POLITECNICO DI TORINO

Master degree course
in Physics of Complex Systems

Master Degree Thesis

Towards a universal model for structured microbial populations

Scientific supervisor
Prof. Anne-Florence Bitbol

Thesis advisor
Prof. Alessandro Pelizzola

Candidate
Irene Lamberti
student number: s253756

Academic year 2018/2019

Summary

The most important ingredients of evolution are variation and selection. Variation appears in nature as alteration in the nucleotide sequence of the genome of an organism (*mutation*) that may happen because of error during DNA replication or because of other type of damage (e.g. radiation) (Ref. [1]). Micro-organisms such as bacteria are asexual and they receive an identical copy of the parent's genome (except if there is a replication error). They can change their genetic material because of mutations but also in other ways, for example by taking up exogenous DNA from the environment (Ref. [2]). The study of the evolution of bacteria populations is of particular interest, e.g. because bacteria can develop resistance to antibiotics (Ref. [3]), one of the biggest threats to global health, food security, and development today (Ref. [4]).

The models for the evolution of microbial populations presented in this work consider the presence of one single mutant in a population of wildtype individual. For simplicity, the size of the population is kept constant and it is neglected the possibility of the appearance of new mutations. Individuals compete between each other because some of them reproduce faster than others. At a certain moment in time, no matter how large the population is, only mutant individuals will be present (*fixation*) or only wildtype individuals (*extinction*). We will focus on the stochastic description of the evolutionary process.

Most of the models studied in evolutionary dynamics consider a *well-mixed* population, i.e. a population in which each individual interact with any other to the same extent. However, this is rarely the case in nature where geographical structure plays a role. For example, during an infection, microbes are subdivided among different organs and among different hosts. Even bacteria on a Petri dish do not all interact one with another to the same extent, but each bacterium competes more with the individuals that are closer in terms of spatial distance, resulting in a smaller effective population size. Individuals belonging to different subpopulations may come into contact thanks to

migration phenomena. These two factors, migrations and local competition, contribute to maintain genetic diversity in a structured population, resulting in a complex evolutionary dynamics.

In order to model the effect of geographical structure, in Ref. [5] individuals are arranged on a graph. This model considers a weighted directed graph where one individual is placed at each node and replacement probabilities are defined on each edge. Specifically, at each time step, an individual is chosen to reproduce; its offspring replaces an individual located on one of the nodes the progenitor is connected to, thus keeping the population size constant. It has been found that some graph structures suppress selection and some others amplify selection. Since sometimes we would like to hinder selection (for example, in the evolution of antibiotic resistance) it is interesting to understand which kind of structures causes selection to be less relevant in the evolutionary process. Important results from this model, however, strongly depend on details of the chosen dynamics (in particular, whether we choose first the individual to reproduce or the individual to be replaced), raising issues on the applicability of the model to experiments. Indeed, in general, there are no strong constraints on the order of death and birth in real populations.

Since this model remains rather artificial and problematic (due to the dependence on the details of the dynamics), we study and generalize a coarse-grained version (Ref. [6]), where a well-mixed subpopulation is placed at each node of a graph and migration probabilities are defined on each edge. We will show that it is possible, within this model, to reduce the dependence of the evolutionary outcome on the dynamics.

The outline of the thesis is the following. First, in Ch. 1 we present a classic model for well-mixed populations called the Moran model, for which we compute the fixation probability and the fixation times and we compare them with simulation results. In Ch. 2 we introduce a model where individuals are arranged on the nodes of a graph and we present a brief review of some theoretical results. Within this model, we investigate the dependence on the dynamics as well as on the initial conditions using numerical simulations. We focus on the star graph which is a well known amplifier of selection in the birth-death case, and show that its property strongly depends on the chosen dynamics, the importance of self-replacement and the initial conditions. Finally, in Ch. 3 we move to the coarse-grained version of the previous model, with subpopulations on a graph, and we study it with both analytical calculations and numerical simulations.

Acknowledgements

I would like to express my gratitude to my supervisor, Anne-Florence, for her guidance, her continuous support and her advice for my future career. I owe very much to her.

I would also like to extend my thanks to all the people at Laboratoire Jean Perrin for their welcome and their interest in my work, and for providing a friendly and stimulating environment.

A thanks goes to all my professors in Turin, and in particular to Profs. Braunstein and Dall'Asta, who kept in touch with me while I was in Paris.

Finally, this thesis would not have been possible without the love and help of my family. In particular, thanks to my dad: our discussions about my scientific research and his advice were always helpful. Thanks to all my friends for their constant encouragement and to Ludovic, for his patience in reading my work and in discussing it with me day after day.

Contents

Chapter 1

The Moran model

1.1 Description of the model

The Moran model is a simple stochastic model allowing to describe the evolution of the composition of well-mixed populations. It considers a population of fixed size N made up of two types of individuals, A and B. At each discrete time step, two events take place: one individual reproduces generating one offspring identical to itself and one individual dies. Because at each time step there is a birth and a death event, the total population size remains constant. The process is schematically represented in Fig. 1.1.

Remark. We consider that the same individual may be chosen for both death and reproduction at the same time step.

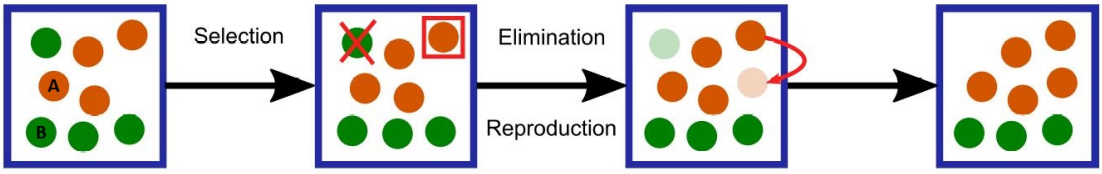


Figure 1.1: Schematic of one step of the Moran process. The population is made up of two types of individuals, type A (orange) and type B (green) (Ref. [3]).

Denoting the number of A individuals by $i = 0, 1, \ldots, N$, we call α_i the transition probability from i to $i+1$ and β_i the transition probability from i to $i-1$; thus, $1 - \alpha_i - \beta_i$ is the probability to stay at i. Fig. 1.2 shows the transition probabilities from one state to another (the probability to

stay in the same state is not drawn). A process of this type is called a "birth-and-death process" (Ref. [7]) and it is a special case of Markov process where transitions are allowed only between adjacent sites (the number of A individuals can change by one at most).

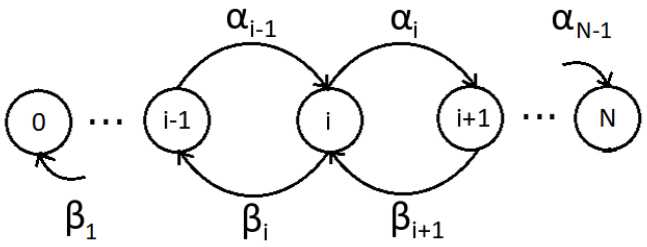


Figure 1.2: A birth-and-death process and its transition probabilities; the probability of staying in the same site is not drawn.

The process can be seen as a random walk in the space of states $i = 0,1,\ldots,N$ with absorbing boundaries 0 (all B individuals) and N (all A individuals). Both boundaries correspond to one type of individual taking over the whole population, which is called *fixation*. We are interested in knowing whether type A will fix or not and, because the process is stochastic, this means that we are interested in the probability for A to fix, which is known as the *fixation probability*.

1.1.1 Fixation probability

Let ρ_i be the probability for type A to reach state N starting at i. We can write the equation

$$\rho_i = \alpha_i \rho_{i+1} + (1 - \alpha_i - \beta_i)\rho_i + \beta_i \rho_{i-1} \tag{1.1}$$

with boundary conditions $\rho_0 = 0$ and $\rho_N = 1$. In words, the probability for i individuals to fix is given by the sum of three terms, each of them corresponding to three different outcomes of the first step: (i) the probability per unit time to go from i to $i+1$ at the first step multiplied by the fixation probability at $i+1$; (ii) the probability per unit time to stay at i multiplied by the fixation probability at i; (iii) the probability per unit time to go from i to $i-1$ multiplied by the fixation probability at $i-1$. Eq. 1.1 can be written in matrix form as

$$\boldsymbol{\rho} = M\boldsymbol{\rho} \tag{1.2}$$

where $\boldsymbol{\rho} = (\rho_1,\ldots,\rho_N)$ is the vector of fixation probabilities and M is the $N \times N$ transition matrix

$$M = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \beta_1 & 1-\alpha_1-\beta_1 & \alpha_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \beta_{N-1} & 1-\alpha_{N-1}-\beta_{N-1} & \alpha_{N-1} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

The matrix M is a right stochastic matrix, which means that its elements are nonnegative and each row sums to unity (Ref. [7]). Eq. 1.2 shows that $\boldsymbol{\rho}$ is an eigenvector of M with unit eigenvalue. Moreover, if we consider the fixation probability $\tilde{\rho}_i$ for type B when starting with i individuals of type A we notice that it also satisfies Eq. 1.1, but with boundary conditions $\tilde{\rho}_0 = 0$ and $\tilde{\rho}_N = 1$. Thus, $\tilde{\boldsymbol{\rho}}$ is also an eigenvector with unit eigenvalue of M. In addition, it is related to $\boldsymbol{\rho}$ by $\tilde{\boldsymbol{\rho}} = 1 - \boldsymbol{\rho}$, because either A or B will finally fix.

In order to express ρ_i, we follow the derivation of Ref. [8]. By introducing $y_i \doteq \rho_i - \rho_{i-1}$ and $\gamma_i \doteq \beta_i/\alpha_i$, Eq. 1.1 can be rewritten as

$$y_{i+1} = \gamma_i y_i. \tag{1.3}$$

Using this relation and the boundary conditions, we can write

$$\begin{aligned} y_1 &= \rho_1 - \rho_0 = \rho_1 \\ y_2 &= \rho_2 - \rho_1 = \gamma_1 \rho_1 \\ &\vdots \\ y_i &= \rho_i - \rho_{i-1} = \rho_1 \prod_{j=1}^{i-1} \gamma_j \\ &\vdots \\ y_N &= \rho_N - \rho_{N-1} = \rho_1 \prod_{j=1}^{N-1} \gamma_j \end{aligned} \tag{1.4}$$

The series $\{y_i : i = 1, \ldots, N\}$ is a telescopic series and it sums to unity:

$$\sum_{i=1}^{N} y_i = (\rho_1 - \rho_0) + (\rho_2 - \rho_1) + \cdots + (\rho_N - \rho_{N-1}) = \rho_N - \rho_0 = 1; \tag{1.5}$$

hence, by substituting Eq. 1.4 into Eq. 1.5 we can write

$$1 = \sum_{i=1}^{N} y_i = \rho_1 \sum_{i=1}^{N} \prod_{j=1}^{i-1} \gamma_j = \rho_1 \left(1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} \gamma_j \right)$$

where $\prod_{j=1}^{0} \gamma_j \doteq 1$. Finally, we obtain the following expression for the fixation probability of one single individual of species A:

$$\rho_1 = \frac{1}{1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} \gamma_j}. \tag{1.6}$$

Starting from i individuals of type A, we can use the relation $\rho_i = \sum_{j=1}^{i} y_j$ and substitute y_i using Eq. 1.4 to obtain

$$\rho_i = \rho_1 \left(1 + \sum_{j=1}^{i-1} \prod_{k=1}^{j} \gamma_k \right)$$

Finally, using Eq. 1.6,

$$\rho_i = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^{j} \gamma_k}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma_k} \tag{1.7}$$

From ρ_i we can obtain the fixation probability $\tilde{\rho}_{N-1}$ of a single individual of type B:

$$\tilde{\rho}_{N-1} = 1 - \rho_{N-1} = \frac{\prod_{k=1}^{N-1} \gamma_k}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma_k} \tag{1.8}$$

so we have that

$$\frac{\tilde{\rho}_{N-1}}{\rho_1} = \prod_{k=1}^{N-1} \gamma_k \tag{1.9}$$

If this product is greater then 1 than $\tilde{\rho}_{N-1} > \rho_1$ and a mutant B in a population of A individuals has a higher probability to fix than a mutant A in a population of B individuals; the other way around if the product is smaller than 1.

1.1.2 Fixation times

Other relevant quantities are the *fixation times*:

- the fixation time t_i^A for A, defined as the average time needed for the system to reach state N starting at state i, provided that N is ultimately reached;
- the fixation time t_i^B for B, defined as the average time needed for the system to reach state 0 starting at state i (note that $N - i$ is the initial number of B individuals), provided that 0 is ultimately reached;

- the unconditional fixation time, defined as the average time needed for the system to reach either 0 or N.

They are equivalent to first passage times of a random walk and they can be computed exactly. Their computation is presented in Appendix A and it follows Ref. [8].

1.1.3 The role of fitness

Until now we have not specified the way in which the individual who reproduces and the individual who dies are chosen. To go in more details, we have to introduce the concept of *fitness* of an individual which, in the context of microbial evolution, is equivalent to the reproduction rate. In the Moran process, the individual who dies is chosen uniformly at random, while the individual to reproduce is chosen proportionally to its fitness.

Let's call r the relative fitness of A with respect to B. We can now explicitly compute the transition probabilities:

$$\alpha_i = \underbrace{\frac{ri}{ri+N-i}}_{\text{prob. that an } A \text{ individual reproduces}} \cdot \underbrace{\frac{N-i}{N}}_{\text{prob. that a } B \text{ individual dies}} \tag{1.10a}$$

$$\beta_i = \underbrace{\frac{N-i}{ri+N-i}}_{\text{prob. that a } B \text{ individual reproduces}} \cdot \underbrace{\frac{i}{N}}_{\text{prob. that an } A \text{ individual dies}} \tag{1.10b}$$

from which it follows that $\gamma_i = \beta_i/\alpha_i = 1/r$. So, by substituting in Eq. 1.7 we have

$$\rho_i = \frac{1+\sum_{j=1}^{i-1} r^{-j}}{1+\sum_{j=1}^{N-1} r^{-j}} = \frac{1-r^{-i}}{1-r^{-N}} \tag{1.11}$$

and for one single mutant

$$\rho_1 = \frac{1-r^{-1}}{1-r^{-N}}. \tag{1.12}$$

Because we are interested in the fixation probability ρ_1 of one single mutant, in the following we will use the expression "fixation probability" specifically for the probability for one single mutant to fix in the population (when not otherwise specified) and we will denote it with ρ.

If A and B are neutral variants with respect to selection (namely, they have the same fitness) then $\alpha_i = \beta_i$ and the process is equivalent to an

unbiased random walk. In this case, the fixation probability for i individuals is

$$\rho_i = i/N \tag{1.13}$$

since each individual is equivalently likely to fix. In Fig. 1.3 the theoretical predictions for the fixation probability (Eq. 1.12) and the fixation times (see Appendix A) of the Moran process are plotted together with simulation results. Appendix B.1 contains some details on the numerical simulations.
In the Moran process, choosing first the individual who reproduces and then the individual who dies or the opposite is completely equivalent, since both choices lead to the same transition probabilities α_i and β_i given by Eqs. 1.10.

