Seed Selection in the Heterogeneous Moran Process

Petros Petsinis[1], Andreas Pavlogiannis[1], Josef Tkadlec[2], Panagiotis Karras[3,1]

AARHUS UNIVERSITY

[1]Department of Computer Science, Aarhus University (AU)

[2]Computer Science Institute, Charles University (CU)

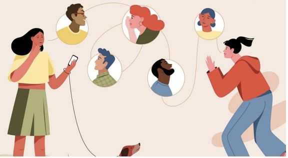
[3]Department of Computer Science, Copenhagen University (KU)

Natural spread through networks

- propagation of information in social networks
- **spread of mutation** in biological networks

Type of Diffusion process

1. **Progressive:**
   independent cascade,
   linear threshold,...

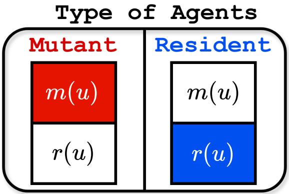


2. **Non-Progressive:**
   Moran, Voter,...

This paper: Non-Progressive model that describes the spread of mutation/novel-trait.

Heterogeneous Moran Process - This Work

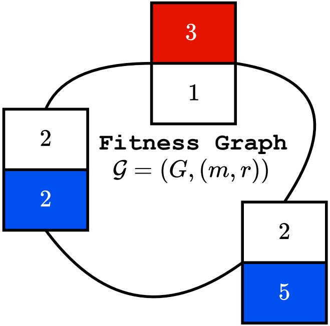
Graph: Population of n agents spread over nodes of graph $G = (V, E, w)$.



Fitness

$f_X(u) = m(u), u \in X$

$f_X(u) = r(u), u \in \neg X$

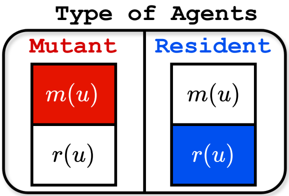


⇒ $t = 0$: mutants $X = S$ appear in the network.

⇒ $t > 0$: repeat **Birth-Death** steps until $X = V$ or $\neg X = V$ ($X = \varnothing$):

1. **Birth:** Pick a node $\boldsymbol{u}$ proportionally to its fitness, $\frac{f_X(u)}{\sum_{v \in V} f_X(v)}$.
2. **Death:** Pick an out-neighbor node $\boldsymbol{v}$ of $\boldsymbol{u}$ proportionally to edge-weight $\boldsymbol{w(u,v)}$ and transfer the trait/type of $\boldsymbol{u}$ on $\boldsymbol{v}$.

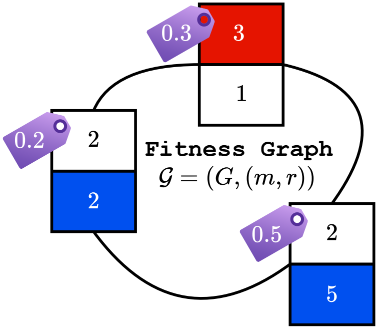
Graph: Population of n agents spread over nodes of graph $G = (V, E, w)$.



Fitness

$f_X(u) = m(u), u \in X$

$f_X(u) = r(u), u \in \neg X$



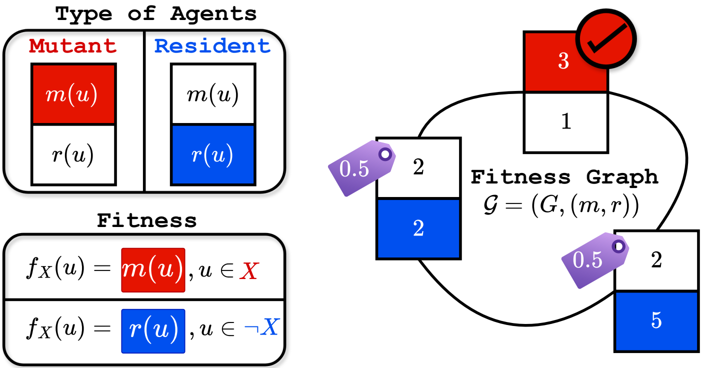
⇒ $t = 0$: mutants $X = S$ appear in the network.

