N)}(v) : v \in \bar{\mathcal{A}}_{s_N}^{i_\ell, t_N}\} = K_N^*(s_N) \right) \rightarrow 1 \quad \text{as } N \rightarrow \infty.$$

From Theorem 2.2 we know that

$$\mathbb{P}(K_N^*(t_N) = K_N^*(s_N)) \rightarrow 1 \quad \text{as } N \rightarrow \infty.$$

Hence we conclude, using the notation (67),

$$\mathbb{P} \left(\eta^{(N)}(i_\ell, t_N) = K_N^*(t_N) + L_{w_N f(N)}^{i_\ell, t_N} \right) \rightarrow 1 \quad \text{as } N \rightarrow \infty.$$

An application of Lemma 12.3 thus concludes the proof. \square

In accordance with the graphical representations of $\xi^{(N)}(t)$ and $K_N^*(t)$ in (8) and (13), the empirical type frequency profile seen from the currently best type (defined in (12)) is represented as

$$X_k^{(N)}(t) = \sum_{i=1}^N \delta_{\eta^{(N)}(i, t) - K_N^*(t)}, \quad k \in \mathbb{N}_0, t \geq 0.$$

The following corollary concludes our second proof of Theorem 2.3 a).

Corollary 12.5. *For (t_N) as in Proposition 12.4 and all $k \in \mathbb{N}_0$, $X_k^{(N)}(t_N)$ converges in probability as $N \rightarrow \infty$ to p_k , with $(p_k)_{k \in \mathbb{N}_0}$ given by (10).*

Proof. For all $k \in \mathbb{N}_0$, the second moment $\mathbb{E}[X_k^{(N)}(t_N)^2]$ is asymptotically equal to the probability that, for J_1, J_2 randomly sampled from $[N]$, the types $\eta^{(N)}(J_1, t_N) - K_N^*(t_N)$ and $\eta^{(N)}(J_2, t_N) - K_N^*(t_N)$ both are equal to k . Proposition 12.4 tells that this probability converges to p_k^2 as $N \rightarrow \infty$. By the same proposition, p_k^2 is the limit of $\mathbb{E}[X_k^{(N)}(t_N)]^2$ as $N \rightarrow \infty$, hence the variance of $X_k^{(N)}(t_N)$ vanishes as $N \rightarrow \infty$. \square

13. THE QUASI-STATIONARY TYPE FREQUENCY PROFILE: PROOF OF THEOREM 2.3B)-D)

Part b). Choosing $k = 0$ in (11) and using the assumption that $p_0 + p_1 + \dots = 1$, we see that (11) implies $p_0 = 1 - \frac{\mu}{\alpha}$. The equivalence of (10) and (11) is then immediate from the identity

$$-p_k + 1 - 2 \sum_{k'=0}^{k-1} p_{k'} = \sum_{k'=k+1}^{\infty} p_{k'} - \sum_{k'=0}^{k-1} p_{k'}.$$

Part c). Abbreviating $1 - \mu/\alpha = \beta$ we get from (10):

$$p_0 = \beta, \quad p_1 = \sqrt{\left(\frac{\beta}{2}\right)^2 + \beta(1-\beta)} - \frac{\beta}{2}.$$

Therefore, $p_0^2 \geq p_1^2 \iff 9\beta^2 \geq \beta^2 + 4\beta(1-\beta) \iff 2\beta \geq 1 - \beta \iff \beta \leq 1/3$, with equality iff $\beta = 1/3$. This proves the assertions (i) and (ii). For checking (iii), we define

$$\mathfrak{k} := \max \left\{ k \in \mathbb{N}_0 : \sum_{k' < k} p_{k'} \leq \sum_{k' > k} p_{k'} \right\}.$$

For $k := \mathfrak{k} + 1$, the l.h.s. of (11) is strictly negative, hence $p_{\mathfrak{k}+1} < p_{\mathfrak{k}}$. Since the l.h.s. of (11) is strictly decreasing in k and thus can be zero for at most one k , it must be strictly positive for $k := \mathfrak{k} - 1$, hence, again because of (11) we have $p_{\mathfrak{k}-2} < p_{\mathfrak{k}-1}$.

The claim concerning the geometric tail of (p_k) follows immediately from (63) combined with Proposition 11.1b).

Part d). The characterisation of $(p_k)_{k \in \mathbb{N}_0}$ in terms of minimal Poisson loads of infinite lineages in a Yule tree has been proved in Proposition 11.1 b).

