

# Fixation Maximization in the Positional Moran Process

Joachim Brendborg,<sup>1</sup> Panagiotis Karras,<sup>1</sup> Andreas Pavlogiannis,<sup>1</sup>  
Asger Ullersted Rasmussen,<sup>1</sup> Josef Tkadlec<sup>2</sup>

<sup>1</sup>Aarhus University, Aabogade 34, Aarhus, Denmark

<sup>2</sup>Harvard University, 1 Oxford Street, Cambridge, USA

panos@cs.au.dk, pavlogiannis@cs.au.dk, tkadlec@math.harvard.edu

## Abstract

The Moran process is a classic stochastic process that models spread dynamics on graphs. A single “mutant” (e.g., a new opinion, strain, social trait etc.) invades a population of “residents” scattered over the nodes of a graph. The mutant fitness advantage  $\delta \geq 0$  determines how aggressively mutants propagate to their neighbors. The quantity of interest is the fixation probability, i.e., the probability that the initial mutant eventually takes over the whole population. However, in realistic settings, the invading mutant has an advantage only in certain locations. E.g., the ability to metabolize a certain sugar is an advantageous trait to bacteria only when the sugar is actually present in their surroundings. In this paper, we introduce the *positional Moran process*, a natural generalization in which the mutant fitness advantage is only realized on specific nodes called *active* nodes, and study the problem of *fixation maximization*: given a budget  $k$ , choose a set of  $k$  active nodes that maximize the fixation probability of the invading mutant. We show that the problem is NP-hard, while the optimization function is not submodular, indicating strong computational hardness. We then focus on two natural limits: at the limit of  $\delta \rightarrow \infty$  (*strong selection*), although the problem remains NP-hard, the optimization function becomes submodular and thus admits a constant-factor greedy approximation; at the limit of  $\delta \rightarrow 0$  (*weak selection*), we show that we can obtain a tight approximation in  $O(m^\omega)$  time, where  $m$  is the number of edges and  $\omega$  is the matrix-multiplication exponent. An experimental evaluation of the new algorithms along with some proposed heuristics corroborates our results.

## 1 Introduction

Several real-world phenomena are described in terms of an *invasion process* that occurs in a network and follows some *propagation model*. The problems considered in such settings aim to optimize the success of the invasion, quantified by some metric. For example, social media campaigns serving purposes of marketing, political advocacy, and social welfare are modeled in terms of *influence spread*, whereby the goal is to maximize the expected influence spread of a message as a function of a set of initial adopters (Kempe, Kleinberg, and Tardos 2003; Tang, Xiao, and Shi 2014; Borgs et al. 2014; Tang, Shi, and Xiao 2015; Zhang et al. 2020); see (Li et al. 2018) for a survey. Similarly, in rumor

propagation (Demers et al. 1987), the goal is to minimize the number of rounds until a rumor spreads to the whole population (Fountoulakis, Panagiotou, and Sauerwald 2012).

A related propagation model is that of the Moran process, introduced as a model of genetic evolution (Moran 1958), and the variants thereof, such as the discrete Voter model (Clifford and Sudbury 1973; Liggett 1985; Antal, Redner, and Sood 2006; Talamali et al. 2021). A single *mutant* (e.g., a new opinion, strain, social trait, etc.) invades a population of *residents* scattered across the nodes of the graph. Contrary to influence and rumor propagation models, the Moran process accounts for *active* resistance to the invasion: a node carrying the mutant trait may not only forward it to its neighbors, but also lose it by receiving the resident trait from its neighbors. This model exemplifies settings in which individuals may switch opinions several times or sub-species competing in a gene pool. A key parameter in this process is the *mutant fitness advantage*, a real number  $\delta \geq 0$  that expresses the intensity by which mutants propagate to their neighbors compared to residents. Large  $\delta$  favors mutants, while  $\delta = 0$  renders mutants and residents indistinguishable.

Success in the Moran process is quantified in terms of the *fixation probability*, i.e., the probability that a randomly occurring mutation takes over the whole population. As the focal model in evolutionary graph theory (Lieberman, Hauert, and Nowak 2005), the process has spawned extensive work on identifying *amplifiers*, i.e., graph structures that enhance the fixation probability (Monk, Green, and Paulin 2014; Giakkoupis 2016; Galanis et al. 2017; Pavlogiannis et al. 2018; Tkadlec et al. 2021). Still, in the real world, the probability that a mutation spreads to its neighborhood is often affected by its position: the ability to metabolize a certain sugar is advantageous to bacteria only when that sugar is present in their surroundings; likewise, people are likely to spread a viewpoint more enthusiastically if it is supported by experience from their local environment (country, city, neighborhood). By default, the Moran process does not account for such positional effects. While optimization questions have been studied in the Moran and related models (Even-Dar and Shapira 2007), the natural setting of positional advantage remains unexplored. In this setting, the question arises: to which positions (i.e., nodes) should we confer an advantage so as to maximize the fixation probability?

**Our contributions.** We define a *positional* variant of the Moran process, in which the mutant fitness advantage is only realized on a subset of nodes called *active* nodes and an associated optimization problem, *fixation maximization* (FM): given a graph  $G$ , a mutant fitness advantage  $\delta$  and a budget  $k$ , choose a set of  $k$  nodes to activate that maximize the fixation probability  $\text{fp}(G^S, \delta)$  of mutants. We also study the problems  $\text{FM}^\infty$  and  $\text{FM}^0$  that correspond to FM at the natural limits  $\delta \rightarrow \infty$  and  $\delta \rightarrow 0$ , respectively. *On the negative side*, we show that FM and  $\text{FM}^\infty$  are NP-hard (Section 3.1). One common way to circumvent NP-hardness is to prove that the optimization function is submodular; however, we show that  $\text{fp}(G^S, \delta)$  is not submodular when  $\delta$  is finite (Section 3.2). *On the positive side*, we first show that the fixation probability  $\text{fp}(G^S, \delta)$  on undirected graphs admits a FPRAS, that is, it can be approximated to arbitrary precision in polynomial time (Section 4.1). We then show that, in contrast to finite  $\delta$ , the function  $\text{fp}^\infty(G^S)$  is submodular on undirected graphs and thus  $\text{FM}^\infty$  admits a polynomial-time constant-factor approximation (Section 4.2). Lastly, regarding the limit  $\delta \rightarrow 0$ , we show that  $\text{FM}^0$  can be solved in polynomial time on any graph (Section 4.3). Overall, while FM is hard in general, we obtain tractability in both natural limits  $\delta \rightarrow \infty$  and  $\delta \rightarrow 0$ .

Due to space constraints, some proofs are in Appendix A.

## 2 Preliminaries

We extend the standard Moran process to its positional variant and define our fixation maximization problem on it.

### 2.1 The Positional Moran Process

**Structured populations.** Standard evolutionary graph theory (Nowak 2006) describes a population structure by a directed weighted graph (network)  $G = (V, E, w)$  of  $|V| = n$  nodes, where  $w$  is a weight function over the edge set  $E$  that defines a probability distribution on each node  $u$ , i.e., for each  $u \in V$ ,  $\sum_{(u,v) \in E} w(u, v) = 1$ . We require that  $G$  is strongly connected when  $E$  is projected to the support of the probability distribution of each node. Nodes represent sites (locations) and each site is occupied by a single agent. Agents are of two types — *mutants* and *residents* — and agents of the same type are indistinguishable from each other. For ease of presentation, we may refer to a node and its occupant agent interchangeably. In special cases, we will consider simple undirected graphs, meaning that (i)  $E$  is symmetric and (ii)  $w(u, \cdot)$  is uniform for every node  $u$ .

**Active node sets and fitness.** Mutants and residents are differentiated in *fitness*, which in turn depends on a set  $S \subseteq V$  of *active nodes*. Consider that mutants occupy the nodes in  $X \subseteq V$ . The *fitness* of the agent at node  $u \in V$  is:

$$f_{X,S}(u) = \begin{cases} 1 + \delta, & \text{if } u \in X \cap S \\ 1, & \text{if } u \notin X \cap S \end{cases} \quad (1)$$

where  $\delta \geq 0$  is the *mutant fitness advantage*. Intuitively, fitness expresses the capacity of a type to spread in the network. The resident fitness is normalized to 1, while  $\delta$  measures the relative competitive advantage conferred by the

mutation. However, the mutant fitness advantage is not realized uniformly in the graph, but only at active nodes. The total fitness of the population is  $F_S(X) = \sum_u f_{X,S}(u)$ .

**The positional Moran process.** We introduce the *positional Moran process* as a discrete-time stochastic process  $(X_t)_{t \geq 0}$ , where each  $X_t \subseteq V$  is a random variable denoting the set of nodes of  $G$  occupied by mutants. Initially, the whole population consists of residents; at time  $t = 0$ , a single mutant appears uniformly at random. That is,  $X_0$  is given by  $\mathbb{P}[X_0 = \{u\}] = 1/n$  for each  $u \in V$ . In each subsequent step, we obtain  $X_{t+1}$  from  $X_t$  by executing the following two stochastic events in succession.

