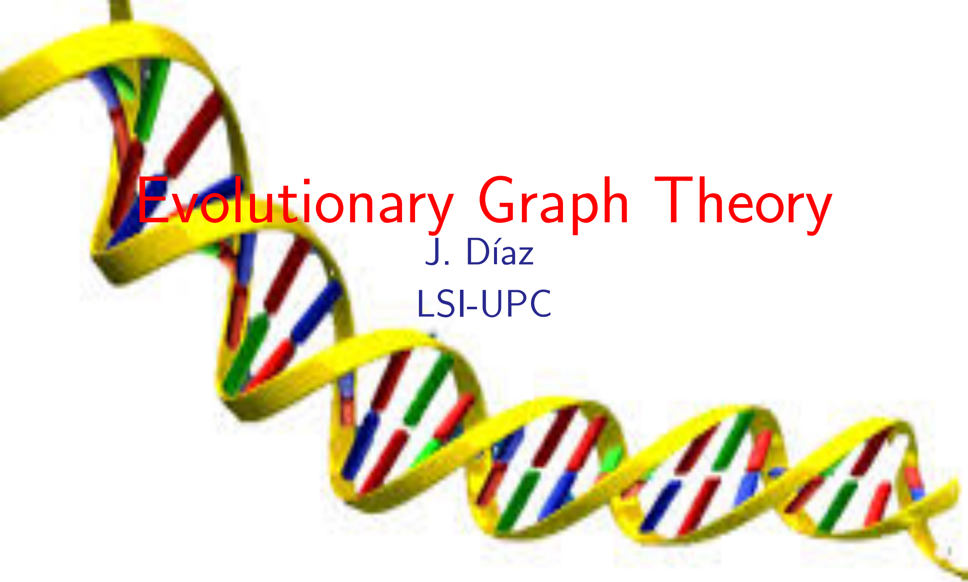


Evolutionary Graph Theory

J. Díaz

LSI-UPC

Nice, May, 2014

Population Genetics Models

Model the forces that produce and maintain genetic evolution within a population.

Mutation: the process by which one individual (gene) changes.
Simulation wants to study the drift of the population: how the frequency of mutants in the total population evolves.

The Moran Process P. Moran: *Random processes in genetics* Cambridge Ph. Soc. 1958

- Start with n individuals. Randomly select one to mutate.
- Select randomly an individual x to replicate.
- Select randomly another y to die.
- Replace y by a clone of x.

Stochastic process. At time t the number mutants evolves in $\{-1, 0, +1\}$.

Evolutionary graph theory (EGT)

Lieberman, Hauert, Nowak: *Evolutionary dynamics on graphs* Nature 2005 (LHN)

EGT studies how the topology of interactions between the population affects evolution.

Graphs have two types of vertices: mutants and non-mutants.

The fitness r of an agent denotes its reproductive rate.
Mutants have fitness $r \in \Theta(1)$, non-mutants have fitness 1.

Mutants and non-mutants extend by cloning one of their neighbors.

Moran process on Evolutionary Graphs

Given a graph $G = (V, E)$, with $|V| = n$, and an $r > 0$, we start with all vertices non-mutant.

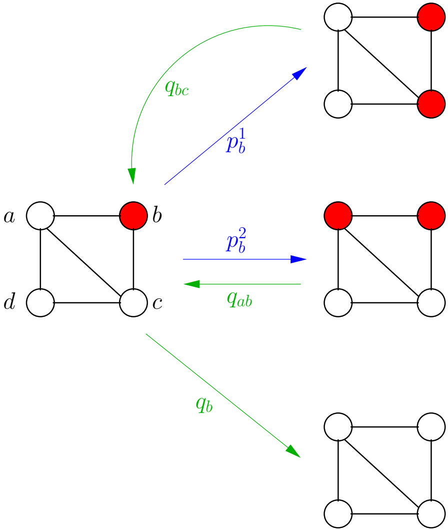
- at $t = 0$ create uniformly at random a mutant in V

At any time $t > 0$, assume we have k mutant and $(n - k)$ non-mutant vertices. Define total fitness at time t by $W_t = kr + (n - k)$:

- Choose u with probability $\frac{r}{W_t}$ if u is mutant and $\frac{1}{W_t}$ otherwise,
- choose uniformly at random a $v \in \mathcal{N}(u)$, and replace v with the clone of u

The process is Markovian, depending on r it tends to one of the two absorbing states: extinction or fixation.

