Information Spread in Structured Populations

Burton Voorhees
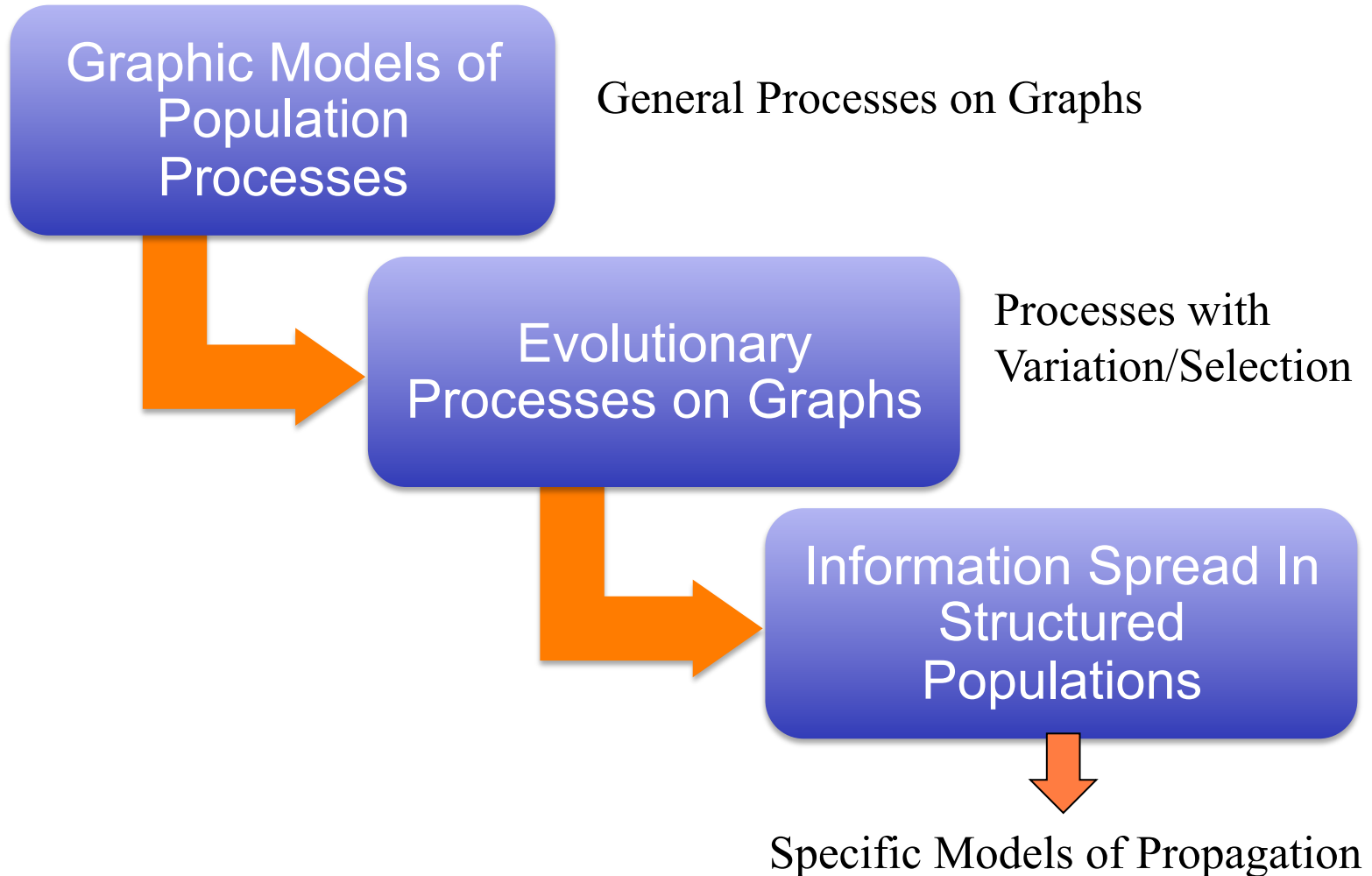## Center for Science
Athabasca University

Immediate References

B. Voorhees (2013) Birth-death fixation probabilities for structured populations. <u>Proceedings of the Royal Society A</u> <u>469</u> 2153
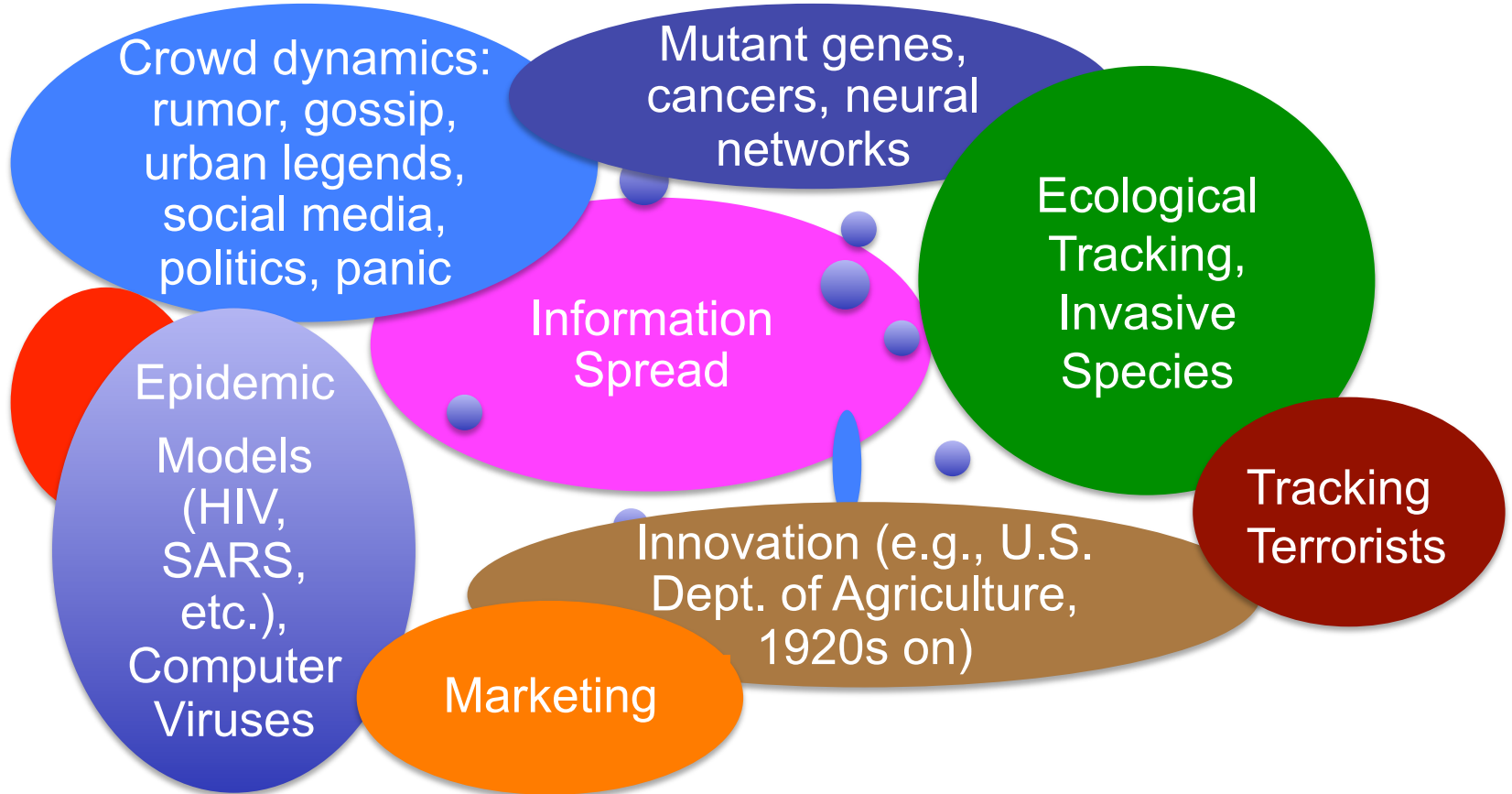
B. Voorhees & A. Murray (2013) Fixation probabilities for simple digraphs. <u>Proceedings of the Royal Society A</u> <u>469</u> 2154

I've also made up a list of general references that's available.

Topic and Sub-Topics

Graphic Models of Population Processes

General Processes on Graphs

Evolutionary Processes on Graphs

Processes with Variation/Selection

Information Spread In Structured Populations

Specific Models of Propagation

Many Areas of Application

Crowd dynamics: rumor, gossip, urban legends, social media, politics, panic

Mutant genes, cancers, neural networks

Ecological Tracking, Invasive Species

Information Spread

Epidemic Models (HIV, SARS, etc.), Computer Viruses

Tracking Terrorists

Innovation (e.g., U.S. Dept. of Agriculture, 1920s on)

Marketing

The Initial Question

Given a genetically homogeneous population of size N, having relative fitness 1 for all members, suppose that a single mutation with fitness r is introduced at time 0. The fitness r may be greater or less than 1.

