Clonal Interference: Efficiently Approximating Fixation Probabilities in the Moran process with Multiple Types

T Schwarz

University of Oxford

Dissertation in candidacy for the

Master of Science in Mathematics and Foundations of Computer Science

Trinity 2022

Abstract

The spatial Moran process is a stochastic process on a graph modelling how an advantageous mutation spreads through a finite population and, as such, has applications ranging from biology, *e.g.* to model cancer initiation and progression like the development of Leukemia [6, 15], to social sciences, where closely related models have been used to model voter behaviour and the spread of ideas trough social networks [18, 26]. The model distinguishes between two types of vertices, namely "heathy" and "mutated", which reproduce along the graph's edges. The key quantity of interest is the probability of all vertices becoming "mutated", the so-called *fixation probability* of the mutation. Both the asymptotic behaviour of this probability as the underlying graph grows and the algorithmic problem of approximating this fixation probability have been well-explored in recent years, most notably by establishing a phase-transition theorem as well as by providing a *fully polynomial randomised approximation scheme* (FPRAS) of the fixation probability [1, 13].

What happens if multiple mutation types, each of different fitness, arise? Then, not only "mutated" cells will compete against "healthy" cells, but different mutation lineages will fight for dominance, as it is the case in Carcinogenesis.[1] How can we approximate the fixation probability of each arising mutation type individually? Is it possible to provide an *efficient* approximation? This dissertation (1) extends the spatial Moran process to account for multiple mutants and (2) answers both questions above to the positive, by providing efficient approximation algorithms and good lower bounds. We further (3) explore what happens if we go beyond the assumption that fitness and number of mutations stay constant, which translates in complexity terms to a *parameterised* computational problem.

In more detail, we introduce the k-type Moran process where type 1 can be thought of as "healthy" vertices and types $2, \ldots, k$ represent the invading mutations, each having their own fitness r_i. If we assume k to be constant and follow previous literature of the 2-type Moran process in assuming the fitness advantages r_i to be constant, we give an FPRAS for the fixation probability of each individual mutation type. That is, for every connect graph of size n and every $\varepsilon > 0$ we provide an algorithm that approximates the fixation probability of every mutation within a factor of $(1 \pm \varepsilon)$ with probability at least $1 - \delta$, and has running time polynomial in n, $1/\varepsilon$, $\log(1/\delta)$, where n is the size of the underlying graph (Corollary 25).

We further discover how the approximation problem changes when we regard both the number of mutants k and their fitnesses as parameters. By introducing a stochastic coupling (Theorem 26) we can reduce the problem of approximating the fixation probability of the strongest mutant for any number of mutants k to the 2-type case, yielding lower bounds independent of k and consequently a *fixed-parameter tractable approximation scheme* (Main Theorem 5), which even becomes an FPRAS under reasonable assumptions on the fitness parameter (Corollary 33). We extend this to approximate the fixation probability of other, sufficiently strong, mutation types in the parameterised problem version (Main Theorem 6 and Corollary 36). The generality of our stochastic coupling becomes evident when we derive the bounds on fixation probabilities obtained in recent work [19] as a corollary of a very restricted subproblem (Corollary 40).

We further resolve an open problem from Monk and Schaik [32] by providing a martingale for every number of $k \geq 2$ types on the complete graph and thereby generalise a known martingale for the $k = 2$ type process [31]. Finally, we provide an implementation of the k-type Moran process.

All results mentioned above are novel results.

---

[1] Carcinogenesis, *i.e.* the formation of cancer. According to Vogelstein *et al.* a typical tumour contains two to eight "driver" mutations and hundreds of "passenger" mutations [36]. In the mentioned social science application, this could be interpreted as multiple competing ideas.

Acknowledgements

I owe a lot to both of my supervisors Prof. Leslie Ann Goldberg and Dr. Marc Roth, who taught me how to guide research in the right way and got me accustomed not only to the standards in research, but also to the particular techniques in this area. I am thankful for many stimulating discussions on the topic and their inspiring enthusiasm.

I am grateful to my parents for their patience and support throughout my time in England.

I thank the *German Academic Scholarship foundation* for financial and academic support, which made my master studies at the University of Oxford possible.

Contents

Appendices

List of Figures

Notation

Throughout this work, we use the following conventions. We denote by $\mathbb{N}$ the natural numbers $\{1, 2, \ldots\}$, which we extend to $\mathbb{N}_0$ if 0 is included. The positive real numbers are $\mathbb{R}^+ := \{r \in \mathbb{R} : r > 0\}$. $\mathbb{R}^{m,n}$ is the set of real $m \times n$ matrices. $\mathbb{R}^*$ is the set of finite tuples whose elements are in $\mathbb{R}$.

The variables i, j, k are always positive integers. Positive integers until k are denoted by $[k] := \{1, 2, \ldots, k\}$.

Disjoin unions are denoted by $\sqcup$. Adding and removing singletons from sets is denoted by $S + s_1 := S \cup \{s_1\}$ and $S - s_1 := S \setminus \{s_1\}$.

All graphs $G := (V, E)$ treated in this work are simple, undirected, connected, and of finite order $n \geq 3$, where $n := |V|$ denotes the graph's *order*. Its edges $\{u, v\} \in E$ are abbreviated by $uv \in E$. For the set of all edges between two vertex sets $V_1, V_2 \subseteq V$, we write $E(V_1, V_2)$. $G[S]$ is the subgraph induced by $S \subseteq V$. The complete graph of order n is denoted by K_n.

We denote by $\operatorname{Exp}(\mu)$ the exponential distribution with parameter μ and recall that the minimum of two exponentially distributed random variables is again exponentially distributed with the parameters summed: $\forall \sigma, \mu > 0 : \quad X \sim \operatorname{Exp}(\sigma), Y \sim \operatorname{Exp}(\mu) \quad \Rightarrow \min(X, Y) \sim \operatorname{Exp}(\sigma + \mu)$. "l.o.e." abbreviates "linearity of expectation". We denote that two distributions are equal by $\overset{\text{distr.}}{=}$.

When talking about the k-type Moran process, we always assume $2 \leq k \leq n$, since there can be at most as many types as vertices and there need to be at least 2 types to make the process interesting.

1 Introduction

Spin systems are widely used in statistical physics and find applications from biology to social sciences to explain a global behaviour of a set of particles based on their local interaction [6, 26, 29]. Among those models is the Moran process, which describes how two asexually reproducing organisms evolve in a finite population over time. Originally proposed by Moran in 1958 [33], it has been generalised by Lieberman, Hauert and Nowak in 2005 to account for a spacial structure among the population, modelled by a finite graph [28]. What is the probability that one organism *fixates*, *i.e.* eventually occupies the whole graph, or goes *extinct*. While explicit formulas for the fixation probability can be derived for certain families of graphs (*e.g.* for regular graphs), there is no efficient way of determining it for arbitrary graphs known. Current approaches would required to solve a system of equations whose size grows exponentially in the graph's order. Thus, the need for efficient randomised approximation schemes arises. Diaz *et al.* [13] resolve this need by giving a fully polynomial randomised approximation scheme (FPRAS) for the fixation probability of the advantageous mutation using a Markov-Chain-Monte-Carlo algorithm. At the heart of this approximation algorithm lies a lower bound of the fixation probability, for which Goldberg, Lapinskas, and Richerby later established a phase transition in the organism's fitness advantage [1].

The models considered so far only describe the competition between *two* types of organisms (in the following described simply as *types*), which has been used to model how malignant cancer cells spread through tissue [9]. In such models, the vertices of the graph represent cells, edges represent a spatial or resource-sharing connection, and each vertex is either "healthy" or "cancerous" [6]. However, it is well-known that in Carcinogenesis more than just one mutation occurs. Rather, different types of mutations "fight" for dominance in a tissue. Thus, the need arises to go beyond the so-called *strong-selection, weak mutation* assumption, "in which the time for beneficial mutations to fix is much shorter than the time for successful beneficial mutations to emerge" [23]. Our goal is twofold: We first extend the spatial Moran process to account for multiple mutations of different fitness advantages, a process known as *Clonal Interference* in genetics [3, 34]. Second, we aim to approximate the fixation probabilities of each of those mutants efficiently.

We therefore go beyond the standard model of spin systems, which is restricted to vertices being only in either of two states [29, 30, 27].

Other applications (and variations of) the Moran process include voter behaviour and the spread of ideas through a social network, where our generalisation to Clonal Interference could be seen in the latter context as the evolution of multiple competing ideas in a social network [26, 30]. For brevity, we stick throughout this work to the biological interpretation introduced above when providing the reader with intuition behind assumptions and definitions.

In the remainder of this section, we first give an intuition behind the Moran process, then describe similar spin systems and efforts to group multiple of these models into a single "meta-model", and subsequently give further biological motivation as well as previous efforts to generalise to multiple mutation types. We then introduce the model formally and generalise it to k types (Section 1.1), describe to what extend general Markov chain theory can be applied (Section 1.2), and then define computational problems arising from the k-type Moran process (Section 1.3).

The spatial Moran process We now give the intuition behind the spatial Moran process, before turning to a rigorous mathematical definition in due course (Section 1.1). The spatial structure is described by a simple, connected graph whose vertices can be though of as cells, each of which is of either type 1 or type 2. Reproduction happens along the edges. The difference between both types is that type-1-vertices reproduce at rate r_1, and type-2-vertices reproduce at rate r_2.

Initially, all cells are of type 1, except for one vertex selected uniformly at random which is of type 2. This vertex embodies the "invading" mutation. At each subsequent time step, one vertex gets selected at random, where type-1-vertices are selected with probability proportional to r_1 and type-2-vertices proportional to r_2.[2] The selected vertex will then *reproduce to* one of its neighbours, selected uniformly at random, *i.e.* copy its type to that neighbour.

Related spin systems There are two to the Moran process closely related Spin systems, both of which assume an infinite lattice as the underlying graph: The *contact process* is a continuous time spin system, where each vertex is in either of two states. Vertices become infected at rate proportional to the number of infected neighbours and, once infected, become healthy at some constant rate. The *voter model* has the same state space. However, state transitions happen according to a different mechanism: Each vertex waits for an exponentially distributed amount of time before choosing to adopt its state to one of its neighbours, chosen according to some probability distribution. This model has been generalised to the *biased* voter model, introduced by Schwartz in 1977 [35], where transitions from type 1 to type 2 are more likely than the reverse transitions. If the underlying lattice is the finite 2-dimensional lattice, the model is called the Williams-Bjerknese model, who introduced it to study the spread of cancer in 1972 [38].

For a rigorous treatment of the described models, see Liggett [30] and Bramson and Griffeath [9]. We will show in Appendix C how Lieberman, Hauert, and Nowak's model can be regarded as the embedded discrete time Markov chain of the Williams-Bjerknes model, generalised to finite graphs.

Even though these models are closely related to the Moran process, only a slight change in the update rules or a change of the underlying graph changes the system's global behaviour significantly, justifying the coexistence of so many models [17].

Grouping these models With multiple, closely related co-exiting models, it would be desirable to group them and study their global behaviour all together. There have been recent efforts to group a family of similar models with local update rules into a "meta-model", where the update rules are treated as a general list of local rules [8], with interesting sequel work (*e.g.* [7]). However, models generalised so far have only a random *initial* state, constrain the set of infected vertices to be increasing, and have only one absorbing state. The Moran process, in contrast, has a random initial state *and* probabilistic update rules. Further, its set of infected vertices is in general not increasing and our generalised Moran process has, as we will see in due course, k absorbing states for k types.

Biological motivation The Moran process in its original form[3] has been used to model how an asexual population reproduces, *e.g.* in the emergence and progression of cancer [5, 6]. To speed up the computational simulation, the Wright-Fisher Model [20, 39] is often used where, roughly speaken, the entire population reproduces at once. For the population size tending to infinity, those models exhibit the same fixation probabilities [37].

However, it became clear that it is not only the mere number of mutants that defines the system's behaviour, but also the population's structure [4]. Thus, the need to account for

---

[2] The attentive reader might notice that it suffices to let type-1-vertices reproduce at rate rate 1 and type-2-vertices at rate r_2/r_1. This is mathematically correct and we will comment on this in Section 1.1. However, with our generalisation to multiple mutations looming, it is more comprehensive to assign each type its own rate, including type 1.

[3] *i.e.* without accounting for the population's spatial structure

spatial structure evolved, motivating the previously mentioned *spatial Moran process* [28]. In Appendix C, we describe how Lieberman, Hauert, and Nowak's model is connected to the so-called Williams-Bjerknes Tumour Growth Model from 1972 and is thereby not an entire novel idea, most certainly not for modelling tumour spreading.

Multiple types in previous literature It became evident that modelling only a single mutation type is not enough to explain phenomena occurring in asexually reproducing populations [34]. To meet that need for modelling multiple mutation types, different models have been proposed in literature. The most prominent ones are due to Beerenwinkel *et al.* [5] and, most recently, Ferreira and Neves [19]. All proposed models, however, account only for a very limited case of the Moran process.

Beerenwinkel *et al.* generalised the Wright-Fisher model [5]. They associate to each fitness type i fitness $(1+s)^i$ for some parameter $s \in \mathbb{R}^+$ and ran simulations.[4] They quote efficiency reasons for choosing the Wright-Fisher model over the more natural Moran process in their simulations. The results stem from simulations and imposing no structure among individual cells.

Ferreira and Neves proposed a model with three mutation types. Unfortunately, their model is constrained to the complete graph, so that the population has only a trivial structure. For this very limited model, they provide bounds on the fixation probabilities of all three types.

It is our mission to fill the gap of generalising the Moran process in its full generality, *i.e.* for an arbitrary connected graph, to multiple mutation types. We further aim to provide good bounds on the fixation probabilities and efficient approximation algorithms. From our general model, we will be able to derive Ferreira and Neves' bounds as a special application of our coupling theorem (Theorem 26), see Appendix B.

1.1 The model

In the following, we use the term *Moran process* to refer to the spatial Moran process introduced by Lieberman, Hauert, and Nowak [28]. We first introduce in Section 1.1.1 formal definitions to describe the 2-type Moran process, which is equivalent to the process considered in previous literature [28, 14], except that our definition is more comprehensive than necessary. It is this more comprehensive notation that allows us to easily generalise to the k-type Moran process, which we introduce in Section 1.1.2. The notation in this section is thus deliberately very formal, to both allow for an easy generalisation from 2 to k types and to allow for concise statements in later theorems.

1.1.1 The conventional 2-type Moran process

We first introduce the notion of a *fitness vector*, capturing the fitness of each type.

Definition 1 (Fitness vector)**.** For $k \geq 2$, a fitness vector $\boldsymbol{r} := (r_1, \ldots, r_k) \in (\mathbb{R}^+)^k$ is a positive, strictly increasing vector: $r_i \leq r_{t+1}$ for any $i \in [k-1]$.

The idea of the strictly increasing property is that we fix the "default"[5] type 1 to be the weakest, since the evolutionary interesting case is when a fitter-than-normal mutation arises. Ordering

---

[4] To be concise, they started initially with all type 1 vertices and added, in addition to the normal reproduction mechanism, a mechanism so that type i can mutate to type $i+1$. Thus, the notion of fixation probability of the strongest mutation is not well-defined since the type of each vertex tends towards ∞. They therefore analysed a different question, namely, the expected time until mutation type k arose.

[5] Initially, all vertices are of type 1. Then, one for each of the types $2, \ldots, k$ is selected at random. This will become clear in the subsequent definition of the process.

$r_2 < \cdots < r_k$, is w.l.o.g. since there is no difference in the Moran process's definition between types $2, \ldots, k$. However, in some occasions, the fitness vector will be non-decreasing or even unordered. In these cases, we call it explicitly the *non-decreasing* or *unordered* fitness vector $\boldsymbol{r}'$ and specify its entries further. This notion avoids extensive remapping of the types when we couple multiple Moran processes in Section 3.

We now introduce the notion of a *type mapping*, which assigns each vertex a unique type.

Definition 2 (type mapping)**.** For any $k \geq 2$ and graph with vertex set V, let $\sigma : V \to [k]$ be the mapping of each vertex to its unique type.

- We denote by σ_t the mapping at time $t \in \mathbb{N}_0$.
- We denote by $\mathrm{R}(\sigma)$ the range (image) of σ. Note that we have $|\,\mathrm{R}(\sigma)\,| \geq 1$, since each vertex belongs to some type. We say that σ is trivial if $|\,\mathrm{R}(\sigma)\,| = 1$.
- For $v \in V$, $j \in [k]$, we denote by $\sigma[v \mapsto j]$ the type mapping obtained from σ by changing v to have type j.
- We denote the preimage of $\{j\}$ by $\sigma^{-1}(j)$ for every type $j \in [k]$.

With the definition of a type mapping at hand, we can regard a Moran process as the evolution of its type mapping over time. More precisely, for a given graph $G = (V, E)$ and fitness vector $\boldsymbol{r} = (r_1, r_2)$, the Moran process M is the Markov chain

$$M := \{\sigma_t | t \in \mathbb{N}_0\}$$

defined on the state space $\Omega := \{V \to [2]\}$, where σ_t are 2-type mappings, with transition probabilities:

$$\begin{aligned}
\forall v \notin \sigma_t^{-1}(1): \qquad \mathbb{P}\left[\sigma_{t+1} = \sigma_t[v \mapsto 1] | \sigma_t\right] &= \sum_{u \in N(v) \cap \sigma_t^{-1}(1)} \frac{r_{\sigma_t(u)}}{W(\sigma_t)} \cdot \frac{1}{\mathsf{deg}(u)} \\
&= \frac{r_1}{W(\sigma_t)} \sum_{u \in N(v) \cap \sigma_t^{-1}(1)} \frac{1}{\mathsf{deg}(u)} \qquad (1) \\
\forall v \notin \sigma_t^{-1}(2): \qquad \mathbb{P}\left[\sigma_{t+1} = \sigma_t[v \mapsto 2] | \sigma_t\right] &= \sum_{u \in N(v) \cap \sigma_t^{-1}(2)} \frac{r_{\sigma_t(u)}}{W(\sigma_t)} \cdot \frac{1}{\mathsf{deg}(u)} \\
&= \frac{r_2}{W(\sigma_t)} \sum_{u \in N(v) \cap \sigma_t^{-1}(2)} \frac{1}{\mathsf{deg}(u)} \qquad (2)
\end{aligned}$$

$$\mathbb{P}\left[\sigma_{t+1} = \sigma_t | \sigma_t\right] = 1 - \sum_{v \notin \sigma_t^{-1}(1)} \mathbb{P}\left[\sigma_{t+1} = \sigma_t[v \mapsto 1] | \sigma_t\right] - \sum_{v \notin \sigma_t^{-1}(2)} \mathbb{P}\left[\sigma_{t+1} = \sigma_t[v \mapsto 2] | \sigma_t\right]. \qquad (3)$$

with $W(\sigma_t) := r_1 \cdot |\sigma_t^{-1}(1)| + r_2 \cdot |\sigma_t^{-1}(2)|$ being the population's overall fitness in state σ_t.

Transition probabilities (1) and (2) result from our intuitive description of the process as follows: At each time step, one vertex $u \in V$ is selected at random, where type j vertices are selected proportionally to r_j for $j \in \{1, 2\}$. Vertex u reproduces to (*i.e.* copies its type to) a neighbour selected u.a.r. Hence, to obtain the probability that any vertex $v \in V$ changes its type, we have to sum over all vertices $u \in N(v)$ that are of a different type the probability that u gets selected (*i.e.* $r_{\sigma_t(u)}/W(\sigma_t)$) and reproduces to its neighbour v (*i.e.* $1/\mathsf{deg}(u)$). Equation (3) simply accounts for the remaining cases when no vertex changes its type.

Properties of the Markov chain We have two absorbing states: For the state σ with $R(\sigma) = \{1\}$, we have $\sigma^{-1}(2) = \emptyset$, so that σ is absorbing by eq. (2). Similarly, σ' with $R(\sigma') = \{2\}$ is absorbing. From any other state σ_t with $R(\sigma_t) = \{1, 2\}$, there will be a pair of adjacent vertices $uv \in E$ such that $\sigma_t(u) \neq \sigma_t(v)$, since G is connected. Thus, there is a positive probability of transitioning to either

$$\sigma_{t+1} = \sigma_t[u \mapsto \sigma_t(v)] \tag{4}$$

$$\text{or} \quad \sigma_{t+1} = \sigma_t[v \mapsto \sigma_t(u)]. \tag{5}$$

Iteratively applying this argument, we can reach eventually with positive probability both the state σ with $R(\sigma) = \{1\}$ as well as the state σ with $R(\sigma) = \{2\}$. Hence, the Moran process is an absorbing Markov chain and we can characterise its state space by stating for any $\sigma \in \Omega$:

- σ is absorbing (and thereby recurrent) if $|R(\sigma)| = 1$,
- else, σ is transient.

Therefore, the Moran process will eventually reach an absorbing state, *i.e.*

$$\forall \sigma_0 : \quad \mathbb{P}\left[\lim_{t\to\infty} |R(\sigma_t)| = 1 \middle| \sigma_0\right] = 1 \tag{6}$$

Let us quickly sketch the proof of this claim.

Proof. The ergodic states are exactly those σ with $|R(\sigma)| = 1$. The claim follows from a basic argument about finite absorbing Markov chains, *e.g.* Theorem 3.1.1 in [25]. □

The question of interest is thus: What is the probability for the Moran process to converge to the "all-have-type-1" (extinction of the mutation) absorbing state and what is the probability to converge to the all-have-type-2 (fixation of the mutation) absorbing state. For the Moran process on a graph G with fitness vector $\boldsymbol{r}$, any state σ, and type $j \in \{1, 2\}$ we define

$$f^j_{G,\boldsymbol{r}}(\sigma) := \mathbb{P}\left[\lim_{t\to\infty} R(\sigma_t) = \{j\} \middle| \sigma_0 = \sigma\right] \tag{7}$$

to be the probability to reach the all-have-type-j absorbing state, given that we start in σ.

Lieberman, Hauert, and Nowak consider the process where all vertices are of the weaker type (type 1) initially, except for one uniformly at random chosen vertex $v \in V$ which has type 2 initially. In the context of cancer, that corresponds to one (random) cell mutating initially. The quantity of interest is then

$$f^j_{G,\boldsymbol{r}} := \frac{1}{|\Theta|} \sum_{\sigma\in\Theta} f^j_{G,\boldsymbol{r}}(\sigma) = \frac{1}{n} \sum_{\sigma\in\Theta} f^j_{G,\boldsymbol{r}}(\sigma) \tag{8}$$

where $\Theta := \left\{\sigma \in \Omega \,\middle|\, |\sigma^{-1}(2)| = 1\right\}$ is the set of all states σ with one type-2 vertex. By eq. (6), we have

$$\sum_{j\in\{1,2\}} f^j_{G,\boldsymbol{r}} = 1. \tag{9}$$

$f^2_{G,\boldsymbol{r}}$ is called the *fixation probability*, *i.e.* the probability for the mutation eventually occupying the whole graph. $f^1_{G,\boldsymbol{r}}$ is the *extinction probability*, *i.e.* the probability that the mutation goes extinct eventually. Note that eq. (9) yields immediately $f^1_{G,\boldsymbol{r}} = 1 - f^2_{G,\boldsymbol{r}}$.

On normalising the fitness r_1 The attentive reader will have realised that the fitness of type 1, r_1, is unnecessary. Since we are only interested in the relative fitness between both mutants, we might just as well fix $r_1 \equiv 1$. Indeed, mathematically, we have

$$\begin{aligned}
\mathbb{P}\left[\sigma_{t+1} = \sigma_t[v \mapsto 1] | \sigma_t\right] &= \sum_{u \in N(v) \cap \sigma_t^{-1}(1)} \frac{r_1}{W(\sigma_t) \cdot \mathsf{deg}(u)} && \text{(by eq. (1))} \\
&= \sum_{u \in N(v) \cap \sigma_t^{-1}(1)} \frac{r_1}{\left(r_1 \cdot |\sigma_t^{-1}(1)| + r_2 \cdot |\sigma_t^{-1}(2)|\right) \cdot \mathsf{deg}(u)} \\
& && \text{(by def. of } W(\sigma_t)\text{)} \\
&= \sum_{u \in N(v) \cap \sigma_t^{-1}(1)} \frac{1}{\left(1 \cdot |\sigma_t^{-1}(1)| + r_2/r_1 \cdot |\sigma_t^{-1}(2)|\right) \cdot \mathsf{deg}(u)} && (10)
\end{aligned}$$

Similarly, for eq. (2),

$$\mathbb{P}\left[\sigma_{t+1} = \sigma_t[v \mapsto 2] | \sigma_t\right] = \sum_{u \in N(v) \cap \sigma_t^{-1}(2)} \frac{r_2/r_1}{\left(1 \cdot |\sigma_t^{-1}(1)| + r_2/r_1 \cdot |\sigma_t^{-1}(2)|\right) \cdot \mathsf{deg}(u)}, \qquad (11)$$

while eq. (3) still captures the case of no state change.

Hence, we could have used simply the fitness vector $\boldsymbol{r}\,' := (1, r_2/r_1)$. This convention is used in previous literature [28, 14, 1]. However, when we generalise to the multi-type Moran process in the subsequent section, it is notationally clearer to assign each type without exception its own fitness and thereby to omit this normalisation. We thus stick to our definition of the fitness vector $\boldsymbol{r}$ (Definition 1), but note that it is possible to convert to the normalised notation as described above.

On the connectedness assumption We noted before that all graphs considered in this work are assumed to be connected. The motivation for this assumption is very simple: Consider a graph consisting of two connected components V_1, V_2. If the mutation starts initially in V_1, it has no chance of ever reaching V_2, since reproduction happens along edges. Hence, the mutation might fixate in V_1, but can never reach V_2 so that the final state might be a mixture between fixation on V_1 and extinction in V_2. For such disconnected graphs, however, we can just consider the Moran process on each connected component individually, so that focusing to *connected* graphs in this work poses no restriction at all.

With our notation-machinery set up for the 2-type Moran process, it is now an easy task to generalise to the k-type process for every $k \geq 2$.

1.1.2 The k-type Moran process

We now generalise the Moran process to k types, for every $k \geq 2$, corresponding to the notion of Clonal interference in evolutionary biology [34]. The generalisation can be stated compactly: Instead of two types, we have k types, each with their own fitness. Initially, we still start with all vertices having type 1 but instead of one random vertex mutating to type 2, we have now $k-1$ vertices mutate each to one of the types $2, \ldots, k$. This set of distinct vertices is chosen uniformly at random. Let us formalise this notion.