To see the equivalence to the characterisation via the eventual minimum in a branching one-sided random walk, we think of the latter as a family of random walks, indexed by the infinite lineages \mathfrak{l} of a Yule tree \mathcal{T} with branching rate α . All the random walkers move on \mathbb{N}_0 , starting at 0 and jumping from k to $k+1$ at rate μ . The (correlated) dynamics of the walkers can thus be seen as driven by a Poisson point process Π with rate μ on \mathcal{T} : each point of Π induces an upwards jump by one, and the (continuous) time of the walks corresponds to the height in \mathcal{T} . Thus the position at time t of the walker that is indexed by an infinite lineage \mathfrak{l} of \mathcal{T} is the number of Poisson points which \mathfrak{l} carries between heights 0 and t . Denoting by $M(t)$ the minimum of the position of all the walkers alive at time t , we see that $M(t)$ increases to the $\mathbb{N}_0 \cup \{\infty\}$ -valued random variable $K := \min\{\Pi(\mathfrak{l}) : \mathfrak{l} \in \mathcal{T}\}$, i.e. the minimum over the numbers of Poisson points carried by the infinite lineages in \mathcal{T} .

This concludes the proof of Theorem 2.3.

14. PROOF OF LEMMATA 6.3, 6.4 AND 6.5

This section is dedicated to the study of the process $(Z^{(N)}(r), r \geq 0)$. The proof of Lemma 6.3 relies essentially on the fact that a stochastic Lotka-Volterra process with large carrying capacity \bar{K} resembles a supercritical process when its size is small and once close to its carrying capacity, stays in a neighborhood of this latter during any time of order $\ln \bar{K}$. This last property is stated in Lemma C.1 in [4], which will be instrumental in the following proof.

Proof of Lemma 6.3. We will prove this result by induction. Let us first consider the case $k = 0$ and introduce, for $\varepsilon \in (0, \alpha - \mu)$, the notation

$$\mathfrak{c}(\alpha, \mu, \varepsilon) := 2(\alpha - \mu - \varepsilon).$$

Notice that for N large enough and $n \leq \mathfrak{c}(\alpha, \mu, \varepsilon)N/f(N)$, the birth and death rates defined in (6.1) obey

$$b_0(n, N) \geq \frac{\alpha}{f(N)} \left(1 - \frac{\varepsilon}{4\alpha}\right) \quad \text{and} \quad d_0(n, N) \leq \frac{\alpha}{f(N)} \left(1 - \frac{\varepsilon}{2\alpha}\right).$$

Thus, if $Z_0(0) \leq \mathfrak{c}(\alpha, \mu, \varepsilon)N/f(N)$, before its hitting time of $\mathfrak{c}(\alpha, \mu, \varepsilon)N/f(N)$, Z_0 stochastically dominates a supercritical branching process with growth rate

$$\frac{\alpha}{f(N)} \frac{\varepsilon}{4\alpha} = \frac{\varepsilon}{4f(N)}$$

and initial state $1/\varepsilon$. From well known results on supercritical branching processes (see for instance [2]), and since $\mathfrak{c}(\alpha, \mu, \varepsilon) < 2(\alpha - \mu)$, we get that

$$\liminf_{N \rightarrow \infty} \mathbb{P} \left(\inf \left\{ r \geq 0, Z_0(r) \geq \mathfrak{c}(\alpha, \mu, \varepsilon) \frac{N}{f(N)} \right\} \leq \frac{6}{\varepsilon} f(N) \ln N \mid Z_0 \geq 1/\varepsilon \right) = 1 - \delta(\varepsilon),$$

where $\delta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. With similar computations, we obtain that

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\inf \left\{ r \geq 0, Z_0(r) \leq \mathfrak{c}(\alpha, \mu, -\varepsilon) \frac{N}{f(N)} \right\} \leq \frac{3}{\varepsilon} f(N) \ln N \right) = 1.$$

By definition, the process Z_0 cannot exceed N . As a consequence, for any $r \geq 0$,

$$b_0(Z_0(r), N) \leq \frac{\alpha}{f(N)} \quad \text{and} \quad d_0(Z_0(r), N) \geq \frac{\alpha}{f(N)} \frac{\mu}{\alpha} = \frac{\mu}{f(N)}.$$

Thus, applying Lemma C.1 in [4], we obtain that for any $R < \infty$,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\sup_{r \leq Rf(N) \ln N} \{Z_0(r)\} \leq \varepsilon N \mid Z_0(0) \leq \mathfrak{c}(\alpha, \mu, -\varepsilon) \frac{N}{f(N)} \right) = 1.$$

Moreover, as long as $Z_0(r) \leq \varepsilon N$, the per capita birth rate of Z_0 satisfies:

$$\frac{\alpha}{f(N)}(1 - \varepsilon) \leq b_0(Z_0(r), N) \leq \frac{\alpha}{f(N)},$$

and since $-\varepsilon/f(N) + 1/2N < 0$ for large N , then the per capita death rate of Z_0 may be bounded as follows

$$\frac{\alpha}{f(N)} \left(\frac{\mu - \varepsilon}{\alpha} + \frac{Z_0(r)f(N)}{2\alpha N} \right) \leq d_0(Z_0(r), N) \leq \frac{\alpha}{f(N)} \left(\frac{\mu}{\alpha} + \frac{Z_0(r)f(N)}{2\alpha N} \right).$$

