Fixation probabilities in evolutionary dynamics under weak selection

Alex McAvoy[1,2,3,*] Benjamin Allen[4,†]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138
[2]Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104
[3]Center for Mathematical Biology, University of Pennsylvania, Philadelphia, PA 19104
amcavoy@sas.upenn.edu

[4]Department of Mathematics, Emmanuel College, Boston, MA 02115
allenb@emmanuel.edu

Abstract

In evolutionary dynamics, a key measure of a mutant trait's success is the probability that it takes over the population given some initial mutant-appearance distribution. This "fixation probability" is difficult to compute in general, as it depends on the mutation's effect on the organism as well as the population's spatial structure, mating patterns, and other factors. In this study, we consider weak selection, which means that the mutation's effect on the organism is small. We obtain a weak-selection perturbation expansion of a mutant's fixation probability, from an arbitrary initial configuration of mutant and resident types. Our results apply to a broad class of stochastic evolutionary models, in which the size and spatial structure are arbitrary (but fixed). The problem of whether selection favors a given trait is thereby reduced from exponential to polynomial complexity in the population size, when selection is weak. We conclude by applying these methods to obtain new results for evolutionary dynamics on graphs.

1 Introduction

Many studies of stochastic evolutionary dynamics concern the competition of two types (traits) in a finite population. Through a series of births and deaths, the composition of the population changes over time. Absent recurring mutation, one of the two types will eventually become fixed

*ORCID: https://orcid.org/0000-0002-9110-4635
†ORCID: https://orcid.org/0000-0002-9746-7613

and the other will go extinct. Such models may incorporate frequency-dependent selection as well as spatial structure or other forms of population structure. The fundamental question is to identify the selective differences that favor the fixation of one type over the other.

This question is typically answered by computing a trait's fixation probability (Haldane, 1927; Moran, 1958; Kimura, 1962; Patwa and Wahl, 2008; Traulsen and Hauert, 2010; Der et al., 2011; Hindersin and Traulsen, 2014; McCandlish et al., 2015; Hindersin et al., 2016) as a function of the initial configuration of the two types. Direct calculation of a mutant's fixation probability is possible in simple models of well-mixed populations (Moran, 1958; Taylor et al., 2004) or spatially structured populations that are highly symmetric (Lieberman et al., 2005; Broom and Rychtář, 2008; Hindersin and Traulsen, 2014) or small (Hindersin and Traulsen, 2015; Cuesta et al., 2018; Möller et al., 2019; Tkadlec et al., 2019). For large populations, fixation probabilities can sometimes be approximated using diffusion methods (Kimura, 1962; Roze and Rousset, 2003; Ewens, 2004; Chen, 2018).

When selection is weak, perturbative methods can be applied to the computation of fixation probabilities (Haldane, 1927; Nowak et al., 2004; Lessard and Ladret, 2007). The first-order effects of selection on fixation probabilities provide information about whether selection favors one trait over another, and, if so, by how much. This perturbative approach is often paired with methods from coalescent theory (Rousset, 2003; Chen, 2013; Van Cleve, 2015; Chen et al., 2016; Allen et al., 2017, 2020).

Our aim in this work is to generalize the weak-selection method for computing fixation probability to a broad class of evolutionary models. Our main result is a first-order weak-selection expansion of a mutant's fixation probability from any initial condition. This result applies to arbitrary forms of spatial structure and frequency-dependent selection, and the expansion can be computed for any particular model and initial configuration by solving a system of linear equations. Under conditions that apply to most models of interest, the size of this system—and hence the complexity of computing this expansion—exhibits polynomial growth in the population size.

Our approach is based on a modeling framework developed by Allen and Tarnita (2014) and Allen and McAvoy (2019), which is described in Section 2. This framework describes stochastic trait transmission in a population of fixed size, N, and spatial structure. This setup leads to a finite Markov chain model of selection. Special cases of this framework include the Moran (Moran, 1958) and Wright-Fisher models (Fisher, 1930; Wright, 1931; Imhof and Nowak, 2006), as well as evolutionary games in graph-structured populations (Ohtsuki et al., 2006; Szabó and Fáth, 2007; Nowak et al., 2009; Allen et al., 2017). We use this framework to define the *degree* of an evolutionary process, which later plays an important role in determining the computational complexity of calculating fixation probabilities.

In Section 3, we establish a connection between sojourn times and stationary probabilities. Specifically, we compare the original Markov chain, which is absorbing, to an amended Markov chain, in which the initial configuration $\boldsymbol{\xi}$ can be re-established from the all-A and all-B states with some probability, u. This amended Markov chain is recurrent, and it has a unique stationary distribution. We show that sojourn times for transient states of the original Markov chain are equal to u-derivatives, at $u = 0$, of stationary probabilities in the amended chain. We also define a set-valued coalescent process that is used in the proof of our main results.

Section 4 proves our main result regarding fixation probabilities. We consider the fixation probability, $\rho_A(\boldsymbol{\xi})$, of type A in a population whose initial configuration of A and B is $\boldsymbol{\xi}$. The intensity of selection, δ, which quantifies selective differences between the two types, is assumed to be sufficiently weak, meaning $\delta \ll 1$. Our main result is a formula for calculating $\rho_A(\boldsymbol{\xi})$ to first order in the selection intensity δ. This formula depends on the first-order effects of selection on marginal trait-transmission probabilities together with a set of sojourn times for neutral drift. The latter can be evaluated by solving a linear system of size $O\left(N^{D+1}\right)$, where D is the degree of the process. This linear system is what bounds the complexity in N of calculating the first-order coefficient, $\frac{d}{d\delta}\Big|_{\delta=0}\rho_A(\boldsymbol{\xi})$.

In Section 5, we extend our main result to the case that the initial configuration of A and B is stochastic rather than deterministic. We derive a formula for $\frac{d}{d\delta}\Big|_{\delta=0}\mathbb{E}_{\mu_A}[\rho_A]$, where μ_A is an arbitrary distribution over initial configurations of A and B (and which can also depend on δ). Section 6 considers relative measures of evolutionary success (Tarnita and Taylor, 2014) obtained by comparing $\mathbb{E}_{\mu_A}[\rho_A]$ to $\mathbb{E}_{\mu_B}[\rho_B]$ when A and B each have their own initial distributions, μ_A and μ_B.

Finally, we apply our results to several well-known questions in evolutionary dynamics, particularly on graph-structured populations. Section 7 discusses the case of constant fecundity, wherein the reproductive rate of an individual depends on only its own type. A large body of research (Lieberman et al., 2005; Broom and Rychtář, 2008; Broom et al., 2011; Voorhees, 2013; Monk et al., 2014; Hindersin and Traulsen, 2015; Kaveh et al., 2015; Cuesta et al., 2017; Pavlogiannis et al., 2018; Möller et al., 2019; Tkadlec et al., 2019) aims to understand the effects of graph structure on fixation probabilities in this context. Our results provide efficient recipes to calculate fixation probabilities under weak selection. Section 8 turns to evolutionary game theory. For a particular prisoner's dilemma game (the "donation game"; Sigmund, 2010), we derive a formula for the fixation probability of cooperation from any starting configuration on an arbitrary weighted graph, generalizing and unifying a number of earlier results (Ohtsuki et al., 2006; Taylor et al., 2007; Chen, 2013; Allen and Nowak, 2014; Chen et al., 2016; Allen et al., 2017).

2 Modeling evolutionary dynamics

We employ a framework previously developed by Allen and Tarnita (2014) and Allen and McAvoy (2019) to represent an evolving population with arbitrary forms of spatial structure and frequency dependence. In addition to being described below, all of the notation and symbols we use are outlined in Table 1. Although we will use the language of a haploid asexual population, our formalism applies equally well to diploid, haplodiploid, or other populations by considering the alleles to be asexual replicators (the "gene's-eye view"), as described in Allen and McAvoy (2019).

As a source of motivation and a tool to illustrate the general model, we consider the Moran process (Moran, 1958) as a running example. The Moran process models evolution in an unstructured population consisting of two types, a mutant (A) and a resident (B), of relative fecundity (reproductive rate) r and 1, respectively. In a population consisting of i individuals of type A and $N-i$ individuals of type B, the probability that type A is selected to reproduce is $ir/\left(ir+N-i\right)$. With probability $(N-i)\,/\,(ir+N-i)$, type B is selected to reproduce. The offspring, which inherits the type of the parent, then replaces a random individual in the population. Throughout our discussion of the framework below, we return to this simple example repeatedly (with more sophisticated examples following in later sections).

2.1 Modeling assumptions

We consider competition between two alleles, A and B, in a population of finite size, N. The state of the population is given by $\mathbf{x} \in \{0,1\}^N$, where $x_i = 1$ (resp. $x_i = 0$) indicates that individual i has type A (resp. B).

Since we are mainly concerned with the probability of fixation rather than the timescale, we may assume without a loss of generality that the population evolves in discrete time. However, the results reported here can also be applied to continuous-time models in a straightforward manner. In what follows, we assume that the population's state is updated in discrete time steps via replacement events. A replacement event is a pair, (R,α), where $R \subseteq \{1,\ldots,N\}$ is the set of individuals who are replaced in a given time step and $\alpha : R \to \{1,\ldots,N\}$ is the offspring-to-parent map. For a fixed replacement event, (R,α), the state of the population at time $t+1$, $\mathbf{x}^{t+1}$, is obtained from the state of the population at time t, $\mathbf{x}^t$, by letting

$$x_i^{t+1} = \begin{cases} x_{\alpha(i)}^t & i \in R, \\ \\ x_i^t & i \notin R. \end{cases} \tag{1}$$

We can express such a transition more concisely by defining the extended mapping

$$\begin{aligned} \widetilde{\alpha} : \{1,\ldots,N\} &\longrightarrow \{1,\ldots,N\} \\ : i &\longmapsto \begin{cases} \alpha(i) & i \in R, \\ \\ i & i \notin R. \end{cases} \end{aligned} \tag{2}$$

We then have $\mathbf{x}^{t+1} = \mathbf{x}^t_{\widetilde{\alpha}}$, where $\mathbf{x}_{\widetilde{\alpha}}$ is the vector whose ith component is $x_{\widetilde{\alpha}(i)}$.

In state $\mathbf{x}$, we denote by $\left\{ p_{(R,\alpha)}(\mathbf{x}) \right\}_{(R,\alpha)}$ the distribution from which the replacement event is chosen. We call this distribution (as a function of $\mathbf{x}$) the *replacement rule*. We assume that this replacement rule depends on an underlying parameter $\delta \geqslant 0$ that represents the intensity of selection. Neutral drift corresponds to $\delta = 0$, and weak selection is the regime $\delta \ll 1$.

Example: Moran process. In the Moran process, a slightly advantageous mutant has fecundity $r = 1+\delta$ for $\delta \ll 1$. The probability that replacement event (R,α) is chosen is

$$p_{(R,\alpha)}(\mathbf{x}) = \begin{cases} \dfrac{x_{\alpha(i)}r + 1 - x_{\alpha(i)}}{\sum_{j=1}^N (x_j r + 1 - x_j)} \dfrac{1}{N} & R = \{i\}, \\ \\ 0 & |R| \neq 1. \end{cases} \tag{3}$$

That is, assuming $R = \{i\}$, the probability that $\alpha(i)$ reproduces is proportional to $x_{\alpha(i)}r + 1 - x_{\alpha(i)}$ (which is r if $\alpha(i)$ has type A and 1 otherwise). The constant of proportionality is the reciprocal of the total population fecundity, $\sum_{j=1}^{N}(x_j r + 1 - x_j)$. If $\alpha(i)$ reproduces, then the probability that the offspring replaces i is simply $1/N$.

For brevity, we write $\mathbb{B} := \{0, 1\}$. Each replacement rule defines a Markov chain on $\mathbb{B}^N$, according to Eq. (1). We let $P_{\mathbf{x}\to\mathbf{y}}$ be the probability of transitioning from state $\mathbf{x}$ to state $\mathbf{y}$ in this Markov chain.

We make three assumptions on the replacement rule. The first is that for every δ, there exists at least one individual who can generate a lineage that takes over the entire population. We state this assumption as an axiom:

Fixation Axiom. There exists $i \in \{1, \dots, N\}$, $m \geqslant 1$, and a sequence $\{(R_k, \alpha_k)\}_{k=1}^{m}$ with

- $p_{(R_k,\alpha_k)}(\mathbf{x}) > 0$ for every $k \in \{1, \dots, m\}$ and $\mathbf{x} \in \mathbb{B}^N$;
- $i \in R_k$ for some $k \in \{1, \dots, m\}$;
- for every $j \in \{1, \dots, N\}$, we have $\widetilde{\alpha}_1 \circ \widetilde{\alpha}_2 \circ \cdots \circ \widetilde{\alpha}_m(j) = i$.

The requirement that $i \in R_k$ for some $k \in \{1, \dots, m\}$ guarantees that the individual at i cannot live forever, since otherwise no evolution would occur (Allen and McAvoy, 2019). Allen and Tarnita (2014) showed that, under the Fixation Axiom, this Markov chain has two absorbing states: the state $\mathbf{A}$ in which $x_i = 1$ for every i (all-A), and the state $\mathbf{B}$ in which $x_i = 0$ for every i (all-B). All other states are transient. We denote the set of all transient states by $\mathbb{B}_{\mathsf{T}}^N := \mathbb{B}^N - \{\mathbf{A}, \mathbf{B}\}$.

Our second assumption is that when $\delta = 0$ (neutral drift), the replacement rule does not depend on the state, $\mathbf{x}$. In this case, we denote the replacement rule by $\left\{p_{(R,\alpha)}^{\circ}\right\}_{(R,\alpha)}$. Note that we have removed the dependence on $\mathbf{x}$. This assumption arises because, under neutral drift, the competing alleles are interchangeable, and so the probabilities of replacement should not depend on how these alleles are distributed among individuals. More generally, in the quantities derived from the replacement rule below (e.g. birth rates and death probabilities), we use the superscript $\circ$ to denote their values under neutral drift.

Our third assumption is that for every $\mathbf{x} \in \mathbb{B}^N$ and every replacement event (R, α), $p_{(R,\alpha)}(\mathbf{x})$ is a smooth function of δ in a small neighborhood of $\delta = 0$. This assumption enables a perturbation expansion in the selection strength, δ.

Example: Moran process. Smoothness in δ is evident from Eq. (3) whenever r is itself a smooth function of δ.

2.2 Quantifying selection

Having outlined the class of models under consideration, we now define quantities that characterize natural selection in a given population state $\mathbf{x} \in \mathbb{B}^N$. For any i and j in $\{1, \dots, N\}$, let $e_{ij}(\mathbf{x})$ be the marginal probability that i transmits its offspring to j in state $\mathbf{x}$, i.e.

$$e_{ij}(\mathbf{x}) := \sum_{\substack{(R,\alpha) \\ j\in R,\, \alpha(j)=i}} p_{(R,\alpha)}(\mathbf{x}). \tag{4}$$

The birth rate (expected offspring number) of i is $b_i(\mathbf{x}) := \sum_{j=1}^{N} e_{ij}(\mathbf{x})$ and the death probability of i is $d_i(\mathbf{x}) := \sum_{j=1}^{N} e_{ji}(\mathbf{x})$. Using these quantities, we can write the expected change in the frequency of A due to selection as (Tarnita and Taylor, 2014; Allen and McAvoy, 2019)

$$\Delta_{\text{sel}}(\mathbf{x}) := \sum_{i=1}^{N} x_i \left(b_i(\mathbf{x}) - d_i(\mathbf{x})\right). \tag{5}$$

Any real-valued function on $\mathbb{B}^N$ is called a *pseudo-Boolean function* (Hammer and Rudeanu, 1968). Since $e_{ij}(\mathbf{x})$ and its derivative with respect to δ at $\delta = 0$ are pseudo-Boolean functions, for every i and j there is a unique multi-linear polynomial representation (Hammer and Rudeanu, 1968; Boros and Hammer, 2002),

$$\left.\frac{d}{d\delta}\right|_{\delta=0} e_{ij}(\mathbf{x}) = \sum_{I\subseteq\{1,\ldots,N\}} c_I^{ij}\mathbf{x}_I, \tag{6}$$

where the c_I^{ij} are a collection of real numbers (Fourier coefficients) indexed by the subsets $I \subseteq \{1,\ldots,N\}$, and $\mathbf{x}_I := \prod_{i\in I} x_i$. Note that $\mathbf{x}_I$ is a scalar, not a state, with $\mathbf{x}_I = 1$ if and only if $x_i = 1$ for each $i \in I$. This representation includes the constant term $c_\varnothing^{ij}$, linear terms of the form $c_{\{k\}}^{ij} x_k$, quadratic terms of the form $c_{\{h,k\}}^{ij} x_h x_k$, and so on up through $c_{\{1,\ldots,N\}}^{ij} x_1 \cdots x_N$. The coefficients c_I^{ij} quantify how genetic assortment among sets of individuals affects the probability that i's offspring replaces j, under weak selection.

We let D_{ij} denote the degree of the above representation, defined as the degree of the highest-order nonzero term:

$$D_{ij} := \max\left\{k \;:\; c_I^{ij} \neq 0 \text{ for some } I \subseteq \{1,\ldots,N\} \text{ with } |I| = k\right\}. \tag{7}$$

(In the trivial case that $c_I^{ij} = 0$ for every $I \subseteq \{1,\ldots,N\}$, we set $D_{ij} = 0$.)

For $I \subseteq \{1,\ldots,N\}$, let $\mathbf{1}_I \in \mathbb{B}^N$ denote the state in which $x_i = 1$ for $i \in I$ and $x_i = 0$ for $i \notin I$. By applying a Möbius transform to Eq. (6) (Grabisch et al., 2000), we can express the coefficients c_I^{ij} as

$$c_I^{ij} = \left.\frac{d}{d\delta}\right|_{\delta=0} \sum_{J\subseteq I} (-1)^{|I|-|J|} e_{ij}(\mathbf{1}_J). \tag{8}$$

This expression provides a recipe for calculating the coefficients c_I^{ij} directly from $e_{ij}(\mathbf{x})$ for a given process (see Section 8).

We define the degree of the overall evolutionary process, under weak selection, to be the maximal degree in Eq. (6) as i and j run over all pairs of sites:

Definition 1. The *degree* of the process under weak selection is $D := \max_{1\leqslant i,j\leqslant N} D_{ij}$.

