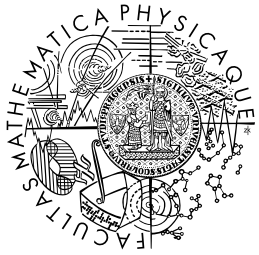


FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Lenka Kopfová

Fixation time in Moran process under strong selection

Computer Science Institute of Charles University

Supervisor of the master thesis: Josef Tkadlec, Ph.D.

Study programme: Computer Science – Discrete Models and Algorithms

Prague 2024

I declare that I carried out this master thesis on my own, and only with the cited sources, literature and other professional sources. I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

I would like to thank my supervisor Josef Tkadlec, Ph.D. for all the time he gave me. For his help, patient guidance and useful discussions throughout the completion of this thesis. And also for the enthusiasm and encouragement he passed on to me in further academic pursuits. I would also like to thank my friends, particularly the co-organisers of the mathematical correspondence seminars, for creating a great environment and motivation during my studies. Last but not least, I would like to thank my family, especially my parents, for their unconditional support and care.

Title: Fixation time in Moran process under strong selection

Author: Lenka Kopfová

Institute: Computer Science Institute of Charles University

Supervisor: Josef Tkadlec, Ph.D., Computer Science Institute of Charles University

Abstract: Moran process is a model used in evolutionary dynamics to study natural selection. In this process, a population of individuals evolves in steps. In one step a random individual is selected with probability proportional to its fitness and spreads to its randomly selected neighbor. The classical course of study is to consider an individual with a hereditary mutation and examine the fate of this mutation in time.
This thesis investigates a modified version of the Moran process that corresponds to the strong selection, as in the dynamics of invasive species. In this process, only the mutant individuals spread and eventually conquer the whole population. The key quantity that we study is the so-called fixation time, which is the expected time until all individuals become mutants.
We give tight upper and lower bounds for fixation time on a general population structure and refine them for some classes. Additionally, we compute the precise fixation times on some specific population structures.

Keywords: evolutionary dynamics, Moran process, graph

Název práce: Čas fixace v Moranově procesu při silné selekci

Autor: Lenka Kopfová

Ústav: Informatický ústav Univerzity Karlovy

Vedoucí diplomové práce: Josef Tkadlec, Ph.D., Informatický ústav Univerzity Karlovy

Abstrakt: Moranův proces je model, který se používá v evoluční dynamice ke studiu přírozeného výběru. V tomto procesu máme populaci jedinců, která se vyvíjí v krocích. V jednom kroku je vybrán náhodný jedinec s pravděpodobností úměrnou své zdatnosti (fitness), a ten se rozšíří do svého náhodně vybraného souseda. Klasickým předmětem studia je uvažovat jedince s dědičnou mutací a zkoumat osud této mutace v čase.
Tato práce se zabývá upravenou verzí Moranova procesu, která odpovídá silné selekci, tak jako je tomu například v dynamice invazivních druhů. V tomto procesu se rozmnožují pouze mutantní jedinci, kteří nakonec ovládnou celou populaci. Klíčovou veličinou, kterou studujeme, je tzv. čas fixace, tedy očekáváná doba než se všichni jedinci stanou mutanty.
V práci ukazujeme těsné horní a dolní odhady na čas fixace pro obecnou strukturu populace a zlepšujeme je pro některé třídy. Kromě toho dokazujeme přesné doby fixace pro některé konkrétní struktury populace.

Klíčová slova: evoluční dynamika, Moranův proces, graf

Contents

1 Introduction

Imagine a homogeneous population of some individuals that evolves in time. The individuals reproduce at some given rate, and the new offsprings gradually replace others. Suppose that, at some point, one individual acquires a hereditary mutation that affects its reproductive rate. This mutation may evolve in many ways; its frequency fluctuates; maybe at some point many individuals have this mutation, and at other times it is rare. But eventually, the mutation either spreads throughout the whole population or goes extinct. Depending on the reproductive advantage of the new mutation and the population structure, we may investigate the probable fates of this mutation. For example, how likely this mutation is to conquer the whole population or how long it will take in expectation. These quantities are called fixation probability and fixation time, respectively.

One of the classical tools to model natural selection and evolution is the so-called Moran process. This process was first introduced by P. A. P. Moran in the late '50s [1] and considers a well-mixed population with n individuals. That is, each individual directly interacts with all the others. The key parameter that affects the outcome is the relative strength of mutant selection r with respect to the background residents. For fixed $r > 1$ and large n it can be shown that the fixation probability tends to $1 - \frac{1}{r}$ and the fixation time scales roughly as $n \log n$.

In 2005 the model was extended in a way that the population structure is represented by a graph [2]. The vertices correspond to the individuals and the edges to the interactions between them. The standard Moran process with the well-mixed population thus corresponds to the complete graph. Since then there has been extensive research regarding fixation probability considering different population structures and different values of the mutant advantage r. For example, the limit $r \to 1$ called *weak selection* corresponds to the setting where mutants have only marginal advantage [3]. For any fixed $r > 1$ there are structures called *amplifiers* that increase the fixation probability of mutants, see [4, 5]. Less is known about the fixation time. The fixation time on undirected graphs is polynomial as shown in [6]. On the other hand in the case of directed graphs it can be exponential [7]. For more information about the Moran process on graphs see [8] for a recent survey.

In this thesis, we study the Moran process in the limit $r \to \infty$. Concretely, we imagine some super advantageous mutation such that once this mutation arrives, only the individuals with this mutation reproduce. This can also be viewed as the arrival of an invasive species into an environment where the species lacks a natural predator. For most sensible population structures this kind of mutation always eventually expands throughout the whole population, thus the fixation probability is equal to 1. Because of that, our main interest will be to investigate the fixation time.

1.1 Preliminaries

Moran process

Our model consists of a graph G (directed or undirected) representing a population structure. The vertices of G denoted by $V(G)$ correspond to the individuals in the population. Two vertices directly interact if one is a neighbor of the other, meaning there is an edge between them. At any point, the individual associated with a given vertex can be either a mutant or a resident. Let us call a setting where a particular subset of vertices are mutants as a *state*. More formally:

Definition 1. *We call a tuple* (G, M) *a* state *where* G *is a directed or undirected graph representing a particular population structure. And* $M \subseteq V(G)$ *is the set of mutant vertices. Then,* $V(G) \setminus M$ *is the set of resident vertices. We will usually denote* $|V(G)| = n$ *the number of individuals in our population.*

The population evolves in time, transiting to different states according to the so-called Moran birth-death process. Considering some fixed value r, every vertex has assigned a fitness value (the rate at which it reproduces). The fitness is $r > 1$ for mutants, and 1 for residents (regardless of the vertex position in the graph G). We will refer to this Moran process with finite $r > 1$ as a finite case. One step of the classical Moran birth-death process consists of two phases – selection and reproduction (sometimes also called birth and death). In the selection phase, an individual u for reproduction is selected proportionally to its fitness. Then, in the reproduction phase, it chooses a uniformly random neighbor v in the graph G into which it spreads. That means if u is a mutant, then v becomes a mutant, and if u is a resident, then v becomes a resident. Note that if both u and v are of the same type, the state doesn't change. Otherwise, the new state is the same except for exactly one vertex.

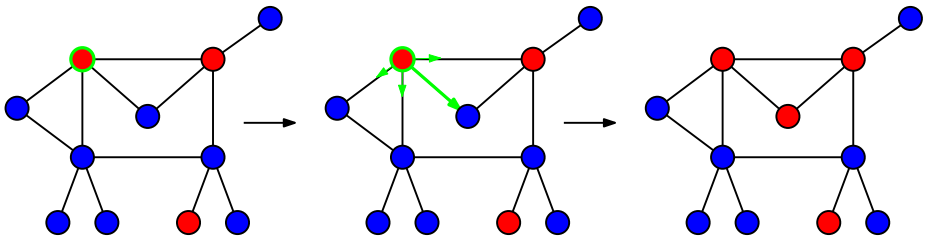


Figure 1.1 One step of the Moran process.

These steps are repeated until the process reaches absorption, that is until all of the vertices become mutants (a state known as *fixation*) or all of them become residents (a state known as *extinction*). Note that on a general graph, the process may never be absorbed. However, we will consider only those population structures such that the absorption happens with probability 1. Once the absorption is ensured, we may investigate some values associated with it. The probability that the process ends in fixation is called *fixation probability*. We can also consider a random variable counting the number of steps until the process reaches absorption. The expected value of this variable is then called *absorption time*. Similarly, we define the *fixation time* to be the expected value of the same random variable conditioned on the fixation event. That means we look only at those trajectories that end in fixation.

Moran process at $r \to \infty$

In this thesis, we consider a modified version of the classical Moran birth-death process, which captures the idea of a strong selection of mutants, meaning that only mutants can reproduce.

We define two different processes that capture this idea. They differ in the probability that a particular mutant vertex is selected for reproduction. The first process corresponds to the scenario when we pick a uniformly random vertex, and then if it is a mutant, it reproduces; otherwise, we pick again.

Definition 2 (Continuous process)**.** *Consider a given graph G and its state (G, M). One step of the* continuous process *is defined as follows:*

1. *(birth) A uniformly random vertex is selected. If it is a mutant, it reproduces and continues to the death phase. However, if it is a resident, nothing happens and we repeat this phase. Thus a particular mutant vertex $m \in M$ is selected for reproduction with probability $\frac{1}{|V(G)|}$ and the probability that some mutant vertex is selected is $\frac{|M|}{|V(G|}$.*

2. *(death) A neighbor u of m is selected uniformly at random and m spreads into u. The probability of picking a given neighbor u of m is thus $\frac{1}{\deg(m)}$.*

Alternatively, we can consider only those steps at which a mutant vertex is picked in the birth phase. This process can also be obtained as the limit of the Moran birth-death process when considering $r \to \infty$.