What is the probility this mutation will become fixed in the population?

The Initial Answer

Patrick Moran (1958) gave an answer using a Markov <u>Birth-Death process</u>: At each time step:

Choose a vertex at random, biased by fitness, to "reproduce."

Choose a vertex accessible from the reproducing vertex to die

Replace the dead vertex by a copy of the reproducing vertex

The Initial Answer: The Math

The "Moran process" is a discrete time Markov process on $S = \{0,1,\ldots,N\}$ where state m corresponds to m mutants and $N - m$ normals in the population. If $p_{m,m\pm1}$ is the probabilities of state transitions $m \rightarrow m\pm1$ and x_m is the probability of fixation starting from state m then $x_0 = 0$ and $x_N = 1$.

The Initial Answer: The Math

The Markov evolution equations are

$$x_0 = 0, \quad x_N = 1$$

$$x_m = p_{m-1,m}x_{m-1} + p_{m,m}x_m + p_{m+1,m}x_{m+1}$$

and the fixation probability $\rho_M = x_1$ is

$$\rho_M = \frac{1}{1 + \sum_{k=1}^{N-1}\prod_{m=1}^{k} g_m}, \quad g_m = \frac{p_{m,m-1}}{p_{m,m+1}}$$

The Initial Answer: The Math

For the Moran process

$$p_{m,m-1} = \frac{1}{N-m+rm}, \quad p_{m,m+1} = \frac{r}{N-m+rm}$$

$$g_m = \frac{1}{r}$$

$$\rho_M = \frac{1-\dfrac{1}{r}}{1-\dfrac{1}{r^N}} = \frac{r^{N-1}}{1+r+\cdots+r^{N-1}}$$

The Next Question

Does population structure influence this result?
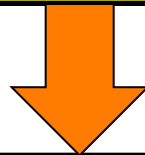
Represent a population of size N by a graph with N vertices together with an edge-weight matrix W such that w_{ij} is the probability that <u>if</u> individual i is chosen for reproduction, <u>then</u> individual j will be chosen for death.

The Next Answers

Early studies (Maruyama, 1974; Slatkin,1981) indicated that population structure did not seem to alter fixation probabilities.

In 2005, however, Lieberman, Hauert, & Nowak published the seminal paper (<u>Nature</u>, <u>433</u>, 312 – 316), demonstrating that certain graphs could <u>suppress</u> or <u>enhance</u> fixation probability relative to the Moran case.

The Isothermal Theorem

<u>Isothermal Theorem</u> (Liberman, et al, 2005)
A graph G with edge weight matrix W is fixation-equivalent to a Moran process if and only if for all j, k

$$\sum_{i=1}^{N} w_{ij} = \sum_{i=1}^{N} w_{ik}$$

I.e., if and only if W is doubly stochastic.

(The sums here are called the "temperatures" of vertices j and k; the hotter the temperature the more exposure a vertex has to change.)

Some Moran-Equivalent Graphs

c = complete graph,
d = cycle

*
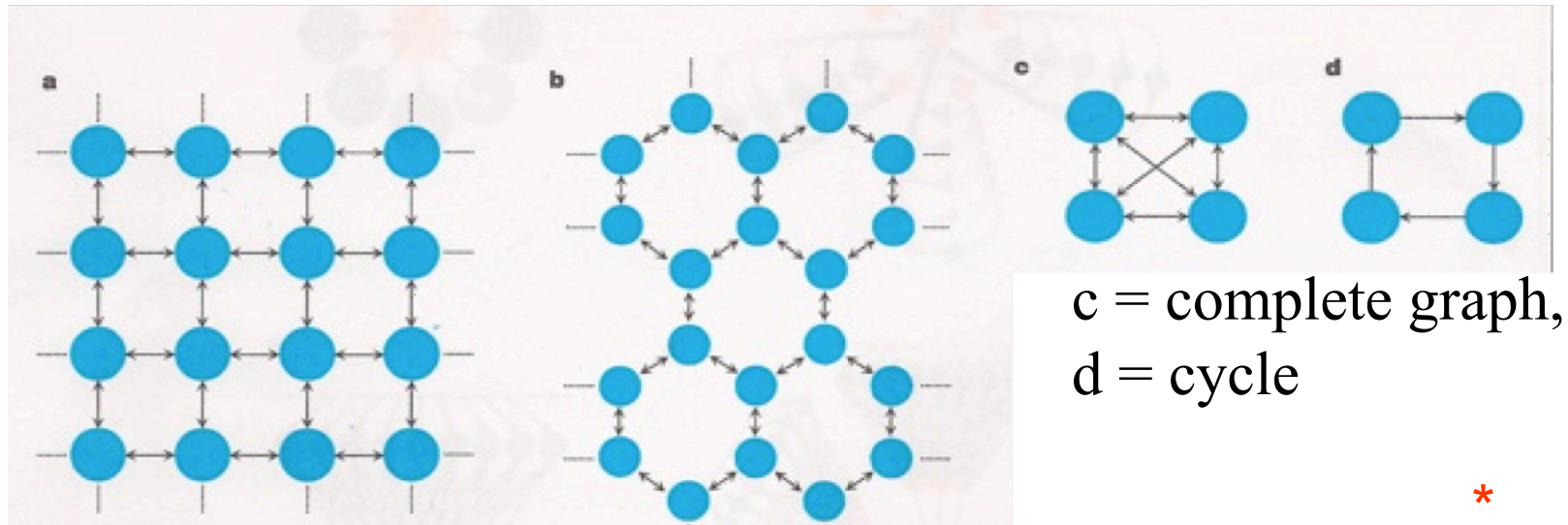
Early work on fixation probabilities based on graphs such as these indicated that population structure had no influence.

*From Liberman, et al (2005)

Suppressing Selection

Some graphs that suppress selection:

*
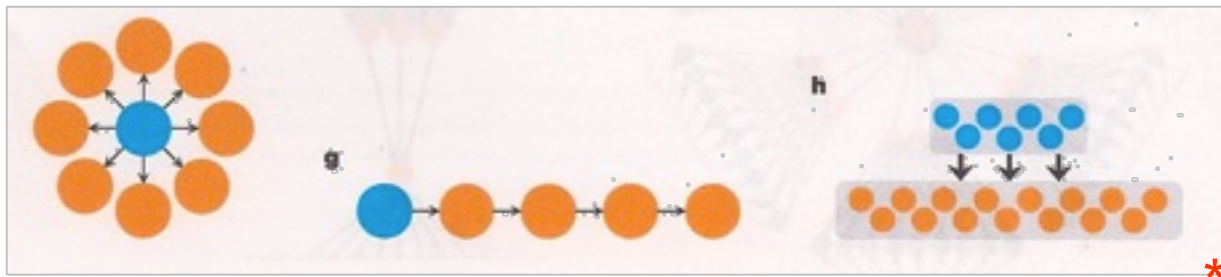
In particular, the line and rooted tree have fixation probability 1/N.
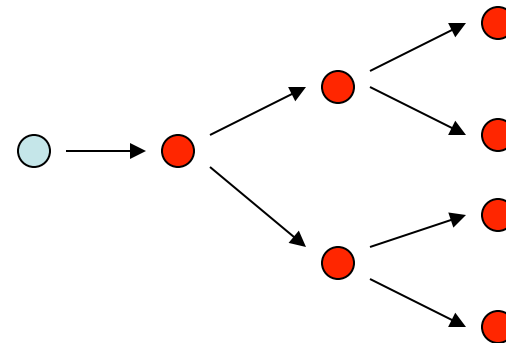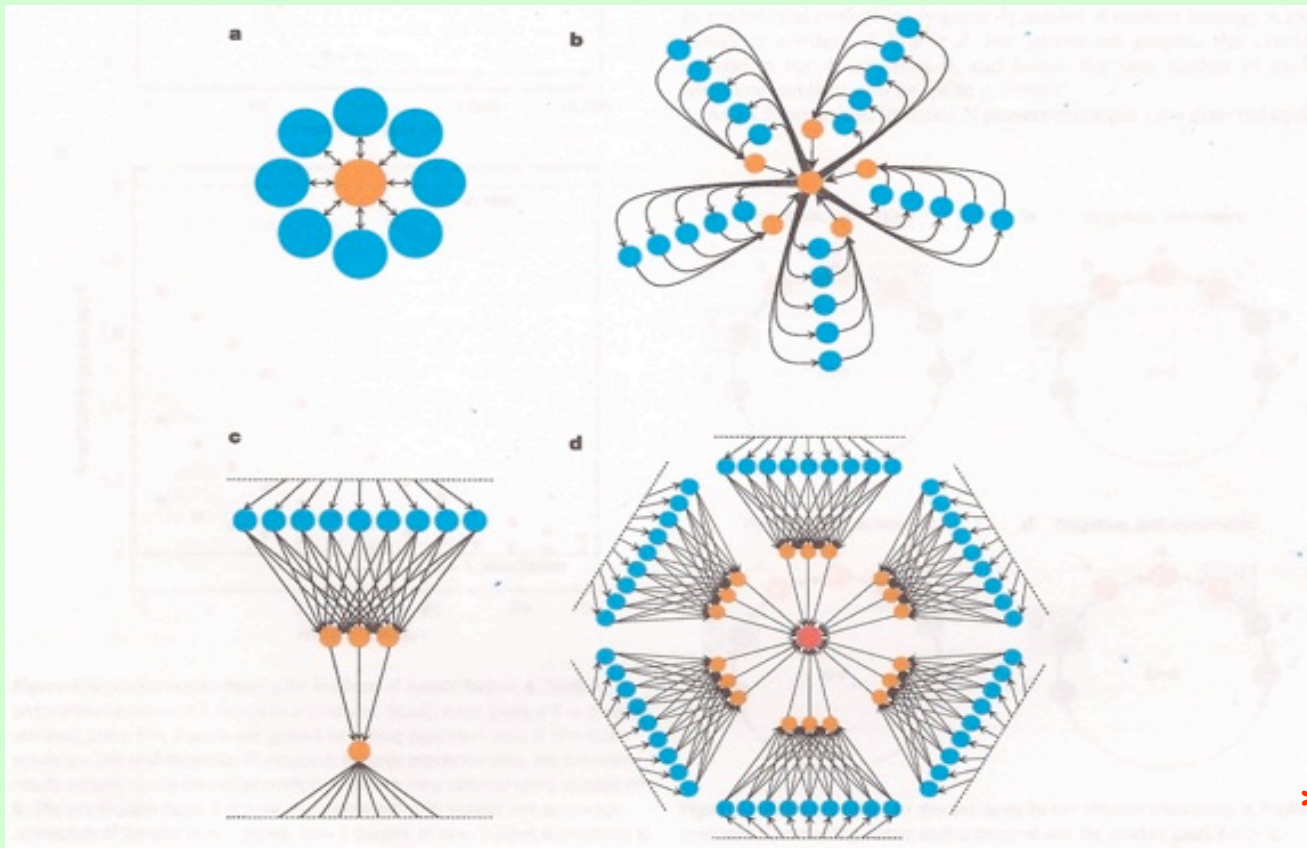