Example of Moran process



where:

$$p_b^1 = \frac{r}{3+r} \cdot \frac{1}{2}$$

$$p_b^2 = \frac{r}{3+r} \cdot \frac{1}{2}$$

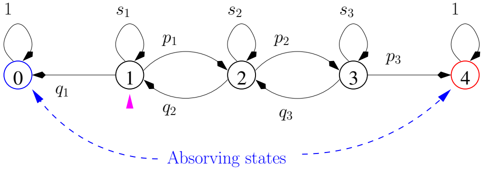
$$q_{ab} = \frac{1}{2+2r} \cdot \frac{5}{6}$$

$$q_{bc} = \frac{1}{(n-1)+r} \cdot \frac{5}{6}$$

$$q_b = \frac{2}{3+r} \cdot \frac{1}{3}$$

Moran Process

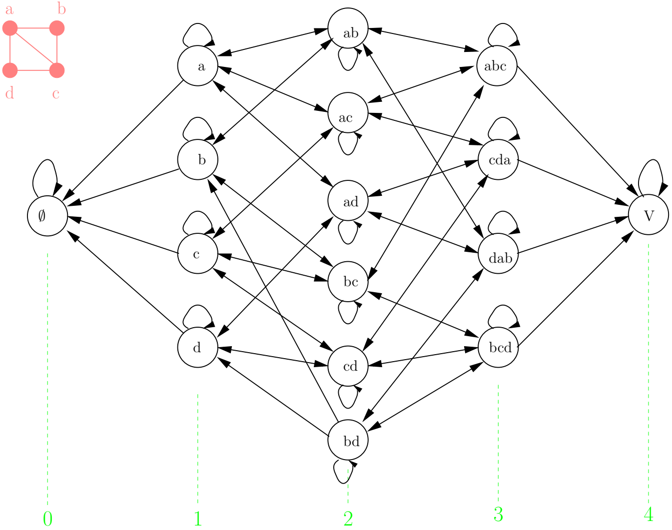
This random process defines discrete, transient Markov chain, on states $\{0, 1, \ldots, n-1, n\}$ with two absorbing states: n fixation (all mutant) and 0 extinction (all non-mutant).



The fixation probability $f_G(r)$ of G is the probability that a single mutant will takes over the whole G. The extinction probability of G is $1 - f_G(r)$.

The Markov chain of configurations

A configuration is a set $S \subseteq V$ of mutants.



Properties of $f_G(r)$

Given $G = (V, E)$ connected and a fitness $r > 0$, for any $S \subset V$ let $f_{G,r}(S)$ denote the fixation probability, when starting with a set S of mutants.
Notice $f_G(r) = \sum_{v \in V} f_{G,r}(\{v\})$.

The case $r = 1$ is denoted neutral drift.

Shakarian, Ross, Johnson, Biosystems 2012
For any $r \geq 1$, $f_G(r) \geq f_G(1)$

Díaz,Goldberg,Mertzios,Richerby,Serna,Spirakis, SODA-2012 (DGMRSS)
For any undirected $G = (V, E)$, $f_G(1) = \frac{1}{n}$.

Bounding $f_G(r)$

Let $G = (V, E)$ be any undirected connected graph, with $|V| = n$.

(DGMRSS)
For any $r \geq 1$, $\frac{1}{n} \leq f_G(r) \leq 1 - \frac{1}{n+r}$, are bounds on the fixation probability for G.

Merzios, Spirakis: ArXive-2014
For any $\epsilon > 0$,

$$f_G(r) \leq 1 - \frac{1}{n^{\frac{3}{4}+\epsilon}}.$$

Open problem: There are not known upper bounds that don't depend on n.
Conjecture: $f_G(r) \leq 1 - \frac{1}{r}$

Questions to study

Given a connected graph $G = (V, E)$ (strongly connected is case of digraphs), and a fitness r:

1.- Is it possible to compute exactly the fixation probability $f_G(r)$?

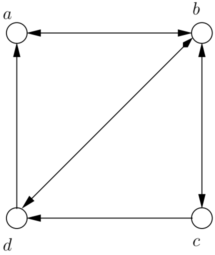
Difficult for some graphs. For a given G the number of constrains and variables is equal to the number of possible configurations of mutants/non-mutants in $G \sim 2^n$.

2.- Given G, is it possible to compute the expected number of steps until arriving to absorption?

