The Computational Complexity of Genetic Diversity

Ioannis Panageas
MIT - SUTD

Previously Georgia Institute of Technology

joint work with **Ruta Mehta**, **Georgios Piliouras** and **Sadra Yazdanbod**

The puzzle of sex: No easy answers

- What is the purpose of sex: A key problem in biology
- Several answers to the problem
- Key high level idea:
  Sex $=$ Gene Mixing $\rightarrow$ Genetic Diversity $=$ Good?

The puzzle of sex: No easy answers

- What is the purpose of sex: A key problem in biology
- Several answers to the problem
- Key high level idea:
  Sex $=$ Gene Mixing $\rightarrow$ Genetic Diversity $=$ Good?
- **This talk:** Focus on diploid species, where evolution is described by *replicator dynamics*.

The puzzle of sex: No easy answers

- What is the purpose of sex: A key problem in biology
- Several answers to the problem
- Key high level idea:
  Sex $=$ Gene Mixing $\rightarrow$ Genetic Diversity $=$ Good?
- **This talk:** Focus on diploid species, where evolution is described by *replicator dynamics*.

Computational Question

Does diversity persist in the limit? Yes/No?

Outline

Basic Terms in Biology

- Gene: A unit that determines some characteristic of the organism, and passes traits to offsprings. Controls the expression of traits, e.g., eye color, blood type.

Basic Terms in Biology

- Gene: A unit that determines some characteristic of the organism, and passes traits to offsprings. Controls the expression of traits, e.g., eye color, blood type.
- Locus: The specific location of a gene on a chromosome (plural loci).

Basic Terms in Biology

- Gene: A unit that determines some characteristic of the organism, and passes traits to offsprings. Controls the expression of traits, e.g., eye color, blood type.
- Locus: The specific location of a gene on a chromosome (plural loci).
- Allele: One of a number of alternative forms of the same gene, found at the same loci. Different alleles can result in different observable traits, such as different eye color, blood type, e.t.c.

Basic Terms in Biology

- Gene: A unit that determines some characteristic of the organism, and passes traits to offsprings. Controls the expression of traits, e.g., eye color, blood type.
- Locus: The specific location of a gene on a chromosome (plural loci).
- Allele: One of a number of alternative forms of the same gene, found at the same loci. Different alleles can result in different observable traits, such as different eye color, blood type, e.t.c.
- Genotype: The genetic constitution of an individual organism.

Basic Terms in Biology

- Gene: A unit that determines some characteristic of the organism, and passes traits to offsprings. Controls the expression of traits, e.g., eye color, blood type.
- Locus: The specific location of a gene on a chromosome (plural loci).
- Allele: One of a number of alternative forms of the same gene, found at the same loci. Different alleles can result in different observable traits, such as different eye color, blood type, e.t.c.
- Genotype: The genetic constitution of an individual organism.
- *Diploid:* Having two copies of each chromosome.

Basic Terms in Biology

- Gene: A unit that determines some characteristic of the organism, and passes traits to offsprings. Controls the expression of traits, e.g., eye color, blood type.
- Locus: The specific location of a gene on a chromosome (plural loci).
- Allele: One of a number of alternative forms of the same gene, found at the same loci. Different alleles can result in different observable traits, such as different eye color, blood type, e.t.c.
- Genotype: The genetic constitution of an individual organism.
- *Diploid:* Having two copies of each chromosome.
- Haploid: Having one copy of each chromosome.

Basic Terms in Biology

- Gene: A unit that determines some characteristic of the organism, and passes traits to offsprings. Controls the expression of traits, e.g., eye color, blood type.
- Locus: The specific location of a gene on a chromosome (plural loci).
- Allele: One of a number of alternative forms of the same gene, found at the same loci. Different alleles can result in different observable traits, such as different eye color, blood type, e.t.c.
- Genotype: The genetic constitution of an individual organism.
- *Diploid:* Having two copies of each chromosome.
- Haploid: Having one copy of each chromosome.
- Panmictic: Every pair of individuals can produce offspring (no male, female distinction).

Setting: Panmictic, Diploid 1-Gene with Sex

The species has one gene: X.

Setting: Panmictic, Diploid 1-Gene with Sex

The species has one gene: X.

Gene X has n possible alleles: $X_1, X_2, \ldots, X_n$.

Setting: Panmictic, Diploid 1-Gene with Sex

The species has one gene: X.
Gene X has n possible alleles: $X_1, X_2, \ldots, X_n$.
Each individual has an unordered genotype of the form $X_i X_j$.

Setting: Panmictic, Diploid 1-Gene with Sex

The species has one gene: X.
Gene X has n possible alleles: $X_1, X_2, \ldots, X_n$.
Each individual has an unordered genotype of the form $X_i X_j$.
Let W_{ij} denote the fitness of genotype $X_i X_j$.
Let x_i the frequencies of allele X_i respectively.

Setting: Panmictic, Diploid 1-Gene with Sex

The species has one gene: X.
Gene X has n possible alleles: $X_1, X_2, \ldots, X_n$.
Each individual has an unordered genotype of the form $X_i X_j$.
Let W_{ij} denote the fitness of genotype $X_i X_j$.
Let x_i the frequencies of allele X_i respectively.