* From Liberman, et al (2005)

Suppressing Selection

Population structures that suppress selection can protect against rapidly reproducing malignant mutations. Skin tissue and colon lining, for example, are hierarchically structured from less to more differentiated cells. Only mutations that appear in the initial stem cells can go to fixation.

Stem
Cell

Skin Cells
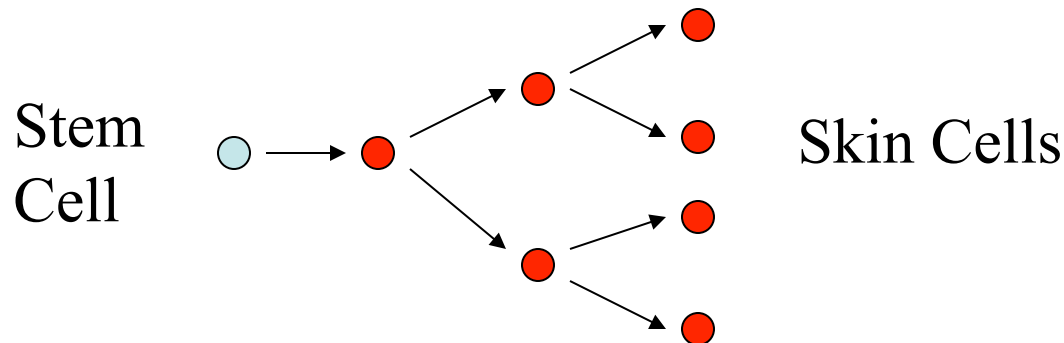
Enhancing Selection

Some graphs that enhance selection:

*From Liberman, et al (2005)

Birth-Death Dynamics on Graphs

A state space approach

A configuration of mutants defines a length N binary vector $\vec{v}$ with v_i equal to 0 or 1 as vertex i is occupied by a normal or mutant. The full state space is the vertex set $V(H^N)$ of the N-Hypercube.
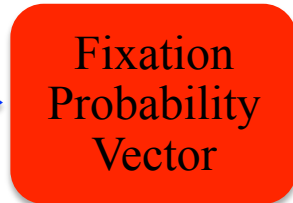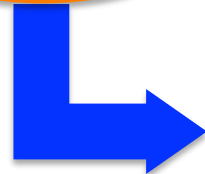
Line of Solution
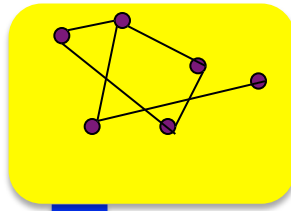
- Population graph (G,W)
- W is weighted adjacency matrix

Construct
transition
matrix T

Subset of N
Hypercube
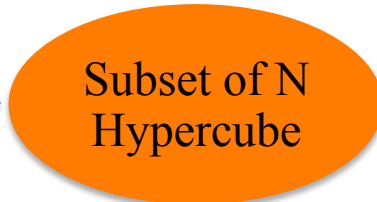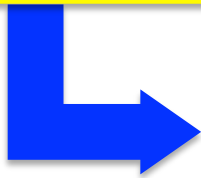
- State Transition Diagram
  (STD,T) Size $2^N \times 2^N$

$$(I-T) \cdot \vec{x} = 0$$

Fixation
Probability
Vector

- Fixation
  Probability

Birth-Death Dynamics on Graphs

Population evolution is given by a Markov process on $V(H^N)$ with transition matrix T. Given T, the steady state solution determines fixation probabilities.

The question is finding T given the edge weight matrix W.

Birth-Death Dynamics on Graphs

Given a population state $\vec{v}$, define vectors

$$\vec{a}(\vec{v}) = \vec{v} \cdot W \, , \ \vec{b}(\vec{v}) = \vec{v}\,' \cdot W \qquad v_i' = 1 - v_i$$

$\underline{a_j(\vec{v})}$: the sum of probabilities that an edge from a mutant vertex terminates at vertex j.

$\underline{b_j(\vec{v})}$: the sum of probabilities that an edge from a normal vertex terminates at vertex j.

$\vec{a}(\vec{v}) + \vec{b}(\vec{v}) = \vec{t}$, where $\vec{t}$ is the $\underline{\text{temperature}}$ vector.

Birth-Death Dynamics on Graphs

Theorem (Transition Matrix Construction):
Let a birth-death process be defined on a graph G with edge weight matrix W.

1. The transition probability
$(v_1,\ldots,v_{j-1},0,v_{j+1},\ldots,v_N) \to (v_1,\ldots,v_{j-1},1,v_{j+1},\ldots,v_N)$
is
$$\frac{r a_j(\vec{v})}{N - m + rm}$$

Birth-Death Dynamics on Graphs

2. The transition probability
$(v_1,\ldots,v_{j-1},1,v_{j+1},\ldots,v_N) \rightarrow (v_1,\ldots,v_{j-1},0,v_{j+1},\ldots,v_N)$
is

$$\frac{b_j(\vec{v})}{N - M + rm}$$

3. The probability that $\vec{v}$ remains unchanged is

$$\frac{r\vec{a}(\vec{v}) \cdot \vec{v} + \vec{b}(\vec{v}) \cdot \vec{v}'}{N - m + rm}$$

The Transition Matrix T

The transition matrix T is row stochastic with maximum eigenvalue 1 and corresponding eigenvector $\vec{\xi} = s\mathbf{1}$ where $\mathbf{1}$ is the 2^N vector of all ones. There are two absorbing states, 0 and 1: the mutation either becomes extinct, or goes to fixation.

The Transition Matrix T

Since **0** and **1** are the only absorbing states the matrix

$$T^* = \lim_{k \to \infty} T^k$$

consists of initial and final non-zero columns with all other entries equal to 0:

$$T^*_{uv} = 0, v \neq 0, 2^N - 1.$$

Birth-Death Dynamics on Graphs

Theorem 1 allows construction of the transition matrix T:V(H^N)➔V(H^N). If $\vec{v}$ is an element of V(H^N) it is a binary number with denary form

$$v = \sum_{k=1}^{N} v_k 2^{k-1}$$
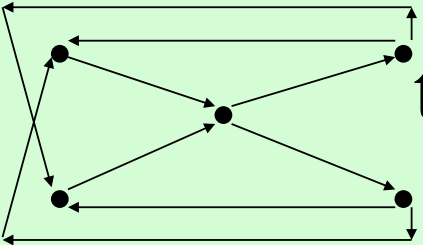
Let x$_v$ be the probability of fixation starting from the state $\vec{v}$.