Example: Moran process. In this particular process, we have

$$e_{ij}\left(\mathbf{x}\right) = \frac{x_i r + 1 - x_i}{\sum_{k=1}^{N}\left(x_k r + 1 - x_k\right)}\frac{1}{N}. \tag{9}$$

If $r = r\left(\delta\right)$ is a smooth function of δ with $r\left(0\right) = 1$, then

$$\left.\frac{d}{d\delta}\right|_{\delta=0} e_{ij}\left(\mathbf{x}\right) = \frac{1}{N^2} r'\left(0\right)\left(x_i - \frac{1}{N}\sum_{k=1}^{N} x_k\right). \tag{10}$$

The degree of the Moran process is thus $D = 1$, with $c_{\varnothing}^{ij} = 0$ and

$$c_{\{k\}}^{ij} = \begin{cases} \dfrac{1}{N^2}\left(1 - \dfrac{1}{N}\right) r'\left(0\right) & k = i, \\ \\ -\dfrac{1}{N^3} r'\left(0\right) & k \neq i. \end{cases} \tag{11}$$

Beyond this linear example, in a degree-two process, $\frac{d}{d\delta}\big|_{\delta=0} e_{ij}\left(\mathbf{x}\right)$ can be represented as a quadratic function of $x_1, \dots, x_N$, with terms only of the form $c_{\varnothing}^{ij}$, $c_{\{k\}}^{ij} x_k$, or $c_{\{h,k\}}^{ij} x_h x_k$. In this case, replacement probabilities under weak selection depend on only pairwise statistics of assortment and not on higher-order associations.

We turn now to fixation probabilities. For $\boldsymbol{\xi} \in \mathbb{B}^N$, let $\rho_A\left(\boldsymbol{\xi}\right)$ (resp. $\rho_B\left(\boldsymbol{\xi}\right)$) be the probability that the state $\mathbf{A}$ (resp. $\mathbf{B}$) is eventually reached after starting in state $\boldsymbol{\xi}$. Since states $\mathbf{A}$ and $\mathbf{B}$ are absorbing and all other states are transient, we have $\rho_B\left(\boldsymbol{\xi}\right) = 1 - \rho_A\left(\boldsymbol{\xi}\right)$ for each $\boldsymbol{\xi} \in \mathbb{B}^N$.

In the case of neutral drift ($\delta = 0$), we let π_i be the probability of fixation for type A when starting from state $\mathbf{1}_{\{i\}}$; that is, $\pi_i = \rho_A\left(\mathbf{1}_{\{i\}}\right)$ when $\delta = 0$. Equivalently, π_i is the probability, under neutral drift, that i is eventually the ancestor of the entire population. These site-specific fixation probabilities are the unique solution to the system of equations (Allen et al., 2015, Theorem 2; Allen and McAvoy, 2019, Theorem 7)

$$\sum_{j=1}^{N} e_{ij}^{\circ}\pi_j = \sum_{j=1}^{N} e_{ji}^{\circ}\pi_i \quad \left(1 \leqslant i \leqslant N\right); \tag{12a}$$

$$\sum_{i=1}^{N} \pi_i = 1. \tag{12b}$$

The quantity π_i can be interpreted as the *reproductive value (RV)* of site i (Fisher, 1930; Taylor, 1990; Maciejewski, 2014; Allen and McAvoy, 2019), in that it quantifies the expected contribution of site i to the future gene pool, under neutral drift. For any state $\mathbf{x} \in \mathbb{B}^N$, the RV-weighted frequency, $\widehat{x} := \sum_{i=1}^{N} \pi_i x_i$, is equal to the probability that A becomes fixed under neutral drift when the process starts in state $\mathbf{x}$ (Allen and McAvoy, 2019, Theorem 7).

Example: Moran process. The reproductive value of every location is $1/N$ due to the fact that the population is unstructured. In a state with i mutants, the fixation probability of A is thus i/N.

In later examples, we will see that this distribution need not be uniform when the population is spatially structured.

Reproductive values provide a natural weighting for quantities characterizing selection; for example, we define the RV-weighted birth rates and death probabilities to be $\widehat{b}_i(\mathbf{x}) := \sum_{j=1}^{N} e_{ij}(\mathbf{x})\,\pi_j$ and $\widehat{d}_i(\mathbf{x}) := \sum_{j=1}^{N} e_{ji}(\mathbf{x})\,\pi_i$, respectively. In state $\mathbf{x}$, the change in reproductive-value-weighted frequency of A due to selection, in one step of the process, is

$$\widehat{\Delta}_{\text{sel}}(\mathbf{x}) := \sum_{i=1}^{N} x_i \left(\widehat{b}_i(\mathbf{x}) - \widehat{d}_i(\mathbf{x})\right) = \sum_{i=1}^{N} \pi_i \sum_{j=1}^{N} \left(x_j - x_i\right) e_{ji}(\mathbf{x}) \tag{13}$$

(Allen and McAvoy, 2019). Since $\widehat{b}_i^{\circ} = \widehat{d}_i^{\circ}$ for $i = 1, \dots, N$, it follows that $\widehat{\Delta}_{\text{sel}}^{\circ}(\mathbf{x}) = 0$ for every $\mathbf{x} \in \mathbb{B}^N$.

3 Stationary distributions, sojourn times, and coalescence

We are ultimately interested in quantifying (to first order in δ) the probability $\rho_A(\boldsymbol{\xi})$ that type A reaches fixation from initial state $\boldsymbol{\xi}$. To do so, we will need to quantify the frequency with which the Markov chain visits a given state $\mathbf{x} \in \mathbb{B}^N$ prior to absorption in state $\mathbf{A}$ or $\mathbf{B}$. We will describe this frequency in two ways: using sojourn times and using the stationary distribution of an amended Markov chain. These two notions are closely connected, as we prove in Proposition 1 below.

We define the sojourn time $t_{\boldsymbol{\xi}}(\mathbf{x})$, for $\mathbf{x} \in \mathbb{B}^N$, to be the expected number of visits to $\mathbf{x}$ prior to hitting $\{\mathbf{A}, \mathbf{B}\}$ when the process begins in state $\boldsymbol{\xi} \in \mathbb{B}_{\mathsf{T}}^N$. These sojourn times $t_{\boldsymbol{\xi}}(\mathbf{x})$ are uniquely determined by the recurrence relation

$$t_{\boldsymbol{\xi}}(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in \{\mathbf{A}, \mathbf{B}\}, \\ 1 + \displaystyle\sum_{\mathbf{y} \in \mathbb{B}^N} t_{\boldsymbol{\xi}}(\mathbf{y})\, P_{\mathbf{y} \to \mathbf{x}} & \mathbf{x} = \boldsymbol{\xi}, \\ \displaystyle\sum_{\mathbf{y} \in \mathbb{B}^N} t_{\boldsymbol{\xi}}(\mathbf{y})\, P_{\mathbf{y} \to \mathbf{x}} & \mathbf{x} \notin \{\mathbf{A}, \mathbf{B}, \boldsymbol{\xi}\}. \end{cases} \tag{14}$$

Since the transition probabilities are continuously differentiable in δ in a neighborhood of $\delta = 0$, so is $t_{\boldsymbol{\xi}}(\mathbf{x})$.

It is also helpful to consider an amended Markov chain that can "reset" in state $\boldsymbol{\xi} \in \mathbb{B}_{\mathsf{T}}^N$ after one of the monoallelic states, $\mathbf{A}$ or $\mathbf{B}$, is reached. We introduce a new parameter $u > 0$, and define, for $\mathbf{x},\mathbf{y}\in\mathbb{B}^N$, the amended transition probabilities

$$P^{\circlearrowright(\boldsymbol{\xi})}_{\mathbf{x}\to\mathbf{y}} := \begin{cases} u & \mathbf{x}\in\{\mathbf{A},\mathbf{B}\}\,,\ \mathbf{y}=\boldsymbol{\xi}, \\ (1-u)\,P_{\mathbf{x}\to\mathbf{y}} & \mathbf{x}\in\{\mathbf{A},\mathbf{B}\}\,,\ \mathbf{y}\neq\boldsymbol{\xi}, \\ P_{\mathbf{x}\to\mathbf{y}} & \mathbf{x}\notin\{\mathbf{A},\mathbf{B}\}\,. \end{cases} \tag{15}$$

Above, $P_{\mathbf{x}\to\mathbf{y}}$ refers to the transition probability in the original Markov chain. Thus the amended chain has the same transition probabilities except that, from either of the monoallelic states $\mathbf{A}$ or $\mathbf{B}$, there is probability u to transition to state $\boldsymbol{\xi}$ (see Fig. 1). Since $\boldsymbol{\xi}$ can be any polymorphic state, and since $\boldsymbol{\xi}$ is the only polymorphic state that can follow $\mathbf{A}$ or $\mathbf{B}$, any interpretation of u as a "mutation" is likely tenuous from a biological standpoint. However, this amended chain plays an integral technical role in deriving our main results, which in turn do apply under much more realistic assumptions of mutant appearance (discussed in Section 5). In particular, the amended chain is clearly aperiodic, and it follows from the Fixation Axiom that it has a single closed communicating class, and all states not in this class are transient. The amended chain therefore has a unique stationary distribution, which we denote by $\left\{\pi_{\circlearrowright(\boldsymbol{\xi})}(\mathbf{x})\right\}_{\mathbf{x}\in\mathbb{B}^N}$, the notation $\circlearrowright(\boldsymbol{\xi})$ indicating regeneration into state $\boldsymbol{\xi}$.

Consider now the Markov chain on the monoallelic states whose transition matrix is

$$\Lambda := \begin{array}{c} \\ \mathbf{A} \\ \mathbf{B} \end{array}\!\!\begin{array}{c} \begin{array}{cc} \mathbf{A} & \mathbf{B} \end{array} \\ \begin{pmatrix} \rho_A(\boldsymbol{\xi}) & \rho_B(\boldsymbol{\xi}) \\ \rho_A(\boldsymbol{\xi}) & \rho_B(\boldsymbol{\xi}) \end{pmatrix} \end{array}. \tag{16}$$

This chain describes the process in which transitions are first from $\mathbf{A}$ or $\mathbf{B}$ to $\boldsymbol{\xi}$, deterministically, and then from $\boldsymbol{\xi}$ to $\mathbf{A}$ with probability $\rho_A(\boldsymbol{\xi})$ and to $\mathbf{B}$ with probability $\rho_B(\boldsymbol{\xi})$. This chain is "embedded" in the amended chain in the sense that when u is small, the amended chain spends most of its time in $\{\mathbf{A},\mathbf{B}\}$, but occasionally it transitions to $\boldsymbol{\xi}$ before returning to $\{\mathbf{A},\mathbf{B}\}$ according to the fixation probabilities. More formally, on a state-by-state basis, the stationary distribution of the embedded chain coincides with $\pi_{\circlearrowright(\boldsymbol{\xi})}$ on the monoallelic states in the limit $u\to 0$ (Fudenberg and Imhof, 2006, Theorem 2), i.e.

$$\lim_{u\to 0}\pi_{\circlearrowright(\boldsymbol{\xi})}(\mathbf{x}) = \begin{cases} \rho_A(\boldsymbol{\xi}) & \mathbf{x}=\mathbf{A}, \\ \rho_B(\boldsymbol{\xi}) & \mathbf{x}=\mathbf{B}, \\ 0 & \mathbf{x}\notin\{\mathbf{A},\mathbf{B}\}\,. \end{cases} \tag{17}$$

The following result, which is key to our methodology, shows that sojourn times of the original chain coincide with the u-derivative, at $u=0$, of the stationary distribution for the amended chain:

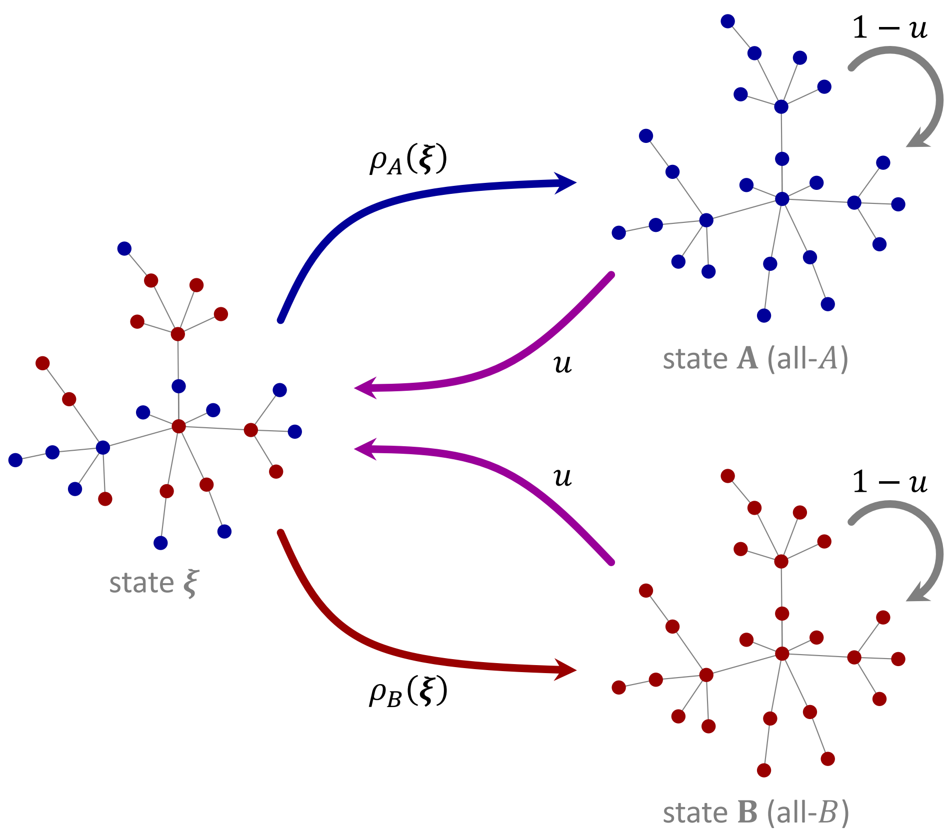

Figure 1: Transitions into a fixed transient state, $\boldsymbol{\xi}$, following absorption. When starting from a non-monomorphic state, the process will eventually reach one of the two absorbing states (all-A or all-B) by the Fixation Axiom. From each absorbing state, the process transitions to $\boldsymbol{\xi}$ with probability $u \geqslant 0$. This "artificial" mutation allows one to focus on the fixation probabilities when the process is started in a fixed initial configuration, $\boldsymbol{\xi}$, of A and B.

Proposition 1. *For every non-monoallelic state* $\mathbf{x} \in \mathbb{B}^N_{\mathsf{T}}$*,*

$$\left.\frac{d}{du}\right|_{u=0} \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{x}\right) = t_{\boldsymbol{\xi}}\left(\mathbf{x}\right). \tag{18}$$

Proof. For $\mathbf{x} \in \mathbb{B}^N_{\mathsf{T}}$, stationarity implies

$$\begin{aligned} \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{x}\right) &= \sum_{\mathbf{y}\in\mathbb{B}^N} \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{y}\right) P^{\circlearrowleft(\boldsymbol{\xi})}_{\mathbf{y}\to\mathbf{x}} \\ &= \sum_{\mathbf{y}\in\mathbb{B}^N_{\mathsf{T}}} \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{y}\right) P_{\mathbf{y}\to\mathbf{x}} + u\delta_{\boldsymbol{\xi},\mathbf{x}}\left(\pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{A}\right) + \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{B}\right)\right), \end{aligned} \tag{19}$$

where the Kronecker symbol $\delta_{\mathbf{x},\boldsymbol{\xi}}$ is equal to 1 if $\mathbf{x} = \boldsymbol{\xi}$ and 0 otherwise. Taking the u-derivative at $u = 0$, and noting that $\lim_{u\to 0}\left(\pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{A}\right) + \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{B}\right)\right) = 1$ by Eq. (17), we obtain

$$\left.\frac{d}{du}\right|_{u=0} \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{x}\right) = \sum_{\mathbf{y}\in\mathbb{B}^N_{\mathsf{T}}} \left.\frac{d}{du}\right|_{u=0} \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{y}\right) P_{\mathbf{y}\to\mathbf{x}} + \delta_{\boldsymbol{\xi},\mathbf{x}}. \tag{20}$$

We observe that $\left.\frac{d}{du}\right|_{u=0} \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbf{x}\right)$ satisfies the same recurrence relation, Eq. (14), as $t_{\boldsymbol{\xi}}\left(\mathbf{x}\right)$. Since this recurrence relation uniquely defines the times $t_{\boldsymbol{\xi}}\left(\mathbf{x}\right)$, we have Eq. (18). □

It follows immediately from Proposition 1 that $t_{\boldsymbol{\xi}}\left(\mathbb{B}^N_{\mathsf{T}}\right)$, the expected time to absorption when starting from state $\boldsymbol{\xi}$, is equal to $\lim_{u\to 0} \pi_{\circlearrowleft(\boldsymbol{\xi})}\left(\mathbb{B}^N_{\mathsf{T}}\right)/u$. With $\mathbb{E}_{\circlearrowleft(\boldsymbol{\xi})}\left[\cdot\right]$ denoting expectation with respect to $\pi_{\circlearrowleft(\boldsymbol{\xi})}$, we have the following result:

Corollary 1. *For any function* $\varphi : \mathbb{B}^N \to \mathbb{R}$ *with* $\varphi\left(\mathbf{A}\right) = \varphi\left(\mathbf{B}\right) = 0$*,*

$$\left.\frac{d}{du}\right|_{u=0} \mathbb{E}_{\circlearrowleft(\boldsymbol{\xi})}\left[\varphi\right] = \sum_{t=0}^{\infty} \mathbb{E}\left[\varphi\left(\mathbf{x}^t\right) \mid \mathbf{x}^0 = \boldsymbol{\xi}\right], \tag{21}$$

and the sum on the right-hand side converges absolutely.

