

Fixation Maximization in the Positional Moran Process

Joachim Brendborg,¹ Panagiotis Karras,¹ Andreas Pavlogiannis,¹
Asger Ullersted Rasmussen,¹ Josef Tkadlec²

¹Aarhus University, Aabogade 34, Aarhus, Denmark

²Harvard University, 1 Oxford Street, Cambridge, USA

{panos, 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 $O(n^{2\omega})$ time, where ω 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 (????); see (?) for a survey. Similarly, in rumor propagation (?), the goal is to minimize the number of rounds until a rumor spreads to the whole population (?).

A related propagation model is that of the Moran process, introduced as a model of genetic evolution (?), and the variants thereof, such as the discrete Voter model (????). A single *mutant* (e.g., a new opinion, strain, social trait, etc.) invades a population of *residents* scattered across the nodes of the graph. Contrariwise 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 δ 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 (?), the process has spawned extensive work on identifying *amplifiers*, i.e., graph structures that enhance the fixation probability (????). Still, in the real world, the probability that a mutation spreads to its neighborhood is often affected by its position: the ability to metabolise 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 (?), 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 δ 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 FM^∞ and 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 FM^∞ 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 δ 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 δ , the function $\text{fp}^\infty(G^S)$ is submodular on undirected graphs and thus FM^∞ admits a polynomial-time constant-factor approximation (Section 4.2). Lastly, regarding the limit $\delta \rightarrow 0$, we show that 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 the ??.

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 (?) 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 δ 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 (?) 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 (?). It is different from models in which agent's fitness depends on which agents occupy the neighboring nodes (so-called frequency dependent fitness) (?).

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 δ 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 δ and infinitely differentiable: the underlying Markov chain is finite and its transition probabilities are rational functions in δ , so $\text{fp}(G^S, \delta)$ too is a rational function in δ , 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 (?). 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) (???), 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 δ .

Fixation maximization. Consider a graph $G = (V, E)$, a mutant fitness advantage δ and a budget $k \in \mathbb{N}$. The problem *Fixation Maximization for the Positional Moran Process*

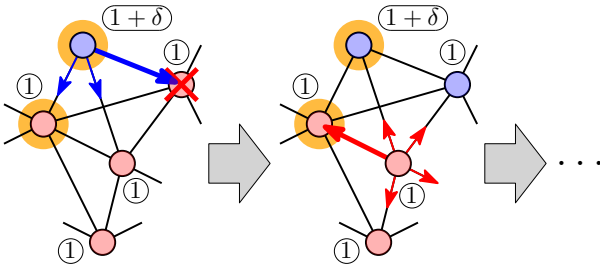


Figure 1: Steps in the positional Moran process. Blue (red) nodes are occupied by mutants (residents). Yellow rings indicate active nodes that realize the mutant fitness advantage.

$\text{FM}(G, \delta, k)$ 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 δ 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 δ , 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 δ , different sets S generally give different fixation probabilities $\text{fp}(G^S, \delta)$, see Fig. 3. For fixed $S \subseteq V$, we can view $\text{fp}(G^S, \delta)$ as a function of δ . 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 δ 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 $\text{fp}'(G^S, 0)$:

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

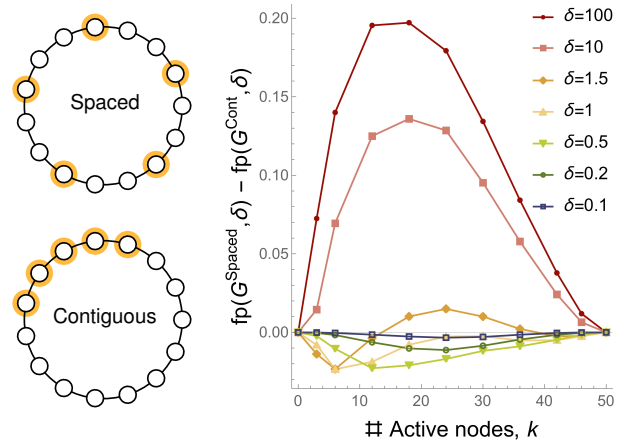


Figure 2: FP differences on a cycle graph on $n = 50$ nodes, under spaced and contiguous activating strategy. E.g. for $\delta = 100$ and $k = 18$ we have $\text{fp}(G^{\text{Spaced}}, \delta) \doteq 0.62$ and $\text{fp}(G^{\text{Cont}}, \delta) \doteq 0.43$, so the difference is almost 20%.

