

Population Genetics Models

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

Mutation: the process by which one individual (gene) changes. 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



Moran process

- Start with a finite set of n individuals, all non-mutant. Select one to mutate.
- iterate until all are mutant or all are no mutants
 - 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 main question: How does population structure affect the outcome of the evolutionary process

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.

For $v \in V$, let $\mathcal{N}(v) \subseteq V$ be the set of neighbors of v in G. $v \in V$ extend (being mutant or non-mutant) by cloning a $u \in \mathcal{N}(v)$.



Moran process on Evolutionary Graphs

- ▶ Given a graph G = (V, E), with |V| = n, and an r > 0, where mutants have fitness r and non-mutants have fitness 1, we start with all vertices non-mutant.
- ▶ The graph also could be directed $\vec{G} = (V, \vec{E})$. In this case we require \vec{G} to be strongly connected.
- Starting for a random mutation on a vertex, the Moran process is a randomized algorithm on G that study the spread of the mutation on V.
- The process is Markovian, depending on r it tends to one of the two absorbing states: extinction, all mutants disappear; fixation all vertices become mutants.

A particular Moran process

- 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

Example of Moran process



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

$$q_{ab} = \frac{1}{2+2r} \cdot \left(\frac{1}{3} + \frac{1}{2}\right) \\ = \frac{1}{2+2r} \cdot \frac{5}{6}$$

$$q_{bc} = \frac{1}{2+2r} \cdot \frac{5}{6}$$
$$q_b = \frac{2}{3+r} \cdot \frac{1}{3}$$

・ロト ・聞ト ・ヨト ・ヨト æ

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 $\phi_G(r)$ of G is the probability that a single mutant will takes over the whole G. The extinction probability of G is $\eta_G = 1 - \phi_G(r)$.

The Markov chain of configurations A configuration is a set $S \subseteq V$ of mutants.



◆□ > ◆□ > ◆豆 > ◆豆 > ̄豆 = のへ⊙

Questions to study

Given a connected graph G = (V, E) (\vec{G} strongly connected), and a fitness r > 0:

1.- Is it possible to compute exactly the fixation probability $\phi_G(r)$? Is it possible to give tight bounds to $\phi_G(r)$?

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 bound or compute the expected number of steps until arriving to absorption?

Consider the computation of the fixation probability:

- Quantitative question: \(\phi_G(r)\)?
 The quantitative question is NP-complete.
- Qualitative question: Given an € > 0, approximate φ_G(r) within €?
 The qualitative question is #P-complete.

Ibsen-Jensen, Chatterjee, Nowak: *Computational complexity of ecological and evolutionary spatial dynamics.* PNAS, 2015

Properties of $f_G(r)$

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

$$\phi_G(r) = \sum_{v \in S} \phi_{G,r}(\{v\}).$$

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

The case r = 1 is denoted neutral drift. For any r > 1, $\phi_G(r) > f_G(1)$

Isothermal graphs (LHN-05)

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 $w_{ij} = 1/\text{deg}^+(i)$ if $(\vec{i}, \vec{j}) \in \vec{E}$ and $w_{ij} = 0$ otherwise. The same definition of W applies to undirected G, with $w_{ij} = 1/\text{deg}(i)$.

The temperature of $i \in V$ is $T_i = \sum_{j \in V} w_{ji}$

A graph \vec{G} is isothermal if $\forall i, j \in V$, $T_i = T_j$.

Examples



$$\begin{array}{cccccc} i & j & k & l \\ i & 0 & 1 & 0 & 0 \\ j & 1/3 & 0 & 1/3 & 1/3 \\ k & 0 & 1/2 & 0 & 1/2 \\ l & 1/2 & 1/2 & 0 & 0 \end{array}$$

$$T_j = 2$$
 and $T_k = 1/3$

$$\begin{array}{cccccc} i & j & k & l \\ i & 0 & 1/3 & 1/3 & 1/3 \\ j & 1/3 & 0 & 1/3 & 1/3 \\ k & 1/3 & 1/3 & 0 & 1/3 \\ l & 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{array}$$