Proof. $\sum_{t=0}^{\infty}\left|\mathbb{E}\left[\varphi\left(\mathbf{x}^t\right) \mid \mathbf{x}^0 = \boldsymbol{\xi}\right]\right|$ is bounded by $t_{\boldsymbol{\xi}}\left(\mathbb{B}^N_{\mathsf{T}}\right) \max_{\mathbf{x}\in\mathbb{B}^N_{\mathsf{T}}} \left|\varphi\left(\mathbf{x}\right)\right|$, so the right-hand side of Eq. (21) converges absolutely. We may therefore rearrange this summation to obtain

$$\begin{aligned} \sum_{t=0}^{\infty} \mathbb{E}\left[\varphi\left(\mathbf{x}^t\right) \mid \mathbf{x}^0 = \boldsymbol{\xi}\right] &= \sum_{t=0}^{\infty} \sum_{\mathbf{x}\in\mathbb{B}^N_{\mathsf{T}}} \mathbb{P}\left[\mathbf{x}^t = \mathbf{x} \mid \mathbf{x}^0 = \boldsymbol{\xi}\right] \varphi\left(\mathbf{x}\right) \\ &= \sum_{\mathbf{x}\in\mathbb{B}^N_{\mathsf{T}}} \varphi\left(\mathbf{x}\right) \sum_{t=0}^{\infty} \mathbb{P}\left[\mathbf{x}^t = \mathbf{x} \mid \mathbf{x}^0 = \boldsymbol{\xi}\right] \end{aligned}$$

$$= \sum_{\mathbf{x} \in \mathbb{B}^N_{\mathsf{T}}} \varphi(\mathbf{x})\, t_{\boldsymbol{\xi}}(\mathbf{x})$$

$$= \left.\frac{d}{du}\right|_{u=0} \mathbb{E}_{\circlearrowright(\boldsymbol{\xi})}[\varphi], \tag{22}$$

as desired. □

In light of Corollary 1, we define the operator $\langle \cdot \rangle_{\boldsymbol{\xi}}$, acting on state functions $\varphi : \mathbb{B}^N \to \mathbb{R}$ with $\varphi(\mathbf{A}) = \varphi(\mathbf{B}) = 0$, by

$$\langle \varphi \rangle_{\boldsymbol{\xi}} := \left.\frac{d}{du}\right|_{u=0} \mathbb{E}_{\circlearrowright(\boldsymbol{\xi})}[\varphi] = \sum_{t=0}^{\infty} \mathbb{E}\left[\varphi(\mathbf{x}^t) \mid \mathbf{x}^0 = \boldsymbol{\xi}\right]. \tag{23}$$

We will use the notation $\langle \cdot \rangle^{\circ}_{\boldsymbol{\xi}}$ to indicate that the above expectations are taken under neutral drift ($\delta = 0$).

For any function $\varphi : \mathbb{B}^N \to \mathbb{R}$ and any $\widetilde{\alpha} : \{1, \ldots, N\} \to \{1, \ldots, N\}$, we define $\varphi_{\widetilde{\alpha}} : \mathbb{B}^N \to \mathbb{R}$ by $\varphi_{\widetilde{\alpha}}(\mathbf{x}) = \varphi(\mathbf{x}_{\widetilde{\alpha}})$. If $\varphi(\mathbf{A}) = \varphi(\mathbf{B}) = 0$, then

$$\begin{aligned} \langle \varphi \rangle^{\circ}_{\boldsymbol{\xi}} &= \varphi(\boldsymbol{\xi}) + \sum_{t=0}^{\infty} \mathbb{E}^{\circ}\left[\varphi(\mathbf{x}^{t+1}) \mid \mathbf{x}^0 = \boldsymbol{\xi}\right] \\ &= \varphi(\boldsymbol{\xi}) + \sum_{t=0}^{\infty} \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)}\, \mathbb{E}^{\circ}\left[\varphi\left(\mathbf{x}^t_{\widetilde{\alpha}}\right) \mid \mathbf{x}^0 = \boldsymbol{\xi}\right] \\ &= \varphi(\boldsymbol{\xi}) + \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \langle \varphi_{\widetilde{\alpha}} \rangle^{\circ}_{\boldsymbol{\xi}}, \end{aligned} \tag{24}$$

which gives the following lemma:

Lemma 1. *For any state function $\varphi : \mathbb{B}^N \to \mathbb{R}$ with $\varphi(\mathbf{A}) = \varphi(\mathbf{B}) = 0$,*

$$\langle \varphi \rangle^{\circ}_{\boldsymbol{\xi}} = \varphi(\boldsymbol{\xi}) + \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \langle \varphi_{\widetilde{\alpha}} \rangle^{\circ}_{\boldsymbol{\xi}}. \tag{25}$$

Allen and McAvoy (2019) introduced the *rare-mutation conditional (RMC)* distribution, defined for a state $\mathbf{x}$ as $\lim_{u \to 0} \mathbb{P}_{\circlearrowright(\boldsymbol{\xi})}\left[\mathbf{X} = \mathbf{x} \mid \mathbf{x} \in \mathbb{B}^N_{\mathsf{T}}\right]$, where $\mathbb{P}_{\circlearrowright(\boldsymbol{\xi})}[\cdot]$ denotes probability with respect to $\pi_{\circlearrowright(\boldsymbol{\xi})}$. Here, we show that this distribution can be equated to the fraction of time spent in state $\mathbf{x}$ out of all transient states:

Corollary 2. *For each $\mathbf{x} \in \mathbb{B}^N_{\mathsf{T}}$*

$$\lim_{u \to 0} \mathbb{P}_{\circlearrowright(\boldsymbol{\xi})}\left[\mathbf{X} = \mathbf{x} \mid \mathbf{x} \in \mathbb{B}^N_{\mathsf{T}}\right] = \frac{t_{\boldsymbol{\xi}}(\mathbf{x})}{t_{\boldsymbol{\xi}}\left(\mathbb{B}^N_{\mathsf{T}}\right)}. \tag{26}$$

Proof. For $\mathbf{x} \in \mathbb{B}_{\mathsf{T}}^N$,

$$\begin{aligned}
\lim_{u\to 0} \mathbb{P}_{\circlearrowright(\boldsymbol{\xi})}\left[\mathbf{X}=\mathbf{x} \mid \mathbf{x}\in\mathbb{B}_{\mathsf{T}}^N\right] &= \lim_{u\to 0}\frac{\pi_{\circlearrowright(\boldsymbol{\xi})}\left(\mathbf{x}\right)}{\pi_{\circlearrowright(\boldsymbol{\xi})}\left(\mathbb{B}_{\mathsf{T}}^N\right)} \\
&= \frac{\lim_{u\to 0}\pi_{\circlearrowright(\boldsymbol{\xi})}\left(\mathbf{x}\right)/u}{\lim_{u\to 0}\pi_{\circlearrowright(\boldsymbol{\xi})}\left(\mathbb{B}_{\mathsf{T}}^N\right)/u} \\
&= \frac{t_{\boldsymbol{\xi}}\left(\mathbf{x}\right)}{t_{\boldsymbol{\xi}}\left(\mathbb{B}_{\mathsf{T}}^N\right)},
\end{aligned} \tag{27}$$

by Proposition 1. □

Finally, we introduce a set-valued Markov chain that will be used in the proof of our main result. The states of this Markov chain are subsets of $\{1,\ldots,N\}$. From a given state $I\subseteq\{1,\ldots,N\}$, a new state I' is determined by choosing a replacement event (R,α) according to the neutral probabilities $p^{\circ}_{(R,\alpha)}$, and setting $I'=\widetilde{\alpha}\,(I)$. This Markov chain, which we denote $\mathcal{C}$, can be understood as a coalescent process (Kingman, 1982; Liggett, 1985; Cox, 1989; Wakeley, 2016). At each time-step, $\mathcal{C}$ transitions from a set of individuals $I\subseteq\{1,\ldots,N\}$ to the set $\widetilde{\alpha}\,(I)$ of parents of these individuals. (In the case that an individual $i\in I$ is not replaced, the "parent" is i itself; that is, $\widetilde{\alpha}\,(i)=i$.) Thus, with $\mathcal{C}_0=\{1,\ldots,N\}$, the state of the process after t steps can be understood as the set of ancestors of the current population, at time t before the present.

By the Fixation Axiom, $\mathcal{C}$ has a single closed communicating class consisting only of singleton subsets. (In biological terms, the population's ancestry eventually converges on a common ancestor. The event that $\mathcal{C}$ first reaches a singleton set is called *coalescence*, and the vertex in this singleton set represents the location of the population's most recent common ancestor.) It follows that $\mathcal{C}$ has a unique stationary distribution concentrated on the singleton subsets. In this stationary distribution, the probability of the singleton set $\{i\}$ in this stationary distribution is given by the reproductive value π_i.

4 Fixation probabilities

We now prove our main results regarding fixation probabilities. First, we show that the weak-selection expansion of a trait's fixation probability has a particular form:

Theorem 1. *For any fixed initial configuration* $\boldsymbol{\xi}\in\mathbb{B}_{\mathsf{T}}^N$,

$$\rho_A\left(\boldsymbol{\xi}\right)=\widehat{\xi}+\delta\left\langle \left.\frac{d}{d\delta}\right|_{\delta=0}\widehat{\Delta}_{\mathrm{sel}}\right\rangle^{\circ}_{\boldsymbol{\xi}}+O\left(\delta^2\right). \tag{28}$$

Theorem 1 generalizes earlier results of Rousset (2003), Lessard and Ladret (2007), Chen (2013), and Van Cleve (2015).

Proof. In the chain defined by Eq. (15), the expected change in the reproductive-value-weighted frequency of A in state $\mathbf{x}$, in one step of the process, is given by

$$\widehat{\Delta}_{\circlearrowright(\boldsymbol{\xi})}(\mathbf{x}) = \begin{cases} -u\left(1-\widehat{\xi}\right) & \mathbf{x}=\mathbf{A}, \\ u\widehat{\xi} & \mathbf{x}=\mathbf{B}, \\ \widehat{\Delta}_{\text{sel}}(\mathbf{x}) & \mathbf{x}\notin\{\mathbf{A},\mathbf{B}\}. \end{cases} \tag{29}$$

Averaging this expected change out over the distribution $\pi_{\circlearrowright(\boldsymbol{\xi})}$ gives

$$\begin{aligned} 0 &= \mathbb{E}_{\circlearrowright(\boldsymbol{\xi})}\left[\widehat{\Delta}_{\circlearrowright(\boldsymbol{\xi})}\right] \\ &= \mathbb{E}_{\circlearrowright(\boldsymbol{\xi})}\left[\widehat{\Delta}_{\text{sel}}\right] - u\pi_{\circlearrowright(\boldsymbol{\xi})}(\mathbf{A})\left(1-\widehat{\xi}\right) + u\pi_{\circlearrowright(\boldsymbol{\xi})}(\mathbf{B})\,\widehat{\xi}. \end{aligned} \tag{30}$$

Differentiating both sides of this equation with respect to u at $u = 0$, applying Eq. (17), and rearranging, we obtain

$$\rho_A(\boldsymbol{\xi}) = \widehat{\xi} + \left\langle \widehat{\Delta}_{\text{sel}} \right\rangle_{\boldsymbol{\xi}}. \tag{31}$$

Since $\widehat{\Delta}^{\circ}_{\text{sel}}(\mathbf{x}) = 0$ for every $\mathbf{x} \in \mathbb{B}^N$, and since the replacement rule is a smooth function of δ around 0, we have

$$\begin{aligned} \left\langle \widehat{\Delta}_{\text{sel}} \right\rangle_{\boldsymbol{\xi}} &= \delta \frac{d}{d\delta}\bigg|_{\delta=0} \left\langle \widehat{\Delta}_{\text{sel}} \right\rangle_{\boldsymbol{\xi}} + O\left(\delta^2\right) \\ &= \delta \left\langle \frac{d}{d\delta}\bigg|_{\delta=0} \widehat{\Delta}_{\text{sel}} \right\rangle^{\circ}_{\boldsymbol{\xi}} + O\left(\delta^2\right). \end{aligned} \tag{32}$$

The interchange of the $\langle\cdot\rangle^{\circ}_{\boldsymbol{\xi}}$ operator with the δ-derivative is justified by Corollary 1. Combining this equation with Eq. (31) completes the proof. $\square$

Alternatively, Theorem 1 can be established by writing

$$\begin{aligned} \rho_A(\boldsymbol{\xi}) &= \lim_{t\to\infty} \mathbb{E}\left[\widehat{x}^t \mid \mathbf{x}^0 = \boldsymbol{\xi}\right] \\ &= \widehat{\xi} + \sum_{t=0}^{\infty} \mathbb{E}\left[\widehat{x}^{t+1} - \widehat{x}^t \mid \mathbf{x}^0 = \boldsymbol{\xi}\right] \\ &= \widehat{\xi} + \sum_{t=0}^{\infty} \mathbb{E}\left[\widehat{\Delta}_{\text{sel}}\left(\mathbf{x}^t\right) \mid \mathbf{x}^0 = \boldsymbol{\xi}\right] \\ &= \widehat{\xi} + \left\langle \widehat{\Delta}_{\text{sel}} \right\rangle_{\boldsymbol{\xi}}. \end{aligned} \tag{33}$$

This calculation recovers Eq. (31), and the rest of the proof follows as above.

Our second main result provides a systematic way to compute the first-order term of the weak-selection expansion, Eq. (28):

Theorem 2. *For any fixed initial configuration $\boldsymbol{\xi} \in \mathbb{B}_{\mathsf{T}}^{N}$,*

$$\rho_A\left(\boldsymbol{\xi}\right) = \widehat{\xi} + \delta\left(\sum_{i=1}^{N}\pi_i\sum_{j=1}^{N}\sum_{\substack{I\subseteq\{1,\dots,N\}\\0\leqslant|I|\leqslant D_{ji}}} c_I^{ji}\left(\eta_{\{i\}\cup I}^{\boldsymbol{\xi}} - \eta_{\{j\}\cup I}^{\boldsymbol{\xi}}\right)\right) + O\left(\delta^2\right), \tag{34}$$

where $\eta^{\boldsymbol{\xi}}$ is the unique solution to the equations

$$\eta_I^{\boldsymbol{\xi}} = \widehat{\xi} - \boldsymbol{\xi}_I + \sum_{(R,\alpha)} p_{(R,\alpha)}^{\circ}\eta_{\widetilde{\alpha}(I)}^{\boldsymbol{\xi}} \qquad \left(1 \leqslant |I| \leqslant D+1\right); \tag{35a}$$

$$\sum_{i=1}^{N}\pi_i\eta_{\{i\}}^{\boldsymbol{\xi}} = 0. \tag{35b}$$

The term D appearing in this system is the degree of the process under weak selection; see Definition 1. To simplify notation, in what follows we occasionally drop the bracket notation when using η (e.g. $\eta_i^{\boldsymbol{\xi}}$ instead of $\eta_{\{i\}}^{\boldsymbol{\xi}}$).

Proof. Let us define

$$\eta_I^{\boldsymbol{\xi}} := \left\langle \widehat{x} - \mathbf{x}_I \right\rangle_{\boldsymbol{\xi}}^{\circ}. \tag{36}$$

From Theorem 1 and Eqs. (6) and (13), we have

$$\begin{aligned} \left.\frac{d}{d\delta}\right|_{\delta=0}\rho_A\left(\boldsymbol{\xi}\right) &= \left\langle \left.\frac{d}{d\delta}\right|_{\delta=0}\widehat{\Delta}_{\text{sel}}\right\rangle_{\boldsymbol{\xi}}^{\circ} \\ &= \sum_{i=1}^{N}\pi_i\sum_{j=1}^{N}\sum_{\substack{I\subseteq\{1,\dots,N\}\\0\leqslant|I|\leqslant D_{ji}}} c_I^{ji}\left\langle\left(x_j - x_i\right)\mathbf{x}_I\right\rangle_{\boldsymbol{\xi}}^{\circ} \\ &= \sum_{i=1}^{N}\pi_i\sum_{j=1}^{N}\sum_{\substack{I\subseteq\{1,\dots,N\}\\0\leqslant|I|\leqslant D_{ji}}} c_I^{ji}\left(\eta_{\{i\}\cup I}^{\boldsymbol{\xi}} - \eta_{\{j\}\cup I}^{\boldsymbol{\xi}}\right), \end{aligned} \tag{37}$$

which proves Eq. (34). Eq. (35b) follows immediately from the definition of η_I in Eq. (36). To obtain Eq. (35a) we apply Lemma 1:

$$\eta_I^{\boldsymbol{\xi}} = \left\langle \widehat{x} - \mathbf{x}_I \right\rangle_{\boldsymbol{\xi}}^{\circ}$$

$$
\begin{aligned}
&= \widehat{\xi} - \boldsymbol{\xi}_I + \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \left\langle \widehat{x_{\widetilde{\alpha}}} - \mathbf{x}_{\widetilde{\alpha}(I)} \right\rangle^{\circ}_{\boldsymbol{\xi}} \\
&= \widehat{\xi} - \boldsymbol{\xi}_I + \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \left\langle \widehat{x} - \mathbf{x}_{\widetilde{\alpha}(I)} \right\rangle^{\circ}_{\boldsymbol{\xi}} \\
&= \widehat{\xi} - \boldsymbol{\xi}_I + \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \eta^{\boldsymbol{\xi}}_{\widetilde{\alpha}(I)},
\end{aligned}
\tag{38}
$$

where the penultimate line follows from the fact that $\widehat{x}$ is a martingale under neutral drift (Allen and McAvoy, 2019, Theorem 7).

To prove uniqueness of the solution to Eq. (35), we consider the coalescent Markov chain $\mathcal{C}$ defined in the previous section. Let $\mathbf{C}$ be the transition matrix for $\mathcal{C}$, and let $\mathbf{p}$ be its stationary distribution in vector form, which satisfies $\mathbf{p}(I) = \pi_i$ if $I = \{i\}$ and $\mathbf{p}(I) = 0$ if $|I| \neq 1$. By uniqueness of the stationary distribution, $\mathbf{C}$ has a simple unit eigenvalue, with corresponding one-dimensional left and right eigenspaces spanned by $\mathbf{p}^T$ and $\mathbf{1}$, respectively. We observe that Eq. (35a) can be written in the form $(\mathbf{I} - \mathbf{C})\,\mathbf{y} = \mathbf{b}$, where $\mathbf{y}$ has entries $\eta^{\boldsymbol{\xi}}_I$ and $\mathbf{b}$ has entries $\widehat{\xi} - \boldsymbol{\xi}_I$. We have already exhibited a solution for $\mathbf{y}$ in Eq. (36); call this solution $\mathbf{y}_0$. By the above remarks about $\mathbf{C}$, the most general solution is $\mathbf{y} = \mathbf{y}_0 + K\mathbf{1}$ for an arbitrary scalar K. Now we impose Eq. (35b), which can be written $\mathbf{p}^T\mathbf{y} = 0$. Since $\mathbf{p}^T\mathbf{y}_0 = 0$ and $\mathbf{p}^T\mathbf{1} = 1$, we must have $K = 0$, which leaves $\mathbf{y} = \mathbf{y}_0$ as the unique solution. □

Remark 1. For a process of degree D, calculating $\left.\frac{d}{d\delta}\right|_{\delta=0}\rho_A$ involves solving a linear system of size $O\left(N^{D+1}\right)$ (Eq. (35)). The complexity of solving for π_i and c^{ij}_I is negligible in comparison. Since the complexity of solving a linear system of n equations is $O\left(n^3\right)$, it follows that calculating $\left.\frac{d}{d\delta}\right|_{\delta=0}\rho_A$ is $O\left(N^{3(D+1)}\right)$.

