Evolutionary Dynamics on Graphs

Leslie Ann Goldberg, University of Oxford

Absorption Time of the Moran Process (2014)

with Josep Díaz, David Richerby and Maria Serna

Approximating Fixation Probabilities in The Generalized Moran Process (2012)

On the fixation probability of superstars (2013)

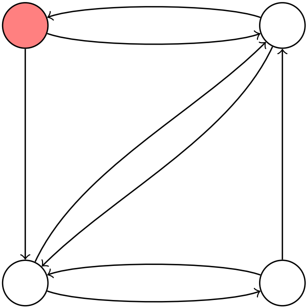
also with George Mertzios and Paul Spirakis

Stochastic Graph Models, Brown University, March 2014

Evolutionary Dynamics on Graphs

Lieberman, Hauert, Nowak; Nature 2005

Each vertex represents an individual



Moran Process

- Pick node i with probability fitness$(i)/W$
- Reproduce to random out-neighbour of i

Strongly connected digraph. Mutant fitness $r > 0$.

Non-mutant fitness 1.

Total fitness $W = r + 3$.

- Initial Configuration:
  - One mutant, chosen u.a.r.
- Final Configurations (given strong connectivity):
  - Extinction: No mutants
  - Fixation: All mutants

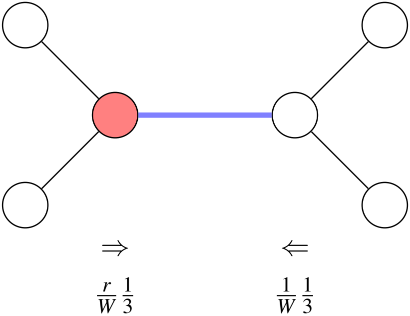
Questions

1. What is the fixation probability of a graph? (exactly, bounds)
2. What is the expected absorption time?
3. Computational Problem: Given a graph, compute its fixation probability?

- Fixation probability and expected absorption time depend on the graph topology and the mutant fitness.
- Can be computed by solving a system of 2^n linear equations!

Some fixation probabilities: regular undirected graph

Random walk on a line (LHN 2005)



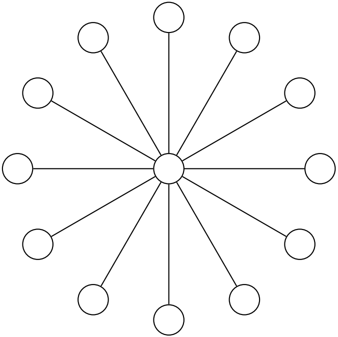
n vertices

$$f_{G,r} = \frac{1-\frac{1}{r}}{1-\frac{1}{r^n}}$$

- $r > 1$: $\lim_{n\to\infty} f_{G,r} = 1 - 1/r$.
- $r < 1$: $f_{G,r}$ is exponentiallly small in n. For $R = 1/r > 1$,

$$f_{G,r} = \frac{R-1}{R^n-1}.$$

The fixation probability of the star



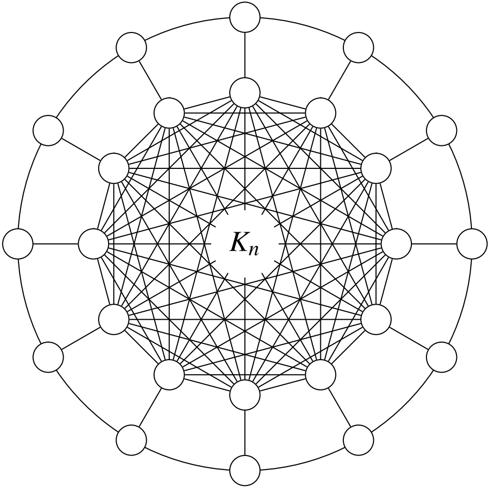
Centre: quickly killed.
Leaf: good.

Approximation of $f_{G,r}$ (for large n)

$$f_{G,r} = \frac{1 - \frac{1}{r^2}}{1 - \frac{1}{r^{2n}}} \sim 1 - \frac{1}{r^2} > 1 - \frac{1}{r}, \quad \text{for } r > 1$$

(LHN 2005; Exact analysis: Broom, Rychtár Proc. Royal Soc. A 2008) highest possible fixation prob?

"Suppressor" (fixation probabilities lower than $1 - 1/r$)



For $1 < r < 4/3$, $\lim_{n\to\infty} f_{G,r} \leq \frac{1}{2}(1 - \frac{1}{r}) + o(1)$.

