

# Thesis Proposal: Modeling Pathways of Cell Differentiation in Genetic Regulatory Networks With Random Boolean Networks

Sheldon Dealy [sdealy@cs.unm.edu](mailto:sdealy@cs.unm.edu)

## Introduction

Genetic regulatory networks are collections of genes that interact together to modify or change one another's behavior. Small modifications to a genetic regulatory network, such as the transient perturbations of a gene's activity, can force a cell to differentiate from one type to another. Differentiation of cell types also occurs as a natural process. Notably, embryonic stem cells differentiate into a variety of cell types.

Understanding genetic regulatory networks, more specifically how a cell type can differentiate, is a puzzle that has yet to be solved. For example, HL60 leukemic cells can be induced to differentiate to a benign state after the application of dimethylsulphoxide.[3] It is thought that with the proper computational model, we can understand the biology better. Through better understanding, predictions about the biology model can be made. Random boolean networks may be able to fill in some of the pieces.

What happens to the cell during differentiation? Once a gene is perturbed, what kinds of predictions can be made about changes occurring to the genetic regulatory network? My research hopes to gain insights into the pathways of cell differentiation using randomly generated boolean networks.

## Boolean Networks

Boolean networks[1] can be thought of as a directed graph where each node evaluates to a binary value of one or zero. The state of the network is a snapshot of the node values taken as a whole. Generating subsequent states in the network is determined by boolean functions.

Boolean functions, one for each node, take as inputs the values of nodes sharing an input edge with the given node. The output value of a given function depends upon the configuration

of boolean values in the input nodes. Random boolean networks are networks of  $k$  inputs and  $n$  nodes whose connecting edges and boolean functions are generated randomly.

When the sequence of states in a boolean network repeats itself, it is called a state-cycle. State cycles in boolean networks are attractors and these attractors can be thought of as representing cell types. Each state in a boolean network state cycle is an analogue for a state produced in a genetic regulatory network. Because there can only be a finite number of states in a boolean network, eventually a sequence of states will be repeated. Thus, by the Pigeon-hole principle, a state cycle will always exist in a boolean network.

The strength of boolean networks is their simplicity. Boolean functions generate transient states which the network passes through before the network reaches another state cycle. The transient states represent a model for cell differentiation as the cell changes from one cell type to another. Like real genetic regulatory networks, I have found that most of the modeled transients are also homeostatic. Homeostatic behavior in genetic regulatory networks is desirable because it contributes to the robustness of the cell types.

The problem with boolean networks is that it is difficult to model the continuous nature of real genetic regulatory networks with discrete values. Boolean networks are synchronous in that each discrete step of the entire network is generated with the boolean functions using the network state from the previous step. The boolean network acts as if there is one central clock directing all activity.

A genetic regulatory network has no discrete clock. Genes in cells are not in lock-step with each other. Likewise cells which are differentiating are not synchronized with other, similar cells. A genetic regulatory network is difficult to label as being precisely in one state or another.

## Research in Progress

Over the last year, I have put together a model of cell differentiation using randomly generated deterministic boolean networks. The model currently generates a network which uses boolean functions to establish state cycles. The model sequentially perturbs nodes in each state to create pathways of differentiation and more state cycles. From the perturbed networks, statistics are harvested about behavior on the transient pathways. With the addition of aggregation and data mining software utilities, I am beginning production-type runs targeted to answer some specific questions.

1. What ratio of perturbations will return to the same cell type (homeostasis), as opposed to changing cell type?
2. What proportion of transients fuse with other transients on the pathway of differentiation?
3. How does perturbing genes in a model regulatory network affect the rest of the genes?
4. What proportion of transients have distinct pathways?

5. More generally, what properties do transients have when differentiating from one cell type to another unique cell type as opposed to those that return homeostatically to the perturbed cell type?
6. How to create a credible mapping from boolean networks to genetic regulatory networks?

## Current Challenges

### Approximating the Biology

I have been trying different ways of evaluating the pathways of differentiation to more approximate the biology model. With professor Kauffman and his colleagues, I am putting together promising mechanisms to map transient pathways in boolean networks to genetic regulatory networks.

One idea is to use the total number of gene changes as a stand-in for transient length by tracking the percentage of time each gene is in a state different from the unperturbed state.

Another kind of mapping I propose is by tracking the *incidence* of gene change over a transient. For instance, a histogram of gene change can be made by counting the number of genes that change state zero times, then the number of genes that change once, then the number of genes that change twice, etc.

### Transient Fusion and Cell Differentiation

It has been found that the Hamming distance between differentiating cells that share the same destination cell type initially grow farther apart and then become more alike as they approach the destination cell type.[2] I have found that transients in random boolean networks which fuse share this behavior. A transient which fuses means that at some state along the transient, a pair of transients share the same final sequence of states. The graph of Hamming distances of states in these transient pairs approximates the biology model before they fuse.

## References

## References

- [1] Kauffman, Stuart(1995). At Home in the Universe: The Search for Laws of Self-Organization and Complexity, Oxford University Press, New York, NY.

- [2] Sui Huang and Don Ingber(2004) Cell Fates as Attractor States. Personal Communication. Harvard Children's Hospital. May 2004.
- [3] To be found. This reference exists, but has been misplaced.