Since the quantities η_I arise as the solution to a system of equations related to the coalescent Markov chain $\mathcal{C}$, it is natural to ask whether the η_I have a coalescent-theoretic interpretation. We show in the following sections that, in some cases when the initial state is chosen from a particular probability distribution, the η_I have a natural interpretation as coalescence times or as branch lengths of a coalescent tree. However, these interpretations do not appear to extend to the case of fixation from an arbitrary initial state $\boldsymbol{\xi}$.

5 Stochastic mutant appearance

Having obtained (in Theorems 1 and 2) a weak-selection expansion for fixation probabilities from a particular state $\boldsymbol{\xi}$, we now generalize to the case that the initial state is sampled from a probability distribution. Specifically, we introduce the probability measures μ_A and μ_B, on $\mathbb{B}^N_{\mathsf{T}}$, to describe the state of the process after type A or B, respectively, has been introduced into the population. We refer to μ_A and μ_B as *mutant-appearance distributions*, although the initial state could just as well arise by some mechanism other than mutation (migration, experimental manipulation, etc.). These distributions can depend on the intensity of selection, δ, and we assume that they are differentiable at $\delta = 0$.

Two mutant-appearance distributions often considered in evolutionary models are *uniform initialization* (a single mutant appears at a uniformly chosen site; Lieberman et al., 2005; Adlam et al., 2015) and *temperature initialization* (a single mutant appears at a site chosen proportionally to the death rate d_i; Allen et al., 2015; Adlam et al., 2015). To define these formally, we define the *complement* $\overline{\mathbf{x}}$ of a state $\mathbf{x} \in \mathbb{B}^N$ by $\overline{x}_i := 1 - x_i$ for $i = 1, \ldots, N$.

Example 1 (Uniform initialization)**.**

$$\mu_A\left(\mathbf{1}_i\right) = \mu_B\left(\overline{\mathbf{1}}_i\right) = \frac{1}{N} \quad \left(1 \leqslant i \leqslant N\right). \tag{39}$$

Example 2 (Temperature initialization)**.**

$$\mu_A\left(\mathbf{1}_i\right) = \frac{d_i\left(\mathbf{B}\right)}{\sum_{j=1}^{N} d_j\left(\mathbf{B}\right)} \quad \left(1 \leqslant i \leqslant N\right); \tag{40a}$$

$$\mu_B\left(\overline{\mathbf{1}}_i\right) = \frac{d_i\left(\mathbf{A}\right)}{\sum_{j=1}^{N} d_j\left(\mathbf{A}\right)} \quad \left(1 \leqslant i \leqslant N\right). \tag{40b}$$

Unlike uniform initialization, temperature initialization opens up the possibility that the mutant-appearance distributions depend on the intensity of selection, δ.

We call a mutant-appearance distribution *symmetric* if it does not distinguish between the two types:

Definition 2. We say that μ_A and μ_B are *symmetric* if $\mu_A\left(\overline{\boldsymbol{\xi}}\right) = \mu_B\left(\boldsymbol{\xi}\right)$ for every $\boldsymbol{\xi} \in \mathbb{B}^N_{\mathsf{T}}$.

Uniform initialization (Example 1) gives symmetric μ_A and μ_B by definition. If mutant initialization is temperature-based (Example 2), then μ_A and μ_B are symmetric when $d_i\left(\mathbf{A}\right) = d_i\left(\mathbf{B}\right)$ for $i = 1, \ldots, N$. This condition is obviously satisfied under neutral drift ($\delta = 0$) but could be violated when $\delta > 0$.

The expected fixation probabilities for A and B, initialized according to μ_A and μ_B, respectively, are

$$\mathbb{E}_{\mu_A}\left[\rho_A\right] = \sum_{\boldsymbol{\xi} \in \mathbb{B}^N_{\mathsf{T}}} \mu_A\left(\boldsymbol{\xi}\right) \rho_A\left(\boldsymbol{\xi}\right); \tag{41a}$$

$$\mathbb{E}_{\mu_B}\left[\rho_B\right] = \sum_{\boldsymbol{\xi}' \in \mathbb{B}^N_{\mathsf{T}}} \mu_B\left(\boldsymbol{\xi}'\right) \rho_B\left(\boldsymbol{\xi}'\right). \tag{41b}$$

Since $\rho_A^{\circ}\left(\boldsymbol{\xi}\right) = \widehat{\xi}$ and $\rho_B^{\circ}\left(\boldsymbol{\xi}\right) = 1 - \widehat{\xi}$ when $\delta = 0$, we have $\mathbb{E}_{\mu_A}^{\circ}\left[\rho_A^{\circ}\right] = \mathbb{E}_{\mu_A}^{\circ}\left[\widehat{\xi}\right]$ and $\mathbb{E}_{\mu_B}^{\circ}\left[\rho_B^{\circ}\right] = 1 - \mathbb{E}_{\mu_B}^{\circ}\left[\widehat{\xi}\right]$. More generally, we have $\mathbb{E}_{\mu_B}\left[\rho_B\right] = 1 - \mathbb{E}_{\mu_B}\left[\rho_A\right]$.

As an immediate consequence of Theorem 2, we obtain the following result for the fixation probability of a given type from a given mutant appearance distribution:

Corollary 3. *If μ_A is a mutant-appearance distribution for A, then*

$$\left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_A}\left[\rho_A\right] = \left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_A}\left[\widehat{\xi}\right] + \sum_{i=1}^{N} \pi_i \sum_{j=1}^{N} \sum_{\substack{I \subseteq \{1,\ldots,N\} \\ 0 \leqslant |I| \leqslant D_{ji}}} c_I^{ji} \left(\eta^{\mu_A}_{\{i\}\cup I} - \eta^{\mu_A}_{\{j\}\cup I}\right), \tag{42}$$

where η^{μ_A} is the unique solution to the equations

$$\eta_I^{\mu_A} = \mathbb{E}^{\circ}_{\mu_A}\left[\widehat{\xi} - \boldsymbol{\xi}_I\right] + \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \eta^{\mu_A}_{\widetilde{\alpha}(I)} \quad (1 \leqslant |I| \leqslant D+1); \tag{43a}$$

$$\sum_{i=1}^{N} \pi_i \eta_i^{\mu_A} = 0. \tag{43b}$$

Similarly, since $\left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_B}\left[\rho_B\right] = -\left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_B}\left[\rho_A\right]$, we have

$$\left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_B}\left[\rho_B\right] = -\left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_B}\left[\widehat{\xi}\right] - \sum_{i=1}^{N} \pi_i \sum_{j=1}^{N} \sum_{\substack{I \subseteq \{1,\ldots,N\} \\ 0 \leqslant |I| \leqslant D_{ji}}} c_I^{ji} \left(\eta^{\mu_B}_{\{i\}\cup I} - \eta^{\mu_B}_{\{j\}\cup I}\right), \tag{44}$$

where η^{μ_B} is the unique solution to the equations

$$\eta_I^{\mu_B} = \mathbb{E}^{\circ}_{\mu_B}\left[\widehat{\xi} - \boldsymbol{\xi}_I\right] + \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \eta^{\mu_B}_{\widetilde{\alpha}(I)} \quad (1 \leqslant |I| \leqslant D+1); \tag{45a}$$

$$\sum_{i=1}^{N} \pi_i \eta_i^{\mu_B} = 0. \tag{45b}$$

In the case of uniform initialization (Example 1), we have $\mathbb{E}_{\mu_A}\left[\xi_i\right] = 1/N$ for all $i \in \{1,\ldots,N\}$. In particular, $\mathbb{E}^{\circ}_{\mu_A}\left[\widehat{\xi} - \xi_i\right] = 0$ for every i. Since $\eta_i^{\mu_A} = 0$ for $i = 1,\ldots,N$ is then a solution to Eq. (43a), and since this solution satisfies Eq. (43b), it must be the unique solution to Eq. (43) in the case that $|I| = 1$ (see the proof of Theorem 2). (This argument is used repeatedly in later examples.) Furthermore, $\mathbb{E}_{\mu_A}\left[\boldsymbol{\xi}_I\right] = 0$ for all $|I| \geqslant 2$. Eq. (43) then simplifies to

$$\eta_I^{\mu_A} = \begin{cases} \dfrac{1}{N} + \displaystyle\sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \eta^{\mu_A}_{\widetilde{\alpha}(I)} & 2 \leqslant |I| \leqslant D+1, \\ 0 & |I| = 1. \end{cases} \tag{46}$$

In this case, $\eta_I^{\mu_A}$ is equal to $1/N$ times the expected number of steps for the coalescent Markov chain $\mathcal{C}$ to reach a singleton set from initial set I. In biological terms, $N\eta_I^{\mu_A}$ is the expected time for the lineages of the individuals in set I to coalesce at a most recent common ancestor.

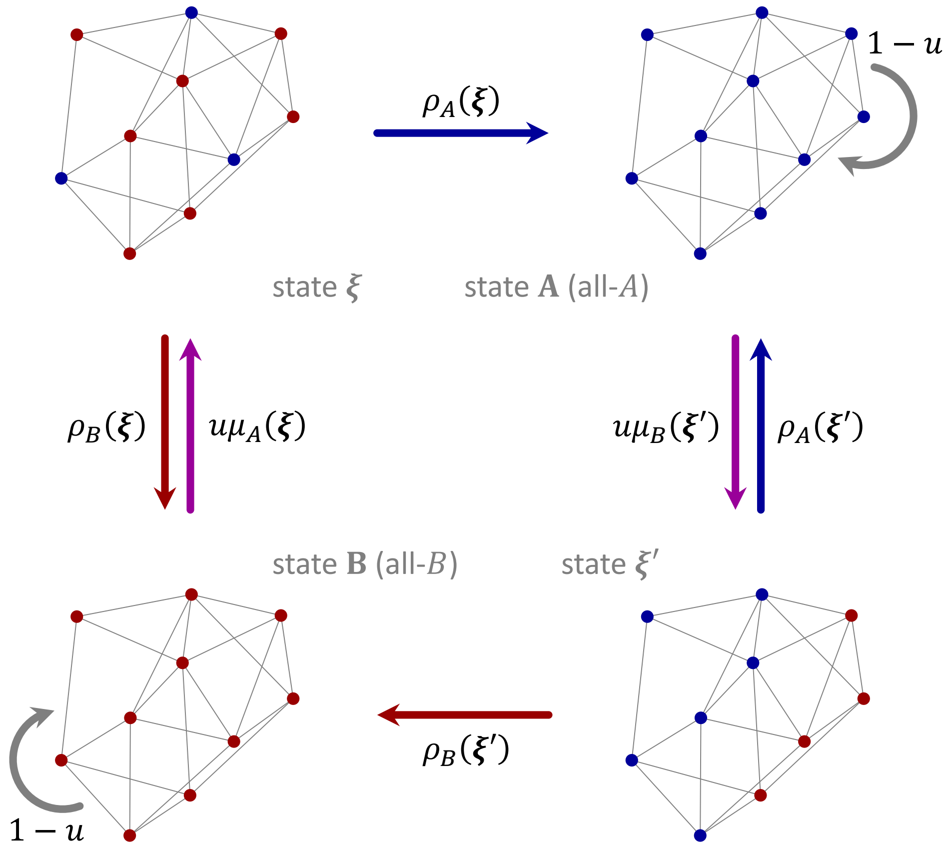

Figure 2: Mutant appearance and fixation in a structured population. In the all-B state, mutants of type A appear based on the mutant-appearance distribution μ_A. In the all-A state, mutants of type B appear based on μ_B. Once the process transitions into a non-monomorphic state, it will eventually reach one of the two monomorphic states.

The overall weak-selection expansion of a trait's fixation probability, in the case of uniform initialization, becomes

$$\mathbb{E}_{\mu_A}\left[\rho_A\right] = \frac{1}{N} + \delta \sum_{i=1}^{N} \pi_i \sum_{j=1}^{N} \sum_{\substack{I \subseteq \{1,\dots,N\} \\ 0 \leqslant |I| \leqslant D_{ji}}} c_I^{ji} \left(\eta_{\{i\}\cup I}^{\mu_A} - \eta_{\{j\}\cup I}^{\mu_A} \right) + O\left(\delta^2\right). \tag{47}$$

6 Comparing fixation probabilities

When evaluating which of two competing types is favored by selection, it is common to compare their fixation probabilities. Specifically, the success of A relative to B is often quantified by whether $\mathbb{E}_{\mu_A}\left[\rho_A\right] > \mathbb{E}_{\mu_B}\left[\rho_B\right]$ (A is favored) or $\mathbb{E}_{\mu_A}\left[\rho_A\right] < \mathbb{E}_{\mu_B}\left[\rho_B\right]$ (B is favored) (Nowak et al., 2004; Allen and Tarnita, 2014; Tarnita and Taylor, 2014). This comparison is natural when the mutant-appearance distributions, μ_A and μ_B, are symmetric.

When μ_A and μ_B are not symmetric, however, it is less natural to compare $\mathbb{E}_{\mu_A}\left[\rho_A\right]$ to $\mathbb{E}_{\mu_B}\left[\rho_B\right]$ directly. We therefore derive an alternative measure of success for the asymmetric case. As in Section 3, we find it helpful to work with an amended Markov chain that possesses a unique stationary distribution. To this end, we suppose that, from the monomorphic state **B**, with probability u the next state is chosen from the distribution μ_A; otherwise, with probability $1-u$, the state stays **B**. Similarly, from the monomorphic state **A**, with probability u the next state is chosen from the distribution μ_B; otherwise, with probability $1-u$, the state stays **B**. The structure of this amended Markov chain is depicted in Fig. 2. Provided $u > 0$, this process has a unique stationary distribution. In analogy to our analysis of mutations into a fixed state in Section 3, we denote this stationary distribution by $\pi_{\circlearrowright(\mu_A,\mu_B)}$.

We now turn to the limit of low mutation. Applying standard results on stationary distributions in this limit (Fudenberg and Imhof, 2006, Theorem 2), we have

$$\lim_{u\to 0} \pi_{\circlearrowright(\mu_A,\mu_B)}(\mathbf{x}) = \begin{cases} \dfrac{\mathbb{E}_{\mu_A}[\rho_A]}{\mathbb{E}_{\mu_A}[\rho_A]+\mathbb{E}_{\mu_B}[\rho_B]} & \mathbf{x}=\mathbf{A}, \\[2ex] \dfrac{\mathbb{E}_{\mu_B}[\rho_B]}{\mathbb{E}_{\mu_A}[\rho_A]+\mathbb{E}_{\mu_B}[\rho_B]} & \mathbf{x}=\mathbf{B}, \\[2ex] 0 & \text{otherwise.} \end{cases} \tag{48}$$

Intuitively, for small but positive mutation rates, the process is almost always in state **A** or **B**, with the fraction of time spent in **A** converging to $\mathbb{E}_{\mu_A}[\rho_A] \,/\, \left(\mathbb{E}_{\mu_A}[\rho_A]+\mathbb{E}_{\mu_B}[\rho_B]\right)$.

We say that weak selection favors A over B if the fraction of time spent in state A is greater under weak selection than it is in the neutral case:

Definition 3. Weak selection favors A over B if

$$\frac{\mathbb{E}_{\mu_A}[\rho_A]}{\mathbb{E}_{\mu_A}[\rho_A]+\mathbb{E}_{\mu_B}[\rho_B]} > \frac{\mathbb{E}^{\circ}_{\mu_A}[\rho^{\circ}_A]}{\mathbb{E}^{\circ}_{\mu_A}[\rho^{\circ}_A]+\mathbb{E}^{\circ}_{\mu_B}[\rho^{\circ}_B]} \tag{49}$$

(or, equivalently, $\mathbb{E}_{\mu_A}[\rho_A] \,/\, \mathbb{E}_{\mu_B}[\rho_B] > \mathbb{E}^{\circ}_{\mu_A}[\rho^{\circ}_A] \,/\, \mathbb{E}^{\circ}_{\mu_B}[\rho^{\circ}_B]$) for all sufficiently small $\delta > 0$.