1. *Birth event:* A single agent is selected for reproduction with probability proportional to its fitness, i.e., an agent on node  $u$  is chosen with probability  $f_{X_t,S}(u)/F_S(X_t)$ .
2. *Death event:* A neighbor  $v$  of  $u$  is chosen with probability  $w(u, v)$ ; its agent is replaced by a copy of the one at  $u$ .

If  $u$  is a mutant and  $v$  a resident, then mutants spread to  $v$ , hence  $X_{t+1} = X_t \cup \{v\}$ ; likewise, if  $u$  is a resident and  $v$  a mutant, then  $X_{t+1} = X_t \setminus \{v\}$ ; otherwise  $u$  and  $v$  are of the same type, and thus  $X_{t+1} = X_t$ .

**Related Moran processes.** Our positional Moran process generalizes the standard Moran process on graphs (Nowak 2006) that forms the special case of  $S = V$  (that is, the mutant fitness advantage holds on the whole graph). Conversely, it can be viewed a special case of the Moran process with two environments (Kaveh, McAvoy, and Nowak 2019). It is different from models in which agent's fitness depends on which agents occupy the neighboring nodes (so-called frequency dependent fitness) (Huang and Traulsen 2010).

**Fixation probability.** If  $\exists \tau \geq 0$  such that  $X_\tau = V$ , we say that the mutants *fixate* in  $G$ , otherwise we say they *go extinct*. As  $G$  is strongly connected, for any initial mutant set  $A \subseteq V$ , the process reaches a homogeneous state almost surely, i.e.,  $\lim_{t \rightarrow \infty} \mathbb{P}[X_t \in \{\emptyset, V\}] = 1$ . The *fixation probability (FP) of such a set  $A$*  is  $\text{fp}(G^S, \delta, A) = \mathbb{P}[\exists \tau \geq 0: X_\tau = V \mid X_0 = A]$ . Thus, the FP of the mutants in  $G$  with active node set  $S$  and fitness advantage  $\delta$  is:

$$\text{fp}(G^S, \delta) = \frac{1}{n} \sum_{u \in V} \text{fp}(G^S, \delta, \{u\}) \quad (2)$$

For any fixed  $G$  and  $S$ , the function  $\text{fp}(G^S, \delta)$  is continuous in  $\delta$  and infinitely differentiable: the underlying Markov chain is finite and its transition probabilities are rational functions in  $\delta$ , so  $\text{fp}(G^S, \delta)$  too is a rational function in  $\delta$ , and since the process is defined for any  $\delta \in [0, \infty)$ ,  $\text{fp}(G^S, \delta)$  has no points of discontinuity for  $\delta \in [0, \infty)$ .

**Computing the fixation probability.** In the neutral setting where  $\delta = 0$ ,  $\text{fp}(G^S, \delta) = \frac{1}{n}$  for any graph  $G$  (Broom et al. 2010). When  $\delta > 0$ , the complexity of computing the fixation probability is unknown even when  $S = V$ ; it is open whether the function can be computed in PTime. For undirected  $G$  and  $S = V$ , the problem has a fully polynomial randomized approximation scheme (FPRAS) (Díaz et al. 2012;

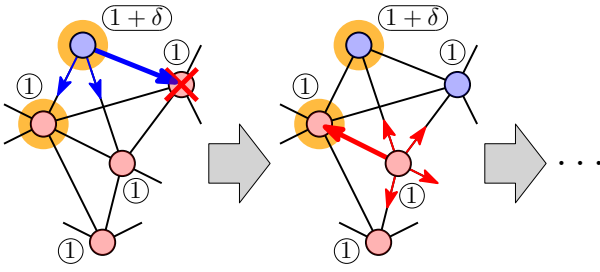


Figure 1: Steps in the positional Moran process. Blue (red) nodes are occupied by mutants (residents). Yellow rings indicate active nodes that realize the mutant fitness advantage.

Chatterjee, Ibsen-Jensen, and Nowak 2017; Ann Goldberg, Lapinskas, and Richerby 2020), as the expected absorption time of the process is polynomial and thus admits fast Monte-Carlo simulations. We also rely on Monte-Carlo simulations for computing  $\text{fp}(G^S, \delta)$ , and show that the problem admits a FPRAS for any  $S \subseteq V$  when  $G$  is undirected.

## 2.2 Fixation Maximization via Node Activation

We study the optimization problem of *fixation maximization* in the positional Moran process. We focus on three variants based on different regimes of the mutant fitness advantage  $\delta$ .

**Fixation maximization.** Consider a graph  $G = (V, E)$ , a mutant fitness advantage  $\delta$  and a budget  $k \in \mathbb{N}$ . The problem *Fixation Maximization for the Moran Process*  $\text{FM}(G, \delta, k)$  calls to determine an active-node set of size  $\leq k$  that maximizes the fixation probability:

$$S^* = \arg \max_{S \subseteq V, |S| \leq k} \text{fp}(G^S, \delta) \quad (3)$$

As we will argue,  $\text{fp}(G^S, \delta)$  is monotonically increasing in  $S$ , hence the condition  $|S| \leq k$  can be replaced by  $|S| = k$ . The associated decision problem is: Given a budget  $k$  and a threshold  $p \in [0, 1]$ , determine whether there exists an active-node set  $S$  of size  $|S| = k$  such that  $\text{fp}(G^S, \delta) \geq p$ . As an illustration, consider a cycle graph on  $n = 50$  nodes and two strategies: one that activates  $k$  “spaced” nodes, and another that activates  $k$  “contiguous” nodes. As Fig. 2 shows, depending on  $\delta$  and  $k$ , either strategy could be better.

**The regime of strong selection.** Next, we consider FM in the limit  $\delta \rightarrow \infty$ . The corresponding fixation probability is:

$$\text{fp}^\infty(G^S) = \lim_{\delta \rightarrow \infty} \text{fp}(G^S, \delta) \quad (4)$$

As we show later,  $\text{fp}(G^S, \delta)$  is monotonically increasing in  $\delta$ , and since it is also bounded by 1, the above limit exists. The corresponding optimization problem  $\text{FM}^\infty(G, k)$  asks for the active node set  $S^*$  that maximizes  $\text{fp}^\infty(G^S)$ :

$$S^* = \arg \max_{S \subseteq V, |S| \leq k} \text{fp}^\infty(G^S) \quad (5)$$

**The regime of weak selection.** Further, we consider the problem  $\text{FM}(G, \delta, k)$  in the limit  $\delta \rightarrow 0$ . For small positive  $\delta$ , different sets  $S$  generally give different fixation probabilities  $\text{fp}(G^S, \delta)$ , see Fig. 3. For fixed  $S \subseteq V$ , we can

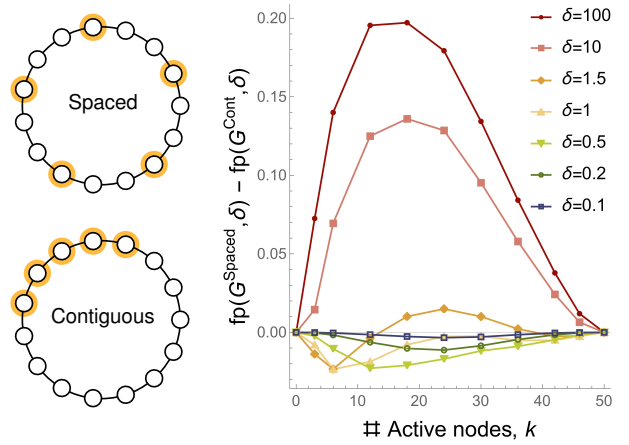


Figure 2: FP differences on a cycle graph on  $n = 50$  nodes, under spaced and contiguous activating strategy. E.g. for  $\delta = 100$  and  $k = 18$  we have  $\text{fp}(G^{\text{Spaced}}, \delta) \doteq 0.62$  and  $\text{fp}(G^{\text{Cont}}, \delta) \doteq 0.43$ , so the difference is almost 20%.

view  $\text{fp}(G^S, \delta)$  as a function of  $\delta$ . By the Taylor expansion of  $\text{fp}(G^S, \delta)$ , and given that  $\text{fp}(G^S, 0) = 1/n$ , we have

$$\text{fp}(G^S, \delta) = \frac{1}{n} + \delta \cdot \text{fp}'(G^S, 0) + \mathcal{O}(\delta^2), \quad (6)$$

where  $\text{fp}'(G^S, 0) = \left. \frac{d}{d\delta} \right|_{\delta=0} \text{fp}(G^S, \delta)$ . Since lower-order terms  $\mathcal{O}(\delta^2)$  tend to 0 faster than  $\delta$  as  $\delta \rightarrow 0$ , on a sufficiently small neighborhood of  $\delta = 0$ , maximizing  $\text{fp}(G^S, \delta)$  reduces to maximizing the derivative  $\text{fp}'(G^S, 0)$  (up to lower order terms). The corresponding optimization problem  $\text{FM}^0(G, k)$  asks for the set  $S^*$  that maximizes  $\text{fp}'(G^S, 0)$ :

$$S^* = \arg \max_{S \subseteq V, |S| \leq k} \text{fp}'(G^S, 0) \quad (7)$$

As  $\text{fp}(G^S, \delta) \geq 1/n$  for all  $S$ , our choice of  $S$  maximizes the *gain* of fixation probability, i.e., the difference  $\Delta \text{fp}(G^S, \delta) = \text{fp}(G^S, \delta) - \frac{1}{n}$ . The set  $S^*$  of Eq. (7) guarantees a gain with relative approximation factor tending to 1 as  $\delta \rightarrow 0$ , i.e., we have the following (simple) lemma.

