Understanding Organism Growth and Cellular Differentiation Through Evolutionary Selection of Dynamical Properties.

by

Yuri Cantor

Thesis Proposal for the degree of Doctor of Philosophy at the Graduate Center of the City University of New York

Committee: Professor Bilal Khan (Advisor)
Professor Nancy Griffeth
Professor Matthew Johnson
Professor Kirk Dombrowski

May, 2016
Contents

1 Abstract 1

2 Introduction 2

2.1 Preliminary Definitions 3

3 Area 1: Cellular Differentiation and Robustness 8

3.1 Introduction 8

3.2 Definitions 9

3.3 Formal Hypothesis 1 10

3.4 Research Plan 10

3.5 Preliminary results 11

4 Area 2: Cellular Differentiation and Adaptivity 12

4.1 Introduction 12

4.2 Definitions 13

4.3 Formal Hypothesis 2 14

4.4 Research Plan 14

4.5 Preliminary results 15

5 Area 3: Symmetry and Growth 16

5.1 Introduction 16

5.2 Definitions 18

5.3 Formal Hypothesis 3 20

5.4 Research Plan 21
5.5 Preliminary results .......................................................... 23

6 Obstacles .................................................................. 24

  6.1 Introduction ................................................................. 24

  6.2 Candidate Approaches .................................................. 25

      6.2.1 Powers of 2 ......................................................... 25

      6.2.2 Sampling techniques ............................................. 26

      6.2.3 Distributed processing .......................................... 27

7 Summary .................................................................. 27

8 Timeline .................................................................. 28
1 Abstract

Computer simulations of networks of cellular automata serve as an experimental approach to exploring relationships between properties of a network and the dynamics of the network. Consequently, these dynamical systems are commonly used to model biological organisms representing cells with the abstraction of cellular automata.

Natural selection describes a process where traits or properties of the biological organism that are better suited to the environment will be selected for. It follows that properties of the dynamics that are better suited to the environment will also be selected for. This proposal studies the impact of Darwinian selection on cellular differentiation and organism growth and focuses on the relationship of organism properties of homogeneity and heterogeneity to the dynamical properties of robustness, adaptivity, and chromatic symmetry.

I propose an experimental approach to show that cellular differentiation increases the expected robustness in an organism’s dynamics, cellular differentiation leads to increasing adaptivity, and organism growth by elongation preserves symmetry in the dynamics of an organism. Further, I will also address practical concerns arising from computational limitations and big data in experimental approaches to studying dynamical systems.
2 Introduction

Biological neural network behavior is difficult to model computationally. These networks are composed of densely connected neurons, have a state of excitation that is continuous in degree, and transmit state asynchronously [50]. There has been extensive research into biological networks and cellular automata (see [44][17] for brief surveys). Cellular automata as described by Von Neumann [40] can be connected to form a network which serves as an abstraction of a biological networks and computational simulation of these simplified networks of cellular automata can lead to understanding the behavior in correlated complex biological networks [28][33][46][19].

In an experimental approach abstractions are necessary to maximize the number of networks and depth of simulation that can be modeled. Further, it is known that decomposing super networks into sub networks can lead to understanding pathways and signaling in biological circuitry [53]. Random Boolean Networks (RBNs) have been used to abstract from continuous state values to Boolean. Kauffman's NK network model abstracts from densely connected to K connections [20][21]. The trivial case where K=1, restricts the nodes to connected along a line. While Conway’s "Game of Life" is example of a Boolean NK network represented using a two dimensional grid where K=8 [18]. Asynchronous transmissions with small temporal tolerances can be modeled as synchronous [39].

This proposal explores theoretical synthetic biology, using the more restricted class of synchronous Boolean cyclic NK networks (where K=2) as a formal basis. This class of networks has received considerable attention [47], and exhibits many of the phenomena seen in more general NK counterparts [45]. Even with these reductions in complexity, fully simulating the dynamics remains computationally intractable except for networks of relatively small size [51].

The relationship of network structural properties to dynamical properties is crucial to understand
how and why biological networks have evolved. Further, this relationship can be used to construct networks with desirable properties [41]. Dynamical systems have been previously be evaluated for their landscape ruggedness [35][34], redundancy [21], and reversibility and surjectivity (reachability and Garden of Eden states [47]) [26][37]. I propose an experimental approach to evaluate the relationship between network structural properties of heterogeneity, homogeneity, and growth to dynamical properties of robustness [9], adaptivity [7], and symmetry [10].

This proposal is divided into four parts. The first three parts correspond to three hypotheses intended to provide insight to the relationship between network structural properties and dynamical properties: cellular differentiation increases the expected robustness in an organism’s dynamics, cellular differentiation leads to increasing adaptivity, and organism growth by elongation preserves symmetry in the dynamics of an organism. The fourth part will address practical concerns to the experimental approach arising from computational limitations and big data in experimental approaches to studying dynamical systems.