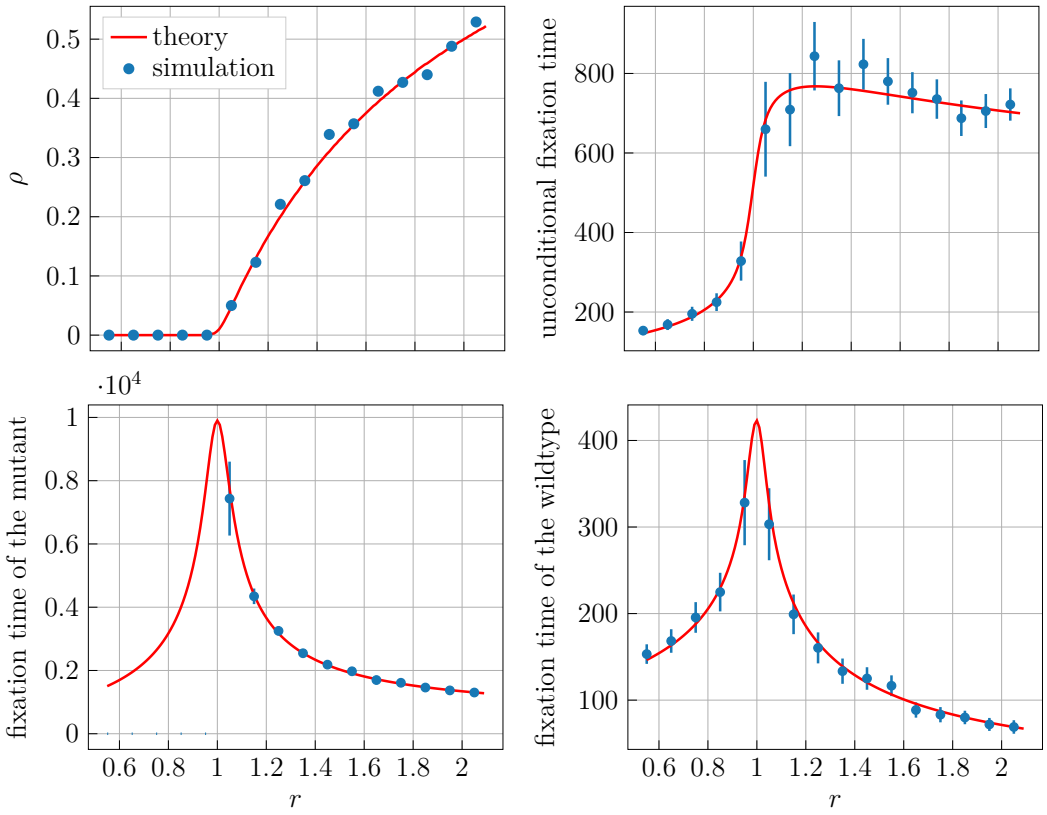


Figure 1.3: Fixation probability ρ and fixation times (in number of time steps of the simulation) as functions of the relative fitness r. The numerical results for the fixation time of a disadvantageous mutant ($r < 1$) are not shown since in this case the mutant is very unlikely to fix. The error bars correspond to 95% confidence interval. $N = 100$, 10^3 samples per data point.

We can study two asymptotic limits of Eq. 1.12 for r close to 1. Let's

assume $r = 1 + s$ with $0 < s \ll 1$; then Eq. 1.12 reads

$$\rho = \frac{1 - (1+s)^{-1}}{1 - (1+s)^{-N}}$$

Because $s \ll 1$ we can expand in Taylor series both the numerator and the denominator:

$$(1+s)^{-1} = 1 - s + \mathrm{o}(s)$$
$$(1+s)^{-N} = 1 - Ns + \mathrm{o}(Ns)$$

If $Ns \ll 1$ we get the following approximation for the fixation probability:

$$\rho \approx \frac{1}{N}. \tag{1.14}$$

If instead $Ns \gg 1$:

$$(1+s)^{-N} = \mathrm{e}^{-N\log(1+s)} = \mathrm{e}^{-N(s+\mathrm{o}(s))}$$

so that we can approximate ρ_1 as

$$\rho \approx \frac{1 - (1-s)}{\exp(-Ns)} \approx s. \tag{1.15}$$

In Fig. 1.4 the two asymptotic limits, Eqs. 1.14 and 1.15, and the exact analytic expression Eq. 1.12 are shown together with the numerical results.

If we forbid the possibility for the same individual to be chosen for both death and reproduction at the same time step then the order with which the two events take place plays a role. The *birth-death* dynamics corresponds to choosing first the individual to reproduce and then the individual who dies; the other way round is called *death-birth* dynamics. In the case of birth-death dynamics the transition probabilities are

$$\alpha_i = \frac{ri}{(r-1)i + N}\frac{N-i}{N-1} \tag{1.16a}$$
$$\beta_i = \frac{N-i}{(r-1)i + N}\frac{i}{N-1} \tag{1.16b}$$

which differ from Eqs. 1.10 through the $N-1$ instead of N at the denominator of the second factor. Nevertheless, because $\forall i\ \gamma_i = 1/r$ as in the standard Moran process discussed above and because ρ_i only depends on $\{\gamma_k : k =$

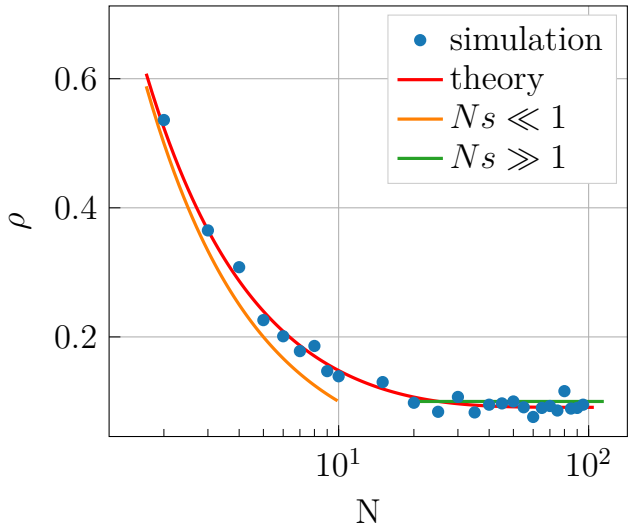


Figure 1.4: The fixation probability ρ as function of the population size N for the Moran process, semi-logarithmic scale. The asymptotic behaviour $\rho \approx 1/N$ for $Ns \ll 1$ is plotted in orange and $\rho \approx s$ for $Ns \gg 1$ is plotted in green; in red the exact solution Eq. 1.12. $r = 1.10$, 10^3 samples per data point.

$1, \ldots, N-1\}$ (see Eq. 1.11), the fixation probability is the same as in the standard Moran process.

Conversely, in the case of death-birth dynamics the transition probabilities are

$$\alpha_i = \frac{N-i}{N} \frac{ri}{(r-1)i + N - 1} \tag{1.17}$$

$$\beta_i = \frac{i}{N} \frac{N-i}{(r-1)i + N - r} \tag{1.18}$$

and hence their ratio is

$$\gamma_i = \frac{1}{r} \frac{(r-1)i + N - r}{(r-1)i + N - 1}$$

which is equal to $1/r$ only for $r = 1$ or in the infinite population limit ($N \to \infty$). Hence, because the fixation probabilities in birth-death and death-birth dynamics depend uniquely on the γ_i's (see Eq. 1.7), they differ in the two dynamics except in the limit $N \to \infty$. All fixation times are function both of the γ_i's and of the transition probabilities themselves (see Appendix A), so their value is independent of the dynamics only in the

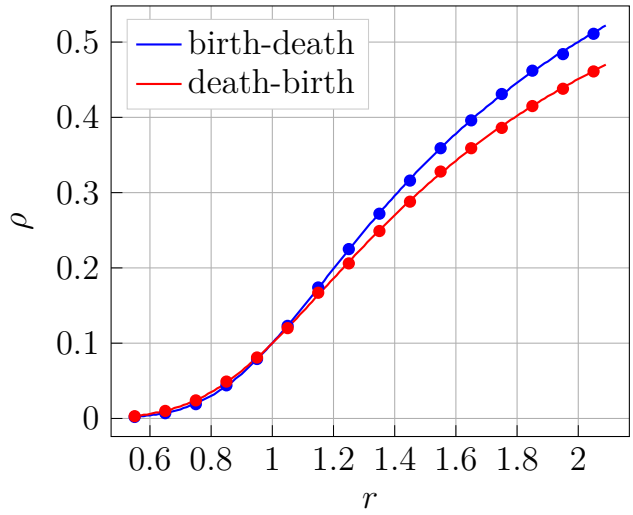


Figure 1.5: Fixation probability ρ versus mutant fitness r in a Moran process where the same individual cannot be chosen for both death and reproduction at the same time step. $N = 10$, 10^5 samples per data point.

infinite population limit. The theoretical predictions and the results from the simulations for the fixation probability are shown in Fig. 1.5 for both birth-death and death-birth dynamics. Note that, as expected from Eq. 1.13, for $r = 1$ the two fixation probabilities in birth-death and death-birth dynamics are equal to $1/N$.

Chapter 2

Evolutionary dynamics on graphs

The Moran model and the variant without self-replacement discussed above do not take into account the structure of a population, namely, the fact that not all interactions between individuals are equivalent, which is almost always the case in real microbial populations.

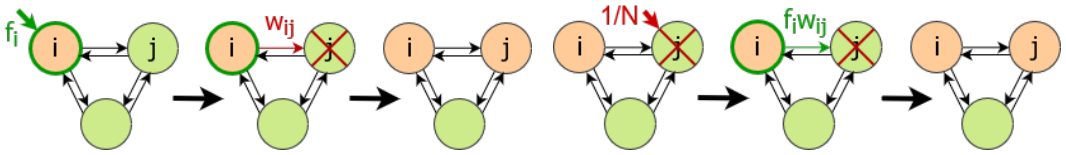
In order to introduce structure, in Ref. [5] each individual is placed on a node of a bi-directed graph of nodes $i = 1, 2, \ldots, N$ defined by its matrix of weights W (positive semi-definite). The nodes i and j can be connected by two edges: one going from i to j and the other from j to i. If $w_{ij} > 0$ there is an edge from node i to j, while if $w_{ij} = 0$ there is no such edge.

Fig. 2.1a shows one step of birth-death dynamics on a graph: an individual i is selected to reproduce with probability proportional to its fitness; its offspring replaces an individual j with probability w_{ij}. As in the Moran process, at each time step the population size is kept constant and, since after reproduction the offspring has to replace another individual, we have the following normalization constraint over w_{ij}:

$$\sum_{j=1}^{N} w_{ij} = 1 \tag{2.1}$$

namely, W is a right stochastic matrix.

Death-birth dynamics is shown in Fig. 2.1b: at each time step an individual j is chosen uniformly at random to die and an individual i is chosen to reproduce with probability proportional to the product between w_{ij} and its fitness; the offspring of i replaces j so to keep the size of the population constant.



(a) birth-death dynamics. (b) death-birth dynamics.

Figure 2.1: Schematic of one step of the dynamics for a structured population. Here, f_i indicates the fitness of the individual at node i.

In the following we will denoted by ρ_{m} the fixation probability of the standard Moran process: $\rho_{\mathsf{m}} \doteq \frac{1-1/r}{1-1/r^N}$. Moreover, in this chapter, sentences like "node i is replaced" or "node i becomes mutant" may be used for simplicity, but they always refer to the individual placed at node i; they should be read as "the individual at node i is replaced" and "the individual at node i is replaced by a mutant individual", respectively.

2.1 Some results for birth-death dynamics

In this section we summarize some results of evolutionary theory on graphs in the context of birth-death dynamics presented in Ref. [5].

2.1.1 Isothermal graphs

To start with, we give a couple of definitions which will be useful in the following: the first one introduces the concept of *temperature of a node* as a measure of how often a node is replaced, the second one defines what it means for a graph to be *ρ-equivalent to the Moran process*.

Definition 2.1.1. Temperature of a node
The temperature of a node j is the sum of the weights of the edges entering j:

$$T_j \doteq \sum_{j=1}^{N} w_{ij}.$$

In the neutral case (when all the individuals are wildtype) the temperature of a node is proportional to the probability with which the node is replaced: a node is "hot" if it is often replaced and it is "cold" if it is replaced rarely.

Definition 2.1.2. ρ-equivalence to the Moran process
An evolutionary graph G is said to be ρ-equivalent to the Moran process if

the fixation probability of one mutant of fitness r is

$$\rho = \frac{1 - r^{-1}}{1 - r^{-N}},$$

equal to the fixation probability ρ_{m} within the Moran process.

So, asking in what kinds of graphs a mutant has the same probability to fix as in a well-mixed population is the same as asking what graphs are ρ-equivalent to the Moran process, and the answer is given by the following theorem:

Theorem 2.1.1. Isothermal Theorem
A graph G is ρ-equivalent to the Moran process if and only if G is isothermal (all nodes have the same temperature).

The proof of this theorem is given in Ref. [9]. Notice that

$$\frac{1}{N}\sum_{j=1}^{N} T_j = \frac{1}{N}\sum_{i,j=1}^{N} w_{ij} = \frac{1}{N}\sum_{i=1}^{N} 1 = 1, \tag{2.2}$$

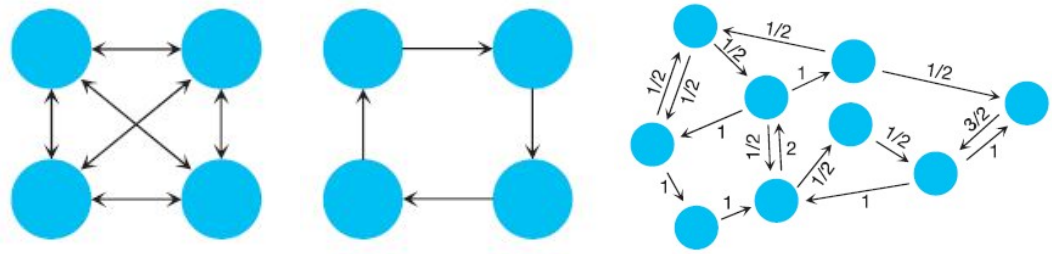
where we have used Eq. 2.1 in the second equivalence. Hence, if G is isothermal then $\forall j\ T_j = 1$ which means that W is doubly stochastic (both rows and columns sum to 1). On the other hand, if W is doubly stochastic then G is isothermal. Hence, the following statements are equivalent:

1. G is ρ-equivalent to the Moran process;

2. G is isothermal;

3. W is doubly stochastic.

Some examples of isothermal graphs are shown in Fig. 2.2. Any symmetric graph is isothermal since $\forall j\ T_j = \sum_{j=1}^{N} w_{ij} = \sum_{j=1}^{N} w_{ji} = 1$, but non-symmetric graphs can also be isothermal, for example the cycle (Fig. 2.2b). Fig. 2.2c shows a generic isothermal graph. The standard Moran model is equivalent to a complete graph with equal weights and self-loops, such as the one shown in Fig. 2.2a (self-loops are not shown), where each edge has weight $1/N$ (being N the total number of nodes). All graphs which are not isothermal have fixation probabilities different from the Moran process.



(a) (b) (c)

Figure 2.2: Examples of isothermal graphs (Ref. [5]): (a) complete graph with equal weights, equivalent to a standard Moran process (self-loops are not shown but they are present); (b) cycle; (c) a more general example of isothermal graph. Whenever the weights of the edges are not shown, each edge departing a given vertex has weight $1/d$, being d the outer degree of that vertex (the outer degree is defined as the number of edges exiting the node).

2.1.2 Suppressors and amplifiers of selection

A class of graphs which are very different from isothermal graphs in terms of fixation probability are the one-rooted graphs. One example is the line, shown in Fig. 2.3 and defined by

$$w_{ij} = \begin{cases} 1 & \text{if } j = i+1 \text{ and } i \neq N \\ 0 & \text{otherwise} \end{cases}$$

where N is the total number of nodes. The node labelled by 0 is called *root*. A root is a node that has no edge leading into it. In one-rooted graphs, the

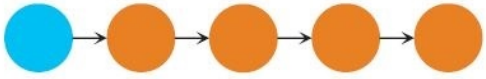


Figure 2.3: An example of one-rooted graph (Ref. [5]): the line.

mutant will fix if and only if it happens to be placed at the root, and this happens with probability $1/N$. Hence, the fixation probability is $\rho = 1/N$ regardless of r, which corresponds to the neutral case of the Moran process. So, selection is completely eliminated and only the stochastic drift, due to

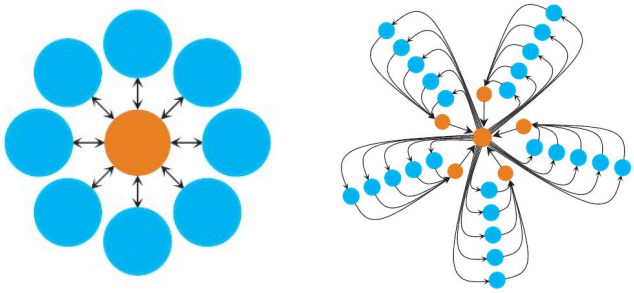
the randomness in the choice of the individuals for reproduction and death, counts.