 $T_j = 1$ and $T_k = 1$

◆□ ▶ < 圖 ▶ < 圖 ▶ < 圖 ▶ < 圖 • 의 Q @</p>

Computing the fixation probability: Easy examples



・ロト ・ 理 ト ・ ヨ ト ・ ヨ ト

3

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

Moran's fixation Theorem

(LHN-05)

1. For a strongly connected digraph \vec{G} s.t. $\forall i, j \in V$ we have $T_i = T_j$ then

$$\phi_{\vec{G}}(r) = rac{1 - 1/r}{1 - 1/r^n}$$

2. If G is undirected, connected and Δ -regular then

$$\phi_G(r)=\frac{1-1/r}{1-1/r^n}.$$

Denote $\rho = \frac{1-1/r}{1-1/r^n}$ the Moran fixation probability.

Consequences: Moran's fixation Theorem

Notice: for large values of n:

- if r < 1 then $\phi_G(r) = \frac{r^n r^{n-1}}{r^n 1} \rightarrow$ exponentially small.
- if r = 1 then $\phi_G(r) = \frac{1}{n}$
- if r > 1 then $\phi_G(r) \rightarrow 1 \frac{1}{r}$.

Examples of G and \vec{G} with $\phi = \rho$

 K_n, Δ_n

Cycle C_n



- 日本 - 4 日本 - 4 日本 - 日本

Amplifiers and suppressors

Given G (directed or undirected) and r, G is said to be:

- an amplifier if $\phi_G(r) > \rho$.
- a suppressor if $\phi_G(r) < \rho$.

Examples for r > 1 (LHN-05),

The star is an amplifier $\phi_G(r) = \frac{1 - \frac{1}{r^2}}{1 - \frac{1}{r^{2n}}} > \rho$

The superstar star is an amplifier $\phi_G(r) = 1 - \frac{1}{kr^4} > \rho$





・ロト ・ 理 ト ・ ヨ ト ・ ヨ ト ・ ヨ

Suppressors

For any r > 1, any directed \vec{G} is "one-rooted" iff $\phi_{\vec{G}} = \frac{1}{n}$. The directed line and the burst have fixation probability $\frac{1}{n} < \rho$,



Suppressors for undirected G

Mertzios, Nikoletseas, Ratopoulos, Spirakis, TCS 2013 **The urchin:** A perfect matching between C_n and K_n . For 1 < r < 4/3, then they proved that $\phi_G(r) \rightarrow \frac{1}{2}(1 - \frac{1}{r}) < \rho$



Amplifiers for the Moran process

Strong amplifiers: graphs for which extinction probability $\rightarrow 0$ as *n* grows.

LHN asked for the existence of strong amplifiers.

Galanis, Goebel,Goldberg,Lapinskas,Richerby, JACM, 2017 answered in the positive:

There exists an infinite family of digraphs, megastar which is strongly amplifier.



The papers formalizes the mathematical techniques for formally analyzing the fixation and extinction probabilities on the Moran process, for any given family of graphs.

General bounds for $\phi_G(r)$

Let G = (V, E) be any undirected connected graph, with |V| = n. D.,Goldberg,Mertzios,Richerby,Serna,Spirakis, SODA-2012

(DGMRSS) For any $r \ge 1$, $\frac{1}{n} \le \phi_G(r) \le 1 - \frac{1}{n+r}$, are bounds on the fixation probability for *G*. Moreover, for r = 1, $\phi_G(1) = \frac{1}{n}$.

Merzios, Spirakis: ICALP-2013 For any $\epsilon > 0$,

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

Absorption time for undirected graphs

Given undirected connected G = (V, E), with |V| = n, run a Moran process $\{S_t\}_{t \ge 0}$, where $\{S_t\}$ set of mutants at time t, i.e. $\phi_G(r) = \sum_{v \in V} \phi_{G,r}(\{v\}).$

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

Whp the Moran's process reaches absorption in a polynomial number of steps.

Theorem DGMRSS

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

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

Sketch of the proof

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

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

For each case r > 1, r = 1, r < 1 at every step

1. compute the evolution of $\mathbf{E} [\phi(S_{t+1}) - \phi(S_t)]$,

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

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

► For
$$r = 1$$
, **E** $[\phi(S_{t+1}) - \phi(S_t)] = 0$

 show that the potential decreases (increases) monotonically for r < 1 (r > 1).