When two individuals mate: e.g., $X_i X_j$ and $X_{i'} X_{j'}$ the possible offspring combinations are $X_i X_j, X_i X_{j'}, X_{i'} X_j, X_{i'} X_{j'}$ and the number of offsprings is proportional to the corresponding entries of fitness (symmetric, positive) matrix W.

Setting: Panmictic, Diploid 1-Gene with Sex

The species has one gene: X.
Gene X has n possible alleles: $X_1, X_2, \ldots, X_n$.
Each individual has an unordered genotype of the form $X_i X_j$.
Let W_{ij} denote the fitness of genotype $X_i X_j$.
Let x_i the frequencies of allele X_i respectively.

When two individuals mate: e.g., $X_i X_j$ and $X_{i'} X_{j'}$ the possible offspring combinations are $X_i X_j, X_i X_{j'}, X_{i'} X_j, X_{i'} X_{j'}$ and the number of offsprings is proportional to the corresponding entries of fitness (symmetric, positive) matrix W.

If two individuals are picked at random and mate, the expected frequencies (of next generation) reduce to (called *replicator dynamics*):

$$x_i' = x_i \frac{(Wx)_i}{x^T W x} \qquad \textit{where } \ (Wx)_i = \sum_j W_{ij} x_j.$$

x is a vector that lies in the simplex of size n, denoted by Δ_n.

Related work

Dynamics on Haploids

- *Chastain, Livnat, Papadimitriou, Vazirani '14:* Strong connection between Game Theory, MWUA and haploid evolution.
- *Mehta, P, Piliouras '15:* In the limit, under mild conditions on fitness matrix, haploid dynamics converges to monomorphic populations (no diversity).

Complexity

- *Nissan '06* and then *Etessami, Lochbihler '08* studied the question of whether a game has an evolutionary stable strategy (ESS).
- *Conitzer '13* showed this to be Σ_2^P-complete.
- See also *Ibsen-Jensen, Chatterjee and Nowak '15*.

Decision Problem

- Input: A $n \times n$ positive symmetric matrix W.

Decision Problem

- Input: A $n \times n$ positive symmetric matrix W.
- Question: Assume that we start at random from Δ_n (space of possible allele frequencies). Is the probability positive that the limit point of replicator dynamics will have support greater than one (aka diverse)?

$$\textit{Replicator dynamics}: \; x_i' = x_i \frac{(Wx)_i}{x^T W x}.$$

Decision Problem

- Input: A $n \times n$ positive symmetric matrix W.
- Question: Assume that we start at random from Δ_n (space of possible allele frequencies). Is the probability positive that the limit point of replicator dynamics will have support greater than one (aka diverse)?

$$\textit{Replicator dynamics}: \; x_i' = x_i \frac{(Wx)_i}{x^T W x}.$$

Losert and Akin '82

If W is positive then we have the following two facts:

Decision Problem

- Input: A $n \times n$ positive symmetric matrix W.
- Question: Assume that we start at random from Δ_n (space of possible allele frequencies). Is the probability positive that the limit point of replicator dynamics will have support greater than one (aka diverse)?

$$\textit{Replicator dynamics}: \; x_i' = x_i \frac{(Wx)_i}{x^T W x}.$$

Losert and Akin '82

If W is positive then we have the following two facts:

- Replicator dynamics converges to a fixed point x^* (limit point).

Decision Problem

- Input: A $n \times n$ positive symmetric matrix W.
- Question: Assume that we start at random from Δ_n (space of possible allele frequencies). Is the probability positive that the limit point of replicator dynamics will have support greater than one (aka diverse)?

$$\textit{Replicator dynamics}: \; x_i' = x_i \frac{(Wx)_i}{x^T W x}.$$

Losert and Akin '82

If W is positive then we have the following two facts:

- Replicator dynamics converges to a fixed point x^* (limit point).
- The update rule is a diffeomorphism on Δ_n.

Our results

Theorem (Main result 1)

Given a $n \times n$ fitness matrix W for a diploid organism with single locus, it is NP-hard to decide if, under evolution (replicator dynamics), diversity will survive (by converging to a specific mixed equilibrium with positive probability) when starting allele frequencies are picked at random from Δ_n.

Our results

Theorem (Main result 1)

Given a $n \times n$ fitness matrix W for a diploid organism with single locus, it is NP-hard to decide if, under evolution (replicator dynamics), diversity will survive (by converging to a specific mixed equilibrium with positive probability) when starting allele frequencies are picked at random from Δ_n.

Theorem (Main result 2)

If the entries of a fitness matrix W are i.i.d. from an atomless (say continuous) distribution then with probability at least $1/3$ (over random entries and initial population), diversity will survive.

Dynamical systems to the rescue

Using techniques from dynamical systems, in particular Center-stable manifold theorem and the two facts mentioned above:

Lemma

The set of initial conditions in Δ_n so that the dynamics converges to "unstable" fixed points is of measure zero.