In general, we can define graphs which suppress selection, even if not completely:

Definition 2.1.3. Suppressor of selection
A graph is a suppressor of selection if $\rho(r) < \rho_{\mathsf{m}}(r)$ for $r > 1$ and $\rho(r) > \rho_{\mathsf{m}}(r)$ for $r < 1$.

Graphs made up of an up-stream population and a down-stream one also suppress selection (Ref. [9]). Conversely, there are graphs which amplify selection:

Definition 2.1.4. Amplifier of selection. A graph is an amplifier of selection if $\rho(r) > \rho_{\mathsf{m}}(r)$ for $r > 1$ and $\rho(r) < \rho_{\mathsf{m}}(r)$ for $r < 1$.



(a) (b)

Figure 2.4: Examples of amplifiers of selection (Ref. [5]). (a) star; (b) superstar of parameter $k = 3$. The orange and blue colours indicate respectively warm and cold nodes.

An example of amplifier of selection is the *star*: a star with N nodes is composed of a central node, the centre, and of $N - 1$ peripheral nodes, the leaves. Its structure is shown in Fig 2.4a: each edge leading from the centre to a leaf has weight $\frac{1}{N-1}$ and each edge leading from a leaf to the centre has weight 1, so to respect the normalization constraint (Eq. 2.1) and the symmetry of the star, in which all the leaves are equivalent. For N large the fixation probability of a mutant of relative fitness r is

$$\rho(r) = \frac{1 - r^{-2}}{1 - r^{-2N}}. \tag{2.3}$$

Therefore, a mutant of fitness r on the star has the fixation probability of a mutant of fitness r^2 in a Moran process (Eq. 1.12): an advantageous mutant on the star is equivalent to a fitter mutant in a Moran process, while a

disadvantageous mutant on the star is equivalent to a less fit mutant in a Moran process. So the star is an amplifier of selection.

Another example of amplifier of selection is the *superstar of parameter* k. Fig. 2.4b shows a superstar with a centre and a number of peripheral structures (here, 5), each of them characterized by some reservoir nodes (in blue) and $k-2$ chain nodes (in orange). In the limit of a large number of peripheral structures and of a large number number of reservoir nodes per peripheral structure, the fixation probability for a mutant of relative fitness r on a superstar of parameter k is

$$\rho(r) = \frac{1 - r^{-k}}{1 - r^{-kN}}, \tag{2.4}$$

which means that we can build arbitrarily strong amplifiers of selection with superstars by increasing k.

2.2 The star

In this section we will study the evolutionary dynamics on the star since it presents interesting amplifying properties while being a rather simple structure. In Fig. 2.5 the fixation probability ρ of a star is plotted versus the mutant fitness r for both birth-death and death-birth dynamics: in the case of birth-death dynamics, the star behaves as an amplifier of selection; contrarily, in the death-birth dynamics it behaves as a suppressor of selection. Some details for the numerical simulations of this chapter can be found in Appendix B.1.

2.2.1 Birth-death dynamics

Let's call P the number of leaves of a star; then each edge leading from the centre to a leaf has weight $1/P$ and each edge leading from a leaf to the centre has weight 1. The smallest interesting example is the star with $P = 2$. In this simple case, we compute exactly the fixation probability by numerically finding the eigenvectors of the matrix of transition probabilities between the different states of the system.

Contrarily to the Moran process, where each state is fully defined by the number of mutants present in the population, for a general graph one needs to specify the exact subset of nodes at which mutants individuals are located in order to define a state. Because each node can hold a mutant or a wildtype

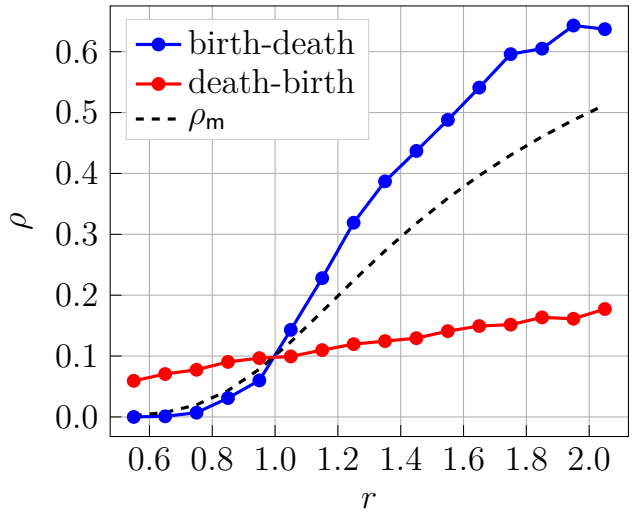


Figure 2.5: Fixation probability ρ vs mutant fitness r for a star with birth-death dynamics (blue) and death-birth dynamics (red). $P = 9$, 10^4 samples per data point.

individual (and there are $P+1$ nodes), the total number of states is $S \doteq 2^{P+1}$. Exact calculations of the fixation probability are possible also for $P > 2$, but they become rapidly heavier since the number of states S grows exponentially with P.

Let us number the states from 1 to S and let state 1 corresponds to only wildtypes and state S to only mutants. By calling τ_{ij} the transition probability from state i to state j we can write the following equation for the fixation probability ρ_i when starting at state i:

$$\rho_i = \sum_{j=1}^{S} \tau_{ij}\rho_j \tag{2.5}$$

or, equivalently, in matrix form

$$\boldsymbol{\rho} = \mathrm{T}\boldsymbol{\rho}, \tag{2.6}$$

where T is the $S \times S$ matrix of elements τ_{ij}. As in the case of the Moran process, both the probabilities for the wildtype to fix and for the mutant to fix satisfy Eq. 2.5 (and hence Eq. 2.6) but with different boundary conditions: for the mutant we have that $\rho_0 = 0$ and $\rho_S = 1$, while for the wildtype we have that $\tilde{\rho}_0 = 1$ and $\tilde{\rho}_S = 0$. Since we are interested in the fixation probability of the mutant, we consider the elements of $\boldsymbol{\rho}$ corresponding to the initial state with the mutant occupying the centre and to the two initial states with the mutant occupying one of the two leaves. Because of the symmetry of the star, the last two initial states hold the same fixation probability.

The two solutions of Eq. 2.6, with the mutant starting on the centre ρ^{centre} and with the mutant starting on a leaf ρ^{leaf}, are plotted in Fig. 2.6 with

a blue and red line, respectively, in the case of birth-death dynamics, and they are in very good agreement with the simulation results (represented by markers). The line coloured in purple shows the linear combination of the two solutions corresponding to a mutant placed uniformly at random on the graph: $\rho^{\mathsf{unif}} = 1/3\,\rho^{\mathsf{centre}} + 2/3\,\rho^{\mathsf{leaf}}$. The amplifying effect of the star with respect to the Moran process is visible even if small: indeed, $\rho^{\mathsf{unif}} > \rho_{\mathsf{m}}$ for $r > 1$, while $\rho^{\mathsf{unif}} < \rho_{\mathsf{m}}$ for $r < 1$. A stronger amplification effect is expected for stars of larger size.

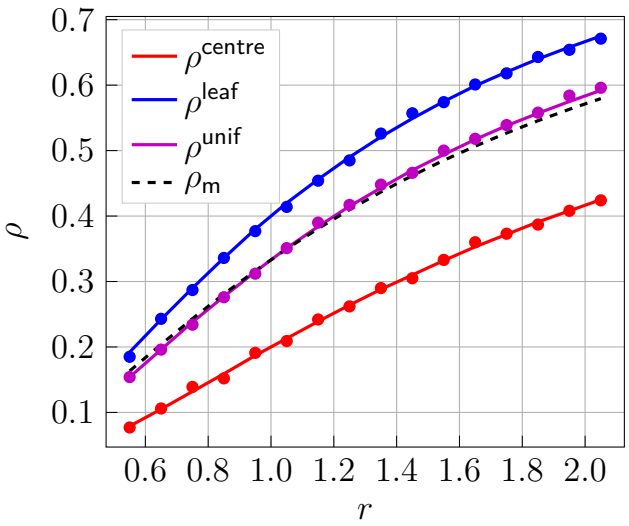


Figure 2.6: Fixation probability ρ versus mutant fitness r for a small star with birth-death dynamics: the solutions of Eq. 2.6 are shown by solid lines and simulation results by markers. The lines in colours correspond to the fixation probability for the star with different initial conditions: in blue, the mutant is placed on a leaf; in red, the mutant is placed on the centre; in purple, the mutant is placed uniformly at random. The black dashed line corresponds to the fixation probability in a Moran process with $N = 3$ individuals. $P = 2$, 10^4 samples per data point.

Fig. 2.6 shows that the fixation probability depends strongly on the initial conditions. In particular, a mutant appearing on a leaf is favoured with respect to a mutant having the same fitness but appearing at the centre. The reason is that the centre is replaced more often than a leaf, i.e. it has a higher temperature. Indeed, using Def. 2.1.1, the temperature of the centre is $T_c = P$ and of a leaf is $T_l = 1/P$.

Realistic initial conditions

Given the importance of the initial conditions on the evolutionary outcome (i.e. what type fixes), it is worth asking what initial conditions are more realistic in a biological sense, namely, which probability distribution should be chosen for the initial position of a new mutant on a general graph.

Uniform initial conditions seem sensible for mutations that arise spontaneously (e.g. due to exposure to radiation), but in nature mutations often arise at division due to errors in DNA replication. In order to model this situation, in Ref. [10] it is proposed to set a mutant on node j with the probability that an individual on node j is replaced. The replacement (death event) arises because of a reproduction event at a neighbouring node. The probability for node j to be replaced (death probability) when all individuals have the same fitness is

$$\sum_{i=1}^{N} \frac{1}{N} w_{ij} = \frac{T_j}{N}. \tag{2.7}$$

since $1/N$ is the probability that individual i reproduces and w_{ij} is the probability that the progeny of i replaces j. Placing the mutant with a probability equal to the replacement probability of a node it is called "temperature initialization":

$$\mathsf{temp} \doteq \frac{1}{N} (T_1, \dots T_N) ; \tag{2.8}$$

it is normalized as shown by Eq. 2.2 and it models the appearance of mutations only if upon division the mutant daughter cell is always the one which migrates.

A more realistic assumption would be that, upon division where a mutation occurs, the mutant cell and the wildtype daughter cells have equal probabilities to stay or to migrate. Because before the mutant appears, all the individuals are wildtype, each of them reproduces with the same probability ($1/N$ if N is the total number of individuals) since there is no fitness difference, so birth is uniformly distributed. At a given node, there is a uniform probability of birth occurring and therefore of a native mutant appearing, plus there is a probability of a mutant born elsewhere to migrate there. The latter is proportional to the temperature of the node considered, which corresponds to the temperature initialization. Hence, the new probability distribution, that we call "mixed", it is just a combination with equal weights of uniform and temperature:

$$\mathsf{mixed} = \frac{1}{2}\mathsf{unif} + \frac{1}{2}\mathsf{temp}, \tag{2.9}$$

where temp is given by Eq. 2.8 and $\mathsf{unif} \doteq \frac{1}{N}(1,\dots,1)$. Consequently,

$$\rho^{\mathsf{mixed}} = \frac{1}{2}\rho^{\mathsf{unif}} + \frac{1}{2}\rho^{\mathsf{temp}}. \tag{2.10}$$

where $\rho^{\mathsf{unif}}, \rho^{\mathsf{temp}}$ and ρ^{mixed} denote the fixation probabilities respectively for uniform, temperature and mixed initial conditions.

Adding self-loops

Still following Ref. [10], we can add self-loops to the star so that the same individual can divide and replace itself; indeed, there is no reason for this event to be forbidden. Moreover, by changing the weight of the self-loops we are able to change the temperature of the nodes. In order to do so, we introduce two parameters x and y ($0 < x, y \leq 1$) such that $1-y$ is the weight of the self-loop of the centre and $1-x$ is the weight of the self-loops on each leaf. The other weights are chosen in order to respect the symmetry of the star and for W to be right stochastic:

$$W = \begin{pmatrix} 1-y & y/P & y/P & \cdots & y/P & y/P \\ x & 1-x & 0 & \cdots & 0 & 0 \\ x & 0 & 1-x & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x & 0 & 0 & \cdots & 1-x & 0 \\ x & 0 & 0 & \cdots & 0 & 1-x \end{pmatrix}. \tag{2.11}$$

The case $x = y = 1$ corresponds to the star without self-loops that we have discusses previously. The temperature of the centre T_c and the temperature of each leaf T_l are functions of x and y:

$$T_c = 1 - y + Px \tag{2.12a}$$
$$T_l = 1 - x + y/P \tag{2.12b}$$

where we have used Def. 2.1.1.

In the top left panel of Fig. 2.7, the results of the simulations are shown for $x = y = 1$ (no loops) for uniform (blue) and temperature (red) initial conditions. Under uniform initial conditions, the star amplifies selection with respect to the unstructured population (black dashed line). Temperature initialization leads to a very different result, due to the large temperature difference between the centre and the leaves. In this case, the selection effect is suppressed for advantageous mutants ($r > 1$) with respect to the

Moran process. In the plot at the top right, the results for $x = 1/P$ and $y = 1/P^2$ show that by mitigating the temperature difference between the centre and the leaves it is possible to obtain amplification for both initial conditions (Ref. [10]). By choosing $y = Px$ we get an isothermal star, such that $T_c = T_l = 1$, so that uniform and temperature initialization coincide. As shown at the bottom of Fig. 2.7 and in agreement with the isothermal theorem (Th. 2.1.1), we then recover the fixation probability ρ_{m} of an unstructured population.

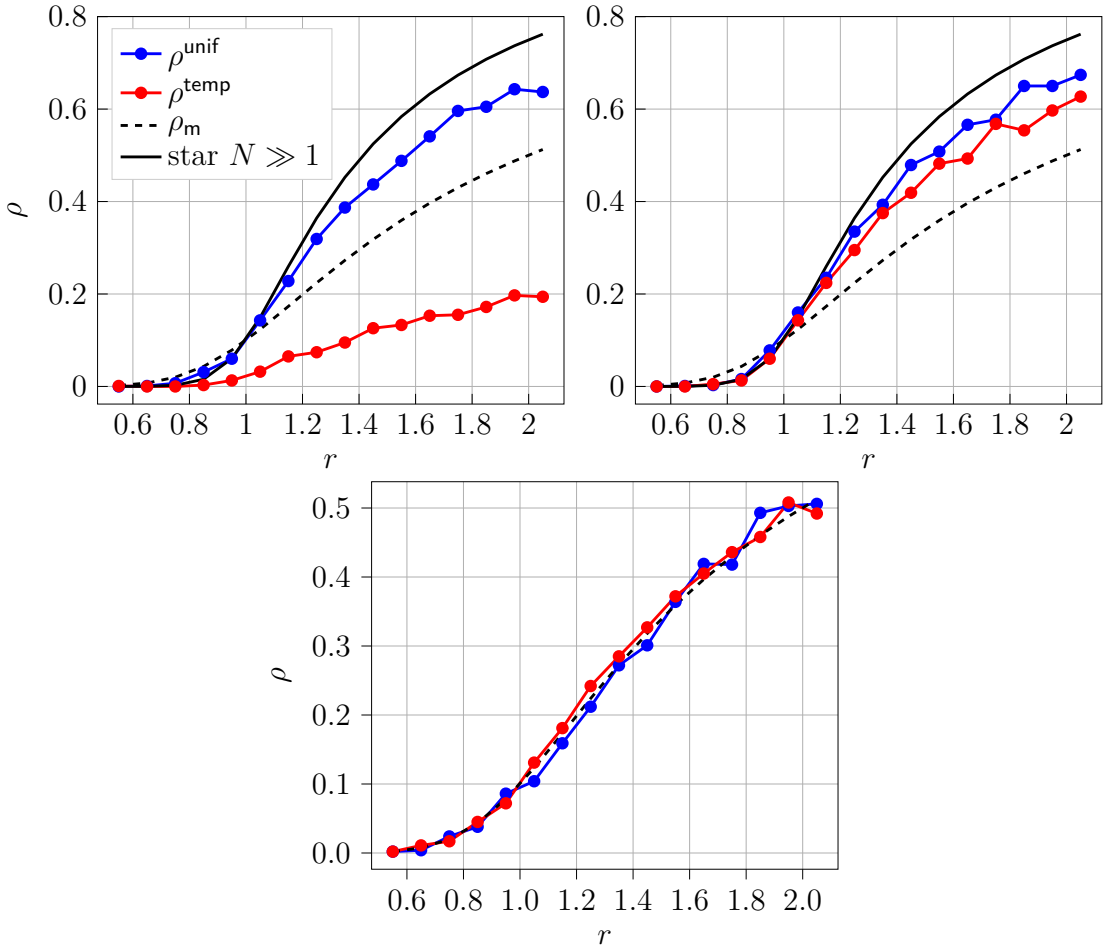


Figure 2.7: Fixation probability ρ versus mutant fitness r for a star with birth-death dynamics. Top left: a star with $x = y = 1$; top right: $x = 1/P$ and $y = 1/P^2$; bottom: $y = Px$. Simulation results for uniform initial conditions in blue and temperature initial conditions in red. In full black line the asymptotic fixation probability for a star with a large number of leaves (Eq. 2.3); in dashed line the fixation probability ρ_{m} of an unstructured population of $N = 10$ individuals. $P = 9$, 10^3 samples per data point.