Approximating $\phi_G(r)$

A FPRAS (Fully Polynomial Time Approximation Scheme) for a function f is a randomized algorithm A such that, given a $0 \le \epsilon \le 1$, for any input x,

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

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 ≥ 1.
- ► There is an FPRAS for computing the extinction probability, for any fixed r < 1.</p>

Worst absorption time for directed graphs

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

D.,Goldberg,Richerby,Serna. RSA 2016 (DGRS)

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

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



The techniques used to approximate fixation in undirected graphs, do not work for directed.

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

(DGRS)

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 *n*th. Harmonic number.

Corollaries

• For \mathcal{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 $\mathbf{E}[\tau] = O(n^2)$.

◆□▶ ◆□▶ ◆ ≧▶ ◆ ≧▶ ─ 差 − のへぐ

Undirected Δ -regular and isoperimetric inequality

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

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

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

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

$$\mathbf{E}\left[\tau\right] \leq \frac{2\Delta nH_n}{i(G)}.$$

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 $\Delta \ge 3$ 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, $\mathbf{E}[\tau] = O(n \log n).$

Main tool in proofs: Domination

Given a Moran's process $\{X_t\}$ on G, intuition says that for any S and any $S' \subset S$, $\phi_S(r) > \phi_{S'}(r)$ 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



Notice that the only possible value for Y_2 to contain X_2 is $\{1, 2, 3\}$, which happens with probability $\frac{r}{2(2r+1)}$ Coupling $\{X_i\}$ and $\{Y_i\}$ fails as for r > 1,

$$\Pr[X_2 \not\subseteq Y_2] \ge \frac{r(r-1)}{2(r+2)(2r+1)} > 0$$

Continuous time process

To use domination for the discrete processes $\{X_i\}$ and $\{Y_i\}$, consider the continuous versions $\tilde{X}[t]$ and $\tilde{Y}[t]$, with exponential clocks,

i.e. any 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 $0 < r \le r'$ and $S \subseteq S'$ then

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

Corollary (*Fixation probability is monotone wrt mutant fitness*) For any \vec{G} and $0 < r \le r'$ then, $f_{\vec{G}}(r) \le f_{\vec{G}}(r')$.

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

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

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.

・ロト ・四ト ・ヨト ・ヨ

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

Moran process and contact process

Contact process: Stochastic infection process played mostly on infinite graphs: SIS model, voter model, and others. In the SIS model, each node can be infected with a probability proportional to the number of infected neighbors, and can heal with a Poisson probability with rate 1. The goal is to study the conditions for epidemic and extinction.

R. Durrett: Some features of the spread of epidemics and information on random graphs PNAS 2010.

- Moran process similar to the discrete finite version of SIS model.
- Voter system is Moran process with r = 1.
- Problems and techniques used for contact model are very different from Moran process. For ex. infection and healing in SIS is very different than from Moran.

Conclusions

Expected absorption time $\mathsf{E}\left[\tau\right]$

G	lower bd.	upper bd.	
Undirected $r < 1$		$O(n^3)$	DGMRSS
Undirected $r = 1$		$O(n^6)$	DGMRSS
Undirected $r > 1$		$O(n^4)$	DGMRSS
Directed $r > 1$		$2^{\Omega(n)}$	DGRS
Dir. Δ-reg.	$\frac{r-1}{r^2}nH_{n-1}$	$n^2\Delta$	DGRS
k _n		$\Theta(n \lg n)$	DGRS
$\sqrt{n} \times \sqrt{n}$ -grid	$\Omega(n \lg n)$	$O(n^{3/2} \lg n)$	DGRS
hypercube	$\Omega(n \lg n)$	$O(n \lg^2 n)$	DGRS

Conclusions

- For the basic LHN model, quite a bit of work has been done. There are stil some open questions as: What happens when we have a hierarchy of different mutants with different fitness?
- Martin Nowak has been proposing different variation of his original model for different epidemiology situations. In particular his I& R graph evolutionary model for ecological problems is a nice generalization of the basic model. Ibsen, Chatterjee, Nowak, PNAS, 2015

Thank you for your attention

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