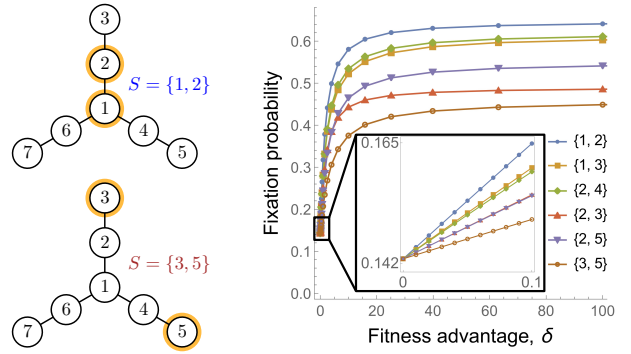


Figure 3: A fixed graph G on 7 nodes and fixation probability $\text{fp}(G^S, \delta)$ for all six non-equivalent subsets S of $k = 2$ nodes; $\text{fp}(G^S, \delta)$ is roughly linear in δ when $\delta \in [0, 0.1]$.

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

Monotonicity. Enlarging the set S or the advantage δ increases the fixation probability. The proof generalizes the standard Moran process monotonicity (?).

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 FM^∞ .

3.1 Hardness of FM and FM[∞]

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

Theorem 1. *FM[∞](G, k) and FM(G, δ, 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 (?). 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 (?), 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 $\delta \geq 0$, 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 $\delta \geq 0$, 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 (??).

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 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 (?).

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 $\mathcal{O}(n^2)$ linear equations, one for each pair of nodes. This can be done in matrix-multiplication time $\mathcal{O}(n^{2\omega})$ for $\omega < 2.373$ (?). 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 and some integer k , the solution to $\text{FM}^0(G, k)$ can be computed in $\mathcal{O}(n^{2\omega})$ 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 ??.

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 . 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 given by $\sum_{i \in [n]} \pi_i = 1$ together with*

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

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

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

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

The intuition behind the quantities π_i , ψ_{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 π_i is in fact the mutant fixation probability starting from the initial configuration $X_0 = \{u_i\}$ (?). 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 $\pi_i = \frac{1/\deg(u_i)}{\sum_{j \in [n]} 1/\deg(u_j)}$ (?).)

The values ψ_{ij} for $i \neq j$ 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_ℓ 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_ℓ was mutant but u_j was not, hence the product $p_{\ell i} \cdot \psi_{\ell j}$. Similar reasoning applies to the product $p_{\ell j} \cdot \psi_{\ell i}$. 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. δ 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 (ψ_{ij}), calibrated by the probability that, when chosen for reproduction, u_i propagates to that neighbor (p_{ij}), who thereafter achieves mutant fixation (π_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 (?). 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 δ . 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 (?), 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 δ .

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

Strong selection. We start with the case of $\text{FM}^\infty(G, K)$. 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