Birth-Death Dynamics on Graphs

Theorem: If x_v is the fixation probability starting from an initial configuration represented by the vector $\vec{v}$ then the Markov process on $V(H^N)$ yields the set of master equations:

$$\left[N + (r-1)m - r\vec{a}(\vec{v}) \cdot \vec{v} - \vec{b}(\vec{v}) \cdot \vec{v}' \right] x_v$$

$$- r \sum_{i=1}^{N} a_i(\vec{v}) v_i' x_{v+2^{N-i}} - \sum_{i=1}^{N} b_i(\vec{v}) v_i x_{v-2^{N-i}} = 0$$

$$x_0 = 0, \ x_{2^N-1} = 1$$

Example

For the graph there are 16 equations

and

$$\rho = \frac{r^4 R(r)}{5P(r)}$$

n	R(r) coefficient of r^n	P(r) coefficient of r^n
0	4704	1152
1	66,408	16,416
2	419,344	108,384
3	560,082	439,424
4	3,801,097	1,226,164
5	6,388,510	2,515,774
6	7,603,301	3,992,261
7	6,468,470	5,131,031
8	3,912,548	5,549,348
9	1,645,520	5,131,031
10	458,496	3,992,261
11	76,320	2,515,774
12	5760	1,226,164
13		439,424
14		108,384
15		16,416
16		1152

General Analytic Solutions

In a 2008 paper, Broom & Rychtar give the exact solution for the n-Star as well as a means of solving for any given line graph. Zang, et al (2012) compute the k-vertex fixation probability for star graphs and give an approximation for the complete bipartite graph.

Another Analytic Solution

About two years ago I found the exact fixation probability for the complete bipartite graph $K_{s,n}$.
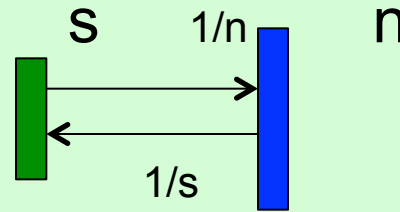
$$\rho_S(s,n) = \left(\frac{r^{n+s-1}}{sr+n}\right)\left[\frac{\left(snr + n^2 - sn + s^2\right)\left(nr + s\right)^{n-s-1}}{P(s,n)}\right]$$

$$P(s,n) = \frac{r^{n+s}(nr+s)^{n-s} - (sr+n)^{n-s}}{r^2 - 1}.$$

These, together with Moran-equivalent cases, are the only known general solutions.

Enhancement in Bipartite Graphs

Typical graphs of $\rho_{K_{s,n}} - \rho_M$ show selection enhancement for r > 1.
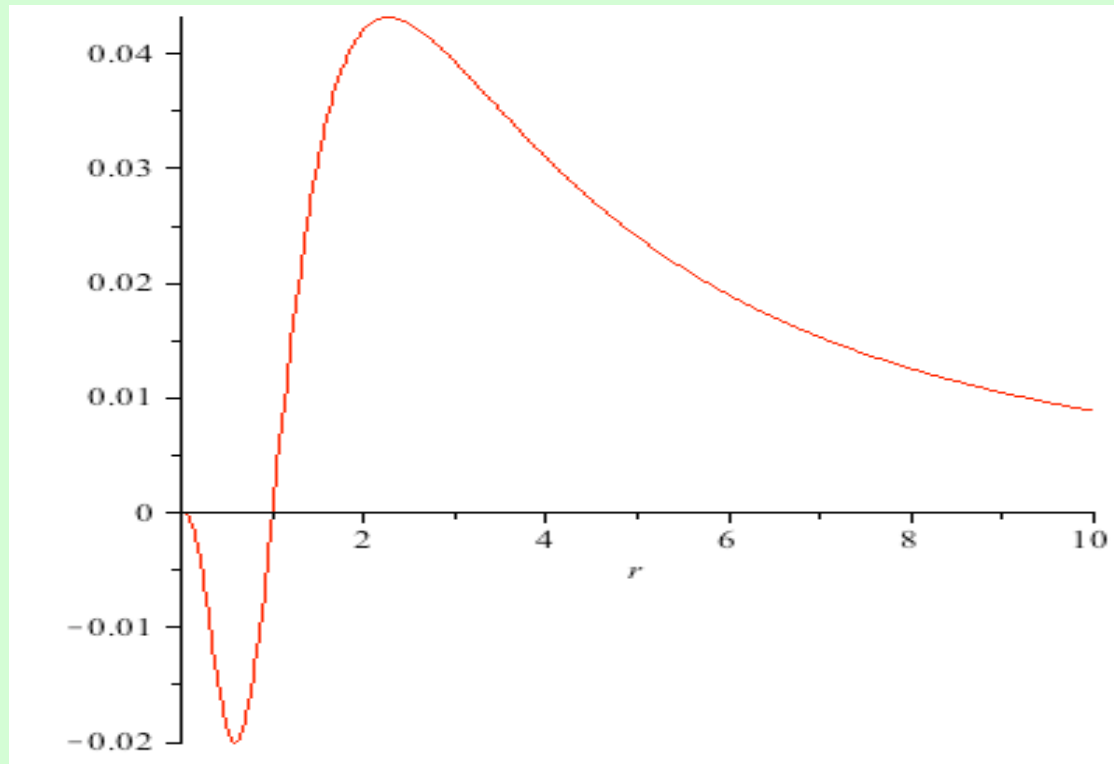
Example: (1,2,n) Funnel Graphs
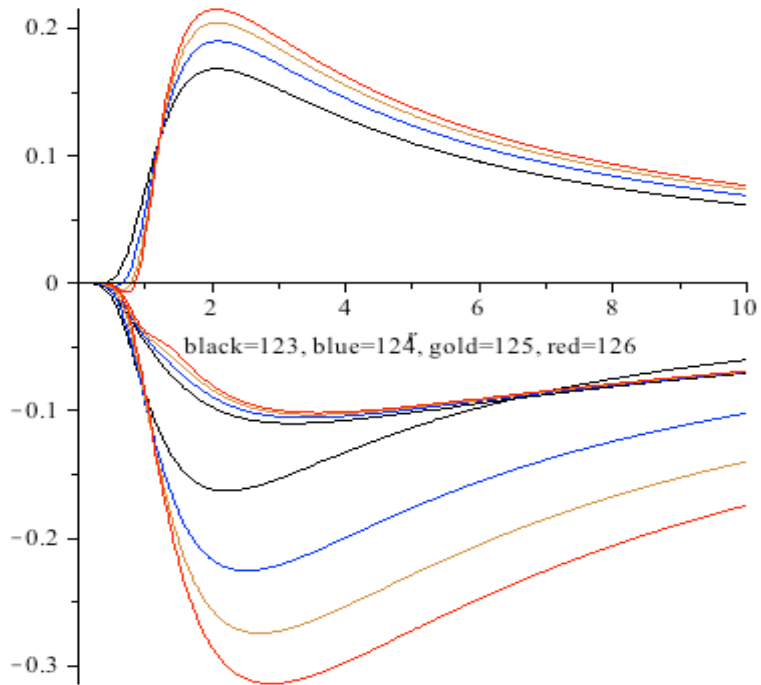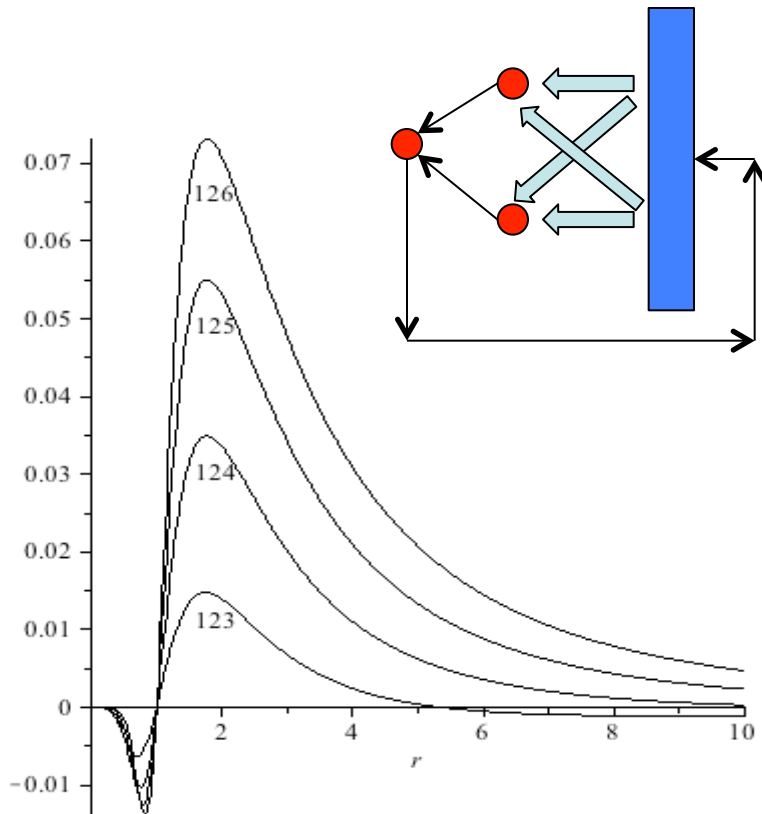
$\rho_G - \rho_M$ for graphs, and for each class in graph.