**Lemma 1.**  $\frac{\Delta \text{fp}(G^{S^{OPT}}, \delta)}{\Delta \text{fp}(G^{S^*}, \delta)} = 1 + \mathcal{O}(\delta)$ .

**Monotonicity.** Enlarging the set  $S$  or the advantage  $\delta$  increases the fixation probability. The proof generalizes the standard Moran process monotonicity (Díaz et al. 2016).

**Lemma 2 (Monotonicity).** Consider a graph  $G = (V, E)$ , two subsets  $S \subseteq S' \subseteq V$  and two real numbers  $0 \leq \delta \leq \delta'$ . Then  $\text{fp}(G^S, \delta) \leq \text{fp}(G^{S'}, \delta')$ .

## 3 Negative Results

We start by presenting hardness results for FM and  $\text{FM}^\infty$ .

### 3.1 Hardness of FM and $\text{FM}^\infty$

We first show NP-hardness for FM and  $\text{FM}^\infty$ . This hardness persists even given an oracle that computes the fixation probability  $\text{fp}(G^S, \delta)$  (or  $\text{fp}^\infty(G^S)$ ) for a graph  $G$  and a set of

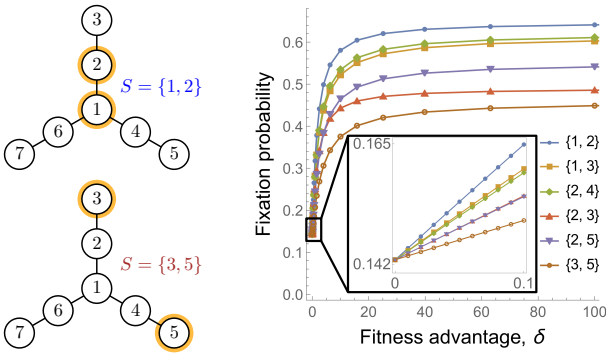


Figure 3: A fixed graph  $G$  on 7 nodes and fixation probability  $\text{fp}(G^S, \delta)$  for all six non-equivalent subsets  $S$  of  $k = 2$  nodes;  $\text{fp}(G^S, \delta)$  is roughly linear in  $\delta$  when  $\delta \in [0, 0.1]$ .

active nodes  $S$ . Given a set of mutants  $A$ , let

$$\text{fp}^\infty(G^S, A) = \lim_{\delta \rightarrow \infty} \text{fp}(G^S, \delta, A) \quad (8)$$

be the fixation probability of a set of mutants  $A$  under strong selection, i.e., in the limit of large fitness advantage. The following lemma states that for undirected graphs, a single active mutant with infinite fitness guarantees fixation.

**Lemma 3.** *Let  $G$  be an undirected graph,  $S$  a set of active nodes, and  $A$  a set of mutants. If  $A \cap S \neq \emptyset$  then  $\text{fp}^\infty(G^S, A) = 1$ .*

The intuition behind Lemma 3 is as follows: consider a node  $u \in A \cap S$ . Then, as  $\delta \rightarrow \infty$ , node  $u$  is chosen for reproduction at a much higher rate than any resident node. Hence, in a typical evolutionary trajectory, the neighbors of node  $u$  are occupied by mutants for most of the time and thus the individual at node  $u$  is effectively guarded from any threat of becoming a resident. In contrast, the mutants are reasonably likely to spread to any other part of the graph. In combination, this allows us to argue that mutants fixate with high probability as  $\delta \rightarrow \infty$ . Notably, this argument relies on  $G$  being undirected. As a simple counterexample, consider  $G$  being a directed triangle  $a \rightarrow b \rightarrow c \rightarrow a$  and  $S = A = \{a\}$ . Then, for the mutants to spread to  $c$ , a mutant on  $b$  must reproduce. Since  $b \notin S$ , a mutant on  $b$  is as likely to reproduce in each round as a resident on  $c$ . If the latter happens prior to the former, node  $a$  becomes resident, and, with no active mutants left, we have a constant (non-zero) probability of extinction.

Due to Lemma 3, if  $G$  is an undirected regular graph with a vertex cover of size  $k$ , under strong selection, we achieve the maximum fixation probability by activating a set  $S$  of  $k$  nodes that forms a vertex cover of  $G$ . The key insight is that, if the initial mutant lands on some node  $u \notin S$ , then the mutant fixates if and only if it manages to reproduce once before it is replaced, as all neighbors of  $u$  are in  $S$  and thus Lemma 3 applies. This is captured in the following lemma.

**Lemma 4.** *For any undirected regular graph  $G$  and  $S \subseteq V$ ,  $\text{fp}^\infty(G^S) \geq \frac{n+|S|}{2n}$  iff  $S$  is a vertex cover of  $G$  (and if so, the equality holds).*

*Proof.* First, note that due to Lemma 3, the fixation probability equals the probability that eventually a mutant is placed on an active node. Due to the uniform placement of the initial mutant, the probability that it lands on an active node is  $\frac{|S|}{n}$ . Let  $A$  be the set of nodes in  $V \setminus S$  that have at least one neighbor not in  $S$ . Thus we can write:

$$\text{fp}^\infty(G^S) = \frac{|S|}{n} + \frac{n-|S|-|A|}{n} p + \sum_{u \in A} \frac{1}{n} \text{fp}^\infty(G^S, \{u\}) \quad (9)$$

where  $p$  is the probability that an initial mutant occupying a non-active node  $u$  whose neighbors are all in  $S$  spreads to any of its neighbors, i.e., reproduces before any of its neighbors replace it;  $u$  reproduces with probability  $p_1 = 1/n$ , while its neighbors replace it with probability  $p_2 = \sum_{(v,u) \in E} 1/n \cdot 1/d = d \cdot 1/n \cdot 1/d = 1/n$ , where  $d$  is the degree of  $G$ , hence  $p = p_1/(p_1 + p_2) = 1/2$ . If  $S$  is a vertex cover of  $G$ , then  $A = \emptyset$ , hence Eq. (9) yields  $\text{fp}^\infty(G^S) = \frac{n+|S|}{2n}$ . If  $S$  is not a vertex cover of  $G$ , then  $|A| \geq 1$ ; for any node  $u \in A$ , since at least one of its neighbors is not in  $S$ ,  $\text{fp}^\infty(G^S, \{u\})$  is strictly smaller than the probability that the mutant on  $u$  reproduces before it gets replaced, which, as we argued, is  $\frac{1}{2}$ , hence Eq. (9) yields  $\text{fp}^\infty(G^S) < \frac{n+|S|}{2n}$ .  $\square$

The hardness for  $\text{FM}^\infty$  is a direct consequence of Lemma 4 and the fact that vertex cover is NP-hard even on regular graphs (see, e.g., (Feige 2003)). Moreover, as  $\text{fp}(G^S, \delta)$  is a continuous function on  $\delta$ , we also obtain hardness for FM (i.e., under finite  $\delta$ ). We thus have the following theorem.

**Theorem 1.**  *$\text{FM}^\infty(G, k)$  and  $\text{FM}(G, \delta, k)$  are NP-hard, even on undirected regular graphs.*

Finally, we note that the hardness persists even given an oracle that computes the fixation probability  $\text{fp}(G^S, \delta)$  (or  $\text{fp}^\infty(G^S)$ ) given a graph  $G$  and a set of active nodes  $S$ .

### 3.2 Non-Submodularity for FM

One standard way to circumvent NP-hardness is to show that the optimization function is submodular, which implies that a greedy algorithm offers constant-factor approximation guarantees (Nemhauser, Wolsey, and Fisher 1978; Krause and Golovin 2014). Although in the next section we prove that  $\text{fp}^\infty(G^S)$  is indeed submodular, unfortunately the same does not hold for  $\text{fp}(G^S, \delta)$ , even for very simple graphs.

**Theorem 2.** *The function  $\text{fp}(G^S, \delta)$  is not submodular.*

*Proof.* Our proof is by means of a counter-example. Consider  $G = K_4$ , that is,  $G$  is a clique on 4 nodes. For  $i \in \{0, 1, 2\}$  denote by  $p_i$  the fixation probability on  $K_4$  when  $i$  of the 4 nodes are active and mutants have fitness advantage  $\delta = 1/3$ . For fixed  $i$ , the value  $p_i$  can be computed exactly by solving a system of  $2^4$  linear equations (one for each configuration of mutants and residents). Solving the three systems using a computer algebra system, we obtain  $p_0 = 1/4$ ,  $p_1 = 38413/137740 < 0.2788$  and  $p_2 = 28984/94153 > 0.3078$ . Hence  $p_0 + p_2 > 0.5578 = 2 \cdot 0.2789 > 2p_1$ , thus submodularity is violated.  $\square$

For  $G = K_4$ , the three systems can even be solved symbolically, in terms of  $\delta > 0$ . It turns out that the submodularity property is violated for  $\delta < \delta^*$ , where  $\delta^* \doteq 0.432$ . In contrast, for  $G = K_3$  the submodularity property holds for all  $\delta \geq 0$  (the verification is straightforward, though tedious).