When μ°_A and μ°_B are symmetric (e.g. temperature or uniform initialization), we have $\mathbb{E}^{\circ}_{\mu_A}[\rho^{\circ}_A] = \mathbb{E}^{\circ}_{\mu_B}[\rho^{\circ}_B]$, and weak selection favors A relative to B if and only if $\mathbb{E}_{\mu_A}[\rho_A] > \mathbb{E}_{\mu_B}[\rho_B]$ for all sufficiently small $\delta > 0$. For general mutant-appearance distributions, μ_A and μ_B, weak selection favors A if

$$\left.\frac{d}{d\delta}\right|_{\delta=0} \frac{\mathbb{E}_{\mu_A}[\rho_A]}{\mathbb{E}_{\mu_A}[\rho_A]+\mathbb{E}_{\mu_B}[\rho_B]} > 0. \tag{50}$$

The left-hand side above has the sign of

$$\mathbb{E}^{\circ}_{\mu_B}[\rho^{\circ}_B] \left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_A}[\rho_A] - \mathbb{E}^{\circ}_{\mu_A}[\rho^{\circ}_A] \left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_B}[\rho_B]\,. \tag{51}$$

Applying Corollary 3, we obtain the condition

$$\mathbb{E}^{\circ}_{\mu_B}\left[\rho^{\circ}_B\right]\frac{d}{d\delta}\bigg|_{\delta=0}\mathbb{E}_{\mu_A}\left[\widehat{\xi}\right]+\mathbb{E}^{\circ}_{\mu_A}\left[\rho^{\circ}_A\right]\frac{d}{d\delta}\bigg|_{\delta=0}\mathbb{E}_{\mu_B}\left[\widehat{\xi}\right]$$
$$+\sum_{i=1}^{N}\pi_i\sum_{j=1}^{N}\sum_{\substack{I\subseteq\{1,\dots,N\}\\ 0\leqslant|I|\leqslant D_{ji}}}c_I^{ji}\left(\eta^{\mu_{AB}}_{\{i\}\cup I}-\eta^{\mu_{AB}}_{\{j\}\cup I}\right)>0, \tag{52}$$

where $\eta^{\mu_{AB}}_I := \mathbb{E}^{\circ}_{\mu_B}\left[\rho^{\circ}_B\right]\eta^{\mu_A}_I+\mathbb{E}^{\circ}_{\mu_A}\left[\rho^{\circ}_A\right]\eta^{\mu_B}_I$ satisfies the system of equations

$$\eta^{\mu_{AB}}_I=\mathbb{E}^{\circ}_{\mu_A}\left[\rho^{\circ}_A\right]\left(1-\mathbb{E}^{\circ}_{\mu_B}\left[\boldsymbol{\xi}_I\right]\right)-\mathbb{E}^{\circ}_{\mu_B}\left[\rho^{\circ}_B\right]\mathbb{E}^{\circ}_{\mu_A}\left[\boldsymbol{\xi}_I\right]$$
$$+\sum_{(R,\alpha)}p^{\circ}_{(R,\alpha)}\eta^{\mu_{AB}}_{\widetilde{\alpha}(I)}\quad\left(1\leqslant|I|\leqslant D+1\right); \tag{53a}$$

$$\sum_{i=1}^{N}\pi_i\eta^{\mu_{AB}}_i=0. \tag{53b}$$

Eq. (53a) is due to the facts that $\mathbb{E}^{\circ}_{\mu_A}\left[\rho^{\circ}_A\right]=\mathbb{E}^{\circ}_{\mu_A}\left[\widehat{\xi}\right]$ and $\mathbb{E}^{\circ}_{\mu_B}\left[\rho^{\circ}_B\right]=1-\mathbb{E}^{\circ}_{\mu_B}\left[\widehat{\xi}\right]$.

Example 3. Consider the case of fixation from a single mutant. Suppose that the initial mutant is located at site i with probability μ_i in both monomorphic states $\mathbf{A}$ and $\mathbf{B}$. That is,

$$\mu_A\left(\mathbf{1}_i\right)=\mu_B\left(\overline{\mathbf{1}}_i\right)=\mu_i\quad\left(1\leqslant i\leqslant N\right). \tag{54}$$

All states not of the form $\mathbf{1}_i$ for $i\in\{1,\dots,N\}$ have probability zero in μ_A, and states not of the form $\overline{\mathbf{1}}_i$ have probability zero in μ_B. Then we have $\mathbb{E}^{\circ}_{\mu_A}\left[\xi_i\right]=\mu_i$ and $\mathbb{E}^{\circ}_{\mu_B}\left[\xi_i\right]=1-\mu_i$. For non-singleton I, we have $\mathbb{E}^{\circ}_{\mu_A}\left[\boldsymbol{\xi}_I\right]=0$ and $\mathbb{E}^{\circ}_{\mu_B}\left[\boldsymbol{\xi}_I\right]=1-\sum_{i\in I}\mu_i$. We also have $\mathbb{E}^{\circ}_{\mu_A}\left[\rho^{\circ}_A\right]=\mathbb{E}^{\circ}_{\mu_B}\left[\rho^{\circ}_B\right]$ by symmetry, and both $\mathbb{E}_{\mu_A}\left[\widehat{\xi}\right]$ and $\mathbb{E}_{\mu_B}\left[\widehat{\xi}\right]$ are independent of δ. In this case, weak selection favors A if

$$\sum_{i=1}^{N}\pi_i\sum_{j=1}^{N}\sum_{\substack{I\subseteq\{1,\dots,N\}\\ 1\leqslant|I|\leqslant D_{ji}}}c_I^{ji}\left(\eta^{\mu_{AB}}_{\{i\}\cup I}-\eta^{\mu_{AB}}_{\{j\}\cup I}\right)>0, \tag{55}$$

where the $\eta^{\mu_{AB}}_I$ (upon rescaling by $1/\,\mathbb{E}^{\circ}_{\mu_A}\left[\rho^{\circ}_A\right]$) satisfy the simplified recurrence relation

$$\eta^{\mu_{AB}}_I=\begin{cases}\displaystyle\sum_{i\in I}\mu_i+\sum_{(R,\alpha)}p^{\circ}_{(R,\alpha)}\eta^{\mu_{AB}}_{\widetilde{\alpha}(I)} & 2\leqslant|I|\leqslant D+1,\\ \\ 0 & |I|=1.\end{cases} \tag{56}$$

7 Application to constant fecundity

Constant (or frequency-independent) fecundity refers to the case in which the fecundity (reproductive rate) of an individual depends on only its type and not on the composition of the rest of the population. A common setup has a mutant (A) of reproductive rate $r > 0$ competing against a resident (B) of reproductive rate 1 (Lieberman et al., 2005; Broom and Rychtář, 2008; Broom et al., 2011; Voorhees, 2013; Monk et al., 2014; Hindersin and Traulsen, 2015; Kaveh et al., 2015; Cuesta et al., 2017; Pavlogiannis et al., 2018; Möller et al., 2019; Tkadlec et al., 2019). The fixation probability of the mutant type is then a function of r, which we can write as $\rho_A(r;\boldsymbol{\xi})$, where $\boldsymbol{\xi} \in \mathbb{B}_{\mathsf{T}}^{N}$ is the initial configuration of the mutant and resident.

A common feature of these models is that the fecundity is interpreted in a relative sense, meaning that r quantifies the reproductive rate of A relative to B. Consequently, the fixation probabilities of both types are invariant under rescaling the reproductive rates of all individuals. In particular, the fixation probability of a mutant of reproductive rate r competing against a resident of reproductive rate 1 is equal (upon dividing by r) to the fixation probability of a mutant of reproductive rate 1 competing against a resident of reproductive rate r^{-1}. By interchanging the roles of A and B, we see that fixation probabilities satisfy the duality

$$\rho_A(r;\boldsymbol{\xi}) = 1 - \rho_A\left(r^{-1};\overline{\boldsymbol{\xi}}\right). \tag{57}$$

For r close to 1, meaning that the mutation has only a small effect on fecundity, a trait's fixation probability can be analyzed using weak-selection methods such as those considered here (Allen et al., 2020). To apply these methods, we define the mutant's *selection coefficient* as $s = r - 1$. We then obtain an expansion of the fixation probability, $\rho_A(1+s;\boldsymbol{\xi})$, around $s = 0$. The selection coefficient s plays a similar role to the selection intensity δ in the foregoing sections, except that s can also be negative, indicating a disadvantageous mutant.

Taking the s-derivative of both sides of Eq. (57) at $s = 0$, it follows that

$$\begin{aligned}\left.\frac{d}{ds}\right|_{s=0} \rho_A(1+s;\boldsymbol{\xi}) &= \left.\frac{d}{ds}\right|_{s=0} \rho_A\left(1+s;\overline{\boldsymbol{\xi}}\right) \\ &= \frac{1}{2}\left.\frac{d}{ds}\right|_{s=0} \left(\rho_A(1+s;\boldsymbol{\xi}) + \rho_A\left(1+s;\overline{\boldsymbol{\xi}}\right)\right). \end{aligned} \tag{58}$$

Applying Theorem 2 (with s in place of δ), we have the following weak-selection expansion for ρ_A:

$$\rho_A(1+s;\boldsymbol{\xi}) = \widehat{\xi} + \frac{s}{2}\sum_{i=1}^{N} \pi_i \sum_{j=1}^{N} \sum_{\substack{I \subseteq \{1,\dots,N\} \\ 0 \leqslant |I| \leqslant D_{ji}}} c_I^{ji} \left(\widetilde{\eta}_{\{i\}\cup I}^{\boldsymbol{\xi}} - \widetilde{\eta}_{\{j\}\cup I}^{\boldsymbol{\xi}}\right) + O\left(s^2\right), \tag{59}$$

where $\widetilde{\eta}_I^{\boldsymbol{\xi}} := \eta_I^{\boldsymbol{\xi}} + \eta_I^{\overline{\boldsymbol{\xi}}}$. These $\widetilde{\eta}_I^{\boldsymbol{\xi}}$ are the unique solution to the recurrence relation

$$\widetilde{\eta}_I^{\boldsymbol{\xi}} = \begin{cases} 1 - \left(\boldsymbol{\xi}_I + \overline{\boldsymbol{\xi}}_I\right) + \sum\limits_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \widetilde{\eta}^{\boldsymbol{\xi}}_{\widetilde{\alpha}(I)} & 2 \leqslant |I| \leqslant D+1, \\ \\ 0 & |I| = 1. \end{cases} \tag{60}$$

Note that $\boldsymbol{\xi}_I + \overline{\boldsymbol{\xi}}_I$ is equal to 1 if all individuals in I have the same type in $\boldsymbol{\xi}$ (i.e. $\xi_i = \xi_j$ for all $i, j \in I$) and 0 otherwise. In particular, $\xi_i + \overline{\xi}_i = 1$ for all $i = 1, \dots, N$, which is why $\widetilde{\eta}_I^{\boldsymbol{\xi}} = 0$ for $|I| = 1$.

If the initial state $\boldsymbol{\xi}$ is drawn from a mutant-appearance distribution, μ_A, we then have

$$\left.\frac{d}{ds}\right|_{s=0} \mathbb{E}_{\mu_A}\left[\rho_A\right] = \left.\frac{d}{ds}\right|_{s=0} \mathbb{E}_{\mu_A}\left[\widehat{\xi}\right] + \frac{1}{2}\sum_{i=1}^{N} \pi_i \sum_{j=1}^{N} \sum_{\substack{I \subseteq \{1,\dots,N\} \\ 0 \leqslant |I| \leqslant D_{ji}}} c_I^{ji} \left(\widetilde{\eta}^{\mu_A}_{\{i\}\cup I} - \widetilde{\eta}^{\mu_A}_{\{j\}\cup I}\right), \tag{61}$$

where

$$\widetilde{\eta}_I^{\mu_A} = \begin{cases} 1 - \mathbb{E}^{\circ}_{\mu_A}\left[\boldsymbol{\xi}_I + \overline{\boldsymbol{\xi}}_I\right] + \sum\limits_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \widetilde{\eta}^{\mu_A}_{\widetilde{\alpha}(I)} & 2 \leqslant |I| \leqslant D+1, \\ \\ 0 & |I| = 1. \end{cases} \tag{62}$$

In particular, for the mutant-appearance distribution in Example 3, in which the initial state has a single mutant whose location i is chosen with probability μ_i (independent of r), we have

$$\mathbb{E}_{\mu_A}\left[\rho_A\right] = \sum_{i=1}^{N} \pi_i \mu_i + \frac{s}{2} \sum_{i=1}^{N} \pi_i \sum_{j=1}^{N} \sum_{\substack{I \subseteq \{1,\dots,N\} \\ 0 \leqslant |I| \leqslant D_{ji}}} c_I^{ji} \left(\widetilde{\eta}^{\mu_A}_{\{i\}\cup I} - \widetilde{\eta}^{\mu_A}_{\{j\}\cup I}\right) + O\left(s^2\right), \tag{63}$$

where, owing to the fact that $\mathbb{E}^{\circ}_{\mu_A}\left[\boldsymbol{\xi}_I + \overline{\boldsymbol{\xi}}_I\right] = \mathbb{E}^{\circ}_{\mu_A}\left[\overline{\boldsymbol{\xi}}_I\right] = 1 - \sum_{i \in I} \mu_i$ when $|I| > 1$,

$$\widetilde{\eta}_I^{\mu_A} = \begin{cases} \sum\limits_{i \in I} \mu_i + \sum\limits_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \widetilde{\eta}^{\mu_A}_{\widetilde{\alpha}(I)} & 2 \leqslant |I| \leqslant D+1, \\ \\ 0 & |I| = 1. \end{cases} \tag{64}$$

Eq. (64) is identical to the recurrence relation for $\eta_I^{\mu_A}$ in Eq. (56).

7.1 Constant fecundity on graphs

Let us now suppose that the population structure is represented by an undirected, unweighted graph. Each vertex is occupied by one individual. As above, we consider a mutant type, A, of fecundity $r = 1 + s$, competing against a resident type, B, of fecundity 1. A robust research program aims to elucidate the effects of graph structure on the mutant's fixation probability (Lieberman et al., 2005; Broom and Rychtář, 2008; Broom et al., 2011; Voorhees, 2013; Monk et al., 2014; Hindersin and Traulsen, 2015; Kaveh et al., 2015; Cuesta et al., 2017; Pavlogiannis et al., 2018; Möller et al., 2019; Tkadlec et al., 2019).

The edge weight between vertices i and j is denoted w_{ij} ($= w_{ji}$). We define the *weighted degree* of vertex i as $w_i := \sum_{j=1}^{N} w_{ij}$. One can define a natural random walk on this graph, moving from vertex i to vertex j with probability $p_{ij}^{(1)} := w_{ij}/w_i$ (where the superscript indicates that this probability is for one step in the random walk). More generally, the probability of moving from i to j in n steps of this random walk is $p_{ij}^{(n)} := \left(p^{(1)}\right)_{ij}^{n}$, i.e. entry (i,j) of the nth power of the transition matrix $p^{(1)}$.

7.1.1 death-Birth updating

Under the *death-Birth* process (Ohtsuki et al., 2006; Hindersin and Traulsen, 2015; Allen et al., 2020; Tkadlec et al., 2020), an individual is first chosen, uniformly-at-random from the population, to be replaced ("death"). A neighbor is then chosen, with probability proportional to the product of edge weight and fecundity, to produce an offspring that fills the vacancy ("Birth"). The term "Birth" is capitalized here to emphasize the fact that selection influences this step (Hindersin and Traulsen, 2015).

For this process, the probability that the offspring of i replaces the occupant of j is

$$e_{ij}\left(1+s;\mathbf{x}\right) = \frac{1}{N}\frac{w_{ij}\left(\left(1+s\right)x_i+1-x_i\right)}{\sum_{k=1}^{N} w_{kj}\left(\left(1+s\right)x_k+1-x_k\right)} = \frac{p_{ji}^{(1)}}{N}\frac{1+sx_i}{1+s\sum_{k=1}^{N} p_{jk}^{(1)} x_k}. \tag{65}$$

Differentiating this expression with respect to s at $s=0$ gives

$$\left.\frac{d}{ds}\right|_{s=0} e_{ij}\left(1+s;\mathbf{x}\right) = \sum_{k=1}^{N} c_k^{ij} x_k, \tag{66}$$

where

$$c_k^{ij} = \begin{cases} \dfrac{p_{ji}^{(1)}}{N}\left(1-p_{ji}^{(1)}\right) & k=i, \\ \\ -\dfrac{p_{ji}^{(1)}}{N} p_{jk}^{(1)} & \text{otherwise}. \end{cases} \tag{67}$$

This process therefore has degree $D=1$.

Under neutral drift, $e_{ij}^{\circ} = p_{ji}^{(1)}/N$, and solving Eq. (12) yields $\pi_i = w_i / \sum_{j=1}^{N} w_j$ as the reproductive value of vertex i under death-Birth updating (see also Maciejewski, 2014; Allen et al., 2017; Allen and McAvoy, 2019). Using Eq. (59) and the reversibility property $\pi_i p_{ij}^{(1)} = \pi_j p_{ji}^{(1)}$, a series of simplifications gives the following result:

Proposition 2. *For the death-Birth process on a weighted graph, with constant mutant fecundity $r = 1 + s$, the fixation probability from arbitrary starting configuration $\boldsymbol{\xi}$ can be expanded under weak selection as*

$$\rho_A\left(1+s;\boldsymbol{\xi}\right) = \widehat{\xi} + \frac{s}{2N}\sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \widetilde{\eta}_{ij}^{\boldsymbol{\xi}} + O\left(s^2\right), \tag{68}$$

where the terms $\widetilde{\eta}_{ij}^{\boldsymbol{\xi}}$ arise as the unique solution to the recurrence relation

$$\widetilde{\eta}_{ij}^{\boldsymbol{\xi}} = \begin{cases} \frac{N}{2}\left(\xi_i + \xi_j - 2\xi_i\xi_j\right) + \frac{1}{2}\sum_{k=1}^{N}\left(p_{ik}^{(1)}\widetilde{\eta}_{kj}^{\boldsymbol{\xi}} + p_{jk}^{(1)}\widetilde{\eta}_{ik}^{\boldsymbol{\xi}}\right) & i \neq j, \\ 0 & i = j. \end{cases} \tag{69}$$

The factor of 2 above is related to the fact that we focus on only those replacement events that influence i and j.

For a single mutant initialized randomly as in Example 3, Eq. (63) becomes

$$\mathbb{E}_{\mu_A}\left[\rho_A\right] = \sum_{i=1}^{N} \pi_i \mu_i + \frac{s}{2N}\sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} \widetilde{\eta}_{ij}^{\mu_A} + O\left(s^2\right), \tag{70}$$

where

$$\widetilde{\eta}_{ij}^{\mu_A} = \begin{cases} N\left(\frac{\mu_i + \mu_j}{2}\right) + \frac{1}{2}\sum_{k=1}^{N}\left(p_{ik}^{(1)}\widetilde{\eta}_{kj}^{\mu_A} + p_{jk}^{(1)}\widetilde{\eta}_{ik}^{\mu_A}\right) & i \neq j, \\ 0 & i = j. \end{cases} \tag{71}$$

Eqs. (70)–(71) generalize a result of Allen et al. (2020), which pertained to the case of uniform initialization, i.e. $\mu_i = 1/N$ for all i.

7.1.2 Birth-death updating

In the *Birth-death* process (Lieberman et al., 2005; Hindersin and Traulsen, 2015), also known as the *invasion process* (Antal et al., 2006), an individual i is selected to reproduce with probability proportional to fecundity; the offspring of i replaces j with probability $p_{ij}^{(1)}$.