Applying again Lemma C.1 in [4], we obtain that for any $R < \infty$,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\frac{f(N)}{N} Z_0(r) \in [\mathfrak{e}(\alpha, \mu, 2\varepsilon), \mathfrak{e}(\alpha, \mu, -2\varepsilon)] \text{ for all } r \leq Rf(N) \ln N \right. \\ \left. \left| \frac{f(N)}{N} Z_0(0) \in [\mathfrak{e}(\alpha, \mu, \varepsilon), \mathfrak{e}(\alpha, \mu, -\varepsilon)] \right. \right) = 1.$$

This proves the lemma for $k = 0$ with $C_0(\varepsilon) = 6/\varepsilon$ and $C_0 = 4\varepsilon$.

Let us now take $g \in \mathbb{N}$ and assume that (27) holds true for $k = 0, \dots, g-1$. Jointly for all these $k = 0, \dots, g-1$ we can take a time frame (which may be as long as we want on the $f(N) \ln N$ -timescale) on which $f(N)Z_k/N \in [\bar{n}_k - C_k\varepsilon, \bar{n}_k + C_k\varepsilon]$, and $Z_g \leq 4\alpha N/f(N)$. During this time interval, the birth rate of the Z_g -population is larger than

$$\frac{\mu}{f(N)} (\bar{n}_{g-1} - C_{g-1}\varepsilon) \frac{N}{f(N)} + Z_g \frac{\alpha}{f(N)} \left(1 - \frac{4\alpha}{f(N)} \right)$$

and smaller than

$$\frac{\mu}{f(N)} (\bar{n}_{g-1} + C_{g-1}\varepsilon) \frac{N}{f(N)} + Z_g \frac{\alpha}{f(N)}.$$

Likewise, the death rate is larger than

$$Z_g \left(\frac{\mu}{f(N)} + \frac{Z_g - 1}{2N} + \sum_{k=0}^{g-1} (\bar{n}_k - C_k\varepsilon) \frac{N}{f(N)} \right)$$

and smaller than

$$Z_g \left(\frac{\mu}{f(N)} + \frac{Z_g - 1}{2N} + \sum_{k=0}^{g-1} (\bar{n}_k + C_k\varepsilon) \frac{N}{f(N)} \right).$$

The remaining part of the proof is the same as in the case $g = 0$, again with an application of [4, Lemma C.1]. \square

Proof of Lemma 6.4. Equation (29) is a property of quasi-stationary distribution (see for instance Proposition 2 in [16]).

The second part of the lemma arises from couplings of the process Z_0 with logistic birth and death processes and applications of results from [6]. To see this in detail, recall the definition of $b_0(\cdot, N)$ and $d_0(\cdot, N)$ in (22) and choose $\varepsilon > 0$. Then for N large enough and any $n \in \mathbb{N}$,

$$b_0(n, N) \leq \frac{\alpha n}{f(N)} \quad \text{and} \quad \frac{n}{f(N)} \left(\mu - \varepsilon + \frac{nf(N)}{2N} \right) \leq d_0(n, N) \leq \frac{n}{f(N)} \left(\mu + \frac{nf(N)}{2N} \right),$$

and for $n \leq \varepsilon N/\alpha$,

$$(\alpha - \varepsilon) \frac{n}{f(N)} \leq b_0(n, N).$$

Now let us consider two auxiliary birth and death processes $Z_0^{(+,N)}$ and $Z_0^{(-,N)}$, where

$$Z_0^{(+,N)} \text{ has jump rates } \begin{cases} nb_0(+, n, N) := \frac{\alpha n}{f(N)} & \text{from } n \text{ to } n+1, \\ nd_0(+, n, N) := \frac{n}{f(N)} \left(\mu - \varepsilon + \frac{nf(N)}{2N} \right) & \text{from } n \text{ to } n-1, \end{cases}$$

$$Z_0^{(-,N)} \text{ has jump rates } \begin{cases} nb_0(-, n, N) := (\alpha - \varepsilon) \frac{n}{f(N)} & \text{from } n \text{ to } n-1, \\ nd_0(-, n, N) := \frac{n}{f(N)} \left(\mu + \frac{nf(N)}{2N} \right) & \text{from } n \text{ to } n-1. \end{cases}$$

In addition we consider the process $Z_0^{(R,-,N)}$ which has the same rates as the process $Z_0^{(-,N)}$ on the set $\{0, \dots, \lfloor \varepsilon N/\alpha \rfloor - 1\}$ and is reflected from below at $N_0 := \lfloor \varepsilon N/\alpha \rfloor$, i.e. jumps from N_0 to $N_0 - 1$ at rate $N_0 d_0(-, N_0, N)$ but never jumps from N_0 to $N_0 + 1$ (see Figure 5).