Examples
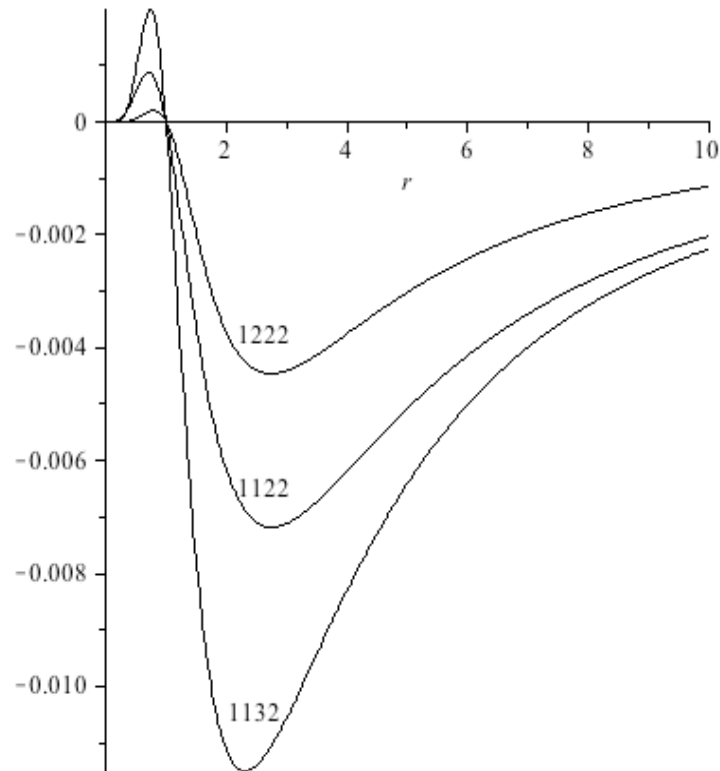
What is particularly interesting about these graphs is that they increase the fixation probability relative to a Moran process for r < 1.

Limitations and a Warning

1. The major limitation is that for a general graph there are $2^N - 2$ equations to solve.
2. This makes it important to develop approximation methods; using, e.g., the thermodynamic analogy, or mean field approaches.
3. But WARNING: Averaging methods will return the Moran result, eliminating information about the influence of graph topology.

An Application of T

Suppose a distribution of mutants (or infected sites, etc.) is observed, represented by the binary vector $(v_1,\ldots,v_N)$. Further, suppose we can estimate the time t that has elapsed since the initial "infection." Then the probability that the initial infected node was u is given by

$$P_{\text{inf}}(u \mid v(t)) = T^t_{uv}$$

Application Questions: Population Polarization

Is it possible to find a graph-based measure of polarization within a population?

For a population represented by a graph (G,W) the spectrum of the Laplacian gives information on the connectivity of G
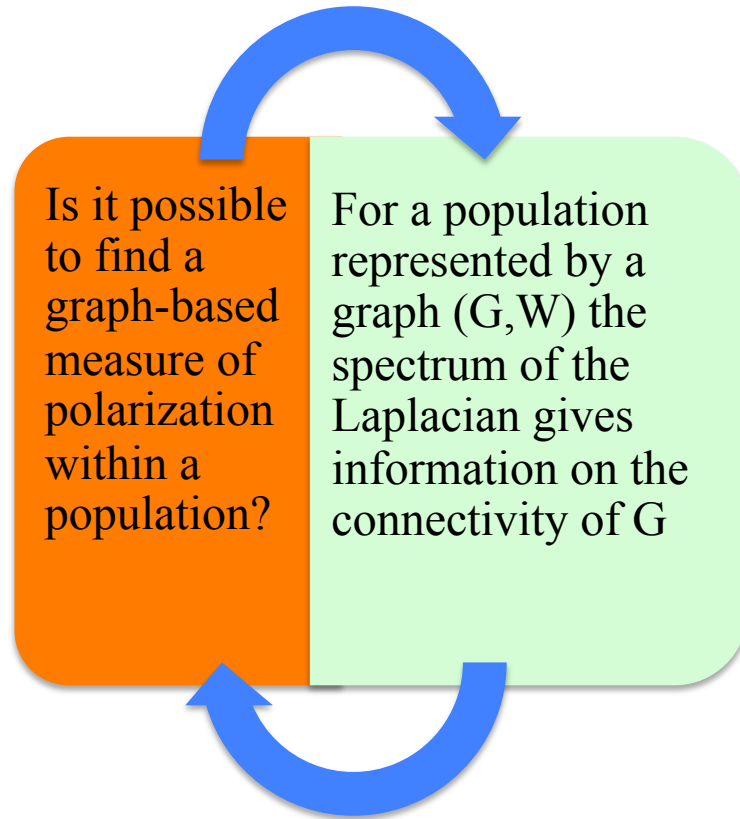
Population Polarization

For an undirected graph the matrix W, the matrix $\Delta = I - W$ is the graph Laplacian. For a directed matrix $\Delta = \Phi^{1/2}(I - W)\Phi^{-1/2}$ where Φ is the diagonal matrix with entries equal to the steady state solutions for W. The first eigenvalue of Δ is always 0. The second eigenvalue (the "algebraic connectivity") gives information about the difficulty of cutting a graph into disconnected parts.

Population Polarization: Examples
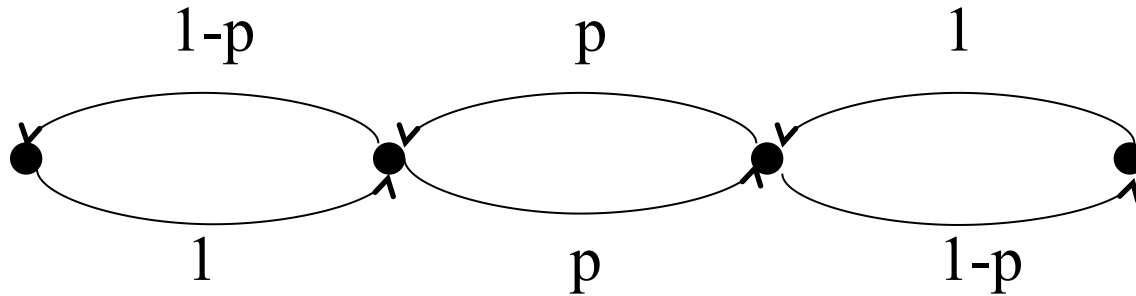
$$W = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & p & 0 & 1-p \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \Delta = \begin{pmatrix} 1 & -\sqrt{1-p} & 0 & 0 \\ -\sqrt{1-p} & 1 & -p & 0 \\ 0 & -p & 1 & -\sqrt{1-p} \\ 0 & 0 & -\sqrt{1-p} & 1 \end{pmatrix}$$

Eigenvalues of Δ: 0, p, 2-p, 2

3,4 Graphs: Limit Cases
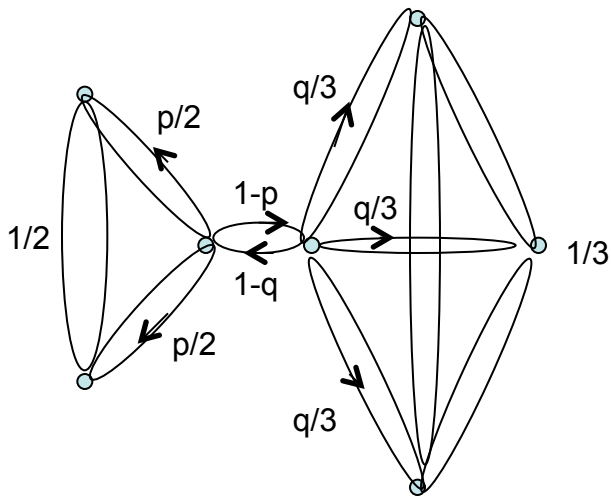
Single Link
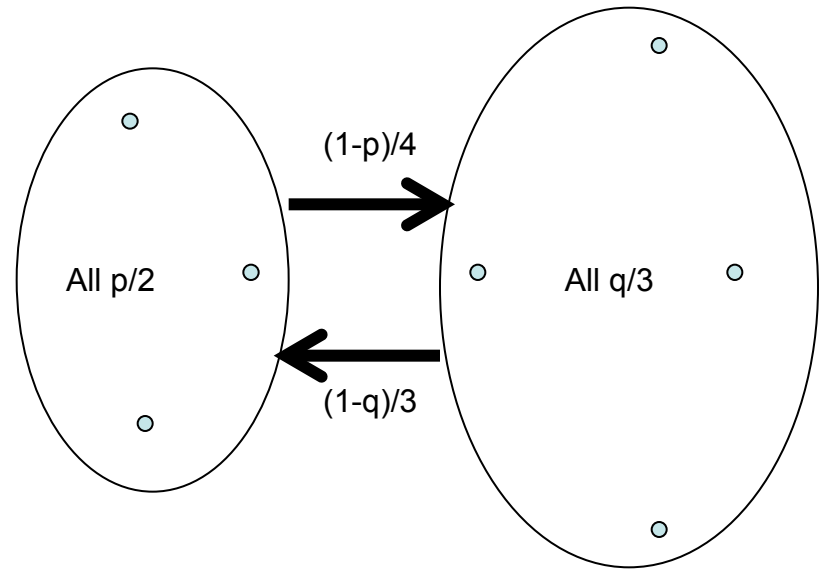
Partial Bipartite