In order to further study how the amplifying property of the star changes depending on the temperature of the nodes, we have performed some simulations using different values for x in (0,1) while keeping y fixed. In the top row of Fig. 2.8 the results are shown in the case of uniform initial conditions: by decreasing the value of x the temperature of the centre is decreased, while the temperature of the leaves is increased according to Eqs. 2.12. As the temperature difference $\Delta T \doteq T_c - T_l$ is reduced, the star gradually loses its amplifying property and, as the centre becomes colder than the leaves, the star becomes a suppressor of selection. Note that for $\Delta T = 0$ we recover the probability ρ_{m} of an unstructured population within the Moran model (Eq. 1.12), as expected from the isothermal theorem (Th. 2.1.1).

The results for the temperature initialization are shown in Fig. 2.8, middle row. We see two regimes: the graph on the left shows that, for large temperature differences, as ΔT decreases the fixation probability increases; however, as ΔT becomes sufficiently small (graph on the right), the results are analogous to the ones with uniform initialization (as the centre becomes colder than the leaves the star becomes a suppressor of selection).

In the bottom row, the results for mixed initial conditions (Eq. 2.9) are shown. The behaviour of the fixation probability is qualitatively similar to both uniform and temperature initial conditions for ΔT close to zero (on the right); for large ΔT the fixation probability increases as the temperature of the centre decreases, as for temperature initialization, but the variation is milder due to the uniform contribution to the probability.

In Fig. 2.9 we have plotted the fixation probability versus ΔT for the three types of initial conditions. The plots are obtained by varying x in (0,1) at fixed y and r. It can be noticed that mixed initialization gives less extreme results with respect to uniform and temperature, e.g. for large ΔT it does not imply very small value of fixation probability as temperature initialization does.

In Fig. 2.10 we have plotted heatmaps of the fixation probability versus x and y for a fixed value of r, showing that the fixation probability does not depend only on the temperature difference ΔT, but it depends on the values of both x and y which determine the matrix of weights W (Eq. 2.11).

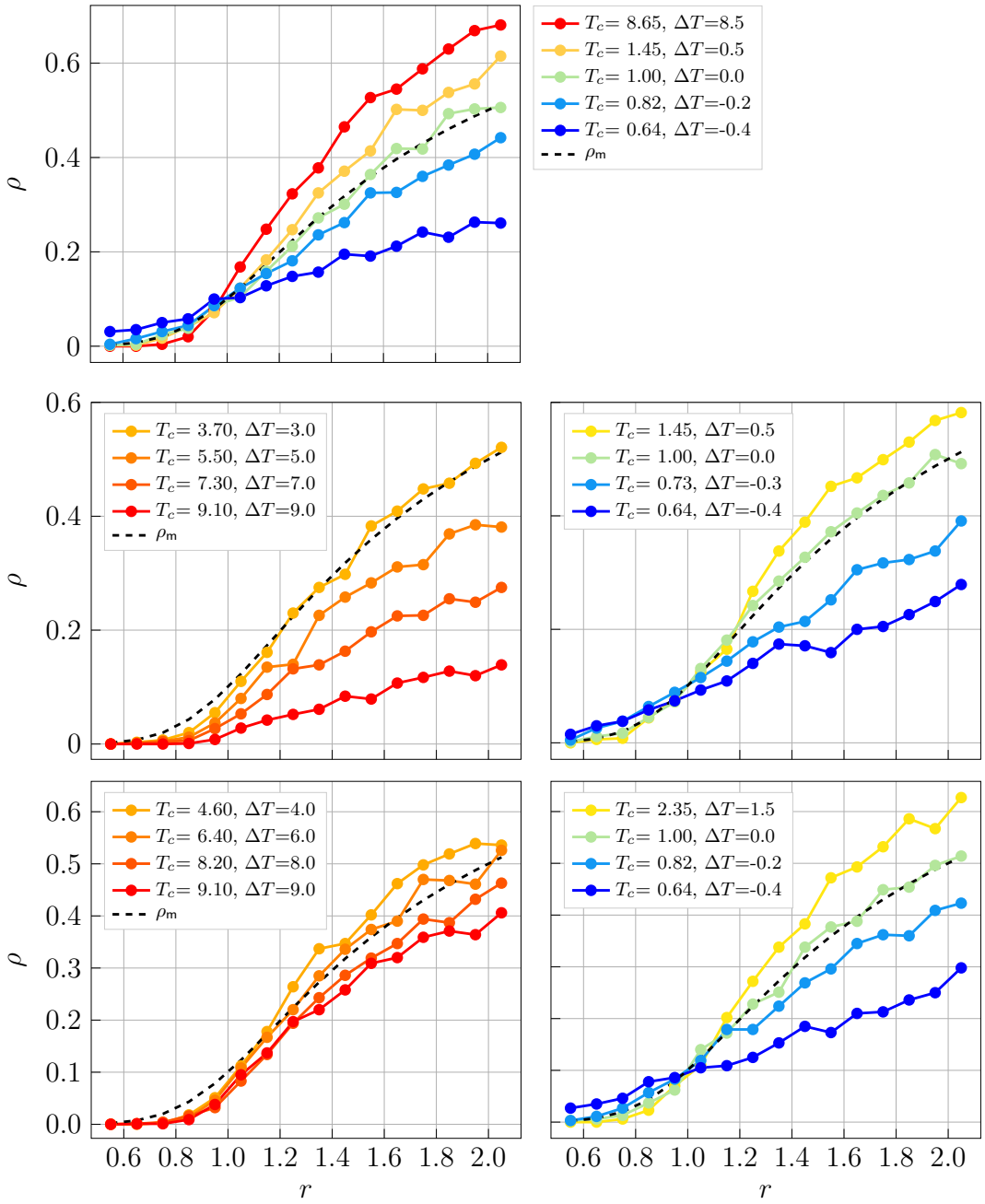


Figure 2.8: Fixation probability ρ vs mutant fitness r for a star with birth-death dynamics: simulation results for different values of the temperature T_c of the centre, obtained by varying x in (0,1), while y is kept fixed. Top row: the results for uniform initial conditions; middle row: the results for temperature initial conditions; bottom row: mixed initialization. $P = 9$, $y = 0.45$, 10^3 samples per data point.

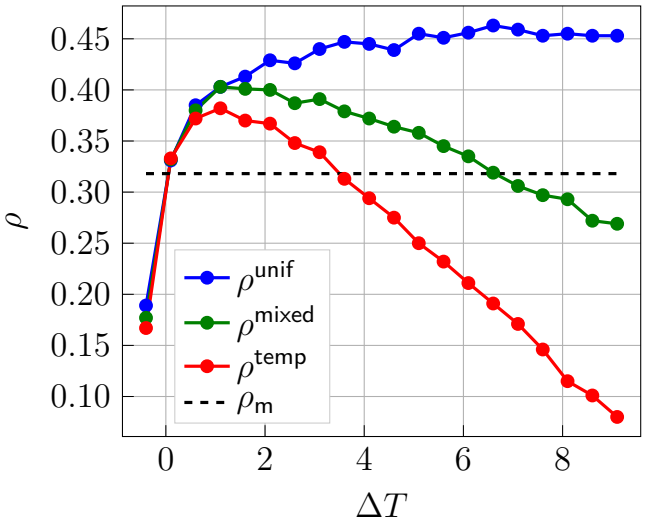


Figure 2.9: Fixation probability ρ versus the temperature difference ΔT between the centre and a leaf, for a star with birth-death dynamics. The value of ΔT is changed by varying x in (0,1) and keeping y fixed, in the case of an advantageous mutant. $P = 9$, $r = 1.45$, $y = 0.45$, 10^4 samples per data point.

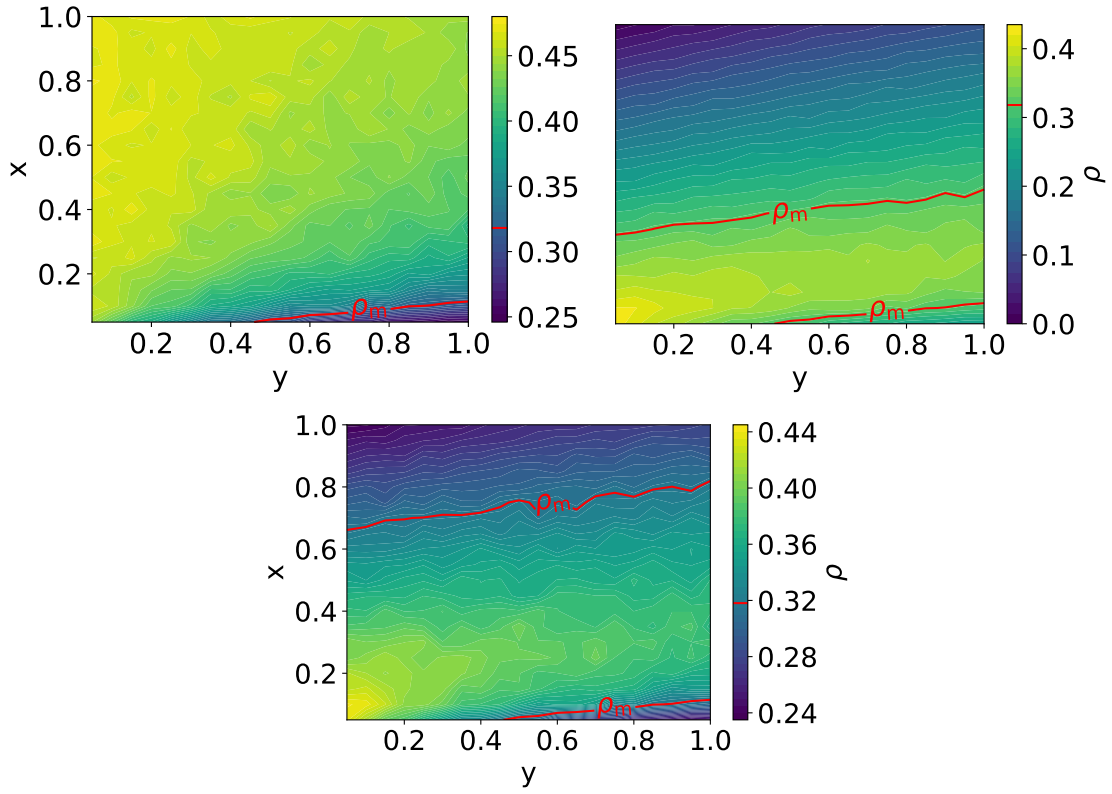


Figure 2.10: Heatmaps of the fixation probability ρ as function of x and y for a star with birth-death dynamics. Top left: uniform initial conditions; top right: temperature initial conditions; bottom: mixed initial conditions. Red lines indicate the probability ρ_{m} of an unstructured population within the Moran model (Eq. 1.12). $P = 9$, $r = 1.45$, 10^5 samples per data point.

2.2.2 Death-birth dynamics

In the previous section we dealt with birth-death dynamics; let us now turn to the alternative choice of death-birth dynamics and study the impact of this choice on the results.
As mentioned at the beginning of this chapter, in this case it is the individual j who dies which is chosen first, uniformly at random, and then the individual i who reproduces is chosen proportionally to the product between w_{ij} and its fitness. Note that there is some freedom in the choice of W, since the probability for an individual at node i to replace another individual at node j, given that this latter dies, is given by $f_i w_{ij}$, where f_i is the fitness of the individual at i. So, contrarily to birth-death, it also depends on the fitness. Hence, there is not a normalization constraint on the w_{ij} which is valid along the whole simulation.

We can choose W to be such that w_{ij} is actually the probability for i to replace j, given that j dies, *in the neutral case* (when the population is fully wildtype). This requirement is equivalent to ask that W is left stochastic ($\sum_{i=1}^{N} w_{ij} = 1$), contrarily to birth-death where the natural choice for W is to be right stochastic (Eq. 2.1). This choice for W is often made in the literature (Ref. [6]), but it has no particular significance and other choices may be justified.

We can check whether all these choices are equivalent or not by looking at the transition probabilities of the process. Let's introduce a variable n_k at each node such that $n_k = 1$ if the node is mutant and $n_k = 0$ if it is wildtype. The probability T_i^+ for node i to become mutant given that it was wildtype reads

$$T_i^+ = \frac{1}{P+1} \frac{\sum_k w_{ki} n_k (1+s)}{\sum_j w_{ji}(1+sn_j)} \tag{2.13}$$

since $1/(P+1)$ is the probability that the individual at i dies and the second factor is the probability that a mutant individual reproduces and replaces i. The only constraints that W must satisfy are the ones imposed by the symmetry of the star: for any $k, j \neq 0$ $w_{0j} = w_{0k}$, $w_{j0} = w_{k0}$, $w_{jk} = 0$, being the centre labeled by 0. In order to remain general, we consider the possibility of self-loops of weight w_{00} on the centre and w_{ii} independent of i on each leaf. Hence, if i is a leaf,

$$T_i^+ = \frac{1}{P+1} \frac{(w_{0i} n_0 + w_{ii} n_i)(1+s)}{w_{0i}(1+sn_0) + w_{ii}(1+sn_i)}; \tag{2.14}$$

if instead i is the centre ($i = 0$):

$$T_0^+ = \frac{1}{P+1}\frac{\sum_k w_{k0} n_k (1+s)}{\sum_j w_{j0}(1+sn_j)}. \tag{2.15}$$

In both cases, T_i^+ depends on the actual weight of the edges, and an analogous result holds for the probability T_i^- for the individual at i to become wildtype given that it was mutant, which is given by

$$T_i^- = \frac{1}{P+1}\frac{\sum_k w_{ki}(1-n_k)}{\sum_j w_{ji}(1+sn_j)}. \tag{2.16}$$

For a star without self-loops ($w_{00} = w_{ii} = 0$) we have

$$T_i^+ = \frac{1}{P+1}\frac{w_{0i} n_0 (1+s)}{w_{0i}(1+sn_0)} = \begin{cases} 0 & \text{if } n_0 = 0 \\ 1/(P+1) & \text{if } n_0 = 1; \end{cases} \tag{2.17}$$

if i is a leaf and

$$T_0^+ = \frac{1}{P+1}\frac{\sum_{k\neq 0} w_{k0} n_k (1+s)}{\sum_{j\neq 0} w_{j0}(1+sn_j)} = \frac{1+s}{P+1}\frac{\sum_k n_k}{\sum_j (1+sn_j)} \tag{2.18}$$

if i is the centre, which are both independents of the actual values of the w_{ij} (the same result holds for T_i^-). Hence, for the star without self-loops, death-birth dynamics is well-defined, since it does not depend on the choice of W (within the constraints imposed by the symmetry). On the contrary, if we allow self-loops the fixation probability depends on the choice of W and there is not an unique choice for it, making the model not well-defined.