Definition 3 (Limit process)**.** *Consider a given graph G and its state (G, M). One step of the* limit process *is defined as follows:*

1. *(birth) A mutant vertex $m \in M$ is selected for reproduction with probability $\frac{1}{|M|}$.*

2. *(death) A neighbor u of m is selected uniformly at random and m spreads into u. The probability of picking u is thus $\frac{1}{\deg(m)}$.*

In Chapter 5, we argue why we think the first process is more natural, and thus, we will consider it to be the default one.

Note that in these processes once a vertex becomes mutant it stays mutant forever. Therefore absorption occurs if and only if fixation occurs. Hence, if some resident vertex is not reachable by a path from any mutant vertex, the process never ends in absorption. We will thus consider only these starting states:

Definition 4. A starting state *is a state (G, M) such that for every resident vertex $u \in V(G) \setminus M$ there exists a path from some mutant vertex v to the vertex u (directed path if G is a directed graph).*

If the process begins in a starting state, with probability 1 it ends in fixation in finite time. Note that for undirected graphs, all states are starting states if and only if the graph G is connected (for disconnected graphs, no states with exactly one mutant are starting states). We will often consider starting states that consist of only one mutant vertex.

Similarly to the classical Moran birth-death process, we can consider the absorption and fixation times. Because in this particular case, the process ends in fixation with probability one, the absorption and fixation time coincide. Both of them count the expected number of steps until all vertices become mutants. We will denote by $\mathrm{T}(G, v)$ the fixation time under the continuous process with starting state $(G, \{v\})$. Similarly, we denote an analogous quantity by $\mathrm{T_L}(G, v)$ for the limit process. We will often study the worst-case scenarios in order to prove upper bounds on the quantities $\mathrm{T}(G, v)$ and $\mathrm{T_L}(G, v)$. Thus we also denote $\mathrm{T}(G) = \max_{v \in V(G)} \mathrm{T}(G, v)$ the maximum expected fixation time over all starting vertices and similarly $\mathrm{T_L}(G) = \max_{v \in V(G)} \mathrm{T_L}(G, v)$.

Note that in each step of both of these processes, the new state either remains the same or gains a new mutant. Hence, we are gaining the mutants one by one, and thus we can divide the steps of the process into stages.

Definition 5. *Let us fix a starting state $(G, \{v\})$. We break the time needed for all the vertices to become mutants into $n - 1$ stages. One stage consists of gaining one more mutant, meaning in stage k, we start with k mutants and end with $k + 1$ mutants. With a slight abuse of notation we denote the probability of gaining a new mutant in stage k as P_k and the expected time until this happens as $\mathbb{E}[t_k] = \frac{1}{P_k}$. Formally, this depends not only on k but also on the current state.*

To distinguish the steps in which the state does change from those when it doesn't, we define the notion of an active edge and an active vertex:

Definition 6. *An* active edge *is an edge between a mutant vertex and a resident vertex. An active vertex is a vertex that is incident to at least one active edge.*

With these definitions, we can show our first result. It states that the limit process is always the same or faster than the continuous one.

Theorem 1. *For any graph G (directed or undirected) and any vertex v we have $\mathrm{T}(G, v) \geq \mathrm{T_L}(G, v)$.*

Proof. Consider any configuration X of nodes currently occupied by mutants. Denote the number of mutants by $|X| = k$. Recall that P_k is the probability of gaining a mutant in a single step in the continuous process. Similarly, for the limit process, denote the probability of gaining a mutant in a single step by P_k^L. We claim that $P_k = \frac{k}{n} \cdot P_k^L$. Indeed, in the continuous process, the probability of picking a particular mutant for reproduction is $\frac{1}{n}$, whereas it is $\frac{1}{k}$ in the limit case. Plugging in $k \leq n$ we find $P_k \leq P_k^L$ and so $\mathbb{E}[t_k^L] \geq \mathbb{E}[t_k]$. As this holds for any k and any mutant configuration X with k mutants, we get the desired result $\mathrm{T}(G) \geq \mathrm{T_L}(G)$.

□

Remark. Throughout this text, we will also use the asymptotic notation $o(\cdot)$, $O(\cdot)$, $\Omega(\cdot)$ and $\Theta(\cdot)$ to denote that some function f is asymptotically strictly smaller than some other function g (denoted $f = o(g)$), asymptotically smaller than or equal to ($f = O(g)$), asymptotically larger than ($f = \Omega(g)$) and asymptotically equal to ($f = \Theta(g)$). We will also use the symbol $\approx$ to denote "approximately

equal to," meaning $f(n) \approx g(n)$ if $f(n) = g(n) + o(g(n))$. For example $\frac{1}{2}n^2 + 3n = o(n^3) = O(n^2) = \Omega(n \log n) = \Theta(n^2)$. For more details, see [9].

In the later chapters, we also use H_n to denote the harmonic number. That is the partial sum of harmonic series, meaning $\mathrm{H}_n = \sum_{k=1}^{n} \frac{1}{k} \approx \log n$.

1.2 Results and organization of the thesis

In this work, we study the behavior of the continuous and limit processes. In the second chapter, we give upper bounds on its fixation times. We show a general upper bound $O(n^3)$ that applies to all spatial structures (that is, for both directed and undirected graphs) and prove it is tight. This polynomial upper bound contrasts with the Moran process with finite $r > 1$. As described in [7], there exists an infinite family of directed graphs with exponential absorption time.

Then, we proceed by improving this upper bound to $O(n^2)$ for regular graphs and to $O(n^2\sqrt{n})$ for undirected graphs. We consider this to be nontrivial and one of the most interesting results of this thesis. Note that this bound for general undirected graphs doesn't hold in the case with finite $r > 1$. As shown in [10], there exists a graph (the double star graph) with fixation time $\Omega(n^3)$. Also, the best-known upper bound for the class of undirected graphs, in this case, is $O(n^{3+\epsilon})$ for arbitrarily small $\epsilon > 0$.

In the third chapter, we shift our attention to lower bounds. We prove the asymptotical lower bound $O(n \log n)$ for both processes. Surprisingly, we also show that the complete graph is not the fastest when counting the number of steps precisely starting with a fixed mutant vertex. Our result also translates to the case with finite (sufficiently large) $r > 1$. This contributes to the open question of whether there exists a faster graph than the complete graph in terms of absorption time (when averaging over all vertices as possible starting mutant vertex) mentioned in [11] or stated as Open Problem 4 in [12].

In chapter four, we compute asymptotically precise fixation times for some specific graphs. We show that the fixation time on the cycle is $\Theta(n^2)$. On the complete graph, it is $\Theta(n \log n)$, the same as in the case of finite $r > 1$ [1]. The star graph is the slowest undirected graph we found with fixation time $\Theta(n^2 \log n)$, again the same as in the finite case. On the other hand, for the double star, the fixation time is asymptotically as fast as for the star graph. This is not true for the finite case, for which it can be shown that the fixation time of the double star is at least $\Omega(n^3)$ [10].

In the final chapter, we discuss the differences between the two notions of time. We show that the two processes are really different by providing an example of a graph on which the fixation times of the processes asymptotically differ. Next, we show that the limit time is not necessarily monotone (see the details in Section 5.2). With these observations, we argue why the continuous process is more reasonable than the limit one.

We conclude with a summary of our results. We also state some open problems, showing the direction of possible future research we find interesting.

2 Upper bounds

2.1 General upper bound

By Theorem 1, the limit process is always faster (or as fast as) the continuous process. Therefore, in this section, we prove the upper bounds for the continuous process, which proves the bounds for both processes.

Theorem 2. *Let G_n be a graph (directed or undirected) with n nodes. Then* $\mathrm{T}(G_n) \leq \frac{1}{2}n^3 - \frac{1}{2}n^2 = \frac{1}{2}n^3 + o(n^2)$.

Proof. We have at least one active vertex in the k-th stage and probability $\frac{1}{n}$ to pick this vertex. Let us denote by a the number of active edges, and by b the number of non-active edges incident to this vertex. We can have at most $k-1$ non-active edges from this vertex, and so $b \leq k-1$. The probability of picking an active edge incident to this vertex is thus $\frac{a}{a+b} \geq \frac{a}{a+k-1} \geq \frac{1}{k}$. In total, that gives us probability $\geq \frac{1}{n \cdot k}$ to gain a mutant in one step. The expected time until this happens is thus at most nk in the k-th stage. When we sum this over all stages, we get that the expected time is at most $\sum_{k=1}^{n-1} nk = n \cdot \frac{(n-1)\cdot n}{2} = \frac{1}{2}n^3 - \frac{1}{2}n^2$. □

Remark. As mentioned, the bound in Theorem 2 applies also to the limit process. Using the same arguments for the limit process, we can improve the constant in the bound to obtain $\mathrm{T_L}(G_n) \leq \frac{(n-1)\cdot n\cdot(2n-1)}{6} = \frac{1}{3}n^3 + o(n^3)$.

We proved that for every graph G_n on n vertices, the process will take at most $O(n^3)$ time. This contrasts with the classical Moran process with finite $r > 1$ which can be on some spatial structures exponential. Even more, the bound in Theorem 2 is exactly tight. We identify the slowest population structure that achieves this bound, concretely, the backward graph.

Definition 7 (Backward graph)**.** *For every n, consider a directed graph B_n defined as follows: vertices are denoted $\{1, 2, \ldots, n\}$, and there are forward and backward edges. Forward edges are of the form $(i, i+1)$ for every $i \leq n-1$. Backward edges lead between every pair of vertices (j, i) such that $i < j$ (see Fig. 2.1).*

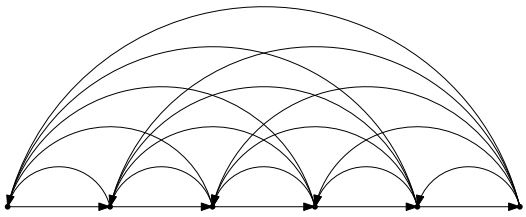


Figure 2.1 Backward graph B_6.

Theorem 3. *For every n there exists a directed graph G_n and an initial mutant node v of G_n such that* $\mathrm{T}(G_n, v) = \frac{1}{2}n^3 - \frac{1}{2}n^2 = \frac{1}{2}n^3 + o(n^2)$.