## 4 Positive Results

We now turn our attention to positive results for fixation maximization in the limits of strong and weak selection.

### 4.1 Approximating the Fixation Probability

We first focus on computing the fixation probability. Although the complexity of the problem is open even in the standard Moran process, when the underlying graph is undirected, the expected number of steps until the standard (non-positional) process terminates is polynomial in  $n$  (Díaz et al. 2012), which yields a FPRAS. A straightforward extension of that approach applies to the expected number  $T(G^S, \delta)$  of steps in the positional Moran process.

**Lemma 5.** *Given an undirected graph  $G = (V, E)$  on  $n$  nodes, a set  $S \subseteq V$  and a real number  $\delta \geq 1$ , we have  $T(G^S, \delta) \leq (1 + \delta)n^6$ .*

As a consequence, by simulating the process multiple times and reporting the proportion of runs that terminated with mutant fixation, we obtain a fully polynomial randomized approximation scheme (FPRAS) for the fixation probability.

**Corollary 1.** *Given a connected undirected graph  $G = (V, E)$ , a set  $S \subseteq V$  and a real number  $\delta \geq 1$ , the function  $\text{fp}(G^S, \delta)$  admits a FPRAS.*

### 4.2 Fixation Maximization under Strong Selection

Here we show that the function  $\text{fp}^\infty(G^S)$  is submodular. As a consequence, we obtain a constant-factor approximation for  $\text{fp}^\infty(G^S)$  in polynomial time.

**Lemma 6** (Submodularity). *For any undirected graph  $G$ , the function  $\text{fp}^\infty(G^S)$  is submodular.*

*Proof.* Consider any set  $S$  of active nodes. Then the positional Moran process  $\{X_t\}$  with mutant advantage  $\delta \rightarrow \infty$  is equivalent to the following process:

1. While  $X_t \cap S = \emptyset$ , perform an update step  $X_t \rightarrow X_{t+1}$  as if  $\delta = 0$ .
2. If  $X_t \cap S \neq \emptyset$ , terminate and report fixation.

Indeed, as long as no active node hosts a mutant, all individuals have the same (unit) fitness and the process is indistinguishable from the Moran process with  $\delta = 0$ . On the other hand, once any one active node receives a mutant, fixation happens with high probability by Lemma 3. All in all, the fixation probability  $p(S) = \text{fp}^\infty(G^S)$  in the limit  $\delta \rightarrow \infty$  can be computed by simulating the neutral process ( $\delta = 0$ ) until either  $X_t = \emptyset$  (extinction) or  $X_t \cap S \neq \emptyset$  (fixation).

To prove the submodularity of  $p(S)$ , it suffices to show that for any two sets  $S, T \subseteq V$  we have:

$$p(S) + p(T) \geq p(S \cup T) + p(S \cap T). \quad (10)$$

Consider any fixed trajectory  $\mathcal{T} = (X_0, X_1, \dots)$ . We say that a subset  $U \subseteq V$  is *good* with respect to  $\mathcal{T}$ , or simply *good*, if there exists a time point  $\tau \geq 0$  such that  $X_\tau \cap U \neq \emptyset$ . It suffices to show that, for any  $\mathcal{T}$ , there are at least as many good sets among  $S, T$  as they are among  $S \cup T$  and  $S \cap T$ . To that end, we distinguish three cases based on how many of the sets  $S \cup T, S \cap T$  are good.

1. Both of them: Since  $S \cap T$  is good, both  $S$  and  $T$  are good and we get  $2 \geq 2$ .
2. One of them: If  $S \cap T$  is good we conclude as before (we get  $2 \geq 1$ ). Otherwise  $S \cup T$  is good, hence at least one of  $S, T$  is good and we get  $1 \geq 1$ .
3. None of them: We have  $0 \geq 0$ .  $\square$

The submodularity lemma leads to the following approximation guarantee for the greedy algorithm (Nemhauser, Wolsey, and Fisher 1978; Krause and Golovin 2014).

**Theorem 3.** *Given an undirected graph  $G$  and integer  $k$ , let  $S^*$  be the solution to  $\text{FM}^\infty(G, k)$ , and  $S'$  the set returned by a greedy maximization algorithm. Then  $\text{fp}^\infty(G^{S'}) \geq (1 - \frac{1}{e}) \text{fp}^\infty(G^{S^*})$ .*

Consequently, a greedy algorithm approximates the optimal fixation probability within a factor of  $1 - 1/e$ , provided it is equipped with an oracle that estimates the fixation probability  $\text{fp}^\infty(G^S)$  given any set  $S$ . For the class of undirected graphs, Lemma 5 and Theorem 3 yield a fully polynomial randomized greedy algorithm for  $\text{FM}^\infty(G, k)$  with approximation guarantee  $1 - 1/e$ .

### 4.3 Fixation Maximization under Weak Selection

Here we show that  $\text{FM}^0$  can be solved in polynomial time. Recall that given an active set  $S \subset V$ , we can write  $\text{fp}(G^S, \delta) = \text{fp}(G^S, 0) + \delta \cdot \text{fp}'(G^S, 0) + \mathcal{O}(\delta^2)$ , while  $\text{FM}^0(G, k)$  calls to compute a set  $S^*$  which maximizes the coefficient  $\text{fp}'(G^S, 0)$  across all sets  $S$  that satisfy  $|S| \leq k$ . We build on the machinery developed in (Allen et al. 2017).

In Lemma 7 below, we show that  $\text{fp}'(G^S, 0) = \sum_{u_i \in S} \alpha(u_i)$  for a certain function  $\alpha: V \rightarrow \mathbb{R}$ . The bottleneck in computing  $\alpha(\cdot)$  is solving a system of  $\mathcal{O}(m)$  linear equations, where  $m$  is the number of edges in  $G$ . This can be done in matrix-multiplication time  $\mathcal{O}(m^\omega)$  for  $\omega < 2.373$  (Alman and Williams 2021). By virtue of the linearity of  $\text{fp}'(G^S, 0)$ , the optimal set  $S^*$  is then given by choosing the top- $k$  nodes  $u_i$  in terms of  $\alpha(u_i)$ . Thus we establish the following theorem.

**Theorem 4.** *Given a graph  $G$  of  $m$  edges and some integer  $k$ , the solution to  $\text{FM}^0(G, k)$  can be computed in  $\mathcal{O}(m^\omega)$  time, where  $\omega < 2.373$  is the matrix multiplication exponent.*

In the rest of this section, we state Lemma 7 and give intuition about the function  $\alpha(\cdot)$ . For the formal proof of Lemma 7, see Appendix.

For brevity, we denote the  $n$  nodes of  $G$  by  $u_1, \dots, u_n$  and we denote by  $p_{ij} = w(u_i, u_j)$  the probability that, if  $u_i$  is selected for reproduction, then the offspring migrates to  $u_j$ .

**Lemma 7.** Let  $G = (V, E)$  be a graph with  $n$  nodes  $u_1, \dots, u_n$  and  $m$  edges. Consider a function  $\alpha: V \rightarrow \mathbb{R}$  defined by  $\alpha(u_i) = \frac{1}{n} \cdot \sum_{j \in [n]} p_{ij} \cdot \pi_j \cdot \psi_{ij}$ , where  $\{\pi_i \mid u_i \in V\}$  is the solution to the linear system

$$\pi_i = \left(1 - \sum_{j \in [n]} p_{ji}\right) \cdot \pi_i + \sum_{j \in [n]} p_{ij} \cdot \pi_j \quad \forall i \in [n], \quad (11)$$

and  $\{\psi_{ij} \mid (u_i, u_j) \in E\}$  is the solution to the linear system

$$\psi_{ij} = \begin{cases} \frac{1 + \sum_{\ell \in [n]} (p_{\ell i} \cdot \psi_{\ell j} + p_{\ell j} \cdot \psi_{i \ell})}{\sum_{\ell \in [n]} (p_{\ell i} + p_{\ell j})} & i \neq j \text{ and } (u_i, u_j) \in E \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

Then  $\text{fp}'(G^S, 0) = \sum_{u_i \in S} \alpha(u_i)$ .

The intuition behind the quantities  $\pi_i$ ,  $\psi_{ij}$ , and  $\alpha(u_i)$  from Lemma 7 is as follows: Consider the Moran process on  $G$  with  $\delta = 0$ . Then for  $i \in [n]$ , the value  $\pi_i$  is in fact the mutant fixation probability starting from the initial configuration  $X_0 = \{u_i\}$  (Allen et al. 2021). Indeed, at any time step, the agent on  $u_i$  will eventually fixate if (i)  $u_i$  is not replaced by its neighbors and eventually fixates from the next step onward, or (ii)  $u_i$  spreads to a neighbor  $u_j$  and fixates from there. The first event happens with rate  $(1 - \sum_{j \in [n]} p_{ji}) \cdot \pi_i$ , while the second event happens with rate  $\sum_{j \in [n]} p_{ij} \cdot \pi_j$ . (We note that for undirected graphs, the system has an explicit solution  $\pi_i = \frac{1/\deg(u_i)}{\sum_{j \in [n]} 1/\deg(u_j)}$  (Broom et al. 2010).)