⇒ $t > 0$: repeat **Birth-Death** steps until $X = V$ or $\neg X = V$ ($X = \varnothing$):

1. **Birth:** Pick a node $\boldsymbol{u}$ proportionally to its fitness, $\frac{f_X(u)}{\sum_{v \in V} f_X(v)}$.
2. **Death:** Pick an out-neighbor node $\boldsymbol{v}$ of $\boldsymbol{u}$ proportionally to edge-weight $\boldsymbol{w(u,v)}$ and transfer the trait/type of $\boldsymbol{u}$ on $\boldsymbol{v}$.

Heterogeneous Moran Process - This Work

Graph: Population of n agents spread over nodes of graph $G = (V, E, w)$.



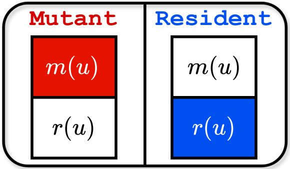
⇒ $t = 0$: mutants $X = S$ appear in the network.

⇒ $t > 0$: repeat **Birth-Death** steps until $X = V$ or $\neg X = V$ ($X = \varnothing$):

1. **Birth:** Pick a node $\boldsymbol{u}$ proportionally to its fitness, $\frac{f_X(u)}{\sum_{v \in V} f_X(v)}$.
2. **Death:** Pick an out-neighbor node $\boldsymbol{v}$ of $\boldsymbol{u}$ proportionally to edge-weight $\boldsymbol{w(u, v)}$ and transfer the trait/type of $\boldsymbol{u}$ on $\boldsymbol{v}$.

Graph: Population of n agents spread over nodes of graph $G = (V, E, w)$.

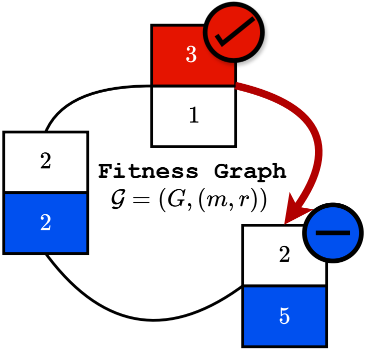
Type of Agents



Fitness

$f_X(u) = m(u), u \in X$

$f_X(u) = r(u), u \in \neg X$



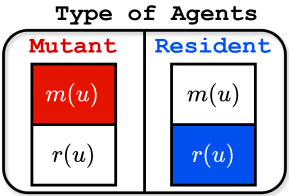
⇒ $t = 0$: mutants $X = S$ appear in the network.

⇒ $t > 0$: repeat **Birth-Death** steps until $X = V$ or $\neg X = V$ ($X = \varnothing$):

1. **Birth:** Pick a node $\boldsymbol{u}$ proportionally to its fitness, $\frac{f_X(u)}{\sum_{v \in V} f_X(v)}$.
2. **Death:** Pick an out-neighbor node $\boldsymbol{v}$ of $\boldsymbol{u}$ proportionally to edge-weight $\boldsymbol{w(u,v)}$ and transfer the trait/type of $\boldsymbol{u}$ on $\boldsymbol{v}$.

Heterogeneous Moran Process - This Work

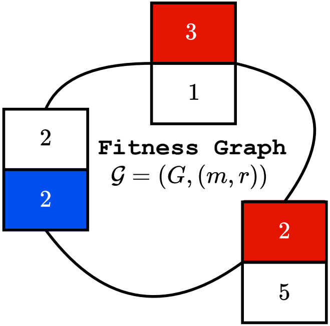
Graph: Population of n agents spread over nodes of graph $G = (V, E, w)$.