Proof. We set $G_n = \mathrm{B}_n$ and v as the node labeled 1 (the leftmost node). It is easy to see that in that case, we have to gain the vertices in the order $2, 3, \ldots, n$. In the k-th stage, we always have one active vertex k with $k-1$ non-active (backward) edges and only one active (forward) edge. Therefore, in k-th stage, the probability of gaining a new mutant is exactly $\frac{1}{nk}$ (that means the bounds used in Theorem 2 are satisfied with equality). And so the expected time until all vertices become mutants is in total exactly $\frac{n^2 \cdot (n-1)}{2} = \frac{1}{2}n^3 + o(n^2)$.

□

Moreover, we can show that asymptotically, the bound is tight also for DAGs. We use the half-path graph as an example of a DAG with the expected fixation time $O(n^3)$.

Definition 8 (Half-path graph)**.** *For every even $n = 2k$, we define a* half-path *graph (denoted as HP_n) as a DAG with vertices $\{1, 2, \ldots, n = 2k\}$, and edges leading as follows: The edges form a path in the first half of the vertices, meaning $(i, i+1)$ is an edge for $i \in \{1, 2, \ldots, k-1\}$. The second half is connected to every vertex in the first half, meaning we have an edge (i, j) for all $i \in \{1, 2, \ldots, k\}$, $j \in \{k+1, k+2, \ldots, 2k\}$ (see Fig. 2.2).*

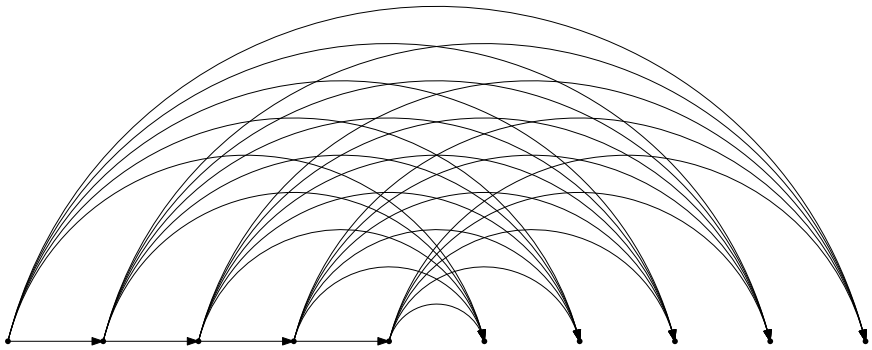


Figure 2.2 Half-path graph HP_{10}.

Theorem 4. *For every even n there exists a directed acyclic graph G_n and an initial mutant node v of G_n such that $\mathrm{T}(G_n, v) = \frac{1}{4}n^3 + o(n^2)$.*

Proof. We set $G_n = \mathrm{HP}_n$ and v as the node labeled 1 (the leftmost node). To get to the state where all vertices are mutants, we must follow the path in the first half of the vertices. This path has length $k-1$, and the probability of gaining one vertex on this path is $\frac{1}{n(k+1)}$, so the expected time is $n(k+1)$. Altogether the total expected time is at least $n(k+1)(k-1) = n(\frac{n}{2}+1)(\frac{n}{2}-1) = \frac{1}{4}n^3 + o(n^2)$.

□

We state two remarks.

First, in the proof, we don't care how long it takes for mutants to claim the second half of the vertices. Just looking at the first half of the vertices, it already takes $\Omega(n^3)$ time in expectation.

Second, for simplicity, the definition of the half-path graph considers only even n, but it is easy to generalize this also for odd n. For $n = 2k+1$, we can do the same construction as for $2k$ and then add the vertex $n+1$ to the second half

of the vertices, meaning that we also add the edges of the form $(i, n+1)$ such that $i \leq k$. Following the same arguments proves the same for this modified graph with odd n.

2.2 A stronger bound for regular graphs

As we showed in the previous section, the expected time can be as large as $\Omega(n^3)$. In this section, we prove stronger upper bounds for particular classes of graphs. First, we look at d-regular undirected graphs.

Theorem 5. *Let G_n be a regular undirected graph with n nodes. Then $\mathrm{T}(G_n) = O(n^2)$.*

Proof. The idea of the proof is as follows. We look at the stages in which the mutants conquer the graph. A stage is fast if there are many active edges compared to the regularity constant d. Hence, we care only about the slow stages with a few active edges. But at the end of a slow stage, we gain a new mutant such that necessarily most of this mutant's neighbors are residents. As these new active edges are all incident to one mutant vertex, their number can decrease only by one per stage. That means that in the next $\frac{d}{2}$ stages, the number of active edges will be large. Hence, we can aggregate the time spent on the one slow stage with the upcoming $\frac{d}{2}$ fast stages and conclude that, on average, we spent linear time per stage, thus giving us the desired bound $O(n^2)$.

More precisely, consider stage k and let e_k denote the number of active edges. As the graph is d-regular, the probability of gaining a mutant in stage k is $\frac{e_k}{d \cdot n}$. The idea is to distinguish two cases: Either e_k is large, and so the probability of gaining a mutant is large as well; Or e_k is small, but then the new mutant will have many active edges, so if we look at the expected time needed to gain several new mutants, the process will again be reasonably fast.

Formally, we distinguish two cases: either $e_k \geq \frac{d}{4}$ or $e_k < \frac{d}{4}$.

1. If $e_k \geq \frac{d}{4}$, then the probability of gaining a new mutant is $\frac{e_k}{d \cdot n} \geq \frac{\frac{d}{4}}{d \cdot n} = \frac{1}{4n}$.
2. If $e_k < \frac{d}{4}$, then $e_{k+1} \geq \frac{3}{4}d$ (because graph G is d-regular) and similarly $e_{k+2} \geq \frac{3}{4}d - 1$ and so on until $e_{k+\frac{d}{2}} \geq \frac{d}{4} + 1$. For us it will be sufficient to know that $\forall i \in \{1, 2, \ldots, \frac{d}{2}\}$ we have $e_{k+i} \geq \frac{d}{4}$. Then $\forall i \in \{1, 2, \ldots, \frac{d}{2}\}$ we get:
$$P_{k+i} = \frac{e_{k+i}}{d \cdot n} \geq \frac{\frac{d}{4}}{d \cdot n} = \frac{1}{4n},$$
and so $\mathbb{E}[t_{k+i}] = \frac{d \cdot n}{e_{k+i}} \leq 4n$. Also, as $\frac{d}{4} > e_k \geq 1$ we have $P_k = \frac{e_k}{d \cdot n} \geq \frac{1}{d \cdot n}$ and so $\mathbb{E}[t_k] \leq d \cdot n$. Now we can sum the expected times for the next $\frac{d}{2} + 1$ stages as:
$$\mathbb{E}[t_k] + \sum_{i=1}^{\frac{d}{2}} \mathbb{E}[t_{k+i}] \leq dn + \sum_{i=1}^{\frac{d}{2}} 4n = dn + 2dn = 3dn.$$

Now, we will compare the expected time of gaining a new mutant or mutants in both of these cases with the case when $e_k = \frac{d}{8}$ for every stage k.

1. In the first case, if $e_k = \frac{d}{8}$, then $P_k = \frac{\frac{d}{8}}{d \cdot n} = \frac{1}{8n}$ and so $\mathbb{E}[t_k] = 8n \geq 4n$ so we see that it is greater or equal to the result in the first case.

2. In the second case we have $P_{k+i} = \frac{\frac{d}{8}}{dn} = \frac{1}{8n}$ and so $\mathbb{E}[t_{k+i}] = 8n$. If we sum this through the $\frac{d}{2} + 1$ stages we get

$$\sum_{i=0}^{\frac{d}{2}} 8n = 4dn + 8n \geq 3dn.$$

   So again, we see that the expected time for gaining the next $\frac{d}{2} + 1$ mutants would be greater if it were the case $e_k = \frac{d}{8}$ for all k.

Altogether, we see that our d-regular graph is not slower than if it was the case that $e_k = \frac{d}{8}$ for every stage k. But if all $e_k = \frac{d}{8}$, then the stage k has expected time $8n$ and so expected time summed through all stages would be $\sum_{k=1}^{n-1} 8n = 8n(n-1) = O(n^2)$. And as our d-regular graph is not slower than this it follows that $\mathrm{T}(G_n) = O(n^2)$.

□

Note that the constant in the big O doesn't depend on the regularity constant d. Also, in general, this asymptotic bound is tight because as we will show in Theorem 10 it holds $T(C_n) = O(n^2)$, where C_n is the cycle on n vertices.

2.3 Relaxing the regularity condition

Next, we relax the condition of regularity to allow vertices to have degrees in some fixed range $[d, D]$.

Theorem 6. *For undirected graph G with minimum degree d and maximum degree D such that $D \leq c \cdot d$ for some constant $c \geq 1$ it holds $T(G_n) = O(n^2)$.*

Proof. Very similar to the previous proof. Again, we distinguish two cases $e_k \geq \frac{d}{4}$ and $e_k < \frac{d}{4}$.

1. If $e_k \geq \frac{d}{4}$, then the probability of gaining a new mutant is at least $\frac{e_k}{D \cdot n} \geq \frac{\frac{d}{4}}{cd \cdot n} = \frac{1}{4cn}$.

2. In the case $e_k < \frac{d}{4}$ let us denote $d + x$ to be the degree of the mutant gained after k–th stage. Clearly $x \geq 0$. Then $e_{k+1} \geq d + x - \frac{d}{4} = \frac{3}{4}d + x$. and $e_{k+\frac{d}{2}} \geq d + x - \frac{d}{4} - \frac{d}{2} = \frac{d}{4} + x$ and so again we get $\forall i \in \{1, 2, \ldots, \frac{d}{2}\}$:

$$P_{k+i} = \frac{e_{k+i}}{(d+x) \cdot n} \geq \frac{\frac{d}{4} + x}{(d+x) \cdot n} \geq \frac{1}{4n}.$$

   In the k–th stage, we have at least one active edge incident to the mutant with a degree at most D. So $P_k \geq \frac{1}{D \cdot n}$ and $\mathbb{E}[t_k] \leq D \cdot n$. If we again sum this together with $\mathbb{E}[t_{k+i}] \leq 4n$ we get similar sum as before:

$$\mathbb{E}[t_k] + \sum_{i=1}^{\frac{d}{2}} \mathbb{E}[t_{k+i}] \leq Dn + \sum_{i=1}^{\frac{d}{2}} 4n = Dn + 2dn \leq cdn + 2dn = (c+2)dn.$$

   The average time spent per one stage in this case is thus at most $\frac{(c+2)dn}{\frac{d}{2}+1}$.