2.1 Preliminary Definitions

Throughout this proposal networks of cellular automata will be referred to as organisms where each node in a network will be a cell of the organism. The application of biological terminology to describe theoretical and synthetic networks is common practice as seen since Von Neumann’s seminal work on cellular automata [40][30]. As noted in the introduction the structure I am focusing on is selected to decrease computational complexity in order to facilitate deeper exploration of state space and exploration of larger organisms.

Structure. Informally, linear cyclic organisms are composed of cells that are directly connected to their two neighboring cells such that the cells form a ring where an initial cell is connected to the
Table 1: Truth table mapping with inputs at time $t$ and resulting output at time $t+1$

<table>
<thead>
<tr>
<th>$s(v_{i-1}, t)$</th>
<th>$s(v_i, t)$</th>
<th>$s(v_{i+1}, t)$</th>
<th>$s(v_i, t + 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>*</td>
<td>0</td>
<td>$b_0$</td>
</tr>
<tr>
<td>0</td>
<td>*</td>
<td>1</td>
<td>$b_1$</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>0</td>
<td>$b_2$</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>1</td>
<td>$b_3$</td>
</tr>
</tbody>
</table>

Table 2: XOR $b_0b_1b_2b_3 = 0110$ truth table mapping

<table>
<thead>
<tr>
<th>$s(v_{i-1}, t)$</th>
<th>$s(v_i, t)$</th>
<th>$s(v_{i+1}, t)$</th>
<th>$s(v_i, t + 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>*</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

next cell and the last cell. Formally, the linear cyclic structure is modeled as an undirected cyclic graph $C = (V, E)$ of size $n$. Vertices are the cells, and are enumerated $V = \{v_0, \ldots, v_{n-1}\}$ where each cell $v_i$ in $V$ is connected in cyclic order to two neighbors such that edges $E = \{(v_i, v_{i+1 \mod n}) | i = 0, \ldots, n-1\}$. The set of possible linear cyclic organisms of size $N$ are a subset of Kaufman’s NK-networks [12] of size $N$ where for each cell the number of inputs is $K = 2$.

Cell state is Boolean, either 0 or 1, and deterministic by fixing a function $f : V \rightarrow F$ that assigns to each cell $v \in V$, a function $f(v)$ from $F = \{g : \{0,1\}\{0,1\} \rightarrow \{0,1\}\}$, the set of all binary Boolean functions. The action of $f$ at a vertex $v_i$ can be thought of as a truth table mapping from the two inputs $K$, $v_i$’s left and right neighbors’ current state, to $v_i$’s state at the next time step. This function differs from the traditional Wolfram model [51][55] in that Wolfram’s model used $K = 3$ and considered the self-input or state of cell $v_i$ as well. The bits $b_0$, $b_1$, $b_2$, $b_3$, as in Table 3, must be either 0 or 1 and the 4-bit binary string $b_0b_1b_2b_3$ is used to name the function $f$.

Note that because there are 2 choices for each of the 4 possible input a cell’s state must be 0 or 1 as a result of the possible the inputs from its neighbors, $|F| = 2^{2^2} = 16$. For example, Table 2 is the XOR function truth table. Note, that the space of possible functions for Wolfram’s model is the more expansive set such that $|F| = 2^{2^3} = 256$ [52]. In Table 3 these rules are defined using Boolean
Table 3: Table of update rules.

Rule numbers expressed as Boolean logic with 2 inputs:
The value of the left neighbor \( v_{i-1} \), and the value of the right neighbor: \( v_{i+1} \)

<table>
<thead>
<tr>
<th>Rule</th>
<th>Boolean Logic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule 0=</td>
<td>0</td>
</tr>
<tr>
<td>Rule 1=</td>
<td>( \neg v_{i-1} \land \neg v_{i+1} )</td>
</tr>
<tr>
<td>Rule 2=</td>
<td>( \neg v_{i-1} \land v_{i+1} )</td>
</tr>
<tr>
<td>Rule 3=</td>
<td>( \neg v_{i-1} )</td>
</tr>
<tr>
<td>Rule 4=</td>
<td>( v_{i-1} \land \neg v_{i+1} )</td>
</tr>
<tr>
<td>Rule 5=</td>
<td>( \neg v_{i+1} )</td>
</tr>
<tr>
<td>Rule 6=</td>
<td>( (v_{i-1} \land \neg v_{i+1}) \lor (\neg v_{i-1} \land v_{i+1}) )</td>
</tr>
<tr>
<td>Rule 7=</td>
<td>( \neg (v_{i-1} \land v_{i+1}) )</td>
</tr>
<tr>
<td>Rule 8=</td>
<td>( v_{i-1} \land v_{i+1} )</td>
</tr>
<tr>
<td>Rule 9=</td>
<td>( (v_{i-1} \land v_{i+1}) \lor (\neg v_{i-1} \land \neg v_{i+1}) )</td>
</tr>
<tr>
<td>Rule 10=</td>
<td>( v_{i+1} )</td>
</tr>
<tr>
<td>Rule 11=</td>
<td>( v_{i+1} \lor (\neg v_{i-1} \land \neg v_{i+1}) )</td>
</tr>
<tr>
<td>Rule 12=</td>
<td>( v_{i-1} )</td>
</tr>
<tr>
<td>Rule 13=</td>
<td>( v_{i-1} \lor (\neg v_{i-1} \land \neg v_{i+1}) )</td>
</tr>
<tr>
<td>Rule 14=</td>
<td>( v_{i-1} \lor v_{i+1} )</td>
</tr>
<tr>
<td>Rule 15=</td>
<td>1</td>
</tr>
</tbody>
</table>

logic and ordered in ascending order according to the Boolean input values. For reference, rule 6 in Table 3 corresponds to the earlier example truth table shown in Table 2.