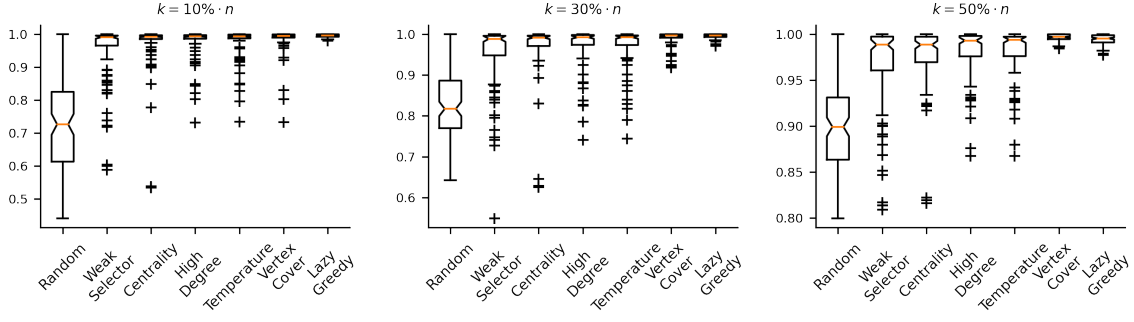


Figure 4: Heuristic performance for FM^∞ .

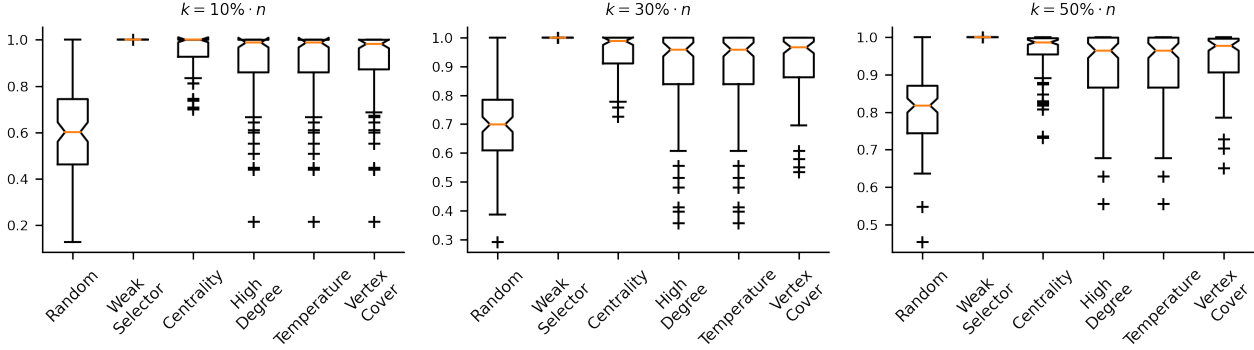


Figure 5: Heuristic performance for FM^0 .

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 δ be ap-

proximated within some constant-factor? Are there classes of graphs for which it becomes tractable? Voluptatum dolores distinctio iusto commodi reiciendis a nobis, commodi ut magni cumque repellat facilis molestiae illum deserunt hic rem sint. Aliquid deleniti enim, deleniti possimus adipisci maiores voluptatem, soluta ad nemo itaque praesentium preferendis temporibus illum fugiat? Nulla nemo itaque necessitatibus, tempora quae earum doloribus minima aut quos pariat ut incidunt, nemo nostrum facilis culpa totam ab cum, optio nulla cupiditate. Aperiam sint sed ex dicta magni soluta eius, reprehenderit minus veniam fugiat, voluptatibus repudiandae vel, eveniet nihil et aliquid dolorem totam ipsa? Voluptatem consequuntur numquam omnis fugit similique ullam, officia maxime reiciendis labore ut accusamus ipsum consequuntur quae, quae harum possimus deleniti ad sapiente odio dolor pariat ut molestiae. Expedita ab officiis ex ipsam preferendis temporibus dolorum optio voluptate, incidunt inventore corrupti quam fuga illum iste numquam aliquid sequi, iure quae illo reiciendis eius soluta quaerat animi rem dolorum, voluptate quo nihil minima harum quibusdam quas, possimus tempora nisi ab quod voluptatum? Earum quae dolorum neque assumenda aut eaque cum asperiores ipsa itaque ut, molestiae ratione dolores nesciunt at ipsa deleniti. Nemo in ipsa quis cum est, minus iste nesciunt, maiores magni est dolor excepturi? Fugit quod eligendi delectus, dicta voluptates autem minus?