Again we can compare it with the situation when $\forall k \in \{1, \ldots, n-1\} : e_k = \frac{d}{8c}$. In this case, the probability of gaining a new mutant in one stage is at most $\frac{\frac{d}{8c}}{dn} = \frac{1}{8cn}$. Hence, the expected time until this happens is at least $8cn$. We want to show that this time is slower than the average time at one stage in the original process. In the first case, we get the inequality $8cn \geq 4cn$, which is always satisfied. In the second case, we want to prove:

$$8cn \geq \frac{(c+2)dn}{\frac{d}{2}+1}.$$

That is equivalent to $8c \cdot dn + 16cn \geq (2c+4)dn$. And as $c \geq 1$, we have $8c \geq 2c+4$, and thus the inequality is satisfied. So again, we see that our graph is not slower than a graph with $e_k = \frac{d}{8c}$ for all k. In this setting, the probability of gaining a mutant at each step is at least $\frac{\frac{d}{8c}}{Dn} \geq \frac{1}{8c^2 n}$ and the expected time for one stage is thus $8c^2 n$. Because the original process is faster than this, we get $\mathrm{T}(G_n) = O(n^2)$.

□

2.4 A stronger bound for undirected graphs

We have seen that for regular undirected graphs, we can get an upper bound of $O(n^2)$. However, this cannot be generalized for all undirected graphs because, as we prove in Theorem 12, there exists an undirected graph, notably the star, on which the continuous process takes $\Omega(n^2 \log n)$ time in expectation. We prove a slightly weaker bound here.

Theorem 7. *Let G_n be an undirected graph with n nodes. Then $\mathrm{T}(G_n) \leq 4n^2\sqrt{n} + o(n^2\sqrt{n}) = O(n^2\sqrt{n})$.*

Proof. Let $d = \sqrt{n}$ be an auxiliary threshold value. Let us call any vertex with degree $\geq d$ a *large vertex*. Vertices that are not large are called *small*. We will prove that, on average, one stage takes roughly $n\sqrt{n}$ steps (up to a constant).

Formally, consider stage k and let e_k denote the number of active edges. We distinguish several cases.

1. Suppose that $e_k \geq \frac{1}{4}d$. For each active edge, the probability that the reproduction event happens along precisely that edge is at least $\frac{1}{n} \cdot \frac{1}{n-1} \geq \frac{1}{n^2}$. Thus, the probability of gaining a new mutant in the next step is at least $P_k \geq \frac{\frac{1}{4}d}{n^2} = \frac{d}{4n^2}$ and the expected time until this happens is $\mathbb{E}[t_k] \leq \frac{4n^2}{d} = 4n\sqrt{n}$, where in the inequality we used $d \geq \sqrt{n}$.

2. Suppose that $e_k < \frac{1}{4}d$. Then, we have three sub-cases depending on the situation of large resident vertices.

   (a) *There exists a large resident vertex with a mutant neighbor.*
   Call the large resident vertex v and its mutant neighbor u. Since $e_k < \frac{1}{4}d$, at least $d - e_k > \frac{3}{4}d$ of v's neighbors are residents. We will

prove that the expected time per stage over the next $1 + \frac{1}{2}d$ stages is at most $2n\sqrt{n}$.

First, we wait until v becomes mutant. The probability that edge (u, v) is selected for reproduction is at least $1/n^2$, which takes at most n^2 steps in expectation.

Second, once v becomes a mutant, we wait until at least $\frac{3}{4}d$ of v's neighbors are mutants. Note that some of them might have become mutants while we were waiting for v. Also, note that once $\frac{3}{4}d$ of v's neighbors are mutants, we have indeed gained at least $1 + \frac{3}{4}d - \frac{1}{4}d = 1 + \frac{1}{2}d$ mutants. As long as at most $\frac{3}{4}d$ of v's neighbors are mutants, at least $d - \frac{3}{4}d = \frac{1}{4}d$ of v's neighbors are residents, and therefore there are at least $\frac{1}{4}d$ active edges incident to v. The probability of gaining a mutant in one step through one of these edges is thus at least $\frac{1}{n} \cdot \frac{\frac{1}{4}d}{d} = \frac{1}{4n}$, and so the expected time is at most $4n$. In total, over the next $1 + \frac{1}{2}d$ stages we spend at most $1 \cdot n^2 + (\frac{1}{2}d) \cdot 4n = n^2 + 2nd$ steps. The average time per stage is thus at most

$$\frac{n^2 + 2nd}{\frac{1}{2}d + 1} \leq \frac{nd^2 + 2nd}{\frac{1}{2}d + 1} = \frac{(\frac{1}{2}d + 1)(2nd)}{\frac{1}{2}d + 1} = 2nd \leq 2n\sqrt{n},$$

where in the two inequalities, we used $n \leq d^2$ and $d \leq \sqrt{n}$.

(b) *Case* (a) *does not occur, and there exists a large resident vertex somewhere in the graph.*

We find and fix some shortest path $v_0, v_1, v_2, \ldots, v_l$ between a mutant vertex and some large resident vertex. That is, v_0 is a mutant vertex, $v_1, v_2, \ldots, v_{l-1}$ are small resident vertices, and v_l is a large resident vertex. Then we wait until $v_1, \ldots, v_l$ and at least half of v_l neighbors become mutants. Note that since we consider the shortest path, vertex v_l initially has no mutant neighbors.

We proceed as in case (a). First, node v_1 becomes mutant in at most n^2 steps in expectation. Since vertices $v_1, \ldots, v_{l-1}$ are small, each vertex $v_2, \ldots, v_l$ becomes mutant in at most $1/(\frac{1}{n} \cdot \frac{1}{d}) = nd$ steps in expectation. Once the large vertex v_l becomes mutant, we wait until at least $d/2$ of its neighbors are mutants. Using the same method as in (a), we aggregate the first "slow" stage (in which v_1 becomes mutant) and the remaining $(l - 1) + \frac{1}{2}d \geq \frac{1}{2}d$ "fast" stages. By the same algebra as in (a), the average time per stage will again be at most $2nd \leq 2n\sqrt{n}$ in expectation.

(c) *Cases* (a) *and* (b) *do not occur; that is, all resident vertices are small.*

Once this happens, we aggregate the time spent from this moment on until all vertices become mutants. Therefore, this case happens only once. Suppose it happens at stage k. We know that at this point $e_k < \frac{1}{4}d$. First, we wait until all these e_k edges are used for reproduction. Each one of them is used after at most n^2 steps in expectations, so in total, we wait at most $e_k \cdot n^2 \leq \frac{1}{4}n^2 d \leq \frac{1}{4}n^2\sqrt{n}$ steps in expectation. From that point on, all active vertices will be small vertices. As we computed in (b), the expected time until we use an edge from a small vertex is at most $nd \leq n\sqrt{n}$.

In each case, we spend, on average, at most $\max(4n\sqrt{n}, 2n\sqrt{n}, n\sqrt{n})$ steps per one stage. Plus, once during the process, we wait at most $\frac{1}{4}n^2\sqrt{n}$ steps in case (c). Since there are $n-1 < n$ stages, the total expected time is at most $(n-1)\cdot 4n\sqrt{n} + \frac{1}{4}n^2\sqrt{n} = 4n^2\sqrt{n} + o(n^2\sqrt{n}) = O(n^2\sqrt{n})$.

□

3 Lower bounds

3.1 Lower bound for the continuous process

First, we prove a general lower bound for both directed and undirected graphs.

Theorem 8. *Let G_n be a graph (directed or undirected) with n nodes and v any initial mutant node. Then* $\mathrm{T}(G_n, v) \geq n \cdot \mathrm{H}_{n-1} = \Omega(n \log n)$.

Proof. To gain a new mutant, we must first pick a mutant vertex. In stage k, the probability of picking a mutant is $\frac{k}{n}$ and therefore $P_k \leq \frac{k}{n}$ and so $\mathbb{E}[t_k] \geq \frac{n}{k}$. When we sum this up over the $n-1$ stages, we get

$$\mathrm{T}(G_n) \geq \sum_{k=1}^{n-1} \mathbb{E}[t_k] \geq \sum_{k=1}^{n-1} \frac{n}{k} = n \log n + o(n \log n).$$

□

3.2 Lower bound for the limit process

Note that as we showed in Theorem 1, the limit process is always faster than the normal process. So, when proving upper bounds, it was sufficient to prove them only for the continuous process. When proving lower bounds, we consider both processes separately. In the next theorem, we prove that the lower bound for the limit process is asymptotically the same as in the previous theorem. As a useful tool in the proof, we will use the notion of the temperature of a vertex.

Definition 9. *Let $G_n = (V, E)$ be a graph. A* temperature *of a vertex u, denoted as* $\mathrm{temp}(u)$*, is defined as* $\mathrm{temp}(u) = \frac{1}{n} \cdot \sum_{v,(v,u)\in E} \frac{1}{\deg(v)}$.

We note that $\sum_{u \in V} \mathrm{temp}(u) = 1$. The idea behind defining the temperature is to measure how often a particular vertex is being replaced. Or, in other words, up to a constant, how long would it take for this vertex to become mutant if all its neighbors were already mutants.

Theorem 9. *Let G_n be a graph (directed or undirected) with n nodes and v any initial mutant node. Then* $\mathrm{T_L}(G_n, v) = \Omega(n \log n)$.