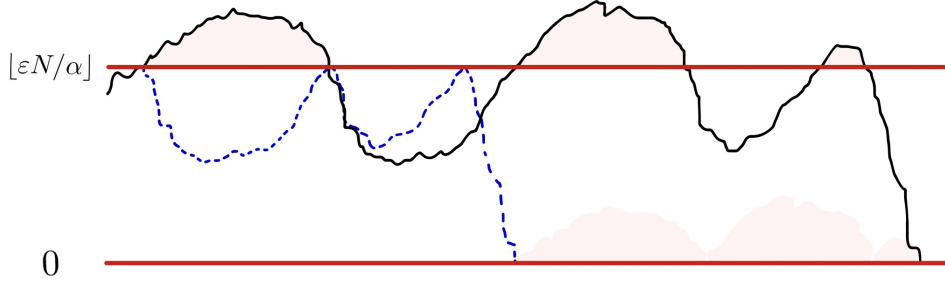


FIGURE 5. This figure schematically displays $Z_0^{(-,N)}$ (path drawn solid) as well as $Z_0^{(R,-,N)}$ (path drawn dashed), illustrating how the total length of the excursions of $Z_0^{(-,N)}$ above level $\lfloor \varepsilon N / \alpha \rfloor$ determines the difference of the times at which the two processes hit 0.

We can couple Z_0 and $Z_0^{(+,N)}$ such that for any $r \geq 0$,

$$Z_0(r) \leq Z_0^{(+,N)}(r).$$

Denoting by $\nu_N^{(\pm)}$ the quasi-stationary distributions of $Z_0^{(\pm,N)}$ and by $\theta_N^{(\pm)}$ the real numbers such that

$$\mathbb{P}_{\nu_N^{(\pm)}} \left(Z_0^{(\pm,N)}(r) > 0 \right) = e^{-r/f(N)\theta_N^{(\pm)}},$$

we thus obtain an upper bound for θ_N , namely $\theta_N \leq \theta_N^{(+)}$. Indeed Z_0 (resp. $Z_0^{(+,N)}$) with an initial state of order $N/f(N)$ may be coupled with Z_0 (resp. $Z_0^{(+,N)}$) with initial distribution ν_N (resp. $\nu_N^{(+)}$) in such a way that they coincide after a time of order $f(N) \ln N$ (the proof is similar to that of Lemma 6.3).

We can also couple Z_0 and $Z_0^{(R,-,N)}$ such that for any r ,

$$Z_0^{(R,-,N)}(r) \leq Z_0(r). \quad (69)$$

Indeed, this relation is fulfilled if $Z_0(r) \geq \lfloor \varepsilon N / \alpha \rfloor$, since $Z_0^{(R,-,N)} \leq \lfloor \varepsilon N / \alpha \rfloor$ by definition. As long as $Z_0^{(R,-,N)} \leq \lfloor \varepsilon N / \alpha \rfloor$, however, the process $Z_0^{(R,-,N)}$ has a smaller birth rate and a larger death rate than the process Z_0 .

The coupling (69) allows us to bound the mean extinction time of Z_0 by that of $Z_0^{(R,-,N)}$. In order to estimate the latter we will prove that the mean extinction times of the processes $Z_0^{(R,-,N)}$ and $Z_0^{(-,N)}$ are of the same order. We will then apply results of [16] to get an equivalent of $\theta_N^{(-)}$, the mean extinction time of the process $Z_0^{(-,N)}$. Let us consider a realization of the process $Z_0^{(-,N)}$, and denote by $(\mathcal{T}_i^{up}, \mathcal{T}_i^{down}, i \in \mathbb{N}_0)$ the successive entrance and exit times of $[\lfloor \varepsilon N / \alpha \rfloor + 1, \infty)$ by the process $Z_0^{(-,N)}$, defined recursively as follows:

$$\begin{aligned} \mathcal{T}_0^{up} &= \mathcal{T}_0^{down} = 0, \\ \mathcal{T}_i^{up} &:= \inf\{r \geq \mathcal{T}_{i-1}^{down}, Z_0^{(-,N)}(r) = \lfloor \varepsilon N / \alpha \rfloor + 1\}, \\ \mathcal{T}_i^{down} &:= \inf\{r \geq \mathcal{T}_i^{up}, Z_0^{(-,N)}(r) = \lfloor \varepsilon N / \alpha \rfloor\}, \quad i \in \mathbb{N}. \end{aligned}$$