In the present work, we will use a left stochastic matrix of weights, consistently with Ref. [6], so that W is defined as

$$W \doteq \begin{pmatrix} 1-y & x & x & \cdots & x & x \\ y/P & 1-x & 0 & \cdots & 0 & 0 \\ y/P & 0 & 1-x & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ y/P & 0 & 0 & \cdots & 1-x & 0 \\ y/P & 0 & 0 & \cdots & 0 & 1-x \end{pmatrix} \tag{2.19}$$

namely, as the transpose of the matrix used in birth-death dynamics. This expression for W arises from the requirement that W is left stochastic and from the symmetries of the star graph.

The numerical and analytical results (obtained along the same lines of Sec. 2.2) for a small star with $P = 2$ leaves are shown in the left panel of Fig. 2.11: in contrast with birth-death, under uniform initialization the star is a suppressor of selection. Moreover, here, a mutant located at the centre is more likely to fix than one located at a leaf, while the opposite result was obtained for birth-death dynamics (Fig. 2.6).

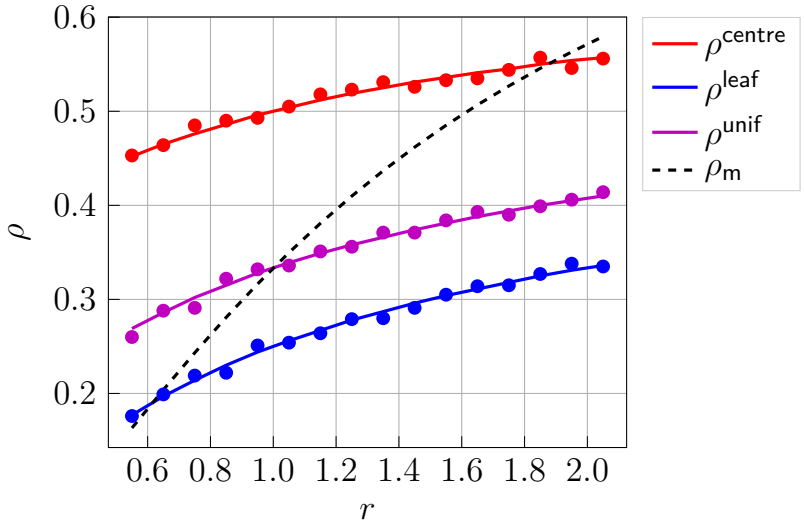


Figure 2.11: Fixation probability ρ versus mutant fitness r for a small star with death-birth dynamics: the solutions of Eq. 2.6 are shown by solid lines and simulation results by markers. The lines in colours correspond to the fixation probability for the star with different initial conditions: in blue, the mutant is placed on a leaf; in red, the mutant is placed on the centre; in purple, the mutant is placed uniformly at random. The black dashed line corresponds to the fixation probability in a Moran process with $N = 3$ individuals. $P = 2$, 10^4 samples per data point.

In partial analogy with what has been done for birth-death, we could define a new temperature in the death-birth framework as

$$T_i \doteq \sum_{j=1}^{P+1} w_{ij} \tag{2.20}$$

such that the reproduction probability of node i in a fully wildtype population is $T_i/(P+1)$. For the same values of x and y the temperature of each node in the death-birth framework is the same as the one defined in the birth-death framework (Eqs. 2.12), hence the centre is again "hotter" than the leaves,

and the star is isothermal if and only if $y = Px$. But here, if a node is "hot", it means that an individual placed on it replaces others (i.e. reproduces) with higher probability than an individual placed on a "cold" node (at least in the neutral situation where all individuals are wildtype), resulting into a higher probability to fix, which explain why ρ^{centre} leads to higher fixation probability than ρ^{leaf} for $r = 1$ as shown in Fig. 2.11. Conversely, in birth-death dynamics we recall that a "hot" node is often replaced. In general, with death-birth dynamics and without assuming that all individuals are wildtype, the reproduction probability of individual i is

$$\frac{1}{P+1}\sum_{j=1}^{P+1}\frac{f_i w_{ij}}{\sum_{k=1}^{N} f_k w_{kj}}$$

where f_j is the relative fitness of individual at node j, so it depends on all fitnesses in the population.
As for birth-death dynamics, we have plotted the numerical results for the fixation probability versus the fitness for different temperatures of the nodes, by keeping y fixed and varying x. The results are shown in Fig. 2.12, where the mixed initial conditions are defined as for birth-death (Eq. 2.9). In the case of uniform initial conditions, as the temperature of the centre decreases, the star changes from suppressor to amplifier of selection, contrarily to what happens for birth-death dynamics (see Fig. 2.8, top row). In the case of temperature initial conditions, for ΔT very large the fixation probability is large for both $r > 1$ and $r < 1$ with respect to a well-mixed population, hence also disadvantageous mutants have a high probability to fix; as ΔT decreases the fixation probability decreases; for small ΔT we recover results similar to the ones for uniform initialization. We notice that for an isothermal star we do not recover the fixation probability of an unstructured population, in contrast with birth-death dynamics for which the isothermal theorem (Th. 2.1.1) holds. Moreover, the behaviour of the fixation probability with respect to ΔT is strongly different from the birth-death case, as can be seen by comparing Fig. 2.9 and Fig. 2.13 where the fixation probability is plotted versus ΔT.

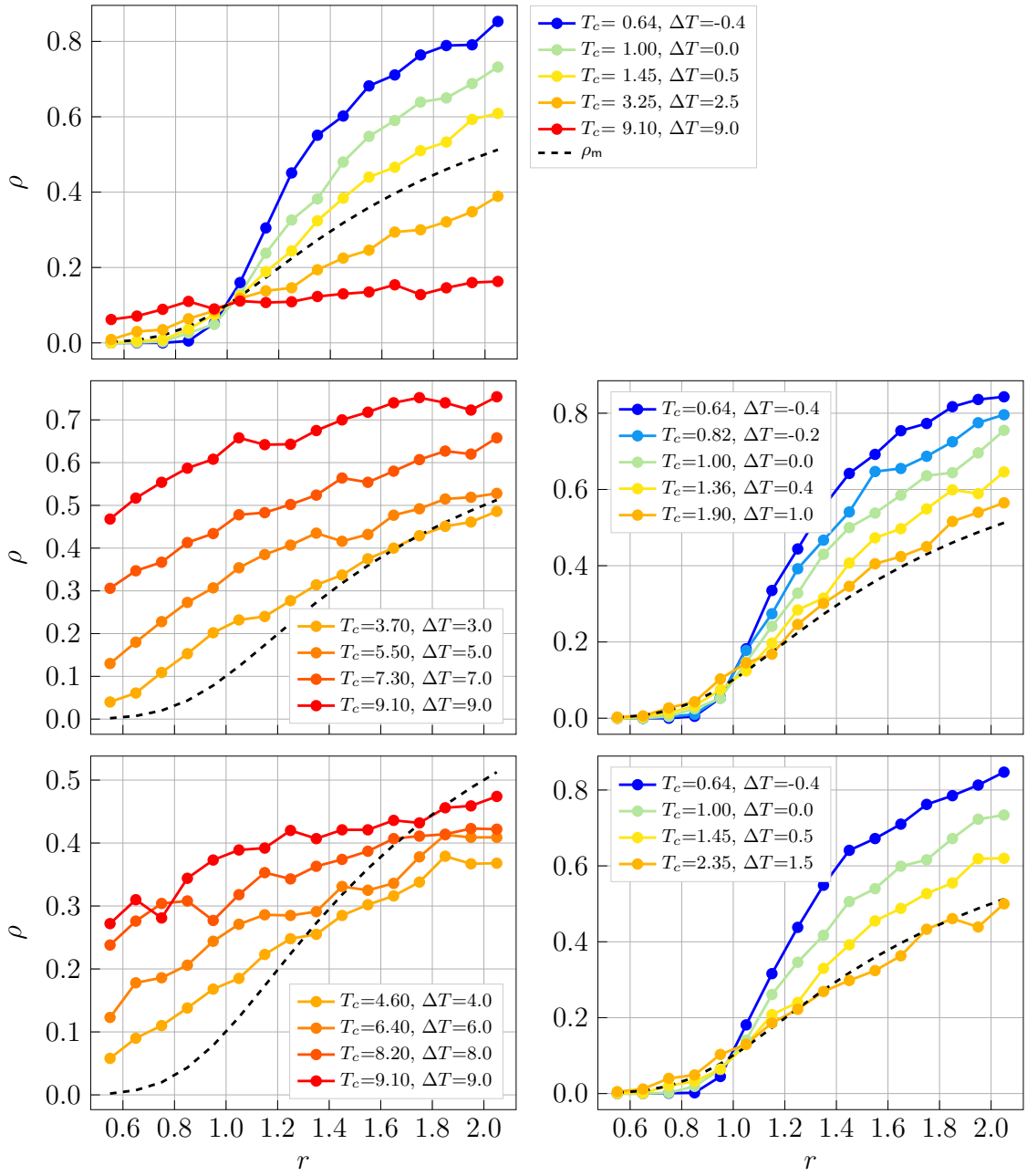


Figure 2.12: Fixation probability ρ vs mutant fitness r for a star with death-birth dynamics: simulation results for different values of the temperature T_c of the centre, obtained by varying x in (0,1), while y is kept fixed. Top row: the results for uniform initial conditions; middle row: the results for temperature initial conditions; bottom row: mixed initialization. $P = 9$, $y = 0.45$, 10^3 samples per data point.

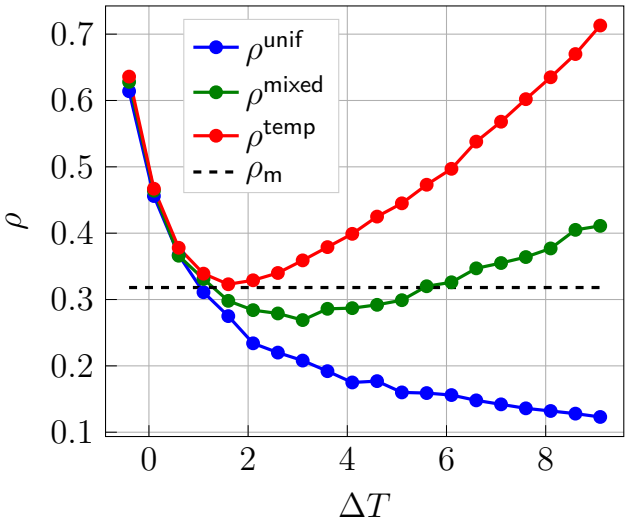


Figure 2.13: Fixation probability ρ versus the temperature difference ΔT between the centre and a leaf, for a star with death-birth dynamics. The value of ΔT is changed by varying x in (0,1) and keeping y fixed, in the case of an advantageous mutant. $P = 9$, $r = 1.45$, $y = 0.45$, 10^4 samples per data point.

Chapter 3

Coarse-graining the model of populations on graphs

A different model (Ref. [6]) with respect to the one presented in the previous chapter takes into account the fact that often in nature populations are divided into relatively well-mixed communities exchanging individuals via migration events. In this kind of model, there are M subpopulations, which we are called "islands", each island being labeled with a number from 1 to M. The total number of individuals on island i is denoted by N_i and the total number of mutant individuals on the same island is denoted by n_i. Each island occupies the node of a bi-directed graph with weights m_{ij}. Hence, this model can be viewed as a coarse-grained version of the previous one, with subpopulations on graphs instead of single individuals on graphs. As before, anytime a death event occurs, a reproduction event also occurs or viceversa (depending on the details of the dynamics), and the newborn individual migrates to replace the individual who dies. Hence, the size of each subpopulation N_i is kept constant.

In birth-death dynamics (also called "invasion process" in Ref. [6]), an individual is chosen for reproduction among all the individuals of the population according to its fitness. Assuming that it belongs to island i, its offspring migrates to island k with probability m_{ik}, where it replaces an individual chosen uniformly at random among the N_k individuals there.

In death-birth dynamics (also called "voter model" in Ref. [6]), an individual is chosen uniformly at random in the entire population to die; assuming that it belongs to island k, with probability m_{ik} a migration event occurs from island i to k. Another option is that the migration event from i to k occurs with probability proportional to the product between m_{ik} and the

total fitness $N_i + sn_i$ of island i. The first choice considers fitness to be relevant only within each island, while the second one takes into account fitness across the islands. We will mainly focus on the second one since, for $N_i = 1\,\forall i$, it allows to recover the model we have described in Ch. 2. Finally, the reproducing individual on island i is chosen according to its fitness within the island. Some details for the numerical simulations of this chapter can be found in Appendix B.2.

In discrete time, we will call $T_i^+(\mathbf{n})$ the transition probability from n_i to $n_i + 1$ and $T_i^-(\mathbf{n})$ the transition probability from n_i to $n_i - 1$, which both depend on the complete state of the system $\mathbf{n} = (n_1, n_2, \ldots, n_M)$. We will also introduce a continuous time description (Ref. [6]) where we will call $W_i^+(\mathbf{n})$ the transition rate from n_i to $n_i + 1$ and $W_i^-(\mathbf{n})$ the transition rate from n_i to $n_i - 1$. In the following we will briefly summarize the derivation of the fixation probability as presented in Ref. [6] for a mutant individual of relative fitness $r = 1 + s$.