Proof. Let us refer to a vertex with temperature $< \frac{3}{n}$ as a *cold* vertex and denote the number of cold vertices by c. The remaining $n - c$ vertices with temperatures $\geq \frac{3}{n}$ are called *hot*. Let S be the sum of temperatures of all hot vertices. Since $S \subseteq V$ we get $S \leq 1$. Therefore

$$\begin{aligned} 1 \geq S &\geq (n-c) \cdot \frac{3}{n} \\ n &\geq 3n - 3c \\ c &\geq \frac{2}{3} n \end{aligned}$$

Let us now look at the process of gaining mutants and stop it at the moment when there are exactly $\frac{n}{2}$ mutants. As $c \geq \frac{2}{3}n$ there are at least $\frac{2}{3}n - \frac{1}{2}n = \frac{1}{6}n$ resident vertices which are cold.

Now, we make the following observation: In stage k, the probability that a vertex u becomes a mutant in the next step is at most $\frac{n}{k} \cdot \mathrm{temp}(u)$. To prove the observation, let M_k^u be the set of all mutant neighbors of u in stage k. Then the probability of gaining u in the next step is $\sum_{v \in M_k^u} \left(\frac{1}{k} \cdot \frac{1}{\deg(v)}\right)$ which is less than or equal to the case when we sum this throughout all neighbors of u to get $\frac{1}{k} \cdot \sum_{v,(v,u) \in E(G_n)} \frac{1}{\deg(v)} = \frac{n}{k} \cdot \mathrm{temp}(u)$. Hence, the observation is proved.

Let us look at the second half of the stages. In stage $k \geq \frac{n}{2}$ the probability that in the next step we gain a fixed cold vertex u is $\leq \frac{n}{k} \cdot \mathrm{temp}(u) \leq \frac{n}{\frac{n}{2}} \cdot \frac{3}{n} = \frac{6}{n}$. Thus, after half of the vertices become mutants, we have at least $\frac{n}{6}$ events (cold vertices becoming mutants) each with probability $\leq \frac{6}{n}$, and we have to wait until all of these events happen. The best scenario is when there are only $\frac{n}{6}$ such vertices, and each of them has probability $\frac{6}{n}$. The expected time until all of these events happen can be then computed by using a coupon collector approach. Suppose that i of these events already happened and let t_i be the expected time until one of the remaining $n-i$ events occurs. In this case, we have the probability that one of the non-happened events occurs to be $p_i = \left(\frac{n}{6} - i\right) \cdot \frac{6}{n}$. Thus $t_i = \frac{1}{p_i} = \frac{n}{6 \cdot \left(\frac{n}{6} - i\right)}$. By linearity of expectation, the expected time until all of the events happen is:

$$t_0 + t_2 + \cdots + t_{n-1} = \sum_{i=0}^{\frac{n}{6}-1} \frac{n}{6 \cdot \left(\frac{n}{6} - i\right)} = \frac{n}{6} \sum_{i=0}^{\frac{n}{6}-1} \frac{1}{\frac{n}{6} - i} = \frac{n}{6} \cdot \mathrm{H}_{\frac{n}{6}} = \Omega(n \log n).$$

Hence, the expected time until all these events happen is $\Omega(n \log n)$.

□

3.3 A graph faster than the complete graph

In this section, we describe a graph and a starting node such that the absorption time is shorter than that on the complete graph in the continuous Moran process. Note that in the continuous process, the fixation time and absorption time are the same quantity because the process always ends with a fixation. Finally, we show that this result translates also to the classical Moran process with sufficiently large finite r.

Consider the following graph G_{11} with 11 vertices. The graph consists of three layers such that the first two layers are independent sets of size 1 and 4 the third one of size 6 is a complete graph. The edges between consecutive layers create a complete bipartite graph. Let us denote the only vertex in the first layer as v. See Fig. 3.1.

Given a graph and a starting node, the fixation or absorption time can be computed exactly using numerical computation (see for example [13]). If we compute this for the graph G_{11} and starting node v we find that $\mathrm{T}(G_{11}, v) \approx 58.4975$ while for the complete graph on 11 vertices it is $\mathrm{T}(K_{11}) \approx 58.5793$. So there exists a graph and a vertex v such that the continuous process on this graph starting from v is faster in expectation than on the complete graph.

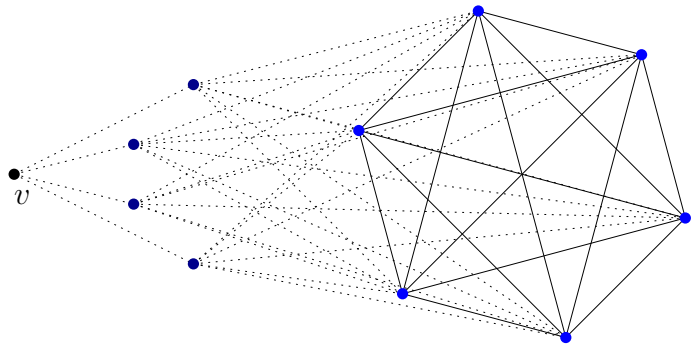


Figure 3.1 A 3-layer graph G_{11} with layers of sizes $1, 4, 6$.

Similarly, for the limit process, we get $\mathrm{T_L}(G_{11}) \approx 28.6494$ while $\mathrm{T_L}(K_{11}) \approx 29.2896$. Because the limit process corresponds to taking the limit of the mutation fitness $r \to \infty$ in the classical Moran process, this result also translates into the Moran process with finite $r > 1$. That is, when taking suitably large r the absorption time on the complete graph K_{11} is larger than on the graph G_{11} in the classical Moran process. In figure Fig. 3.2 we can see the comparison between these two absorption times plotted as a function of the mutation fitness r.

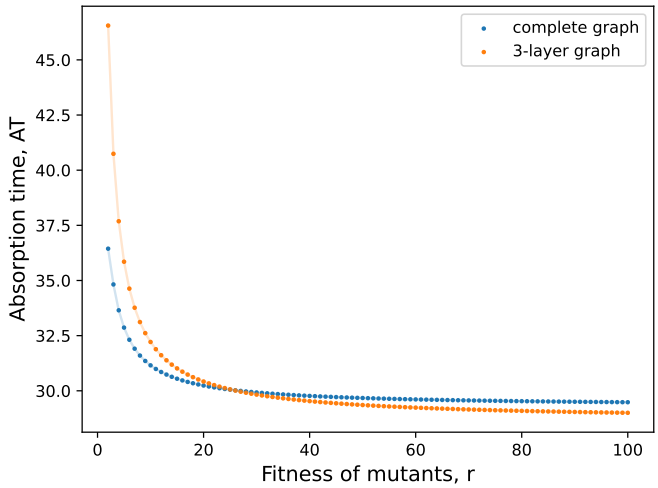


Figure 3.2 Comparison between absorption times on complete graph K_{11} and the 3-layer graph G_{11}. The plotted values of r are integers between 2 and 100.

The question of whether the complete graph is the fastest in terms of absorption or fixation time has been raised before. For example, in [12], the stated Open Problem 4 asks the question of whether there exist graphs that are faster than the complete graph in terms of the absorption time. Here, we found a graph and a particular starting mutant vertex such that it is indeed faster. Note that the overall absorption time (which is average through all vertices) on this particular graph G_{11} is not faster than on the complete graph.

4 Specific graphs

In this chapter, we show how fast is the continuous process for some specific graphs. We analyze cycles, cliques, star graphs, double stars, and total order graphs.

4.1 Cycle

Theorem 10 (Cycle)**.** *Let* C_n *be a cycle with* n *nodes. Then* $\mathrm{T}(\mathrm{C}_n) = \Theta(n^2)$.

Proof. At each stage, the mutants cover a consecutive segment of the cycle. Hence, at every time, there are exactly two active edges connecting the ends of the mutant segment with the rest of the cycle. To pick an active edge we need to select the incident mutant which is done with probability $\frac{1}{n}$ and then pick this particular edge. But because the degree of each node is 2 this has probability $\frac{1}{2}$. The two active edges are thus picked with probability $2 \cdot \frac{1}{n} \cdot \frac{1}{2} = \frac{1}{n}$ independently of the stage number. The expected time until we gain a new mutant is then always n. Because we have $n-1$ stages, in total the expected time is $n(n-1) = \Theta(n^2)$. □

4.2 Clique

Theorem 11 (Complete graph)**.** *Let* K_n *be a complete graph with* n *nodes. Then* $\mathrm{T}(\mathrm{K}_n) = 2(n-1)\,\mathrm{H}_{n-1} = \Theta(n \log n)$.