The organisms in this proposal are **synchronous** where every cellular state is instantaneously determined at each time interval according to:

\[
s(v_i, t + 1) = f(v_i) \left( s(v_{i-1} \mod n), t \right), s(v_{i+1} \mod n), t \right)
\]

for each \( i = 0, \ldots n - 1 \) and \( t \geq 0 \). *Note that results from synchronous models can be transformed and be realized in asynchronous models with synchronization stays within small tolerances locally [39]. This implies that research into synchronous organisms has implications for asynchronous organisms.*

Homogeneity and heterogeneity have previously been studied with respect to connectivity [32] and functions [23]; however, in this proposal they will refer only to functions. If every cell in an organism has the same function assigned to it, then the organism is **homogeneous**. There are 16 unique
homogeneous organisms of any given size. However, if not every cell in an organism has the same function assigned to it, then the organism is heterogeneous. There are $16^n - 16$ heterogeneous organisms of size $n$. Formally, an organism is homogeneous if $|\text{Im}(f)| = 1$; otherwise it is heterogeneous. *There has been research into the distance between functions using hamming distance as a measure of proximity* [27][55]. However, here I simplify and look only at the number of different functions instead of the specificity of what those functions are and their distance from each other.

An **organism state**, as in Figure 3, is the ordered set of its cell states at time $t$. This set of states for example in Figure 1, can be expressed as the sequence of states $v_0v_1v_2v_3v_4 = 00010$.

**Dynamics.** The dynamics of the organism are represented as a directed graph $S$ connecting organism states, as in Figure 1, to their **successor states**, seen in Figure 2, representing the organism’s phase space, in which $(X, Y)$ is an edge if it can be said that whenever the organism is in state $X$ at time $t$, it is necessarily (*absent noise*) in state $Y$ at time $t + 1$. Formally, the dynamics is a directed graph $S = (2^V, D)$ whose vertex set consists of all possible states of the organism (i.e. the power set of $V$), and whose edge set $D$ includes every ordered pair $(X, Y)$ for which $s^+(t) = X \implies s^+(t + 1) = Y$. If an organism state $X$ is connected to organism state $Y$ by a directed edge from $X$ to $Y$, then $Y$ is the successor state of $X$. **Not only have successor functions have been studied for extensively, but their resulting dynamics and dynamical components have been the focus of research problems** [52].
In the dynamics, a **Garden of Eden state** is a state that cannot be reached from any other state by a directed edge in the dynamics space of an organism. In other words, no state in the dynamics of the organism has as its successor a Garden of Eden state. *This dynamical component has drawn attention as part of the problem of reachability: to determine if a state is a Garden of Eden state [47].* Research into Garden of Eden states involves developing algorithms for inverting the successor function or reversibility [26]/[37], and Wuensche and Lesser introduced a reverse algorithm [59].

An **attractor state** is an organism state that occurs in a cycle in the dynamics space of an organism. In other words, if X is an attractor state then there exists some discrete number of time intervals j such that when starting at state X and time t and computing the successor states at times t + 1, . . . , t + j then at the state at time t = t + j will be X. An attractor is the set of all states that are in the same cycle in the dynamics space of an organism and attractor length is the number of states in an attractor. *Here the term attractor encompasses the classes of attractors defined by Wolfram including: fixed point, periodic, and strange [51]. In Hopfield networks attractors are the content-addressable memory [25].*

A **tributary state** is a state that is not in an attractor. When starting at a tributary state and computing the successor states eventually an attractor state will be encountered because the state space has a discrete number of possible states. A tributary is the set of states composing a path of tributary states that lead to the first encountered attractor state when computing forward from a Garden of Eden state. An attractor state can have many tributaries or no tributaries. **Basins of attraction** [56] are formed by these tributaries.

An attractor with its tributaries composes a single **connected component** of the dynamics graph of an organism. The graph of an organism’s dynamics space is made up of these components. Figures 4, 5, 6 show an example decomposition of the size 12 homogeneous XOR organism. The
complete dynamics graph for the homogeneous XOR organism of size 12 has 6 attractors of length 2 (Figure 4), 60 attractors of length 4 (Figure 5), and 4 attractors of length 1 (Figure 6). In the Figures, nodes are labeled with organism state which is composed of the Boolean state of the cells. Directed edges lead from an organism state to its successor state. Black edges connect tributaries to their successor state and blue edges connect attractor states to their successor state [10].