Then if we ignore the excursions above $\lfloor \varepsilon N / \alpha \rfloor$ and glue together the points beginning and ending these excursions (that is the point \mathcal{T}_i^{up} and \mathcal{T}_i^{down}), we obtain a realization of the process $Z_0^{(R,-,N)}$, which is almost surely smaller than the process $Z_0^{(-,N)}$ at any time. To make this “glueing of excursions” formal, let us introduce the process $Z_0^{(C,-,N)}$ via

$$Z_0^{(C,-,N)}(r) := Z_0^{(-,N)} \left(r - \sum_{j=1}^i \left(\mathcal{T}_j^{down} - \mathcal{T}_j^{up} \right) \right)$$

for r in the random time interval

$$\mathcal{T}_i^{up} - \sum_{j=1}^{i-1} \left(\mathcal{T}_j^{down} - \mathcal{T}_j^{up} \right) \leq r \leq \mathcal{T}_{i+1}^{up} - \sum_{j=1}^i \left(\mathcal{T}_j^{down} - \mathcal{T}_j^{up} \right).$$

This process obeys

$$Z_0^{(C,-,N)} \stackrel{\mathcal{L}}{=} Z_0^{(R,-,N)} \quad \text{and} \quad Z_0^{(C,-,N)}(r) \leq Z_0^{(-,N)}(r), \quad \forall r \geq 0, \quad a.s.$$

Thus, if we prove that the mean time spent above $\lfloor \varepsilon N / \alpha \rfloor$ by the process $Z_0^{(-,N)}$ is negligible with respect to its mean extinction time $\theta_N^{(-)}$, we may deduce that the mean extinction time of the process $Z_0^{(R,-,N)}$ is equivalent to $\theta_N^{(-)}$, which in turn entails that $\theta_N^{(-)} \leq \theta_N$.

As $\nu_N^{(-)}$ is a quasi-stationary distribution, we have for every $r \geq 0$ (see for instance [16]),

$$\begin{aligned} \nu_N^{(-)}(\lfloor \varepsilon N / \alpha \rfloor, \infty) &= \mathbb{P}_{\nu_N^{(-)}} \left(Z_0^{(-,N)}(r) \in (\lfloor \varepsilon N / \alpha \rfloor, \infty) \right) \left(\mathbb{P}_{\nu_N^{(-)}} \left(\mathcal{H}_0^{(N)} > r \right) \right)^{-1} \\ &= \mathbb{P}_{\nu_N^{(-)}} \left(Z_0^{(-,N)}(r) \in (\lfloor \varepsilon N / \alpha \rfloor, \infty) \right) e^{r/f(N)\theta_N^{(-)}}. \end{aligned}$$

Hence, the expected time spent by the process $Z_0^{(-,N)}$ in the set $(\lfloor \varepsilon N / \alpha \rfloor, \infty)$ is

$$\begin{aligned} \mathbb{E}_{\nu_N^{(-)}} \left[\int_0^\infty \mathbf{1}_{\{Z_0^{(-,N)}(r) > \lfloor \varepsilon N / \alpha \rfloor\}} dr \right] &= \int_0^\infty \mathbb{P}_{\nu_N^{(-)}} \left(Z_0^{(-,N)}(r) > \lfloor \varepsilon N / \alpha \rfloor \right) dr \\ &= \nu_N^{(-)}(\lfloor \varepsilon N / \alpha \rfloor, \infty) \int_0^\infty e^{-r/f(N)\theta_N^{(-)}} dr = \nu_N^{(-)}(\lfloor \varepsilon N / \alpha \rfloor, \infty) f(N) \theta_N^{(-)}, \end{aligned}$$

where we applied Fubini's Theorem.

Now from Theorem 3.7 in [6], we get that the total variation distance between $\nu_N^{(-)}$ and a Gaussian law centered at $2(\alpha - \mu - \varepsilon)N/f(N)$ and with variance $N/((\alpha - \varepsilon)f(N))$ is of order $\sqrt{f(N)/N}$. We deduce that

$$\lim_{N \rightarrow \infty} \nu_N^{(-)}(\lfloor \varepsilon N / \alpha \rfloor, \infty) = 0,$$

which allows us to conclude that $\theta_N^{(-)} \leq \theta_N \leq \theta_N^{(+)}$.