Dynamical systems to the rescue

Using techniques from dynamical systems, in particular Center-stable manifold theorem and the two facts mentioned above:

Lemma

The set of initial conditions in Δ_n so that the dynamics converges to "unstable" fixed points is of measure zero.

A fixed point p is "unstable" if the corresponding Jacobian of the rule at p has an eigenvalue greater than 1.

Dynamical systems to the rescue

Using techniques from dynamical systems, in particular Center-stable manifold theorem and the two facts mentioned above:

Lemma

The set of initial conditions in Δ_n so that the dynamics converges to "unstable" fixed points is of measure zero.

A fixed point p is "unstable" if the corresponding Jacobian of the rule at p has an eigenvalue greater than 1.

Easy remark

We need to characterize the "stable" fixed points to answer our question!

High level intuition

Tossing a coin, 3 possible steady (*e*quilibrium) states:

- Tail “Stable”

- Head “Stable”

- Landing on its edge “Unstable”

Do the set of “stable” fixed points contain mixed fixed points (support greater than one)?

Lemma

A fixed point p is "stable" iff the strategy profile (p, p) is a Nash equilibrium of the 2-player symmetric coordination game (W, W) and also T_W is negative semi-definite. We also call it stable Nash.

T_W is the resulting matrix if we subtract first row/column from all other rows/columns of W and then remove the first row/column.

Lemma

A fixed point p is "stable" iff the strategy profile (p, p) is a Nash equilibrium of the 2-player symmetric coordination game (W, W) and also T_W is negative semi-definite. We also call it stable Nash.

T_W is the resulting matrix if we subtract first row/column from all other rows/columns of W and then remove the first row/column.

Definition

Naturally, we define strict stable Nash.

Lemma

A fixed point p is "stable" iff the strategy profile (p, p) is a Nash equilibrium of the 2-player symmetric coordination game (W, W) and also T_W is negative semi-definite. We also call it stable Nash.

T_W is the resulting matrix if we subtract first row/column from all other rows/columns of W and then remove the first row/column.

Definition

Naturally, we define strict stable Nash.

Lemma

The set of fixed points with positive measure of attraction is sandwiched between strict stable Nash and stable Nash fixed points.

Proof of Theorem 2

It is easy to prove via inclusion-exclusion arguments (avoiding correlations):

Lemma

If the entries of a positive symmetric matrix W are i.i.d. from an atomless (say continuous) distribution then with probability at least $1/3 - o(1)$ we have that every diagonal entry W_{ii} is dominated by W_{ij}, W_{ji} for some j (different for every i).

Proof of Theorem 2

It is easy to prove via inclusion-exclusion arguments (avoiding correlations):

Lemma

If the entries of a positive symmetric matrix W are i.i.d. from an atomless (say continuous) distribution then with probability at least $1/3 - o(1)$ we have that every diagonal entry W_{ii} is dominated by W_{ij}, W_{ji} for some j (different for every i).

Hence, we don't have pure symmetric Nash equilibrium!

Proof of Theorem 2

It is easy to prove via inclusion-exclusion arguments (avoiding correlations):

Lemma

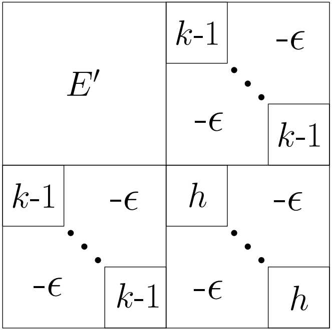
If the entries of a positive symmetric matrix W are i.i.d. from an atomless (say continuous) distribution then with probability at least $1/3 - o(1)$ we have that every diagonal entry W_{ii} is dominated by W_{ij}, W_{ji} for some j (different for every i).

Hence, we don't have pure symmetric Nash equilibrium!

Therefore all the Nash stable and strict Nash stable fixed points must be mixed!

Proof of Theorem 1

Use same reduction for both notions of stability. The reduction is from clique. Let E' be the adjacency matrix of input graph G, so that all zero entries are replaced by $-h$ (a large number depending on the graph size). Define the following fitness matrix W:



- If G has a clique of size k then (W, W) has a mixed strict Nash stable.
- If (W, W) has a mixed Nash stable, then G has a clique of size k.
- We can make W positive, by adding the same number $c > 0$ to every entry ("stability" does not change!).

Future work

- Introduce mutations.
- Changing environments.
- Generalize results for more genes, the equations are more complicated!
- More connections between theoretical computer science and evolution.

Thank you!

Center-Stable Manifold theorem

Center-Stable Manifold theorem (informally)

If the rule of the dynamics is a diffeomorphism then

- For every fixed point $\mathbf{p}$, there exists an open ball $B_{\mathbf{p}}$ so that if $\mathbf{x}(n)$ is inside $B_{\mathbf{p}}$ for all $n \geq 0$ then $\mathbf{x}(0)$ belongs to a (local) center stable manifold $W_{sc}(\mathbf{p})$ which has dimension equal to the dimension of the space spanned by eigenvectors of the Jacobian (at $\mathbf{p}$) with absolute value ≤ 1.