We have already defined both fitness vector $\boldsymbol{r}$ (Definition 1) and type mapping (Definition 2) for k types. The k-type Moran process with fitness $\boldsymbol{r}$ on graph $G = (V, E)$ of order n (with $2 \leq k \leq n$) is thus the Markov chain

$$M := \{\sigma_t \,| t \in \mathbb{N}_0\}$$

defined on the state space $\Omega := \{V \to [k]\}$. For the evolution, every vertex of type $j \in [k]$ is reproducing at rate r_j to copy its type to one of its neighbours:

$$\forall v \notin \sigma_t^{-1}(j): \qquad \mathbb{P}\left[\sigma_{t+1} = \sigma_t[v \mapsto j] | \sigma_t\right] = \sum_{u \in N(v) \cap \sigma_t^{-1}(j)} \frac{r_j}{W(\sigma_t)} \cdot \frac{1}{\mathsf{deg}(u)} \tag{12}$$

with $W(\sigma_t) := \sum_{j \in [k]} r_j \cdot |\sigma_t^{-1}(j)|$ being the population's overall fitness in state σ_t. The probability of remaining in the same state is

$$\mathbb{P}\left[\sigma_{t+1} = \sigma_t | \sigma_t\right] = 1 - \sum_{j \in [k]} \sum_{v \notin \sigma_t^{-1}(j)} \mathbb{P}\left[\sigma_{t+1} = \sigma_t[v \mapsto j] | \sigma_t\right]. \tag{13}$$

These transition probabilities are consistent with the 2-type Moran process, since eqs. (12) and (13) are equal to eqs. (1) to (3) for $k = 2$.

We have thus k-absorbing states: each $\sigma : V \mapsto [k]$ with $|\,\mathrm{R}(\sigma)\,| = 1$. The fixation probability of type $j \in [k]$ given that we start in some state σ is the probability of type j fixating, *i.e.* that eventually, all vertices have type j. The fixation probability in general is the average of this quantity over all possible starting states. Formally, we define the fixation probability as follows.

Definition 3 (Fixation probability)**.** For any graph G, k-type fitness vector $\boldsymbol{r}$, state σ, and type $j \in [k]$, we define the *fixation probability from state* σ

$$f^j_{G,\boldsymbol{r}}(\sigma) := \mathbb{P}\left[\lim_{t \to \infty} R(\sigma_t) = \{j\} \middle| \sigma_0 = \sigma\right].$$

The *fixation probability* is

$$f^j_{G,\boldsymbol{r}} := \frac{1}{|\Theta|} \sum_{\sigma \in \Theta} f^j_{G,\boldsymbol{r}}(\sigma) = \frac{(n-k+1)!}{n!} \sum_{\sigma \in \Theta} f^j_{G,\boldsymbol{r}}(\sigma) \quad \in [0,1],$$

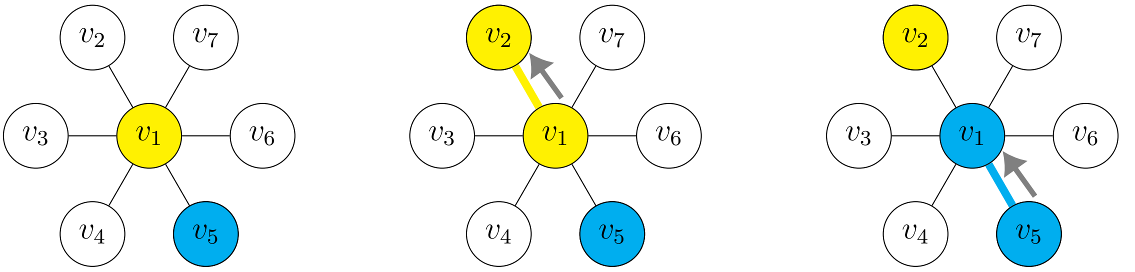
where $\Theta := \left\{\sigma \in \Omega \;\middle|\; \forall j \in [k] \setminus \{1\} : |\sigma^{-1}(j)| = 1\right\}$ is the set of all states with exactly one vertex per type $i \in [k] \setminus \{1\}$.

This definition is consistent with the definition of the fixation probability in previous literature for $k = 2$, where $f^2_{G,\boldsymbol{r}}$ is called the fixation probability (in our case, fixation of type 2) and $f^1_{G,\boldsymbol{r}}$ is called the extinction probability (in our case, fixation of type 1) [14, 22, 1]. As in the 2-type case, all fixation probabilities sum to 1: Since we will eventually reach one of the k absorbing states, we have

$$\sum_{j \in [k]} f^j_{G,\boldsymbol{r}} = 1. \tag{14}$$

We provide in Appendix D an efficient implementation of the k-type Moran process.

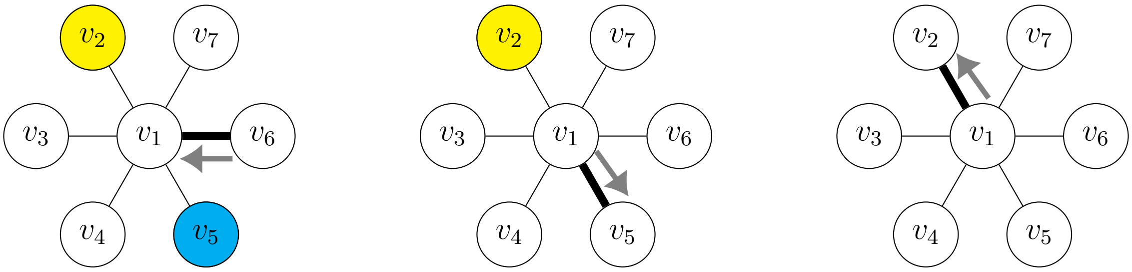
Example of the k-type Moran process Consider Figure 1 for an example of the $k = 3$-type Moran process on a star graph in Figure 1, where type 1 fixates after 5 steps.



(a) Time $t = 0$: Initially, all vertices are of type 1 (white) except for one type 2 vertex (yellow) and one type 3 vertex (blue).

(b) Time $t = 1$: Vertex v_1 gets chosen for reproduction to v_2. Hence, v_2 also becomes a type 2 vertex (yellow).

(c) Time $t = 2$: Vertex v_5 gets chosen for reproduction to its (only) neighbour v_1, which thus becomes a type 3 vertex (blue).



(d) Time $t = 3$: Vertex v_6 reproduces to v_1.

(e) Time $t = 4$: Vertex v_1 reproduces to v_5. Hence, the last type-3-vertex has died, so that type 3 is now extinct.

(f) Time $t = 5$: Vertex v_1 reproduces to v_2. All vertices are now of type 1, so that type 1 has fixated.

Figure 1: A possible evolution of the 3-type Moran process on the star graph. In this case, type 1 (white) fixates after 5 time steps.

1.2 Applicability of the theory of absorbing Markov chains

The overall goal is to compute the fixation probabilities. In the realm of absorbing Markov chains, this corresponds to finding the probability of reaching each of the absorbing states, given that we start in a certain initial state. More precisely, we can compute these quantities by computing the *fundamental matrix*. Let us enumerate the vertices in some order $v_1, \dots, v_n$ and represent every state $\sigma \in \Omega : V \to [k]$ by the vector $\boldsymbol{\sigma} := (\sigma(v_1), \dots, \sigma(v_n)) \in [k]^n$. If we denote by $\boldsymbol{j}$ the all-j-vector for $j \in [k]$ and enumerate the state space by $\mathbf{1}, \mathbf{2}, \dots, \boldsymbol{k}$ followed by all $n!/(n-k+1)!$ many states that have exactly one of each of the types $2, 3, \dots, k$, and then all other remaining states in any order, the $k^n \times k^n$ transition matrix simplifies to

$$P = \left(\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline R & Q \end{array}\right) \tag{15}$$

where $\mathbf{I}$ is the $k \times k$ identity and $\mathbf{0}$ the $k \times (k^n - k)$ 0 matrix. Both $R \in \mathbb{R}^{k^n-k,k}$ and $Q \in \mathbb{R}^{k^n-k,k^n-k}$ have to be computed based on the underlying topology represent by the graph. According to Kemeny and Snell [25], we can compute the matrix $B \in \mathbb{R}^{k^n-k,k}$ by

$$B = (I - Q)^{-1} R \tag{16}$$

where $(I-Q)^{-1}$ is the so-called fundamental matrix. B_{ij} is then the probability that the process starting in the i-th transient state[6] ends up in the j-th absorbing state (Theorem 3.3.7 in [25]).

[6] since the first k states are the absorbing states, this corresponds to the $i + k$-th state in our enumeration

We can now compute for every $j \in [k]$:

$$f^j_{G,\boldsymbol{r}} = \frac{(n-k+1)!}{n!} \sum_{i \in [n!/(n-k+1)!]} B_{ij}. \tag{17}$$

However, even in the simplest case of $k = 2$, this requires the inversion of a matrix whose size is exponential in n. Therefore, the existing approach on absorbing Markov chains is useless for large graphs, motivating the use of approximation schemes.

1.3 Computational problems

We now define the computational problems for calculating the fixation probabilities.

The first problem Moran-$\boldsymbol{r}$ assumes $\boldsymbol{r}$ to be constant, and thereby follows previous literature [14, 1]. We then regard the number of mutants k and the fitness vector $\boldsymbol{r}$ as parameters for p-Moran. We refine this problem to approximate only the strongest type's fixation probability in p-MoranMax and define the promised problem version p-Moran-c-ThFittest to approximate the fixation probability of sufficiently strong mutation types.

Formally, by denoting for any vector $\boldsymbol{v} \in \mathbb{R}^d$ by $|\boldsymbol{v}| := d$ its dimension, we define the following computational problems.[7]

Moran-$\boldsymbol{r}$ (where $\boldsymbol{r}$ is a fitness vector)	**Input:** Graph G and type $j \in \{2, \ldots, \lvert\boldsymbol{r}\rvert\}$. **Output:** $f^j_{G,\boldsymbol{r}}$.
p-MoranMax	**Input:** Graph G and fitness vector $\boldsymbol{r}$. **Parameter:** $\boldsymbol{r}$. **Output:** $f^k_{G,\boldsymbol{r}}$ where $k := \lvert\boldsymbol{r}\rvert$.
p-Moran	**Input:** Graph G, fitness vector $\boldsymbol{r}$, and type $j \in \{2, \ldots, \lvert\boldsymbol{r}\rvert\}$. **Parameter:** $\boldsymbol{r}$. **Output:** $f^j_{G,\boldsymbol{r}}$.
p-Moran-c-ThFittest (for some $c \in \mathbb{N}$)	**Input:** Graph G, fitness vector $\boldsymbol{r}$, and type $j \in \{\max(2, \lvert\boldsymbol{r}\rvert - c + 1), \ldots, \lvert\boldsymbol{r}\rvert\}$. **Parameter:** $\boldsymbol{r}$. **Output:** $f^j_{G,\boldsymbol{r}}$.

We will only attempt to approximate $f^j_{G,\boldsymbol{r}}$ for $j > 1$, since this quantity can be for $j = 1$ exponentially small in n, as it is the case for $k = 2$ on the complete graph [13].

Recall that a *Randomised Approximation Scheme* (RAS) for a Problem on input I with parameter P is a randomised algorithm A that takes as input an instance I and error tolerance $\varepsilon \in (0, 1)$ such that

$$(1 - \varepsilon)\textsc{Problem}(I) \leq A(I, \varepsilon) \leq (1 + \varepsilon)\textsc{Problem}(I) \tag{18}$$

holds with probability at least 3/4. Problem(I) denotes the solution of Problem on input I.

If the running time of algorithm A is bounded by $p_1(\lVert I\rVert) \cdot p_2(1/\varepsilon)$ for some polynomials p_1, p_2, we say that A is a *Fully Polynomial Randomised Approximation Scheme* (FPRAS), where $\lVert\cdot\rVert$ denotes the encoding length.

If the running time of algorithm A is bounded by $g(P) \cdot p_1(\lVert I\rVert) \cdot p_2(1/\varepsilon)$ for any computable function g and some polynomials p_1, p_2, we say that A is a *Fixed-Parameter Tractable Randomised Approximation Scheme* (FPTRAS). Choosing $\boldsymbol{r}$ as a parameter, is syntactic sugar since

---

[7] Recall that the fitness vector (Definition 1) is strictly increasing.

a parameterisation of a problem must be a computable function from the input to a *number*. In our case, the number we are interested in is actually $|\boldsymbol{r}| + \max_{j\in[k]} \frac{r_i}{r_i - r_{i-1}}$.

On using the parameterised complexity framework When providing an FPRAS for MORAN-$\boldsymbol{r}$, we will see later that its running time has a factor of $n^{\Theta(k)}$, *i.e.* it is exponential in k with base n (Section 3.1). Can we do better? Since we still assume that the number of mutants (k) is much smaller than the graph's order (n), we are interested in algorithms with running time of the form

$$g(k) \cdot \mathrm{poly}(n) \tag{19}$$

for any function $g(k)$, which is independent of n. Hence, we are interested in so-called *fixed-parameter tractable* algorithms due to Downey and Fellows [16] (see also the excellent textbook by Cygan *et al.* [12]). Since we are interested in *approximation* algorithms, we use the concept of Fixed-Parameter Tractable Randomised Approximation schemes, introduced by Arvind and Raman [2].

For the input, we assume the graph is encoded as an adjacency matrix and thus

$$\|G\| \in \Theta\left(n^2\right). \tag{20}$$

The computational problem considered in previous work is equivalent to MORAN-$\boldsymbol{r}$ for $\boldsymbol{r} \in \mathbb{R}^1$, since the fitness is assumed to be constant [13, 14, 1].

1.4 Results

For the problem MORAN-$\boldsymbol{r}$, *i.e.* where $\boldsymbol{r}$ is a constant, we provide an FPRAS.

Main Theorem 4. MORAN-$\boldsymbol{r}$ *has an FPRAS.*

Thus, we have a fully polynomial randomised approximation algorithm. Once we consider $\boldsymbol{r}$ to be non-constant, we loose the polynomial running time since this algorithm's running time is exponential in the dimension of $\boldsymbol{r}$, *i.e.* in k. Since we consider the number of mutants, k, to be much smaller than the graph's order, we turn to the parameterised problems and find an FPTRAS for P-MORANMAX.

Main Theorem 5. P-MORANMAX *has an FPTRAS.*

We then generalise this to the promised problem P-MORAN-c-THFITTEST.

Main Theorem 6. P-MORAN-c-THFITTEST *has an FPTRAS.*

The running times of the approximation algorithms for both P-MORANMAX and P-MORAN-c-THFITTEST are even independent of k, so that we get FPRASes under reasonable assumptions on the components of $\boldsymbol{r}$ (Corollary 33 and 36).

The key ingredient for both FPTRASes is a stochastic coupling relying on the continuous time generalisation of the Moran process (Theorem 26) which is interesting on its own since it allows to translate a phase transition in the $k = 2$ case to the strongest mutant for any $k \geq 2$ (Theorem 31) and yields the bounds provided by Ferreira and Neves recently [19] as a simple corollary, as we show in Appendix B.

When considering the special case of the complete graph K_n, we will see that the state space can be simplified to the evolution of a vector $\boldsymbol{S} \in (\mathbb{N}_0)^k$, denoted by $\{\boldsymbol{S}_t : t \geq 0\}$. Denoting the state changes by $\boldsymbol{X}_t$, we can define a function $f(\boldsymbol{S}_t)$ (Definition 37) on the state space which is a martingale:

Main Theorem 7. *The sequence* $\{f(\boldsymbol{S}_t) : t \geq 0\}$ *is a martingale with respect to the sequence* $\{\boldsymbol{X}_t : t \geq 0\}$*, where we define* $\boldsymbol{X}_0 \equiv \boldsymbol{0}$.

This resolves an open problem raised by Monk and Schaik.

1.5 Outline

This dissertation is organised as follows. In Section 2, we derive upper bounds on the expected absorption time for every mutant $j \in [k]$. We then derive lower bounds on the fixation probabilities in Section 3 by establishing a stochastic coupling and provide efficient approximation algorithms. In Section 4, we provide a martingale for the Moran process on the complete graph, resolving an open problem posed by Monk and Schaik. We conclude by motivating research into several further directions in Section 5.

2 Bounding the absorption time

In this section, we provide upper bounds on the expected absorption times (Definition 8) in the k-type Moran process.

We will first upper bound the expected absorption time for type k (Section 2.1) and then, inductively, upper bound the expected absorption time for any type $j < k$ (Section 2.2). We finally answer the question of whether simulating so-called *active steps* only, a practice employed successfully for the 2-type Moran process, leads with the present techniques to an improvement of the upper bound (Section 2.3).

Throughout this section, the fitness vector $\boldsymbol{r}$ is assumed to be non-decreasing, so that type k is the strongest type in the k-type Moran process.

2.1 Absorption time for mutant k

We now bound the absorption time of type k using a potential function technique, generalising an idea by Goldberg, Lapinskas, and Richerby [1] to the $k \geq 2$ case.

Formally, the absorption time of type $j \in [k]$ is the first time when type j either fixates or goes extinct:

Definition 8 (Absorption time)**.** Let $M = \{\sigma_t\}_{t \in \mathbb{N}_0}$ be the k-type Moran process. We define the stopping times

$$\tau_j := \min\left\{t \in \mathbb{N}_0 : \ \mathrm{R}\left(\sigma_t\right) = \{j\} \ \vee \ j \notin \mathrm{R}\left(\sigma_t\right)\right\} \tag{21}$$

for any type $j \in [k]$. We say τ_j is *the absorption* time of type j.

Definition 9. We define the potential function Ψ for any state $\sigma : V \to [k]$:

$$\Psi(\sigma) := \sum_{v \in \sigma^{-1}(k)} \frac{1}{\mathsf{deg}(v)}. \tag{22}$$

We further recall the definition of *drift* from [1]. While drift is not needed in our approximation algorithm results later, we include it in Lemma 11, so that the interested reader can easily compare it to the case of $k = 2$ mutants in previous literature (Lemma 8 in [1]).

Definition 10 (Drift, Definition 7 in [1])**.** For any graph $G = (V, E)$ and any pair of disjoint vertex sets $S_1 \sqcup S_2 \subseteq V$, their drift is defined by

$$\mathrm{Dr}\left(A, B\right) := \sum_{(x,y) \in E(A,B)} \frac{1}{\mathsf{deg}(x)\mathsf{deg}(y)}. \tag{23}$$

Lemma 11. *For the k-type Moran process on any connected graph $G = (V, E)$ with fitness $\boldsymbol{r}$ where $r_{k-1} < r_k$, we have*

$$\forall \sigma : V \to [k] \text{ s.t. } 0 < \left|\left\{\sigma^{-1}(k)\right\}\right| < n :$$
$$\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) | \sigma_t = \sigma\right] > \frac{r_k - r_{k-1}}{n r_k} \mathrm{Dr}\left(\sigma^{-1}(k), V \setminus \sigma^{-1}(k)\right) > \left(1 - \frac{r_{k-1}}{r_k}\right) n^{-3},$$

i.e. *for any state σ where type-k-fixation has not yet happened, $\Psi(\sigma)$ is strictly increasing in expectation with the given lower bound.*

Proof. Let G, σ, i be arbitrary as defined in the theorem. Denoting by $W(\sigma) > 0$ the state's overall fitness, we have

$$
\begin{aligned}
&\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) | \sigma_t = \sigma\right] && (24)\\
&= \sum_{\substack{xy\in E\\ \sigma(x)=k\\ \sigma(y)\neq k}} \mathbb{P}\left[x \text{ reproduces to } y\right] \left(\Psi(\sigma[y \mapsto k]) - \Psi(\sigma)\right) \\
&\qquad\qquad + \mathbb{P}\left[y \text{ reproduces to } x\right] \left(\Psi\left(\sigma[x \mapsto \sigma(y)]\right) - \Psi(\sigma)\right) && (25)\\
&= \sum_{\substack{xy\in E\\ \sigma(x)=k\\ \sigma(y)\neq k}} \frac{r_k}{W(\sigma)} \frac{1}{\mathsf{deg}(x)} \left(\Psi\left(\sigma[y \mapsto k]\right) - \Psi(\sigma)\right) + \frac{r_{\sigma(y)}}{W(\sigma)} \frac{1}{\mathsf{deg}(y)} \left(\Psi\left(\sigma\left[x \mapsto \sigma(y)\right]\right) - \Psi(\sigma)\right) && (26)\\
&= \frac{1}{W(\sigma)} \sum_{\substack{xy\in E\\ \sigma(x)=k\\ \sigma(y)\neq k}} \frac{r_k}{\mathsf{deg}(x)\mathsf{deg}(y)} - \frac{r_{\sigma(y)}}{\mathsf{deg}(x)\mathsf{deg}(y)} && (27)\\
&= \frac{1}{W(\sigma)} \sum_{\substack{xy\in E\\ \sigma(x)=k\\ \sigma(y)\neq k}} \left(r_k - r_{\sigma(y)}\right) \frac{1}{\mathsf{deg}(x)\mathsf{deg}(y)} && (28)\\
&\geq \frac{r_k - r_{k-1}}{W(\sigma)} \sum_{\substack{xy\in E\\ \sigma(x)=k\\ \sigma(y)\neq k}} \frac{1}{\mathsf{deg}(x)\mathsf{deg}(y)} && (r_{\sigma(y)} \leq r_{k-1})\\
&> \frac{r_k - r_{k-1}}{n r_k} \sum_{\substack{xy\in E\\ \sigma(x)=k\\ \sigma(y)\neq k}} \frac{1}{\mathsf{deg}(x)\mathsf{deg}(y)} && (W(\sigma) < n r_k)\\
&\geq \frac{r_k - r_{k-1}}{n r_k} \operatorname{Dr}\left(\sigma^{-1}(k), V \setminus \sigma^{-1}(k)\right) && (29)\\
&> \left(1 - \frac{r_{k-1}}{r_k}\right) n^{-3}. && (30)
\end{aligned}
$$

Where eq. (25) follows since any change to $\sigma^{-1}(k)$ and thereby any change to $\Psi(\sigma)$ comes from a vertex reproducing along an edge of the cut set $E\left(\sigma^{-1}(k), V \setminus \sigma^{-1}(k)\right)$. Since $0 < |\sigma^{-1}(k)| < n$, this set is indeed non-empty. Eq. (30) follows since the aforementioned cut set is non-empty and since each vertex degrees is less than n. ☐

Upper bounding $\mathbb{E}\left[\tau_k\right]$

The technique used to upper bound $\mathbb{E}\left[\tau_k\right]$ relies on a technique called *drift analysis* used to bound the expected absorption time of supermartingales, introduced by Hajek [24]. Goldberg, Lapinskas, and Richerby applied these techniques to study the 2-type process. While generalising their analysis to the k-type process, we will rely on Lemma 45 in [1], which we recall here with notation adapted to this work, to make references in the following proof easier.

Lemma 12 (Lemma 45 in [1])**.** *Let $M := \{\sigma_t \mid t \in \mathbb{N}_0\}$ be a Markov chain with finite state space Ω. Let $c_1, c_2 > 0$, let $\Psi : \Omega \to \mathbb{R}_{\geq 0}$ be a function and $\tau \geq 0$ be a stopping time with $\tau \leq \min\{t \mid \Psi(\sigma_t) = 0 \vee \Psi(\sigma_t) \geq c_1\}$. Suppose that:*

- *(i) from every state $S_1 \in \Omega$ with $0 < \Psi(S_1) < c_1$, there exists a path in M from S_1 to some state S_2 with $\Psi(S_2) = 0$ or $\Psi(S_2) \geq c_1$;*

(ii) for all $t \geq 0$, if $\Psi(\sigma_t) < c_1$, then $\Psi(\sigma_{t+1}) \leq c_1 + 1$; and

(iii) for all $t \geq 0$ and all $\sigma \in \Omega$ such that the events $\tau > t$ and $\tau_i = \tau$ are consistent, we have $\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) \mid \sigma_t = \sigma\right] \geq c_2$.

Then we have

$$\mathbb{E}\left[\tau\right] \leq (c_1 - \Psi(\sigma_0) + 1)/c_2.$$

Since we have already upper bounded the expected drift of the potential function Ψ in Lemma 11, we can now apply Lemma 12 to upper bound $\mathbb{E}\left[\tau_k\right]$.

Corollary 13. *For the k-type Moran process with $r_{k-1} < r_k$ on any graph $G = (V, E)$ of order n and any initial state $\sigma_0 : V \to [k]$, we have for the absorption time of type k:*

$$\mathbb{E}\left[\tau_k\right] < \frac{r_k}{r_k - r_{k-1}}\left(n+1\right)n^3, \tag{31}$$

Proof. Let $M := \{\sigma_t | i \in \mathbb{N}_0\}$ be the Markov chain corresponding to the k-type Moran process with state space Ω. Note that we can rewrite τ_k (Definition 8) by

$$\tau_k := \min\left\{t : \Psi(\sigma_t) = 0 \;\vee\; \Psi(\sigma_t) = c_1\right\} \tag{32}$$

where $c_1 := \sum_{v \in V} 1/\mathsf{deg}(v)$. If σ_0 is absorbing, $\tau_k = 0$ and the lemma holds trivially.

Otherwise, not that by definition of the Moran process, Ω is finite. Note also that Ψ is non-negative for every state. We now show that all three requirements $(i) - (iii)$ of the supermartingale lemma (Lemma 45 in [1]) are satisfied.
For (i), suppose $0 < \Psi(\sigma) < c_1$ for some $\sigma \in \Omega$. Since any absorbing state σ' fulfils $\Psi(\sigma') \in \{0, c_1\}$, we know that σ cannot be absorbing. Since $0 < \Psi(\sigma)$, there exists at least one type-k vertex. Since G is connected, this vertex can reproduce with positive probability to all of its neighbours, which in turn reproduce to their neighbours, and so forth until the entire graph is of type k. Call that new state σ^*. Then, $\Psi(\sigma^*) = \sum_{v \in V} 1/\mathsf{deg}(v) = c_1$ by definition of c_1.
For (ii), observe that the maximum potential is c_1.
For (iii), we have by Lemma 11

$$\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) | \sigma_t = \sigma\right] > c_2 > 0. \tag{33}$$

where $c_2 := \left(1 - \frac{r_{k-1}}{r_k}\right) n^{-3}$. We get by Lemma 45 of [1]:

$$\begin{aligned} \mathbb{E}\left[\tau_k\right] &\leq \frac{c_1 - \Psi\left(\sigma_0\right) + 1}{c_2} && (34) \\ &< \frac{r_k}{r_k - r_{k-1}}\left(n+1\right)n^3, && (35) \end{aligned}$$

where the last step follows since $c_1 < n$ — recall that G is connected and thus does not contain isolated vertices. □

We now apply Markov's inequality to turn the upper bound on $\mathbb{E}\left[\tau_k\right]$ from Corollary 13 into an upper bound on τ_k being much larger than its expectation in Corollary 14.

Corollary 14. *For the k-type Moran process with $r_{k-1} < r_k$ on any graph $G = (V, E)$ or order n and any initial state $\sigma_0 : V \to [k]$, we have for the absorption time of type k:*

$$\forall N \geq 1 : \quad \mathbb{P}\left[\tau_k \geq N\right] < \frac{r_k}{N\left(r_k - r_{k-1}\right)}\left(n+1\right)n^3. \tag{36}$$

Note that Corollary 14 yields only useful results for $N > \mathbb{E}\left[\tau_k\right]$.

2.2 Absorption time for any mutant j

After having bounded the expected absorption time for the strongest mutant, we want to extend our results to any mutant j. In order to get polynomial running times of the approximation algorithms defined in Section 3, we aim for bounds polynomial in n and k.