A direct application of Remark 3.3 in [6] yields that, as $N \rightarrow \infty$,

$$\theta_N^{(-)} \sim f(N)(\alpha - \varepsilon - \mu)^2 \sqrt{2\pi\mu \frac{f(N)}{2N}} \exp \left(2 \left(\alpha - \varepsilon - \mu + \mu \ln \frac{\mu}{\alpha - \varepsilon} \right) \frac{N}{f(N)} \right)$$

and

$$\theta_N^{(+)} \sim f(N)(\alpha + \varepsilon - \mu)^2 \sqrt{2\pi(\mu - \varepsilon) \frac{f(N)}{2N}} \exp \left(2 \left(\alpha + \varepsilon - \mu + (\mu - \varepsilon) \ln \frac{\mu - \varepsilon}{\alpha} \right) \frac{N}{f(N)} \right).$$

This concludes the proof of (7), and thus completes the proof of Lemma 6.4. \square

Proof of Lemma 6.5. To prove part a) of the lemma, we will zoom on a small window before the extinction time of $Z_0 = Z_0^{(N)}$, namely after the last hitting time of $1/\varepsilon^2$. The strategy of the proof consists in showing that this time is short (having a duration of order $\ln 1/\varepsilon$) and that on the way to extinction, Z_0 will feed the Z_1 population by producing many individuals that carry one mutation.

To begin with, we consider the process Z_0 conditioned to reach 0 before $\lfloor 1/\varepsilon^2 \rfloor$. Its transition matrix \widehat{P} arises from that of the unconditioned Z_0 as the harmonic transform

$$\widehat{P}(i, i-1) := \frac{h(i-1)}{h(i)} \frac{\mu + \frac{f(N)(n-1)}{2N}}{\alpha \left(1 - \frac{i}{N}\right) + \mu + \frac{f(N)(n-1)}{2N}},$$

where

$$h(i) := \mathbb{P} \left(H_0^{(N)} < H_{\lfloor 1/\varepsilon^2 \rfloor}^{(N)} \mid Z_0(0) = i \right), \quad i = 0, \dots, \lfloor 1/\varepsilon^2 \rfloor, \quad (70)$$

and $H_z^{(N)}$ is the first hitting time of z , cf. (28). The form of the harmonic functions for birth and death chains in terms of ratios of upward and downward jump rates is well known; for (70) this leads to the expression

$$h(i) = 1 - (1 - h(1)) \sum_{j=1}^i \prod_{k=1}^{j-1} \rho_k, \quad \text{where} \quad \rho_k := \frac{\mu + \frac{f(N)(k-1)}{2N}}{\alpha \left(1 - \frac{k}{N}\right)}, \quad 1 \leq k < N.$$

Hence we need to study the quantity

$$\frac{h(i-1)}{h(i)} = \frac{1 - (1 - h(1)) \sum_{j=1}^{i-1} \prod_{k=1}^{j-1} \rho_k}{1 - (1 - h(1)) \sum_{j=1}^i \prod_{k=1}^{j-1} \rho_k}. \quad (71)$$

In fact a lower bound will be sufficient in our case. We notice that the expression in (71) is non-decreasing if any of the ρ_k 's is increased, and that $\rho_k \leq \mu/\alpha$ for any k . We thus get

$$\frac{h(i-1)}{h(i)} \geq \frac{1 - (1 - h(1)) \sum_{j=1}^{i-1} \prod_{k=1}^{j-1} \mu/\alpha}{1 - (1 - h(1)) \sum_{j=1}^i \prod_{k=1}^{j-1} \mu/\alpha} = \frac{1 - (1 - h(1)) \frac{\alpha}{\alpha - \mu} \left(1 - \left(\frac{\mu}{\alpha}\right)^{i-1}\right)}{1 - (1 - h(1)) \frac{\alpha}{\alpha - \mu} \left(1 - \left(\frac{\mu}{\alpha}\right)^i\right)}.$$

The last step consists in finding an equivalent of $1 - h(1)$ for large N . As for any $k \leq 1/\varepsilon^2$,

$$\frac{\mu}{\alpha} \leq \rho_k \leq \frac{\mu + \frac{f(N)}{2N\varepsilon^2}}{\alpha \left(1 - \frac{1}{N\varepsilon^2}\right)},$$

we obtain from classical results on hitting probabilities for Galton-Watson process (see [2])

$$\left(1 - \frac{\mu + \frac{f(N)}{2N\varepsilon^2}}{\alpha \left(1 - \frac{1}{N\varepsilon^2}\right)}\right) \left(1 - \left(\frac{\mu + \frac{f(N)}{2N\varepsilon^2}}{\alpha \left(1 - \frac{1}{N\varepsilon^2}\right)}\right)^{1/\varepsilon^2}\right)^{-1} \leq 1 - h(1) \leq \left(1 - \frac{\mu}{\alpha}\right) \left(1 - \left(\frac{\mu}{\alpha}\right)^{1/\varepsilon^2}\right)^{-1}.$$