Mertzios, Nikoletseas, Raptopoulos, Spirakis, TCS 2013

Absorption time

Theorem When $r > 1$ and the initial single mutant is chosen uniformly at random, the absorption time of the Moran process on an n-vertex undirected graph G satisfies

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

Dominate the absorption time: a process that gets a new mutant (u.a.r.) if it ever goes extinct.

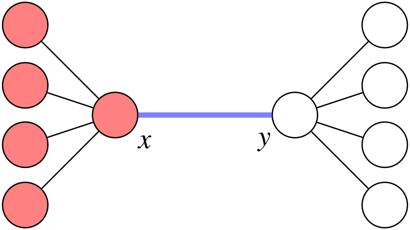
- Potential function for set S of mutants.

$$\phi(S) = \sum_{x \in S} \frac{1}{\deg x}$$

- $\phi(\{v\}) = \frac{1}{\deg v} \geq \frac{1}{n}$
- $\phi(V) \leq n$
- $$\mathbb{E}[\phi(S_t) - \phi(S_{t-1})] \geq \left(1 - \frac{1}{r}\right)\frac{1}{n^3}.$$

A bad configuration S_{t-1}

$$\mathbb{E}[\phi(S_t) - \phi(S_{t-1})] \geq \left(1 - \frac{1}{r}\right)\frac{1}{n^3}.$$



$$\mathbb{E}[\phi(S_t) - \phi(S_{t-1})] =$$

$$\frac{r-1}{W}\,\frac{1}{\deg(x)\deg(y)}$$

Actual absorption time $O(n^3)$: $O(n^3)$ to get LHS half full, then every $\Theta(n^2)$ steps x fires a mutant into y. This has to happen $\Theta(n)$ times before spreads to RHS leaf. Then $O(n^3)$ to fill RHS. Probably $O(n^3)$ for all undirected graphs.

Theorem When $r < 1$ and the initial single mutant is chosen uniformly at random, the absorption time of the Moran process on an n-vertex undirected graph G satisfies

$$\mathbb{E}[\tau] \leq \tfrac{1}{1-r} n^3.$$

Proof:

$$\mathbb{E}[\phi(X_{i+1}) - \phi(X_i) \mid X_i = S] < -\left(\frac{1-r}{n^3}\right).$$

Now the process quickly goes exctinct.

Theorem When $r = 1$ and the initial single mutant is chosen uniformly at random, the absorption time of the Moran process on an n-vertex undirected graph G satisfies

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

Martingale argument. At each step, probability that the potential moves is at least n^{-2}. If the potential moves, it changes by at least n^{-1}. Study a process Z_t, depending on t and ϕ_t, which increases in expectation until a stopping time when the process absorbs. $E[Z_\tau] \geq E[Z_0]$, so we get bound on $E[\tau]$.

Computational Problem: Given a graph, compute its fixation probability.

FPRAS for a function f: A randomized algorithm g such that, for any input X and any $\varepsilon \in (0,1)$,

$$\Pr\Big((1-\varepsilon)f(X) \leq g(X) \leq (1+\varepsilon)f(X)\Big) \geq \frac{3}{4}.$$

The running time of g is at most poly$(|X|, \varepsilon^{-1})$.

Corollary of absorption time bounds

- For fixed $r \geq 1$ there is an FPRAS for approximating the fixation probability.
- For fixed $r < 1$ there is an FPRAS for approximating the extinction probability.

Ingredients: Tail bounds on absorption times via Markov's inequality, upper and lower bounds on fixation probability.

- For $r < 1$ we FPRAS extinction probability because we don't have a positive polynomial lower bound on the fixation probability

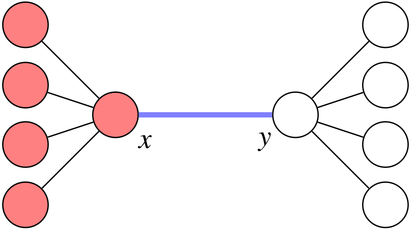
regular undirected graphs, $r > 1$

Theorem. The expected absorption time on a connected Δ-regular n-vertex undirected graph is at most

$$\tfrac{r}{r-1} n^2 \Delta.$$

$\phi(S) = \sum_{x \in S} \frac{1}{\deg x}$

$\phi(V) = n/\Delta$



$$\mathbb{E}[\phi(S_t) - \phi(S_{t-1})] = \frac{r-1}{W} \frac{1}{\deg(x)\deg(y)} = \Theta\left(\frac{1}{\Delta^2 n}\right)$$