The values  $\psi_{ij}$  for  $(u_i, u_j) \in E$  also have intuitive meaning: They are equal to the expected total number of steps during the Moran process (with  $\delta = 0$ ) in which  $u_i$  is mutant and  $u_j$  is not, when starting from a random initial configuration  $\mathbb{P}[X_0 = \{u_i\}] = 1/n$ . To sketch the idea behind this claim, suppose that in a single step, a node  $u_\ell$  spreads to node  $u_i$  (this happens with rate  $p_{\ell i}$ ). Then the event “ $u_i$  is mutant while  $u_j$  is not” holds if and only if  $u_\ell$  was mutant but  $u_j$  was not, hence the product  $p_{\ell i} \cdot \psi_{\ell j}$ . Similar reasoning applies to the product  $p_{\ell j} \cdot \psi_{i \ell}$ . The term 1 in the numerator comes out of the  $1/n$  probability that the initial mutant lands on  $u_i$  (in which case indeed  $u_i$  is mutant and  $u_j$  is not).

Finally, the function  $\alpha(\cdot)$  also has a natural interpretation: The contribution of an active node  $u_i$  to the fixation probability grows vs.  $\delta$  at  $\delta = 0$  to the extent that a mutant  $u_i$  is likely to be chosen for reproduction, spread to a resident neighbor  $u_j$ , and fixate from there; that is, by the total time that  $u_i$  is mutant but a neighbor  $u_j$  is not ( $\psi_{ij}$ ), calibrated by the probability that, when chosen for reproduction,  $u_i$  propagates to that neighbor ( $p_{ij}$ ), who thereafter achieves mutant fixation ( $\pi_j$ ); this growth rate is summed over all neighbors and weighted by the rate  $1/n$  at which  $u_i$  reproduces when  $\delta = 0$ .

## 5 Experiments

Here we report on an experimental evaluation of the proposed algorithms and some other heuristics. Our data set

consists of 110 connected subgraphs of community graphs and social networks of the Stanford Network Analysis Project (Leskovec and Krevl 2014). These subgraphs were chosen randomly, and varied in size between 20-170 nodes. Our evaluation is not aimed to be exhaustive, but rather to outline the practical performance of various heuristics, sometimes in relation to their theoretical guarantees.

In particular, we use 7 heuristics, each taking as input a graph, a budget  $k$ , and (optionally) the mutant fitness advantage  $\delta$ . The heuristics are motivated by our theoretical results and by the related fields of influence maximization and evolutionary graph theory.

1. *Random*: Simply choose  $k$  nodes uniformly at random. This heuristic serves as a baseline.
2. *High Degree*: Choose the  $k$  nodes with the largest degree.
3. *Centrality*: Choose the  $k$  nodes with the largest betweenness centrality.
4. *Temperature*: The temperature of a node  $u$  is defined as  $\mathcal{T}(u) = \sum_v w(v, u)$ ; this heuristic chooses the  $k$  nodes with the largest temperature.
5. *Vertex Cover*: Motivated by Lemma 4, this heuristic attempts to maximize the number of edges with at least one endpoint in  $S$ . In particular, given a set  $A \subseteq V$ , let

$$c(A) = \{ |(u, v) \in E : u \in A \text{ or } v \in A| \}. \quad (13)$$

The heuristic greedily maximizes  $c(\cdot)$ , i.e., we start with  $S = \emptyset$  and perform  $k$  update steps

$$S \leftarrow S \cup \arg \max_{u \notin S} c(S \cup \{u\}). \quad (14)$$

6. *Weak Selector*: Choose the  $k$  nodes that maximize  $\text{fp}'(G^S, 0)$  using the (optimal) weak-selection method.
7. *Lazy Greedy*: A simple greedy algorithm starts with  $S = \emptyset$ , and in each step chooses the node  $u$  to add to  $S$  that maximizes the objective function. For  $\text{FM}(G, \delta, k)$  and  $\text{FM}^\infty(G, k)$  this process requires to repeatedly evaluate  $\text{fp}(G^{S \cup \{u\}}, \delta)$  (or  $\text{fp}(G^{S \cup \{u\}}, \delta)$ ) for every node  $u$ . This is done by simulating the process a large number of times (recall Lemma 5), and becomes a computational bottleneck when we require high precision. As a workaround we suggest a lazy variant of the greedy algorithm (Minoux 1978), which is faster but requires submodularity of the objective function. In effect, this algorithm is a correct implementation of the greedy heuristic in the limit of strong selection (recall Lemma 6), while it may still perform well (but without guarantees) for finite  $\delta$ .

In all cases ties are broken arbitrarily. Note that the heuristics vary in the amount of information they have about the graph and the invasion process. In particular, Random has no information whatsoever, High Degree only considers the direct neighbors of a node, Temperature considers direct and distance-2 neighbors of a node, Centrality and Vertex Cover consider the whole graph, while Weak Selector and Lazy Greedy are the only ones informed about the Moran process.

For each graph  $G$ , we have chosen values of  $k$  corresponding to 10%, 30% and 50% of its nodes, and have evaluated the above heuristics in their ability to solve  $\text{FM}^\infty(G, K)$  (strong

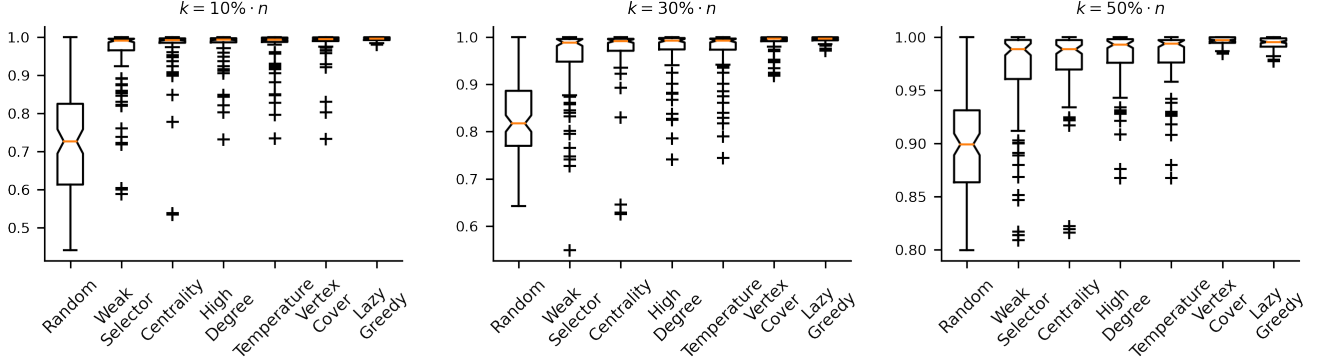


Figure 4: Heuristic performance for  $FM^\infty$ .

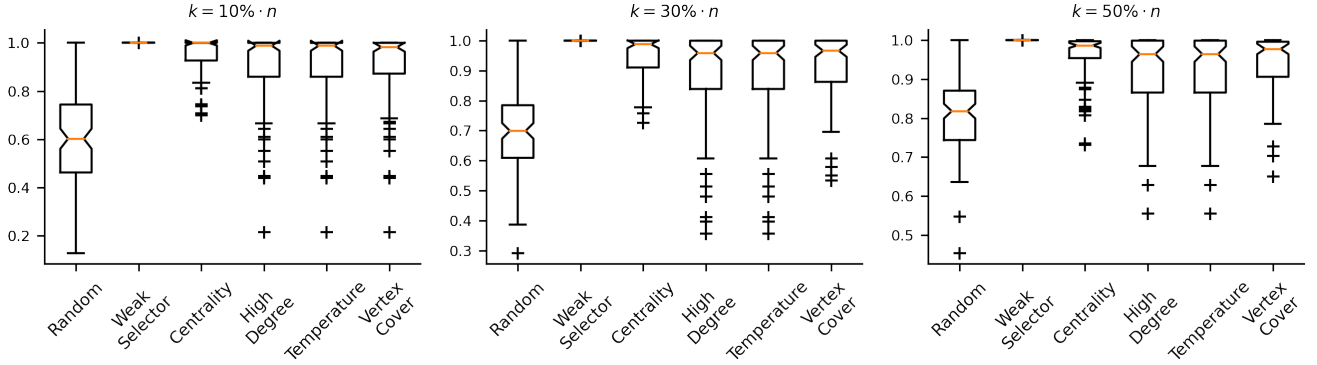


Figure 5: Heuristic performance for  $FM^0$ .

selection) and  $FM^0(G, k)$  (weak selection). We have not considered other values of  $\delta$  as evaluating  $fp(G^S, \delta)$  precisely via simulations requires many repetitions and becomes slow.