Recalling that the absorption time of any type $j \in [k]$ is the time when type j either fixated or went extinct (Definition 8), we can describe our idea is as follows: Since we already have an upper bound for the expected absorption time of the strongest mutant, we know that after time τ_k, either type k fixated (so that types $1, \dots, k-1$ went extinct and have thus been absorbed by this time) or type k went extinct. In that case, there is some other mutant type $j \in [k-1]$ the strongest mutant from time τ_k onwards. We can thus re-apply our bond on the absorption time of the strongest mutant, which is now type j. Using linearity of expectation makes our idea mathematically sound.

We define the following stopping time to turn our aforementioned inductive idea into a rigorous proof.

Definition 15 (τ_j^{ind})**.** For any type $j \in [k]$ in the k-type Moran process $\{\sigma_t \mid t \in \mathbb{N}_0\}$, we define

$$\tau_j^{\mathsf{ind}} := \min\Big\{t \in \mathbb{N}_0 : \ \mathrm{R}(\sigma_t) \cap \{j, \dots, k\} = \emptyset \ \vee \ \mathrm{R}(\sigma_t) \in \big\{\{j\}, \dots, \{k\}\big\}\Big\}.$$

τ_j^{ind} is thus the first point in time when either all types $j, \dots, k$ are extinct or one of them has fixated. For the special case $j = k$, we have $\tau_k^{\mathsf{ind}} = \tau_k$.

Lemma 16. *τ_j^{ind} is monotone decreasing in j,* i.e.

$$\forall j \in [k-1]: \quad \tau_j^{ind} \geq \tau_{j+1}^{ind}. \tag{37}$$

Proof. For any $j \in [k-1]$, let $\tau_j^{\mathsf{ind}} = t$ for some $t \in \mathbb{N}_0$ and any Moran process $\{\sigma_t \mid t \in \mathbb{N}_0\}$. We must have by Definition 15 either of the following cases.

Case i) $\mathrm{R}(\sigma_t) \cap \{j, \dots, k\} = \emptyset$. Thus, $\mathrm{R}(\sigma_t) \cap \{j+1, \dots, k\} = \emptyset$.

Case ii) $\mathrm{R}(\sigma_t) = \{j\}$. Thus, $\mathrm{R}(\sigma_t) \cap \{j+1, \dots, k\} = \emptyset$.

Case iii)$\mathrm{R}(\sigma_t) \in \big\{\{j+1\}, \dots, \{k\}\big\}$**.**

In all cases,

$$\mathrm{R}(\sigma_t) \cap \{j+1, \dots, k\} = \emptyset \ \vee \ \mathrm{R}(\sigma_t) \in \big\{\{j+1\}, \dots, \{k\}\big\} \tag{38}$$

holds. Since τ_{j+1} is by Definition 15 the minimum time such that Equation (38) holds, we conclude $\tau_{j+1}^{\mathsf{ind}} \leq t = \tau_j^{\mathsf{ind}}$. ☐

With the monotonicity of τ_j^{ind} proved formally, we can now conduct our induction proof to upper bound $\mathbb{E}\left[\tau_j^{\mathsf{ind}}\right]$. We will subsequently turn this into an upper bound on $\mathbb{E}[\tau_j]$ (Corollary 18).

Theorem 17. *For every k-type Moran process $\{\sigma_t \mid t \in \mathbb{N}_0\}$ with strictly increasing fitness vector $\boldsymbol{r}$ on any graph $G = (V, E)$ of order n, we have for any type $j \in [k]$:*

$$\forall \text{ initial states } \sigma_0 : V \to [k]. \quad \mathbb{E}\left[\tau_j^{ind}\right] \leq \sum_{i=j}^{k} \frac{r_i}{r_i - r_{i-1}} (n+1) n^3.$$

Proof by induction. **I.H.(j):**

$$\forall \text{ initial states } \sigma_0 : V \to [k]. \quad \mathbb{E}\left[\tau_j^{\mathsf{ind}}\right] \le \sum_{i=j}^{k} \frac{r_i}{r_i - r_{i-1}}(n+1)n^3.$$

B.C. (j=k): Let $\sigma_0 : V \to [k]$ be arbitrary. We apply Corollary 13 to prove the claim

$$\begin{aligned} \mathbb{E}\left[\tau_k^{\mathsf{ind}}\right] &= \mathbb{E}\left[\tau_k\right] && \text{(by Definition 15)} \\ &\le \frac{r_k}{r_k - r_{k-1}}(n+1)n^3. && \text{(by Corollary 13)} \end{aligned}$$

I.S. $(1 \le j \le k-1)$: Assume $I.H.(j+1)$. Let $\sigma_0 : V \to [k]$ be arbitrary. We have

$$\begin{aligned} \mathbb{E}\left[\tau_j^{\mathsf{ind}}\right] &= \mathbb{E}\left[\tau_j^{\mathsf{ind}} - \tau_{j+1}^{\mathsf{ind}} + \tau_{j+1}^{\mathsf{ind}}\right] && (39) \\ &= \mathbb{E}\left[\tau_j^{\mathsf{ind}} - \tau_{j+1}^{\mathsf{ind}}\right] + \mathbb{E}\left[\tau_{j+1}^{\mathsf{ind}}\right]. && \text{(l.o.e.)} \end{aligned}$$

For the first part, we get

$$\begin{aligned} & \mathbb{E}\left[\tau_j^{\mathsf{ind}} - \tau_{j+1}^{\mathsf{ind}}\right] && (40) \\ = & \sum_{\substack{\sigma:V\to[k] \\ j\in \mathrm{R}(\sigma)}} \mathbb{E}\left[\tau_j^{\mathsf{ind}} - \tau_{j+1}^{\mathsf{ind}} \mid \sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right] \mathbb{P}\left[\sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right] && (41) \\ + & \sum_{\substack{\sigma:V\to[k] \\ j\notin \mathrm{R}(\sigma)}} \underbrace{\mathbb{E}\left[\tau_j^{\mathsf{ind}} - \tau_{j+1}^{\mathsf{ind}} \mid \sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right]}_{=0} \mathbb{P}\left[\sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right] && (42) \end{aligned}$$

where eq. (42) is 0 since if $j \notin \mathrm{R}\left(\sigma_{\tau_{j+1}}^{\mathsf{ind}}\right)$, we must have $\tau_j^{\mathsf{ind}} = \tau_{j+1}^{\mathsf{ind}}$ by Definition 15.

In eq. (41), we note that for $j \in \sigma_{\tau_{j+1}^{\mathsf{ind}}}$, we must have $\mathrm{R}\left(\sigma_{\tau_{j+1}^{\mathsf{ind}}}\right) \cap \{j+1, \ldots, k\} = \emptyset$. We can thus simplify eq. (41) to eq. (44):

$$\begin{aligned} & \sum_{\substack{\sigma:V\to[k] \\ j\in \mathrm{R}(\sigma)}} \mathbb{E}\left[\tau_j^{\mathsf{ind}} - \tau_{j+1}^{\mathsf{ind}} \mid \sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right] \mathbb{P}\left[\sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right] && (43) \\ = & \sum_{\substack{\sigma:V\to[j] \\ j\in \mathrm{R}(\sigma)}} \mathbb{E}\left[\tau_j^{\mathsf{ind}} - \tau_{j+1}^{\mathsf{ind}} \mid \sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right] \mathbb{P}\left[\sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right] && (44) \\ = & \sum_{\substack{\sigma:V\to[j] \\ j\in \mathrm{R}(\sigma)}} \mathbb{E}\left[\tau_j' \mid \sigma_0' = \sigma\right] \mathbb{P}\left[\sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right] && (45) \\ \le & \frac{r_j}{r_j - r_{j-1}}(n+1)n^3 \underbrace{\sum_{\substack{\sigma:V\to[j] \\ j\in \mathrm{R}(\sigma)}} \mathbb{P}\left[\sigma_{\tau_{j+1}^{\mathsf{ind}}} = \sigma\right]}_{\le 1} && (46) \\ \le & \frac{r_j}{r_j - r_{j-1}}(n+1)n^3. && (47) \end{aligned}$$

Since j is the strongest type in $\sigma : V \to [j]$, we obtain eq. (45) for $M' := \{\sigma_t' \mid t \in \mathbb{N}_0\}$ being the Moran process on G with j types and fitness vector consisting of the first j entries of $\boldsymbol{r}$. Hence, Corollary 13 applies on M', yielding eq. (46).

Hence,

$$\begin{aligned}
\mathbb{E}\left[\tau_j^{\mathsf{ind}}\right] &= \mathbb{E}\left[\tau_j^{\mathsf{ind}} - \tau_{j+1}^{\mathsf{ind}}\right] + \mathbb{E}\left[\tau_{j+1}^{\mathsf{ind}}\right] && \text{(l.o.e.)}\\
&\leq \frac{r_j}{r_j - r_{j-1}}(n+1)n^3 + \mathbb{E}\left[\tau_{j+1}^{\mathsf{ind}}\right] && \text{(by eq. (47))}\\
&\leq \frac{r_j}{r_j - r_{j-1}}(n+1)n^3 + \sum_{i=j+1}^{k} \frac{r_i}{r_i - r_{i-1}}(n+1)n^3 && (I.H.(j+1))\\
&= \sum_{i=j}^{k} \frac{r_i}{r_i - r_{i-1}}(n+1)n^3, && (48)
\end{aligned}$$

concluding our proof.

□

Corollary 18. *For the k-type Moran process with strictly increasing fitness vector $\boldsymbol{r}$ and graph $G = (V, E)$ of order n, we have for any type $j \in [k]$:*

$$\forall \sigma : V \to [k]. \quad \mathbb{E}\left[\tau_j\right] \leq \sum_{i=j}^{k} \frac{r_i}{r_i - r_{i-1}}(n+1)n^3.$$

Proof. Let $j \in [k]$, $\sigma : V \to [k]$ be arbitrary. By Definition 15, we have $\tau_j \leq \tau_j^{\mathsf{ind}}$. Therefore,

$$\mathbb{E}\left[\tau_j\right] \leq \mathbb{E}\left[\tau_j^{\mathsf{ind}}\right] \leq \sum_{i=j}^{k} \frac{r_i}{r_i - r_{i-1}}(n+1)n^3, \tag{49}$$

by Theorem 17. □

We have thus established an upper bound on $\mathbb{E}\left[\tau_j\right]$ for any $j \in [k]$. This bound is very useful for designing *efficient* approximation algorithms later: Notably, this bound depends only polynomially on n, so that increasing the graph's size increases the expected absorption times at most polynomially, which we can exploit later when constructing a fully *polynomial* approximation scheme (Section 3.1). If we impose reasonable conditions on the fitnesses r_i for any $i \in \{j-1, \dots, k\}$, the bound is also polynomial in k, a property we will exploit when considering the *parameterised* problems (Sections 3.2 and 3.3).

2.3 On the advantage of active-step-only simulation

The content of this section is a remark on the applicability of a possible method to speed up the process and as such is not needed for later results. Thus, the reader is invited to read the present section but may skip to Section 3 without compromising their understanding of the approximation algorithms.

Since not every transition changes the Moran process's state,[8] it is natural to simulate only those transitions that do so. For the $k = 2$ process, the potential Ψ changes if and only if the process's state changes. However, this is not the case for $k > 2$: If σ_t transitions to a different state σ_{t+1} but the set of k-type vertices remains unchanged ($\sigma_t^{-1}(k) = \sigma_{t+1}^{-1}(k)$), then the potential Ψ does not change. Hence, a change in the process's state does not always translate to a change in the potential Ψ for $k > 3$.

---

[8] Whenever u reproduces to w with $\sigma_t(u) = \sigma_t(w)$, we have $\sigma_t = \sigma_{t+1}$.

Can we asymptotically improve the lower bound on the expected drift of Ψ (Lemma 11), by simulating only the active steps? For the $k = 2$ process with the present choice of Ψ (Definition 9) as potential function, this is possible as shown in previous work [1, 10]. For $k > 2$ and the given potential function, we answer this question to the negative in this section. This might seem surprising at first in the light of Lemma 20.

We first define *active steps* more formally.

Definition 19 (Active steps)**.** For a Moran process $M = \{\sigma_t \mid t \in \mathbb{N}_0\}$, the process

$$M^{\mathsf{active}} := \{\sigma_t \in M \mid \ \sigma_t \neq \sigma_{t-1}, t \in \mathbb{N}\}$$

is called the *active-steps-only Moran process.* We enumerate its states by $0, 1, 2, \ldots$ yielding

$$M^{\mathsf{active}} = \{\sigma_l^{\mathsf{active}} \mid l \in \mathbb{N}_0\}.$$

Since M will eventually converge, M^{active} is finite. The steps of M^{active} are referred to as the *active steps of* M.

Lemma 20. *For the k-type Moran process executing only active steps on any graph $G = (V, E)$ with maximum degree Δ, minimum degree δ, and $r_{k-1} < r_k$, we have*

$$\begin{aligned}
&\forall \sigma : V \to [k] \text{ s.t. } 0 < \left|\left\{\sigma^{-1}(k)\right\}\right| < n.\ \forall t \in \mathbb{N}_0 : \\
\frac{r_k - r_{k-1}}{(r_k + r_{k-1})\Delta} \leq &\mathbb{E}\left[\Psi(\sigma_{t+1}^{\mathsf{active}}) - \Psi(\sigma_t^{\mathsf{active}}) \;\middle|\; \left(\sigma_t^{\mathsf{active}}\right)^{-1}(k) \neq \left(\sigma_{t+1}^{\mathsf{active}}\right)^{-1}(k) \ \wedge \ \sigma_t^{\mathsf{active}} = \sigma\right] \\
\leq &\frac{r_k - r_1}{(r_k + r_1)\delta}.
\end{aligned}$$

Since $\Delta < n$, this would be a quadratic improvement in n compared to Lemma 11, if every active step changed the set $\sigma^{-1}(k)$. We will first prove Lemma 20 and then show why we cannot use this apparent speedup in general.

Proof of Lemma 20. Let G, σ be arbitrary as defined in the lemma. We follow the proof structure of Lemma 67 in [1]. By definition of active steps, it suffices to prove that for all $t \geq 0$,

$$\begin{aligned}
\frac{r_k - r_{k-1}}{(r_k + r_{k-1})\Delta} \leq &\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) \;\middle|\; (\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k) \ \wedge \ \sigma_t = \sigma\right] \\
\leq &\frac{r_k - r_1}{(r_k + r_1)\delta}.
\end{aligned}$$

Let $i \geq 0$ be arbitrary. We have for $C := E\left(\sigma_t^{-1}(k), V \setminus \ \sigma_t^{-1}(k)\right)$

$$\begin{aligned}
&\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) \;\middle|\; (\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k) \ \wedge \ \sigma_t = \sigma\right] &(50)\\
&= \sum_{xy \in C} \mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) \;\middle|\; \mathcal{E}_{xy} \ \wedge \ \sigma_t = \sigma\right] \cdot \mathbb{P}\left[\mathcal{E}_{xy} \;\middle|\; (\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k) \ \wedge \ \sigma_t = \sigma\right] &(51)\\
&= \sum_{xy \in C} \left(\frac{r_k/\mathsf{deg}(x)}{r_k/\mathsf{deg}(x) + r_{\sigma(y)}/\mathsf{deg}(y)} \frac{1}{\mathsf{deg}(y)} - \frac{r_{\sigma(y)}/\mathsf{deg}(y)}{r_k/\mathsf{deg}(x) + r_{\sigma(y)}/\mathsf{deg}(y)} \frac{1}{\mathsf{deg}(x)}\right) \\
&\qquad \cdot \mathbb{P}\left[\mathcal{E}_{xy} \;\middle|\; (\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k) \ \wedge \ \sigma_t = \sigma\right] &(52)\\
&= \sum_{xy \in C} \frac{r_k - r_{\sigma(y)}}{r_k \mathsf{deg}(y) + r_{\sigma(y)} \mathsf{deg}(x)} \mathbb{P}\left[\mathcal{E}_{xy} \;\middle|\; (\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k) \ \wedge \ \sigma_t = \sigma\right], &(53)
\end{aligned}$$

where Equation (51) follows since $(\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k)$ iff $\bigcup_{xy\in C}\mathcal{E}_{xy}$ and since $(\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k)$ implies $\sigma_{t+1} \neq \sigma_t$.

Since for all $xy \in C$, we have

$$\frac{r_k - r_{k-1}}{(r_k + r_{k-1})\,\Delta} \leq \frac{r_k - r_{\sigma(y)}}{r_k \mathsf{deg}(y) + r_{\sigma(y)}\mathsf{deg}(x)} \leq \frac{r_k - r_1}{(r_k + r_1)\,\delta}\,, \tag{54}$$

and since

$$\sum_{xy\in C} \mathbb{P}\left[\mathcal{E}_{xy} \,\middle|\, (\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k) \ \wedge \ \sigma_t = \sigma\right] = 1, \tag{55}$$

we can conclude

$$\frac{r_k - r_{k-1}}{(r_k + r_{k-1})\Delta} \leq \mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) \,\middle|\, (\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k) \ \wedge \ \sigma_t = \sigma\right] \leq \frac{r_k - r_1}{(r_k + r_1)\delta}. \tag{56}$$

□

We noted that for Lemma 20 to improve our bound, we would need that every active step changes the set $\sigma^{-1}(k)$ with sufficiently high probability. However, this is not the case. We can even construct explicitly a family of graphs for which concentrating on active steps only will not give an (asymptotic) better bound on the expected change of Ψ.

Lemma 21. *For every $k \geq 3$, there is an infinite family of graphs $\mathcal{G}$ such that there exists for every $G \in \mathcal{G}$ a state σ with $0 < |\{\sigma^{-1}(k)\}| < n$ and*

$$\mathbb{P}\left[(\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k) \,\middle|\, \sigma_t \neq \sigma_{t+1} \ \wedge \ \sigma_t = \sigma\right] < \frac{9(r_2 + r_k)}{r_1 + r_2}\frac{1}{n^2},$$

in the k-type Moran process with fitness vector $\boldsymbol{r}$, where n denotes the order of G.[9]

Proof. Let $k \geq 3$ be arbitrary and consider the k-type Moran process. For any $n \geq 4$ such that $l := \frac{n+1}{3} \in \mathbb{N}$, we define $G_n := (V, E)$ to be the graph consisting of the vertex sets $V := S_1 \sqcup S_2 \sqcup S_3$ of size $|S_1| = l$, $|S_2| = l$, and $|S_3| = l - 1$. Its edge set E consists of the edges of the complete bipartite graph between S_1, S_2, the edges of the complete graph on S_3, and a single edge connecting some vertex $x \in S_2$ to some vertex $y \in S_3$, see Figure 2. We define the state $\sigma : V \to [k]$ as:

- $\sigma(v) = 1 \quad \forall v \in S_1$
- $\sigma(v) = 2 \quad \forall v \in S_2$
- $\sigma(v) = k \quad \forall v \in S_3$.

To simplify notation, we define for any $j \in [k]$ the cut set $C_j \subseteq E$ by

$$C_j := E\left(\sigma^{-1}(j), \left(\sigma^{-1}(j)\right)^C\right). \tag{57}$$

Since C_k is the singleton $\{xy\}$, we know that the set $\sigma^{-1}(k)$ changes iff reproduction happens from x to y or vice-versa. We also know that (for $\sigma = \sigma_t$) the event $\sigma_t \neq \sigma_{t+1}$ happens iff reproduction happens along any edge in $\bigcup_{j\in[k]} C_j$.

---

[9] We recall that all graphs in this work are connected. We thus show that there exists an infinite family of connected graphs.

Denoting by W the population fitness in state $\sigma_t = \sigma$ and by v the vertex chosen for reproduction to another vertex w, we have

$$\mathbb{P}\left[\{v, w\} \in E(S_1, S_2)\right] = \frac{r_1|S_1| + r_2(|S_2| - \frac{1}{|S_1|+1})}{W}, \tag{58}$$

where we have the term $|S_2| - \frac{1}{|S_1|+1}$ since every vertex in $S_2 - x$ has only edges to S_1, and x has only one neighbour not in S_1, which is chosen with probability $1/\mathsf{deg}(x) = 1/(|S_1| + 1)$ in case $v = x$.
Further, we have

$$\mathbb{P}\left[\{v, w\} = \{x, y\}\right] = \frac{r_2\frac{1}{|S_1|+1} + r_k\frac{1}{|S_3|+1}}{W}. \tag{59}$$

Since

$$\bigcup_{j\in[k]} C_j = E(S_1, S_2) + \{x, y\}, \tag{60}$$

we have

$$\begin{aligned}
&\mathbb{P}\left[(\sigma)_t^{-1}(k) \neq (\sigma)_{t+1}^{-1}(k) \;\middle|\; \sigma_t \neq \sigma_{t+1} \,\wedge\, \sigma_t = \sigma\right] && (61)\\
&= \mathbb{P}\left[vw \in C_k \;\middle|\; vw \in \bigcup_{j\in k} C_j\right] && (62)\\
&= \mathbb{P}\left[\{v, w\} = \{x, y\} \;\middle|\; \{v, w\} \in E(S_1, S_2) + \{x, y\}\right] && \text{(using eq. (60))}\\
&= \frac{r_2\frac{1}{|S_1|+1} + r_k\frac{1}{|S_3|+1}}{r_1|S_1| + r_2|S_2| + r_k\frac{1}{|S_3|+1}} && \text{(using eqs. (58) and (59))}\\
&< \frac{r_2 + r_k}{r_1 + r_2}\frac{1}{l^2} && \text{(by choice of } |S_1|, |S_2|, |S_3|)\\
&< \frac{9(r_2 + r_k)}{r_1 + r_2}\frac{1}{n^2}. && \text{(by definition of } l)
\end{aligned}$$

We can therefore define our infinite family of graphs as

$$\mathcal{G} := \{G_n \,|n \geq 5 \;\wedge n \mod 3 = 2\}. \tag{63}$$

□

Corollary 22. *There exists an infinite family of graphs $\mathcal{G}$ such that for any $G \in \mathcal{G}$ of order $n \geq 8$, there is a state σ with $0 < |\{\sigma^{-1}(k)\}| < n$, such that for any $t \geq 0$*

$$\mathbb{E}\left[\Psi(\sigma_{t+1}^{\mathsf{active}}) - \Psi(\sigma_t^{\mathsf{active}}) \;\middle|\; \sigma_t^{\mathsf{active}} = \sigma\right] \leq \frac{27(r_k - r_1)(r_2 + r_k)}{(r_k + r_1)(r_1 + r_2)} \cdot \frac{1}{n^2(n-5)}.$$

Proof. It suffices to prove that for any $t \geq 0$:

$$\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) \;\middle|\; \sigma_t \neq \sigma_{t+1} \,\wedge\, \sigma_t = \sigma\right] \leq \frac{27(r_k - r_1)(r_2 + r_k)}{(r_k + r_1)(r_1 + r_2)} \cdot \frac{1}{n^2(n-5)}. \tag{64}$$

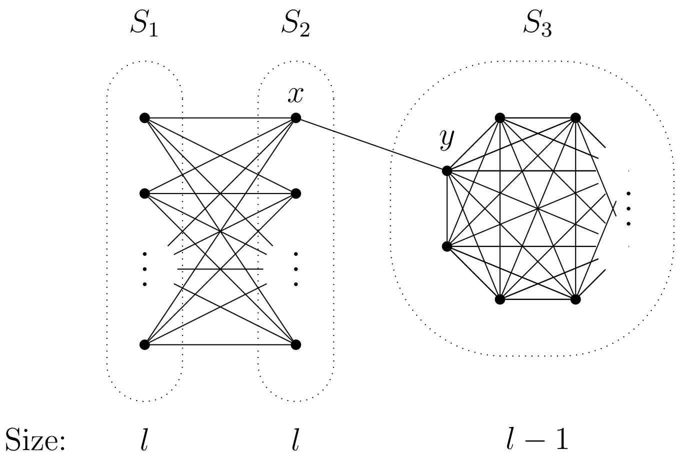


Figure 2: The graph $G_n \in \mathcal{G}$, where $l := \frac{n+1}{3}$.

Let $t \geq 0$ be arbitrary. We take $\mathcal{G}$ and σ as in Lemma 21 (see Figure 2) and have for any $G_n \in \mathcal{G}$

$$\begin{aligned}
&\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) \;\middle|\; \sigma_t \neq \sigma_{t+1} \;\wedge\; \sigma_t = \sigma\right] && (65)\\
=&\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) \;\middle|\; \sigma_t^{-1}(k) \neq \sigma_{t+1}^{-1}(k)\; \sigma_t \neq \sigma_{t+1} \;\wedge\; \sigma_t = \sigma\right]\\
&\quad\cdot \mathbb{P}\left[\sigma_t^{-1}(k) \neq \sigma_{t+1}^{-1}(k) \;\middle|\; \sigma_t \neq \sigma_{t+1} \;\wedge\; \sigma_t = \sigma\right] && (66)\\
&+ \underbrace{\mathbb{E}\left[\Psi(\sigma_{t+1}) - \Psi(\sigma_t) \;\middle|\; \sigma_t^{-1}(k) = \sigma_{t+1}^{-1}(k)\; \sigma_t \neq \sigma_{t+1} \;\wedge\; \sigma_t = \sigma\right]}_{=0}\\
&\quad\cdot \mathbb{P}\left[\sigma_t^{-1}(k) = \sigma_{t+1}^{-1}(k) \;\middle|\; \sigma_t \neq \sigma_{t+1} \;\wedge\; \sigma_t = \sigma\right] && (67)\\
\leq&\frac{r_k - r_1}{(r_k + r_1)\cdot\delta} \cdot \frac{9(r_2 + r_k)}{r_1 + r_2}\frac{1}{n^2} && \text{(by Lemmas 20 and 21)}\\
<&\frac{27(r_k - r_1)(r_2 + r_k)}{(r_k + r_1)(r_1 + r_2)} \cdot \frac{1}{n^2(n-5)} && \text{(since } \delta(G_n) = \tfrac{n-5}{3} \text{ and } n \geq 8)
\end{aligned}$$

where we get the "$= 0$" part in eq. (67) since the potential Ψ changes iff $\sigma^{-1}(k)$ changes (Definition 9).

□

We thus have by Corollary 22 an *upper bound* of $\Theta\left(\frac{1}{n^3}\right)$ on the expected drift of the active-steps only Moran process. Since Lemma 11 gave a $\Theta\left(\frac{1}{n^3}\right)$ *lower bound* on the expected drift of Ψ, we can conclude that simulating active steps only does not improve our bound on the expected absorption time asymptotically. We have thus shown that with the present techniques, it its not possible to get an asymptotically better bound on the expected absorption time by active-step-only-simulation for $k > 3$, contrasting the $k = 2$ case [1, 10].

We emphasise that this does not rule out the existence of a better bound using different techniques or a novel potential function. We will comment further on this when discussing future research in Section 5.

3 Fixation probability bounds and algorithms

Since we try to approximate the fixation probabilities $f^j_{G,\boldsymbol{r}}$ for $j \in \{2, \dots, k\}$ within a *factor* of $(1 \pm \varepsilon)$ with high probability,[10] the task becomes harder the smaller $f^j_{G,\boldsymbol{r}}$ is. To that end, the goal of this section is to provide lower bounds on the fixation probabilities $f^j_{G,\boldsymbol{r}}$ and to provide efficient approximation algorithms. In doing so, we first derive a rough lower bound in Section 3.1. This bound is enough to provide an FPRAS for Moran-$\boldsymbol{r}$, which we recall from Section 1.3 to be defined for *fixed*, increasing $\boldsymbol{r}$ as

Moran-$\boldsymbol{r}$
(where $\boldsymbol{r}$ is a fitness vector)
Input: Graph G and type $j \in \{2, \dots, |\boldsymbol{r}|\}$.
Output: $f^j_{G,\boldsymbol{r}}$.