Fitness

$f_X(u) = m(u), u \in X$

$f_X(u) = r(u), u \in \neg X$



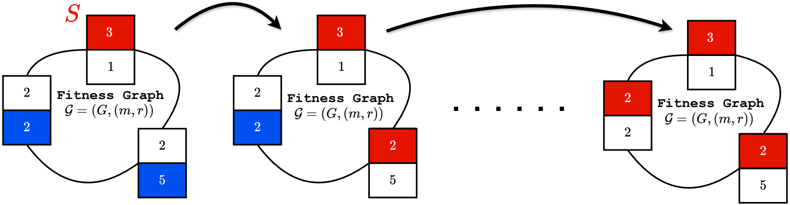
⇒ $t = 0$: mutants $X = S$ appear in the network.

⇒ $t > 0$: repeat **Birth-Death** steps until $X = V$ or $\neg X = V$ ($X = \emptyset$):

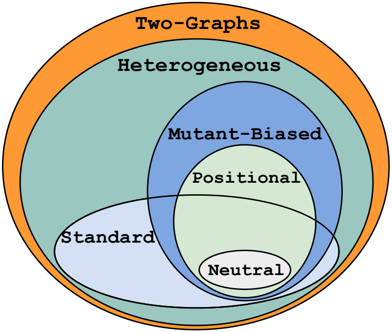
1. **Birth:** Pick a node $\boldsymbol{u}$ proportionally to its fitness, $\frac{f_X(u)}{\sum_{v \in V} f_X(v)}$.
2. **Death:** Pick an out-neighbor node $\boldsymbol{v}$ of $\boldsymbol{u}$ proportionally to edge-weight $\boldsymbol{w(u,v)}$ and transfer the trait/type of $\boldsymbol{u}$ on $\boldsymbol{v}$.

Setting Parameters: Fitness graph $\mathcal{G}$ and a seed set of mutants S.

Fixation Probability: The probability $\text{fp}_{\mathcal{G}}(S)$ that a seed set of mutants S leads to fixation.



Relation to other Moran Processes



Neutral: $m(u) = r(u) = \delta, \ \forall u \in V$.

Standard: $m(u) = 1 + \delta$ and $r(u) = 1, \ \forall u \in V, \delta \geq -1$.

Positional: $m(u) = 1 + \delta, \ \forall u \in A \subseteq V$ and $r(u) = 1, \ \forall u \in V, \delta \geq 0$.

Mutant-Biased: $m(u) \geq r(u), \ \forall u \in V$.

Heterogeneous: $m(u)$ and $r(u), \ \forall u \in V$.

Two-Graphs: type-dependent fitness graphs $\mathcal{G}_m$ and $\mathcal{G}_r$.

Seed Selection - Fixation Maximization

Optimization Problem: Given a fitness graph $\mathcal{G}$ and a budget k, which k nodes S^* should initiate the mutant invasion so as to maximize the fixation probability?

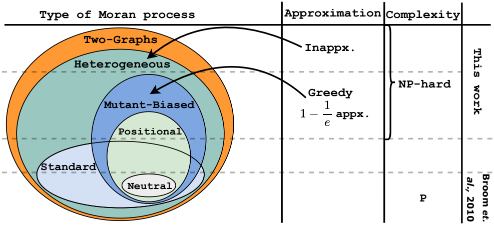
$$S^* = \arg\max_{S,|S|=k} \mathrm{fp}_{\mathcal{G}}(S)$$

Optimization Problem: Given a fitness graph $\mathcal{G}$ and a budget k, which k nodes S^* should initiate the mutant invasion so as to maximize the fixation probability?

$$S^* = \arg\max_{S,|S|=k} \mathrm{fp}_{\mathcal{G}}(S)$$

Results Overview:

1. **FPRAS:** for undirected mutant-biased $\mathcal{G}$.
2. **Inapproximability: NP**-hard to distinguish between maximum fixation probability ϵ and $1-\epsilon$.
3. **NP-hardness** of finding $S^* = \arg\max_{S,|S|=k} \mathrm{fp}_{\mathcal{G}}(S)$ on mutant-biased $\mathcal{G}$.
4. **Approximations** for mutant-biased $\mathcal{G}$; proving monotonicity and submodularity.