Isothermal graphs (LHN)

Given a directed $\vec{G} = (V, \vec{E})$, $\forall i \in V$ let $\deg^+(i)$ be its outgoing degree:
Define the stochastic matrix $W = [w_{ij}]$, where $w_{ij} = 1/\deg^+(i)$ if $(\vec{i,j}) \in \vec{E}$ and $w_{ij} = 0$ otherwise.
The same definition of W applies to undirected G, with $w_{ij} = 1/\deg(i)$.

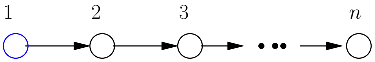
The temperature of $i \in V$ is $T_i = \sum_{j \in V} w_{ji}$
A graph $\vec{G}$ is isothermal if $\forall i, j \in V$, $T_i = T_j$.



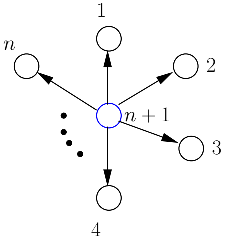
$$W = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$$

$T_b = 2$ and $T_c = 1/3$

Computing the fixation probability



If $\vec{G}$ is a digraph with a single source then $f_{\vec{G}}(r) = \frac{1}{n}$.



Isothermal Theorem (LHN)
For a strongly connected graph $\vec{G}$ s.t. $\forall i, j \in V$ we have $T_i = T_j$ (i.e. W is bi-stochastic) then

$$f_{\vec{G}}(r) = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^n}} \equiv \rho$$

Undirected graphs

The isothermal theorem also applies to undirected graphs.
Given G undirected and connected, then
G is Δ-regular iff W is bi-stochastic.

If G is undirected and connected then
$f_G(r) = \rho = \frac{1-1/r}{1-1/r^n}$ *iff* G *is* Δ*-regular*.

For example, if G is C_n or K_n then $f_G(r) = \rho$.

Notice:

- if $r > 1$ then $\lim_{n\to\infty} f_G(r) = 1 - \frac{1}{r}$.
- if $r < 1$ then $f_G(r) = \frac{r^n - r^{n-1}}{r^n - 1} \to$ exponentially small.

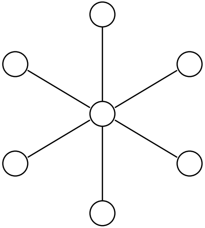
Amplifiers and suppressors

Given G (directed or undirected) and r, G is said to be an amplifier if $f_G(r) > \rho$. G is said to be a suppressor if $f_G(r) < \rho$.

The star

(LHN), (Broom, Rychtá. Proc.R. Soc. A 2008)

For $r > 1$ $f_G(r) = \frac{1-\frac{1}{r^2}}{1-\frac{1}{r^{2n}}} > \rho$

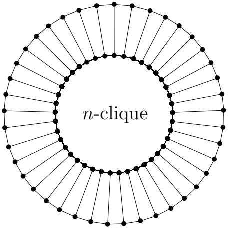


The star is an amplifier

Suppressors

The directed line and the burst have fixation probability $\frac{1}{n} < \rho$, therefore they are examples of suppressors.

How about non-directed graphs as suppressors?
Mertzios, Nikoletseas,Ratopoulos,Spirakis, TCS 2013
The urchin



For $< r < 4/3$
$\lim_{n\to\infty} f_G(r) = \frac{1}{2}(1 - \frac{1}{r}) < \rho$

The urchin is an undirected graph suppressor

Absorption time for undirected graphs

Given undirected connected $G = (V, E)$, with $|V| = n$, run a Moran process $\{S_t\}_{t\geq 0}$, where $\{S_t\}$ set of mutants at time t.

Define the absorption time $\tau = \min\{t \mid S_t = \emptyset \vee S_t = V\}$.

Theorem DGMRSS

Given G undirected, for the Moran process $\{S_t\}$ starting with $|S_1| = 1$:

1. If $r < 1$, then $\mathbf{E}[\tau] \leq \frac{r}{r-1} n^3$,
2. if $r > 1$, then $\mathbf{E}[\tau] \leq \frac{r}{r-1} n^4$,
3. if $r = 1$, then $\mathbf{E}[\tau] \leq n^6$.

Sketch of the proof

We bound $\mathbf{E}[\tau]$ using a potential function that decreases in expectation until absorption.