3 Area 1: Cellular Differentiation and Robustness

3.1 Introduction

The evolutionary trend in organisms has been away from homogeneity and towards heterogeneity and complex structures. Since natural selection is a driving force for evolution, there must be some properties of the underlying networks [49] and the dynamics of those networks that are being selected for [13] which drive evolution towards heterogeneity. Consequently, I propose the following hypothesis: Cellular differentiation increases the expected robustness in an organism’s dynamics.

When exploring the dynamics of an organism over infinite time, most of the time will be spend cycling in attractors. As such, in this proposal the focus is only on attractor state perturbations and the effects of those perturbations. Note, this assumption decreases the number of states whose perturbations are computed and explored thereby also minimizing the computational overhead of
exploring both the state space and the perturbations.

3.2 Definitions

A **mutation** is the perturbation of a random cell state in the attractor state. In other words, it is a bit flip of a single bit in the organism state [27]. The organism state that results from a mutation of a single random cell state is a **mutation state**. Wuensche describes graphs of these perturbations as attractor jump-graphs [57]. Mutation states are connected to the originating attractor state by a **mutation edge**. Each organism state of an organism of size n has n possible resulting mutations states and outgoing mutation edges. *This type of perturbation is one of many that have been explored. Perturbations can also be made in the cell function [54] resulting in a change in the wiring or circuitry of the organism or in the timing or synchronicity of the signal [8]. There has been research into the distance between cell states and cell functions, using the hamming distance between two binary values [27][55]. For example, the distance D between a state 00000 and another state 00111 is D = 3 since there are three bits with differing values.*

A mutation state is an organism state that is either in an attractor or connected to an attractor through directed edges in the dynamics graph. If that attractor or connected attractor is the originating attractor then the mutation edge is called **robust**. Otherwise the mutation edge is not robust. *Robustness is the ability to resist external influences [6][11], and in the case of the organism dynamics, robustness is the ability to conserve the dynamical topology [14][22][43]. Research where the perturbation is in the synchronicity of signals applies the definition of robustness to be qualitative changes resulting from perturbations in synchronicity in the update scheme [8].*

**Attractor robustness** is computed by dividing the number of robust mutation edges by the total number of mutation edges of the attractor states. An attractor of length L in the dynamics
space of an organism of size \( n \) has \( n \times L \) mutation edges. Therefore, the robustness of that attractor would be the number of robust mutation edges \( r/n \times L \).

**Organism robustness** \( \rho \) is the average attractor robustness of all the attractors in an organism’s dynamics space. If the total number of attractors in the dynamics space of an organism is \( \alpha \), and the robustness of attractor \( A \) is \( r_A \) then

\[
\rho = (1/\alpha) \sum_{A=1}^{\alpha} r_A.
\]

Note, that organism robustness could be computed as the average number of edges that are robust or \( R/E \) where \( R \) is the number of robust mutation edges and \( E \) is the total number of mutation edges. However, in this proposal, I apply the former definition such that organism robustness is the average attractor robustness for all attractors in an organism’s dynamics space.

### 3.3 Formal Hypothesis 1

Applying the definitions, hypothesis 1 can be posed formally as follows. Let the average robustness of the set of all heterogeneous organisms \( X \) at any fixed size \( N \) be \( Avg(\rho(X(N))) \) and let the average robustness of the set of all homogeneous organisms \( Y \) at the fixed size \( N \) be \( Avg(\rho(Y(N))) \) then for \( N \geq 2 \)

\[
Avg(\rho(X(N))) > Avg(\rho(Y(N))).
\]

### 3.4 Research Plan

The experimental methodology for testing this hypothesis will follow entail 3 components: programming, simulation and data collection, and analysis.

The first stage is to develop a C program that sequentially computes the full organism state
space, the number of attractors, and organism robustness. The program will also generate data files of dynamics graph with mutation edges that can be rendered using graphviz [15]. There is existing software, in particular Discrete Dynamics lab [58] is a tool that has been extensively used to model, simulate, and study dynamical systems. Since the focus in this proposal is to explore properties that differ from the properties studied before, the benefit of building my own code base is that it will be highly customize-able. Developing a program instead of using the existing software further enables optimization for larger organisms (exhaustive search of state space is limited, $n \leq 12$ [55]) and to develop novel algorithms for simulation (e.g. the later discussed simulation by multiple successor steps).

The second stage is to simulate and collect data for analysis. The program will be run on organism sizes that can be fully explored. The program will explore all 16 unique homogeneous organisms for sizes $n = 2 \ldots 16$. For organism sizes $n = 2 \ldots 16$ the space of possible heterogeneous organisms $(n^{16} - 16)$ is too large to fully explore. Therefore the program will also be run on a randomly sampled set of 1000 heterogeneous organisms for sizes $n = 2 \ldots 16$.