Proof. In the k-th stage, we have the probability $\frac{k}{n}$ of picking a mutant vertex. Every mutant vertex then has $n-k$ active edges. Hence, the probability of picking one is $\frac{n-k}{n-1}$. In the k-th stage, the probability of gaining a mutant is thus $\frac{k \cdot (n-k)}{n \cdot (n-1)}$ and the expected time until this happens is $\mathbb{E}[t_k] = \frac{n \cdot (n-1)}{k \cdot (n-k)} = (n-1) \cdot (\frac{1}{k} + \frac{1}{n-k})$. Summing this over all stages, we get:

$$\mathrm{T}(\mathrm{K}_n) = \sum_{k=1}^{n-1} (n-1) \cdot \left(\frac{1}{k} + \frac{1}{n-k} \right) = (n-1) \cdot 2 \cdot \sum_{k=1}^{n-1} \frac{1}{k} = 2n \log n + o(n \log n).$$

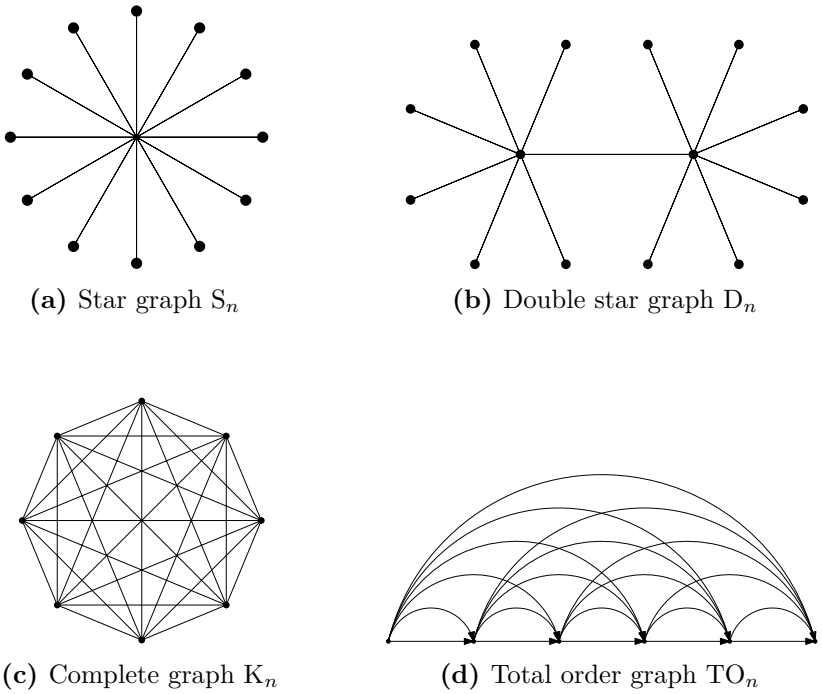
□

4.3 Star and double star graphs

Definition 10 (Star graph)**.** *For every* $n = k+1$*, we define a* star *graph (denoted as* S_n*) as an undirected graph with vertices* $\{1, 2, \ldots, n = k+1\}$*. We call the vertex* $k+1$ *to be the center and all the other vertices to be leaves. The edges connect all leaves to the center, meaning that for every* $i \in \{1, \ldots, k\}$ *the pair* $(i, k+1)$ *is an edge (see Fig. 4.1a).*

The double star graph can be obtained by gluing two copies of the same star graph together.

Definition 11 (Double star graph)**.** *For every even $n = 2k + 2 \geq 4$, we define a* double star *graph (denoted as D_n) as an undirected graph with vertices $\{1, 2, \ldots, n = 2k + 2\}$. This graph is obtained by taking two star graphs, one with center $k + 1$ and leaves $1, 2, \ldots, k$ and the other one with center $2k + 2$ and leaves $k + 2, \ldots, 2k + 1$ and joining them by an edge connecting their centers $(k + 1, 2k + 2)$ (see Fig. 4.1b).*



(a) Star graph S_n **(b)** Double star graph D_n

(c) Complete graph K_n **(d)** Total order graph TO_n

Figure 4.1 Specific graphs.

Next, we prove the asymptotic fixation times on star and double star graphs. Note that the star graph is the slowest unoriented graph we found for the continuous process. We also prove that the double star is asymptotically as fast as the star. However, for the Moran process with finite $r > 1$, this is not true. In the finite case, it can be proven that the star takes $\Theta(n^2 \log n)$ time and double star at least $\Omega(n^3)$.

Theorem 12 (Star graph)**.** *For a star graph S_n with n nodes and any initial mutant node v we have $\mathrm{T}(\mathrm{S}_n, v) = \Theta(n^2 \log n)$.*

Proof. The first stage takes n steps on average. After that, it is always the case that mutants occupy the center of the star and one leaf, no matter where the initial mutant started. In the k-th stage ($k \geq 2$), there are $k - 1$ mutant leaves and $n - k$ non-mutant leaves. The probability of gaining a new vertex is therefore $\frac{1}{n} \cdot \frac{n-k}{n-1}$. The expected time until this happens is thus $\frac{n \cdot (n-1)}{n-k}$. If we sum this up over all the stages, we get the following:

$$\mathrm{T}(\mathrm{S}_n) = \sum_{k=1}^{n-1} \frac{n \cdot (n-1)}{n-k} = n \cdot (n-1) \cdot \sum_{k=1}^{n-1} \frac{1}{n-k}$$
$$= n \cdot (n-1) \cdot \sum_{k=1}^{n-1} \frac{1}{k} = n^2 \log n + o(n^2 \log n).$$

□

Theorem 13 (Double star)**.** *For a double star graph* D_n *with* $n = 2k+2 \geq 4$ *nodes and any initial mutant node* v *we have* $\mathrm{T}(\mathrm{D}_n, v) = \Theta(n^2 \log n)$.

Proof. Let us denote the two centers of the double star as x_1 and x_2. WLOG, we start in the left half of the graph. Let us divide the process into two phases.

The first phase will end when we gain the vertex x_2, and the rest of the process will be the second phase. The expected time of the first phase is surely $\leq \frac{n}{k+2} \cdot \mathrm{T}(\mathrm{S}_{k+2}) \leq 2 \cdot \mathrm{T}(\mathrm{S}_n) = \Theta(n^2 \log n)$, because we can look at the left star with center x_1 and its neighbors (including x_2) as a star graph with $k+2$ vertices. The constant $\frac{n}{k+2}$ is there to adjust to the fact that the probability of picking a particular vertex in the double star is $\frac{1}{n}$ whilst it is $\frac{1}{k+2}$ in the star S_{k+2}.

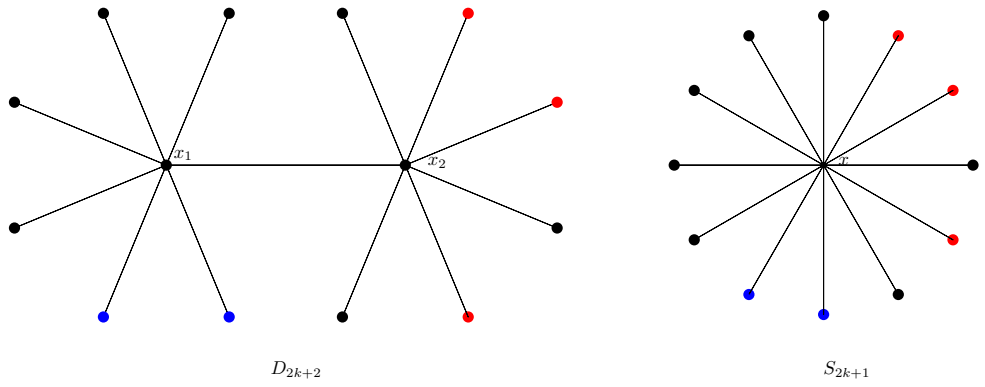


Figure 4.2 A double star D_{2k+2} and its comparison to star S_{2k+1} in the second phase.

In every stage in the second phase, the vertices x_1 and x_2 are already mutants and also some leaves. Suppose there are a mutant leaves next to x_1 and b mutant leaves next to x_2. We want to compare this process with the process on a star graph S_{2k+1} with mutant center x and $a+b$ mutant leaves. Then, the probability of gaining one more mutant in the next step in the double star case is:

$$\frac{1}{n} \cdot \frac{k-a}{k+1} + \frac{1}{n} \cdot \frac{k-b}{k+1} = \frac{2k-a-b}{n(k+1)}.$$

On the other hand in the star graph S_{2k+1} in stage $m+n+1$ the probability is:

$$\frac{1}{n-1} \cdot \frac{2k-a-b}{2k}.$$

We would like to know when the first probability is larger, for that we can compare only the denominators:

$$\begin{aligned}\frac{1}{n(k+1)} &\geq \frac{1}{(n-1)\cdot 2k}\\ (n-1)\cdot 2k &\geq n(k+1)\\ 2nk-2k &\geq nk+n\\ nk-2k-n &\geq 0\\ (n-2)\cdot(k-1) = 2k\cdot(k-1) &\geq 2\end{aligned}$$

The last inequality is true whenever $k > 1$. The probability of gaining a mutant in the double–star case is thus always greater than the probability of gaining a mutant in the star graph. So altogether, we see that the second phase has expected time at most $\mathrm{T}(\mathrm{S}_{2k+1}) = \mathrm{T}(\mathrm{S}_{n-1}) = \Theta(n^2 \log n)$. Both phases of the double star graph together thus have expected time $O(n^2 \log n)$.

For the lower bound it is sufficient to notice that the double star contains the star graph S_{k+2} as a subgraph (with center x_1 and its neighbors) as we did in the first phase above. The fixation time of the double star is thus at least the expected time until mutants gain this subgraph S_{k+2}. This we already computed in the upper bound of the first phase and it is $\frac{n}{k+2} \cdot \mathrm{T}(\mathrm{S}_{k+2}) = \Theta(n^2 \log n)$.

□

4.4 Total order graph

Definition 12 (Total order graph)**.** *For every n, we define a* total order *graph (denoted as TO_n) as a directed graph with vertices $\{1, 2, \ldots, n\}$ and edges (i, j) for every $i, j \in \{1, 2, \ldots, n\}$ such that $i < j$ (see Fig. 4.1d).*

Theorem 14 (Total order)**.** *For a total order graph TO_n with n nodes and a mutant node v being the first node we have $\mathrm{T}(\mathrm{TO}_n, v) = \Theta(n^2)$.*

Proof. First, we prove the lower bound $\mathrm{T}(\mathrm{TO}_n) = \Omega(n^2)$. The probability that in a single step, the first resident vertex becomes mutant is $\frac{1}{n\cdot(n-1)}$. The expected time until this happens is $n \cdot (n-1) = \Omega(n^2)$. Thus $\mathrm{T}(\mathrm{TO}_n) = \Omega(n^2)$ as required.