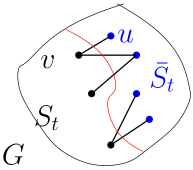
Define the potential function $\phi(S) = \sum_{v \in S} \frac{1}{\deg(v)}$
Notice $\phi(\{v\}) \geq 1/n$ and $0 \leq \phi(S_\tau) \leq n$

Use the following result from MC (Hajek, Adv Appl. Prob. 1983)

If $\{X_t\}_{t \geq 0}$ is a MC with state space Ω and there exist constants $k_1, k_2 > 0$ and a $\phi : \Omega \to \mathbb{R}^+ \cup \{0\}$ s.t.
(1) $\phi(S) = 0, \exists S \in \Omega$,
(2) $\phi(S) \leq k_1$,
(3) $\mathbf{E}[\phi(X_t) - \phi(X_{t+1}) \mid X_t = S] \geq k_2, \forall t \geq 0$ s.t. $\phi(S) > 0$,
then $\mathbf{E}[\tau] \leq k_1/k_2$, where $\tau = \min\{t \mid \phi(S) = 0\}$.

Sketch of the proof

To compute evolution of
$\mathbf{E}\left[\phi(S_{t+1}) - \phi(S_t)\right]$.
To show that the potential decreases (increases) monotonically for $r < 1$ ($r > 1$), consider the contribution of each (u, v) in the cut for $S_{t+1} = S_t \cup \{v\}$ and to $S_{t+1} = S_t \backslash \{v\}$.



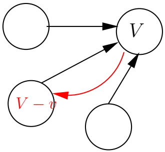
1. For $r < 1$, $\mathbf{E}\left[\phi(S_{t+1}) - \phi(S_t)\right] < \frac{r-1}{n^3} < 0$.
2. For $r > 1$, $\mathbf{E}\left[\phi(S_{t+1}) - \phi(S_t)\right] \geq (1 - \frac{1}{r})\frac{1}{n^3}$.
3. For $r = 1$, $\mathbf{E}\left[\phi(S_{t+1}) - \phi(S_t)\right] = 0$.

Domination argument for $r < 1$

For any fixed initial $S \subset V$:

Let $\{Y_i\}_{i \geq 0}$ be a stochastic process as Moran's, except if it arrives to state V, u.a.r. choose v and exit to state $V \backslash \{v\}$.

Let $\tau' = \min\{i \mid Y_i = \emptyset\}$



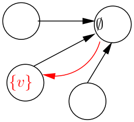
Then,

$$\mathbf{E}\left[\tau \mid X_0 = S\right] \leq \mathbf{E}\left[\tau' \mid Y_0 = S\right] \leq \frac{1}{1-r} n^3 \phi(S)$$

$\Rightarrow \mathbf{E}\left[\tau\right] \leq \frac{1}{1-r} n^3.$

Domination argument for $r > 1$

For any fixed initial $S \subset V$:
Define a process $\{Y_i\}_{i \geq 0}$ as in Moran's, except if arrives to state $\emptyset$, u.a.r. choose v and exit to state $\{v\}$.



Let $\tau' = \min\{i \mid Y_i = V\}$

Then,

$$\mathbf{E}[\tau | X_0 = S] \leq \mathbf{E}\left[\tau' | Y_0 = S\right] \leq \frac{rn^3}{r-1}(\phi(G) - \phi(S))$$

$\Rightarrow \mathbf{E}[\tau] \leq \frac{r}{r-1} n^4.$

Proof for $r = 1$

For undirected $G = (V, E)$ with $r = 1$,

$$\mathbf{E}[\tau] \leq \phi(V)^2 n^4 \leq n^6.$$

In this case $\mathbf{E}[\phi(S_t) - \phi(S_{t-1})]$ does not change
$\Rightarrow$ Use a martingale argument

At each t, the probability that ϕ changes is $\geq 1/n^2$, and it changes by $\leq 1/n$.
Dominate by process $Z_t(\phi_t)$, which increases in expectation until stopping time, when the process absorbs.

Then $\mathbf{E}[Z_\tau] \geq \mathbf{E}[Z_0]$ and we get a bound for $\mathbf{E}[\tau]$.

Aproximating $f_G(r)$

A FPRAS for a function f: A randomized algorithm A such that, given a $0 \leq \epsilon \leq 1$, for any input x,

$$\mathbf{Pr}\left[(1-\epsilon)f(x) \leq A(x) \leq (1+\epsilon)f(x)\right] \geq \frac{3}{4},$$

with a running time $\leq \text{poly}(|x|, 1/\epsilon)$.