Finally, in the third stage data collected from the simulations will be analyzed at each size $n$ for homogeneous organism and for heterogeneous organisms to compare average robustness $\rho$ and average number of attractors $\alpha$. Finally, the data will be used to compare dynamical properties between the sets of homogeneous and heterogeneous organisms.

3.5 Preliminary results

The number of attractors found from fully exploring all 16 homogeneous organisms of sizes $= 2 \ldots 16$ can be divided into two classes: Class 1 organisms where the number of attractors remains bound by some integer $b$ as organism size increases (see figure 1), and Class 2 organisms where the
Figure 7: Homogeneous organisms with bounded (left) and unbounded (right) numbers of attractors. The number of attractors grows unbounded (see figure 2). Given this preliminary result, the proposal is to compare each of these classes of homogeneous organisms separately to the randomly sampled heterogeneous organisms.

4 Area 2: Cellular Differentiation and Adaptivity

4.1 Introduction

Starting from a single cell, every organism is homogeneous. Therefore, this proposal seeks to identify a property that when selected for would lead to the initial cellular differentiation from a homogeneous organism to a "minimally" heterogeneous organism. In biological networks, specialization of cells in heterogeneous organisms helps facilitate their adaptivity to environments. In an organism’s dynamics the corollary to a group of specialized cells is an attractor. Consequently, the informally proposed hypothesis is that cellular differentiation results in an increase in number of attractors and in turn to an increase in adaptivity.
Note that robustness and adaptivity are related properties both in nature and as dynamical properties. Biologically, an organism can either be robust to a stimulus or adapt and evolve; as such, there has been research into how an organism can be robust, while still evolving and adapting [3]. In the previous area I propose exploring the dynamical property of robustness by varying the organism property of homogeneity and heterogeneity while organism size is static. In this area, I propose to focus on the dynamical property of adaptivity with respect to the organism property of homogeneity and heterogeneity during growth.

Related work has involved researching spread [4] and influence [24] in the dynamics of social networks. In particular the standing ovation model [38] and similar consensus problems are of interest because the correlation in the dynamics of the organism is a collapse of adaptivity. However, it is known that searching the dynamics space for specific attractor structures is computationally challenging. Specifically, finding singleton attractors which have attractor length 1 is known to be NP-hard [42][60][1].

4.2 Definitions

As mentioned in the preliminary definitions and introduction, there has been research into the hamming distance between functions in an organism [27][55]. However, here I use the simple count of different functions as a measure of homogeneity and heterogeneity of an organism. Therefore, a homogeneous organism would have 0 different functions or formally $|\text{Im}(f)| = 1$, while a heterogeneous organism would have at least 1 different function or $|\text{Im}(f)| > 1$. Given these two distinct classes of organisms, the border case is of interest. Specifically, an organism with 1 different function, formally $|\text{Im}(f)| = 2$, where every cell determines its state using the same function except for one cell. I refer to this border case class of organism as **minimal heterogeneous**.
Given the earlier definition of an attractor attractor, I can now define **Adaptivity** as the number of attractors \( \alpha(X) \) in the dynamics space of an organism. *Attractor count is a measure of interest that has been extensively studied [28][51][17]. I use the term adaptivity due to its relationship to evolvability, defined by Aldana et al. to mean that an organism can acquire new functions and adapt to new environments [3]. Further, if attractor count is considered memory [25], then organisms with greater memory are more likely to be adaptive to new environments that require shifting memory.*

### 4.3 Formal Hypothesis 2

Applying the definitions, hypothesis 2 can be posed formally as follows. Let the robustness of a single homogeneous organism \( Y \) at size \( N \) be \( \rho(Y_N) \) and the robustness of a single minimally heterogeneous organism \( X \) at size \( N \) be \( \rho(X_N) \). Then as the organisms \( X \) and \( Y \) increase in size from \( N = 2 \ldots \infty \),

\[
\exists i \mid i \geq 1 \text{ and } \rho(X_{N+i}) \geq \rho(Y_{N+i}).
\]

### 4.4 Research Plan

The experimental methodology for testing this hypothesis will follow entail 3 components: programming, simulation and data collection, and analysis.

The first stage will be to extend the C program from area 1 that sequentially explores the organism state space to compute the number of attractors. The program will also generate data files of dynamics graph with mutation edges that can be rendered using graphviz [15] and charted using gnuplot [36]. In order to explore larger networks, I will expand the program to sampling the state space and estimate attractor counts.

The second stage will simulate and collect data for analysis. The program will be run on organism sizes that can be fully explored and sample the state space of for larger sizes. The program will
explore the homogeneous XOR organisms for sizes $n \geq 2 \ldots 16$. The program will also explore every possible minimally heterogeneous organisms of the same size as the homogeneous XOR organisms simulated.

In the third stage, analysis of the resulting number of attractors discovered for the homogeneous XOR organism will be compared to the number of attractors found for each individual minimally heterogeneous organism at every size organism whose dynamics space is fully explored and every at every size organism whose dynamics space is sampled.