Regular digraphs. $r > 1$.

indegree $=$ outdegree $= \Delta$

- The fixation probability does not depend on the graph. The probability that the next reproduction happens along (u, v) is $\frac{r}{W}\frac{1}{\Delta}$ if u is a mutant and $\frac{1}{W}\frac{1}{\Delta}$ if u is not. There are exactly as many edges from mutants to non-mutants as from non-mutants to mutants.
- The expected number of "active steps" tends to $n(1 + \frac{1}{r})$ as $n \to \infty$. This does not depend on the graph (assuming regularity)
- The expected absorption time does depend on the graph.

Theorem. The expected absorption time of the Moran process on a strongly connected Δ-regular n-vertex digraph G satisfies

$$\left(\frac{r-1}{r^2}\right) n\, H_{n-1} \leq \mathbb{E}[\tau] \leq n^2 \Delta.$$

H_n is the n'th Harmonic number $\sum_{j=1}^{n} \frac{1}{j} \sim \ln n$.

The idea

- Consider a Markov chain with state space $\{0, \ldots, n+1\}$ which starts at one (one mutant), has a rightward drift (corresponding to r), goes (deterministically) to state $n+1$ from states 0 and n (absorption) and from there to state 1 (repeating the process). Let γ_j^k be the number of visits to state j between visits to state k.
- Solve recurrences to find $\mathbb{E}[\gamma_j^{n+1}]$, which is the expected number of active steps when the Moran process has j mutants. This does not depend on the digraph. For every j it is between $1 - 1/r^2$ and $1 + 1/r$.
- For the given digraph, find bounds on the expected amount of time that the process hovers at the (best/worst) j-mutant state.

- Use Wald's equality to calculate the total amount of time spent with j mutants. The random variable giving the time that you sit there is the same each time (or at least the bound is the same — use domination) and the number of times that you go there is independent of that. $E[X_1 + \cdots + X_N] = E[N]E[X_1]$.

Consequences

- Undirected clique. $\Theta(n \log n)$ (upper and lower bounds)
- Undirected or directed cycle. $\Theta(n^2)$ (upper and lower bounds)

A connected Δ-regular undirected graph

The isoperimetric number of G is a discrete analog of the Cheeger isoperimetric constant defined by Buser 1978.

$$i(G) = \min \left\{ \frac{|\partial S|}{|S|} \,\middle|\, S \subseteq V(G), 0 < |S| \leq \frac{|V(G)|}{2} \right\},$$

where ∂S is the set of edges between vertices in S and vertices in $V(G) \setminus S$.

Corollary.

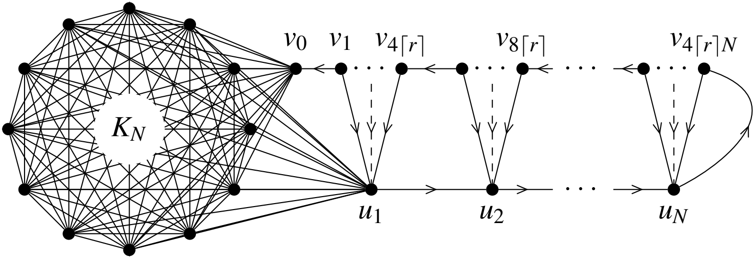
$$\mathbb{E}[\tau] \leq 2\Delta\, n\, H_n / i(G).$$

Consequences of $\mathbb{E}[\tau] \leq 2\Delta\, n\, H_n / i(G)$

- $\sqrt{n}$ by $\sqrt{n}$ grid: $\mathbb{E}[\tau] = O(n^{3/2} \log n)$
  $i(G) = \Theta(1/\sqrt{n})$
- hypercube $\mathbb{E}[\tau] = O(n \log^2 n)$
  $i(G) = 1$
- For $\Delta \geq 3$, almost all Δ-regular n-vertex undirected graphs G (as n tends to infinity) have $O(n \log n)$ expected absorption time.
  Bollobas: There is a positive number $\eta < 1$ such that , for almost all Δ-regular n-vertex undirected graphs G (as n tends to infinity), $i(G) \geq (1 - \eta)\Delta/2$.

Exponentially large absorption time

undirected graphs have $E[\tau] = O(n^4)$ but this is not true for digraphs



WHP, start in the clique, which gets at least half-full. It doesn't go extinct quickly (random walk against drift) and it doesn't fix quickly (against drift along the path).