**Strong selection.** We start with the case of  $FM^\infty(G, K)$ . Since different graphs  $G$  and budgets  $k$  generally result in highly variant fixation probabilities, in order to get an informative aggregate measure of each heuristic, we divide the fixation probability it obtains by the maximum fixation probability obtained across all heuristics for the same graph and budget. This normalization yields values in the interval  $[0, 1]$ , and makes comparison straightforward. Fig. 4 shows our results. We see that the Lazy Greedy algorithm performs best for small budgets (10%, 30%), while its performance is matched by Vertex Cover for budget 50%. The high performance of Lazy Greedy is expected given its theoretical guarantee (Theorem 3). We also observe that, apart from Weak Selector, the other heuristics perform quite well on many graphs, though there are several cases on which they fail, appearing as outlier points in the box plots. As expected, our baseline Random heuristic performs quite poorly; this result indicates that the node activation set  $S$  has a significant impact on the fixation probability. Finally, recall that Weak Selector is optimal for the regime of weak selection ( $\delta \rightarrow 0$ ). The fact that Weak Selector underperforms for strong selection ( $\delta \rightarrow \infty$ ) indicates an intricate relationship between fitness advantage and fixation probability.

**Weak selection.** We collect the results on weak selection

in Fig. 5, using the normalization process above. Since the derivative satisfies  $fp'(G^S, 0) = \sum_{u_i \in S} \alpha(u_i)$  (Section 4.3), Lazy Greedy is optimal and coincides with Weak Selector, and is thus omitted from the figure. Naturally, the Weak Selector always outperforms others, while the Random heuristic is weak. The other heuristics have mediocre performance, with a clear advantage of Centrality over the rest, which becomes clearer for larger budget values.

## 6 Conclusion

We introduced the positional Moran process and studied the associated fixation maximization problem. We have shown that the problem is NP-hard in general, but becomes tractable in the limits of strong and weak selection. Our results only scratch the surface of this new process, as several interesting questions are open, such as: Does the strong-selection setting admit a better approximation than the one based on submodularity? Can the problem for finite  $\delta$  be approximated within some constant-factor? Are there classes of graphs for which it becomes tractable?

## References

- Allen, B.; Lippner, G.; Chen, Y.-T.; Fotouhi, B.; Momeni, N.; Yau, S.-T.; and Nowak, M. A. 2017. Evolutionary dynamics on any population structure. *Nature*, 544(7649): 227–230.
- Allen, B.; Sample, C.; Steinhagen, P.; Shapiro, J.; King, M.; Hedspeth, T.; and Goncalves, M. 2021. Fixation probabil-

- ities in graph-structured populations under weak selection. *PLOS Computational Biology*, 17(2): 1–25.
- Alman, J.; and Williams, V. V. 2021. A Refined Laser Method and Faster Matrix Multiplication. In Marx, D., ed., *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, 522–539. SIAM.
- Ann Goldberg, L.; Lapinskas, J.; and Richerby, D. 2020. Phase transitions of the Moran process and algorithmic consequences. *Random Structures & Algorithms*, 56(3): 597–647.
- Antal, T.; Redner, S.; and Sood, V. 2006. Evolutionary Dynamics on Degree-Heterogeneous Graphs. *Phys. Rev. Lett.*, 96: 188104.
- Borgs, C.; Brautbar, M.; Chayes, J.; and Lucier, B. 2014. Maximizing Social Influence in Nearly Optimal Time. In *SODA*, 946–957.
- Broom, M.; Hadjichrysanthou, C.; Rychtář, J.; and Stadler, B. 2010. Two results on evolutionary processes on general non-directed graphs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2121): 2795–2798.
- Chatterjee, K.; Ibsen-Jensen, R.; and Nowak, M. A. 2017. Faster Monte-Carlo Algorithms for Fixation Probability of the Moran Process on Undirected Graphs. In Larsen, K. G.; Bodlaender, H. L.; and Raskin, J.-F., eds., *42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017)*, volume 83 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 61:1–61:13. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-95977-046-0.
- Clifford, P.; and Sudbury, A. 1973. A Model for Spatial Conflict. *Biometrika*, 60(3): 581–588.
- Demers, A.; Greene, D.; Hauser, C.; Irish, W.; Larson, J.; Shenker, S.; Sturgis, H.; Swinehart, D.; and Terry, D. 1987. Epidemic Algorithms for Replicated Database Maintenance. In *Proceedings of the Sixth Annual ACM Symposium on Principles of Distributed Computing, PODC '87*, 1–12. New York, NY, USA: Association for Computing Machinery. ISBN 089791239X.
- Díaz, J.; Goldberg, L. A.; Mertzios, G. B.; Richerby, D.; Serna, M.; and Spirakis, P. G. 2012. Approximating Fixation Probabilities in the Generalized Moran Process. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '12*, 954–960. USA: Society for Industrial and Applied Mathematics.
- Díaz, J.; Goldberg, L. A.; Richerby, D.; and Serna, M. 2016. Absorption time of the Moran process. *Random Structures & Algorithms*, 49(1): 137–159.
- Even-Dar, E.; and Shapira, A. 2007. A Note on Maximizing the Spread of Influence in Social Networks. In Deng, X.; and Graham, F. C., eds., *Internet and Network Economics*, 281–286. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-77105-0.
- Feige, U. 2003. Vertex cover is hardest to approximate on regular graphs. Technical Report MCS03-15, The Weizmann Institute of Science.
- Fountoulakis, N.; Panagiotou, K.; and Sauerwald, T. 2012. Ultra-Fast Rumor Spreading in Social Networks. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '12*, 1642–1660. USA: Society for Industrial and Applied Mathematics.
- Galanis, A.; Göbel, A.; Goldberg, L. A.; Lapinskas, J.; and Richerby, D. 2017. Amplifiers for the Moran process. *Journal of the ACM (JACM)*, 64(1): 5.
- Giakkoupis, G. 2016. Amplifiers and Suppressors of Selection for the Moran Process on Undirected Graphs. *arXiv preprint arXiv:1611.01585*.
- Huang, W.; and Traulsen, A. 2010. Fixation probabilities of random mutants under frequency dependent selection. *Journal of Theoretical Biology*, 263(2): 262–268.
- Kaveh, K.; McAvoy, A.; and Nowak, M. A. 2019. Environmental fitness heterogeneity in the Moran process. *Royal Society open science*, 6(1): 181661.
- Kempe, D.; Kleinberg, J.; and Tardos, E. 2003. Maximizing the Spread of Influence through a Social Network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137–146.
- Kötzing, T.; and Krejca, M. S. 2019. First-hitting times under drift. *Theoretical Computer Science*, 796: 51–69.
- Krause, A.; and Golovin, D. 2014. Submodular function maximization. *Tractability*, 3: 71–104.
- Leskovec, J.; and Krevl, A. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. [/url=http://snap.stanford.edu/data](http://snap.stanford.edu/data).
- Li, Y.; Fan, J.; Wang, Y.; and Tan, K. 2018. Influence Maximization on Social Graphs: A Survey. *IEEE TKDE*, 30(10): 1852–1872.
- Lieberman, E.; Hauert, C.; and Nowak, M. A. 2005. Evolutionary dynamics on graphs. *Nature*, 433(7023): 312–316.
- Liggett, T. 1985. *Interacting Particle Systems*. Classics in mathematics. Springer New York. ISBN 9783540960690.
- McAvoy, A.; and Allen, B. 2021. Fixation probabilities in evolutionary dynamics under weak selection. *Journal of Mathematical Biology*, 82(3): 1–41.
- Minoux, M. 1978. Accelerated greedy algorithms for maximizing submodular set functions. In Stoer, J., ed., *Optimization Techniques*, 234–243. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-35890-9.
- Monk, T.; Green, P.; and Paulin, M. 2014. Martingales and fixation probabilities of evolutionary graphs. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 470(2165): 20130730.
- Moran, P. A. P. 1958. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(1): 60–71.
- Nemhauser, G. L.; Wolsey, L. A.; and Fisher, M. L. 1978. An Analysis of Approximations for Maximizing Submodular Set Functions—I. *Math. Program.*, 14(1): 265–294.
- Nowak, M. A. 2006. *Evolutionary dynamics: exploring the equations of life*. Cambridge, Massachusetts: Belknap Press of Harvard University Press. ISBN 0674023382 (alk. paper).

Pavlogiannis, A.; Tkadlec, J.; Chatterjee, K.; and Nowak, M. A. 2018. Construction of arbitrarily strong amplifiers of natural selection using evolutionary graph theory. *Communications Biology*, 1(1): 71.

Talamali, M. S.; Saha, A.; Marshall, J. A. R.; and Reina, A. 2021. When less is more: Robot swarms adapt better to changes with constrained communication. *Science Robotics*, 6(56): eabf1416.

Tang, Y.; Shi, Y.; and Xiao, X. 2015. Influence Maximization in Near-Linear Time: A Martingale Approach. In *SIGMOD*, 1539–1554.

Tang, Y.; Xiao, X.; and Shi, Y. 2014. Influence Maximization: Near-optimal Time Complexity Meets Practical Efficiency. In *SIGMOD*.

Tkadlec, J.; Pavlogiannis, A.; Chatterjee, K.; and Nowak, M. A. 2021. Fast and strong amplifiers of natural selection. *Nature Communications*, 12(1): 4009.