### 4.5 Preliminary results

The initial results of the dynamics space of homogeneous XOR organism exhibit a pattern of collapse in adaptivity at sizes which are a power of 2 (see figures 8,9,10 and the chart in figure 11). Therefore, for sizes $n > 16$ simulations will focus on organism sizes which are powers of 2.
5 Area 3: Symmetry and Growth

5.1 Introduction

In areas 2, adaptivity is explored with the assumption has been that organisms will grow incrementally by a single cell from size $n$ to $n+1$. However, growth rate is a network property that will have its own relationship to dynamical properties. If growth rate impacts the dynamical properties then natural selection would prefer a rate that increased or at least retained the desirable dynamical properties. In area 2, the preliminary results show that for the homogeneous XOR organism adaptivity is adversely impacted by incremental growth towards sizes that are powers of 2.

To introduce the idea of how organism growth could preserve symmetry, let’s look at the example of a starfish. The starfish has clear rotational or radial symmetry. How can a starfish grow while preserving rotational symmetry? There is the possibility to remain the same size while growing another arm. And, there is the possibility to simply grow larger, where each part of the organism is elongated. Both of these preserve its rotational symmetry. This proposal will look at both possible
symmetry preserving patterns of growth.

In this area, I propose to explore how dynamical properties of symmetry can be preserved during growth [5] in order to understand which patterns of growth might be selected for in organisms. Previous work in symmetry of cellular automata has been explored by focusing on symmetry of the functions [29][52], time symmetry referring to the successor function computing forwards and backwards [16], and an organism’s dynamic state pattern symmetry resulting from an initial state [48].

In area 1 and 2, I take a perspective of the organism as a whole and observe the global dynamics. Since the focus in this area is on symmetry of the organism, in this area I will focus on the cell dynamics as they relate to the global organism dynamics. As before in the assumption posited in area 1, I will focus only on attractor states since from any initial state after an infinite number of discrete time intervals, the organism will spend the majority of its time in attractor states. Observing each individual cell of the organism as the organism orbits in an attractor, the individual cells flip on and off in a period that divides the length of the attractor. Now, rather than the instantaneous state of an organism in this area I consider the state of each cell over unbounded time by examining the periodicity of its binary oscillations as the organism orbits in an attractor. The periodic transition from 0 to 1 of the cell states may result in symmetric pattern of state changes of cells in the organism cells during the orbit of an attractor. In this area, I propose to explore this symmetry of cell periodicity in an attractor as a dynamical property in relation to organism growth [10].

The focus in this area is symmetry. Here we will focus on rotational symmetry. Using the example of a starfish that has clear rotational symmetry, I propose to look at two forms of growth which could preserve that rotational symmetry. The first method of growth is cell by cell elongation of
each "arm" simultaneously. The second method of growth is the reduplication or growth of an entire new "arm" all at once. Taking inspiration from the biological organism, a starfish grows through elongation rather than adding another arm: informally the proposed hypothesis is that growth by elongation is more likely to preserve symmetry in dynamical properties than growth by reduplication.

5.2 Definitions

Color is determined by the periodicity of a cell’s state as the organism traverses an attractor. Color is assigned to an individual cell. Note that periodicity of cell state is dependent upon the attractor being traversed; as such, for every attractor, each cell can have a different periodicity and coloring. Because the periodicity has an upper bound of the length \( l \) of the attractor and a lower bound of 1 the possible colors for a cell will vary dependent upon the length of an attractor. As defined earlier, the organisms are cyclic. Therefore, for each attractor in the dynamics of an organism, as each cell is colored with respect to its periodicity, the organism itself takes on a pattern of corresponding colors. In Figure 12 on the left, the homogeneous XOR organism of size 12 orbits an attractor of length 4 and the. In Figure 12 in the middle, each of the 12 cells’ states are expanded outwards towards infinity as the organism orbits in that attractor of length 4. In Figure 12 on the right, the periodicity of the cells of the homogeneous XOR organism of size 12 can be seen as the organism orbits an attractor of length 4, and corresponding colors are assigned to the organism.

The pattern of colors arranged in a ring can result in symmetry. In particular, the coloring could be rotationally symmetry. Rotational symmetry in the case of coloring means that after a certain amount of rotation the coloring would be the same. Segment size \( l \) is the number of cells which when rotated result in the same coloring of an organism. Since rotation by segment size \( l \) results in
the same coloring, segment size divides organism size. **Foldedness** $k$ is the number of times segment size divides into organism size $k = N/l$. **Trivial rotational symmetry** is defined as a segment size $l = 1$ and foldedness $k = n$ because it is the case where all cells have the same periodicity resulting in an organism where all cells have the same color. This proposal is only interested in non-trivial cases of symmetry.

In this area, **Chromatic symmetry** is defined to be non-trivial rotational symmetry in the coloring of the organism cells as it traverses an attractor. A rotationally symmetric organism has $k$ segments of size $l \geq 1$ where each segment has the same pattern of coloring. In Figure 12 (on the right), the coloring of the homogeneous XOR organism of size $n = 12$ as it orbits the attractor of length 4 shown in Figure 12 (on the left) exhibits chromatic symmetry with segment size $l = 6$ and foldedness $k = 2$. This means that any the organism coloring can be rotated from any initial position by a segment of 6 cells to the right or left and the resulting organism coloring will be the same as the initial organism coloring [10].