Fixation Probability Minus Moran Probability for Partial Bipartite (3,4), r=2
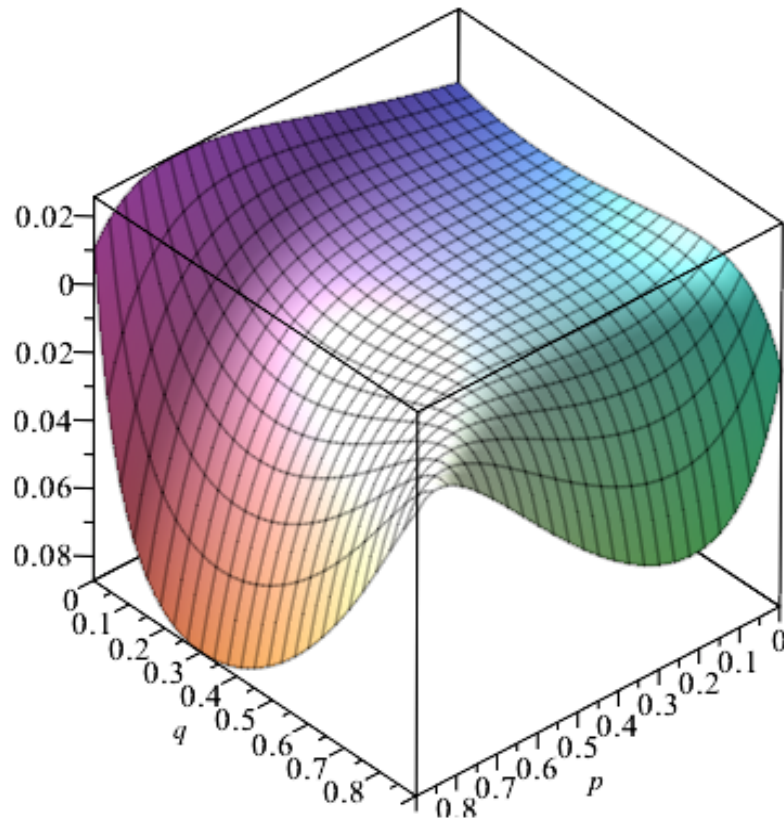
Graph only goes to p = 9/10, q = 9/10

Fixation Probabilities 3,4 Partial Bipartite Graph, r = 2
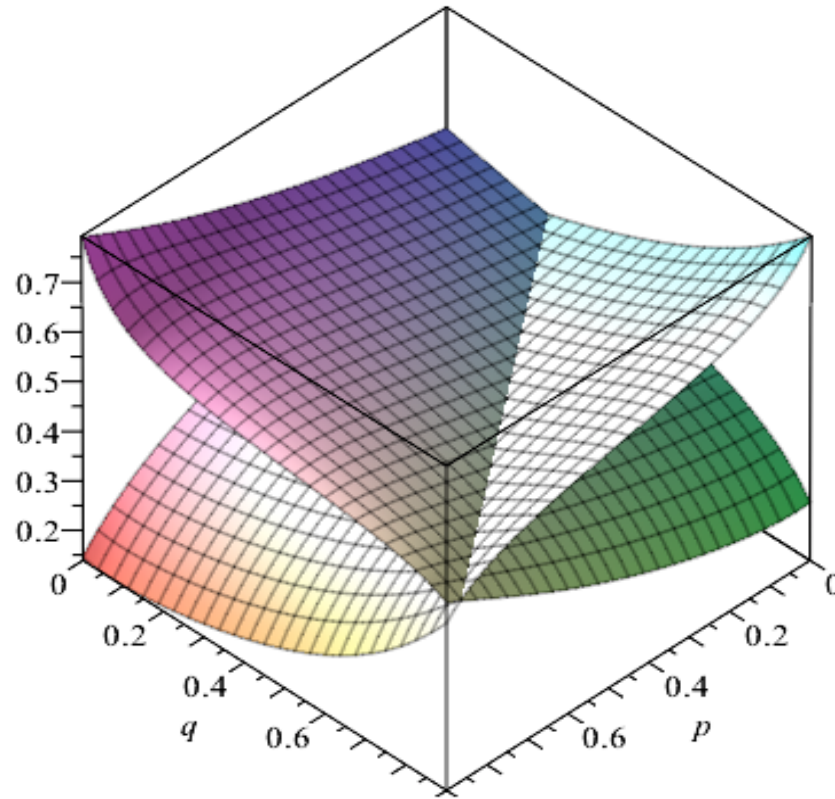
Lower sheet is fixation probability of starting from a vertex on the 3 side, upper sheet from starting at a vertex on the 4 side.

Fixation Probability Minus Moran Probability for Single Link (3,4), r=2
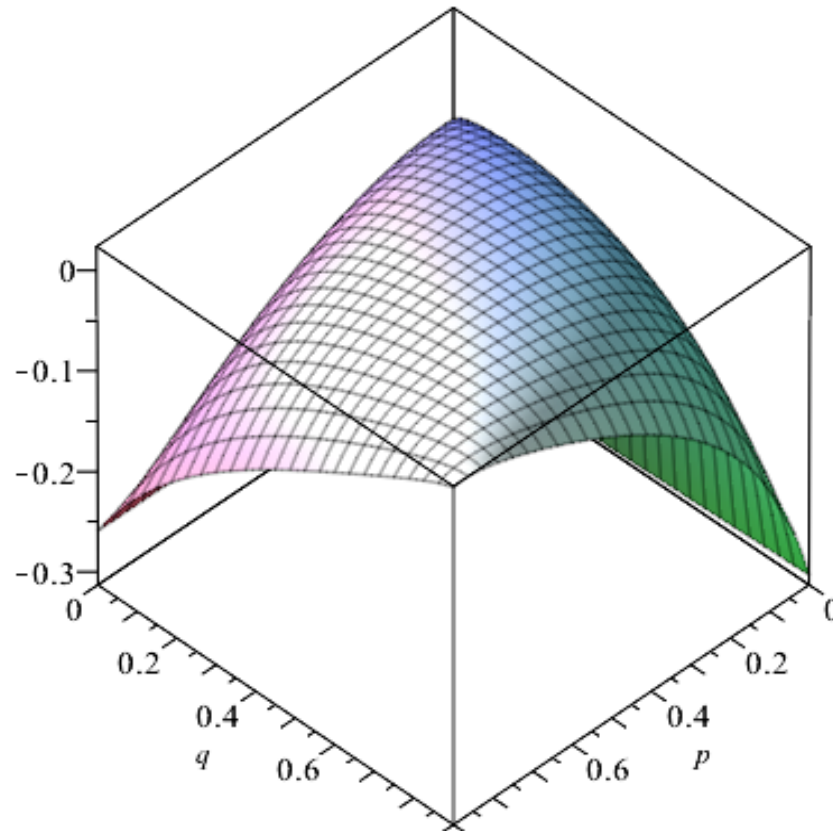
Graph only goes to p = 9/10, q = 9/10