We thus get that

$$1 - h(1) = \frac{\alpha - \mu}{\alpha} + \delta(1/N) \left(\frac{\mu}{\alpha}\right)^{1/\varepsilon^2},$$

where $\delta(1/N) \rightarrow 0$ as $N \rightarrow \infty$, which entails that for any $1 \leq i \leq 1/\varepsilon^2$,

$$\frac{h(i-1)}{h(i)} = \frac{\alpha}{\mu} + \delta(1/N)$$

and consequently

$$\widehat{P}(i, i-1) = \frac{\alpha}{\alpha + \mu} + \delta(1/N).$$

In other words, the process Z_0 conditioned to reach 0 before $1/\varepsilon^2$ behaves as a subcritical branching process with individual birth and death rates $\mu/f(N)$ and $\alpha/f(N)$. With a probability close to one when ε is small, it thus takes a time smaller than $4f(N)/(\alpha-\mu)\ln(1/\varepsilon)$ to reach 0. We will now prove that during this time, on the way of extinction, a number of order $1/\varepsilon^2$ of individuals of the Z_1 population are produced by mutation from Z_0 individuals.

Let us denote by $\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}((k,l))$ the probability for the process (Z_0, Z_1) to perform its first jump to (k,l) when starting in (i,j) . We then have

$$\begin{aligned} \mathbb{P}_{(i,j)}^{(Z_0,Z_1)}\left((i-1,j+1)\middle|H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\right) &= \frac{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}\left((i-1,j+1), H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\right)}{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}\left(H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\right)} \\ &= \frac{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}((i-1,j+1))}{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}\left(H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\right)} \mathbb{P}_{(i,j)}^{(Z_0,Z_1)}\left(H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\middle|(i-1,j+1)\right) \\ &= \frac{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}((i-1,j+1))}{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}\left(H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\right)} \mathbb{P}_{(i-1,j+1)}^{(Z_0,Z_1)}\left(H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\right) = \mathbb{P}_{(i,j)}^{(Z_0,Z_1)}((i-1,j+1)) \frac{h(i-1)}{h(i)}, \end{aligned}$$

with $h(i)$ as in (70). The last equality is a consequence of the autonomy of the law of Z_0 , cf. Remark 6.1. In the same way, we can prove that

$$\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}\left((i-1,j)\middle|H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\right) = \mathbb{P}_{(i,j)}^{(Z_0,Z_1)}((i-1,j)) \frac{h(i-1)}{h(i)}.$$

This entails that

$$\frac{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}\left((i-1,j+1)\middle|H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\right)}{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}\left((i-1,j)\middle|H_0^{(N)} < H_{1/\varepsilon^2}^{(N)}\right)} = \frac{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}((i-1,j+1))}{\mathbb{P}_{(i,j)}^{(Z_0,Z_1)}((i-1,j))}.$$

In other words, conditioning on the event that Z_0 is on its way to extinction does not modify the proportion of those deaths in the process Z_0 which lead to a creation of a new individual in the Z_1 population. Hence, if we denote by $\mathcal{M}(\varepsilon)$ the number of mutations from Z_0 to Z_1 after the last visit of Z_0 to $\lfloor 1/\varepsilon^2 \rfloor$, we have

$$\mathcal{M}(\varepsilon) \geq \sum_{i=1}^{\lfloor 1/\varepsilon^2 \rfloor} \text{Be}^{(i)}\left(\frac{\mu}{\mu + \frac{f(N)(i-1)}{2N}}\right),$$

where the $\text{Be}^{(i)}(\lambda_i)$'s are independent Bernoulli random variables with parameter λ_i . Indeed, in the considered time interval there are at least $1/\varepsilon^2$ deaths of individuals in the Z_0 population, and the parameter of $\text{Be}^{(i)}$ is the probability that a jump from i to $i-1$ in the Z_0 population leads to the arrival of a new individual in the Z_1 population. For ε small enough, the random variable $\mathcal{M}(\varepsilon)$ is thus larger than $1/(2\varepsilon^2)$ with a probability tending to 1 as $N \rightarrow \infty$. The evolution of the Z_1 population after the last visit of Z_0 to $\lfloor 1/\varepsilon^2 \rfloor$ can thus with high probability be coupled to the offspring of $\lfloor 1/(2\varepsilon^2) \rfloor$ immigrants arriving in a time interval of length $4f(N)/(\alpha-\mu)\ln(1/\varepsilon)$, with their offspring suffering an individual death rate by competition from Z_0 individuals which is smaller than $1/N\varepsilon^2$, as well as a reduction of their individual birth rate (compared to α) that is smaller than $\alpha/(f(N)\varepsilon^2N)$. This proves part a) of Lemma 6.5.