![Figure 12: Attractor of size 4 in dynamics of homogeneous XOR organism size $n=12$.](image)

**Growth.** Growth patterns have been extensively studied [2], for example for modeling morphogenesis in plants using L-systems [31]. Growth in areas 1 and 2 focused on incremental organism
growth by a single cell from size $n$ to size $n+1$. **Reduplication** is growth by incrementally increasing the number of segments from $k$ to $k+1$. This increase in segments means that the organism growth is from size $n$ to size $n+l$. **Elongation** is growth by incrementally increasing the segment length from $l$ to $l+1$. This increase in segment length means that every segment increases in length by 1 and in turn the organism growth is from size $n$ to size $n+k$. In Figure 13 the example of a homogeneous XOR organism of size $n = 10$ in the orbit of an attractor of length 6 in its dynamics graph **Elongation** gives a starting point from which the two forms of growth can be explored. Note, that in the example the chromatic symmetry has segment size $l = 5$ and foldedness $k = 2$. Figure 14 represents growth that occurs by segment elongation from a size $n = 10$ organism to a size $n = 12$ organism, where the growth is an increased by increasing the segment size $l = 5$ to $l = 6$. Figure 15 represents growth that occurs by segment reduplication from a size $n = 10$ organism to a size $n = 15$ organism, where the growth is an increased by increasing the foldedness $k = 2$ to $k = 3$. Note that in Figure 15 the segment size remains $l = 5$ [10].

### 5.3 Formal Hypothesis 3

Applying the definitions, hypothesis 3 can be posed formally as follows. Let $A$ be a homogeneous XOR organism of size $N$ which has nontrivial rotational chromatic symmetry of segment length $L_{A_i}$ as the organism $A$ traverses attractor $i$. Let $\text{Probability}(\text{sym}(R))$ be the probability that there exists a homogeneous organism $R$ of size $N + L_{A_i}$ which has nontrivial rotational chromatic symmetry of segment length $L_{A_i}$ and let $\text{Probability}(\text{sym}(E))$ be the probability that there exists a homogeneous organism $E$ of size $N + (N/L_{A_i})$ which has nontrivial rotational chromatic symmetry
Figure 13: Coloring in orbit of attractor length 6 in dynamics of homogeneous XOR organism of size $n=10$ of segment length $L_{A_1} + 1$, then:

$$\text{Probability}(\text{sym}(E)) > \text{Probability}(\text{sym}(R))$$

5.4 Research Plan

The first stage will further extend the C program from area 2 to explore the dynamics of organisms compute the chromatic symmetry of the organism across its attractors.

Because symmetry will be tested at larger organism sizes, the program will have to be able to sample the dynamics space and handle larger numbers. Therefore, the program will be written
Figure 14: Coloring in orbit of attractor length 4 in dynamics of homogeneous XOR organism of size n=12

Figure 15: Coloring in orbit of attractor length 3 in dynamics of homogeneous XOR organism of size n=15

using the GNU MP big number library.

The second stage will simulate the organism dynamics space and collect data. The program will explore the dynamics space of all homogeneous organisms and identify their possible colorings. Next, selecting from the homogeneous organisms a subset that exhibit non-trivial chromatic symmetry, the
program will be run focusing on sizes which correspond to patterns of growth that entail elongation and reduplication.

In the third stage analyzing the data collected will require comparing the probability of preserving symmetry by each growth pattern (elongation and re-duplication) in the sampled dynamics across the simulated organism sizes.

5.5 Preliminary results

Attractor length bounds periodicity; therefore, organisms with attractor lengths of 1 are restricted to the trivial case of chromatic symmetry. By sampling the dynamics of all the homogeneous organism from 1000 random start states for sizes $n = 2 \ldots 50$, preliminary data collection of attractor lengths can be used to identify organisms whose coloring will be restricted to the trivial case of chromatic symmetry. Figure 16 shows that only a subset of homogeneous organisms have increasing attractor lengths and are therefore more likely to exhibit non-trivial chromatic symmetry.

![Attractor Lengths at Organism Size](image)

Figure 16: Attractor lengths of homogeneous organisms of sizes $N = 2 \ldots 50$
The next step is to determine the relative frequency of the chromatic symmetry for these organisms. The simulation and data collection will focus on only those homogeneous organisms that have a non-zero relative frequency of chromatic symmetry.

6 Obstacles

6.1 Introduction

The experimental approach I propose taking faces obstacles that accompany big data. As the size of a network grows linearly, the state space grows exponentially. Consequently, even at relatively small network sizes, exhaustive search of the state space becomes computationally intractable. Being unable to exhaustively search the state space results in the inability to generate the complete dynamics including identifying attractors, robustness, and chromatic symmetry. As the size of a network grows linearly, the possible organisms grow exponentially. Further, the time and computational limitations of simulating each of these networks becomes computationally intractable even at small network sizes.