Letting $|\mathbf{x}| := \sum_{i=1}^{N} x_i$ denote the abundance of type A in state $\mathbf{x}$, the probability that i replaces j in this state is

$$e_{ij}\left(1+s;\mathbf{x}\right) = \frac{\left(1+s\right)x_i + 1 - x_i}{\sum_{k=1}^{N}\left(\left(1+s\right)x_k + 1 - x_k\right)} p_{ij}^{(1)} = \frac{1 + sx_i}{N + s\left|\mathbf{x}\right|} p_{ij}^{(1)}. \tag{72}$$

Differentiating this expression with respect to s at $s = 0$ gives

$$\frac{d}{ds}\bigg|_{s=0} e_{ij}\left(1+s;\mathbf{x}\right) = \sum_{k=1}^{N} c_{k}^{ij} x_{k}, \tag{73}$$

where

$$c_{k}^{ij} = \begin{cases} \dfrac{N-1}{N^2} p_{ij}^{(1)} & k = i, \\ \\ -\dfrac{1}{N^2} p_{ij}^{(1)} & \text{otherwise.} \end{cases} \tag{74}$$

Under neutral drift, $e_{ij}^{\circ} = p_{ij}^{(1)} / N$, and Eq. (12) yields a reproductive value of $\pi_i = w_i^{-1} / \sum_{j=1}^{N} w_j^{-1}$ for Birth-death updating (see also Maciejewski, 2014; Allen et al., 2017; Allen and McAvoy, 2019). A series of simplifications based on Eq. (59) and the relation $\pi_i p_{ji}^{(1)} = \pi_j p_{ij}^{(1)}$ gives the following result:

Proposition 3. *For the Birth-death process on a weighted graph, with constant mutant fecundity $r = 1 + s$, the fixation probability from arbitrary starting configuration $\boldsymbol{\xi}$ can be expanded under weak selection as*

$$\rho_A(1+s) = \widehat{\xi} + \frac{s}{2N} \sum_{i,j=1}^{N} \pi_i p_{ji}^{(1)} \widetilde{\eta}_{ij}^{\boldsymbol{\xi}} + O\left(s^2\right), \tag{75}$$

where the terms $\widetilde{\eta}_{ij}^{\boldsymbol{\xi}}$ arise as the unique solution to

$$\widetilde{\eta}_{ij}^{\boldsymbol{\xi}} = \begin{cases} \dfrac{N\left(\xi_i + \xi_j - 2\xi_i\xi_j\right) + \sum_{k=1}^{N}\left(p_{ki}^{(1)}\widetilde{\eta}_{kj}^{\boldsymbol{\xi}} + p_{kj}^{(1)}\widetilde{\eta}_{ik}^{\boldsymbol{\xi}}\right)}{\sum_{k=1}^{N}\left(p_{ki}^{(1)} + p_{kj}^{(1)}\right)} & i \neq j, \\ \\ 0 & i = j. \end{cases} \tag{76}$$

The factor $\sum_{k=1}^{N}\left(p_{ki}^{(1)} + p_{kj}^{(1)}\right)$ is analogous to the factor of 2 in the corresponding expression for death-Birth updating (due to considering the effects of a replacement rule on only i and j). However, death is not necessarily uniform under Birth-death updating, which results in a slightly more complicated scaling factor.

For the mutant-appearance distribution of Example 3, Eq. (63) simplifies to

$$\mathbb{E}_{\mu_A}\left[\rho_A\right] = \sum_{i=1}^{N} \pi_i \mu_i + \frac{s}{2N} \sum_{i,j=1}^{N} \pi_i p_{ji}^{(1)} \widetilde{\eta}_{ij}^{\mu_A} + O\left(s^2\right), \tag{77}$$

where

$$\widetilde{\eta}_{ij}^{\mu_A} = \begin{cases} \dfrac{N\left(\mu_i + \mu_j\right) + \sum_{k=1}^{N} \left(p_{ki}^{(1)} \widetilde{\eta}_{kj}^{\mu_A} + p_{kj}^{(1)} \widetilde{\eta}_{ik}^{\mu_A} \right)}{\sum_{k=1}^{N} \left(p_{ki}^{(1)} + p_{kj}^{(1)} \right)} & i \neq j, \\ \\ 0 & i = j. \end{cases} \tag{78}$$

8 Application to evolutionary game theory

We now move from constant fecundity to evolutionary games (frequency-dependent fecundity) in structured populations (Blume, 1993; Nowak and May, 1992; Ohtsuki et al., 2006; Szabó and Fáth, 2007; Nowak et al., 2009). In this setting, individuals interact with one another and receive a net payoff based on the types (strategies) of those with whom they interact. In state $\mathbf{x} \in \mathbb{B}^N$, we let $u_i\left(\mathbf{x}\right)$ denote the payoff (or utility) of player i. This payoff is converted into relative fecundity, F_i, by letting $F_i\left(\mathbf{x}\right) = \exp\left\{\delta u_i\left(\mathbf{x}\right)\right\}$, where $\delta > 0$ is the selection intensity parameter. (An alternative convention, $F_i\left(\mathbf{x}\right) = 1 + \delta u_i\left(\mathbf{x}\right)$, is equivalent under weak selection, since both satisfy $F_i\left(\mathbf{x}\right) = 1 + \delta u_i\left(\mathbf{x}\right) + O\left(\delta^2\right)$.) The replacement rule then depends directly on the fecundity vector, $\mathbf{F} \in \left(0, \infty\right)^N$, i.e. $e_{ij}\left(\mathbf{x}\right) = e_{ij}\left(\mathbf{F}\left(\mathbf{x}\right)\right)$. Furthermore, under weak selection, there is no loss of generality in assuming that the state-to-fecundity mapping, $\mathbf{x} \mapsto \mathbf{F}\left(\mathbf{x}\right)$, is deterministic (see McAvoy et al., 2020). Therefore, for every $i, j = 1, \dots, N$, we can write

$$\left.\frac{d}{d\delta}\right|_{\delta=0} e_{ij}\left(\mathbf{x}\right) = \sum_{k=1}^{N} \left(\left.\frac{\partial}{\partial F_k}\right|_{\mathbf{F}=\mathbf{1}} e_{ij}\left(\mathbf{F}\right) \right) u_k\left(\mathbf{x}\right). \tag{79}$$

Since the fecundity derivative $\left.\frac{\partial}{\partial F_k}\right|_{\mathbf{F}=\mathbf{1}} e_{ij}\left(\mathbf{F}\right)$ does not depend on $\mathbf{x}$, the degree of the process under weak selection is controlled by the utility functions $\left\{u_i\left(\mathbf{x}\right)\right\}_{i=1}^{N}$. Let $m_k^{ij} := \left.\frac{\partial}{\partial F_k}\right|_{\mathbf{F}=\mathbf{1}} e_{ij}\left(\mathbf{F}\right)$ be the marginal effect of the fecundity of k on i replacing j (McAvoy et al., 2020), and suppose that individual k's payoff is $u_k\left(\mathbf{x}\right) = \sum_{I \subseteq \left\{1,\dots,N\right\}} p_I^k \mathbf{x}_I$. We then have

$$\left.\frac{d}{d\delta}\right|_{\delta=0} e_{ij}\left(\mathbf{x}\right) = \sum_{k=1}^{N} m_k^{ij} u_k\left(\mathbf{x}\right) = \sum_{I \subseteq \left\{1,\dots,N\right\}} \left(\sum_{k=1}^{N} m_k^{ij} p_I^k \right) \mathbf{x}_I, \tag{80}$$

which, by the uniqueness of the representation of Eq. (6), gives $c_I^{ij} = \sum_{k=1}^{N} m_k^{ij} p_I^k$. Therefore, generically, the degree of the process coincides with the maximal degree of the payoff functions when each payoff function is viewed as a multi-linear polynomial in $x_1, \dots, x_N$ (see also Ohtsuki, 2014; McAvoy and Hauert, 2016).

8.1 Additive games

Additive games are a special class of games for which the conditions to be favored by selection can be written in a simplified form. An evolutionary game is additive if its payoff function, $u_i\left(\mathbf{x}\right)$, is of degree one (linear) in $\mathbf{x}$ for every $i = 1, \ldots, N$. In this case we can write

$$\left.\frac{d}{d\delta}\right|_{\delta=0} e_{ij}(\mathbf{x}) = c_0^{ij} + \sum_{k=1}^{N} c_k^{ij} x_k \tag{81}$$

for some collection of constants c_k^{ij} with $i, j, k \in \{1, \ldots, N\}$. If we further assume that the all-B state has the same replacement probabilities as the neutral process—that is, $p_{(R,\alpha)}(\mathbf{B}) = p^{\circ}_{(R,\alpha)}$ for all (R, α) and all $\delta > 0$—it then follows that $c_0^{ij} = 0$ for all i, j. In this case, Theorem 2 gives

$$\left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_A}[\rho_A] = \left.\frac{d}{d\delta}\right|_{\delta=0} \mathbb{E}_{\mu_A}\left[\widehat{\xi}\right] + \sum_{i,j,k=1}^{N} \pi_i c_k^{ji} \left(\eta_{ik}^{\mu_A} - \eta_{jk}^{\mu_A}\right), \tag{82}$$

where the terms $\eta_{ij}^{\mu_A}$ arise as the unique solution to the equations

$$\eta_{ij}^{\mu_A} = \mathbb{E}^{\circ}_{\mu_A}\left[\widehat{\xi} - \xi_i \xi_j\right] + \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \eta^{\mu_A}_{\widetilde{\alpha}(i)\widetilde{\alpha}(j)} \quad (1 \leqslant i, j \leqslant N); \tag{83a}$$

$$\sum_{i=1}^{N} \pi_i \eta_{ii}^{\mu_A} = 0. \tag{83b}$$

Let us now consider the mutant-appearance distributions of Example 3; i.e. a single mutant appears at site i with probability μ_i. The weak-selection fixation probability of type A can then be calculated as

$$\mathbb{E}_{\mu_A}[\rho_A] = \sum_{i=1}^{N} \pi_i \mu_i + \delta \sum_{i,j,k=1}^{N} \pi_i c_k^{ji} \left(\eta_{ik}^{\mu_A} - \eta_{jk}^{\mu_A}\right) + O\left(\delta^2\right), \tag{84}$$

with $\eta_{ij}^{\mu_A}$ as above. Furthermore, the condition for weak selection to favor A, Eq. (55), becomes

$$\sum_{i,j,k=1}^{N} \pi_i c_k^{ji} \left(\eta_{ik}^{\mu_{AB}} - \eta_{jk}^{\mu_{AB}}\right) > 0, \tag{85}$$

where the terms $\eta_{ij}^{\mu_{AB}}$ arise as the unique solution to

$$\eta_{ij}^{\mu_{AB}} = \begin{cases} \mu_i + \mu_j + \sum_{(R,\alpha)} p^{\circ}_{(R,\alpha)} \eta^{\mu_{AB}}_{\widetilde{\alpha}(i)\widetilde{\alpha}(j)} & i \neq j, \\ 0 & i = j. \end{cases} \tag{86}$$

8.2 Graph-structured populations with death-Birth updating

Evolutionary games on a graphs provide well-known models for social interactions in structured populations (Santos and Pacheco, 2005; Ohtsuki et al., 2006; Taylor et al., 2007; Szabó and Fáth, 2007; Santos et al., 2008; Chen, 2013; Débarre et al., 2014; Allen et al., 2017, 2019; McAvoy et al., 2020). As in Section 7.1, we suppose that the population structure is represented by a weighted, undirected graph with adjacency matrix $\left(w_{ij}\right)_{i,j=1}^{N}$. We adopt the notation of Section 7.1 for the weighted degree $w_i := \sum_{j=1}^{N} w_{ij}$ and the step probability $p_{ij}^{(1)} := w_{ij}/w_i$.

As in Section 7.1.1, we consider death-Birth updating: an individual is chosen uniformly at random for death, and then a neighboring individual is chosen, with probability proportional to the product of fecundity and edge weight, to reproduce into the vacancy. With this update rule, the marginal probability that i transmits an offspring to j is

$$e_{ij}\left(\mathbf{x}\right) = \frac{1}{N}\frac{F_i\left(\mathbf{x}\right) w_{ij}}{\sum_{\ell=1}^{N} F_\ell\left(\mathbf{x}\right) w_{\ell j}}. \tag{87}$$

Differentiating with respect to δ at $\delta = 0$ gives

$$\begin{aligned} \left.\frac{d}{d\delta}\right|_{\delta=0} e_{ij}\left(\mathbf{x}\right) &= \frac{1}{N}\frac{u_i\left(\mathbf{x}\right) w_{ij} w_j - w_{ij}\sum_{\ell=1}^{N} u_\ell\left(\mathbf{x}\right) w_{\ell j}}{w_j^2} \\ &= \frac{p_{ji}^{(1)}}{N}\left(u_i\left(\mathbf{x}\right) - \sum_{\ell=1}^{N} p_{j\ell}^{(1)} u_\ell\left(\mathbf{x}\right)\right). \end{aligned} \tag{88}$$

We now focus on a particular game called the *donation game* (Sigmund, 2010), in which type A pays a cost of c to donate b to every neighbor; type B donates nothing and pays no cost. For $b > c > 0$, this is a special case of the prisoner's dilemma, with A playing the role of cooperators and B playing the role of defectors.

There are two main conventions for aggregating the payoffs received from these game interactions (Maciejewski et al., 2014). The first is to take the edge-weighted sum of the payoffs received from all others; this method is called *accumulated payoffs*, and leads to $u_i\left(\mathbf{x}\right) = -w_i c x_i + \sum_{k=1}^{N} w_{ik} b x_k$. The second is to take the edge-weighted average (i.e. to normalize the sum by the weighted degree); this method is called *averaged payoffs* and leads to $u_i\left(\mathbf{x}\right) = -c x_i + b\sum_{k=1}^{N} p_{ik}^{(1)} x_k$. In both cases, the game is additive according to the definition of the previous subsection.

Substituting the respective payoff functions into Eq. (88) yields

$$\left.\frac{d}{d\delta}\right|_{\delta=0} e_{ij}\left(\mathbf{x}\right) = \sum_{k=1}^{N} c_k^{ij} x_k, \tag{89}$$

where

$$c_k^{ij} = \begin{cases} \dfrac{p_{ji}^{(1)}}{N}\left(-c\left(p_{ik}^{(0)} - p_{jk}^{(1)}\right) + b\left(p_{ik}^{(1)} - p_{jk}^{(2)}\right)\right) & \text{(averaged)}, \\ \dfrac{p_{ji}^{(1)}}{N}\left(-c\left(p_{ik}^{(0)} w_i - p_{jk}^{(1)} w_k\right) + b\left(w_{ik} - \sum_{\ell=1}^{N} p_{j\ell}^{(1)} w_{\ell k}\right)\right) & \text{(accumulated)}. \end{cases} \tag{90}$$

Above, we have used $p_{ij}^{(n)}$ to denote the probability that an n-step random walk from i terminates at j; note in particular that $p_{ij}^{(0)}$ equals 1 if $i = j$ and 0 otherwise. We further define $\eta_{(n)}^{\boldsymbol{\xi}} := \sum_{i,j=1}^{N} \pi_i p_{ij}^{(n)} \eta_{ij}^{\boldsymbol{\xi}}$ as the expectation of $\eta_{ij}^{\boldsymbol{\xi}}$ when i and j are sampled from the two ends of a stationary n-step random walk, where $n \geqslant 0$. We can then state the following result:

Proposition 4. *For the donation game on an arbitrary weighted, connected graph, with death-Birth updating, the fixation probability of cooperators from an arbitrary configuration $\boldsymbol{\xi}$ can be expanded under weak selection as*

$$\rho_A(\boldsymbol{\xi}) = \widehat{\xi} + \frac{\delta}{N}\left(-c\eta_{(2)}^{\boldsymbol{\xi}} + b\left(\eta_{(3)}^{\boldsymbol{\xi}} - \eta_{(1)}^{\boldsymbol{\xi}}\right)\right) + O\left(\delta^2\right), \tag{91}$$

for averaged payoffs, and

$$\begin{aligned} \rho_A(\boldsymbol{\xi}) = \widehat{\xi} + \frac{\delta}{N}\Bigg(& -c\sum_{i,j=1}^{N} \pi_i \left(p_{ij}^{(2)} - p_{ij}^{(0)}\right) w_j \eta_{ij}^{\boldsymbol{\xi}} \\ & + b\sum_{i,j,k=1}^{N} \pi_i \left(p_{ij}^{(2)} - p_{ij}^{(0)}\right) w_{jk} \eta_{ik}^{\boldsymbol{\xi}}\Bigg) + O\left(\delta^2\right), \end{aligned} \tag{92}$$

for accumulated payoffs. In both cases, the terms $\eta_{ij}^{\boldsymbol{\xi}}$ arise as the unique solution to

$$\eta_{ij}^{\boldsymbol{\xi}} = \frac{N}{2}\left(\widehat{\xi} - \xi_i \xi_j\right) + \frac{1}{2}\sum_{k=1}^{N}\left(p_{ik}^{(1)} \eta_{kj}^{\boldsymbol{\xi}} + p_{jk}^{(1)} \eta_{ik}^{\boldsymbol{\xi}}\right) \quad (i \neq j); \tag{93a}$$

$$\eta_{ii}^{\boldsymbol{\xi}} = N\left(\widehat{\xi} - \xi_i\right) + \sum_{j=1}^{N} p_{ij}^{(1)} \eta_{jj}^{\boldsymbol{\xi}}; \tag{93b}$$

$$\sum_{i=1}^{N} \pi_i \eta_{ii}^{\boldsymbol{\xi}} = 0. \tag{93c}$$

Proof. Theorem 2 gives

$$\rho_A(\boldsymbol{\xi}) = \widehat{\xi} + \delta \sum_{i,j,k=1}^{N} \pi_i c_k^{ji}\left(\eta_{ik}^{\boldsymbol{\xi}} - \eta_{jk}^{\boldsymbol{\xi}}\right) + O\left(\delta^2\right), \tag{94}$$

where the $\eta_{ij}^{\boldsymbol{\xi}}$ are the unique solution to Eq. (93). The result then follows from applying Eq. (90) and simplifying using the reversibility property $\pi_i p_{ij}^{(n)} = \pi_j p_{ji}^{(n)}$, noting that $\eta_{(0)}^{\boldsymbol{\xi}} = \sum_{i=1}^{N} \pi_i \eta_{ii}^{\boldsymbol{\xi}} = 0$ by Eq. (93c). □

Proposition 4 generalizes one of the main results of Allen et al. (2017) (who considered only uniform initialization) to the case of an arbitrary initial state.

8.2.1 Homogeneous (regular) graphs

In the case of a regular graph, we can obtain the weak-selection expansion of fixation probabilities in closed form. Suppose the graph is unweighted (meaning each edge weight is either 0 or 1), has no self-loops ($w_{ii} = 0$ for each i) and is regular of degree d ($w_i = d$ for all i). For regular graphs, accumulated and averaged payoffs are equivalent upon rescaling all payoffs by a factor of d; we consider averaged payoffs here.