However, its running time which is exponential in k due to a factor of $n^{\Theta(k)}$, can be significantly improved, allowing us to tackle harder problems: In Section 3.2, establish a method to reduce to the 2-mutant case. This reduction is made possible by a stochastic coupling (Theorem 26) which provides insight into how the process behaves if the fitnesses and initial state changes and is therefore interesting for its own sake. The coupling allows us to provide good lower bounds on $f^k_{G,\boldsymbol{r}}$ and thereby to give an FPTRAS for p-MoranMax. We then use the coupling to improve our bounds for every type $j \in \{2, \dots, k\}$ in Section 3.3, allowing for an FPTRAS for p-Moran-c-ThFittest, the promise problem of p-Moran. We show that under reasonable assumptions on the fitness vector $\boldsymbol{r}$, both FPTRASes become even FPRASes, showing that they are still efficient for large k.

3.1 A generous lower bound on the fixation probability

In this section, we first give a rough lower bound on $f^j_{G,\boldsymbol{r}}$ for any type $j \in \{2, \dots, k\}$. Since this lower bound is for fixed $\boldsymbol{r}$ of the form $\frac{1}{\text{poly}(n)}$ we can use it subsequently for providing an efficient FPRAS for Moran-$\boldsymbol{r}$. Recall that all graphs considered in this work are connected.

Lemma 23. *For every type $j \in \{2, \dots, k\}$ in the k-type Moran process with strictly increasing fitness vector $\boldsymbol{r}$ on any graph G of maximum degree Δ, we have*

$$f^j_{G,\boldsymbol{r}} > (k-1)! \left(\frac{r_j}{r_k}\right)^{k-2} \cdot \left(\frac{1}{\Delta}\right)^{k-2} \frac{\log n}{n^{2k-3}} \tag{68}$$

Proof. Let $j \in \{2, \dots, k\}$ be arbitrary. Recall from Definition 3 that

$$f^j_{G,\boldsymbol{r}} := \frac{(n-k+1)!}{n!} \sum_{\sigma \in \Theta} f^j_{G,\boldsymbol{r}}(\sigma),$$

where $\Theta := \left\{\sigma \in \Omega \;\middle|\; \forall j \in \{2, \dots, k\} : |\sigma^{-1}(j)| = 1\right\}$. We define the set

$$\Theta' := \left\{\sigma \in \Theta \;\middle|\; G\left[\sigma^{-1}(\{2, \dots, k\})\right] \text{ is connected}\right\}. \tag{69}$$

For any $\sigma' \in \Theta'$ let $S := \sigma'^{-1}(\{j, \dots, k\})$. Since $G[S]$ is connected, there must be a spanning tree T in $G[S]$. We root this tree at v_j, where $v_j \in S$ is the unique vertex having type $\sigma'(v_j) = j$.

---

[10] To be concise, this statement holds by using a basic probabilistic amplification technique: By our definition of FPRAS, the approximation should be within $(1 \pm \varepsilon) \cdot f^j_{G,\boldsymbol{r}}$ with probability at least 3/4. However, by taking the Median of sufficiently many approximations, we get the *with high probability* statement. $\mathcal{O}(\log(1/\delta))$ repetitions are enough to amplify the success probability to be at least $1 - \delta$, see *e.g.* Theorem 4.3.4 in [21].

If the first $k-2$ edges chosen for reproduction form some tree traversal of T starting at its roots v_j, then after $k-2$ steps, all vertices in S have type j, and all other vertices in $V \setminus S$ have type 1. Call this state σ^*. The probability to reach this σ^* from σ' within these first $k-2$ steps is at least

$$\left(\frac{r_j}{r_k \cdot n \cdot \Delta}\right)^{k-2}, \tag{70}$$

which follows by a rough upper bound that the population's overall fitness is at most $r_k \cdot n$ at any point in time.

We have further

$$f^j_{G,\boldsymbol{r}}(\sigma^*) = f^j_{G,\boldsymbol{r}^+}(\sigma^+), \tag{71}$$

where $\boldsymbol{r}^+ \in (\mathbb{R}^+)^2$, $\sigma^+ : V \to [2]$ are obtained from $\boldsymbol{r}$, σ^* by remapping[11] type j to type 2 and rescaling $r_1^+ := 1$, $r_2^+ := \frac{r_j}{r_1}$. Since $j \geq 2$, we have $r_2^+ > 1$. We are thus in the 2-mutant case and get by applying Corollary 8 in [14] (subset domination) and Theorem 2 in [1]

$$f^j_{G,\boldsymbol{r}^+}(\sigma^+) \geq \frac{\log n}{n}. \tag{72}$$

We can conclude therefore

$$\begin{aligned}
f^j_{G,\boldsymbol{r}} &:= \frac{(n-k+1)!}{n!} \sum_{\sigma \in \Theta} f^j_{G,\boldsymbol{r}}(\sigma) && \text{(by Definition 3)} \\
&\geq \frac{(n-k+1)!}{n!} \sum_{\sigma' \in \Theta'} f^j_{G,\boldsymbol{r}}(\sigma') && \text{(since } \Theta' \subseteq \Theta\text{)} \\
&\geq \frac{(n-k+1)!}{n!} \sum_{\sigma' \in \Theta'} \left(\frac{r_j}{r_k \cdot n \cdot \Delta}\right)^{k-2} \cdot \frac{\log n}{n} && \text{(by eqs. (70) to (72))} \\
&= \frac{(n-k+1)!}{n!} (n-k+2)(k-1)! \left(\frac{r_j}{r_k}\right)^{k-2} \left(\frac{1}{\Delta}\right)^{k-2} \cdot \frac{\log n}{n^{k-1}} && (73) \\
&> (k-1)! \left(\frac{r_j}{r_k}\right)^{k-2} \left(\frac{1}{\Delta}\right)^{k-2} \frac{\log n}{n^{2k-3}}, && (74)
\end{aligned}$$

where the penultimate step follows by bounding the cardinality of Θ' as follows: There are at least $n - |S| + 1$ many connected subsets of V of size $|S|$, since G is connected.[12] How many $\sigma' \in \Theta'$ are there for a given S such that $S = \sigma'^{-1}(\{2, \ldots, k\})$ (recall that $|S| = k-1$)? There

---

[11] This is indeed well-defined since all vertices in state σ^* are either of type 1 or of type j, and thus the remapping yields a 2-type process.

[12] We sketch a constructive proof of this claim. Extend the spanning tree T of $G[S]$ to a spanning tree T' of G. For each vertex $v \in V \setminus S$, there is a unique path P_v from T to v via T', by the spanning tree property. $S \cup V(P_v)$ is thus a unique set for any $v \in V \setminus S$. We delete vertices from $S \cup V(P_v)$ in decreasing order (if vertices have the same distance, take any of them) of their shortest path distance to v in T, until we obtain a set of size $|S|$, call this set S_v. Uniqueness: If we deleted all vertices of S, uniqueness follows by uniqueness of P_v and our order of deletion. Otherwise, the set contains P_v as well as some other vertices of S. No other S_v contains P_v but not $V \setminus (P_v \cup S)$, and thus uniqueness follows. By our order of deletion, the set must be connected. Thus, there are at least $n - |S| + 1$ connected sets of size $|S|$ in G.

are $|S|!$ many permutations for assigning to each of the types $2, \dots, k$ a single vertex in S. We can therefore conclude that

$$\begin{aligned} |\Theta'| &\geq (n - |S| + 1) \cdot |S|! \qquad (75) \\ &= (n - k + 2)(k - 1)! \,. \qquad (76) \end{aligned}$$

□

We define Algorithm 1 to simulate the Moran process, depending on two functions N, T which determine the number of simulations as well as the maximal number of simulation steps depending on the input. We hide the data structure related details for the sake of clearness, but will elaborate on them in the running time analysis (Lemma 24).

Algorithm 1 (N, T): Simulating the Moran process, where N, T are functions on $(0,1) \times \mathbb{N} \times \mathbb{R}^* \times \mathbb{N} \to \mathbb{N}$

```
Input: Graph $G$, fitness vector $\boldsymbol{r}$, type $j \in \{2, \dots, k\}$, and error tolerance $\varepsilon \in (0,1)$, where $n := |G|$, $k := |\boldsymbol{r}|$
Output: Estimate for $f^j_{G,\boldsymbol{r}}$
count ← 0
$\hat{N} \leftarrow N(\varepsilon, n, \boldsymbol{r}, j)$                ▷ Calculate the number of runs
$\hat{T} \leftarrow T(\varepsilon, n, \boldsymbol{r}, j)$                ▷ Calculate the number of steps per run
perform at most $\hat{T}$ simulation steps of
    repeat $\hat{N}$ times
        Sample $\{v_2, \dots, v_k\} \subset V$ u.a.r.
        $\sigma_0[v_i] \leftarrow i \quad \forall i \in \{2, \dots, k\}$
        $\sigma_0[v] \leftarrow 1 \quad \forall i \in V \setminus \{v_2, \dots, v_k\}$
        Simulate $\{\sigma_i\}$ step by step until fixation
        if $\mathrm{R}(\sigma_{\hat{T}}) = \{j\}$ then
            count ← count + 1
        end if
    end
simulation end
if simulation was interrupted before completion then
    return “error”
end if
return count/$\hat{N}$
```

We now analyse the running time of Algorithm 1.

Lemma 24. *If $\hat{N} \leq \hat{T}$ and N, T are time-constructible, Algorithm 1 has running time*

$$\mathcal{O}\left(T(\varepsilon, n, \boldsymbol{r}, j) \cdot k \cdot \log n + n^2\right),$$

where $k := |\boldsymbol{r}|$.

Proof. We first preprocess the graph to get an adjacency array per vertex in time $\mathcal{O}(n^2)$. We then maintain a self-balancing search tree T_j per type $j \in [k]$ (for $k := |\boldsymbol{r}|$), which contains the set $\sigma^{-1}(j)$ in some order, as well as counters for $|\sigma^{-1}(j)|$.

Each step of the Moran process can be simulated in time $\mathcal{O}(k \cdot \log n)$: For sampling the vertices v, w where v shall reproduce to w, we first sample the type of v, where the probability for every type $j \in [k]$ is proportional to r_j and $|\sigma^{-1}(j)|$. For the sampled j, we sample u.a.r.

an index $i \in [|\sigma^{-1}(j)|]$. This is possible in time $\mathcal{O}(k)$. The i-th vertex on T_j is our chosen v, which we can access in the balanced binary tree T_j in time $\mathcal{O}(\log |T_i|) \subseteq \mathcal{O}(\log n)$. Sampling $w \in N(v)$ u.a.r. works via the adjacency array of v in time $\mathcal{O}(1)$. Updating the tree data structures of w's old and new type takes, including rebalancing the trees, time $\mathcal{O}(\log n)$ in *e.g.* AVL trees [11].

Since Algorithm 1 simulates at most $T(\varepsilon, n, \boldsymbol{r}, j)$ steps, we can conclude that Algorithm 1 has an overall running time of

$$\mathcal{O}\left(T(\varepsilon, n, \boldsymbol{r}, j) \cdot k \cdot \log n + n^2\right) , \tag{77}$$

concluding our proof.

We note that our analysis is independent of $N(\varepsilon, n, \boldsymbol{r}, j)$, since the code part involving $\hat{N}$ (Line 5) is inside the part which executes at most $\hat{T}$ simulation steps (Lines 4 to 14). By requiring $\hat{N} \le \hat{T}$, and by assuming time-constructibility of both functions, we can ensure that the computation of $\hat{N}$, $\hat{T}$ in Lines 2 and 3 works in the required running time. ☐

With our lower bound on $f^j_{G,\boldsymbol{r}}$ from Lemma 23 and upper bound on $\mathbb{E}[\tau_j]$ from Corollary 18, we can now provide an FPRAS for Moran-$\boldsymbol{r}$.

Main Theorem 4 (restated). Moran-$\boldsymbol{r}$ *has an FPRAS.*

Proof. We show that Algorithm 1 with the functions[13]

$$N(\varepsilon, n, \boldsymbol{r}, j) := \left\lceil 2^{-1}\varepsilon^{-2}\left(\frac{r_k}{r_j}\right)^{2k-4} \frac{1}{((k-1)!)^2}\frac{n^{6k-10}}{\log^2(n)}\ln(16)\right\rceil$$

$$T(\varepsilon, n, \boldsymbol{r}, j) := \left\lceil \sum_{i=j}^{k} \frac{r_i}{r_i - r_{i-1}}\right\rceil (n+1)n^3 8\left(N(\varepsilon, n, \boldsymbol{r}, j)\right)^2 \qquad \text{(where } k := |\boldsymbol{r}|\text{)}$$

is an FPRAS for Moran-$\boldsymbol{r}$. We abbreviate the values of those functions by $\hat{N}, \hat{T}$, as in the algorithm.

We first show correctness. Note first that both terms $\hat{N}, \hat{T}$ are well-defined, since by Definition 1, $\boldsymbol{r}$ is strictly increasing and all components are non-zero.
We first analyse Lines 5 to 13 as if they were not interrupted. We define X_i to be the indicator for the i-th simulation reaching type-j-fixation and

$$p := \frac{1}{N(\varepsilon, n, \boldsymbol{r}, j)} \sum_{i=0}^{N(\varepsilon,n,\boldsymbol{r},j)} X_i, \tag{78}$$

so that we have $\mathbb{E}[p] = f^j_{G,\boldsymbol{r}}$. By a Chernoff bound we get for any $\varepsilon > 0$

$$\mathbb{P}\left[\left|p - f^j_{G,\boldsymbol{r}}\right| > \varepsilon f^j_{G,\boldsymbol{r}}\right] \le 2\exp\left(-2\varepsilon^2\left(f^j_{G,\boldsymbol{r}}\right)^2 \hat{N}\right) \le \frac{1}{8} \tag{79}$$

where we used in the last step that $f^j_{G,\boldsymbol{r}} > (k-1)!\left(\frac{r_j}{r_k}\right)^{k-2} \cdot \left(\frac{1}{\Delta}\right)^{k-2} \frac{\log n}{n^{2k-3}}$ (Lemma 23) and the definition of $\hat{N}$.

---

[13] Note that $N(\varepsilon, n, \boldsymbol{r}, j), T(\varepsilon, n, \boldsymbol{r}, j)$ depend also on type j which we do not use in the definition of N, T in this theorem. In later theorems, however, this parameter will be used. Including j in the signature of N, T already here allows for more concise statements later.

Further, by Corollary 18,

$$\mathbb{E}\left[\tau_j\right] \leq \sum_{i=j}^{k} \frac{r_i}{r_i - r_{i-1}}(n+1)n^3, \tag{80}$$

Therefore, the probability that type j has not reach fixation in a single simulation (Line 9) by time $\hat{T}/\hat{N}$ is

$$\mathbb{P}\left[\tau_j \geq \hat{T}/\hat{N}\right] \leq \frac{\mathbb{E}\left[\tau_j\right]}{\hat{T}/\hat{N}} \leq \frac{1}{8\hat{N}}. \tag{81}$$

The probability that all simulations together take more than $\hat{T}$ steps is bounded by the probability that any of the $\hat{N}$ many simulations takes more than $\hat{T}/\hat{N}$ steps, which is by a union bound at most $\frac{1}{8}$.

Therefore, the probability that the simulation is interrupted and "error" is returned is at most $\frac{1}{8}$. Given that the simulation is not interrupted, the probability of returning a wrong numerical error is at most $\frac{1}{8}$ by Section 3.1. We have thus by a union bound[14] that with probability at least $\frac{3}{4}$, the algorithm returns a value within a factor $1 \pm \varepsilon$ of $f^j_{G,\boldsymbol{r}}$. It remains to show that the running time is bounded by a polynomial in the size of the input and $1/\varepsilon$. Recall that by definition of Moran-$\boldsymbol{r}$, $\boldsymbol{r}$ and thereby $k := |\boldsymbol{r}|$ is constant. By Lemma 24, the running time is in

$$\mathcal{O}\left(T(\varepsilon, n, \boldsymbol{r}, j) \cdot k \cdot \log n + n^2\right). \tag{82}$$

We have

$$T(\varepsilon, n, \boldsymbol{r}, j) \in \mathcal{O}\left(\frac{n^{12k-16}}{\log^4 n}\right) \cdot \varepsilon^{-4}, \tag{83}$$

so that the overall running time is polynomially in n and $1/\varepsilon$ under the assumption of $k := |\boldsymbol{r}|$ being constant. □

If there is an FPRAS for a particular problem, then the probability of the approximation algorithm being within a $1 \pm \varepsilon$ factor can be amplified from $\frac{3}{4}$ to $1 - 1/\delta$ for every $\delta > 0$ by increasing the running time growing only by a polynomially factor in $\log 1/\delta$. For the sake of illustration, we state in Corollary 25 the result of this amplification technique (see *e.g.* Theorem 4.3.4 in [21]) for the FPRAS from Main Theorem 4.

Corollary 25. *For every $\varepsilon, \delta > 0$, there is an Algorithm that approximates* Moran-$\boldsymbol{r}$ *within a factor $(1 \pm \varepsilon)$ with probability at least*

$$1 - 1/\delta.$$

The algorithm's running time is polynomial in $1/\varepsilon$, $\log 1/\delta$, and $\|I\|$ where I is the input.

---

[14] More formally, we denote by $\mathcal{A}$ the event that the simulation gets interrupted and by $\mathcal{B}$ the event that the output is a numerical value which is outside of our error tolerance. We have $\mathbb{P}[\mathcal{A}] \leq 1/8$. Further, $\mathbb{P}\left[\mathcal{B} \mid \overline{\mathcal{A}}\right] \leq 1/8$. Since $\mathcal{B} \subseteq \overline{\mathcal{A}}$, we have $\mathbb{P}[\mathcal{B}] \leq \mathbb{P}\left[\mathcal{B} \mid \overline{\mathcal{A}}\right]$ so that we can union bound $\mathbb{P}[\mathcal{A} \vee \mathcal{B}] \leq \mathbb{P}[\mathcal{A}] + \mathbb{P}[\mathcal{B}] \leq 1/4$.

3.2 Improving the lower bound on f^k: Reducing to the 2-type case

The lower bound of Lemma 23 is exponential in $k := |\boldsymbol{r}|$ with base $1/n$, rendering it useless when k becomes large. In particular, we need to improve this bound to approximate the parametrised problems efficiently. We start with the easiest case, namely the fixation probability of the strongest mutant.[15] We hope to achieve a lower bound which is not exponentially small in k with base $1/n$.

We improve the lower bound on $f^k_{G,\boldsymbol{r}}$ by reducing more efficiently to the 2-mutant case. For doing so, we need to transition to the continuous time analogue of the k-type Moran process. Intuitively, the necessity stems from us being able to regard each vertex reproducing at its own speed in the continuous time process, whereas vertices would need to "arrange among themselves" who is to reproduce at the next time step in the discrete time process. For an extensive discussion and a concrete example for the necessity of continuous time, see Diaz *et al*. [14]. The coupling connects the k-type Moran process to previous literature and thereby provides great insight into the process, such as a monotonicity result (Corollary 27). With the help of our coupling, we then present an analogue of the phase transition in the 2-type case of $f^2_{G,(1,r)}$ proved by Goldberg, Lapinskas, and Richerby [1], by proving a phase transition of $f^k_{G,\boldsymbol{r}}$ in the monotonicity of $\boldsymbol{r}$. With these improved bounds, we provide an FPTRAS for p-MoranMax and give sufficient conditions on r_{k-1}, r_k to obtain an FPRAS.

3.2.1 The continuous time Moran process

The continuous Time Moran process $\tilde{M} = \{\tilde{\sigma}_t | t \in \mathbb{R}_{\geq 0}\}$,[16] as considered by Diaz *et al*. [14], is the Markov process obtained if we equip each $v \in V$ with a "clock" $t_v \sim \operatorname{Exp}\left(r_{\tilde{\sigma}(v)}\right)$ where $r_{\tilde{\sigma}(v)}$ is the vertex's fitness. All exponential distributions are independent. The minimum of all $dt := \min_{v\in V} t_v$ gets to reproduce. The motivation for taking the exponential distribution is twofold: First, it is memoryless, allowing us to provide an inductive argument in our coupling. Second, we have[17] for every $v \in V$

$$\mathbb{P}\left[dt = t_v\right] = \frac{r_{\tilde{\sigma}(v)}}{\sum_{u\in V} r_{\tilde{\sigma}(u)}} \tag{84}$$

which is the probability of v being chosen for reproduction in the discrete case. Therefore, the discrete time Moran process is indeed the embedded discrete time Markov chain of this continuous process. Once a vertex is chosen to reproduce, it will reproduce to one of its neighbours, selected uniformly at random. Equivalently, we could say reproduction from any vertex v to any $u \in N(v)$ happens at rate $r_{\tilde{\sigma}(v)}/\mathsf{deg}(v)$.

With the continuous time process introduces, we now prove a general coupling result.

Theorem 26 (Stochastic Domination, superset)**.** *For any $k \geq 2$ and $x, y \in [k]$ with $x \leq y$, let $\tilde{M} := \{\tilde{\sigma}_t \mid t \in \mathbb{R}_{\geq 0}\}$ be a k-type continuous time Moran process with non-decreasing fitness vector $\boldsymbol{r}$ on any graph $G = (V, E)$. Let $\tilde{M}' := \{\tilde{\sigma}'_t \mid t \in \mathbb{R}_{\geq 0}\}$ be any other k-type continuous time Moran process on G with the unordered fitness vector $\boldsymbol{r}'$ such that $\forall i \in \{x, \dots, y\} : r'_i = r_i$, $\forall i \in [k] \setminus \{x, \dots, y\} : r'_i \geq r_l$ where $l := k-1$ if $x = k$ and $l := k$ otherwise. Visually, this*

---

[15] Intuitively, since it is the *strongest* type, we expect its fixation probability to be the highest among types $2, \dots, k$.

[16] The symbol ˜ denotes that we talk about continuous time.

[17] Recall that, by essential properties of the exponential distribution, the minimum of a set S of exponential distributions is exponentially distributed with parameter being the sum of the parameters in S.

translates to the following two cases, depending on whether $x < k$ or $x = k$.

$$\boldsymbol{r}' = \begin{pmatrix} r'_1 \\ \vdots \\ r'_{x-1} \\ r'_x \\ \vdots \\ r'_y \\ r'_{y+1} \\ \vdots \\ r'_k \end{pmatrix} \begin{array}{l} \geq r_k \\ \vdots \\ \geq r_k \\ = r_x \qquad (x < k) \\ \vdots \\ = r_y \\ \geq r_k \\ \vdots \\ \geq r_k \end{array} \qquad \boldsymbol{r}' = \begin{pmatrix} r'_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ r'_{k-1} \\ r'_x \end{pmatrix} \begin{array}{l} \geq r_{k-1} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \geq r_{k-1} \\ = r_x \qquad (x = k). \end{array} \tag{85}$$

Then there exists a coupling between $\tilde{M}$ and $\tilde{M}'$ such that at any point in time $t \in \mathbb{R}_{\geq 0}$,

$$\forall j \in \{x, \ldots, y\} : \; \tilde{\sigma}_t^{-1}(j) \supseteq \tilde{\sigma}_t'^{-1}(j) \tag{86}$$

implies

$$\forall j \in \{x, \ldots, y\} : \; \tilde{\sigma}_{t+dt}^{-1}(j) \supseteq \tilde{\sigma}_{t+dt}'^{-1}(j), \tag{87}$$

where dt is the time it takes for the first reproduction to happen in $\tilde{M}$ or $\tilde{M}'$ after time t.

Proof. We construct this coupling explicitly. Assume eq. (86) holds. We will construct a coupling such that eq. (87) holds. We slightly abuse notation and denote by $r_v := r_{\tilde{\sigma}_t(v)}$, $r'_v := r'_{\tilde{\sigma}'_t(v)}$ the fitnesses of v in $\tilde{M}$, $\tilde{M}'$ at time t.

We define the random variables

$$\begin{aligned} \forall v \in V : \quad & t_v \sim \operatorname{Exp}\left(\min(r_v, r'_v)\right) && (88) \\ \forall v \in V \text{ with } r_v > r'_v : \quad & t_v^1 \sim \operatorname{Exp}\left(r_v - r'_v\right) && (89) \\ \forall v \in V \text{ with } r_v < r'_v : \quad & t_v^2 \sim \operatorname{Exp}\left(r'_v - r_v\right), && (90) \end{aligned}$$

where each random variable can be seen as an exponentially distributed clock. All these random variables form the "clock set"

$$C := \{t_v | v \in V\} \cup \{t_v^1 | r_v > r'_v\} \cup \{t_v^2 | r_v < r'_v\}. \tag{91}$$

The "earliest clock" is the random variable

$$dt := \min C. \tag{92}$$

We evolve the two chains as follows, where we denote throughout this proof by v the vertex chosen for reproduction and by w the vertex which v reproduces to.

If $\tau = t_v$ for some $v \in V$: Select a neighbour $w \in N(v)$ u.a.r. and evolve the chains synchronously, *i.e.*

$$\begin{aligned} \tilde{\sigma}_{t+dt} &:= \tilde{\sigma}_t[w \mapsto \tilde{\sigma}_t(v)] && (93) \\ \tilde{\sigma}'_{t+dt} &:= \tilde{\sigma}'_t[w \mapsto \tilde{\sigma}'_t(v)]. && (94) \end{aligned}$$

If $\tau = t_v^1$ for some $v \in V$: Thus, we must have $r_v > r'_v$. Select a neighbour $w \in N(v)$ u.a.r. and evolve only $\tilde{M}$, *i.e.*

$$\begin{aligned} \tilde{\sigma}_{t+dt} &:= \tilde{\sigma}_t[w \mapsto \tilde{\sigma}_t(v)] && (95) \\ \tilde{\sigma}'_{t+dt} &:= \tilde{\sigma}'_t. && (96) \end{aligned}$$

If $\tau = t_v^2$ for some $v \in V$: Thus, we must have $r_v < r'_v$. Select a neighbour $w \in N(v)$ u.a.r. and evolve only $\tilde{M}'$, *i.e.*

$$\begin{aligned}\tilde{\sigma}_{t+dt} &:= \tilde{\sigma}_t && (97)\\ \tilde{\sigma}'_{t+dt} &:= \tilde{\sigma}'_t[w \mapsto \tilde{\sigma}'_t(v)]. && (98)\end{aligned}$$

We now show that eq. (87) indeed holds. Suppose first that both chains evolve synchronously. It suffices to show that for all $j \in \{x, \dots, y\}$ (i) if $v \in \tilde{\sigma}_t'^{-1}(j)$ reproduces to w, then also $w \in \tilde{\sigma}_{t+dt}^{-1}(j)$ as well as (ii) if $w \notin \tilde{\sigma}_{t+dt}^{-1}(j)$ gets replaced, then also $w \notin \tilde{\sigma}_{t+dt}'^{-1}(j)$.
Let $j \in \{x, \dots, y\}$ be arbitrary. For (i), suppose $v \in \tilde{\sigma}_t'^{-1}(j)$ reproduces to w. Since $\tilde{\sigma}_t^{-1}(j) \supseteq \tilde{\sigma}_t'^{-1}(j)$ by assumption (eq. (86)), we must have $v \in \tilde{\sigma}_t^{-1}(j)$. Since $\tilde{M}$ evolves as well, we have $w \in \tilde{\sigma}_{t+dt}^{-1}(j)$, concluding case (i). For (ii), suppose $w \notin \tilde{\sigma}_{t+dt}^{-1}(j)$ gets replaced. Thus, by the evolution of $\tilde{M}$, we must have $v \notin \tilde{\sigma}_t^{-1}(j)$. Since $\tilde{\sigma}_t^{-1}(j) \supseteq \tilde{\sigma}_t'^{-1}(j)$ (eq. (86)), we must have $v \notin \tilde{\sigma}_t'^{-1}(j)$. Since $\tilde{M}'$ evolves as well, this means $w \notin \tilde{\sigma}_{t+dt}'^{-1}(j)$.