Zhang, K.; Zhou, J.; Tao, D.; Karras, P.; Li, Q.; and Xiong, H. 2020. *Geodemographic Influence Maximization*, 2764–2774. New York, NY, USA: Association for Computing Machinery. ISBN 9781450379984.

## A Appendix

**Lemma 2** (Monotonicity). *Consider a graph  $G = (V, E)$ , two subsets  $S \subseteq S' \subseteq V$  and two real numbers  $0 \leq \delta \leq \delta'$ . Then  $\text{fp}(G^S, \delta) \leq \text{fp}(G^{S'}, \delta')$ .*

*Proof.* Fix pairs  $(S, \delta)$  and  $(S', \delta')$ . The key property is that, for any two configurations  $X \subseteq X'$  and any node  $u \in V$ , we have  $f_{X,S}(u) \leq f_{X',S'}(u)$ : Indeed, the inequality is strict if and only if  $u \in (X' \setminus X) \cap (S' \setminus S)$ . As a consequence, Lemma 5 from (Díaz et al. 2016) applies and yields a desired coupling between the processes with parameters  $(S, \delta)$  and  $(S', \delta')$ .  $\square$

**Lemma 3.** *Let  $G$  be an undirected graph,  $S$  a set of active nodes, and  $A$  a set of mutants. If  $A \cap S \neq \emptyset$  then  $\text{fp}^\infty(G^S, A) = 1$ .*

*Proof.* Due to Lemma 2, it suffices to prove the statement for when  $A = \{u\}$  is a singleton set. Our proof is by a stochastic domination argument. In particular, we construct a simple Markov chain  $\mathcal{M}$  and a coupling between  $\mathcal{M}$  and the Moran process on  $G$  from  $X = \{u\}$  such that the probability of a random walk starting from a particular start state  $s_u$  of  $\mathcal{M}$  has at least probability  $\text{fp}^\infty(G^S, A)$  of getting absorbed in a final state  $s_n$ .

Let  $\ell = 2 + |\{(u, v) \in E\}|$ , i.e.,  $\ell$  equals 2 plus the number of neighbors of  $u$ .  $\mathcal{M}$  consists of the following set of states

$$S = \{s_0, s_u, s_{u^*}\} \cup \{s_i : \ell \leq i \leq n\}.$$

The transition probability function  $p : S \times S \rightarrow [0, 1]$  is as

follows, for constants  $c_1, c_2$  that depend on  $G$  but not on  $\delta$ .

$$\begin{aligned} p(s_u, s_0) &= \frac{c_1}{1 + \delta} & p(s_u, s_{u^*}) &= 1 - p(s_u, s_0) \\ p(s_{u^*}, s_\ell) &= p(s_i, s_{i+1}) = c_2 & \text{for } \ell \leq i < n \\ p(s_{u^*}, s_u) &= p(s_i, s_u) = 1 - c_2 & \text{for } \ell \leq i \leq n \end{aligned}$$

while for every state  $s$ , the remaining probability mass is added as a self-loop probability  $p(s, s)$ .

We now sketch the coupling between  $\mathcal{M}$  and the positional Moran process on  $G$ . The coupling guarantees the following correspondence between the states of  $\mathcal{M}$  and the configuration  $X$  of the Moran process.

1. For the state  $s_u$ , we have  $u \in X$ .
2. For the state  $s_{u^*}$ , we have (i)  $u \in X$ , and (ii) for every  $(u, v) \in E$ , we have  $v \in X$ .
3. For each state  $s_i$ , with  $\ell \leq i \leq n$ , we have (i)  $u \in X$ , (ii) for every  $(u, v) \in E$ , we have  $v \in X$ , and (iii)  $|X| \geq i$ .

We argue that the transition probabilities preserve the above correspondence.

1. While  $u \in X$ , the probability for a resident neighbor  $v$  of  $u$  to reproduce and place an offspring on  $u$  is bounded by  $\propto \frac{1}{1+\delta}$ . On the other hand, the probability that  $u$  reproduces in successive rounds until it places a mutant offspring to each of its neighbors can be exponentially small in  $n$  but independent of  $\delta$ , as, since  $u \in S$ , the fitness of  $u$  is at least a fraction  $1/n$  of the total population fitness regardless of the remaining nodes. It follows that, from a configuration  $X$  with  $u \in X$ , the probability that  $u$  turns resident before every neighbor of  $v$  turns mutant is upper-bounded by  $\frac{c}{1+\delta}$ , for some constant  $c$  that depends on  $G$  but not on  $\delta$ .
2. Given any configuration  $X$  with  $u \in X$ , the probability that a resident reproduces is at most  $\propto \frac{1}{1+\delta}$ . Similarly, the probability that a mutant reproduces and replaces a resident is at least  $\propto \frac{1}{1+\delta}$ . Thus, the probability that the latter event happens before the former event is lower-bounded by some constant  $c_2$  that depends on  $G$  but not on  $\delta$ .

Thus, we have a coupling in which the probability that a random walk in  $\mathcal{M}$  starting at  $s_u$  gets absorbed in  $s_n$  is at least as large as  $\text{fp}^\infty(G^S, \{u\})$ . It is straightforward to verify that the probability of the former event tends to 1 as  $\delta \rightarrow \infty$ . Thus  $\text{fp}^\infty(G^S, \{u\})$ , as desired.  $\square$

**Theorem 1.**  *$\text{FM}^\infty(G, k)$  and  $\text{FM}(G, \delta, k)$  are NP-hard, even on undirected regular graphs.*

*Proof.* Following Lemma 4, the hardness for  $\text{FM}^\infty$  follows directly by reducing the question of whether a graph  $G$  has a vertex cover of size  $k$  to the problem of whether  $\text{FM}^\infty$  can achieve fixation probability of at least  $\frac{n+k}{2n}$ . For the hardness of  $\text{FM}$ , recall that for any  $S \subseteq V$ , the function  $\text{fp}(G^S, \delta)$  is continuous on  $\delta$ . A crude but slightly more detailed analysis in the proof of Lemma 4 shows that if we have an edge

$(u, v) \in E$  with  $u, v \notin S$ , then the fixation probability from  $u$  is

$$\text{fp}^\infty(G^S, \{u\}) < \frac{1}{2} - \frac{1}{2n^4}.$$

In turn, if  $S$  is not a vertex cover of  $G$  then

$$\text{fp}^\infty(G^S) < c = \frac{n + |S|}{2n} - \frac{1}{2n^5}.$$

Due to the continuity of  $\text{fp}(G^S, \delta)$ , there exists a large enough  $\delta^*$  such that if  $G$  has a vertex cover  $S$  of size  $k$  then  $\text{fp}(G^S, \delta^*) \geq c$ . On the other hand, due to the monotonicity (Lemma 2), if  $G$  has no vertex cover of size  $k$ , then we have

$$\max_{S \subseteq V, |S|=k} \text{fp}(G^S, \delta^*) \leq \max_{S \subseteq V, |S|=k} \text{fp}^\infty(G^S) < c.$$

Thus for such a fitness advantage  $\delta^*$ , we can achieve fixation probability  $c$  iff  $G$  has a vertex cover of size  $k$ .  $\square$

**Lemma 5.** *Given an undirected graph  $G = (V, E)$  on  $n$  nodes, a set  $S \subseteq V$  and a real number  $\delta \geq 1$ , we have  $T(G^S, \delta) \leq (1 + \delta)n^6$ .*

*Proof.* We follow the proof strategy of Theorem 11 from (Díaz et al. 2012). Given a set  $X$  of nodes occupied by mutants, consider a potential function defined by  $\phi(X) = \sum_{u \in X} \frac{1}{\deg(u)}$ , where  $\deg(u)$  is the number of edges incident to  $u$ . Note that  $0 \leq \phi(X) \leq n$ , since  $0 \leq |X| \leq n$  and for each  $u \in V$  we have  $\deg(u) \geq 1$ . Moreover, given a configuration  $X_t$  at time-point  $t$ , let  $\Delta_t = \phi(X_{t+1}) - \phi(X_t)$  be a random variable that measures the increase of the potential function  $\phi$  in a single step. We make two claims:

1.  $\mathbb{E}[\Delta_t | X_t] \geq 0$ : Let  $S$  be a set of edges  $(u, v)$  such that  $u$  is occupied by a mutant and  $v$  by a resident, and let  $F = \sum_{u \in V} f_{X_t, S}(u)$  be the total fitness of the population. In a single step, the set of nodes occupied by mutants changes either by a mutant replacing a resident neighbor, or by a resident replacing a mutant neighbor. Thus

$$\begin{aligned} \mathbb{E}[\Delta_t | X_t] &= \sum_{(u, v) \in S} \frac{f_{X_t, S}(u)}{F} \cdot \frac{1}{\deg(u)} \cdot \frac{+1}{\deg(v)} \\ &\quad + \sum_{(u, v) \in S} \frac{f_{X_t, S}(v)}{F} \cdot \frac{1}{\deg(v)} \cdot \frac{-1}{\deg(u)} \\ &= \sum_{(u, v) \in S} \frac{f_{X_t, S}(u) - f_{X_t, S}(v)}{F \deg(u) \deg(v)} \geq 0, \end{aligned}$$

where the last inequality holds since the fitness of any mutant is at least as large as that of any neighboring resident.