3.1 Fixed points of the probability generating function equation

Let $P(\mathbf{n}, t)$ be the probability to observe the system in state $\mathbf{n}$ at time t. Moreover, let the vector $a_i^{\pm}\mathbf{n}$ be the vector with components n_k for all $k \neq i$ and $n_i \pm 1$. The master equation reads

$$\begin{aligned}\frac{\partial P}{\partial t}(\mathbf{n}, t) = &\sum_{i=1}^{M} P(a_i^-\mathbf{n}, t) W_i^+(a_i^-\mathbf{n}) + P(a_i^+\mathbf{n}, t) W_i^-(a_i^+\mathbf{n}) \\ &- \sum_{i=1}^{M} P(\mathbf{n}, t) \left[W_i^+(\mathbf{n}) + W_i^-(\mathbf{n}) \right]\end{aligned} \tag{3.1}$$

3.1.1 Birth-death dynamics

Consider now the birth-death case. The transition probability $T_i^+(\mathbf{n})$ is given by

$$T_i^+(\mathbf{n}) = \underbrace{\frac{N_i - n_i}{N_i}}_{(2)} \sum_{k=1}^{M} \underbrace{m_{ki} \frac{n_k(1+s)}{\sum_j N_j + s\sum_j n_j}}_{(1)} \tag{3.2}$$

where (1) is the probability for a mutant to reproduce on island k and to migrate to i, which is then summed over all the islands k, and (2) is the

probability that, given that a death event occurs on island i (because an individual in i is being replaced), a wildtype individual dies. Analogously:

$$T_i^-(\mathbf{n}) = \frac{n_i}{N_i} \sum_{k=1}^{M} m_{ki} \frac{N_k - n_k}{\sum_j N_j + s \sum_j n_j}. \tag{3.3}$$

To describe the process in continuous time, in the framework of Eq. 3.1, we can consider a total event rate (reproduction rate) μ' proportional to the total fitness of the population, i.e. $\mu' = \mu \left(\sum_j N_j + s \sum_j n_j \right)$ where μ stands for the division rate of one wildtype individual. This yields:

$$W_i^+(\mathbf{n}) = \mu' \frac{N_i - n_i}{N_i} \sum_{k=1}^{M} m_{ki} \frac{n_k(1+s)}{\sum_j N_j + s \sum_j n_j} \tag{3.4a}$$

$$W_i^-(\mathbf{n}) = \mu' \frac{n_i}{N_i} \sum_{k=1}^{M} m_{ki} \frac{N_k - n_k}{\sum_j N_j + s \sum_j n_j} \tag{3.4b}$$

which allows to eliminate the nonlinearity due to the term $s \sum_j n_j$ at the denominator of Eqs. 3.2 and 3.3 so that, finally, we get

$$W_i^+(\mathbf{n}) = \mu(1+s) \frac{N_i - n_i}{N_i} \sum_{k=1}^{M} m_{ki} n_k \tag{3.5a}$$

$$W_i^-(\mathbf{n}) = \mu \frac{n_i}{N_i} \sum_{k=1}^{M} m_{ki} (N_k - n_k) \tag{3.5b}$$

Recall that in birth-death dynamics the m_{ki} satisfy the normalization constraint

$$\sum_{i=1}^{M} m_{ki} = 1. \tag{3.6}$$

Following Ref. [6], let us now derive an equation for the probability generating function (PGF) starting from Eq. 3.1. The PGF is defined as

$$\phi(\mathbf{z}, t) \doteq \sum_{\mathbf{n}} P(\mathbf{n}, t) \mathbf{z}^{\mathbf{n}} \tag{3.7}$$

where $\mathbf{z}^{\mathbf{n}} \doteq z_1^{n_1} z_2^{n_2} \cdots z_M^{n_M}$ for short. By multiplying Eq. 3.1 by $\mathbf{z}^{\mathbf{n}}$ and by summing over $\mathbf{n}$ we obtain the following equation:

$$\frac{\partial \phi}{\partial t}(\mathbf{z}, t) = \sum_{i=1}^{M} (z_i - 1) \langle \mathbf{z}^{\mathbf{n}} W_i^+(\mathbf{n}) \rangle + \left(\frac{1}{z_i} - 1 \right) \langle \mathbf{z}^{\mathbf{n}} W_i^-(\mathbf{n}) \rangle \tag{3.8}$$

where $\langle\cdot\rangle$ indicates a mean over $\mathbf{n}$ with probability $P(\mathbf{n},t)$. For simplicity, in the following we consider $N_i = N\ \forall i$. By substituting Eqs. 3.5 and by taking into account Eq. 3.6, we finally obtain the following partial differential equation for the PGF:

$$c\frac{\partial\phi}{\partial t} = \sum_i (z_i - 1)\sum_k m_{ki}\left[Nz_k\partial_k - Nc\partial_i - (z_i - c)\partial_i z_k\partial_k\right]\phi \tag{3.9}$$

where $c \doteq 1/(1+s)$ and $\partial_i \doteq \frac{\partial}{\partial z_i}$. In order to obtain such a relatively simple equation it is fundamental to get rid of the nonlinearity due to the denominator of the transition probabilities in Eqs. 3.4, as we have done by introducing an event rate μ'. Note that this point is overlooked in Ref. [6].
The stationary solution of Eq. 3.9 can be expressed as

$$\phi_s(\mathbf{z}) = \pi_0 + \pi_f(z_1 z_2 \dots z_M)^N \tag{3.10}$$

where π_0 and π_f stand respectively for the probability for the mutant to get extinct (loss probability) and for the fixation probability. Indeed, from the definition of PGF (Eq. 3.7),

$$\pi_0 = \phi_s(0) = P(\mathbf{0},t)$$
$$\pi_f = \prod_{i=1}^{M}\frac{1}{N!}\partial_k^N\phi = P(\mathbf{N},t)$$

where $\mathbf{N} \doteq (N,\dots,N)$; accordingly, $\pi_0 + \pi_f = 1$ since $\phi(\mathbf{1},t) = 1$ (see Eq. 3.7). The probability distribution corresponding to ϕ_s is $P_s(\mathbf{n}) = \pi_0\delta_{\mathbf{n},\mathbf{0}} + \pi_f\delta_{\mathbf{n},\mathbf{N}}$, which justifies the ansatz in Eq. 3.10.
Eq. 3.9 always admits the fixed point (1,1,...,1) corresponding to the normalization of the probability: $\phi(\mathbf{1},t) = \sum_{\mathbf{n}} P(\mathbf{n},t) = 1$. Suppose now that there exists another fixed point $\boldsymbol{\zeta} \doteq (\zeta_1,\dots,\zeta_M)$, such that at any time t

$$\frac{\partial\phi}{\partial t}(\boldsymbol{\zeta},t) = 0$$

than we could deduce that

$$\pi_0 + \pi_f(\zeta_1\zeta_2\dots\zeta_M)^N = \phi(\boldsymbol{\zeta},t=0)$$

and so, using the fact that $\pi_0 = 1 - \pi_f$,

$$\pi_f = \frac{1-\phi(\boldsymbol{\zeta},t=0)}{1-(\zeta_1\zeta_2\dots\zeta_M)^N}. \tag{3.11}$$

In the case where the mutant appears uniformly at random on one of the islands, the initial condition is

$$P(\mathbf{n}, t=0) = \frac{1}{M}\sum_{i=1}^{M} \delta_{\mathbf{n},\mathbf{e}_i}$$

where $\mathbf{e}_i$ is a vector of all zeros except for a one at position i, so

$$\phi(\mathbf{z}, t=0) = \frac{1}{M}\sum_{i=1}^{M} z_i \tag{3.12}$$

and Eq. 3.11 reads

$$\pi_f = \frac{1 - \frac{1}{M}\sum_{i=1}^{M}\zeta_i}{1-(\zeta_1\zeta_2\ldots\zeta_M)^N}. \tag{3.13}$$

Hence, the calculation of the fixation probability reduces to the problem of finding (if it exists) a non-trivial fixed point of Eq. 3.9.
Note that Eq. 3.13 evaluated at $\mathbf{z}_\mathsf{m} \doteq (c, c, \ldots, c)$ yields

$$\pi_f = \frac{1-c}{1-c^{MN}}, \tag{3.14}$$

which is the fixation probability of a well-mixed population of MN individuals within the Moran process (Eq. 1.12). The PGF equation at $\mathbf{z}_\mathsf{m}$ reads

$$c\frac{\partial\phi}{\partial t}(\mathbf{z}_\mathsf{m}, t) = Nc(c-1)\left[\sum_{k=1}^{M}(1-T_k)\partial_k\right]\phi : \tag{3.15}$$

hence, $\mathbf{z}_\mathsf{m}$ is indeed a fixed point if and only if $\forall k\ T_k = 1$. This corresponds to the isothermal theorem (Th. 2.1.1).

The star without self-loops

For a star without loops and birth-death dynamics, where we label with 0 the centre and with $k = 1, \ldots, P$ the P leaves, we have

$$\begin{aligned} m_{k0} &= 1 \text{ for } k > 0 \\ m_{0k} &= 1/P \text{ for } k > 0 \\ m_{kj} &= 0 \text{ for any } k, j > 0 \\ m_{kk} &= 0 \text{ for any } k. \end{aligned}$$

Moreover, because of the symmetry of the graph, the fixed point of Eq. 3.9 has the form $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, \zeta_1, \ldots, \zeta_1)$. By substituting into Eq. 3.9 we find

$$\zeta_0 = \frac{c(P+c)}{Pc+1}$$
$$\zeta_1 = \frac{c^2}{\zeta_0}$$

We have compared the resulting theoretical prediction for the fixation probability (from Eq. 3.13) with the simulation results for islands of size one performed as in Ch. 2; they are shown in Fig. 3.1. We find a very good agreement between the analytical prediction and the simulation results. Note that simulations are made in the discrete time framework (see above).

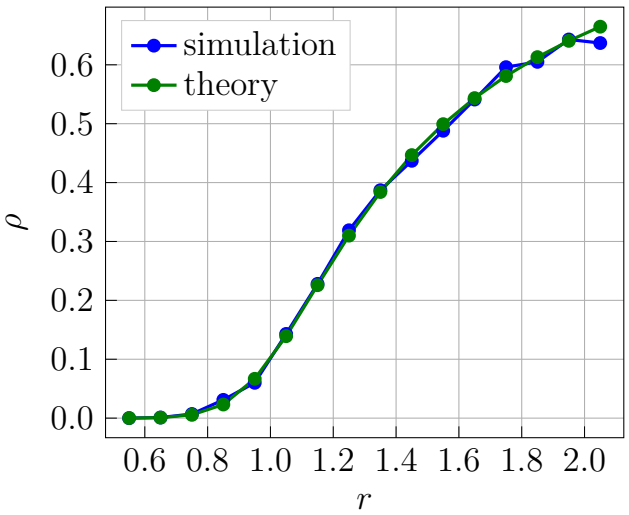


Figure 3.1: Fixation probability ρ versus mutant fitness r for a star without loops and birth-death dynamics. Simulations of the model described in Ch. 2 for islands of size one. $P = 9$, $N_i = 1\ \forall i$, 10^3 samples per data point.

3.1.2 Death-birth dynamics

In death-birth dynamics, the transition probability $T_i^+(\mathbf{n})$ is

$$T_i^+(\mathbf{n}) = \underbrace{\frac{N_i - n_i}{\sum_j N_j}}_{(1)} \sum_{k=1}^{M} \underbrace{m'_{ki}}_{(2)} \underbrace{\frac{n_k(1+s)}{N_k + s n_k}}_{(3)} \tag{3.16}$$

where (1) is the probability for a wildtype individual to die on island i, while (2) is the probability that, given that a death event occurs on island i, a migration event occurs from island k to i. It takes into account the total

fitness of island k:

$$m'_{ki} \doteq \frac{m_{ki}(N_k + sn_k)}{\sum_j m_{ji}(N_j + sn_j)} \tag{3.17}$$

with the convention that $\sum_{k=1}^{M} m_{ki} = 1$. Finally, (3) is the probability that, given that a reproduction event happens in island k, it is a mutant who reproduces. Similarly,

$$T_i^-(\mathbf{n}) = \frac{n_i}{\sum_j N_j} \sum_{k=1}^{M} m'_{ki} \frac{N_k - n_k}{N_k + sn_k}. \tag{3.18}$$

Recall that if $m'_{ki} = m_{ki}$ is also possible and it corresponds to a model in which fitness is important in determining selection only within the individuals of the same island.

To describe the process in continuous time, we introduce a total event rate (total death rate), which is proportional to the total population size (μ being the death rate of an individual):

$$\mu' \doteq \mu \sum_{i=1}^{M} N_i. \tag{3.19}$$

We have

$$W_i^+(\mathbf{n}) = \mu' T_i^+(\mathbf{n}) = \mu(N_i - n_i) \frac{\sum_k m_{ki} n_k (1+s)}{\sum_j m_{ji}(N_j + sn_j)}$$
$$W_i^-(\mathbf{n}) = \mu' T_i^-(\mathbf{n}) = \mu n_i \frac{\sum_k m_{ki}(N_k - n_k)}{\sum_j m_{ji}(N_j + sn_j)}$$

where we have used the expressions of μ' and m'_{ki} given respectively by Eqs. 3.19 and 3.17. As before, for simplicity, let us now consider the case of islands of the same size $N_i = N \ \forall i$. Then, the previous equations read

$$W_i^+(\mathbf{n}) = \mu(N - n_i) \frac{\sum_k m_{ki} n_k (1+s)}{N + s\sum_j m_{ji} n_j} \tag{3.20a}$$

$$W_i^-(\mathbf{n}) = \mu n_i \frac{\sum_k m_{ki}(N - n_k)}{N + s\sum_j m_{ji} n_j}. \tag{3.20b}$$

where we have used $\sum_j m_{ji} = 1$. In contrast with the birth-death case, here there is no way to get rid of the term $s\sum_j m_{ji} n_j$ at the denominator by introducing a suitable event rate. But if s is sufficiently small for that term to be negligible, we obtain the same transition rates as for birth-death in the

case of islands of the same size (see Eq. 3.5). Note that this equivalence does not hold for islands of different size. However, because the normalization of the m_{ki} is different from the one of birth-death, even in this approximation we get a slightly different PGF equation:

$$c\frac{\partial\phi}{\partial t} = \sum_i (z_i - 1)\left[N\sum_k m_{ki}z_k\partial_k - Nc\partial_i - (z_i - c)\partial_i\sum_k m_{ki}z_k\partial_k\right]\phi. \quad (3.21)$$

As for birth-death dynamics, finding a non-trivial fixed point of the PGF equation allows to compute the fixation probability through Eq. 3.13, in the approximation for small s. Note that if s is not small, then it is impossible to write a simple equation on the PGF. This point was not discussed in the earlier literature (Ref. [6]).

The star without self-loops

As for birth-death dynamics, we can compute the fixed points of Eq. 3.21 in the case of a star without loops, for which we have

$$\begin{aligned} m_{k0} &= 1/P \text{ for } k > 0 \\ m_{0k} &= 1 \text{ for } k > 0 \\ m_{kj} &= 0 \text{ for any } k, j > 0 \\ m_{kk} &= 0 \text{ for any } k. \end{aligned}$$

The fixed points we find are

$$\begin{aligned} \zeta_0 &= \frac{c(Pc+1)}{P+c} \\ \zeta_1 &= \frac{c^2}{\zeta_0} \end{aligned}$$

These fixed points are exact, but Eq. 3.21 is only valid in the approximation of small s. The comparison between this approximated solution and the simulation results for a star with islands of size one is shown in Fig. 3.2 in the top row.

For the simulations the model presented in Ch. 2 has been used. We see that as $s = r - 1$ increases in absolute value the error due to the approximation also increases, consistently with our analytical understanding. In the bottom row, on the left, simulation results are shown for $N_i = 10\ \forall i$, obtain within the model presented in this chapter and where fitness matters across

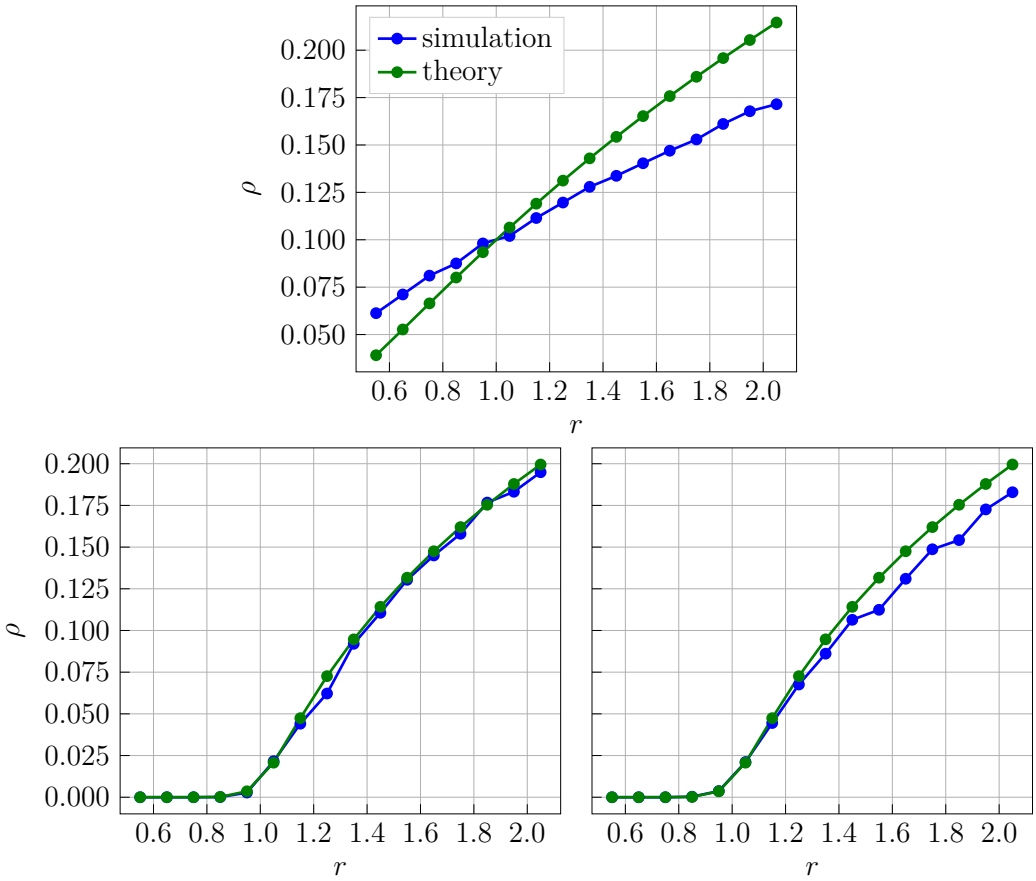


Figure 3.2: Fixation probability ρ versus mutant fitness r for a star without self-loops and death-birth dynamics ($P = 9$). Top row: $N_i = 1\ \forall i$ (10^5 samples per data point, simulations within the model of Ch. 2). Bottom: $N_i = 10\ \forall i$ (10^4 samples per data point), on the left, fitness matters across the islands, according to Eq. 3.17; on the right, fitness matters only within each island, i.e. $m'_{ki} = m_{ki}$.

the islands (see Eq. 3.17). We see that the approximation is much better in this case. On the right, the same results are shown for $m'_{ki} = m_{ki}$ implying that fitness matters only within each island, and the theoretical predictions are then less accurate especially for advantageous mutants (an explanation is given at the end of this section). As in birth-death case, simulations are made in the discrete time framework (see above).