Suppose now that the chains do not evolve synchronously. We define for any type $i \in [k]$ the sets

$$\begin{aligned}S_{i,<} &:= \{v \in V : \tilde{\sigma}_t(v) = i \wedge r_v < r'_v\} && (99)\\ S_{i,>} &:= \{v \in V : \tilde{\sigma}_t(v) = i \wedge r_v > r'_v\} && (100)\\ S_{i,=} &:= \{v \in V : \tilde{\sigma}_t(v) = i \wedge r_v = r'_v\}, && (101)\end{aligned}$$

which form a partition of V:

$$V = \bigsqcup_{i \in [k]} S_{i,<} \sqcup S_{i,>} \sqcup S_{i,=}. \tag{102}$$

We perform a case distinction on the vertex v chosen for reproduction. Since we assume that the chains evolve asynchronously, we must have $v \notin S_{i,=}$.

Case $v \in S_{i,<}$ for $i \in \{1, \dots, x-1\}$ if $x > 1$. Since $r_v < r'_v$ and since we assume that only one chain evolves, only $\tilde{M}'$ evolves. Since $\tilde{\sigma}_t(v) = i < x$, we have $v \notin \tilde{\sigma}_t^{-1}(\{x, \dots, y\})$. By eq. (86), this implies $v \notin \tilde{\sigma}_t'^{-1}(\{x, \dots, y\})$. We can conclude for every $j \in \{x, \dots, y\}$

$$\begin{aligned}\tilde{\sigma}_{t+dt}^{-1}(j) &= \tilde{\sigma}_t^{-1}(j) && \text{(since } \tilde{M} \text{ does not evolve)}\\ &\supseteq \tilde{\sigma}_t'^{-1}(j) && \text{(by assumption (86))}\\ &\supseteq \tilde{\sigma}_{t+dt}'^{-1}(j). && \text{(since } v \notin \tilde{\sigma}_t'^{-1}(\{x, \dots, y\}) \text{ reproduces)}\end{aligned}$$

Case $v \in S_{i,<}$ for $i \in \{x, \dots, y\}$. Thus, only $\tilde{M}'$ evolves and $r_v = r_i < r'_v$. We must have $v \notin \tilde{\sigma}_t'^{-1}(\{x, \dots, y\})$, for if we had $v \in \tilde{\sigma}_t'^{-1}(j)$ for some $j \in \{x, \dots, y\}$, we had by eq. (86) $v \in \tilde{\sigma}_t^{-1}(j)$, which implies $r_v = r'_v$ by our definition of $\boldsymbol{r}'$, contradicting $r_v < r'_v$.
We can therefore conclude

$$\begin{aligned}\tilde{\sigma}_{t+dt}^{-1}(j) &= \tilde{\sigma}_t^{-1}(j) && \text{(since } \tilde{M} \text{ does not evolve)}\\ &\supseteq \tilde{\sigma}_t'^{-1}(j) && \text{(by assumption (86))}\\ &\supseteq \tilde{\sigma}_{t+dt}'^{-1}(j). && \text{(since } v \notin \tilde{\sigma}_t'^{-1}(\{x, \dots, y\}) \text{ reproduces)}\end{aligned}$$

Case $v \in S_{i,<}$ for $i \in \{y+1, \dots, k\}$ if $y < k$. Thus, only $\tilde{M}'$ evolves and $r_y \le r_i = r_v < r'_v$. By definition of $\boldsymbol{r}'$ and since $\boldsymbol{r}$ is non-decreasing, this means $v \notin \tilde{\sigma}_t'^{-1}(\{x, \dots, y\})$, for if $v \in \tilde{\sigma}_t'^{-1}(\{x, \dots, y\})$ we had $r'_v \le r_y$ contradicting $r_y < r'_v$. We conclude

$$\begin{aligned}\tilde{\sigma}_{t+dt}^{-1}(j) &= \tilde{\sigma}_t^{-1}(j) && \text{(since } \tilde{M} \text{ does not evolve)}\\ &\supseteq \tilde{\sigma}_t'^{-1}(j) && \text{(by assumption (86))}\\ &\supseteq \tilde{\sigma}_{t+dt}'^{-1}(j). && \text{(since } v \notin \tilde{\sigma}_t'^{-1}(\{x, \dots, y\}) \text{ reproduces)}\end{aligned}$$

Case $v \in S_{i,>}$ for $i \in \{1, \ldots, x-1\}$ if $x > 1$. We show that this case is impossible. We have

$$\begin{aligned} r'_v < r_v = r_i & \qquad \text{(by def. of } S_{i,>}\text{)} \\ \le r_{x-1} & \qquad \text{(since } \boldsymbol{r} \text{ is non-decreasing)} \\ \le r_l. & \qquad \text{(by choice of } l\text{)} \end{aligned}$$

Further, since $i < x$, we have $v \notin \tilde{\sigma}_t^{-1}(\{x, \ldots, y\})$. By eq. (86), this implies $v \notin \tilde{\sigma}_t'^{-1}(\{x, \ldots, y\})$, and thus by definition of $\boldsymbol{r}'$, we must have $r'_v \ge r_l$. But this contradicts $r'_v < r_l$, making this case is impossible.

Case $v \in S_{i,>}$ for $i \in \{x, \ldots, y\}$. Thus, $v \in \tilde{\sigma}_t^{-1}(\{x, \ldots, y\})$ and $r_v > r'_v$, so that only $\tilde{M}$ evolves. We distinguish between two cases.
If $v \in \tilde{\sigma}_t'^{-1}(\{x, \ldots, y\})$, then we must have $\tilde{\sigma}(v) = \tilde{\sigma}'(v)$ by eq. (86). But then, $r_v = r'_v$, contradicting $r_v > r'_v$.
Else, $v \notin \tilde{\sigma}_t'^{-1}(\{x, \ldots, y\})$. We distinguish further:

If $x < k$. Then,

$$\begin{aligned} r'_v \ge r_l & \qquad \text{(by def. of } \boldsymbol{r}'\text{)} \\ \ge r_k & \qquad \text{(by def. of } l \text{ for } x < k\text{)} \\ \ge r_y & \qquad \text{(since } \boldsymbol{r} \text{ is non-decreasing)} \\ \ge r_v. & \qquad \text{(since } v \in \tilde{\sigma}_t^{-1}(\{x, \ldots, y\}) \text{ and } \boldsymbol{r} \text{ is non-decreasing)} \end{aligned}$$

However, this contradicts $r_v > r'_v$.

Else $x = k$. Thus, $\{x, \ldots, y\} = \{k\}$. Then,

$$\begin{aligned} \tilde{\sigma}_{t+dt}^{-1}(k) \supseteq \tilde{\sigma}_t^{-1}(k) & \qquad \text{(since } v \in \tilde{\sigma}_t^{-1}(k) \text{ reproduces.)} \\ \supseteq \tilde{\sigma}_t'^{-1}(k) & \qquad \text{(by assumption eq. (86))} \\ = \tilde{\sigma}_{t+dt}'^{-1}(k). & \qquad \text{(since } \tilde{M}' \text{ does not evolve)} \end{aligned}$$

We have covered all sub-cases, concluding the case $v \in S_{i,>}$ for $i \in \{x, \ldots, y\}$.

Case $v \in S_{i,>}$ for $i \in \{y+1, \ldots, k\}$ if $y < k$. We show that this case is impossible. We have $x \le y < k$ and $r_v > r'_v$. Further,

$$\begin{aligned} r_l = r_k & \qquad \text{(by def. of } l \text{ for } x < k\text{)} \\ \ge r_v & \qquad \text{(since } \boldsymbol{r} \text{ is non-decreasing)} \\ > r'_v. & \qquad \text{(by def. of } S_{i,>}\text{)} \end{aligned}$$

$r_l > r'_v$ implies by def. of $\boldsymbol{r}'$ that $v \in \tilde{\sigma}_t'^{-1}(\{x, \ldots, y\})$. By eq. (86), this implies $\tilde{\sigma}_t(v) = \tilde{\sigma}'_t(v)$ and thereby $r_v = r'_v$, contradicting $r_v > r'_v$.

We can therefore conclude that eq. (87) does indeed hold at time $t + dt$.

It remains to show that the coupling is indeed faithful up to time dt. Let us denote the original chains with $\hat{M}$ (the one with fitness $\boldsymbol{r}$) and $\hat{M}'$ (the one with fitness $\boldsymbol{r}'$). We first show that $\tilde{M}$ is indeed a faithful copy of $\hat{M}$. Recall that $\tilde{M}$ evolves if and only if $dt = t_v$ or $dt = t_v^1$ for some vertex v. By definition of t_v, t_v^1, we have for every vertex $v \in V$:

If $r_v \le r'_v$ then $\tilde{M}$ chooses v for reproduction if only if $dt = t_v$. Hence, the clock of v is given by the distribution $t_v \sim \operatorname{Exp}(\min(r_v, r'_v)) \overset{\text{distr.}}{=} \operatorname{Exp}(r_v)$.

Otherwise, $r_v > r'_v$ and $\tilde{M}$ evolves if and only if $dt = t_v$ or $dt = t_v^1$ so that the clock of v is given by

$$\begin{aligned}
\min(t_v, t_v^1) &\sim \min\Big(\operatorname{Exp}\left(\min(r_v, r'_v)\right), \operatorname{Exp}\left(r_v - r'_v\right)\Big) && \text{(def. of } t_v,\ t_v^1\text{)}\\
&\overset{\text{distr.}}{=} \operatorname{Exp}\left(\min(r_v, r'_v) + r_v - r'_v\right) && \text{(min of exponential distributions)}\\
&\overset{\text{distr.}}{=} \operatorname{Exp}\left(r_v\right). && \text{(since } r_v > r'_v\text{)}
\end{aligned}$$

Thus, the clock of v follows an $\operatorname{Exp}(r_v)$ distribution in $\tilde{M}$, as it does by definition in $\hat{M}$, so that $\tilde{M}$ is indeed a faithful copy of $\hat{M}$ in the time span $[t, t+dt]$.

To show that $\tilde{M}'$ is a faithful copy of $\hat{M}'$ we use the same argument as before, with the difference that the case distinction is for every vertex $v \in V$:

If $r_v \geq r'_v$ then $\tilde{M}'$ chooses v for reproduction if only if $dt = t_v$. Hence, the clock of v is given by the distribution $t_v \sim \operatorname{Exp}\left(\min\left(r_v, r'_v\right)\right) \overset{\text{distr.}}{=} \operatorname{Exp}\left(r'_v\right)$.

Otherwise, $r_v < r'_v$ and $\tilde{M}'$ evolves if and only if $dt = t_v$ or $dt = t_v^2$ so that the clock of v is given by

$$\begin{aligned}
\min(t_v, t_v^2) &\sim \min\Big(\operatorname{Exp}\left(\min(r_v, r'_v)\right), \operatorname{Exp}\left(r'_v - r_v\right)\Big) && \text{(def. of } t_v,\ t_v^2\text{)}\\
&\overset{\text{distr.}}{=} \operatorname{Exp}\left(\min(r_v, r'_v) + r'_v - r_v\right) && \text{(min of exponential distributions)}\\
&\overset{\text{distr.}}{=} \operatorname{Exp}\left(r'_v\right). && \text{(since } r_v < r'_v\text{)}
\end{aligned}$$

Since $\tilde{M}'$ evolves if and only if $dt = t_v$ or $dt = t_v^2$ for some vertex v, $\tilde{M}'$ is indeed a faithful copy of $\hat{M}'$ in the time span $[t + dt]$.

Since the exponential distribution is memoryless, it is possible to apply this argument inductively beyond time $t + dt$. ☐

Recall from the definition of the continuous time Moran process and Equation (84), that when regarding each evolution of the continuous time Moran process as a discrete time step, the transition probabilities are as in the discrete time Moran process. More formally, if we are currently in state σ and denote by $dt := \min_{v \in V} t_v$ the minimum of the clock set,[18] we have for every $v \in V$:

$$\mathbb{P}\left[dt = t_v\right] = \frac{r_{\sigma(v)}}{\sum_{u \in V} r_{\sigma(u)}}.$$

The probability for any $w \in N(v)$ to be chosen is $1/\deg v$ as in the discrete time process. Hence, the discrete time Moran process is just the embedded discrete time Markov chain of the continuous time Markov process. Therefore, the fixation probabilities are exactly the same in both the discrete and continuous time version. When we state properties of the fixation probabilities, the statements are thus valid for both discrete and continuous time.

We get some immediate corollaries from Theorem 26. They cover just special cases, and by no means the entirety of the insights gained from Theorem 26.

Corollary 27 (Monotonicity of f^k in r_k)**.** *For the (continuous time) k-type Moran process on any graph $G = (V, E)$, initial state $\sigma : V \to [k]$, non-decreasing fitness vector $\boldsymbol{r}$ and vector $\boldsymbol{r}'$ where $\forall i \in [k-1] : r_i = r'_i$, $r_k \leq r'_k$:*

$$f^k_{G,\boldsymbol{r}}(\sigma) \leq f^k_{G,\boldsymbol{r}'}(\sigma).$$

[18] Recall from Section 3.2.1 that $t_v \sim \operatorname{Exp}\left(r_{\tilde{\sigma}(v)}\right)$.

Proof. We apply Theorem 26 for $x = 1, y = k - 1$. Note that $\boldsymbol{r'}$ fulfils the condition of Theorem 26, since all components are equal except for the k-th component, where we have $r'_k \geq r_k =: r_l$ since $x = 1 < k$. We choose for both $\tilde{M} := \{\tilde{\sigma}_t \mid t \in \mathbb{R}_{\geq 0}\}$ and $\tilde{M}' := \{\tilde{\sigma}'_t \mid t \in \mathbb{R}_{\geq 0}\}$ the initial state $\tilde{\sigma}_0 = \tilde{\sigma}'_0 = \sigma$, so that we get by inductively applying Theorem 26 for any point in time $t \in \mathbb{R}_{\geq 0}$:

$$\forall j \in [k-1] : \tilde{\sigma}_t^{-1}(j) \supseteq \tilde{\sigma}_t'^{-1}(j) \tag{103}$$

$$\Rightarrow \bigcup_{j \in [k-1]} \tilde{\sigma}_t^{-1}(j) \supseteq \bigcup_{j \in [k-1]} \tilde{\sigma}_t'^{-1}(j) \qquad \text{(taking unions)}$$

$$\tilde{\sigma}_t^{-1}(k) \subseteq \tilde{\sigma}_t'^{-1}(k). \qquad \text{(taking complements)}$$

Therefore, whenever type k fixates in $\tilde{M}$, say at time $t = \tau$, we have $V = \tilde{\sigma}_\tau^{-1}(k) \subseteq \tilde{\sigma}_\tau'^{-1}(k)$, so $\tilde{\sigma}_\tau'^{-1}(k) = V$, and thus type k fixates in $\tilde{M}'$, too. Formally, this implies that the event "k fixates in $\tilde{M}$" implies the event "k fixates in $\tilde{M}'$", hence

$$f^k_{G,\boldsymbol{r}}(\sigma) \leq f^k_{G,\boldsymbol{r'}}(\sigma).$$

□

Corollary 27 can be seen as generalisations of Corollary 7 in [14].

If we take $x = y = k$, we can reduce by Theorem 26 the problem of finding a lower bound for $f^k_{G,\boldsymbol{r}}$ in the k-type Moran process to finding a lower bound for $f^2_{G,\boldsymbol{r'}}$ in a 2-type Moran process. The latter has been well studied so that we can then apply results from previous literature. We formalise this in Corollary 28.

Corollary 28. *For the (continuous time) k-type Moran process for any $k \geq 2$ on any graph $G = (V, E)$, state $\sigma : V \to [k]$, and non-increasing fitness vector $\boldsymbol{r} \in (\mathbb{R}^+)^k$, we must have*

$$f^k_{G,\boldsymbol{r}}(\sigma) \geq f^2_{G,\boldsymbol{r'}}(\sigma'),$$

where the fitness vector $\boldsymbol{r'} \in (\mathbb{R}^+)^2$ is defined by $r'_1 := r'_{k-1}$, $r'_2 := r_k$ and the type mapping $\sigma' : V \to [2]$ is defined by $\sigma'^{-1}(2) := \tilde{\sigma}^{-1}(k)$, $\sigma'^{-1}(1) = V \setminus \sigma'^{-1}(2)$.

Proof. We define the fitness vector $\boldsymbol{r^*} \in (\mathbb{R}^+)^k$ by $\forall i \in [k-1] : r^*_i := r_{k-1}$, $r^*_k := r_k$. We further define the continuous time Moran processes $\tilde{M} := \{\tilde{\sigma}_t \mid t \in \mathbb{R}_{\geq 0}\}$ on G with fitness vector $\boldsymbol{r}$ as well as $\tilde{M}^* := \{\tilde{\sigma}^*_t \mid t \in \mathbb{R}_{\geq 0}\}$ on G with fitness vector $\boldsymbol{r^*}$. Both processes have initial state σ: $\tilde{\sigma}_0 := \sigma$, $\tilde{\sigma}^*_0 := \sigma$.
We apply Theorem 26 on $\tilde{M}$, $\tilde{M}^*$ with $x = y = k$. Since $x = k$ implies $r_l = r_{k-1}$, $\boldsymbol{r^*}$ satisfies the condition for the unordered fitness vector in Theorem 26. Further, eq. (86) is trivially satisfied since both processes have the same initial state. Applying Theorem 26 inductively, we get for any point in time $t \in \mathbb{R}_{\geq 0}$:

$$\tilde{\sigma}_t^{-1}(k) \supseteq \left(\tilde{\sigma}''_t\right)^{-1}(k). \tag{104}$$

In particular, whenever type k fixates in $\tilde{M}^*$, say at time $t = \tau$, we have $\tilde{\sigma}_\tau^{-1}(k) \supseteq \left(\tilde{\sigma}^*_\tau\right)^{-1}(k) = V$, so $\tilde{\sigma}_\tau^{-1}(k) = V$, and thus type k fixates in $\tilde{M}$, too. Therefore,

$$f^k_{G,\boldsymbol{r}}(\sigma) \geq f^k_{G,\boldsymbol{r^*}}(\sigma). \tag{105}$$

In $\tilde{M}^*$, there is no difference between types 1 to $k-1$: all have the same fitness. Thus, we can w.l.o.g. aggregate types $1, \ldots, k-1$ as type 1 vertices, as well as remap type k to type 2. Formally, this yields the 2-type process $\tilde{M}' := \{\sigma'_t \mid t \in \mathbb{R}_{\geq 0}\}$ with fitness vector $\boldsymbol{r'}$: $r'_1 := r_{k-1}$,

$r'_2 := r_k$ and initial state $\tilde{\sigma}'_0 := \sigma'$ where $\sigma'^{-1}(2) := \sigma^{-1}(k)$, $\sigma'^{-1}(1) := V \setminus \sigma'^{-1}(2)$. Since we remapped type k to type 2, we can conclude

$$\begin{aligned} f^k_{G,\boldsymbol{r}}(\sigma) &\geq f^k_{G,\boldsymbol{r}^*}(\sigma) && \text{(by Equation (105))} \\ &= f^2_{G,\boldsymbol{r}'}(\sigma') && \text{(since type } k \text{ in } \tilde{M}^* \text{ gets remapped to type 2 in } \tilde{M}'\text{)} \end{aligned}$$

□

Since Corollary 28 holds for any initial state σ, it must in particular hold for those that map a single vertex to each of the types $2, \ldots, k$. Hence, it must also hold if any of these states is selected as initial state with equal probability. We can thus reduce the problem of finding a lower bound for f^k in the k-type process to f^2 in a 2-type process:

Corollary 29. *For the (continuous time) k-type Moran process for any $k \geq 2$ on any graph $G = (V, E)$ and non-decreasing fitness vector $\boldsymbol{r} \in (\mathbb{R}^+)^k$, we must have*

$$f^k_{G,\boldsymbol{r}} \geq f^2_{G,\boldsymbol{r}'},$$

where the fitness vector $\boldsymbol{r'} \in (\mathbb{R}^+)^2$ is defined by $r'_1 := r'_{k-1}$, $r'_2 := r_k$.

Proof. We define the function $g : \{V \to [k]\} \to \{V \to [2]\}$ mapping states from the k-type Moran process to states of the 2-type Moran process for any $\sigma : V \to [k]$ by

$$(g(\sigma))^{-1}(2) := (\sigma)^{-1}(k) \tag{106}$$

$$(g(\sigma))^{-1}(1) := V \setminus (\sigma)^{-1}(k). \tag{107}$$

Applying Definition 3, we have for $\Theta := \left\{\sigma : V \to [K] \,\middle|\, \forall j \in \{2, \ldots, k\} : |\sigma^{-1}(j)| = 1\right\}$

$$\begin{aligned} f^k_{G,\boldsymbol{r}} &= \frac{(n-k+1)!}{n!} \sum_{\sigma \in \Theta} f^j_{G,\boldsymbol{r}}(\sigma) && \text{(by Definition 3)} \\ &\geq \frac{(n-k+1)!}{n!} \sum_{\sigma \in \Theta} f^2_{G,\boldsymbol{r}'}(g(\sigma)), && (108) \end{aligned}$$

where the last step followed by Corollary 28. For $k > 2$, the function g is not injective on Θ, since for any $\sigma \in \Theta$, all of the $k-2$ many singletons mapped to types $2, \ldots, k-1$ under σ get aggregated to type 2 under $g(\sigma)$. We thus have that

$$\frac{(n-1)!}{(n-(k-2)-1)!} = \frac{(n-1)!}{(n-k+1)!} \tag{109}$$

many $\sigma \in \Theta$ have the same image $g(\sigma)$. We can therefore simplify

$$\begin{aligned} \Theta' &:= \{g(\sigma) : \sigma \in \Theta\} \\ &= \left\{g(\sigma) : \sigma \in \Theta \,\middle|\, |\sigma^{-1}(k)| = 1\right\} && \text{(by non-injectivity argument above)} \\ &= \left\{\sigma' : V \to [2] \,\middle|\, |\sigma'^{-1}(2)| = 1\right\}. && (110) \end{aligned}$$

Continuing eq. (108), we have

$$\frac{(n-k+1)!}{n!} \sum_{\sigma \in \Theta} f^2_{G,\boldsymbol{r}'}(g(\sigma)) = \frac{(n-k+1)!}{n!} \frac{(n-1)!}{(n-k+1)!} \sum_{\sigma' \in \Theta'} f^2_{G,\boldsymbol{r}'}(\sigma')$$

(by eq. (109) and def. of Θ')

$$= \frac{1}{n} \sum_{\sigma' \in \Theta'} f^2_{G,\boldsymbol{r'}}(\sigma')$$

$$= f^2_{G,\boldsymbol{r'}}. \qquad \text{(by Definition 3 and eq. (110))}$$

We can thus conclude

$$f^k_{G,\boldsymbol{r}} \geq f^2_{G,\boldsymbol{r'}}. \tag{111}$$

□

Another corollary from Theorem 26 yields the bounds from Ferreira and Neves [19]. We derive this in Appendix B.

We can prove an almost symmetric version of Theorem 26.

Theorem 30 (Stochastic Domination, subset)**.** *For any $k \geq 2$ and $y \in [k]$, let $\tilde{M} := \{\tilde{\sigma}_t \mid t \in \mathbb{R}_{\geq 0}\}$ be a k-type continuous time Moran process with non-decreasing fitness vector $\boldsymbol{r}$ on a graph $G = (V, E)$. Let $\tilde{M}' := \{\tilde{\sigma}'_t \mid t \in \mathbb{R}_{\geq 0}\}$ be any other k-type continuous time Moran process on G with the unordered fitness vector $\boldsymbol{r'}$ such that $\forall i \in \{1, \ldots, y\} : r'_i = r_i$, $\forall i \in \{y+1, \ldots, k\} : r'_i \leq r_{y+1}$. Visually, this translates to*

$$\boldsymbol{r'} = \begin{pmatrix} r'_1 \\ \vdots \\ r'_y \\ r'_{y+1} \\ \vdots \\ r'_k \end{pmatrix} \begin{matrix} = r_1 \\ \vdots \\ = r_y \\ \leq r_{y+1} \\ \vdots \\ \leq r_{y+1}. \end{matrix} \tag{112}$$

Then there exists a coupling between $\tilde{M}$ and $\tilde{M}'$ such that at any point in time $t \in \mathbb{R}_{\geq 0}$,

$$\forall j \in \{1, \ldots, y\} : \ \tilde{\sigma}_t^{-1}(j) \subseteq \tilde{\sigma}_t'^{-1}(j) \tag{113}$$

implies

$$\forall j \in \{1, \ldots, y\} : \ \tilde{\sigma}_{t+dt}^{-1}(j) \subseteq \tilde{\sigma}_{t+dt}'^{-1}(j), \tag{114}$$

where dt is the time it takes for the first reproduction to happen in $\tilde{M}$ or $\tilde{M}'$ after time t.

Since the proof is very similar to the proof of Theorem 26, we defer the proof to Appendix A.

We can now translate the phase transition result [1] with the help of Theorems 26 and 30 to a phase transition of $f^k_{G,\boldsymbol{r}}$ in the k-type Moran process:

Theorem 31. *For the (continuous time) k-type Moran process for any $k \geq 2$ with vector $\boldsymbol{r} \in (\mathbb{R}^+)^k$ on any sufficiently large connected graph $G = (V, E)$, we have*

$$f^k_{G,\boldsymbol{r}} > \frac{(\log n)^C}{n} \qquad \text{if } \boldsymbol{r} \text{ is strictly increasing } (r_1 < \ldots < r_k) \tag{i}$$

$$f^k_{G,\boldsymbol{r}} = \frac{1}{n} \qquad \text{if } \boldsymbol{r} \text{ is constant } (r_1 = \ldots = r_k) \tag{ii}$$

$$f^k_{G,\boldsymbol{r}} < \frac{1}{n^C} \qquad \text{if } \boldsymbol{r} \text{ is strictly decreasing } (r_1 > \ldots > r_k) \tag{iii}$$

for all $C > 0$.