2. If  $\emptyset \subsetneq X_t \subsetneq V$  then  $\mathbb{P}[\Delta_t \leq -1/n | X_t] \geq \frac{1}{(1+\delta)n^2}$ : If the population is not homogeneous then there exists at least one edge  $(u, v) \in E$  with  $u \notin X_t$  and  $v \in X_t$ . With probability  $p_u = f_{X_t, S}(u)/F \geq \frac{1}{(1+\delta)n}$  the agent at  $u$  is selected for reproduction and with probability  $p_{u \rightarrow v} = 1/\deg(u) \geq 1/n$  its offspring migrates to  $v$ , which changes the potential function by  $-1/\deg(v) \leq -1/n$ .

By Item 1, the potential function  $\phi$  gives rise to a submartingale. Moreover, the function  $\phi$  is bounded by  $B = n$  and by Item 2, we have  $V = \text{Var}[\phi(X_{t+1}) | X_t] \geq p_u \cdot p_{u \rightarrow v} \cdot (1/n)^2 \geq \frac{1}{(1+\delta)n^4}$ . Thus, the standard martingale machinery of drift analysis applies (Kötzing and Krejca 2019). Namely, the rescaled function  $\phi \cdot (\phi - 2B) + B^2$  satisfies the conditions of the upper additive drift theorem (with initial value at most  $B^2$  and step-wise drift at least  $V$ ). The expected time until termination is thus at most

$$T(G^S, \delta) \leq \frac{B^2}{V} \leq (1 + \delta)n^6. \quad \square$$

**Lemma 7.** *Let  $G = (V, E)$  be a graph with  $n$  nodes  $u_1, \dots, u_n$  and  $m$  edges. Consider a function  $\alpha: V \rightarrow \mathbb{R}$  defined by  $\alpha(u_i) = \frac{1}{n} \cdot \sum_{j \in [n]} p_{ij} \cdot \pi_j \cdot \psi_{ij}$ , where  $\{\pi_i | u_i \in V\}$  is the solution to the linear system*

$$\pi_i = \left(1 - \sum_{j \in [n]} p_{ji}\right) \cdot \pi_i + \sum_{j \in [n]} p_{ij} \cdot \pi_j \quad \forall i \in [n], \quad (11)$$

and  $\{\psi_{ij} | (u_i, u_j) \in E\}$  is the solution to the linear system

$$\psi_{ij} = \begin{cases} \frac{1 + \sum_{\ell \in [n]} (p_{\ell i} \cdot \psi_{\ell j} + p_{\ell j} \cdot \psi_{i \ell})}{\sum_{\ell \in [n]} (p_{\ell i} + p_{\ell j})} & i \neq j \text{ and } (u_i, u_j) \in E \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

Then  $\text{fp}'(G^S, 0) = \sum_{u_i \in S} \alpha(u_i)$ .

*Proof.* Given a set  $S$ , we write  $\lambda_i = 1$  to indicate that  $u_i \in S$  (and  $\lambda_i = 0$  otherwise). Given any configuration  $X \subset V$ , we write  $x_i = 1$  to indicate that  $u_i \in X$  (and  $x_i = 0$  otherwise). We also define a function  $\phi(X) = \sum_{i \in [n]} \pi_i \cdot x_i$  and we let

$$\Delta(X, \delta) = \mathbb{E}[\phi(X_{t+1}) - \phi(X_t) | X_t = X] \quad (15)$$

be its expected change in a single step of the Moran process with fitness advantage  $\delta$  at nodes  $S$ . Let  $\bar{x} = \frac{1}{n} \sum_{i \in [n]} \lambda_i x_i$ . Then we can write

$$\Delta(X, \delta) = \sum_{i, j \in [n]} \frac{1 + \delta \cdot \lambda_i x_i}{n(1 + \delta \cdot \bar{x})} p_{ij} (x_i - x_j) \pi_j. \quad (16)$$

Finally, the derivative of  $\Delta(X, \delta)$  at  $\delta = 0$  is

$$\begin{aligned} \Delta'(X) &= \frac{d}{d\delta} \Big|_{\delta=0} \Delta(X, \delta) \\ &= \frac{1}{n} \sum_{i, j \in [n]} (\lambda_i \cdot x_i - \bar{x}) \cdot p_{ij} \cdot (x_i - x_j) \cdot \pi_j \\ &= \frac{1}{n} \sum_{i, j \in [n]} \lambda_i \cdot x_i \cdot p_{ij} \cdot (x_i - x_j) \cdot \pi_j \\ &\quad - \frac{\bar{x}}{n} \sum_{i, j \in [n]} p_{ij} \cdot (x_i - x_j) \cdot \pi_j \\ &= \frac{1}{n} \sum_{i, j \in [n]} \lambda_i \cdot p_{ij} \cdot \pi_j \cdot x_i (1 - x_j), \end{aligned} \quad (17)$$

where in the last equality we used that  $x_i^2 = x_i$  and that the second sum vanishes, since all the terms that involve any fixed  $x_i$  sum up to  $x_i \cdot \sum_{j \in [n]} (p_{ij} \cdot \pi_j - p_{ji} \cdot \pi_i) = 0$  (due to Eq. (11)).

Now consider the neutral positional Moran process ( $\delta = 0$ ) starting with  $\mathbb{P}[X_0 = \{u_i\}] = 1/n$  for  $i \in [n]$ . For any  $i, j \in [n]$  and  $t \geq 0$ , consider an event  $Y_{ij}(t)$  defined as “at time  $t$  we have  $u_i \in X_t$  and  $u_j \notin X_t$ ” and let  $\psi_{ij} = \sum_{t=0}^{\infty} \mathbb{P}[Y_{ij}(t)]$  be the total expected time spent with  $u_i$  being a mutant and  $u_j$  not. Then by (McAvoy and Allen 2021, Theorem 1) we have

$$\begin{aligned} \text{fp}'(G^S, 0) &= \sum_{t=0}^{\infty} \sum_{X \subset V} \mathbb{P}[X_t = X] \cdot \Delta'(X) \\ &= \frac{1}{n} \sum_{i,j \in [n]} \lambda_i \cdot p_{ij} \cdot \pi_j \cdot \psi_{ij} = \sum_{u_i \in S} \alpha(u_i), \end{aligned}$$

It remains to show that the quantities  $\psi_{ij}$  satisfy the linear system of Eq. (12). For  $i = j$  we clearly have  $\psi_{ii} = 0$ . For  $i \neq j$  we write

$$\psi_{ij} = \mathbb{P}[Y_{ij}(0)] + \sum_{t=0}^{\infty} \mathbb{P}[Y_{ij}(t+1)]. \quad (18)$$

The first term is simply  $1/n$ . Let  $e_{ij} = \frac{1}{n} p_{ij}$  be the probability that, in a single step of the standard Moran process ( $\delta = 0$ ), node  $u_i$  is selected for reproduction and places its offspring on  $u_j$ . Then we can rewrite each term of the sum in terms of events at time  $t$  as follows:

$$\begin{aligned} \mathbb{P}[Y_{ij}(t+1)] &= \sum_{l \in [n]} e_{li} \cdot \mathbb{P}[Y_{lj}(t)] + \sum_{l \in [n]} e_{lj} \cdot \mathbb{P}[Y_{il}(t)] \\ &\quad + \left(1 - \sum_{l \in [n]} (e_{li} + e_{lj})\right) \cdot \mathbb{P}[Y_{ij}(t)]. \end{aligned}$$

Summing over  $t$  and plugging this into Eq. (18) we get

$$\psi_{ij} = \frac{\frac{1}{n} + \sum_{l \in [n]} e_{li} \cdot \psi_{lj} + \sum_{l \in [n]} e_{lj} \cdot \psi_{il}}{\sum_{l \in [n]} e_{li} + \sum_{l \in [n]} e_{lj}},$$

thus we arrive at the the linear system of Eq. (12).

This concludes the proof.  $\square$

**Graph from weak-selection experiments.** Fig. 6 illustrates a challenging graph for  $\text{FM}^0$ , along with the activation choices of different heuristics.

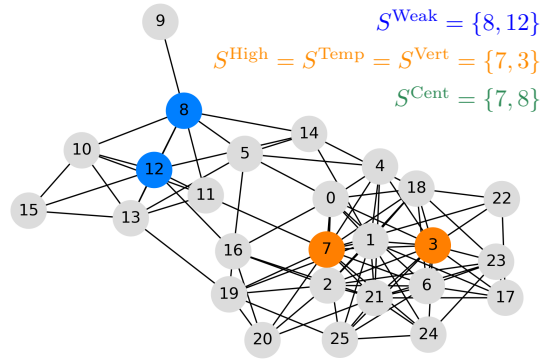


Figure 6: Node activation for weak selection on a challenging graph of  $n = 26$  nodes, that shows as an outlier in Fig. 5. The budget is  $k = \lfloor 10\% \cdot n \rfloor = 2$ . The choices of the different heuristics are shown.