In order to better understand the results shown in Fig. 3.2, we have computed the error made on the transition rates, due to the small-s approximation at the denominator (Eqs. 3.20), in the two cases (selection within and among the islands) in the simple cases where a loss or a fixation event occur.

The transition rates in the approximation for small s are denoted by $\tilde{W}_i^+$ and $\tilde{W}_i^-$.

In a loss event, we consider $n_i = 0\ \forall i \neq l$ and $n_l = 1$, where $l \neq 0$ hence l is a leaf-index (this is the most probable configuration with a single mutant individual for a star with multiple leaves). In the case of selection across the islands, according to Eqs. 3.20 we have

$$W_i^+ = \mu N \frac{m_{li}(1+s)}{N + s m_{li}} \tag{3.22}$$

$$W_l^- = \mu \frac{\sum_k m_{kl} N}{N} = \mu \tag{3.23}$$

where we have used $\sum_k m_{ki} = 1$ (we do not consider W_l^+ and W_i^- since they vanish) and that $m_{ll} = 0$ since we consider a star without self-loops. Independently of s, $W_l^- = \mu$, so no error is made on its value when focusing on the small s regime and neglecting the terms in s at the denominator. Contrarily,

$$\tilde{W}_i^+ = \mu N \frac{m_{li}(1+s)}{N}$$

and the relative error is

$$\frac{\left|\tilde{W}_i^+ - W_i^+\right|}{W_i^+} = m_{li}\frac{s}{N}. \tag{3.24}$$

In the case of selection within the islands, the transition rates read:

$$W_i^+ = \mu(N_i - n_i) \sum_{k=1}^{M} m_{ki} \frac{n_k(1+s)}{N_k + s n_k} \tag{3.25a}$$

$$W_i^- = \mu n_i \sum_{k=1}^{M} m_{ki} \frac{N_k - n_k}{N_k + s n_k} \tag{3.25b}$$

and, in the case of a star without self-loops, for a loss event Eqs. 3.25 yield

$$W_i^+ = \mu N \frac{m_{li}(1+s)}{N+s}$$

$$W_l^- = \mu \frac{\sum_k m_{kl} N}{N} = \mu.$$

Again, no error is made on the value of W_l^- by making the small-s approximation, since it does not depend on s, while the relative error on W_i^+ is

$$\frac{\left|\tilde{W}_i^+ - W_i^+\right|}{W_i^+} = \frac{s}{N}. \tag{3.26}$$

Analogously, we can compute the error in the case of a fixation event, where $n_i = N \ \forall i \neq l$ and $n_l = N - 1$. In this case, we find that W_i^+ is affected by the same relative error s in both cases, while the error on W_i^- in the case of selection across the islands is

$$\frac{\left|\tilde{W}_i^- - W_i^-\right|}{W_i^-} = s\left(1 - \frac{m_{li}}{N}\right)$$

while in the case of selection within the islands

$$\frac{\left|\tilde{W}_i^- - W_i^-\right|}{W_i^-} = s\left(1 - \frac{1}{N}\right).$$

Since $m_{li} \leq 1$, in a loss event the error made with respect to the simulations with selection across the islands is smaller (or equal to) the error made with respect to the case in which selection matters only within the islands. The opposite holds for a fixation event. However, loss events are much more likely than fixation events (even for a very fit mutant, the fixation probability is not higher than 0.2, see Fig. 3.2). Loss events often happen because a single mutant individual is present in the population and it is replaced by a wildtype individual before having the possibility to reproduce. The most probable configuration (under uniform initialization) is that the mutant is placed on a leaf. In this case, since $m_{li} = 1/P$, the relative error made in a loss event with respect to the simulations with selection across the islands is a factor P smaller than the one made with respect to the simulations with selection only within the island. This is in agreement with the results shown in Fig. 3.3, where the fixation probability is plotted versus the island size N in the case of selection across the islands (left) and in the case of selection only within the islands (right): the error on the fixation probability due to the small-s approximation in the first case is about half of the error in the second case.

Notice that the predominance of loss events over fixation events can explain also why we observe that the approximation becomes better if N increases, as shown in Fig. 3.3, and why it worsens for larger values of s, especially in the case where fitness matters only within the islands, shown in Fig. 3.2, bottom row, right panel. Indeed, in a loss event the relative error (Eqs. 3.24 and 3.26) is inversely proportional to N and directly proportional to s.

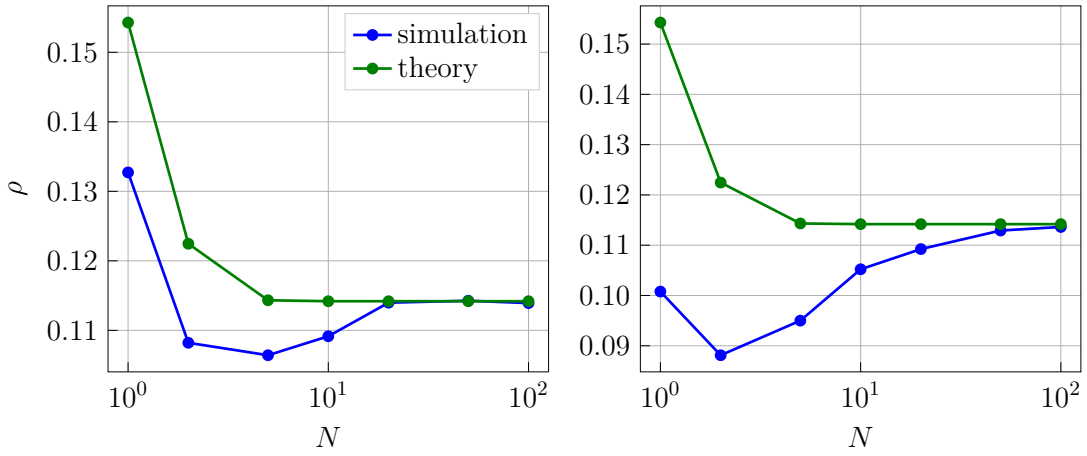


Figure 3.3: Fixation probability ρ versus island size N for a star without self-loops and death-birth dynamics. On the left, fitness matters across the islands, according to Eq. 3.17; on the right, fitness matters only within each island, i.e. $m'_{ki} = m_{ki}$. $P = 9$, 10^5 samples per data point.

3.2 Quasi-fixed points of the probability generating function equation

For the star with self-loops Eq. 3.9 for birth-death and Eq. 3.21 for death-birth have no fixed points except for the trivial one. But in both equations, all terms entering in the sum have a coefficient proportional to N or to Nc (where $c = 1/r$) except for the terms that contain a second derivative. In the regime where both N and Nc are large, we may thus consider the terms with the second derivative negligible compared to the others. If this is the case, the fixed points satisfy the following system of algebraic equations

$$\sum_i m_{ki} \left(\zeta_i - 1\right) = cT_k \left(1 - 1/\zeta_k\right) \tag{3.27}$$

for birth-death and

$$\sum_i m_{ki} \left(\zeta_i - 1\right) = c \left(1 - 1/\zeta_k\right) \tag{3.28}$$

for death-birth, which can be solved numerically. Note that the two equations are identical if $T_k = 1 \; \forall k$ (the temperature being given by Eq. 2.1.1 for birth-death and Eq. 2.20 for death-birth), hence for an isothermal star. In this case, Eqs. 3.27 and 3.28 admits $\zeta_i = c \; \forall i$ as nontrivial solution, corresponding to

the fixation probability of a well-mixed population within the Moran process (see Eq. 3.14).

In Fig. 3.4 we can see the results for birth-death on the top row and for death-birth on the bottom row for a star with self-loops of parameters $x = 0.1$ and $y = 0.1$ (Eqs. 2.11 and 2.19), for $N = 1$ on the left and for $N = 10$ on the right. For both dynamics the approximation is better if the size of the subpopulations is larger and the approximation is worse for death-birth dynamics than for birth-death. Indeed, the theoretical results for birth-death dynamics are affected only by the approximation for the quasi-fixed points (consisting in neglecting the second derivatives in Eqs. 3.9 and 3.21, valid for large N), while the theoretical results for death-birth dynamics are affected also by the approximation on the transition rates (small-s approximation in Eqs. 3.20).

Note that $x = y = 0.1$ implies that a reproducing individual has probability 0.9 to replace an individual in the same subpopulation and probability 0.1 to migrate. This situation of "strong" self-loops or rare migrations is biologically relevant, since in real structured populations the migration rate is in general low compared to the reproduction rate. We notice that in this case the difference between birth-death and death-birth is reduced with respect to the star without self-loops, as it is shown in Fig. 3.5. On the left, the fixation probability for the two dynamics is plotted versus r in the case of a star without self-loops; on the right the same is done for a star with "strong" loops ($x = y = 0.1$). This is in agreement with the fact that by adding "strong" self-loops of the same weight on the centre and on the leaves, we have reduced the temperature difference between the nodes. Indeed, by using Eqs. 2.12, we obtain

$$\Delta T \doteq T_c - T_l = (P+1)\left(x - \frac{y}{P}\right)$$

and for $x = y = \varepsilon$

$$\Delta T = (P+1)\left(1 - \frac{1}{P}\right)\varepsilon.$$

Hence, the stronger are the self-loops (i.e. the smaller is ε) the closer the star is to isothermality and we know that, for an isothermal star, Eqs. 3.27 and 3.28 together with Eq. 3.13 predicts the same fixation probability independently of the dynamics (which is equal to the fixation probability of a well-mixed population within the Moran process, see above). Indeed, Fig. 3.5 shows that for $x = y = 0.1$ the fixation probability of a star is not far from the one for the Moran process (black dashed line).

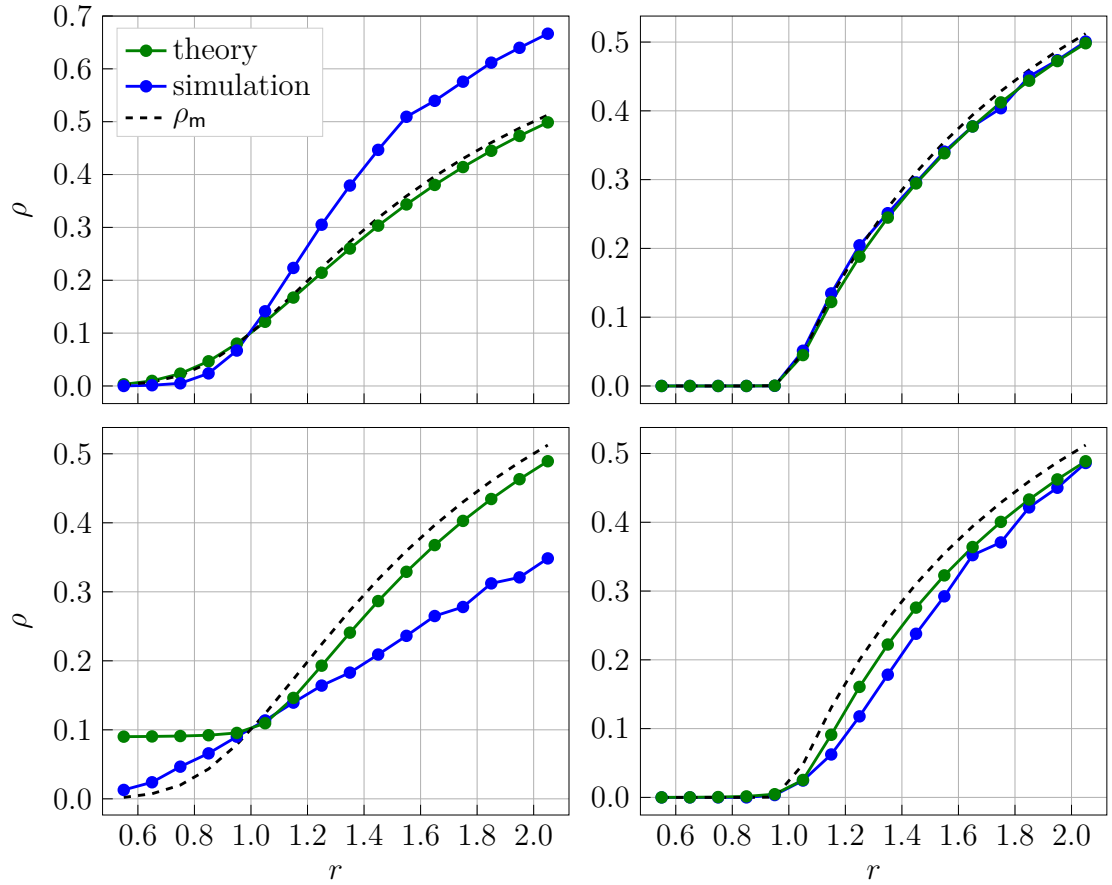


Figure 3.4: Fixation probability ρ versus mutant fitness r for a star with self-loops. Top row: birth-death dynamics; bottom row: death-birth dynamics. Left: $N = 1$; right: $N = 10$. $P = 9$, $x = 0.1$, $y = 0.1$, 10^4 samples per data point.

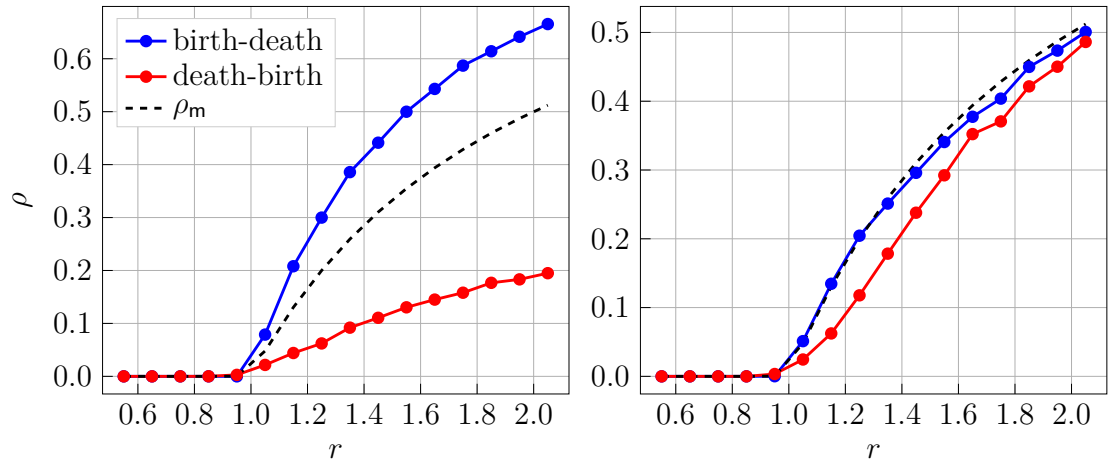


Figure 3.5: Fixation probability ρ versus mutant fitness r for a star without self-loops (left) and with self-loops of parameters $x = y = 0.1$ (right). $P = 9$, $N = 10$, 10^4 samples per data point.