Proof. We first prove claim (i). In fact, the requirement could be relaxed to $\boldsymbol{r}$ being non-decreasing and $r_{k-1} < r_k$. Since $\boldsymbol{r}$ is non-decreasing, it satisfies our definition of a fitness vector (Definition 1). We have for the 2-dimensional fitness vectors $\boldsymbol{r'}$ with $r'_1 := r_{k-1}$, $r'_2 := r_k$ and $\boldsymbol{r^*}$ with $r^*_1 := 1$, $r^*_2 := r_k/r_{k-1}$

$$\begin{aligned} f^k_{G,\boldsymbol{r}} &\geq f^2_{G,\boldsymbol{r'}} && \text{(by Corollary 29)} \\ &= f^2_{G,\boldsymbol{r^*}}. && \text{(by rescaling as in eqs. (10) and (11))} \end{aligned}$$

Since $r_k/r_{k-1} > 1$, the third case of Theorem 2 in [1] applies. Recall from Goldberg, Lapinskas and Richerby that this theorem applies only to graphs of order bigger than some constant depending only on r_k/r_{k-1}, say for $|G| > n_{r_k/r_{k-1}}$,[19] yielding

$$f^2_{G,\boldsymbol{r^*}} \geq \frac{\log(n)^C}{n} \tag{115}$$

for $|G| > n_{r_k/r_{k-1}}$ for some constant $n_{r_k/r_{k-1}}$ depending only on $\frac{r_k}{r_{k-1}}$ as in Theorem 2 of [1], showing claim (i).

We show (ii) by an argument similar to Lemma 1 in [13] by first expanding the number of types to n. We sample a permutation of V u.a.r. to obtain $v_1, \ldots v_n$ and suppose each vertex v_i has its own type $i \in [n]$ initially, where all types have the same fitness: $\boldsymbol{r'} := (r_1, \ldots, r_1) \in (\mathbb{R}^+)^n$. Eventually, one of the types will fixate, yielding

$$\sum_{j \in [n]} f^j_{G,\boldsymbol{r'}} = 1. \tag{116}$$

Since each fitness is equal and the permutation was taken u.a.r., we have for any type $j \in [n]$: $f^j_{G,\boldsymbol{r'}} = \frac{1}{n}$. In particular, $f^k_{G,\boldsymbol{r'}} = \frac{1}{n}$. Since $r_1 = \ldots = r_k$, we can obtain the k-type Moran process from the n-type process above by mapping types $k+1, \ldots, n$ to type 1. Since we took an arbitrary permutation of V as the initial state, the remapped type k-process is indeed initialised as required in Definition 3. Since type k is not remapped, we still have in the type k process : $f^k_{G,\boldsymbol{r}} = \frac{1}{n}$, showing claim (ii).

We now prove claim (iii). In fact, the requirement could be relaxed to $\boldsymbol{r}$ being non-increasing and $r_{k-1} > r_k$. We require some permutations of the types.[20] and define the Moran processes

$$\begin{aligned} M^a &:= \{\tilde{\sigma}^a_t |\, t \in \mathbb{R}_{\geq 0}\} && \text{with non-increasing fitness vector } \boldsymbol{r^a} := \boldsymbol{r} \\ M^b &:= \{\tilde{\sigma}^b_t |\, t \in \mathbb{R}_{\geq 0}\} && \text{with non-decreasing fitness vector } \boldsymbol{r^b} := (r_k, \ldots, r_1) \\ M^c &:= \{\tilde{\sigma}^c_t |\, t \in \mathbb{R}_{\geq 0}\} && \text{with non-decreasing fitness vector } \boldsymbol{r^c} := (r_k, r_{k-1} \ldots, r_{k-1}) \\ M^d &:= \{\tilde{\sigma}^d_t |\, t \in \mathbb{R}_{\geq 0}\} && \text{with increasing fitness vector } \boldsymbol{r^d} := (r_k, r_{k-1}) \\ M^e &:= \{\tilde{\sigma}^e_t |\, t \in \mathbb{R}_{\geq 0}\} && \text{with decreasing fitness vector } \boldsymbol{r^e} := (r_{k-1}, r_k), \end{aligned}$$

where M^b is obtained from M^a by inverting the order of the types, *i.e.* type i maps to type $k-i+1$ for any $i \in [k]$. M^c is obtained from M^b by increasing the fitnesses of types $3, \ldots, k$. M^d is obtained from M^c by realising that types $2, \ldots, k$ have the same fitness, so all ypes $2, \ldots, k$ get mapped to type 2. M^e is obtained from M^d by flipping types. Since $r_{k-1} > r_k$ by assumption, r^d is indeed increasing and r^e decreasing. Except for the M^b to M^c mapping, each

---

[19] In the notation of [1], the constant depends only on r, which is in our case r_k/r_{k-1}. For more details, see Theorem 2 of [1].

[20] The reason for this rather extensive notation is that $\boldsymbol{r}$ is non-increasing and our previous theorems work only for $\boldsymbol{r}$ being non-decreasing. Indeed, this leads to a notational overhead in this and only this section. In all other proofs, it is a very convenient convention that simplifies their proof to a great extend without limiting their scope.

type has the same fitness as the type that it gets mapped to in the new Moran process. We can thus just "trace" the types through the mapping to obtain *e.g.* $(\tilde{\sigma}_t^a)^{-1}(i) = \left(\tilde{\sigma}_t^b\right)^{-1}(k-i+1)$ for all $i \in [k]$ at every point in time $t \in \mathbb{R}_{\geq 0}$.

For the M^b to M^c mapping, we can apply Theorem 30 with $y := 1$ since $\boldsymbol{r}^b$ is non-decreasing and $r_1^b = r_1^c$, and obtain by inductive application for every point in time $t \in \mathbb{R}_{\geq 0}$:

$$\left(\tilde{\sigma}_t^b\right)^{-1}(1) \subseteq \left(\tilde{\sigma}_t^c\right)^{-1}(1). \tag{117}$$

Tracing type k in the initial Markov chain M^a until M^e and applying eq. (117), we have for every point in time $t \in \mathbb{R}_{\geq 0}$:

$$(\tilde{\sigma}_t^a)^{-1}(k) \subseteq (\tilde{\sigma}_t^e)^{-1}(2). \tag{118}$$

Thus, whenever type k fixates in process M^a, type 2 must fixate in process M^e as well, yielding

$$f_{G,\boldsymbol{r^a}}^k \leq f_{G,\boldsymbol{r^e}}^2. \tag{119}$$

Process M^e is simply the 2-type process and we can normalise its fitness to $(1, \frac{r_k}{r_{k-1}})$. Since $r_{k-1} > r_k$, this is equivalent to the first case in Theorem 2 of [1], yielding

$$\begin{aligned} f_{G,\boldsymbol{r}}^k &= f_{G,\boldsymbol{r^a}}^k && \text{(by def of } M^a\text{)} \\ &\leq f_{G,\boldsymbol{r^e}}^2 && \text{(by eq. (119))} \\ &\leq \frac{1}{n^C} && \text{(by Theorem 2 of [1])} \end{aligned}$$

for every $C > 0$ and $|G| > n_{r_k/r_{k-1}}$, for some constant $n_{r_k/r_{k-1}}$ depending only on $\frac{r_k}{r_{k-1}}$ as in Theorem 2 of [1], concluding the proof of case (iii).

□

Theorem 31 gave a lower bound on the fixation probability of the strongest mutant, $f_{G,\boldsymbol{r}}^k$, which is independent of k. We can thus regard $\boldsymbol{r}$ now as a parameter, to get an FPTRAS for p-MoranMax in Main Theorem 5. Since Theorem 31 has the caveat requiring the graph to be sufficiently large, we introduce Algorithm 2 to also deal with small graphs.

Algorithm 2 (N, T): Approximating $f_{G,\boldsymbol{r}}^k$ in the Moran process, where N, T are functions on $(0,1) \times \mathbb{N} \times \mathbb{R}^* \times \mathbb{N} \to \mathbb{N}$

Input: Graph G, fitness vector $\boldsymbol{r}$, and error tolerance $\varepsilon \in (0,1)$, where $n := |G|$, $k := |\boldsymbol{r}|$
Output: Approximation of $f_{G,\boldsymbol{r}}^k$

```
if $n > n_{r_k/r_{k-1}}$ then                                              ▷ For sufficiently large $G$
    $f_{G,\boldsymbol{r}}^k \leftarrow$ ALGORITHM 1$(N,T)$ on input $G, \boldsymbol{r}, j := k$        ▷ Simulate the Markov chain
    return $f_{G,\boldsymbol{r}}^k$
else                                                                        ▷ For small $G$
    $f_{G,\boldsymbol{r}}^k \leftarrow$ Find exact solution for $f_{G,\boldsymbol{r}}^k$      ▷ "Brute-Force", see details in proof.
    return $f_{G,\boldsymbol{r}}^k$
end if
```

Main Theorem 5 (restated). p-MoranMax *has an FPTRAS.*

Proof. We show that Algorithm 2 with

$$N: \ (\varepsilon, n, \boldsymbol{r}, j) \mapsto \left\lceil 2^{-1}\varepsilon^{-2}\frac{n^2}{\log^2 n}\ln(16)\right\rceil \tag{120}$$

$$T:\ (\varepsilon, n, \boldsymbol{r}, j) \mapsto \left\lceil \frac{r_k}{(r_k - r_{k-1})} \right\rceil (n+1)\, n^3 8 \left(N(\varepsilon, n, \boldsymbol{r}, j)\right)^2, \tag{121}$$

is an FPTRAS for p-MoranMax and abbreviate the functions' values by $\hat{N}, \hat{T}$. Note that both $\hat{N}, \hat{T}$ are well-defined, positive integers since by Definition 1, $\boldsymbol{r}$ is strictly increasing and positive.

We first focus on the $n > n_{r_k/r_{k-1}}$ case. For correctness, the analysis proceeds as in the proof of Main Theorem 4. The difference is in the application of the Chernoff bound (eq. (79)) where we get instead for any $\varepsilon > 0$

$$\mathbb{P}\left[\left|p - f^k_{G,\boldsymbol{r}}\right| > \varepsilon f^k_{G,\boldsymbol{r}}\right] \leq 2\exp\left(-2\varepsilon^2 \left(f^k_{G,\boldsymbol{r}}\right)^2 \hat{N}\right) \leq \frac{1}{8} \tag{122}$$

where we used in the last step that $f^k_{G,\boldsymbol{r}} > \frac{\log n}{n}$ (Theorem 31) and the definition of $\hat{N} = N(\varepsilon, n, \boldsymbol{r}, j)$ (eq. (120)).

Further, the probability that type k has not reached absorption in a single simulation within $\hat{T}/\hat{N}$ steps is by Corollary 14

$$\mathbb{P}\left[\tau_k \geq \hat{T}/\hat{N}\right] \leq \frac{\hat{N}}{\hat{T}} \frac{r_k}{(r_k - r_{k-1})} (n+1)\, n^3 \leq \frac{1}{8\hat{N}}, \tag{123}$$

where the last step followed by eq. (121).

The probability that all simulations together take more than $\hat{T}$ steps is bounded by the probability that any of the $\hat{N}$ many simulations takes more than $\hat{T}/\hat{N}$ steps, which is by a union bound at most $\frac{1}{8}$. Thus, with probability at least $\frac{3}{4}$, the algorithm returns a value within a factor $1 \pm \varepsilon$ of $f^k_{G,\boldsymbol{r}}$.

It remains to show that the running time is bounded by a polynomial in the size of the input and $1/\varepsilon$. For $n > n_{r_k/r_{k-1}}$, recall that by definition of Moran-$\boldsymbol{r}$, $\boldsymbol{r}$ and thereby $k := |\boldsymbol{r}|$ is assumed to be constant and thereby not part of the input. By Lemma 24, the running time is in

$$\mathcal{O}\left(\hat{T} \cdot k \cdot \log n + n^2\right). \tag{124}$$

We have by definition for some constant C

$$\hat{T} = T(\varepsilon, n, \boldsymbol{r}, j) \leq \frac{r_k}{r_k - r_{k-1}} \varepsilon^{-4} \cdot C \frac{n^8}{\log^4 n}, \tag{125}$$

so that the running time is indeed bounded by

$$\underbrace{C\frac{n^8}{\log^3 n}}_{p_1(n)} \cdot \underbrace{\varepsilon^{-4}}_{p_2(1/\varepsilon)} \cdot \underbrace{k\frac{r_k}{r_k - r_{k-1}}}_{g(\boldsymbol{r})}. \tag{126}$$

Hence, we have indeed an FPTRAS.

For $n \leq n_{r_k/r_{k-1}}$, we just calculate $f^k_{G,\boldsymbol{r}}$ exactly, *e.g.* using techniques from Section 1.2. The running time vanishes in the function $g(\boldsymbol{r})$, since $n_{r_k/r_{k-1}}$ depends only on r_k, r_{k-1}. □

By allowing for a slightly worse lower bound, we get a lower bound on $f^k_{G,\boldsymbol{r}}$ that holds for any non-decreasing fitness vector $\boldsymbol{r}$ (as opposed to strictly increasing) and get guarantees for graphs of any order, so that we do not have to deal with $n_{r_k/r_{k-1}}$:

Corollary 32. *For the (continuous time) k-type Moran process for any $k \geq 2$ with non-decreasing fitness vector $\boldsymbol{r} \in (\mathbb{R}^+)^k$ on any graph $G = (V, E)$, we have*

$$f^k_{G,\boldsymbol{r}} \geq \frac{1}{n}.$$

Proof. In the proof of Theorem 31 (i), the difference is that $r_2^* := r_k/r_{k-1} \geq 1$, instead of a strict inequality, so that Corollary 2 of [13] applies, yielding the lower bound $1/n$. ☐

Can we turn the FPTRAS even into an FPRAS for p-MoranMax, *i.e.* in an approximation algorithm with running time also polynomial in $\boldsymbol{r}$? The answer is yes, if we make realistic assumptions. First, we assume that r_k is not some artificially large value such as n^k. Second, we assume that r_k and r_{k-1} are not too close. Intuitively, the second assumption makes sense for the following reason: Suppose we have $k = 3$ types and start initially with one vertex u of type 2 and one vertex v of type 3. We work on a very symmetric graph such that with sufficiently high probability, v and u start under the same conditions.[21] Then, it will be very hard to distinguish whether type 2 or type 3 rather wins if r_2 is very close to r_3, say $r_2 = r_3 - 2^{-10n}$, since they start under very similar conditions. In the $k = 2$ case, this was not a problem since type 1 and type 2 always differ significantly in the initial state even when their fitnesses are close, since type 1 starts with $n - 1$ vertices. We now make both requirements of r_k being not "too large" and r_{k-1} being not "too close" to r_k formal.

Corollary 33. *For*

$$\begin{aligned} r_k &\in \mathcal{O}\left(\operatorname{poly}(n,k)\right) \\ r_k - r_{k-1} &\in \Omega\left(\operatorname{poly}(n,k)\right) \end{aligned}$$

there is an FPRAS for p-MoranMax.

Note that Corollary 33 applies to many fitness vectors $\boldsymbol{r}$, such as the very basic, linearly growing vector $\boldsymbol{r} = (1, 2, \ldots, k)$.

Proof of Corollary 33. We choose for *Algorithm* 1

$$\begin{aligned} N &: \ (\varepsilon, n, \boldsymbol{r}, j) \mapsto \left\lceil 2^{-1}\varepsilon^{-2} n^2 \ln(16) \right\rceil \\ T &: \ (\varepsilon, n, \boldsymbol{r}, j) \mapsto \left\lceil \frac{r_k}{(r_k - r_{k-1})} \right\rceil (n+1)\, n^3 8 \left(N(\varepsilon, n, \boldsymbol{r}, j)\right)^2 . \end{aligned}$$

The claim follows from the $n > n_{r_k/r_{k-1}}$ case in the proof of Main Theorem 5 by noting that Corollary 32 applies to graphs of *any* order. For the running time, we have

$$T(\varepsilon, n, \boldsymbol{r}, j) \leq \frac{r_k}{r_k - r_{k-1}} \varepsilon^{-4} \cdot C n^8 \tag{127}$$

for some constant C. Therefore, by Lemma 24, the overall running time is bounded by

$$\underbrace{C n^8 \log n \cdot k \frac{r_k}{r_k - r_{k-1}}}_{p_1(n,k)} \cdot \underbrace{\varepsilon^{-4}}_{p_2(1/\varepsilon)} . \tag{128}$$

By our assumption on r_k, r_{k-1}, we have $p_1(n,k) \in \operatorname{poly(n,k)}$. Further, $p_2(1/\varepsilon) \in \operatorname{poly}(1/\varepsilon)$ so that we have indeed an FPRAS. ☐

We now have improved a rough lower bound on $f^j_{G,\boldsymbol{r}}$ for $j \in \{2, \ldots, k\}$ to a better lower bound when $j = k$, which is independent of k. Can we do so for other j? This will be answered in Section 3.3.

---

[21] Graph-theoretically, this can be translated to working on a graph $G = (V, E)$ with a high number of automorphisms $|\mathrm{Aut}(G)|$ and requiring that for u, v chosen u.a.r. from the set of distinct pairs of V, there is with sufficiently high probability an automorphism $\varphi \in \mathrm{Aut}(G)$ such that $\varphi(v) = u$

3.3 Filling the gap: Lower bounds for the fixation probabilities of types $j \in \{2, \ldots, k\}$

Recall the lower bounds on $f^j_{G,\boldsymbol{r}}$ for $j \in \{2, \ldots, k\}$ we have proved so far: For types $j \in \{2, \ldots, k\}$, we have a bound which is exponentially small in k (Lemma 23). We were able to improve this for $j = k$ to a bound independent of k.

What about the other types $2, \ldots, k-1$? We will now provide in Theorem 34 an improved lower bound for these types. Notably, this bound coincides with Theorem 31 (i) for the one "extreme" case $j = k$ and with Lemma 23 for the other "extreme" case $j = 2$. Hence, the closer j is to k, the better the bound of Theorem 34. We exploit this property when giving an FPTRAS and FPRAS for the promise problem p-Moran-c-ThFittest in Main Theorem 6 and Corollary 36.

Theorem 34. *For every type $j \in \{2, \ldots, k\}$ in the k-type Moran process with strictly increasing fitness vector $\boldsymbol{r}$ on any sufficiently large graph G with maximum degree Δ, we have*

$$f^j_{G,\boldsymbol{r}} > (k-j+1)! \left(\frac{r_j}{r_k}\right)^{k-j} \left(\frac{1}{\Delta}\right)^{k-j} \frac{\log n}{n^{2k-2j+1}}. \tag{129}$$

Proof. Let $j \in \{2, \ldots, k\}$ be arbitrary. Recall from Definition 3 that

$$f^j_{G,\boldsymbol{r}} := \frac{(n-k+1)!}{n!} \sum_{\sigma \in \Theta} f^j_{G,\boldsymbol{r}}(\sigma),$$

where $\Theta := \left\{\sigma \in \Omega \;\middle|\; \forall j \in \{2, \ldots, k\} : |\sigma^{-1}(j)| = 1\right\}$. We define the set

$$\Theta' := \left\{\sigma \in \Theta \;\middle|\; G\left[\sigma^{-1}(\{j, \ldots, k\})\right] \text{ is connected}\right\}. \tag{130}$$

For any $\sigma' \in \Theta'$ let $S := \sigma'^{-1}(\{j, \ldots, k\})$. Since $G[S]$ is connected, there must be a spanning tree T in $G[S]$. We root this tree at v_j, where $v_j \in S$ is the unique vertex having type $\sigma'(v_j) = j$. If the first $k-j$ edges chosen for reproduction form some tree traversal of T starting at its roots v_j, then after $k-j$ steps, all vertices in S have type j, and all other vertices have type at most $j-1$. Call this state σ^*. The probability to reach σ^* from σ' within these first $k-j$ steps is at least

$$\left(\frac{r_j}{r_k \cdot n \cdot \Delta}\right)^{k-j}, \tag{131}$$

by a rough upper bound that the population's overall fitness is at most $r_k \cdot n$ at any point in time.

We can further bound

$$f^j_{G,\boldsymbol{r}}(\sigma^*) \geq f^j_{G,\boldsymbol{r}^+}(\sigma^+), \tag{132}$$

where $\boldsymbol{r}^+ \in (\mathbb{R}^+)^2$ with $r^+_1 := r_{j-1}$, $r^+_2 := r_j$ and $\sigma^+ : V \to [2]$ with $(\sigma^+)^{-1}(1) := V \setminus S$ and $(\sigma^+)^{-1}(2) := S$. This follows by noting that in state σ^*, all vertices are of type $\{1, \ldots, j\}$, so that we can truncate $\boldsymbol{r}$ and σ^* to the j-mutant case and apply Corollary 28, yielding the fitness vector (r_{k-1}, r_j). We can further apply Corollary 8 in [14] (subset domination) and obtain, e.g. by the third case of Theorem 2 in [1]

$$f^j_{G,\boldsymbol{r}^+}(\sigma^+) \geq \frac{\log n}{n}. \tag{133}$$

We can conclude therefore

$$\begin{aligned}
f^j_{G,\boldsymbol{r}} &:= \frac{(n-k+1)!}{n!}\sum_{\sigma\in\Theta} f^j_{G,\boldsymbol{r}}(\sigma) && \text{(by Definition 3)}\\
&\geq \frac{(n-k+1)!}{n!}\sum_{\sigma'\in\Theta'} f^j_{G,\boldsymbol{r}}(\sigma') && \text{(since } \Theta'\subseteq\Theta)\\
&\geq \frac{(n-k+1)!}{n!}\sum_{\sigma'\in\Theta'} \left(\frac{r_j}{r_k\cdot n\cdot\Delta}\right)^{k-j}\cdot\frac{\log n}{n} && \text{(by eqs. (131) to (133))}\\
&> (k-j+1)!\left(\frac{r_j}{r_k}\right)^{k-j}\left(\frac{1}{\Delta}\right)^{k-j}\frac{\log n}{n^{2k-2j+1}}, && (134)
\end{aligned}$$

where the last step is due to the following cardinality bound on Θ': There are at least $n-|S|+1$ many connected subsets of V of size $|S|$, since G is connected. How many $\sigma'\in\Theta'$ are there for a given S such that $S=\sigma'^{-1}(\{j,\dots,k\})$ (recall that $|S|=k-j+1$)? There are $|S|!$ many permutations for assigning to each of the types $j,\dots,k$ a single vertex in S. We must then assign types $2,\dots,j-1$ to vertices in $V\setminus S$, for which there are

$$\frac{(n-|S|)!}{(n-|S|-(j-2))!}$$

possibilities. All remaining vertices get type 1. We can therefore conclude that

$$\begin{aligned}
|\Theta'| &\geq (n-|S|+1)\cdot|S|!\cdot\frac{(n-|S|)!}{(n-|S|-(j-2))!} && (135)\\
&= \frac{(k-j+1)!(n-k+j)!}{(n-k+1)!}. && (136)
\end{aligned}$$

☐

Note that this proof uses Corollary 28 and thereby relies on the coupling result (Theorem 26).

We get again a slight loser bound when relaxing $\boldsymbol{r}$ to be non-decreasing, which applies to graphs of any order.

Theorem 35. *For every type $j\in\{2,\dots,k\}$ in the k-type Moran process with non-decreasing fitness vector $\boldsymbol{r}$ on any graph G with maximum degree Δ, we have*

$$f^j_{G,\boldsymbol{r}} > (k-j+1)!\left(\frac{r_j}{r_k}\right)^{k-j}\left(\frac{1}{\Delta}\right)^{k-j}\frac{1}{n^{2k-2j+1}}. \tag{137}$$

Proof. The proof proceeds as the proof of Theorem 34, except that Corollary 2 in [13] applies so that Equation (133) becomes

$$f^j_{G,\boldsymbol{r}^+}(\sigma^+) \geq \frac{1}{n}. \tag{138}$$

☐

We now turn our lower bound into an efficient approximation algorithm for p-Moran-c-ThFittest.

Main Theorem 6 (restated). p-Moran-c-ThFittest *has an FPTRAS.*

Proof. We show that Algorithm 2 with the functions

$$N: (\varepsilon, n, \boldsymbol{r}, j) \mapsto \left\lceil 2^{-1}\varepsilon^{-2}\left(\frac{r_k}{r_j}\right)^{2k-2j} \frac{1}{((k-j+1)!)^2}\frac{n^{6k-6j+2}}{\log^2(n)}\ln(16)\right\rceil \tag{139}$$

$$T: (\varepsilon, n, \boldsymbol{r}, j) \mapsto \left\lceil \sum_{i=j}^{k} \frac{r_i}{r_i - r_{i-1}}\right\rceil (n+1)n^3 8\left(N(\varepsilon, n, \boldsymbol{r}, j)\right)^2 \qquad \text{(where } k := |\boldsymbol{r}|\text{)}$$

is an FPTRAS for p-Moran-c-ThFittest and abbreviate the functions' values by $\hat{N}, \hat{T}$. Note that both $\hat{N}, \hat{T}$ are well-defined, positive integers since by Definition 1, $\boldsymbol{r}$ is strictly increasing and positive.

The $n \le n_{r_k/r_{k-1}}$ case is covered as in the proof of Main Theorem 5.

Let $n > n_{r_k/r_{k-1}}$. For correctness, the analysis proceeds as in the proof of Main Theorem 4. The difference is in the application of the Chernoff bound (eq. (79)) where we get instead for any $\varepsilon > 0$

$$\mathbb{P}\left[\left|p - f^j_{G,\boldsymbol{r}}\right| > \varepsilon f^j_{G,\boldsymbol{r}}\right] \le 2\exp\left(-2\varepsilon^2\left(f^j_{G,\boldsymbol{r}}\right)^2\hat{N}\right) \le \frac{1}{8} \tag{140}$$

where we used in the last step that $f^j_{G,\boldsymbol{r}} > (k-j+1)!\left(\frac{r_j}{r_k}\right)^{k-j}\left(\frac{1}{\Delta}\right)^{k-j}\frac{\log n}{n^{2k-2j+1}}$ (Theorem 34) and the definition of $N(\varepsilon, n, \boldsymbol{r}, j)$.

Further, the probability that type k has not reached absorption in a single simulation within $\hat{T}/\hat{N}$ steps is at most

$$\mathbb{P}\left[\tau_j \ge \hat{T}/\hat{N}\right] \le \frac{\mathbb{E}[\tau_j]}{\hat{T}/\hat{N}} \le \frac{1}{8\hat{N}}, \tag{141}$$

where we used in the last step that by Corollary 18,

$$\mathbb{E}[\tau_j] \le \sum_{i=j}^{k}\frac{r_i}{r_i - r_{i-1}}(n+1)n^3, \tag{142}$$

and the definition of $T(\varepsilon, n, \boldsymbol{r}, j)$.

The probability that all simulations together take more than $\hat{T}$ steps is bounded by the probability that any of the $\hat{N}$ many simulations takes more than $\hat{T}/\hat{N}$ steps, which is by a union bound at most $\frac{1}{8}$. Thus, with probability at least $\frac{3}{4}$, the algorithm returns a value within a factor $1 \pm \varepsilon$ of $f^k_{G,\boldsymbol{r}}$.