Stochastic Domination

Conjecture. (Shakarian, Roos, Johnson, Biosystems 2012)
Fixation probability is monotonic in r.

Intuitions

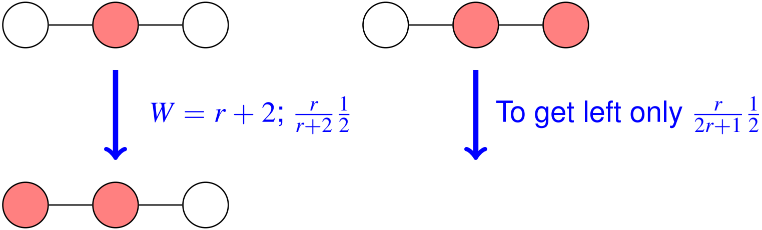
- "The Moran process has a higher probability of reaching fixation from S than from some subset of S and it will do so in fewer steps."
- "Modifying the process by allowing all transitions that create new mutants but forbidding some transitions that remove mutants should make fixation faster and more probable."

Domination

Goal. Couple the Moran process $(Y_t)_{t\geq 1}$ with another copy $(Y'_t)_{t\geq 1}$ of the process where $Y_1 \subseteq Y'_1$. The coupling would be designed so that $Y_1 \subseteq Y'_1$ would ensure that $Y_t \subseteq Y'_t$ for all $t > 1$.

The snag

$Y_1 = \{2\}$ $\quad$ $Y_1' = \{2, 3\}$



There is no coupling with $Y_2 \subseteq Y_2'$.

When vertex 3 becomes a mutant it becomes more likely to reproduce so it “slows down” all of the other mutants in the graph.

Continuous time process

- Vertex v has fitness $r_v \in \{1, r\}$. Waiting time is exponential with parameter r_v (independently of other vertices).
- Probability density function

$$f(t) = \begin{cases} r_v e^{-r_v t}, & \text{if } t \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

- Cumulative distribution function (probability of being $\leq t$)

$$F(t) = \begin{cases} 1 - e^{-r_v t}, & \text{if } t \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

- The waiting time until the first vertex fires is exponential with parameter $W = \sum_v r_v$
- The probability that v fires first is r_v / W

The Moran process is recovered by taking the sequence of configurations each time a vertex fires

Coupling Lemma

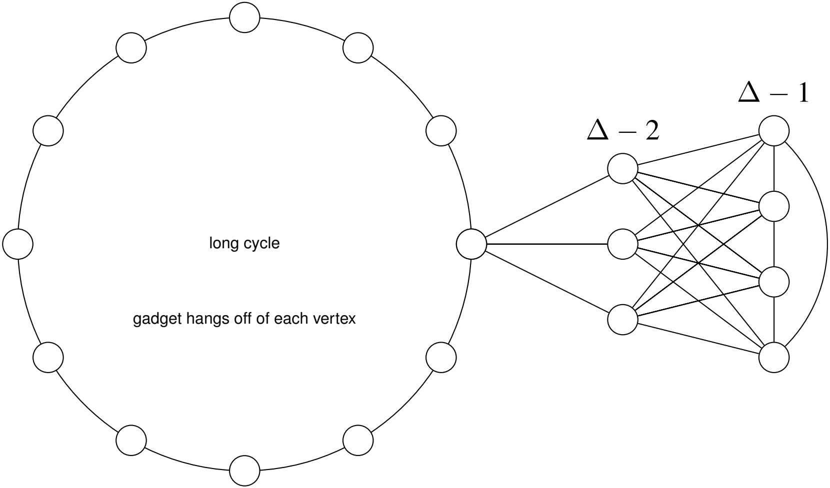
Let $G = (V, E)$ be any digraph, let $Y \subseteq Y' \subseteq V(G)$ and $1 \leq r \leq r'$. Let $\widetilde{Y}[t]$ and $\widetilde{Y}'[t]$ ($t \geq 0$) be continuous-time Moran processes on G with mutant fitness r and r', respectively, and with $\widetilde{Y}[0] = Y$ and $\widetilde{Y}'[0] = Y'$. There is a coupling between the two processes such that $\widetilde{Y}[t] \subseteq \widetilde{Y}'[t]$ for all $t \geq 0$.

Consequence: If $0 < r \leq r'$ and $S \subseteq S' \subseteq V$, then $f_{G,r}(S) \leq f_{G,r'}(S')$.