In developing the programs, I propose the following three approaches to mitigating big data obstacles: First, using algorithms which make use of patterns to increase efficiency in memory and search by computing, storing, and comparing the successor state at multiplicative or exponential time increments instead of single increments. Second, using sampling techniques to estimate properties of the dynamics and the organism. Third, using distributed processing to explore the state space and sets of possible organisms in parallel
6.2 Candidate Approaches

6.2.1 Powers of 2

Successor states are computed sequentially. Starting from a state, every next state is compared against every previously viewed successor state of the initial state. Attractors are determined once a previously seen state is seen again.

I propose to implement a more efficient algorithm than the trivial sequential successor computation with every state computed being compared with every previously computed successor state. If P is the number of states from an initial state to the first state in an attractor + the length of the attractor L, then the memory required is P states and the number of comparisons to discover an attractor is \( \frac{P^2}{2} \).

It is possible to lower the memory required for detecting an attractor by only storing a state if it is at a successor step that is a power of 2. A side effect of the lower memory fewer results in fewer comparison operations \( \log_2 P \) where \( p \) is the path explored so to the current successor state) to check if the state has been visited before for each successor computation since there are fewer states being stored. This algorithm requires \( P(\log_2 P) \) comparisons and \( \log_2 P \) memory.

For XOR organisms, I propose to work on developing an algorithm for simulating more than one successor state at a time with a goal of computing forward in powers of 2 successor states at once. The basis for this algorithm is the pattern observed in the preliminary results for area 2. The core component of this algorithm relies on a proof that organisms of size \( n \) is a power of 2 in any state will collapse to a single attractor after a set number \( C \) of successor computations. With a proof of what the state of the organism will be after a set number of successor computations, then the algorithm can be modified such that the incremental successor computations to operate in \( C \) steps.
instead of steps of size 1. Additional steps in the algorithm will be necessary to compute states that need to be computed at intervals that differ from exact powers of $C$.

### 6.2.2 Sampling techniques

Random sampling is accomplished by picking a random state between 0 and $\left(2^N\right) - 1$, instead of sequentially starting at each subsequent state in the state space and tracing it to an attractor. Attractors are determined in the same way however, by computing the successor states from the initial state until a previously seen state is seen again. For random sampling, I select a set $D$ of random states and trace them until I identify the attractor they fall into to.

For capture recapture, the program will randomly sample a number of states $D$, twice as $2$ separate intervals. The set of attractors identified from the first set of $D$ starting states are labeled as Capture 1. The set of attractors identified from the second set of $D$ starting states are labeled as Capture 2. And, the set of attractors seen in both Capture 1 and Capture 2 are labeled as Recapture.

Capture Recapture has limitations which occur as the size of the state space and the number of attractors increase to such a degree that the sampled Capture 1 and Capture 2 have no overlap. Further Capture Recapture assumes that the size and distribution of the basin of the attractors are the same.

I propose developing a capture recapture technique which takes into account the varying attractor sizes to estimate attractor counts of differing sized attractors and the total number of attractors in an organisms dynamics space.
6.2.3 Distributed processing

I propose to use a message passing interface to distribute sets of initial states to trace into attractors to machines in a cluster. I will implement the distributed processing in the forensics lab cluster at John Jay College of The City University of New York. Data computed by the simulation will then be returned to a centralized server where it can be compressed and bulk inserted to a centralized database that keeps track of the states explored. The bottleneck occurs at the centralized database. Future work would be to decentralize the database to improve access time to the database for reading and writing.

7 Summary

This proposal studies the impact of Darwinian selection on cellular differentiation and organism growth. The proposal seeks a greater understanding of natural selection through exploring the relationship of cellular differentiation to robustness and adaptivity and the relationship of growth to chromatic symmetry. This proposal will follow an experimental approach to testing three hypotheses:

- Cellular differentiation increases the expected robustness in an organism’s dynamics.

- Cellular differentiation leads to increasing adaptivity

- Organism growth by elongation preserves symmetry in the dynamics of an organism.

The proposed experimental approach faces obstacles related to big data. I propose three approaches to addressing these obstacles: algorithms, sampling and distributed processing.
8 Timeline

<table>
<thead>
<tr>
<th>Date</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 2011</td>
<td>Research Plan 1</td>
</tr>
<tr>
<td>November 2014</td>
<td>Research Plan 2</td>
</tr>
<tr>
<td>November 2015</td>
<td>Research Plan 3</td>
</tr>
<tr>
<td>May 2016</td>
<td>Dissertation Proposal</td>
</tr>
<tr>
<td>June-September 2016</td>
<td>Thesis writing</td>
</tr>
<tr>
<td>September 2016</td>
<td>Thesis defense</td>
</tr>
</tbody>
</table>

Table 4: Schedule of Tasks

Table 4 describes the time-line to complete the dissertation.
References