Comparison
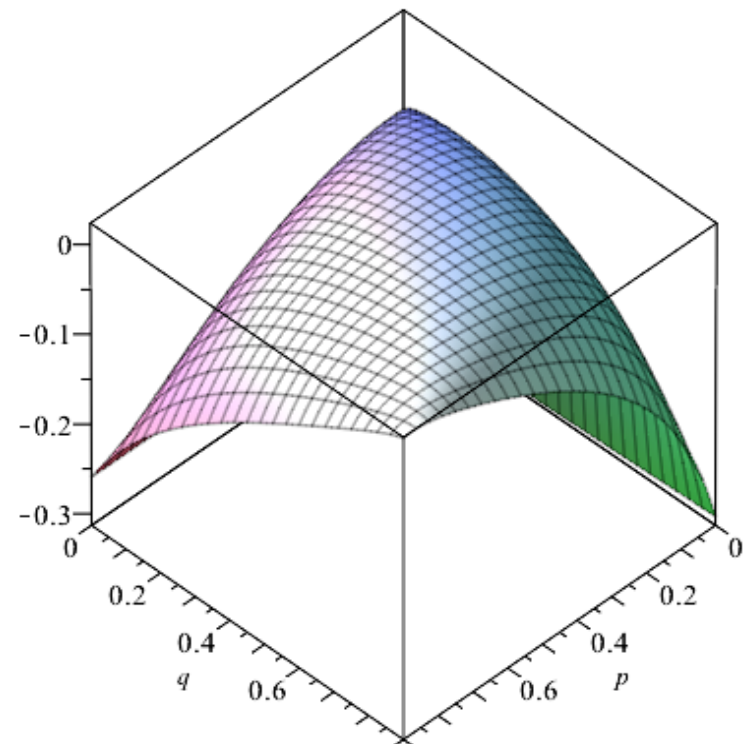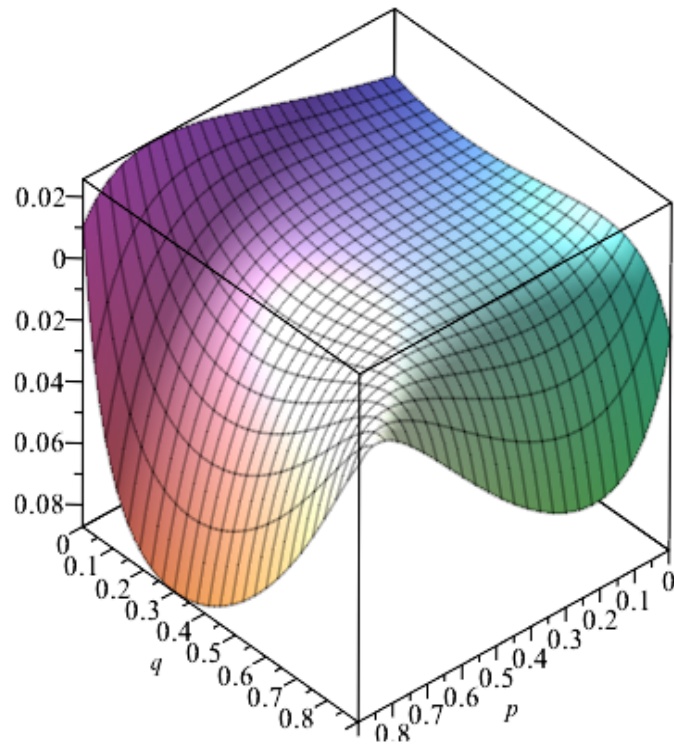
Graphs goes to p = 9/10, q = 9/10

Comparison
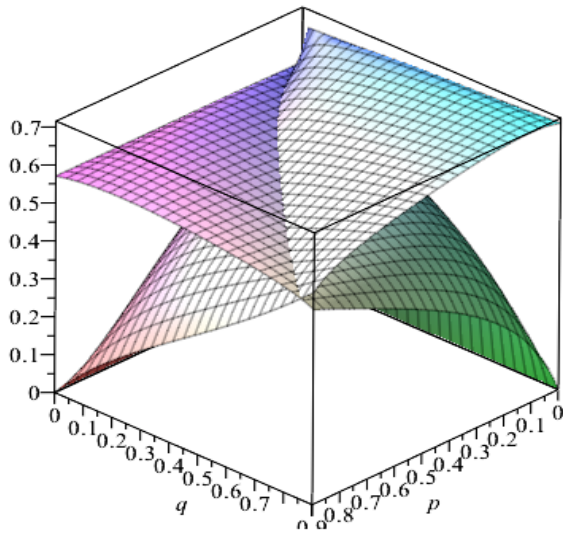
Fixation Probabilities for vertex on 2 side and vertex on 3 side.
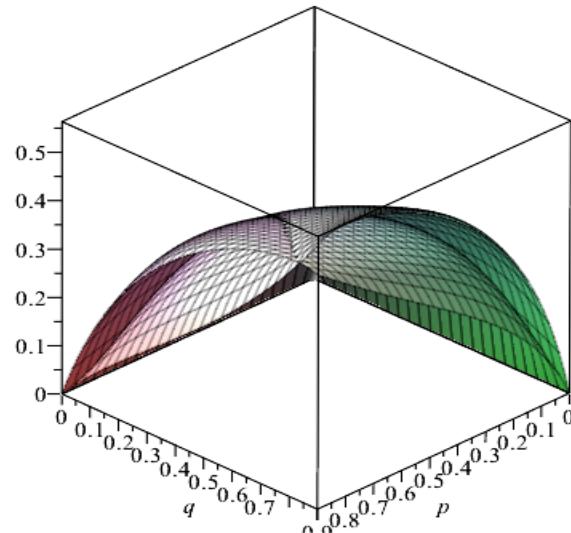
Fixation Probabilities linking vertices.
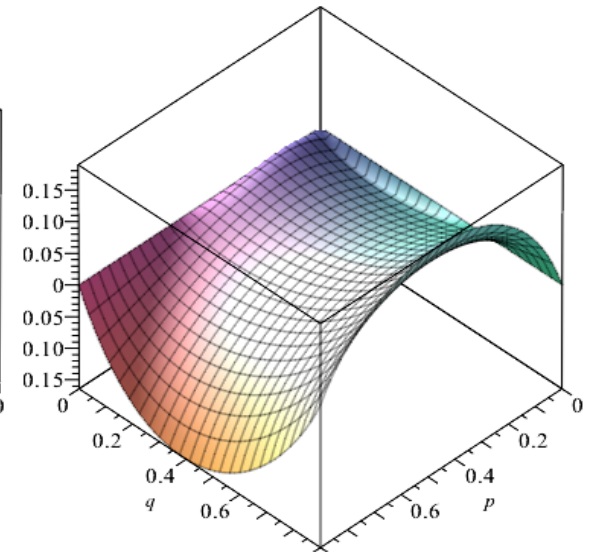
Difference between single link vertices fixation probabilities.

Graphs goes to p = 9/10, q = 9/10

Laplacian Roots

$$\left(x-\frac{1}{n-1}\right)^{n-2}\left(x-\frac{1}{m-1}\right)^{m-2}\left\{\left[x^2-p'q'\right]\left(x-\frac{n-2}{n-1}\right)\left(x-\frac{m-2}{m-1}\right)\right.$$

$$\left.-\frac{p}{n-1}\left(x-\frac{m-2}{m-1}\right)x-\frac{q}{m-1}\left(x-\frac{n-2}{n-1}\right)x+\frac{pq}{(n-1)(m-1)}\right\}$$