Corollary to absorption bounds

- There is an FPRAS for computing the fixation probability, for any fixed $r \geq 1$.
- There is an FPRAS for computing the extinction probability, for any fixed $r < 1$.

Absorption time Δ-regular graphs, $r > 1$

Díaz,Goldberg,Richerby,Serna. ArXive 2014
Recall the upper bound for absorption time undirected G is $\frac{r}{r-1}n^4$.

Theorem If $G = (V, E)$ is a connected Δ-regular graph with $|V| = n$, the upper bound to the expected absorption time is

$$\mathbf{E}[\tau] \leq \frac{r}{r-1} n^2 \Delta.$$

Sketch of proof
For any $\emptyset \subseteq S \subseteq V$, use $\phi(S) = \sum_{v \in S} \frac{1}{\deg(v)} = \frac{|S|}{\Delta}$ and $\phi(V) = \frac{n}{\Delta}$.

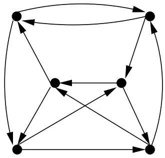
$\mathbf{E}[\phi(S_{t+1}) - \phi(S_t)] = \frac{r-1}{W_{t+1}} \frac{1}{\deg(u)\deg(v)} = \Theta(\frac{1}{\Delta^2 n})$

Δ-regular digraphs

Δ-regular digraph: $\forall v$, $\deg^-(v) = \deg^+(v) = \Delta$.

Recall for regular digraphs:

- Fixation probability is ρ, independent of the particular topology of the graph.
- As $n \to \infty$, $\rho \to 1 - \frac{1}{r}$, therefore the expected number of active steps $\to n(1 - \frac{1}{r})$, independently of the graph.



Expected absorption time for regular digraphs, $r > 1$

The expected absorption time does depend on the graph.

Theorem Let G be a strongly connected Δ-regular n-vertex digraph. Then the expected absorption time is

$$(\frac{r-1}{r^2})nH_{n-1} \leq \mathbf{E}[\tau] \leq n^2\Delta,$$

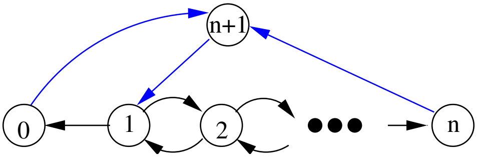
where H_n is the nth. Harmonic number.

Corollaries

- For K_n $(\Delta = n-1) \Rightarrow \mathbf{E}[\tau] = \Omega(n \log n)$ and $\mathbf{E}[\tau] = O(n^3)$.
- For $C_n \Rightarrow \mathbf{E}[\tau] = \Omega(n \log n)$ and $\mathbf{E}[\tau] = O(n^2)$.

Glimpse of proof

Dominate the process by a Markov chain:



Solve difference equation to find the expected number of active steps going from state j to state $n+1$.

Compute bound on the time you spend in each state j.

Undirected Δ-regular and isoperimetric inequality

Given an undirected graph $G = (V, E)$, the isoperimetric number (Harper, J. Comb. Theory 1966) is defined as

$$i(G) = \min_S \left\{ \frac{|\delta S|}{S} \mid S \subset V, 0 < |S| \leq |V|/2 \right\},$$

where δS is the set of edges in the cut between S and $V \backslash S$.

Proposition If G is Δ-regular undirected (*good expander*)

$$\mathbf{E}[\tau] \leq \frac{2\Delta n H_n}{i(G)}.$$

For some Δ-reg. G the isoperimetric bound improves the general theorem.

Applications of the isoperimetric result

- The K_n has $i(G) = \Theta(1/\sqrt{n}) \Rightarrow$
  $\mathbf{E}[\tau] = \Theta(n \log n)$ ($\mathbf{E}[\tau] = O(n^3)$).
- The $\sqrt{n} \times \sqrt{n}$-grid has $i(G) = \Theta(1/\sqrt{n}) \Rightarrow$
  $\mathbf{E}[\tau] = O(n^{3/2} \log n)$ ($\mathbf{E}[\tau] = O(n^2)$).
- The C_n has $i(G) = 4/n \Rightarrow$
  $\mathbf{E}[\tau] = O(n^2 \log n)$ ($\mathbf{E}[\tau] = O(n^2)$).