The complexity of computing $\mathrm{fp}_{\mathcal{G}}(S)$ is OPEN even for the standard model.

Lemma 1 - Expected Time

For undirected mutant-biased fitness graphs, the expected time to a homogeneous state ($X = V$ *or* $\neg X = V$) *is* $\mathcal{O}\left(n^2 \frac{m_{\max}}{r_{\min}}\right)^3$.

Approximate $\mathrm{fp}_{\mathcal{G}}(S)$ via monte-carlo simulations in P-time.

Theorem 1 - Inapproximation

*For any $0 < \epsilon < 1/2$, it is **NP**-hard to distinguish between instances with $\max_S \mathrm{fp}_\mathcal{G}(S) \leq \epsilon$ and those with $\max_S \mathrm{fp}_\mathcal{G}(S) > 1 - \epsilon$.*

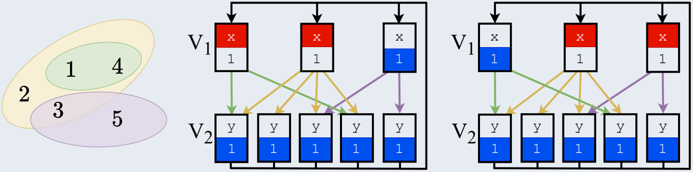
Inapproximability

Theorem 1 - Inapproximation

For any $0 < \epsilon < 1/2$, *it is* ***NP****-hard to distinguish between instances with* $\max_S \mathrm{fp}_\mathcal{G}(S) \leq \epsilon$ *and those with* $\max_S \mathrm{fp}_\mathcal{G}(S) > 1 - \epsilon$.

Proof.

Reduction from **Set Cover**; **NP**-hard to distinguish between maximum fixation probability $\boldsymbol{\leq \epsilon}$ (¬Set Cover) and $\boldsymbol{> 1 - \epsilon}$ (Set Cover). There exist $y = 1/\mathcal{O}(n^3)$ and $x = \mathcal{O}(n^{10})$ such that:



$\mathrm{fp}_\mathcal{G}(S) \leq \epsilon$ $\qquad$ $\mathrm{fp}_\mathcal{G}(S) > 1 - \epsilon$

□

NP-hardness for Mutant-Biased $\mathcal{G}$

Theorem 2 - NP-hardness

*For mutant-biased $\mathcal{G}$, it is **NP**-hard to distinguish between instances with $\max_S \mathrm{fp}_{\mathcal{G}}(S) \leq 1 - n^{2n}$ and those with $\max_S \mathrm{fp}_{\mathcal{G}}(S) > 1 - n^{2n}$.*

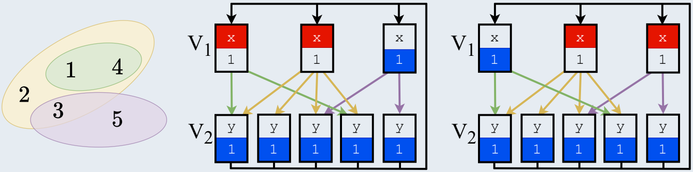
NP-hardness for Mutant-Biased $\mathcal{G}$

Theorem 2 - NP-hardness

*For mutant-biased $\mathcal{G}$, it is **NP**-hard to distinguish between instances with $\max_S \mathrm{fp}_{\mathcal{G}}(S) \le 1 - n^{2n}$ and those with $\max_S \mathrm{fp}_{\mathcal{G}}(S) > 1 - n^{2n}$.*

Proof.