Noting that $\pi_i = 1/N$ for all i, and $\widehat{\xi} = |\boldsymbol{\xi}|/N$, for regular graphs, Eq. (93) for $\eta_{ij}^{\boldsymbol{\xi}}$ can be written as

$$\begin{aligned}
\eta_{ij}^{\boldsymbol{\xi}} = \frac{1}{2}\left(|\boldsymbol{\xi}| - N\xi_i\xi_j\right) + \frac{1}{2}\sum_{k=1}^{N}\left(p_{ik}^{(1)}\eta_{kj}^{\boldsymbol{\xi}} + p_{jk}^{(1)}\eta_{ik}^{\boldsymbol{\xi}}\right) \qquad (95a)\\
+ \frac{\delta_{ij}}{2}\left(|\boldsymbol{\xi}| - N\xi_i + 2\sum_{k=1}^{N} p_{ik}^{(1)}\left(\eta_{kk}^{\boldsymbol{\xi}} - \eta_{ik}^{\boldsymbol{\xi}}\right)\right),\\
\sum_{i=1}^{N}\eta_{ii}^{\boldsymbol{\xi}} = 0, \qquad (95b)
\end{aligned}$$

where δ_{ij} is the Kronecker delta function. Multiplying by $\frac{1}{N}p_{ij}^{(n)}$ and summing over i and j leads to a recurrence relation for $\eta_{(n)}^{\boldsymbol{\xi}}$:

$$\begin{aligned}
\eta_{(n+1)}^{\boldsymbol{\xi}} - \eta_{(n)}^{\boldsymbol{\xi}} = \frac{1}{2}\sum_{i,j=1}^{N} p_{ij}^{(n)}\xi_i\xi_j - \frac{1}{2}|\boldsymbol{\xi}|\\
+ \frac{1}{2N}\sum_{i=1}^{N} p_{ii}^{(n)}\left(N\xi_i - |\boldsymbol{\xi}| + 2\sum_{k=1}^{N} p_{ik}^{(1)}\left(\eta_{ik}^{\boldsymbol{\xi}} - \eta_{kk}^{\boldsymbol{\xi}}\right)\right). \qquad (96)
\end{aligned}$$

We now let $n \to \infty$, taking a running average in the case that the random walk is periodic. In this limit, $p_{ij}^{(n)}$ converges to $\pi_j = 1/N$ for each i and j. Simplifying and applying Eq. (95b), Eq. (96) then becomes

$$\begin{aligned}
0 &= \frac{1}{2N}\sum_{i,j=1}^{N}\xi_i\xi_j - \frac{1}{2}|\boldsymbol{\xi}| + \frac{1}{2N^2}\sum_{i=1}^{N}\left(N\xi_i - |\boldsymbol{\xi}|\right) + \frac{1}{N^2}\left(\sum_{i,k=1}^{N} p_{ik}^{(1)}\eta_{ik}^{\boldsymbol{\xi}} - \sum_{k=1}^{N}\eta_{kk}^{\boldsymbol{\xi}}\right)\\
&= \frac{1}{2N}|\boldsymbol{\xi}|^2 - \frac{1}{2}|\boldsymbol{\xi}| + \frac{1}{N}\eta_{(1)}^{\boldsymbol{\xi}}, \qquad (97)
\end{aligned}$$

which leads to

$$\eta^{\boldsymbol{\xi}}_{(1)} = \frac{1}{2} |\boldsymbol{\xi}| \left(N - |\boldsymbol{\xi}|\right). \tag{98}$$

Applying Eq. (96), and noting that $p^{(2)}_{ii} = 1/d$ for each i, we see that

$$\eta^{\boldsymbol{\xi}}_{(2)} = \frac{1}{2} \left(|\boldsymbol{\xi}| \left(N - |\boldsymbol{\xi}| - 1\right) + \sum_{i,j=1}^{N} p^{(1)}_{ij} \xi_i \xi_j \right), \tag{99a}$$

$$\eta^{\boldsymbol{\xi}}_{(3)} = \frac{1}{2} \left(|\boldsymbol{\xi}| \left(\frac{d+1}{d} \left(N - |\boldsymbol{\xi}|\right) - 2 \right) + \sum_{i,j=1}^{N} \left(p^{(1)}_{ij} + p^{(2)}_{ij} \right) \xi_i \xi_j \right). \tag{99b}$$

Substituting into Eq. (91), we obtain the following closed-form result (a corollary to Proposition 4):

Corollary 4. *For the donation game on an unweighted regular graph with death-Birth updating, the fixation probability of cooperators from arbitrary initial configuration* $\boldsymbol{\xi}$ *can be expanded under weak selection as*

$$\rho_A\left(\boldsymbol{\xi}\right) = \frac{|\boldsymbol{\xi}|}{N} + \frac{\delta}{2N} \Bigg(-c \left(|\boldsymbol{\xi}| \left(N - |\boldsymbol{\xi}| - 1\right) + \sum_{i,j=1}^{N} p^{(1)}_{ij} \xi_i \xi_j \right) + b \left(|\boldsymbol{\xi}| \left(\frac{N - |\boldsymbol{\xi}|}{d} - 2 \right) + \sum_{i,j=1}^{N} \left(p^{(1)}_{ij} + p^{(2)}_{ij} \right) \xi_i \xi_j \right) \Bigg) + O\left(\delta^2\right). \tag{100}$$

Corollary 4 is equivalent to the main result of Chen et al. (2016). In particular, when the initial state contains only a single type A individual ($|\boldsymbol{\xi}| = 1$), we have $\sum_{i,j=1}^{N} p^{(1)}_{ij} \xi_i \xi_j = 0$ and $\sum_{i,j=1}^{N} p^{(1)}_{ij} \xi_i \xi_j = \sum_{i=1}^{N} p^{(2)}_{ii} \xi_i = 1/d$, leading to

$$\rho_A\left(\boldsymbol{\xi}\right) = \frac{1}{N} + \frac{\delta}{2N} \left(-c\left(N-2\right) + b\left(\frac{N}{d} - 2 \right) \right) + O\left(\delta^2\right). \tag{101}$$

This result holds regardless of which vertex contains the initial type A individual, as was first proven by Chen (2013).

8.3 Comparing population structures

A large body of literature is devoted to the question of whether—and to what extent—population structure can promote the evolution of cooperation (Nowak and May, 1992; Hauert and Doebeli, 2004; Santos and Pacheco, 2005; Ohtsuki et al., 2006; Taylor et al., 2007; Santos et al., 2008; Nowak et al., 2009; Tarnita et al., 2009; Débarre et al., 2014; Allen et al., 2017). The donation game with $b > c > 0$ provides an elegant model for studying this question (Sigmund, 2010):

Type A (representing cooperation) pays cost c to give benefit b to its partners; type B pays no costs and gives no benefits. Population structures can then be compared according to whether or not they increase A's chance of becoming fixed, depending on the benefit, b, and cost, c. Each population structure has a "critical benefit-to-cost ratio," $(b/c)^*$, such that weak selection increases A's fixation probability if and only if $(b/c)^* > 0$ and $b/c > (b/c)^*$ (Ohtsuki et al., 2006; Nowak et al., 2009; Allen et al., 2017).

A lower critical benefit-to-cost can then be interpreted as "better for the evolution of cooperation" (Nathanson et al., 2009). Such quantities have also been used to formally order population structures. For example, Peña et al. (2016) state that "two different models of spatial structure and associated evolutionary dynamics can be unambiguously compared by ranking their relatedness or structure coefficients: the greater the coefficient, the less stringent the conditions for cooperation to evolve. Hence, different models of population structure can be ordered by their potential to promote the evolution of cooperation in a straightforward way." While there is indeed an unambiguous comparison of population structures based on critical benefit-to-cost ratios, a comparison based on which is "better for the evolution of cooperation" is more subtle.

Consider the donation game with accumulated payoffs on a graph, with death-Birth updating and uniform mutant-appearance distribution. By Eq. (92),

$$\mathbb{E}_{\text{unif}}\left[\rho_A\right] = \frac{1}{N} + \frac{\delta}{N}\left(-c\sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} w_j \eta_{ij}^{\text{unif}} + b\sum_{i,j,k=1}^{N} \pi_i \left(p_{ij}^{(2)} - p_{ij}^{(0)}\right) w_{jk}\eta_{ik}^{\text{unif}}\right) + O\left(\delta^2\right), \tag{102}$$

where

$$\eta_{ij}^{\text{unif}} = \begin{cases} \dfrac{1}{2}\left(1 + \displaystyle\sum_{k=1}^{N}\left(p_{ik}^{(1)}\eta_{kj}^{\text{unif}} + p_{jk}^{(1)}\eta_{ik}^{\text{unif}}\right)\right) & i \neq j, \\ 0 & i = j. \end{cases} \tag{103}$$

The critical benefit-to-cost ratio is

$$\left(\frac{b}{c}\right)^* = \frac{\sum_{i,j=1}^{N} \pi_i p_{ij}^{(2)} w_j \eta_{ij}^{\text{unif}}}{\sum_{i,j,k=1}^{N} \pi_i \left(p_{ij}^{(2)} - p_{ij}^{(0)}\right) w_{jk}\eta_{ik}^{\text{unif}}}. \tag{104}$$

In Fig. 3, we apply this result to first compute the critical benefit-to-cost ratios for two heterogeneous population structures of size $N = 50$. Specifically, we give examples of graphs Γ_1 (Fig. 3**A**) and Γ_2 (Fig. 3**B**) such that $0 < (b/c)^*_{\Gamma_1} < (b/c)^*_{\Gamma_2}$, which means that the condition for cooperation to be favored on Γ_1 is less strict than that of Γ_2. However, this ranking alone does not imply that Γ_1 is unambiguously better for the evolution of cooperation than Γ_2. For example, when $b = 10$ and $c = 1$, which corresponds to $b/c > (b/c)^*_{\Gamma_1}, (b/c)^*_{\Gamma_2}$, weak selection boosts the fixation probability of cooperators on Γ_2 more than it does on Γ_1, based on the magnitudes of the first-order effects

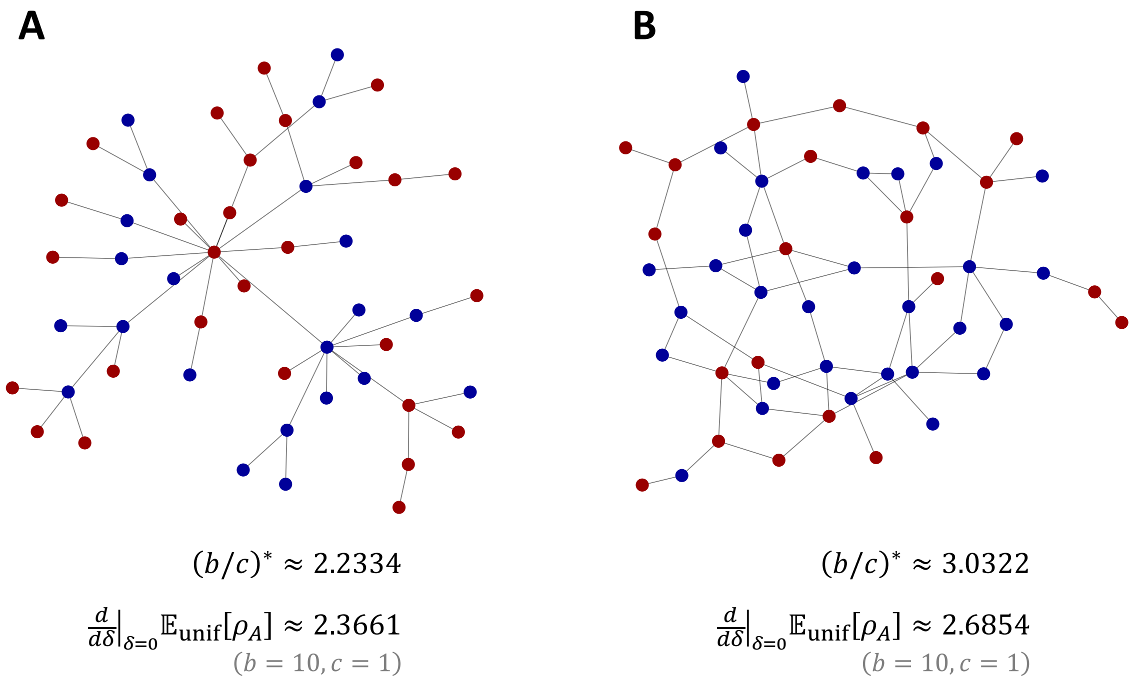

Figure 3: Two heterogeneous population structures of size $N = 50$ evolving based on death-Birth updating. The population structure depicted in **B** is unequivocally better for the evolution of cooperation than that of **A** when $b = 10$ and $c = 1$ because selection results in a greater improvement to a rare cooperator's fixation probability. This result holds despite the fact that the critical benefit-to-cost ratio of **B** is greater than it is for **A**, which means that the condition under which cooperators can thrive on **B** is stricter than that of **A**. It follows that a comparison between population structures based on which one is better for the evolution of cooperation cannot be made based on the critical ratios alone.

of selection. Therefore, for this particular cooperative social dilemma, Γ_2 more strongly supports the evolution of cooperation than Γ_1. It follows that the critical benefit-to-cost ratio provides only part of the story when comparing two population structures based on their abilities to support the emergence of cooperation.

9 Discussion

In this study, we have analyzed the fixation probability of a mutant type under weak selection, for a broad class of evolutionary models and arbitrary initial conditions. The main result, Theorem 2, gives a first-order expansion of the fixation probability of A, $\rho_A(\boldsymbol{\xi})$, in the selection intensity δ, for any initial configuration, $\boldsymbol{\xi}$. This expansion has three main ingredients: *(i)* reproductive value, π_i, which quantifies the expected contribution of i to future generations; *(ii)* neutral sojourn times, $\eta_I^{\boldsymbol{\xi}}$, which may be interpreted in terms of the mean number of steps in which all individuals in I have type A prior to absorption, given that the initial state of the population is $\boldsymbol{\xi}$; and *(iii)* Fourier coefficients, c_I^{ij}, of first-order effects of the probability that i replaces j in one update step.

It follows from Theorem 2 that the complexity of calculating this first-order expansion is

$O\left(N^{3(D+1)}\right)$, where D is the degree of the process. Actually, by the work of Le Gall (2012), this complexity is (in theory) $O\left(N^{2.373(D+1)}\right)$. This bound can be further improved in some cases by taking into account structural properties of the population, as we observed in the case of death-Birth updating on a regular graph. In any case, for fixed degree D, the system size exhibits polynomial growth in N, whereas the number of states in the evolutionary process grows exponentially in N.

The neutral sojourn times, $\eta_I^{\boldsymbol{\xi}}$ and variants thereof, play a central role in our method. Their interpretation is therefore a question of interest. From Eq. (36), we can see that $-\eta_I^{\boldsymbol{\xi}} = \langle \mathbf{x}_I - \widehat{x} \rangle_{\boldsymbol{\xi}}^{\circ}$ is a measure of the tendency for all individuals in I have type A, under neutral drift from initial state $\boldsymbol{\xi}$. In the case of a uniform initial distribution μ_A (Example 1), $\eta_I^{\mu_A}$ is proportional to the expected time for the coalescent process $\mathcal{C}$ to reach a singleton set (coalesce) starting from set I.

The utility of our framework is illustrated by the application to evolutionary dynamics on graphs in Sections 7 and 8. In particular, Proposition 4 provides the weak-selection expansion of fixation probabilities for the donation game with arbitrary graph and initial configuration. This result unifies and generalizes the main results of Chen et al. (2016) (who considered only regular graphs) as well as Allen et al. (2017) (who considered only uniform initialization).

From these results, one can derive many of the well-known results on critical benefit-to-cost ratios for cooperation to be favored in social dilemmas (Ohtsuki et al., 2006; Taylor et al., 2007; Chen, 2013; Allen and Nowak, 2014; Fotouhi et al., 2018). Moreover, they provide more information than just *when* weak selection favors a particular trait; they also determine *how much*, based on the magnitude of $\frac{d}{d\delta}\Big|_{\delta=0}\rho_A$, which can lead to more nuanced comparisons of population structures based on their ability to promote a trait (Section 8.3). The magnitude of $\frac{d}{d\delta}\Big|_{\delta=0}\rho_A$ has been explored considerably less than its sign, and our results allow this question to be explored for quite a large class of evolutionary update rules, population structures, and initial configurations.

Our results on fixation probabilities apply to finite populations of a given size. It would be interesting to connect these results to the considerable body of theory for large populations (Kimura, 1962; Roze and Rousset, 2003; Traulsen et al., 2006; Cox and Durrett, 2016; Chen, 2018), by analyzing the large-population ($N \to \infty$) asymptotics of our results such as Eq. (34). However, a number of challenges arise. First, unless one places a bound on the degree of the process, the number of terms in first-order part of Eq. (34) grows exponentially with N. Second, since Eq. (34) is itself an expansion for weak selection ($\delta \ll 1$), one must consider the relationship between N and δ as $N \to \infty$ and $\delta \to 0$. These two limits are known to be non-interchangeable, even in the relatively simple case of two-player games in a well-mixed population (Sample and Allen, 2017). However, neither of these challenges appears insurmountable, and addressing them is an important goal for future work.

Acknowledgments

We thank Krishnendu Chatterjee, Joshua Plotkin, Qi Su, and John Wakeley for helpful discussions and comments on earlier drafts. We are also grateful to the anonymous referees for many valuable suggestions. This work was supported by the Army Research Laboratory (grant W911NF-18-2-0265), the John Templeton Foundation (grant 61443), the National Science Foundation (grant DMS-1715315), the Office of Naval Research (grant N00014-16-1-2914), and the Simons Foundation (Math+X Grant to the University of Pennsylvania).