Bolobás, Eur. J. Comb. 1988: *For $\Delta \geq 3$ there is a number $0 < \nu < 1$ such that, as $n \to \infty$, for almost all undirected Δ-regular G, $i(G) = \nu\Delta/2$.*

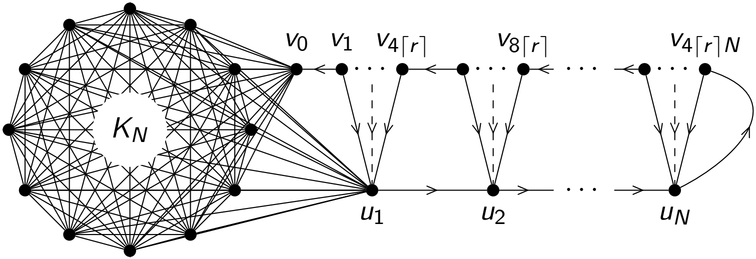
- Bollobás result $\Rightarrow$ for almost all undirected Δ-regular G,
  $\mathbf{E}[\tau] = O(n \log n)$.

Worst absorption time for directed graphs

Recall the absorption time of undirected graphs $\mathbf{E}[\tau] \leq O(n^4)$.

Theorem There is an infinite family of strongly connected digraphs such that the expected absorption time for an n vertex graph is

$$\mathbf{E}[\tau] = 2^{\Omega(n)}.$$



Domination

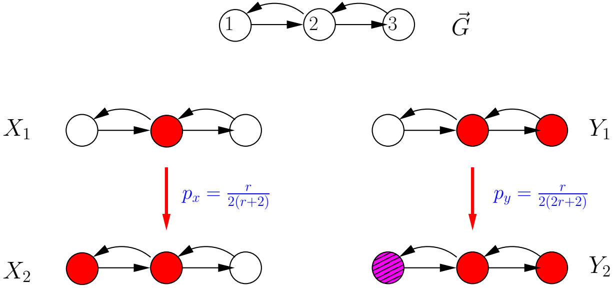
Given a Moran's process $\{X_t\}$ on G, intuition says that for any S and any $S' \subset S$, $f_S(r) > f_{S'}(r)$ and $\tau(S) < \tau(S')$.

$\therefore$ To analyze $\{X_t\}$, we can couple it with a process $\{Y_t\}$, which is easier to analyze (*for instance by allowing transitions that create new mutants but forbidding some of the transitions removing mutants*).

Then we must ensure that for every $t > 1$, if $X_1 \subseteq Y_1 \Rightarrow X_t \subseteq Y_t$.

NOT ALWAYS TRUE for discrete Moran's processes

Counterexample



Coupling $\{X_i\}$ and $\{Y_i\}$ fails as for $r > 1$,

$$\mathbf{Pr}\left[X_2 \not\subseteq Y_2\right] > 0$$

Continuous time process

To use domination for the discrete processes $\{X_i\}$ and $\{Y_i\}$, consider the continuous versions $\tilde{X}[t]$ and $\tilde{Y}[t]$, where vertex v with fitness $r_v \in \{1, r\}$ waits an amount of time which follows an exponential distribution with parameter r_v.

The discrete Moran process is recovered by taking the sequence of configurations each time a vertex reproduces.

Notice: in continuous time, each v reproduces at a rate given by r_v, independently of the other vertices, while in discrete time the population "coordinates" before deciding who is next to reproduce.

Coupling Lemma and consequences

Coupling Lemma For $\vec{G} = (V, \vec{E})$, let $X \subseteq Y$ and $1 \leq r \leq r'$. Let $\tilde{X}[t]$ and $\tilde{Y}[t]$ ($t \geq 0$) be the continuous-time Moran process on G with mutant fitness r and r', and with $\tilde{X}[0] = X$ and $\tilde{Y}[0] = Y$. There is a coupling between the two processes s. t. $\tilde{X}[t] \subseteq \tilde{Y}[t]$, $\forall t \geq 0$.

Theorem For any $\vec{G}$, if $0 < r \leq r'$ and $S \subseteq S'$ then

$$f_{\vec{G},r}(S) \leq f_{\vec{G},r'}(S').$$

Corollary *(Monotonicity)*
For any $\vec{G}$ and $0 < r \leq r'$ then, $f_{\vec{G}}(r) \leq f_{\vec{G}}(r')$.

Corollary *(Subset domination)*
For any $\vec{G}$ and $0 < r$ then, if $S \subseteq S'$ then $f_{\vec{G},r}(S) \leq f_{\vec{G},r'}(S')$.

Thank you for your attention