We now show that the running time is bounded by polynomials $p_1(n)$, $p_2(1/\varepsilon)$ and some function $g(\boldsymbol{r})$. Recall that by definition of p-Moran-c-ThFittest, $\boldsymbol{r}$ and thereby $k := |\boldsymbol{r}|$ are parameters and $j \in \{\max(2, k-c+1), \dots, k\}$.

Since $j \ge k - c + 1$, we have

$$\hat{N} = \left\lceil \frac{\ln(16)}{2}\varepsilon^{-2}\left(\frac{r_k}{r_j}\right)^{2c-2}\frac{1}{(c!)^2}n^{6c-4}\right\rceil \tag{143}$$

$$\hat{T} := \left\lceil \sum_{i=k-c+1}^{k}\frac{r_i}{r_i - r_{i-1}}\right\rceil (n+1)n^3 8\left(N(\varepsilon, n, \boldsymbol{r}, j)\right)^2 \qquad \text{(where } k := |\boldsymbol{r}|\text{)}$$

We have by definition for some constant D

$$\hat{T} = T(\varepsilon, n, \boldsymbol{r}, j) \le \sum_{i=k-c+1}^{k}\frac{r_i}{r_i - r_{i-1}}\cdot\left(\frac{r_k}{r_j}\right)^{4c-4}\varepsilon^{-4}\cdot D\frac{n^{12c-4}}{\log^4 n}, \tag{144}$$

so that the running time is by Lemma 24 indeed bounded by

$$\underbrace{D\frac{n^{12c-4}}{\log^3 n}}_{p_1(n)} \cdot \underbrace{\varepsilon^{-4}}_{p_2(1/\varepsilon)} \underbrace{\cdot k \sum_{i=k-c+1}^{k} \frac{r_i}{r_i - r_{i-1}} \cdot \left(\frac{r_k}{r_j}\right)^{4c-4}}_{g(\boldsymbol{r})} . \tag{145}$$

Hence, we have an FPTRAS for p-Moran-c-ThFittest. ☐

Under similar assumptions as in Corollary 33, we even get an FPRAS:

Corollary 36. *For*

$$\begin{aligned} \forall i \in \{k-c+1, \dots, k\} : \qquad\qquad r_i &\in \mathcal{O}\left(\mathrm{poly}(n,k)\right) \\ r_i - r_{i-1} &\in \Omega\left(\mathrm{poly}(n,k)\right) \end{aligned}$$

p-Moran-c-ThFittest *has an FPRAS for* p-Moran-c-ThFittest.

Proof. We choose in Algorithm 1

$$N : (\varepsilon, n, \boldsymbol{r}, j) \mapsto \left\lceil 2^{-1}\varepsilon^{-2}\left(\frac{r_k}{r_j}\right)^{2k-2j} \frac{1}{((k-j+1)!)^2} n^{6k-6j+2}\ln(16)\right\rceil \tag{146}$$

$$T : (\varepsilon, n, \boldsymbol{r}, j) \mapsto \left\lceil \sum_{i=j}^{k} \frac{r_i}{r_i - r_{i-1}} \right\rceil (n+1)n^3 8(N(\varepsilon, n, \boldsymbol{r}, j)))^2. \qquad \text{(where } k := |\boldsymbol{r}|\text{)}$$

By the condition on r_i, we have for every $i \in \{k-c+1, \dots, k\}$

$$\frac{r_k}{r_i}, \frac{r_i}{r_i - r_{i-1}} \in \mathcal{O}\left(\mathrm{poly}(n,k)\right), \tag{147}$$

So that the overall running time is, analogously to Corollary 33, bounded by

$$\underbrace{Dn^{12c-4}k \sum_{i=k-c+1}^{k} \frac{r_i}{r_i - r_{i-1}} \cdot \left(\frac{r_k}{r_j}\right)^{4c-4}}_{p_1(n,k)} \cdot \underbrace{\varepsilon^{-4}}_{p_2(1/\varepsilon)} \tag{148}$$

for some constant D. Since $p_1(n,k) \in \mathrm{poly}(n,k)$ by eq. (147), we have indeed an FPRAS for p-Moran-c-ThFittest. ☐

4 Martingale approach for the complete graph

We mentioned in the beginning that computing explicit formulas by solving the recurrence equation is a very tedious procedure and we even doubt that there exist explicit formulas for the k-type process already on the complete graph K_n. However, Monk, Green, and Paulin [31] gave martingale proofs for the explicit formula in the 2-type process on K_n. Since this approach is much more compact than explicitly solving the non-linear recurrence relation, the hope is to get, at least in special cases, an explicit formula for the fixation probabilities by using a martingale approach. Monk and Schaik [32] note that

> *We can consider birth-death processes with more than two competing species, each with different fitnesses [...]. Whether or not martingale analysis is applicable to any of these Moran process extensions is an open research question. [...] Finding such an expectation may be quite laborious or impossible, depending on the complexity of the stochastic process and the exploitable symmetries in it.*

In this section, we answer the above question by giving a martingale for the k-type Moran process on K_n. This generalises the martingale by Monk, Green, and Paulin [31] from 2 to any number of k types. Hence, the graph of interest throughout this section is the complete graph $G = K_n$.[22] Since the complete graph imposes a trivial structure among its vertices, the state-space can be simplified by only counting the number of vertices per type: We denote for any state $\sigma_t : V \to [k]$ by

$$\boldsymbol{S}_t := \begin{pmatrix} \left|\sigma_t^{-1}(1)\right| \\ \vdots \\ \left|\sigma_t^{-1}(k)\right| \end{pmatrix} \in (\mathbb{N}_0)^k, \tag{149}$$

The vector $\boldsymbol{S}_t$ denotes thus the state at time t, which we index by $i \in [k]$: $S_{t,i} = \left|\sigma_t^{-1}(i)\right|$. By definition, we have at any time $t \in \mathbb{N}_{\geq 0}$:

$$\sum_{i \in [k]} S_{t,i} = n. \tag{150}$$

For any $t \geq 0$, we can express the state $\boldsymbol{S}_t$ by its difference to the previous state:

$$\boldsymbol{S}_{t+1} = \boldsymbol{S}_t + \boldsymbol{X}_{t+1}, \tag{151}$$

where $\boldsymbol{X}_t \in \{-1, 0, +1\}^k$ with at most two entries being non-zero and $\sum_{i \in [k]} X_{t,i} = 0$. We denote by $W(\boldsymbol{S})$ the population's overall fitness in state S. Note that by definition of the Moran process, we have for every $t \in \mathbb{N}$, $i, j \in [k]$ with $i \neq j$ and every state $\boldsymbol{S}$:

$$\mathbb{P}\left[X_{t,i} = 1 \ \wedge \ X_{t,j} = -1 \mid \boldsymbol{S}_{t-1} = \boldsymbol{S}\right] = \frac{r_i S_i}{W(\boldsymbol{S})} \frac{S_j}{n-1}. \tag{152}$$

Further, $\boldsymbol{X}_t$ is zero if and only if no pair i, j of its components is non-zero:

$$\mathbb{P}\left[\boldsymbol{X}_t = \boldsymbol{0} \mid \boldsymbol{S}_{t-1}\right] = 1 - \sum_{\substack{i,j \in [k] \\ i \neq j}} \mathbb{P}\left[X_{t,i} = 1 \ \wedge \ X_{t,j} = -1 \mid \boldsymbol{S}_{t-1} = \boldsymbol{S}\right]. \tag{153}$$

Let us formalise our definition of $\boldsymbol{S}$ and define a function on $\boldsymbol{S}$.

---

[22] Recall that every graph considered in this work has no self-loops.

Definition 37. For the vector $\boldsymbol{S}$ defined by $\boldsymbol{S} := (|\sigma^{-1}(1)|, \dots, |\sigma^{-1}(k)|)$ for some state $\sigma : V \to [k]$, we define

$$f(\boldsymbol{S}) := \prod_{i \in [k]} r_i^{-S_i},$$

where $r_i > 0$ is the fitness of type i. If $\sigma = \sigma_t$, we write $\boldsymbol{S}_t$.

We claim that $f(\boldsymbol{S})$ is a martingale.

Main Theorem 7 (restated). *The sequence $\{f(\boldsymbol{S}_t) : t \geq 0\}$ is a martingale with respect to the sequence $\{\boldsymbol{X}_t : t \geq 0\}$, where we define $\boldsymbol{X}_0 \equiv \mathbf{0}$.*

Proof. Let $t \in \mathbb{N}_{\geq 0}$ be arbitrary. Note first that all r_i are strictly positive and finite and that $S_i \in \{0, \dots, n\}$ for any $i \in [n]$, so that $f(\boldsymbol{S}_t)$ is bounded and thus $\mathbb{E}[|f(\boldsymbol{S}_t)|] < \infty$.

For every $t \in \mathbb{N}$, we have

$$\begin{aligned} \mathbb{E}[f(\boldsymbol{S}_t) \mid \boldsymbol{X}_0, \dots, \boldsymbol{X}_{t-1}] &= \mathbb{E}[f(\boldsymbol{X}_t) \cdot f(\boldsymbol{S}_{t-1}) \mid \boldsymbol{X}_0, \dots, \boldsymbol{X}_{t-1}] && (154) \\ &= \mathbb{E}[f(\boldsymbol{X}_t) \mid \boldsymbol{X}_0, \dots, \boldsymbol{X}_{t-1}] \cdot f(\boldsymbol{S}_{t-1}), && (155) \end{aligned}$$

where the first equation follows by basic properties of the exponential function.[23] It thus suffices to show $\mathbb{E}[f(\boldsymbol{X}_t) \mid \boldsymbol{X}_0, \dots, \boldsymbol{X}_{t-1}] = 1$. Since $\boldsymbol{X}_t$ is either non-zero at exactly two locations or $\mathbf{0}$, we have for any state $\boldsymbol{S}$.

$$\begin{aligned} &\mathbb{E}[f(\boldsymbol{X}_t) \mid \boldsymbol{X}_0, \dots, \boldsymbol{X}_{t-1}] && (156) \\ &= \sum_{\substack{i,j \in [k] \\ i \neq j}} \mathbb{P}[X_{t,i} = 1 \ \wedge \ X_{t,j} = -1 \mid \boldsymbol{S}_{t-1} = \boldsymbol{S}] \cdot \frac{r_j}{r_i} + \mathbb{P}[\boldsymbol{X}_t = \mathbf{0} \mid \boldsymbol{S}_{t-1} = \boldsymbol{S}] \cdot 1 \\ & && \text{(by def. of } f(\boldsymbol{X}_t)) \\ &= \sum_{\substack{i,j \in [k] \\ i \neq j}} \frac{r_i S_i}{W(\boldsymbol{S})} \frac{S_j}{n-1} \frac{r_j}{r_i} + \left(1 - \sum_{\substack{i,j \in [k] \\ i \neq j}} \frac{r_i S_i}{W(\boldsymbol{S})} \frac{S_j}{n-1} \right) && \text{(by eqs. (152) and (153))} \end{aligned}$$

This quantity is 1 if and only if

$$\sum_{\substack{i,j \in [k] \\ i \neq j}} \frac{r_i S_i}{W(\boldsymbol{S})} \frac{S_j}{n-1} \frac{r_j}{r_i} = \sum_{\substack{i,j \in [k] \\ i \neq j}} \frac{r_i S_i}{W(\boldsymbol{S})} \frac{S_j}{n-1} \tag{157}$$

holds. Multiplying by $W(\boldsymbol{S}) \cdot (n-1)$ yields

$$\sum_{\substack{i,j \in [k] \\ i \neq j}} r_j S_i S_j = \sum_{\substack{i,j \in [k] \\ i \neq j}} r_i S_i S_j \tag{158}$$

which holds by symmetry of i, j. We can thus conclude that

$$\mathbb{E}[f(\boldsymbol{S}_t) \mid \boldsymbol{X}_0, \dots, \boldsymbol{X}_{t-1}] = f(\boldsymbol{S}_{t-1}) \tag{159}$$

for any $t \in \mathbb{N}$, concluding the proof that $\{f(S_n) : n \geq 0\}$ is a martingale with respect to the sequence $\{X_n : n \geq 0\}$. ☐

---

[23] More precisely, this step is due to the exponential function being a homomorphism from the additive group of $\mathbb{R}$ to the multiplicative group of $\mathbb{R}^+$.

We define the stopping time

$$\tau := \min\left\{ t \in \mathbb{N}_0 \;\middle|\; \boldsymbol{S}_t \in \left\{ \begin{pmatrix} n \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ n \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ n \end{pmatrix} \right\} \right\}. \tag{160}$$

Since $|f(\boldsymbol{S}_t)|$ is bounded for every $t \leq \tau$, the optional stopping theorem applies, yielding

$$\mathbb{E}\left[f(\boldsymbol{S}_\tau)\right] = \mathbb{E}\left[f(\boldsymbol{S}_0)\right]. \tag{161}$$

Since $\boldsymbol{S}_\tau$ is with probability $f^j_{G,\boldsymbol{r}}$ the vector $n \cdot \boldsymbol{e}_j$, where $\boldsymbol{e}_j$ denotes the j-th basis vector in the standard basis of $\mathbb{R}^k$ for every type $j \in [k]$, eq. (161) translates into

$$\prod_{j\in[k]} r_j^{-S_{0,j}} = \prod_{j\in[k]} f^j_{G,\boldsymbol{r}} \cdot r_j^{-n}. \tag{162}$$

The martingale yields thus in addition to the trivial equation

$$\sum_{j\in[k]} f^j_{G,\boldsymbol{r}} = 1 \tag{163}$$

a second equation to determine the fixation probabilities, of which we have k many.

4.1 Applying the martingale: Solving for the fixation probabilities in a special case

We now demonstrate the usefulness of the martingale in the simple case where all invading mutations have an equal fitness advantage, *i.e.* $r_2 = \ldots = r_k$. We let each of the types start with any number of vertices initially so that the initial state is an arbitrary vector $\boldsymbol{S}_0 \in \mathbb{N}^k$ with the only requirement $\sum_{j\in[k]} S_{0,j} = n$. We need the following lemma.

Lemma 38. *For the k-type Moran process on $G = K_n$ with fitness $\boldsymbol{r}$ satisfying $r_1 < r_2 = r_3 = \ldots = r_k$, and any starting state $\boldsymbol{S}_0$, we have*

$$\frac{f^i_{G,\boldsymbol{r}}}{f^j_{G,\boldsymbol{r}}} = \frac{S_{0,i}}{S_{0,j}}.$$

for every $i, j \in \{2, \dots, k\}$.

Proof. Consider the process where each of the $l := S_{0,2} + S_{0,3} + \ldots + S_{0,k}$ many type $2, \dots, k$ vertices start with their own type and enumerate those by $1, \dots, l$. We remap the former type 1 to type $l + 1$, and are not interested in it for the remainder of this prove anymore.

Consider the event

$$\mathcal{A}_x := \text{“The new type } x \text{ fixates”}$$

for any $x \in [l]$. All of these l many new types have the same fitness (namely $r_2 = \ldots = r_k$), so that their fixation probabilities are equal:

$$\forall x, y \in [l]: \quad \mathbb{P}\left[\mathcal{A}_x\right] = \mathbb{P}\left[\mathcal{A}_y\right].$$

Since eventually one and only one type fixates in the Moran process, the events $\mathcal{A}_x, \mathcal{A}_y$ are mutually exclusive for every $x, y \in [l]$, $x \neq y$. Therefore, for any initial type $i \in \{2, \dots, k\}$ we have

$$f^i_{G,\boldsymbol{r}} = S_{0,i} \cdot \mathbb{P}\left[\mathcal{A}_1\right], \tag{164}$$

since type i fixates if and only if any of its $S_{0,i}$ initial vertices fixates. Since these events are mutually exclusive and all equal to $\mathcal{A}_1$ as argued above, we can simply multiply $S_{0,i} \cdot \mathbb{P}[\mathcal{A}_1]$.

The same argument holds for type $j \in [k]$, so that we get

$$\frac{f^i_{G,\boldsymbol{r}}}{f^j_{G,\boldsymbol{r}}} = \frac{S_{0,i} \cdot \mathbb{P}[\mathcal{A}_1]}{S_{0,j} \cdot \mathbb{P}[\mathcal{A}_1]} = \frac{S_{0,i}}{S_{0,j}}. \tag{165}$$

□

From Lemma 38 we get $k-2$ many equations:

$$\forall j \in \{2, \dots, k-1\}: \quad \frac{f^j_{G,\boldsymbol{r}}}{f^{j+1}_{G,\boldsymbol{r}}} = \frac{S_{0,j}}{S_{0,j+1}}. \tag{166}$$

Together with eqs. (162) and (163) we thus have k equations for the k unknowns $f^1_{G,\boldsymbol{r}}, \dots, f^k_{G,\boldsymbol{r}}$. Solving this non-singular system of equations yields the following exact formulas for the fixation probabilities.

Corollary 39. *For the k-type Moran process on $G = K_n$ with fitness $\boldsymbol{r}$ satisfying $r_1 < r_2 = r_3 = \ldots = r_k$, and every starting state $\sigma_0 : V \to [k]$, the fixation probabilities are*

$$f^1_{G,\boldsymbol{r}}(\sigma_0) = \frac{\left(\frac{r_1}{r_2}\right)^n - \left(\frac{r_1}{r_2}\right)^l}{\left(\frac{r_1}{r_2}\right)^n - 1},$$

$$\forall j \in \{2, \dots, k\}: \quad f^j_{G,\boldsymbol{r}}(\sigma_0) = \frac{\left|\sigma_0^{-1}(j)\right| \left(\left(\frac{r_1}{r_2}\right)^{i+j} - 1\right)}{l\left(\left(\frac{r_1}{r_2}\right)^n - 1\right)},$$

where $l := \sum_{i=2}^{k} \left|\sigma_0^{-1}(i)\right|$.

In this section, we have thus resolved a question posed by Monk and Schaik by providing a martingale for any number of $k \geq 2$ types. We have further demonstrated how it can be used in a particular instance of the Moran process, but note that the martingale's usefulness depends on having further equations to calculate the k unknown fixation probabilities. This leads to further research questions described in the subsequent Section 5.

5 Outlook

In this dissertation, we generalised the spatial Moran process to any number of k mutants and provided an FPRAS for Moran-$\boldsymbol{r}$, the analogue computational problem which has been efficiently approximated in the 2-type process.

We further introduced the parameterised computational problem p-Moran, which treats the fitness vector $\boldsymbol{r}$ and its dimension, k, as general parameters. We are able to approximate this problem in the case of p-MoranMax as well as in the promised version p-Moran-c-ThFittest by FPTRASes, and in some instances even by FPRASes independent of k. At the heart of these algorithms lies a stochastic coupling which provides bounds on the fixation probabilities in a very general setting and is thus interesting on its own; as a narrow subcase, we applied it to derive bounds given by Ferreira and Neves in recent work. We motivate further research into the parameterised problem p-Moran for weak mutations: Does p-Moran have an FPTRAS if $j = 2$?

In Section 2.3, we proved that the present techniques do not yield asymptotic speed-up if only active steps are simulated, distinguishing the $k = 2$ from the $k > 2$ type Moran process. Further research could ask: Do other potential functions provide a speedup? One could try a potential function of the form

$$\Psi'(\sigma) := \sum_{j\in[k]} \sum_{v\in\sigma^{-1}(j)} \frac{1}{\mathsf{deg}(v)} \cdot p(j), \tag{167}$$

where $p(j)$ is some function in j. We suggest $p(j)$ to be monotonically increasing to give more weight to the highest mutation type while maintaining $p(k)$ small enough to keep a small bound in Lemma 12.

In Section 4, we gave a martingale for every k-type Moran process, thereby solving an open problem by Monk and Schaik. We motivate further research into finding more equations which can be used in combination with our martingale (Definition 37) to give exact formulas for the fixation probabilities on the complete graph without restrictions on the fitness $\boldsymbol{r}$.

References

[1] Leslie Ann Goldberg, John Lapinskas, and David Richerby. Phase transitions of the Moran process and algorithmic consequences. *Random Structures & Algorithms*, 56(3):597–647, 2020. doi:10.1002/rsa.20890.

[2] V. Arvind and Venkatesh Raman. Approximation Algorithms for Some Parameterized Counting Problems. In *Proceedings of the 13th International Symposium on Algorithms and Computation*, ISAAC '02, pages 453–464, Berlin, Heidelberg, November 2002. Springer-Verlag.

[3] Ann-Marie Baker, Trevor A Graham, and Nicholas A Wright. Pre-tumour clones, periodic selection and clonal interference in the origin and progression of gastrointestinal cancer: Potential for biomarker development. *The Journal of Pathology*, 229(4):502–514, 2013. doi:10.1002/path.4157.

[4] N. H. Barton. The probability of fixation of a favoured allele in a subdivided population. *Genetical Research*, 62(2):149–157, October 1993. doi:10.1017/S0016672300031748.

[5] Niko Beerenwinkel, Tibor Antal, David Dingli, Arne Traulsen, Kenneth W. Kinzler, Victor E. Velculescu, Bert Vogelstein, and Martin A. Nowak. Genetic Progression and the Waiting Time to Cancer. *PLOS Computational Biology*, 3(11):e225, November 2007. doi:10.1371/journal.pcbi.0030225.

[6] Niko Beerenwinkel, Roland F. Schwarz, Moritz Gerstung, and Florian Markowetz. Cancer Evolution: Mathematical Models and Computational Inference. *Systematic Biology*, 64(1):e1–e25, January 2015. doi:10.1093/sysbio/syu081.

[7] Béla Bollobás, Hugo Duminil-Copin, Robert Morris, and Paul Smith. Universality of two-dimensional critical cellular automata, June 2020. URL: http://arxiv.org/abs/1406.6680, arXiv:1406.6680.

[8] Béla Bollobás, Paul Smith, and Andrew Uzzell. Monotone Cellular Automata in a Random Environment. *Combinatorics, Probability and Computing*, 24(4):687–722, July 2015. doi:10.1017/S0963548315000012.

[9] Maury Bramson and David Griffeath. On the Williams-Bjerknes tumour growth model: II. *Mathematical Proceedings of the Cambridge Philosophical Society*, 88(2):339–357, September 1980. doi:10.1017/S0305004100057650.

[10] Krishnendu Chatterjee, Rasmus Ibsen-Jensen, and Martin A. Nowak. Faster monte-carlo algorithms for fixation probability of the moran process on undirected graphs. In *42nd International Symposium on Mathematical Foundations of Computer Science*, volume 83 of *LIPIcs*, pages 61:1–61:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPIcs.MFCS.2017.61.

[11] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, 3rd Edition.* MIT Press, 2009. URL: http://mitpress.mit.edu/books/introduction-algorithms.

[12] Marek Cygan, Fedor V. Fomin, Łukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michał Pilipczuk, and Saket Saurabh. *Parameterized Algorithms.* Springer International Publishing, Cham, 2015. doi:10.1007/978-3-319-21275-3.

[13] Josep Díaz, Leslie Ann Goldberg, George B. Mertzios, David Richerby, Maria Serna, and Paul G. Spirakis. Approximating Fixation Probabilities in the Generalized Moran Process. *Algorithmica*, 69(1):78–91, May 2014. `doi:10.1007/s00453-012-9722-7`.

[14] Josep Díaz, Leslie Ann Goldberg, David Richerby, and Maria Serna. Absorption time of the Moran process. *Random Structures & Algorithms*, 49(1):137–159, 2016. `doi:10.1002/rsa.20617`.

[15] Khanh N. Dinh, Seth J. Corey, and Marek Kimmel. Application of the Moran Model in Estimating Selection Coefficient of Mutated CSF3R Clones in the Evolution of Severe Congenital Neutropenia to Myeloid Neoplasia. *Frontiers in Physiology*, 11:806, September 2020. `doi:10.3389/fphys.2020.00806`.

[16] Rod G. Downey and Michael R. Fellows. Fixed-Parameter Tractability and Completeness I: Basic Results. *SIAM Journal on Computing*, 24(4):873–921, August 1995. `doi:10.1137/S0097539792228228`.

[17] Rick Durrett, Thomas Liggett, and Yuan Zhang. The contact process with fast voting. *Electronic Journal of Probability*, 19(none), January 2014. `doi:10.1214/EJP.v19-3021`.

[18] Juan Fernández-Gracia, Krzysztof Suchecki, José J. Ramasco, Maxi San Miguel, and Víctor M. Eguíluz. Is the Voter Model a Model for Voters? *Physical Review Letters*, 112(15):158701, April 2014. `doi:10.1103/PhysRevLett.112.158701`.

[19] Eliza M. Ferreira and Armando G. M. Neves. Fixation probabilities for the Moran process with three or more strategies: General and coupling results. *Journal of Mathematical Biology*, 81(1):277–314, July 2020. `doi:10.1007/s00285-020-01510-0`.

[20] R. A. Fisher. *The Genetical Theory of Natural Selection.* Oxford University Press, Oxford, 1930.

[21] Leslie Ann Goldberg. Probability and Computing, Oxford 2021-22, Lecture Notes, March 2022. URL: `https://www.cs.ox.ac.uk/files/13232/LectureNotes.pdf`.

[22] Leslie Ann Goldberg, John Lapinskas, Johannes Lengler, Florian Meier, Konstantinos Panagiotou, and Pascal Pfister. Asymptotically Optimal Amplifiers for the Moran Process, August 2018. `arXiv:1611.04209`, `doi:10.48550/arXiv.1611.04209`.

[23] Yipei Guo and Ariel Amir. The effect of weak clonal interference on average fitness trajectories in the presence of macroscopic epistasis. *Genetics*, 220(4):iyac028, April 2022. `doi:10.1093/genetics/iyac028`.

[24] Bruce Hajek. Hitting-Time and Occupation-Time Bounds Implied by Drift Analysis with Applications. *Advances in Applied Probability*, 14(3):502–525, 1982. `doi:10.2307/1426671`.

[25] John G Kemeny and J. Laurie Snell. *Finite Markov Chains,.* Van Nostrand, Princeton, N.J., 1960.

[26] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146, New York, NY, USA, August 2003. Association for Computing Machinery. `doi:10.1145/956750.956769`.

[27] John B. Kogut. An introduction to lattice gauge theory and spin systems. *Reviews of Modern Physics*, 51(4):659–713, October 1979. doi:10.1103/RevModPhys.51.659.

[28] Erez Lieberman, Christoph Hauert, and Martin A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, January 2005. doi:10.1038/nature03204.

[29] Thomas M Liggett. *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes.* Springer, 1999. URL: https://ebookcentral.proquest.com/lib/concordiaab-ebooks/detail.action?docID=3098239.

[30] Thomas M. Liggett and Thomas M. Liggett. *Interacting Particle Systems.* Classics in Mathematics. Springer, Berlin New York, 2005.

[31] T. Monk, P. Green, and M. Paulin. Martingales and fixation probabilities of evolutionary graphs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 470(2165):20130730, May 2014. doi:10.1098/rspa.2013.0730.

[32] Travis Monk and André van Schaik. Martingales and the characteristic functions of absorption time on bipartite graphs. *Royal Society Open Science*, 8(10), 2021. doi:10.1098/rsos.210657.

[33] P. A. P. Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(1):60–71, January 1958. doi:10.1017/S0305004100033193.

[34] Su-Chan Park and Joachim Krug. Clonal interference in large populations. *Proceedings of the National Academy of Sciences*, 104(46):18135–18140, November 2007. doi:10.1073/pnas.0705778104.