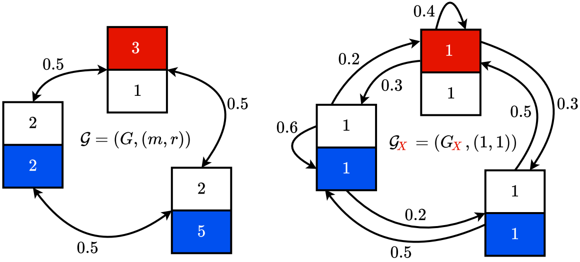
Reduction from **Set Cover**; **NP**-hard to distinguish between maximum fixation probability $\le \mathbf{1 - n^{2n}}$ ($\neg$Set Cover) and $> \mathbf{1 - n^{2n}}$ (Set Cover). There exist $y = \mathcal{O}(1)$ and $x = 2^{\mathcal{O}(n \log n)}$ such that:



$\mathrm{fp}_{\mathcal{G}}(S) \le 1 - n^{2n}$ $\qquad$ $\mathrm{fp}_{\mathcal{G}}(S) > 1 - n^{2n}$ $\square$

Loopy Process: In each time t, with mutants $X_t = X$, the Birth-Death process runs on $\mathcal{G}_X = (G_X, (\mathbf{1}, \mathbf{1}))$ with:

$$w_X(u,v) = \begin{cases} \frac{f_X(u)}{f_{\max}} \cdot w(u,v), & \text{if } u \neq v \\ 1 - \frac{f_X(u)}{f_{\max}}(1 - w(u,v)), & \text{if } u = v \end{cases}$$



Lemma 2 - Loopy Process

For any seed set, the Heterogeneous and Loopy Moran processes share the same fixation probability.

Monotonicity and Submodularity

Corollary 1- Monotonicity (Two-Graphs)

For any mutant-biased $\mathcal{G}$ and any two seed sets $S \subseteq S'$, it holds $\mathrm{fp}_{\mathcal{G}}(S) \leq \mathrm{fp}_{\mathcal{G}}(S')$ [Melissourgos et al., 2022, Corollary 6].

Proof.

Using Loopy-Process, we prove Heterogeneous ⊂ Two-Graphs. □

Lemma 3 - Submodularity

For any mutant-biased $\mathcal{G}$, function $\mathrm{fp}_{\mathcal{G}}(S)$ is submodular.

Proof.

Loopy-Processes $\mathrm{fp}_{\mathcal{G}}(S)$, $\mathrm{fp}_{\mathcal{G}}(T)$, $\mathrm{fp}_{\mathcal{G}},(S \cup T)$ and $\mathrm{fp}_{\mathcal{G}}(S \cap T)$. At time t, node u reproduces with equal probability in all cases; examine the probability that $|X_{t+1}| \geq |X_t|$ and prove:

$$\mathrm{fp}_{\mathcal{G}}(S) + \mathrm{fp}_{\mathcal{G}}(T) \geq \mathrm{fp}_{\mathcal{G}}(S \cup T) + \mathrm{fp}_{\mathcal{G}}(S \cap T).$$ □

Approximations

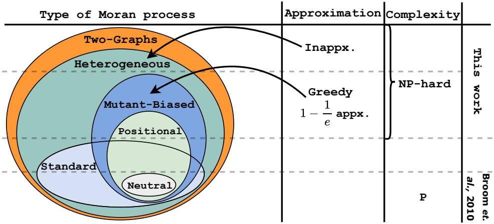
Corollary 2 - Approximations

For undirected mutant-biased $\mathcal{G}$, function $\mathrm{fp}_{\mathcal{G}}(S)$ is:

Monotone + ***Submodular***

$$\Downarrow$$

(1-1/e) *greedy approximation algorithm [Nemhauser, 1978]*



1. **FPRAS:** for undirected mutant-biased $\mathcal{G}$.
2. **Inapproximability: NP**-hard to distinguish between maximum fixation probability ϵ and $1 - \epsilon$.
3. **NP-hardness** of finding $S^* = \arg\max_{S, |S|=k} \mathrm{fp}_{\mathcal{G}}(S)$ on mutant-biased $\mathcal{G}$.
4. **Approximations** for mutant-biased $\mathcal{G}$; proving monotonicity and submodularity.

Thank you!