Proves conjecture and also gives subset domination: “adding more mutants can’t decrease the fixation probability”

Using the domination

- Recall that the expected absorption time of the Moran process on a strongly connected Δ-regular n-vertex digraph G is at most $n^2\Delta$.
- For each $\Delta > 2$ we construct an infinite family of connected Δ-regular undirected graphs for which the expected absorption time is $\Omega(n^2)$.



- Prob(quickly extinct) $\leq$ Prob(extinct) $\leq \frac{1}{r}$ (regular)
- Prob(quickly fix) is small.
  - Domination: start with a cycle-vertex on (and possibly a vertex in its gadget)
  - Domination: don't turn cycle vertices off
  - We know how the cycle behaves!

Fixation probabilities: A lower bound for a strongly connected n-vertex digraph for $r \geq 1$

$$f_{G,r} \geq f_{G,1} = \frac{1}{n} \sum_{v \in V} f_{G,1}(x) = \frac{1}{n}$$

(The sum adds up to 1: Consider n different kinds of mutants — some will take over.)

(Recall that if $r < 1$ the clique has exponentially small fixation probability, so there is no such polynomial lower bound)

An upper bound for a connected n-vertex undirectred graph for $r > 0$

$$f_{G,r} \leq 1 - \frac{1}{n+r}$$

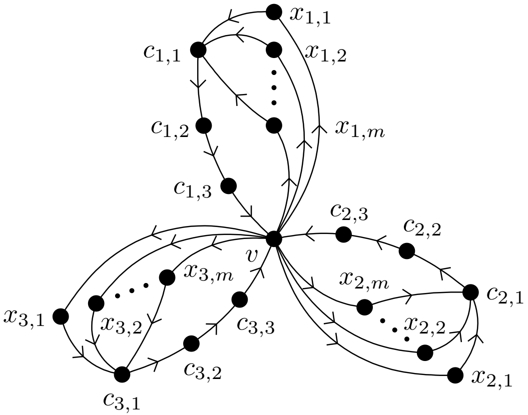
(This is an upper bound on the probability that the first active step creates a second mutant.)

Mertzios, Spirakis 2014 For any $\varepsilon > 0$,

$$f_{G,r} \leq 1 - \frac{1}{n^{3/4+\varepsilon}}$$

There are no known upper bounds that don't depend on n even though we think the true upper bound is something like $1 - 1/r$ (bounded below 1).

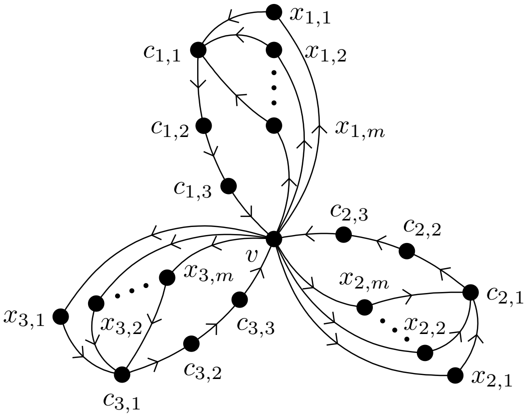
“Amplifiers”



LHM 2005 Superstar $S^k_{\ell,m}$. k is the “amplification factor” (chain length $k-2$, here 3), ℓ leaves (here 3), reservoir size m.

Claim: $$\lim_{\ell,m\to\infty} f_{S^k_{\ell,m},r} = \frac{1-r^{-k}}{1-r^{-kn}}.$$

$k = 5$



Claim: $\lim_{\ell\to\infty} f_{S^k_{\ell,m(\ell)},r} = 1 - r^{-k}$

$$\lim_{\ell\to\infty} f_{S^5_{\ell,m(\ell)},r} \leq 1 - \frac{r+1}{2r^5 + r + 1} = 1 - \frac{1}{\Theta(r^4)} < 1 - r^{-5}.$$

“Fixation probabilities on superstars, revisited and revised”

Jamieson-Lane and Hauert, 22 Dec 2013

New claim: Let $N \sim \ell m(\ell)$ be the number of vertices.
Taking $k = (N)^{1/6} + 3$

$$\lim_{\ell \to \infty} f_{S^k_{\ell,m(\ell)},r} \geq \frac{1}{1 + \frac{1}{(r^2 - r)^{12} N}} = 1 - \frac{1}{(r^2 - r)^{12} N} + O(1/N^2)$$

No rigorous proof