Conclusions

The Moran process is a model for the evolution of a well-mixed population for which it is possible to compute analytically the fixation probability and the fixation times (Fig. 1.3). In this case birth-death and death-birth dynamics are completely equivalent since the transition probabilities are the same independently of the dynamics (Eqs. 1.10). However, if we remove the possibility for an individual to replace itself, the fixation probability (and the fixation times) depends on the choice of the dynamics, and the two dynamics are equivalent only in the limit of an infinite population size. This type of model has some limits: the fact that the population size is kept constant is far from reality and population structure is completely neglected.

A model like the one proposed in Ref. [5] introduces a structure in the population by considering each individual to be placed on a node of a graph and edge weights specify replacement probabilities. However, the population size is still fixed and, in general, the choice of the dynamics strongly affects the fixation probability (Fig. 2.5). Moreover, the initial conditions used for the placement of the mutant also impact the evolutionary outcome (Fig. 2.7) as it was observed in Ref. [10], where uniform and temperature initial conditions were considered. We generalized these results and proposed as realistic initial conditions a linear combination of the two, which takes into account mutation arising at division and the possibility for the mutant individual to stay in the same node or to migrate to a neighbouring node.

By adding self-loops, we allow individuals to replace themselves and the amplifying property of the star changes according to the weight of the self-loops. While in birth-death dynamics the choice of the normalization for the edge weights w_{ij} is natural and corresponds to the requirement for W to be right stochastic (Eq. 2.11), in death-birth dynamics the choice is not unique and, in the case of a star with self-loops, leads to different results for the fixation probability.

We have studied a coarse-grained version of the model of populations on graphs (Ref. [6]), which considers a well-mixed subpopulation (island)

at each node of a graph and where edge weights specify the probability to migrate in another subpopulation. In the case of a star without self-loops and birth-death dynamics it is possible to compute analytically the fixation probability (Ref. [6]) and for population of size one we recover the same results of the previous model (Fig. 3.1). However, for death-birth dynamics an approximation on the transition rates is necessary, and it only holds for small fitness differences (Fig. 3.2). In this case we can consider two different versions of death-birth dynamics: one for which selection (based on fitness) acts across the islands, and the other for which selection acts only within each island. Only the first one gives back the model with one individual per node in the case of islands of size one (Fig. 3.2, top row). The analytical formula for the fixation probability is a better approximation for the first version of death-birth dynamics with respect to the second (Fig. 3.2, bottom row) and for both the error decreases as the size of the islands increases (in the range that we have considered for the fitness of the mutant).

Adding self-loops to this model is particularly relevant, since it allows individuals to stay in the same subpopulations after division. In this case, it is possible to derive an analytical expression for the star only in the approximation of large island size (Fig. 3.4). A biologically relevant case is the one that considers self-loops of large weight, so that migration events become rare, and reproducing individuals mostly replace other individuals of the same subpopulation. The result is that the dependence of the fixation probability on the choice of the dynamics reduces significantly (Fig. 3.5).

The coarse-grained version of the model proposed by Ref. [5] seems more realistic, since it reduces the artifacts due to the choice of the dynamics, but it still imposes a strong constraint on the size of the population. In the future it would be interesting to improve the model with subpopulations, by relaxing the constraint on the size of the population by allowing stochastic fluctuations around a mean value, and to implement a Gillespie algorithm in order to sample the stochastic trajectory.

Appendices

Appendix A

Fixation times

Beyond the fixation probability, there are other quantities of interests in the evolutionary process: the fixation times. We recall here the three different fixation times in the case of a population of size N composed by two different types of individuals, A and B.

1. The unconditional fixation time t_i is the average time needed for the system to reach one of the two absorbing states $i = 0$ or $i = N$;

2. the conditional fixation time t_i^A is the average time needed for i individuals of species A to fix ($i = N$);

3. the conditional fixation time t_i^B is the average time needed for $N - i$ individuals of species B to fix ($i = 0$).

In the following we detail the calculations of these quantities, following Ref. [8].

Unconditional fixation time

We can write an equation for t_i in the same way as for the fixation probability (Eq. 1.1):

$$t_i = 1 + \beta_i t_{i-1} + (1 - \alpha_i - \beta_i) t_i + \alpha_i t_{i+1} \tag{A.1}$$

with boundary conditions $t_0 = t_N = 0$. In the same spirit of the derivation of the fixation probability we define a new variable $z_i \doteq t_i - t_{i-1}$ so that Eq. A.1 reads

$$z_{j+1} = \gamma_j z_j - 1/\alpha_j \tag{A.2}$$

with the usual definition $\gamma_j \doteq \beta_j/\alpha_j$. So we have the following iterative scheme

$$\begin{aligned}
z_1 &= t_1 - t_0 = t_1 \\
z_2 &= t_2 - t_1 = \gamma_1 t_1 - 1/\alpha_1 \\
z_3 &= t_3 - t_2 = \gamma_2\gamma_1 t_1 - \gamma_1/\alpha_1 - 1/\alpha_2 \\
&\vdots \\
z_k &= z_k - z_{k-1} = t_1 \prod_{j=1}^{k-1} \gamma_j - \sum_{j=1}^{k-1} \alpha_j^{-1} \prod_{m=j+1}^{k-1} \gamma_m \\
&\vdots \\
z_N &= z_N - z_{N-1} = t_1 \prod_{j=1}^{N-1} \gamma_j - \sum_{j=1}^{N-1} \alpha_j^{-1} \prod_{m=j+1}^{N-1} \gamma_m
\end{aligned} \tag{A.3}$$

and because of the telescopic nature of the series we have $\sum_{k=i+1}^{N} z_k = -t_i$. In the case $i = 1$ and using Eq. A.3 we get

$$t_1 = -t_1 \sum_{k=1}^{N-1} \prod_{j=1}^{k} \gamma_j + \sum_{k=1}^{N-1} \sum_{j=1}^{k} \frac{1}{\alpha_j} \prod_{m=j+1}^{k} \gamma_m$$

so that, by rearranging the terms and by using Eq. 1.6 for ρ_1, we get the following expression for t_1

$$t_1 = \rho_1 \sum_{k=1}^{N-1} \sum_{j=1}^{k} \alpha_j^{-1} \prod_{m=j+1}^{k} \gamma_j. \tag{A.4}$$

The average unconditional fixation time in the case of i individuals of species A is finally given by

$$t_i = -t_1 \sum_{k=i}^{N-1} \prod_{j=1}^{k} \gamma_j + \sum_{k=i}^{N-1} \sum_{j=1}^{k} \alpha_j^{-1} \prod_{m=j+1}^{k} \gamma_m. \tag{A.5}$$

Conditional fixation times

We can compute both t_i^A and t_i^B using a similar method. We start from the equation

$$\rho_i t_i^A = \rho_{i-1}\beta_i \left(1 + t_{i-1}^A\right) + \rho_i \left(1 - \alpha_i - \beta_i\right) \left(1 + t_i^A\right) + \rho_{i+1}\alpha_i \left(1 + t_{i+1}^A\right) \tag{A.6}$$

where ρ_i is the probability of fixation of i individuals of type A. Eq. A.6 is similar to Eq. A.1, except that here we have to multiply by the probability that the fixation of species A occurs since t_i^A is a conditional fixation time. If we define $\theta_i^A \doteq \rho_i t_i^A$ for brevity and $w_i \doteq \theta_i^A - \theta_{i-1}^A$ we get an equation with the same structure of Eq. A.2:

$$w_{i+1} = \gamma_i w_i - \frac{\rho_i}{\alpha_i}$$

and using the same iteration approach as before we obtain

$$w_k = \theta_i^A - \theta_{k-1}^A = \theta_1^A \prod_{j=1}^{k-1} \gamma_j - \sum_{j=1}^{k-1} \frac{\rho_j}{\alpha_j} \prod_{m=j+1}^{k-1} \gamma_m$$

and as boundary conditions we have $\theta_0^A = 0$ because $\rho_0 = 0$ and $\theta_N^A = 0$ because $t_N^A = 0$. Furthermore, we have that $\sum_{k=i+1}^{N} w_k = -\theta_i^A$ and for $i = 1$ we get

$$t_1^A = \sum_{k=1}^{N-1} \sum_{j=1}^{k} \frac{\rho_j}{\alpha_j} \prod_{m=j+1}^{k} \gamma_j \tag{A.7}$$

and finally for any i we have

$$t_i^A = -t_1^A \frac{\rho_1}{\rho_i} \sum_{k=i}^{N-1} \prod_{j=1}^{k} \gamma_j + \frac{1}{\rho_i} \sum_{k=i}^{N-1} \sum_{j=1}^{k} \frac{\rho_j}{\alpha_j} \prod_{m=j+1}^{k} \gamma_m. \tag{A.8}$$

Using similar calculations we can compute the fixation time t_{N-1}^B of a single B mutant, for which we get

$$t_{N-1}^B = \sum_{k=1}^{N-1} \sum_{j=1}^{k} \frac{\tilde{\rho}_{N-j}}{\beta_{N-j}} \prod_{m=j+1}^{k} \frac{1}{\gamma_{N-m}} \tag{A.9}$$

where $\tilde{\rho}_j \doteq 1 - \rho_j$. As before, we can also compute the fixation time for $N - i$ individuals of species B, which reads

$$t_i^B = -t_{N-1}^B \frac{\tilde{\rho}_{N-1}}{\tilde{\rho}_i} \sum_{k=N-i}^{N-1} \prod_{m=1}^{k} \frac{1}{\gamma_{N-m}} + \sum_{k=N-i}^{N-1} \sum_{j=1}^{k} \frac{\tilde{\rho}_{N-j}}{\tilde{\rho}_i} \frac{1}{\beta_{N-m}} \prod_{m=j+1}^{k} \frac{1}{\gamma_{N-m}}. \tag{A.10}$$

Appendix B

Numerical simulations

In the following we will use monospaced font to refer both to the code variable and its value.

B.1 One individual per node

For the simulations of Ch. 2, first we define a two-dimensional vector `W` of dimensions $\mathtt{N} \times \mathtt{N}$, where `N` is an integer variable storing the total population size. For the Moran process `W` is such that $\mathtt{W[i][j]} = 1/\mathtt{N}$. For the star with $\mathtt{N} = P + 1$ leaves and x, y parameters, `W` is defined according to Eq. 2.11 for birth-death dynamics and according to Eq. 2.19 for death-birth dynamics.

For each sample, we define a binary vector `v` of length `N`, which keeps track of the type of the individual at each node: a wildtype is encoded by 0 and a mutant by 1. The vector is initialized by all zeros and a single one. The index of this latter is drawn according to the chosen distribution (uniform, temperature or mixed) between 0 and $\mathtt{N}-1$. An integer variable `m`, storing the sum of the elements in `v`, keeps track of the number of mutants. The fitness of the mutant is stored in `r` and the number of time steps `t` is initialized with 0.

At each time step, two integers between 0 and $\mathtt{N} - 1$ are drawn: the index `i` of the individual chosen for reproduction and the index `j` of the individual chosen for death. In birth-death dynamics `i` is drawn first, with a probability distribution proportional to the fitness, then `j` is drawn with probability `W[i][j]`. In death-birth dynamics `j` is drawn first, with probability $1/\mathtt{N}$, and then `i` is drawn with probability proportional to $\mathtt{r}\,\mathtt{W[i][j]}$ if $\mathtt{v[i]} = 1$ and `W[i][j]` if $\mathtt{v[i]} = 0$. Next, `v`, `m` and `t` are updated. The process ends when $\mathtt{m} = \mathtt{N}$ or $\mathtt{m} = 0$. A Boolean variable encoding the outcome of the process (the

mutant fixes or not) is saved together with the number `t` of time steps.

The fixation probability is computed as the ratio between the times the mutant has fixed and the total number of simulated random trajectories. The mean fixation times are also computed with an error bar corresponding to the 95% confidence interval.

B.2 Subpopulations on a graph

A similar code has been used for the simulations of Ch. 3 concerning the coarse-grained version of the model, with some differences. In this case, the number of leaves of the star is `P`, and a constant vector `N` of length `P`+1 stores the total number of individuals on each island. Another vector `n` of length `P` + 1 stores the number of mutants on each island. This latter is initialized with all zeros and a single one; the index of the element 1 is drawn uniformly at random between 0 and `P`. The two-dimensional vector representing the matrix of weights is defined as before and it has dimensions (`P` + 1) × (`P` + 1).

The mutant fitness is 1 + `s` and `t` is the number of time steps, initialized by 0. At each time step, the type of the individual who reproduces and of the individuals who dies are stored in two Boolean variables, `tb` and `td`, which take value 1 for mutant and 0 for wildtype. The total number of individuals is `Ntot` and the total number of mutants is `ntot`, and they are equal to the sum of the elements in `N` and `n` respectively.

At each time step of birth-death dynamics, the index `k` of the island where a reproduction event occurs is drawn with probability proportional to the fitness of the island `N`[`k`] + `s` `n`[`k`]. Then, within the island, the value of `tb` is chosen with probability proportional to the fitness (`n`[`k`](1 + `s`) for the mutant and `N`[`k`] − `n`[`k`] for the wildtype). The index `i` of the island where death occurs is chosen with probability proportional to `W`[`k`][`i`] and the value of `td` is drawn with probability proportional to `N`[`i`] − `n`[`i`] for the wildtype and `n`[`i`] for the mutant.

At each time step of death-birth dynamics, the index `i` where a death event occurs is chosen uniformly at random between 0 and `P`. The value of `td` is chosen with probability proportional to `N`[`i`] − `n`[`i`] for the wildtype and `n`[`i`] for the mutant. Then, in the case in which fitness matters across the islands, the index `k` of the island where reproduction occurs is chosen with probability proportional to `W`[`k`][`i`](`N`[`k`] + `s` `n`[`k`]); else, `k` is chosen according to `W`[`k`][`i`]. The value of `tb` is chosen with probability proportional to the fitness (`n`[`k`](1 + `s`) for the mutant and `N`[`k`] − `n`[`k`] for the wildtype).

At the end of each time step, `n`, `ntot` and `t` are updated. The process ends when `ntot` = `Ntot` or `ntot` = 0. The outcome of the process together with the number of time steps are stored. Finally, as before, the fixation probability and the mean fixation times are computed over all the simulated trajectories.

Bibliography

[1] Wikipedia. Mutation — Wikipedia, the free encyclopedia, 2019. [Online; accessed 27-June-2019].

[2] Wikipedia. Bacteria — Wikipedia, the free encyclopedia, 2019. [Online; accessed 27-June-2019].

[3] Loïc Marrec and Anne-Florence Bitbol. Quantifying the impact of a periodic presence of antimicrobial on resistance evolution in a homogeneous microbial population of fixed size. *Journal of theoretical biology*, 457:190–198, 2018.

[4] World Health Organization. Antibiotic resistance, 2018. [Online; accessed 5-July-2019].

[5] Erez Lieberman, Christoph Hauert, and Martin A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433:312–316, Jan 2005.

[6] Bahram Houchmandzadeh and Marcel Vallade. The fixation probability of a beneficial mutation in a geographically structured population. *New Journal of Physics*, 13(7):073020, jul 2011.

[7] N. G. Van Kampen. *Stochastic processes in Physics and Chemistry*. Elsevier, 2007.

[8] Arne Traulsen and Christoph Hauert. *Stochastic Evolutionary Game Dynamics*, chapter 2, pages 25–61. John Wiley Sons, Ltd, 2010.

[9] Martin A. Nowak. *Evolutionary dynamics: exploring the equations of life*. Belknap / Harvard, 2006.

[10] B. Adlam, K. Chatterjee, and M. A. Nowak. Amplifiers of selection. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150114, 2015.

[11] Bahram Houchmandzadeh and Marcel Vallade. Alternative to the diffusion equation in population genetics. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 82:051913, 11 2010.