Second, we prove the upper bound $\mathrm{T}(\mathrm{TO}_n) = O(n^2)$. It is sufficient to prove the bound for n of the form $n = 2^k$, so suppose our n is a power of 2. We divide our n vertices into $k+1$ blocks from left to right, and we index them from zero (as block $0, 1, 2, \ldots, k$). The sizes of the blocks will be $1, 1, 2, 2^2, 2^3, \ldots, 2^{k-1}$. We begin with the first vertex; therefore, the zero block is already mutant. In step i, we wait until the vertices in the i-th block become mutants. There are 2^{i-1} vertices in the block i and 2^{i-1} vertices in the previous blocks $0, 1 \ldots, i-1$. Therefore, before any mutant in the i-th block is present, there are $2^{i-1} \cdot 2^{i-1}$ active edges incident to i-th block. The probability of using one particular edge is at least $\frac{1}{n^2}$ and so the probability of gaining one vertex from the i-th block is at least $\frac{2^{i-1}\cdot 2^{i-1}}{n^2}$. After gaining one mutant we have $2^{i-1} \cdot (2^{i-1} - 1)$ active edges incident to i-th block and generally after gaining m mutants this number of active edges is $2^{i-1} \cdot (2^{i-1} - m)$. The probability of gaining one more mutant in this situation

is thus at least $\frac{2^{i-1}\cdot(2^{i-1}-m)}{n^2}$ so the expected time until this happens is at most $\frac{n^2}{2^{i-1}\cdot(2^{i-1}-m)}$. Together the expected time until all vertices in the i-th block become mutants is then at most:

$$\sum_{m=0}^{2^{i-1}-1} \frac{n^2}{2^{i-1}\cdot(2^{i-1}-m)} = \frac{n^2}{2^{i-1}} \cdot \sum_{m=0}^{2^{i-1}-1} \frac{1}{(2^{i-1}-m)} = \frac{n^2}{2^{i-1}} \cdot \sum_{m=1}^{2^{i-1}} \frac{1}{m} =$$
$$= \frac{n^2}{2^{i-1}} \cdot \mathrm{H}_{2^{i-1}} \approx \frac{n^2}{2^{i-1}} \cdot (i-1)$$

And if we sum this over all the blocks, the expected time until all vertices become mutants is at most:

$$\sum_{i=1}^{k} \frac{n^2}{2^{i-1}} \cdot (i-1) = n^2 \cdot \sum_{i=0}^{k-1} \frac{i}{2^i} \leq n^2 \cdot \sum_{i=0}^{\infty} \frac{i}{2^i} = 2n^2 = O(n^2).$$

□

5 Comparison of the two notions of time

In this chapter, we show that the two notions of time (limit time and continuous time) are substantially different. In particular, we prove that on some graphs, the continuous time is asymptotically larger than the limit one. Moreover, we show that the limit time is not monotone in the sense described below.We interpret this as an indication that perhaps the continuous time is more natural.

To illustrate these two claims, we use the lollipop graph as an example.

Definition 13 (Lollipop graph)**.** *For every n, consider an undirected graph L_n defined as follows: vertices are denoted $\{1, 2, \ldots, n\}$, and there are two parts of the graph. The first part consists of vertices $1, 2, \ldots, \sqrt{n}$ and a path between them, meaning there are edges $(i, i+1); \forall i \in \{1, 2, \ldots, \sqrt{n} - 1\}$. The second part of the graph is a clique, meaning there are edges $(i, j); \forall i, j \in \{\sqrt{n} + 1, \ldots, n\}$. These two parts are then connected by an edge $(\sqrt{n}, \sqrt{n} + 1)$ (see Fig. 5.1).*

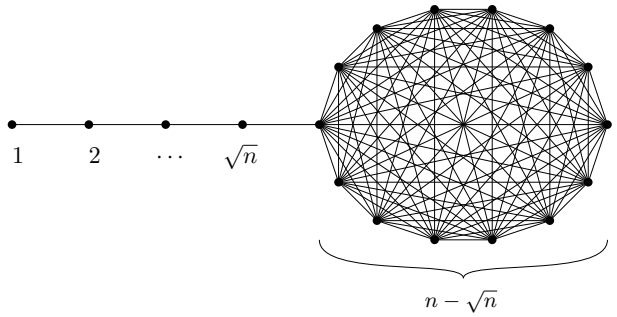


Figure 5.1 A lollipop graph L_n on n nodes consists of a path of length $\sqrt{n}$ that is connected to a clique of size $n - \sqrt{n}$.

5.1 The two notions of time are different

First, we show that the limit time and the continuous time are substantially different on this lollipop graph. The intuition is that when we fix a particular state with k mutants, depending on the value of k the two processes become more similar as k gets larger. That is for the continuous process the probability of picking a mutant in stage k is $\frac{1}{n}$ in comparison with $\frac{1}{k}$ for the limit process. Hence, if k is large (of order n), then these processes have probabilities scaled just by some constant, but when k is small (sublinear to n) there is an asymptotic difference between these probabilities. Hence, the limit process can be asymptotically faster in these steps. With this intuition, the lollipop graph is chosen exactly in a way to ensure this as we will see in the proof of the following theorem.

Theorem 15. *There exists a graph G_n and an initial mutant node v of G_n such that $\mathrm{T}(G_n, v) = \Theta(n\sqrt{n})$ while $\mathrm{T}_{\mathrm{L}}(G_n, v) = \Theta(n \log n)$.*

Proof. We set $G_n = \mathrm{L}_n$ and v as the node labeled 1 (the leftmost node). To get to the state where all vertices are mutants, we must first follow the path and then spread through the whole clique.

First, consider the continuous process. Vertex i in the first part will propagate with probability $\frac{1}{2n}$ if $i \neq 1$ and with probability $\frac{1}{n}$ if $i = 1$. Thus, the expected time until the node $\sqrt{n}+1$ becomes mutant is $n + \sum_{i=2}^{\sqrt{n}} 2n = 2n\sqrt{n} - n$. Then, similarly to Theorem 11, we can compute the expected time until all nodes become mutants. When there are k mutants already in the clique, the probability of gaining a new one is $\frac{k}{n} \cdot \frac{n-\sqrt{n}-k}{n-\sqrt{n}-1}$, thus the expected time until this happens is $\frac{n}{k} \cdot \frac{n-\sqrt{n}-1}{n-\sqrt{n}-k} = n \cdot \frac{n-\sqrt{n}-1}{n-\sqrt{n}} \cdot \left(\frac{1}{k} + \frac{1}{n-\sqrt{n}-k}\right)$. Summing this over all k we get:

$$n \cdot \frac{n-\sqrt{n}-1}{n-\sqrt{n}} \cdot \sum_{k=1}^{n-\sqrt{n}-1} \left(\frac{1}{k} + \frac{1}{n-\sqrt{n}-k}\right) = \\ = n \cdot \frac{n-\sqrt{n}-1}{n-\sqrt{n}} \cdot 2\,\mathrm{H}_{n-\sqrt{n}-1} \approx 2n \log n.$$

For the continuous process, the first part of the graph thus takes $\Theta(n\sqrt{n})$ steps, and the second part takes $\Theta(n \log n)$ steps in expectation. Altogether, the process takes $\Theta(n\sqrt{n})$ steps in expectation as claimed.

For the limit process, we proceed similarly. In the first part, vertex i will propagate with probability $\frac{1}{2i}$ if $i \neq 1$ and with probability 1 if $i = 1$. Thus, in total the time until the node $\sqrt{n}+1$ becomes mutant is in expectation $1 + \sum_{i=2}^{\sqrt{n}} 2i = \sqrt{n} \cdot (\sqrt{n}-1) - 1 = n - \sqrt{n} - 1 = \Theta(n)$.

In the second part, the expected time can be computed in exactly the same way as for the continuous process, only with the difference that when we have k mutants in the clique already, the probability of picking a mutant node in the clique is $\frac{k}{\sqrt{n}+k}$ instead of $\frac{k}{n}$. Following the same algebraic modifications, we get that the total expected time for the limit process in the second part is

$$\frac{n-\sqrt{n}-1}{n-\sqrt{n}} \cdot \sum_{k=1}^{n-\sqrt{n}-1} (\sqrt{n}+k) \cdot \left(\frac{1}{k} + \frac{1}{n-\sqrt{n}-k}\right) = \\ \frac{n-\sqrt{n}-1}{n-\sqrt{n}} \cdot \left(2\sqrt{n} \cdot \mathrm{H}_{n-\sqrt{n}-1} + \sum_{k=1}^{n-\sqrt{n}-1} \left(1 + \frac{k}{n-\sqrt{n}-k}\right)\right) = \\ \frac{n-\sqrt{n}-1}{n-\sqrt{n}} \cdot \left(2\sqrt{n} \cdot \mathrm{H}_{n-\sqrt{n}-1} + \sum_{k=1}^{n-\sqrt{n}-1} \frac{n-\sqrt{n}}{k}\right) = \\ \frac{n-\sqrt{n}-1}{n-\sqrt{n}} \cdot (n+\sqrt{n}) \cdot \mathrm{H}_{n-\sqrt{n}-1} = \\ \approx n \log n.$$

For the limit process, the first part of the graph thus takes $\Theta(n)$ step and the second part takes $\Theta(n \log n)$ steps in expectation. Altogether, the limit process takes $\Theta(n \log n)$ steps in expectation as claimed.

□

5.2 Limit time doesn't have to be monotone

Next, we prove that the limit time is not necessarily monotone. That is, we show that if we add additional mutants to the starting configuration, the expected

time until fixation might increase, even asymptotically. When considering fixation probability rather than the absorption time, this kind of monotonicity is sometimes called *subset domination* and it is known that it is satisfied by the Moran process with finite $r > 1$ on undirected graphs [7].

Theorem 16. *There exists a directed graph G_n, an initial mutant node v of G_n and a subset of vertices $\{v\} \subseteq X \subseteq V(G_n)$ such that $\mathrm{T_L}(G_n, v) = \Theta(n \log n)$ while $\mathrm{T_L}(G_n, X) = \Theta(n\sqrt{n})$.*

Proof. Let us consider a directed lollipop graph L'_n, which is obtained from L_n by orienting the edges in the first part from left to right, meaning we will have directed edges $(i, i+1); \forall i \in \{1, 2, \ldots, \sqrt{n}\}$ on the path. The edges in the clique remain undirected. Then we set $G_n = \mathrm{L}'_n$,, the starting node v as the node labeled 1 (the leftmost node), and $X = \{\sqrt{n}+1, \ldots, n\} \cup \{v\}$. In other words, X is a set containing all vertices of the clique and the starting node.