Table 1: Glossary of Notation

Symbol	Description	Introduced
$\mathbf{A}$	Monomorphic state in which all individuals have type A	Section 2.1
α	Parentage map in a replacement event	Section 2.1
$\widetilde{\alpha}$	Extension of the parentage map, α, to $\{1,\dots,N\}$	Section 2.1
$\mathbf{B}$	Monomorphic state in which all individuals have type B	Section 2.1
$\mathbb{B}$	Boolean domain $\{0,1\}$ (1 for A, 0 for B)	Section 2.1
$\mathbb{B}^N$	Set of configurations of types in the population	Section 2.1
$\mathbb{B}_{\mathsf{T}}^N$	Set of non-monomorphic configurations of types in the population, i.e. $\mathbb{B}^N - \{\mathbf{A},\mathbf{B}\}$	Section 2.1
$b_i(\mathbf{x})$	Expected offspring number of i in state $\mathbf{x}$	Section 2.2
$(b/c)^*$	Critical benefit-to-cost ratio for cooperation to evolve	Section 8.3
c_I^{ij}	Fourier coefficients of $\frac{d}{d\delta}\Big\rvert_{\delta=0} e_{ij}(\mathbf{x})$	Section 2.2
δ	Selection strength	Section 2.1
$d_i(\mathbf{x})$	Death probability of i in state $\mathbf{x}$	Section 2.2
$\Delta_{\mathrm{sel}}(\mathbf{x})$	Expected change in the frequency of A due to selection	Section 2.2
D_{ij}	Fourier degree of $\frac{d}{d\delta}\Big\rvert_{\delta=0} e_{ij}(\mathbf{x})$ as a multi-linear polynomial	Section 2.2
D	Fourier degree of the evolutionary process at $\delta=0$	Section 2.2
$e_{ij}(\mathbf{x})$	Marginal probability that i transmits its offspring to j in state $\mathbf{x}$	Section 2.2
μ_A, μ_B	Mutant-appearance distributions of types A and B, respectively	Section 5
μ_i	Probability that a single mutant appears at location i	Section 6
m_k^{ij}	Marginal effect of the fecundity of k on i replacing j	Section 8
N	Number of individuals (population size)	Section 2.1
$p_{(R,\alpha)}(\mathbf{x})$	Probability of choosing replacement event (R,α) in state $\mathbf{x}$	Section 2.1
π_i	Reproductive value of location i	Section 2.2
$\pi_{\circlearrowleft(\boldsymbol{\xi})}$	Mutation-selection stationary distribution (which depends on a fixed configuration, $\boldsymbol{\xi}$)	Section 3
$p_{ij}^{(n)}$	Probability of moving from vertex i to vertex j in n steps of a random walk on a graph	Section 7.1
R	Set of replaced positions in a replacement event	Section 2.1
r	Relative reproductive rate of a mutant in a constant-fecundity process	Section 7
ρ_A, ρ_B	Fixation probabilities of types A and B, respectively	Section 2.2

Table 1: Glossary of Notation

Symbol	Description	Introduced
s	Selection coefficient of a mutant in a constant-fecundity process	Section 7
u	Probability of regenerating a fixed transient state, $\boldsymbol{\xi}$, following $\mathbf{A}$ or $\mathbf{B}$	Section 3
$u_i(\mathbf{x})$	Payoff to i in state $\mathbf{x}$ in an evolutionary game	Section 8
w_{ij}	Edge weight between vertices i and j in a graph	Section 7.1
w_i	Degree of vertex i in a graph	Section 7.1
$\mathbf{x}$	Configuration of types in the population	Section 2.1
x_i	Type occupying i (1 for A, 0 for B)	Section 2.1
$\boldsymbol{\xi}$	Initial configuration of types in the population	Section 2.2
$^\circ$	Indicates the absence of selection	Section 2.1
$\widehat{\ }$	Indicates weighting by reproductive values	Section 2.2

References

B. Adlam, K. Chatterjee, and M. A. Nowak. Amplifiers of selection. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 471(2181):20150114, 2015. doi: 10.1098/rspa.2015.0114.

B. Allen and A. McAvoy. A mathematical formalism for natural selection with arbitrary spatial and genetic structure. *Journal of Mathematical Biology*, 78(4):1147–1210, 2019. doi: 10.1007/s00285-018-1305-z.

B. Allen and M. A. Nowak. Games on graphs. *EMS Surveys in Mathematical Sciences*, 1(1): 113–151, 2014. doi: 10.4171/emss/3.

B. Allen and C. E. Tarnita. Measures of success in a class of evolutionary models with fixed population size and structure. *Journal of Mathematical Biology*, 68(1-2):109–143, 2014. doi: 10.1007/s00285-012-0622-x.

B. Allen, C. Sample, Y. Dementieva, R. C. Medeiros, C. Paoletti, and M. A. Nowak. The Molecular Clock of Neutral Evolution Can Be Accelerated or Slowed by Asymmetric Spatial Structure. *PLOS Computational Biology*, 11(2):e1004108, 2015. doi: 10.1371/journal.pcbi.1004108.

B. Allen, G. Lippner, Y.-T. Chen, B. Fotouhi, N. Momeni, S.-T. Yau, and M. A. Nowak. Evolutionary dynamics on any population structure. *Nature*, 544(7649):227–230, 2017. doi: 10.1038/nature21723.

B. Allen, G. Lippner, and M. A. Nowak. Evolutionary games on isothermal graphs. *Nature Communications*, 10(1):5107, 2019. doi: 10.1038/s41467-019-13006-7.

B. Allen, C. Sample, R. Jencks, J. Withers, P. Steinhagen, L. Brizuela, J. Kolodny, D. Parke, G. Lippner, and Y. A. Dementieva. Transient amplifiers of selection and reducers of fixation for death-Birth updating on graphs. *PLOS Computational Biology*, 16(1):e1007529, 2020. doi: 10.1371/journal.pcbi.1007529.

T. Antal, S. Redner, and V. Sood. Evolutionary Dynamics on Degree-Heterogeneous Graphs. *Physical Review Letters*, 96(18), 2006. doi: 10.1103/physrevlett.96.188104.

L. E. Blume. The Statistical Mechanics of Strategic Interaction. *Games and Economic Behavior*, 5(3):387–424, 1993. doi: 10.1006/game.1993.1023.

E. Boros and P. L. Hammer. Pseudo-Boolean optimization. *Discrete Applied Mathematics*, 123 (1-3):155–225, 2002. doi: 10.1016/s0166-218x(01)00341-9.

M. Broom and J. Rychtář. An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 464(2098):2609–2627, 2008. doi: 10.1098/rspa.2008.0058.

M. Broom, J. Rychtář, and B. T. Stadler. Evolutionary Dynamics on Graphs - the Effect of Graph Structure and Initial Placement on Mutant Spread. *Journal of Statistical Theory and Practice*, 5 (3):369–381, 2011. doi: 10.1080/15598608.2011.10412035.

Y.-T. Chen. Sharp benefit-to-cost rules for the evolution of cooperation on regular graphs. *The Annals of Applied Probability*, 23(2):637–664, 2013. doi: 10.1214/12-aap849.

Y.-T. Chen. Wright-fisher diffusions in stochastic spatial evolutionary games with death–birth updating. *The Annals of Applied Probability*, 28(6):3418–3490, 2018. doi: 10.1214/18-aap1390.

Y.-T. Chen, A. McAvoy, and M. A. Nowak. Fixation Probabilities for Any Configuration of Two Strategies on Regular Graphs. *Scientific Reports*, 6(1), 2016. doi: 10.1038/srep39181.

J. T. Cox. Coalescing random walks and voter model consensus times on the torus in $\mathbf{Z}^d$. *Annals of Probability*, 17(4):1333–1366, 1989. doi: 10.1214/aop/1176991158.

J. T. Cox and R. Durrett. Evolutionary games on the torus with weak selection. *Stochastic Processes and their Applications*, 126(8):2388–2409, 2016. doi: 10.1016/j.spa.2016.02.004.

F. A. Cuesta, P. G. Sequeiros, and Á. L. Rojo. Suppressors of selection. *PLOS One*, 12(7): e0180549, 2017. doi: 10.1371/journal.pone.0180549.

F. A. Cuesta, P. G. Sequeiros, and Á. L. Rojo. Evolutionary regime transitions in structured populations. *PLOS One*, 13(11):e0200670, 2018. doi: 10.1371/journal.pone.0200670.

F. Débarre, C. Hauert, and M. Doebeli. Social evolution in structured populations. *Nature Communications*, 5, 2014. doi: 10.1038/ncomms4409.

R. Der, C. L. Epstein, and J. B. Plotkin. Generalized population models and the nature of genetic drift. *Theoretical Population Biology*, 80(2):80–99, 2011. doi: 10.1016/j.tpb.2011.06.004.

W. J. Ewens. *Mathematical Population Genetics*. Springer New York, 2004. doi: 10.1007/978-0-387-21822-9.

R. A. Fisher. *The Genetical Theory of Natural Selection*. Clarendon Press, 1930. doi: 10.5962/bhl.title.27468.

B. Fotouhi, N. Momeni, B. Allen, and M. A. Nowak. Conjoining uncooperative societies facilitates evolution of cooperation. *Nature Human Behaviour*, 2(7):492–499, 2018. doi: 10.1038/s41562-018-0368-6.

D. Fudenberg and L. A. Imhof. Imitation processes with small mutations. *Journal of Economic Theory*, 131(1):251–262, 2006. doi: 10.1016/j.jet.2005.04.006.

M. Grabisch, J.-L. Marichal, and M. Roubens. Equivalent Representations of Set Functions. *Mathematics of Operations Research*, 25(2):157–178, 2000. doi: 10.1287/moor.25.2.157.12225.

J. B. S. Haldane. A mathematical theory of natural and artificial selection, part V: Selection and mutation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 23, pages 838–844. Cambridge University Press, 1927. doi: 10.1017/S0305004100015644.

P. L. Hammer and S. Rudeanu. *Boolean Methods in Operations Research and Related Areas*. Springer Berlin Heidelberg, 1968. doi: 10.1007/978-3-642-85823-9.

C. Hauert and M. Doebeli. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature*, 428(6983):643–646, 2004. doi: 10.1038/nature02360.

L. Hindersin and A. Traulsen. Counterintuitive properties of the fixation time in network-structured populations. *Journal of The Royal Society Interface*, 11(99), 2014. doi: 10.1098/rsif.2014.0606.

L. Hindersin and A. Traulsen. Most Undirected Random Graphs Are Amplifiers of Selection for Birth-Death Dynamics, but Suppressors of Selection for Death-Birth Dynamics. *PLOS Computational Biology*, 11(11):e1004437, 2015. doi: 10.1371/journal.pcbi.1004437.

L. Hindersin, M. Möller, A. Traulsen, and B. Bauer. Exact numerical calculation of fixation probability and time on graphs. *Biosystems*, 150:87–91, 2016. doi: 10.1016/j.biosystems.2016.08.010.

L. A. Imhof and M. A. Nowak. Evolutionary game dynamics in a Wright-Fisher process. *Journal of Mathematical Biology*, 52(5):667–681, 2006. doi: 10.1007/s00285-005-0369-8.

K. Kaveh, N. L. Komarova, and M. Kohandel. The duality of spatial death-birth and birth-death processes and limitations of the isothermal theorem. *Royal Society Open Science*, 2(4):140465–140465, 2015. doi: 10.1098/rsos.140465.

M. Kimura. On the Probability of Fixation of Mutant Genes in a Population. *Genetics*, 47(6):713–719, 1962.

J. F. C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13(3):235–248, 1982. doi: 10.1016/0304-4149(82)90011-4.

F. Le Gall. Faster Algorithms for Rectangular Matrix Multiplication. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*. IEEE, 2012. doi: 10.1109/focs.2012.80.

S. Lessard and V. Ladret. The probability of fixation of a single mutant in an exchangeable selection model. *Journal of Mathematical Biology*, 54(5):721–744, 2007. doi: 10.1007/s00285-007-0069-7.

E. Lieberman, C. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, 2005. doi: 10.1038/nature03204.

T. M. Liggett. *Interacting Particle Systems*. Springer New York, 1985. doi: 10.1007/978-1-4613-8542-4.

W. Maciejewski. Reproductive value in graph-structured populations. *Journal of Theoretical Biology*, 340:285–293, 2014. doi: 10.1016/j.jtbi.2013.09.032.

W. Maciejewski, F. Fu, and C. Hauert. Evolutionary game dynamics in populations with heterogenous structures. *PLoS Computational Biology*, 10(4):e1003567, 2014. doi: 10.1371/journal.pcbi.1003567.

A. McAvoy and C. Hauert. Structure coefficients and strategy selection in multiplayer games. *Journal of Mathematical Biology*, 72(1):203–238, 2016. doi: 10.1007/s00285-015-0882-3.

A. McAvoy, B. Allen, and M. A. Nowak. Social goods dilemmas in heterogeneous societies. *Nature Human Behaviour*, 4(8):819–831, 2020. doi: 10.1038/s41562-020-0881-2.

D. M. McCandlish, C. L. Epstein, and J. B. Plotkin. Formal properties of the probability of fixation: Identities, inequalities and approximations. *Theoretical Population Biology*, 99:98–113, 2015. doi: 10.1016/j.tpb.2014.11.004.

M. Möller, L. Hindersin, and A. Traulsen. Exploring and mapping the universe of evolutionary graphs identifies structural properties affecting fixation probability and time. *Communications Biology*, 2(1):137, 2019. doi: 10.1038/s42003-019-0374-x.

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):20130730–20130730, 2014. doi: 10.1098/rspa.2013.0730.

P. A. P. Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(01):60, 1958. doi: 10.1017/s0305004100033193.

C. G. Nathanson, C. E. Tarnita, and M. A. Nowak. Calculating Evolutionary Dynamics in Structured Populations. *PLoS Computational Biology*, 5(12):e1000615, 2009. doi: 10.1371/journal.pcbi.1000615.

M. A. Nowak and R. M. May. Evolutionary games and spatial chaos. *Nature*, 359(6398):826–829, 1992. doi: 10.1038/359826a0.

M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983):646–650, 2004. doi: 10.1038/nature02414.

M. A. Nowak, C. E. Tarnita, and T. Antal. Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537):19–30, 2009. doi: 10.1098/rstb.2009.0215.

H. Ohtsuki. Evolutionary dynamics of n-player games played by relatives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1642):20130359–20130359, 2014. doi: 10.1098/rstb.2013.0359.

H. Ohtsuki, C. Hauert, E. Lieberman, and M. A. Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092):502–505, 2006. doi: 10.1038/nature04605.

Z. Patwa and L. M. Wahl. The fixation probability of beneficial mutations. *Journal of The Royal Society Interface*, 5(28):1279–1289, 2008. doi: 10.1098/rsif.2008.0248.

A. Pavlogiannis, J. Tkadlec, K. Chatterjee, and M. A. Nowak. Construction of arbitrarily strong amplifiers of natural selection using evolutionary graph theory. *Communications Biology*, 1(1), 2018. doi: 10.1038/s42003-018-0078-7.

J. Peña, B. Wu, and A. Traulsen. Ordering structured populations in multiplayer cooperation games. *Journal of The Royal Society Interface*, 13(114):20150881, 2016. doi: 10.1098/rsif.2015.0881.

F. Rousset. A Minimal Derivation of Convergence Stability Measures. *Journal of Theoretical Biology*, 221(4):665–668, 2003. doi: 10.1006/jtbi.2003.3210.

D. Roze and F. Rousset. Selection and drift in subdivided populations: a straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics*, 165(4):2153–2166, 2003.

C. Sample and B. Allen. The limits of weak selection and large population size in evolutionary game theory. *Journal of Mathematical Biology*, 75(5):1285–1317, 2017. doi: 10.1007/s00285-017-1119-4.

F. C. Santos and J. M. Pacheco. Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters*, 95(9):098104, 2005. doi: 10.1103/physrevlett.95.098104.

F. C. Santos, M. D. Santos, and J. M. Pacheco. Social diversity promotes the emergence of cooperation in public goods games. *Nature*, 454(7201):213–216, 2008. doi: 10.1038/nature06940.

K. Sigmund. *The calculus of selfishness*. Princeton University Press, 2010. doi: 10.1515/9781400832255.

G. Szabó and G. Fáth. Evolutionary games on graphs. *Physics Reports*, 446(4-6):97–216, 2007. doi: 10.1016/j.physrep.2007.04.004.

C. E. Tarnita and P. D. Taylor. Measures of Relative Fitness of Social Behaviors in Finite Structured Population Models. *The American Naturalist*, 184(4):477–488, 2014. doi: 10.1086/677924.

C. E. Tarnita, H. Ohtsuki, T. Antal, F. Fu, and M. A. Nowak. Strategy selection in structured populations. *Journal of Theoretical Biology*, 259(3):570–581, 2009. doi: 10.1016/j.jtbi.2009.03.035.

C. Taylor, D. Fudenberg, A. Sasaki, and M. A. Nowak. Evolutionary game dynamics in finite populations. *Bulletin of Mathematical Biology*, 66(6):1621–1644, 2004. doi: 10.1016/j.bulm.2004.03.004.

P. D. Taylor. Allele-Frequency Change in a Class-Structured Population. *The American Naturalist*, 135(1):95–106, 1990. doi: 10.1086/285034.

P. D. Taylor, T. Day, and G. Wild. Evolution of cooperation in a finite homogeneous graph. *Nature*, 447(7143):469–472, 2007. doi: 10.1038/nature05784.

J. Tkadlec, A. Pavlogiannis, K. Chatterjee, and M. A. Nowak. Population structure determines the tradeoff between fixation probability and fixation time. *Communications Biology*, 2, 2019. doi: 10.1038/s42003-019-0373-y.

J. Tkadlec, A. Pavlogiannis, K. Chatterjee, and M. A. Nowak. Limits on amplifiers of natural selection under death-birth updating. *PLOS Computational Biology*, 16(1):e1007494, 2020. doi: 10.1371/journal.pcbi.1007494.

A. Traulsen and C. Hauert. Stochastic Evolutionary Game Dynamics. In *Reviews of Nonlinear Dynamics and Complexity*, pages 25–61. Wiley-Blackwell, 2010. doi: 10.1002/9783527628001.ch2.

A. Traulsen, J. M. Pacheco, and L. A. Imhof. Stochasticity and evolutionary stability. *Physical Review E*, 74(2):021905, 2006. doi: 10.1103/physreve.74.021905.

J. Van Cleve. Social evolution and genetic interactions in the short and long term. *Theoretical Population Biology*, 103:2–26, 2015. doi: 10.1016/j.tpb.2015.05.002.

B. Voorhees. Birth-death fixation probabilities for structured populations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 469(2153):20120248, 2013. doi: 10.1098/rspa.2012.0248.

J. Wakeley. *Coalescent Theory: An Introduction*. Macmillan Learning, 2016.

S. Wright. Evolution in Mendelian populations. *Genetics*, 16:97–159, 1931.