$$p'=1-p, \quad q'=1-q$$

Single Link Case

$$(x+1)(x-1+p+q)(x-p/(n-1))^{n-1}(x-q/(m-1))^{m-1}$$

(n,m) Partial Bipartite

Note: x = $\lambda-1$

Laplacian Roots, Single Link Case
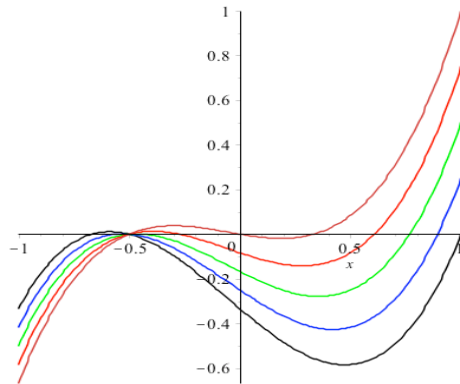
P=0
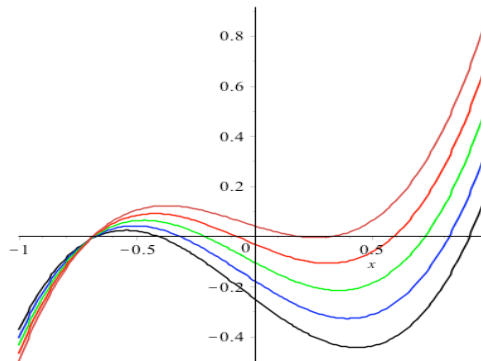
P=1/4
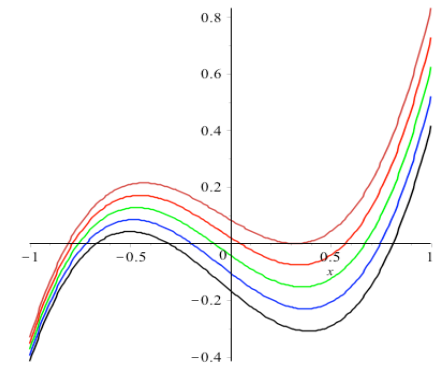
P=1/2
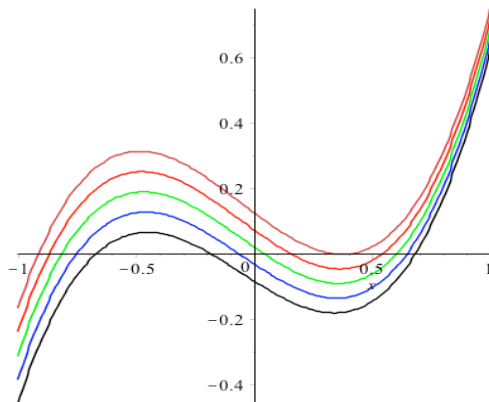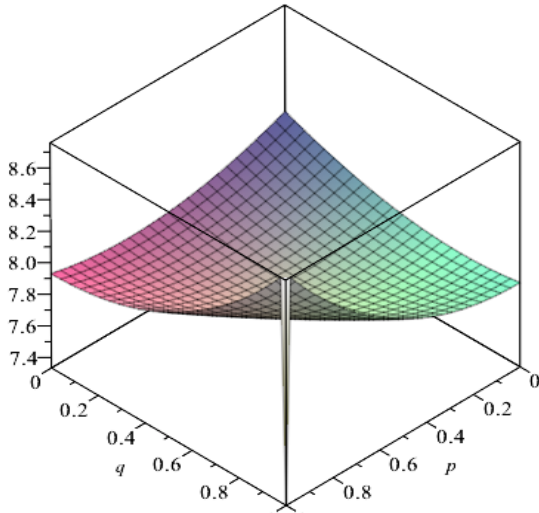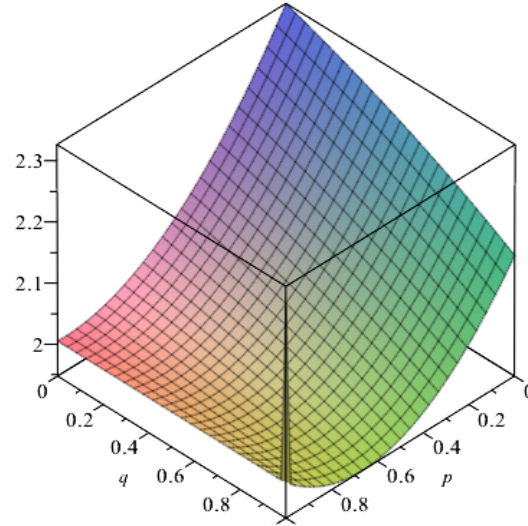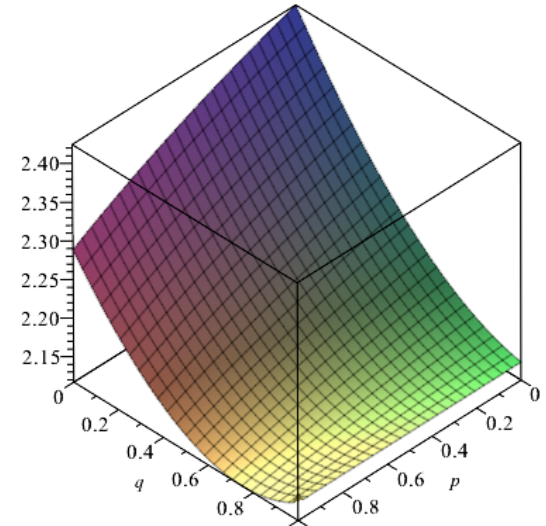
P=3/4
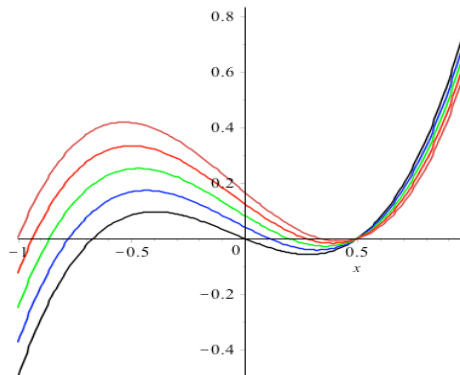
P=1
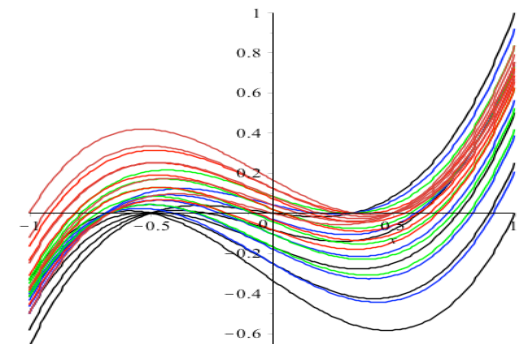
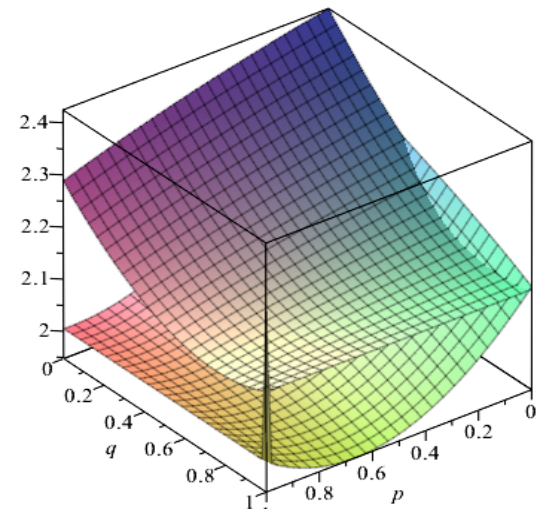# *Measures of Connectivity: exp(W) for 3,4 Partial Bipartite*

Trace of exp(W)

s=3 vertex centrality
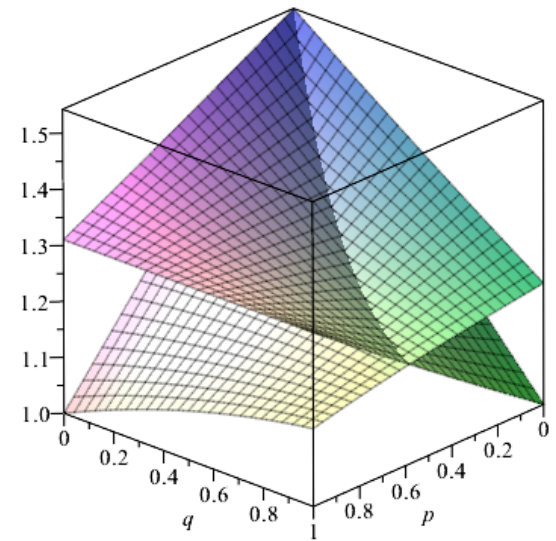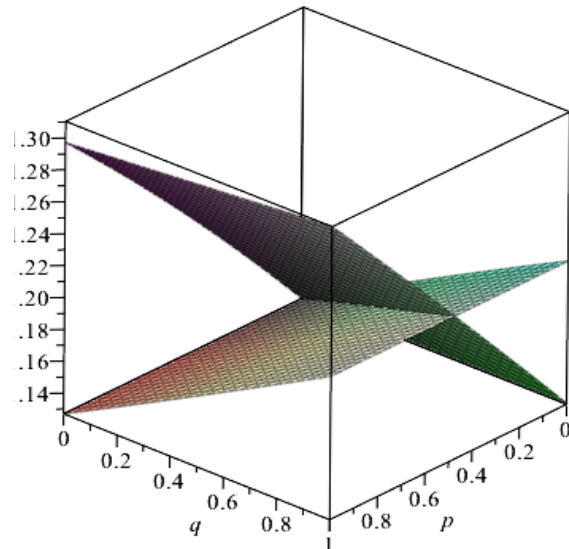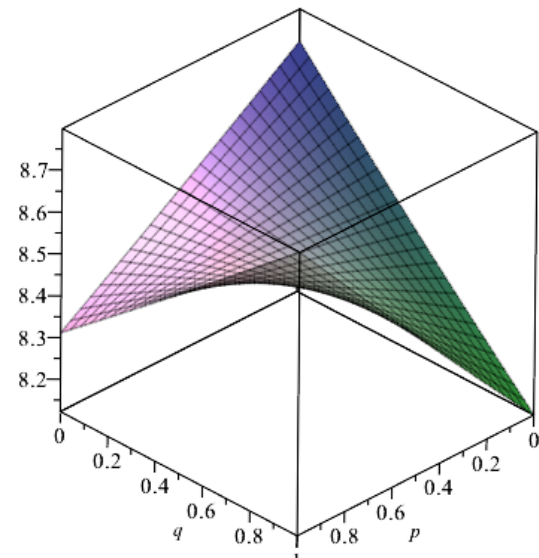
n=4 vertex centrality

Minimum of trace
At p = 1/3, q = 1/2

Exp(W) is called the matrix
"communicability."

Measures of Connectivity: exp(W) for 3,4 Single Link

Trace of exp(W)

Unlinked vertex centralities (lower is from s = 3 side, upper from n = 4 side

Centralities of linked vertices (lower from 3 side, upper from 4 side

Other Update Paradigms

In all of this, I've used a birth-death update paradigm. Other approaches are possible and the approach chosen can influence the results obtained.

Other Update Paradigms

Some possibilities:
1. Birth-Death
2. Death-Birth
3. Voter Models
4. Probabilistic Voter Models

Basic Questions for Study

1. How does interaction structure influence the spread of information in a population?
2. What is the probability of a mutation taking over a population, of a (computer) virus spreading, or of a rumor going viral?
3. If a distribution of mutants is observed, where did it most likely originate? E.g., where did this mutant, virus, rumor, innovation, etc., come from?

Thank You