It is easy to see that $\mathrm{T_L}(\mathrm{L}'_n, v)$ is almost the same as $\mathrm{T_L}(\mathrm{L}_n, v)$, which we computed in the proof of Theorem 15. The only difference is that along the path we gain a new vertex with probability $\frac{1}{n}$ instead of $\frac{1}{2n}$. Asymptotically this is the same. Hence $\mathrm{T_L}(\mathrm{L}'_n, v) = \Theta(n \log n)$.

Now let us look at $\mathrm{T_L}(\mathrm{L}'_n, X)$. Since no edge is going out of the mutant clique, the only way to turn all vertices into mutants is to follow the path in the first half of the graph. As the out-degree of the nodes in this path is one, to infect the $(i+1)$-th node on this path, we need to pick its mutant predecessor i. If node i is the last mutant along the path, it is picked with probability $\frac{1}{n-\sqrt{n}+i}$. Thus the expected time until it happens is $n - \sqrt{n} + i$. Summing this over all vertices on the path, we get that the total time until all vertices become mutants in this setting is:

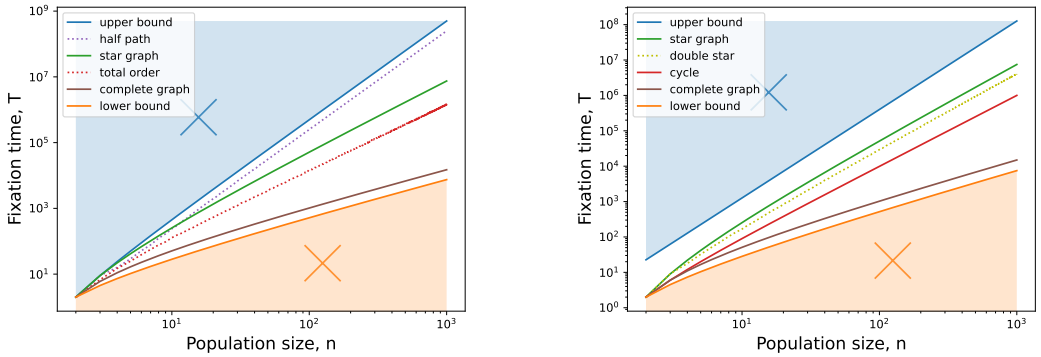
$$\mathrm{T_L}(\mathrm{L}'_n, X) = \sum_{i=2}^{\sqrt{n}} (n - \sqrt{n} + i) = (n - \sqrt{n}) \cdot (\sqrt{n} - 1) + \frac{\sqrt{n} \cdot (\sqrt{n}+1)}{2} - 1 = \Theta(n\sqrt{n}).$$

□

Let us remark that considering the directed lollipop, starting with the clique being mutant using the limit process is asymptotically as fast as considering the undirected lollipop and using the continuous process with only one mutant vertex. That makes sense because, for the continuous case, the first part of the lollipop was the asymptotically slowest part. And as we previously observed the probabilities of picking a mutant vertex in k-th stage between those two processes become more similar as k gets larger. As we start with $n - \sqrt{n} + 1$ mutants in the limit case, these probabilities are already quite similar.

Conclusion

In this thesis, we studied the Moran process in the limit $r \to \infty$. We introduced two possible notions of time, the continuous one and the limit one, and stated that we will use mostly the continuous one. We proved the upper bound of $O(n^3)$ for the fixation time on arbitrary spatial structures. And we also identified the exact slowest graph. Then we improved the bound to $O(n^2)$ for regular graphs and to $O(n^2\sqrt{n})$ for all undirected graphs. Note that for fixed finite $r > 1$ some undirected graphs are slower than this, namely $\Omega(n^3)$. We continued by showing the lower bound $O(n \log n)$ for both processes. Moreover, we provided an example of a graph and a starting vertex such that its fixation time is faster than on the complete graph, both in the setting of limit $r \to \infty$ and for sufficiently large finite r. Then we computed the fixation times on some specific graphs and obtained $\Theta(n^2)$ for cycles, $\Theta(n \log n)$ for complete graphs, $\Theta(n^2 \log n)$ for stars and double stars, and $\Theta(n^2)$ for total order graphs. In the end, we discussed the differences between the two notions of time and argued why we think the continuous one is more natural. The results are summarized in Fig. 5.2. The lower orange area corresponds to everything below the lower bound and the upper blue area corresponds to everything above the proved upper bound. Thus there is no graph with fixation time in those areas.



(a) directed graphs **(b)** undirected graphs

Figure 5.2 Summary of the proved fixation times. Lines correspond to the plotting of the exact results proved in this thesis, dots correspond to numerical simulations. Both plots are log-scaled.

5.3 Open questions

To conclude we discuss open questions. In general, we proved some upper and lower bounds and for some of them we were able to show that they are tight. However, there are still some bounds with a notable gap that could be improved.

We proved in Theorem 7 that the continuous process on every undirected graph takes $O(n^2\sqrt{n})$ time in expectation. We also showed in Theorem 12 an example of a graph (the star graph) on which it takes $\Theta(n^2 \log n)$. So there is a notable gap between this lower and upper bound. Is it really so that the star

graph is the slowest one among the undirected graphs? If so, can it be proven that the upper bound is also $O(n^2 \log n)$? And if not, what is the slowest undirected graph instead? Note that double star, which is the slowest known undirected graph in case of finite $r > 1$, is also only $O(n^2 \log n)$. The computer simulations on small graphs don't show any other good candidate for the slowest graph, however, that doesn't imply anything when considering the asymptotics for large n.

Open problem 1. *What is asymptotically the slowest undirected graph for the continuous process?*

Another direction is to consider the fastest graphs. As we showed in Section 3.3 there exists a graph and a node v such that the absorption time when starting from this node v is shorter than that on the complete graph. Both when considering the continuous process and the Moran process with sufficiently large finite r. There appear to be many interesting questions worth exploring in this direction. For instance:

Open problem 2. *What is the fastest graph for the continuous process?*

We found a particular 3-layer graph G_{11} on 11 vertices that is faster than K_{11}. This result then translates into the classical Moran process with sufficiently large r. Thus answering this open question might also help with understanding the finite case which is generally more complicated.

Maybe a bit easier question might be to answer what is the fastest graph among a particular class of graphs. The construction of a graph G_{11} may be generalized to an arbitrary 3-layer graph consisting of three layers with sizes 1, a, and $n - 1 - a$. Let us call this graph $G_{n,a}$. We again consider the first two layers to be independent sets and the third layer to be a complete graph. The edges between the consecutive layers create a complete bipartite graph. Thus the graph G_{11} corresponds to this generalized 3-layer graph $G_{11,4}$. Hence, we may ask this simplified version of Open problem 2:

Open problem 3 (Subproblem of Open problem 2)**.** *If we fix the number of vertices n, what is the best value of a? Meaning for which a is the 3-layer graph $G_{n,a}$ fastest?*

The numerical simulations for small n seem to suggest that the optimal value of a could be of order $\sqrt{n}$.

The other interesting direction would be to prove or disprove that in general for every n the fastest 3-layer graph is indeed faster than the complete graph. If this is true we could also ask how much faster. We know, as proved in Chapter 3, that it cannot be asymptotically faster, but is it faster by a constant factor? And if so, what is the constant?

Bibliography

1. Moran, Patrick Alfred Pierce et al. The statistical processes of evolutionary theory. *The statistical processes of evolutionary theory*. 1962.
2. Lieberman, Erez; Hauert, Christoph; Nowak, Martin A. Evolutionary dynamics on graphs. *Nature*. 2005, vol. 433, no. 7023, pp. 312–316.
3. Allen, Benjamin; Lippner, Gabor; Chen, Yu-Ting; Fotouhi, Babak; Momeni, Naghmeh; Yau, Shing-Tung; Nowak, Martin A. Evolutionary dynamics on any population structure. *Nature*. 2017, vol. 544, no. 7649, pp. 227–230.
4. Adlam, Ben; Chatterjee, Krishnendu; Nowak, Martin A. Amplifiers of selection. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2015, vol. 471, no. 2181, p. 20150114.
5. Pavlogiannis, Andreas; Tkadlec, Josef; Chatterjee, Krishnendu; Nowak, Martin A. Construction of arbitrarily strong amplifiers of natural selection using evolutionary graph theory. *Communications biology*. 2018, vol. 1, no. 1, p. 71.
6. Díaz, Josep; Goldberg, Leslie Ann; Mertzios, George B; Richerby, David; Serna, Maria; Spirakis, Paul G. Approximating fixation probabilities in the generalized moran process. *Algorithmica*. 2014, vol. 69, pp. 78–91.
7. Díaz, Josep; Goldberg, Leslie Ann; Richerby, David; Serna, Maria. Absorption time of the Moran process. *Random Structures & Algorithms*. 2016, vol. 49, no. 1, pp. 137–159.
8. Díaz, Josep; Mitsche, Dieter. A survey of the modified Moran process and evolutionary graph theory. *Computer Science Review*. 2021, vol. 39, p. 100347.
9. Arora, Sanjeev; Barak, Boaz. *Computational complexity: a modern approach*. Cambridge University Press, 2009.
10. Ann Goldberg, Leslie; Lapinskas, John; Richerby, David. Phase transitions of the Moran process and algorithmic consequences. *Random Structures & Algorithms*. 2020, vol. 56, no. 3, pp. 597–647.
11. Tkadlec, Josef; Pavlogiannis, Andreas; Chatterjee, Krishnendu; Nowak, Martin A. Population structure determines the tradeoff between fixation probability and fixation time. *Communications biology*. 2019, vol. 2, no. 1, p. 138.
12. Tkadlec, Josef. *A role of graphs in evolutionary processes*. 2020. PhD thesis.
13. Hindersin, Laura; Möller, Marius; Traulsen, Arne; Bauer, Benedikt. Exact numerical calculation of fixation probability and time on graphs. *Biosystems*. 2016, vol. 150, pp. 87–91.