[35] Diane L. Schwartz. Applications of Duality to a Class of Markov Processes. *The Annals of Probability*, 5(4):522–532, 1977. URL: https://www.jstor.org/stable/2243074.

[36] Bert Vogelstein, Nickolas Papadopoulos, Victor E. Velculescu, Shibin Zhou, Luis A. Diaz, and Kenneth W. Kinzler. Cancer Genome Landscapes. *Science (New York, N.Y.)*, 339(6127):1546–1558, March 2013. doi:10.1126/science.1235122.

[37] Ewens J. Warren. *Mathematical Population Genetics 1*, volume 27 of *Interdisciplinary Applied Mathematics.* Springer, 2004.

[38] Trevor Williams and Rolf Bjerknes. Stochastic Model for Abnormal Clone Spread through Epithelial Basal Layer. *Nature*, 236(5340):19–21, March 1972. doi:10.1038/236019a0.

[39] Sewall Wright. Evolution in Mendelian Populations. *Genetics*, 16(2):97–159, 1931.

Appendices

A Proof for the subset stochastic domination result

We now prove the following stochastic domination result.

Theorem 30 (restated). *For any $k \geq 2$ and $y \in [k]$, let $\tilde{M} := \{\tilde{\sigma}_t \mid t \in \mathbb{R}_{\geq 0}\}$ be a k-type continuous time Moran process with non-decreasing fitness vector $\boldsymbol{r}$ on a graph $G = (V, E)$. Let $\tilde{M}' := \{\tilde{\sigma}'_t \mid t \in \mathbb{R}_{\geq 0}\}$ be any other k-type continuous time Moran process on G with the unordered fitness vector $\boldsymbol{r}'$ such that $\forall i \in \{1, \ldots, y\} : r'_i = r_i$, $\forall i \in \{y+1, \ldots, k\} : r'_i \leq r_{y+1}$. Visually, this translates to*

$$\boldsymbol{r}' = \begin{pmatrix} r'_1 \\ \vdots \\ r'_y \\ r'_{y+1} \\ \vdots \\ r'_k \end{pmatrix} \begin{matrix} = r_1 \\ \vdots \\ = r_y \\ \leq r_{y+1} \\ \vdots \\ \leq r_{y+1}. \end{matrix} \tag{168}$$

Then there exists a coupling between $\tilde{M}$ and $\tilde{M}'$ such that at any point in time $t \in \mathbb{R}_{\geq 0}$,

$$\forall j \in \{1, \ldots, y\} : \; \tilde{\sigma}_t^{-1}(j) \subseteq \tilde{\sigma}_t'^{-1}(j) \tag{169}$$

implies

$$\forall j \in \{1, \ldots, y\} : \; \tilde{\sigma}_{t+dt}^{-1}(j) \subseteq \tilde{\sigma}_{t+dt}'^{-1}(j), \tag{170}$$

where dt is the time it takes for the first reproduction to happen in $\tilde{M}$ or $\tilde{M}'$ after time t.

Proof of Theorem 30. We proceed as in the proof of Theorem 26 by constructing this coupling explicitly. We point the necessary modifications to the proof of Theorem 26 out.

The random variables t_v, t_v^1, t_v^2, clock set C, and $dt := \min C$ are defined as in Theorem 26. Both chains evolve as in Theorem 26. Suppose eq. (169) holds.

We now show that eq. (170) indeed holds. Suppose first that both chains evolve synchronously. It suffices to show that for all $j \in \{1, \ldots, y\}$ (i) if $v \in \tilde{\sigma}_t^{-1}(j)$ reproduces to w, then also $w \in \tilde{\sigma}_{t+dt}'^{-1}(j)$ as well as (ii) if $w \notin \tilde{\sigma}_{t+dt}'^{-1}(j)$ dies, then also $w \notin \tilde{\sigma}_{t+dt}^{-1}(j)$.
Let $j \in \{1, \ldots, y\}$ be arbitrary. For (i), suppose $v \in \tilde{\sigma}_t^{-1}(j)$ reproduces to w. Since $\tilde{\sigma}_t^{-1}(j) \subseteq \tilde{\sigma}_t'^{-1}(j)$ by assumption (eq. (169)), we must have $v \in \tilde{\sigma}_t'^{-1}(j)$. Since $\tilde{M}'$ evolves as well, we have $w \in \tilde{\sigma}_{t+dt}'^{-1}(j)$, concluding case (i). For (ii), suppose $w \notin \tilde{\sigma}_{t+dt}'^{-1}(j)$ dies. Thus, by the evolution of $\tilde{M}'$, we must have $v \notin \tilde{\sigma}_t'^{-1}(j)$. Since $\tilde{\sigma}_t^{-1}(j) \subseteq \tilde{\sigma}_t'^{-1}(j)$ (eq. (169)), we must have $v \notin \tilde{\sigma}_t^{-1}(j)$. Since $\tilde{M}$ evolves as well, this means $w \notin \tilde{\sigma}_{t+dt}^{-1}(j)$.

Suppose now that the chains do not evolve synchronously. To that end, we define the partitions $S_{i,<}, S_{i,>}, S_{i,=}$ for $i \in [k]$ as in Theorem 26. We perform a case distinction on the vertex v chosen for reproduction.

Case $v \in S_{i,<}$ for $i \in \{1, \ldots, y\}$. Thus, $r_v < r'_v$. We show that this case is impossible. Since $\tilde{\sigma}_t(v) = i \leq y$, we must have by eq. (169) $\tilde{\sigma}'_t(v) = i$. Thus, $r_v = r'_v$, contradicting $r_v < r'_v$.

Case $v \in S_{i,<}$ for $i \in \{y+1, \ldots, k\}$. Note that if $y = k$, this case holds trivially. Otherwise, we show that this case is impossible. We have

$$\begin{aligned} r_{y+1} &\leq r_v && \text{(since } y+1 \leq \tilde{\sigma}_t(v) \text{ and } \boldsymbol{r} \text{ is non-decreasing)} \\ &< r'_v && \text{(by def. of } S_{i,<}) \end{aligned}$$

However, any fitness in $\boldsymbol{r}'$ is by definition at most r_{y+1}, contradicting $r_{y+1} < r'_v$.

Case $v \in S_{i,>}$ for $i \in \{1, \dots, y\}$. Thus, $r_v > r'_v$. We show that this case is impossible. Since $\tilde{\sigma}_t(v) = i \leq y$, we must have by eq. (169) $\tilde{\sigma}'_t(v) = i$. Thus, $r_v = r'_v$, contradicting $r_v > r'_v$.

Case $v \in S_{i,>}$ for $i \in \{y+1, \dots, k\}$. Note that if $y = k$, this case holds trivially. Since we assume that only one chain evolves and since $r_v > r'_v$, only $\tilde{M}$ evolves. Since $i > y$, have $v \notin \tilde{\sigma}_t^{-1}(\{1, \dots, y\})$, allowing us to conclude for any $j \in \{1, \dots, y\}$:

$$\begin{aligned} \tilde{\sigma}_{t+dt}^{-1}(j) &\subseteq \tilde{\sigma}_t^{-1}(j) && \text{(since } v \notin \tilde{\sigma}_t^{-1}(\{j\}) \text{ reproduces.)} \\ &\subseteq \tilde{\sigma}_t'^{-1}(j) && \text{(by assumption eq. (169))} \\ &= \tilde{\sigma}_{t+dt}'^{-1}(j). && \text{(since } \tilde{M}' \text{ does not evolve)} \end{aligned}$$

We can therefore conclude that eq. (170) does indeed hold at time $t + dt$.

The proof that we constructed indeed faithful copies of $\tilde{M}$, $\tilde{M}'$ is identical to the faithfulness proof of Theorem 26. ☐

B Ferreira and Neves's bounds on K_n with 3 types: A corollary from Theorem 26

Ferreira and Neves analyse the $k = 3$ case on the complete graph and obtain bounds for the fixation probabilities [19]. These bounds follow just as a very niche case of Theorem 26. This section section provides these bounds as a corollary of Theorem 26, demonstrating the theorem's generality.

Corollary 40. *In the 3-type Moran process on the complete graph $G = K_n$ with non-increasing fitness vector $\boldsymbol{r}$, we have for every state $\sigma : V \to [3]$*

$$\begin{aligned} \frac{1 - (r_3/r_1)^{|\sigma^{-1}(1)|}}{1 - (r_3/r_1)^n} &\leq f^1_{K_n,\boldsymbol{r}} \leq \frac{1 - (r_2/r_1)^{|\sigma^{-1}(1)|}}{1 - (r_2/r_1)^n} \\ \frac{1 - (r_3/r_2)^{|\sigma^{-1}(2)|}}{1 - (r_3/r_2)^n} &\leq f^2_{K_n,\boldsymbol{r}} \leq \frac{1 - (r_1/r_2)^{|\sigma^{-1}(2)|}}{1 - (r_1/r_2)^n} \\ \frac{1 - (r_2/r_3)^{|\sigma^{-1}(3)|}}{1 - (r_2/r_3)^n} &\leq f^3_{K_n,\boldsymbol{r}} \leq \frac{1 - (r_1/r_3)^{|\sigma^{-1}(3)|}}{1 - (r_1/r_3)^n}. \end{aligned}$$

Proof. We prove the case of $f^1_{K_n,\boldsymbol{r}}$ here. The other cases follow analogously. For the first inequality, adopting the notation from Theorem 26, we set $x = y = 1$ and $\boldsymbol{r}' := (r_1, r_3, r_3)$. Both chains $\tilde{M} := \{\tilde{\sigma}_t \mid t \in \mathbb{R}_{\geq 0}\}$ and $\tilde{M}' := \{\tilde{\sigma}'_t \mid t \in \mathbb{R}_{\geq 0}\}$ start with initial state σ: $\tilde{\sigma}_0 := \sigma$, $\tilde{\sigma}'_0 := \sigma$. Applying Theorem 26 inductively, we get

$$\forall t \geq 0 : \quad \tilde{\sigma}_t^{-1}(1) \supseteq \tilde{\sigma}_t'^{-1}(1) \tag{171}$$

and thereby

$$f^1_{K_n,\boldsymbol{r}} \geq f^1_{K_n,\boldsymbol{r}'}, \tag{172}$$

which holds in both the discrete and continuous time processes, as argued in Section 3.2.1. Since we have $r'_2 = r'_3$, we can in $\tilde{M}'$ simply remap type 3 to type 2, without changing the behaviour of type 1. We thus have a 2-type Moran process, with fitness vector (r_1, r_3) and $|\sigma^{-1}(1)|$ many initial type-1 vertices. The type 2-process on K_n has been well studied, so that we know *e.g.* from [13] that in this process, type 1 fixates with probability

$$\frac{1 - (r_3/r_1)^{|\sigma^{-1}(1)|}}{1 - (r_3/r_1)^n}, \tag{173}$$

yielding by Equation (172)

$$\frac{1-(r_3/r_1)^{|\sigma^{-1}(1)|}}{1-(r_3/r_1)^n} \leq f^1_{K_n,\boldsymbol{r}}. \tag{174}$$

For the upper bound on $f^1_{K_n,\boldsymbol{r}}$, take in Theorem 26 the fitness of $\tilde{M}$ to be (r_1, r_2, r_2) and the fitness of $\tilde{M}'$ to be (r_1, r_2, r_3). We see that $\tilde{M}$ can just be seen as a 2-type process where type 1 fixates with probability

$$\frac{1-(r_2/r_1)^{|\sigma^{-1}(1)|}}{1-(r_2/r_1)^n}, \tag{175}$$

so that an inductive application of Theorem 26 yields

$$f^1_{K_n,\boldsymbol{r}} \leq \frac{1-(r_2/r_1)^{|\sigma^{-1}(1)|}}{1-(r_2/r_1)^n}. \tag{176}$$

This shows the lower and upper bound on $f^1_{K_n,\boldsymbol{r}}$.

The proofs for $f^2_{K_n,\boldsymbol{r}}$ and $f^3_{K_n,\boldsymbol{r}}$ are symmetric; we take $x = y = 2$ and $x = y = 3$ respectively in the application of Theorem 26. ☐

Corollary 40 are indeed the bounds from Ferreira and Neves (Theorem 4 in [19]). Note that our Theorem 26, in contrast, works for arbitrary graphs and arbitrary number of mutants.

C Connection of the Lieberman, Hauert, and Nowak model to the biased voter model and the Williams-Bjerknes tumour growth model

In this section, we will consider the Moran process generalised to continuous time (Section 3.2.1). This has been done for $k = 2$ by Diaz *et al.* to establish a coupling theorem [14]. We show how closely related the spatial Moran process proposed by Lieberman, Hauert, and Nowak [28] is to the *biased* voter and Williams-Bjerknes model.

The Williams-Bjerknes process and biased voter model The Williams-Bjerknes process is the spin system on the finite 2-dimensional grid graph G, whose vertices can be in either state 1 or state 2. If we denote by S_t the set of type 2-vertices at time $t \in \mathbb{R}$, then $(S_t)_{t\geq 0}$ is the Markov process whose jump rates are given by[24]

$$\begin{aligned} S &\mapsto S \cup \{v\} \qquad \text{at rate } \kappa \cdot |S \cap N(v)| \\ S &\mapsto S \setminus \{v\} \qquad\quad \text{at rate } |S \cap N(v)| \end{aligned} \tag{177}$$

for every $v \in V$ and any constant $\kappa > 1$. This has been described by Williams and Bjerknes in 1972 [38] and further formalised by Bramson and Griffeath [9]. For G being the infinite sized is the d-dimensional lattice ($d \geq 1$), this model is known as the biased voter model, introduced independently by Schwartz in 1977 [35]. The difference in the underlying graph stems from Williams and Bjerknes considering tumour growth and Schwartz working on interacting particle systems.

---

[24] We use the convention that "rate 0" means "not happening".

To see the connection to the model introduced by Lieberman, Hauert, and Nowak, let G be the d-dimensional lattice for any $d \geq 1$. Thus, each vertex has degree $2d$. Consider the 2-type continuous time Moran process with fitness vector $(2d, r \cdot 2d)$ for any $r > 1$ on G.[25] By the definition of the continuous time Moran process, reproduction happens from u to v at rate $\frac{r_{\sigma(v)} \cdot 2d}{2d} = r_{\sigma(v)}$ for any $\{u, v\} \in E$. If we denote by S the set of type-2 vertices, we have thus

$$\begin{aligned} S &\mapsto S \cup \{v\} && \text{at rate } r \cdot |S \cap N(v)| \\ S &\mapsto S \setminus \{v\} && \text{at rate } |S \cap N(v)| \end{aligned} \tag{178}$$

which is equivalent to (177) with $r = \kappa$. Since the Moran process as proposed by Lieberman, Hauert, and Nowak in 2005 is just the embedded discrete time Markov chain of the process considered above, we conclude that for regular graphs, their model is just the discrete version of the biased voter model generalised to finite graph.

D Code to simulate the k-type Moran process

In this section, we provide code in `C++` to simulate the k-type Moran process on every connected graph. In Appendix E, we use this code to provide simulation results on the complete graph for different values of k.

```cpp
//  Simulating the Moran process with k types
//  Author: Tassilo Schwarz

#include "simulation.hpp"

/**
Simulate one run of the k-type Moran process on a connected graph of order n.
 @param fitness r_1, ..., r_k: double vector of fitness of each of the k types
 @param number_initial_types number of each type in the uar initialisation. Must sum to n.
 @param degrees array of the vertex degrees
 @param adjacency_list adjacency list (in the algorithmic sense)  of the graph, encoded as a n x n array (in the data structure sense). This provides both algorithmic and data structural speedup. adjacency_list[i][j] for j>= degrees[i] are ignored.
 @param neighbour_distr List of uniform distribution over neighbours, as returned by function @b convert_graph
 @returns type of fixated mutant
 */
int simulate(array<double,k> fitness,array<int,k> &number_initial_types,array<int,n> &degrees,array<array<int, n>, n> &adjacency_list,array<uniform_int_distribution<>,n> &neighbour_distr){
    // size of graph must be bigger than number of types
    assert(n>k);

    // verify that number_initial_types is well-defined.
```

---

[25] We normalised the first component of the fitness vector to $2d$, as described in Section 1.1.

```
    assert(accumulate(begin(number_initial_types), end(number_initial_types), 0)==n);

    // the pseudo random generator
    random_device rd;
    mt19937 gen(rd());

    array<int,n> sigma; // the type mapping, representing the current state.

    // initialise the state by selecting number_initial_types of each type
    int init_counter=0;
    for(int j=0;j<k;j++){
        for(int i = 0;i<number_initial_types[j];i++){
            sigma[init_counter++]=j;
        }
    }
    shuffle ( begin(sigma),end(sigma) ,gen);

    // fitness of each vertex
    array<double,n> fitness_per_vertex;
    for(int i=0;i<n;i++){
        fitness_per_vertex[i] = fitness[sigma[i]];
    }

    // #vertices per each type
    array<int,k> type_counter;
    for(int i=0;i<k;i++){
        type_counter[i]=number_initial_types[i];
    }

    // weighted types. i.e. #type-i * r_i for any i in [k]
    array<double,k> weighted_types;
    for(int i=0;i<k;i++){
        weighted_types[i]=number_initial_types[i]*fitness[i];
    }

    // 2d array of vertices of each type at a given time. Used for efficiency.
    array<array<int,n>,k> type_lists={{-1}};
    array<int,k> type_lists_idx={0};
    for(int i=0;i<n;i++){
        type_lists[sigma[i]][type_lists_idx[sigma[i]]++]=i;
    }

    // check that updates make sense
    for(int i=0;i<k;i++){
        assert(type_lists_idx[i]==type_counter[i]);
    }

    // evolve the process until it converged
```

```
    while(true){
        // reproduce v to u.

        // 1. select v
        discrete_distribution<int> distr_v(begin(fitness_per_vertex), end(fitness_per_vertex));
        int reproduction_v = distr_v(gen);

        // 2. select u
        int target_u = adjacency_list[reproduction_v][neighbour_distr[reproduction_v](gen)];

        // 3. update the state, if changed
        if(sigma[target_u] != sigma[reproduction_v]){
            int reproduction_v_type = sigma[reproduction_v];
            int old_type_u = sigma[target_u];

            type_counter[reproduction_v_type]++;
            assert(type_counter[reproduction_v_type]<=n);

            type_counter[old_type_u]--;
            assert(type_counter[old_type_u] >=0);

            sigma[target_u] = reproduction_v_type;

            fitness_per_vertex[target_u] = fitness[reproduction_v_type];

            // terminate, if fixated
            if(type_counter[sigma[reproduction_v]]==n){
                return sigma[reproduction_v];
            }
        }
    }
    return -1;
}

/**
 Converts and adjacency matrix to a degree list and adjacency list
 @param adjacency_matrix Reference to the n x n binary adjacency matrix (input)
 @param degrees Reference to list of degrees (output)
 @param adjacency_list Reference to list adjacency list, encoded as a n x n array for efficiency (output)
 @param neighbour_distr List of Uniform[degrees[i]] distributions, for i in [n]. The i-th distribution is uniform on 0...degrees[i]-1 and thus can be used to sample from neighbours
 */
void convert_graph(array<array<int, n>, n> &adjacency_matrix,array<int, n> &degrees,array<array<int, n>, n> &adjacency_list,array<uniform_int_distribution<>,n> &neighbour_distr){
```

```

    for(int i=0;i<n;i++){
        int list_idx=0;
        int degree_cnt=0;
        for(int j=0;j<n;j++){
            if(adjacency_matrix[i][j]==1){
                adjacency_list[i][list_idx++]=j;
                degree_cnt++;
            }
        }
        degrees[i]=degree_cnt;
        neighbour_distr[i]=uniform_int_distribution<>(0,degree_cnt-1);
    }
}


/**
 Simulates N runs of the Moran process on a connected graph.
 @param N number of simulations
 @param fitness fitnesses r_1, ... , r_k
 @param number_initial_types number of each type initially. Must sum to n.
 @param simulation_results Vector storing #fixations per type
 */
void simulate_N_runs(int N,array<double,k> fitness,array<int,k> &number_initial_types,array<array<int, n>, n> &adjacency_matrix,array<int,k> &simulation_results){
    simulation_results = {0};


    array<int,n> degrees={0};
    array<array<int, n>, n> adjacency_list={0};
    array<uniform_int_distribution<>,n> neighbour_distr;

    convert_graph(adjacency_matrix, degrees, adjacency_list,neighbour_distr);

    for(int i=0;i<N;i++){
        simulation_results[simulate(fitness,number_initial_types,degrees, adjacency_list,neighbour_distr)]++;
        if(i%(N/20)==0 && i>0){
            cout<<"["<<setw(2)<<"] progress: "<<1.0*i/N<<endl;
        }
    }
}
```

The associated header file *"simulation.hpp"* is as follows.

```
//  Simulating the Moran process with k types. Header file.
//  Author: Tassilo Schwarz

#ifndef simulation_hpp
#define simulation_hpp
```

```

#include <stdio.h>
#include <iostream>
#include <stdlib.h>     /* srand, rand */
#include <time.h>       /* time */
#include <iomanip>
#include <map>
#include <algorithm>    // random_shuffle
#include <random>
#include <array>
#include <list>
#include <string>

#define k 15 // the number of mutants
#define n 1000 // the number of nodes. n has to be greater than k

using namespace std;

int simulate(array<double,k> fitness,array<int,k>
 ↪  &number_initial_types,array<int,n> &degrees,array<array<int, n>, n>
 ↪  &adjacency_list,array<uniform_int_distribution<>,n> &neighbour_distr);
void convert_graph(array<array<int, n>, n> &adjacency_matrix,array<int, n>
 ↪  &degrees,array<array<int, n>, n>
 ↪  &adjacency_list,array<uniform_int_distribution<>,n> &neighbour_distr);
void simulate_N_runs(int N,array<double,k> fitness,array<int,k>
 ↪  &number_initial_types,array<array<int, n>, n>
 ↪  &adjacency_matrix,array<int,k> &simulation_results);
#endif /* simulation_hpp */
```

E Simulating the k-type Moran process

Since the primary focus of this dissertation is of theoretical nature, this section should rather be understood as a demonstration of the k-type process than as a simulation-based research work. For the latter, more simulation runs would be needed.

The code in Appendix D can be used to simulate the k-type Moran process on every connected graph of interest to observe empirical values for the fixation probabilities. We did so on the complete graph with different values for k, and chose for every type $j \in [k]$ the fitness value $1 + (j-1) \cdot 0.05$. We conducted 5000 simulations per value of k and plotted the fixation probability of the three strongest types as well as of the weakest type in Figure 3. While f^1 decays, f^k grows, which would also be our understanding from the 2-type process. We see that the other, almost strongest types also grow slightly. Intuitively, this behaviour can be explained as follows: Type 1 has no fitness advantage, its only advantage stems from occupying $n-k+1$ vertices initially. However, as k grows, this advantage declines. The fixation probabilities of types k, $k-1$, $k-2$ increases, since their fitness values increase with growing k. It is important to note that those results hold only for small values of k compared to n.

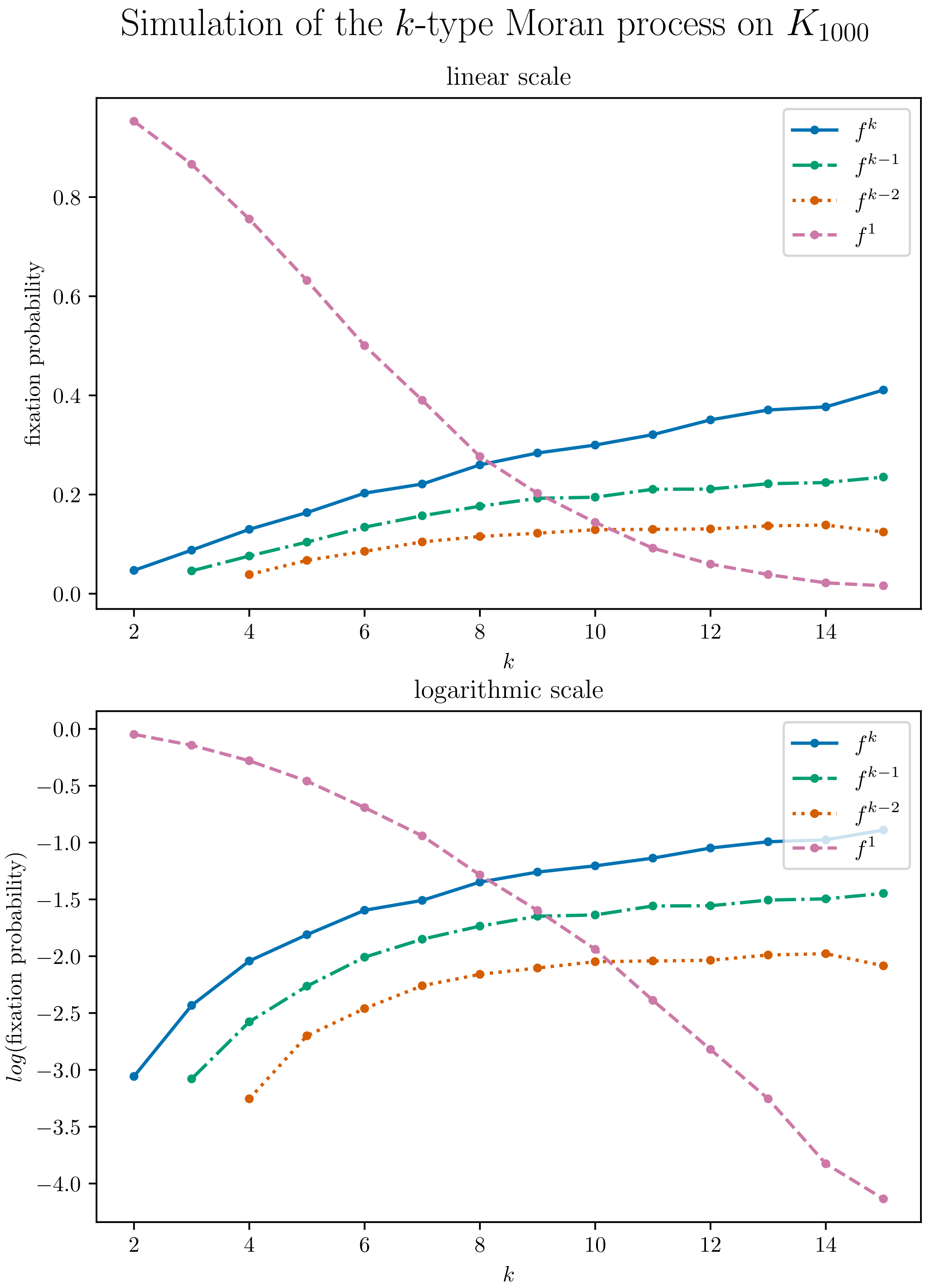


Figure 3: Simulation of the k-type Moran process on the complete graph with $n = 1000$ vertices. For each $k \in \{2, \ldots, 15\}$, the process was simulated 5000 times using the code in Appendix D. The fitness values for each of the types $1, 2, \ldots, k$ were chosen to be $1.0, 1.05, \ldots 1 + (k-1) \cdot 0.5$.