To prove part b), in view of Lemmata 6.3 and 6.4 it suffices to restrict to the event $\{\underline{k}_N > 0\}$. We can then work iteratively along the extinction times of the $Z_0, \dots, Z_{\underline{k}_N-1}$ populations, applying part a) by induction. \square

ACKNOWLEDGEMENT

We are grateful to Ralph Neininger for pointing us to [1] in connection with Proposition 11.1, see Remark 11.2b). We also thank Antar Bandyopadhyay for related discussions, and Mariana Gomez for showing us the link between our model and tournament selection. Our sincere thanks go to the Mathematical Research Institute Oberwolfach where core parts of this project were carried out in a joint stay as Oberwolfach research fellows in summer 2022. AGC and AW acknowledge the hospitality of INRAE and Institut Fourier in October 2021; during this joint visit (funded by the French program Investissement d’avenir), substantial preparatory studies were performed. AGC acknowledges the hospitality of Goethe University during his visit to Frankfurt in September 2022.

REFERENCES

- [1] D. J. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equations. *The Annals of Applied Probability*, 15(2):1047–1110, 2005.
- [2] K. B. Athreya and P. E. Ney. *Branching processes*. Springer-Verlag Berlin, Mineola, NY, 1972. Reprint of the 1972 original [Springer, New York; MR0373040].
- [3] T. Blickle and L. Thiele. A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computation*, 4(4):361–394, 1996.
- [4] N. Champagnat, S. Méléard, and V. C. Tran. Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer. *The Annals of Applied Probability*, 31(4):1820–1867, 2021.
- [5] N. Champagnat and D. Villemonais. General criteria for the study of quasi-stationarity. *arXiv preprint arXiv:1712.08092*, 2017.
- [6] J.-R. Chazottes, P. Collet, and S. Méléard. Sharp asymptotics for the quasi-stationary distribution of birth-and-death processes. *Probab. Theor. Rel. Fields*, 164(1-2):285–332, 2016.
- [7] F. Cordero, S. Hummel, and E. Schertzer. General selection models: Bernstein duality and minimal ancestral structures. *The Annals of Applied Probability*, 32(3):1499–1556, 2022.
- [8] A. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. How often does the ratchet click? Facts, heuristics, asymptotics. *Trends in stochastic analysis*, 353:365–390, 2009.
- [9] S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.
- [10] A. González Casanova and C. Smadi. On λ -fleming–viot processes with general frequency-dependent selection. *Journal of Applied Probability*, 57(4):1162–1197, 2020.
- [11] A. González Casanova and D. Spanò. Duality and fixation in ξ -wright–fisher processes with frequency-dependent selection. *The Annals of Applied Probability*, 28(1):250–284, 2018.
- [12] J. Haigh. The accumulation of deleterious genes in a population - Muller’s ratchet. *Theoretical population biology*, 14(2):251–267, 1978.
- [13] Y. Hu. How big is the minimum of a branching random walk? *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 52(1):233–260, 2016.
- [14] S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theoretical population biology*, 51(3):210–237, 1997.
- [15] M. Mariani, E. Pardoux, and A. Velleret. Metastability between the clicks of the Muller ratchet. *arXiv preprint arXiv:2007.14715*, 2020.
- [16] S. Méléard and D. Villemonais. Quasi-stationary distributions and population processes. *Probability Surveys*, 9:340–410, 2012.
- [17] J. J. Metzger and S. Eule. Distribution of the fittest individuals and the rate of Muller’s ratchet in a model with overlapping generations. *PLoS computational biology*, 9(11):e1003303, 2013.
- [18] H. J. Muller. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1):2–9, 1964.

- [19] C. Neuhauser and S. M. Krone. The genealogy of samples in models with selection. *Genetics*, 145(2):519–534, 1997.
- [20] P. Pfaffelhuber, P. R. Staab, and A. Wakolbinger. Muller's ratchet with compensatory mutations. *The Annals of Applied Probability*, 22(5):2108–2132, 2012.
- [21] C. Pokalyuk and P. Pfaffelhuber. The ancestral selection graph under strong directional selection. *Theoretical population biology*, 87:25–33, 2013.
- [22] J. M. Smith and J. Maynard-Smith. *The evolution of sex*, volume 4. Cambridge University Press Cambridge, 1978.

ADRIAN GONZÁLEZ CASANOVA, UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO (UNAM), INSTITUTO DE MATEMÁTICAS, CIRCUITO EXTERIOR, CIUDAD UNIVERSITARIA, 04510, MÉXICO, 11
E-mail address: `adriangcs@matem.unam.mx`

CHARLINE SMADI, UNIV. GRENOBLE ALPES, INRAE, LESSEM, 38000 GRENOBLE, FRANCE AND UNIV. GRENOBLE ALPES, CNRS, INSTITUT FOURIER, 38000 GRENOBLE, FRANCE
E-mail address: `charline.smadi@inrae.fr`

ANTON WAKOLBINGER, GOETHE-UNIVERSITÄT, INSTITUT FÜR MATHEMATIK, 60629 FRANKFURT AM MAIN, GERMANY
E-mail address: `wakolbinger@math.uni-frankfurt.de`