

A Further Analysis of Genetic Nets in Bioinformatics

Introduction

This report includes an analysis of Stuart Kaufman's computational model for dynamic gene networks. In this network, genes are represented as either "on" or "off," the regulatory interactions are defined by boolean functions, states are updated iteratively, and the internal dynamics of the network are derived from the different nodes interacting with each other over time. In this model, a gene network is represented as a graph of nodes with evaluative functions based on the aforementioned boolean functions. Kaufman's model for gene nets works as a strong manner through which to understand the behavior of complex biological systems. The guiding questions of this project are as follows:

1. What happens if you try multiple trials with the same networks, but starting with different randomly-assigned start states?
2. What happens if two well connected subnetworks have limited connection between each other?
3. Do almost identical networks behave similarly -- for example if only one gene has a different Boolean function?

My particular work pertains to Question (2), wherein I tried to grasp how connectedness between internal subnetworks impacts the behavior of the overall network.

Methods

Question 1

In Lizzie's approach to Question 1, her strategy was to conduct 10 iterations of evolution on a genetic net of size 10, randomly assigning a new start state with each iteration. Implementing and answering this question covers her portion of the project, the results of which are listed and discussed below.

Question 2

In order to approach my portion of the project, Question 2, I defined a default Kaufman network that functioned as a control and a modified network that fit the terms of the question. The modified network uses inputs of *size*, which specifies the number of nodes in the network, and *overlap*, which specifies the connectedness. The modified gene net has 2 subnetworks, consisting of *nodes*[0:*size*/2] and *nodes*[*size*/2:]. The nodes in each subnetwork are assigned

neighbors within their own subnetwork. Afterwards, I use overlapping nodes in order to distinguish connectedness among the subnetworks. In each subnetwork, I randomly set aside a number of nodes corresponding to the *overlap* input. For these nodes, I assign their neighbor to be a random node in the other subnetwork. The maximum possible *overlap* is $size/2$, where every node in each subnetwork has 1 neighbor that is in the other subnetwork, and another neighbor within its own subnetwork. I set this maximum because at and beyond this point, there are no longer distinguishable subnetworks.

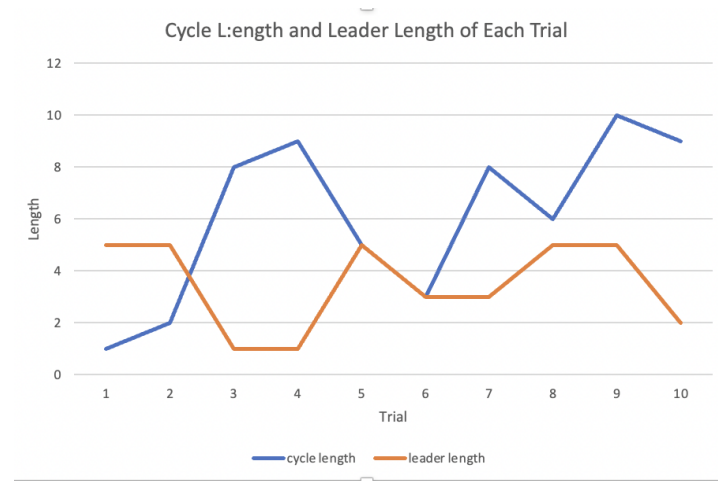
With the networks defined, I initialized 1 default Kaufman network of size 64 nodes as the control; this is a control in that it does not have manually defined subnetworks. Afterwards, I initialized 33 test networks, each of size 64 nodes, each corresponding to the range of possible overlapping nodes spanning from 0 to 32, inclusive. In order to judge the behavior of the subnetworks, I performed 200 trials of evolution on each network and measured the mean cycle length, median cycle length, and mean path length to achieve a cycle. Here, a cycle refers to the number of consecutive values that remain the same across a past state of a genetic net and following new states after the net has evolved. Thus, the cycle length is the number of values that cycle between states. The path length refers to the number of nodes needed in order to reach the cycle. I took these values by simulating 200 rounds of evolution on each network and tracking the changes in cycle length and path length.

Question 3

In Lily's approach to Question 3, she tests the behavior of the genetic nets by creating networks ranging from size 10 to 49. For each size of network, she uses identical networks to run trials. Among those identical networks, she implements between 1 and $network_size/10$ changes in order to test for the impact of change on the behavior of near identical networks. The results of her experiment are discussed below.

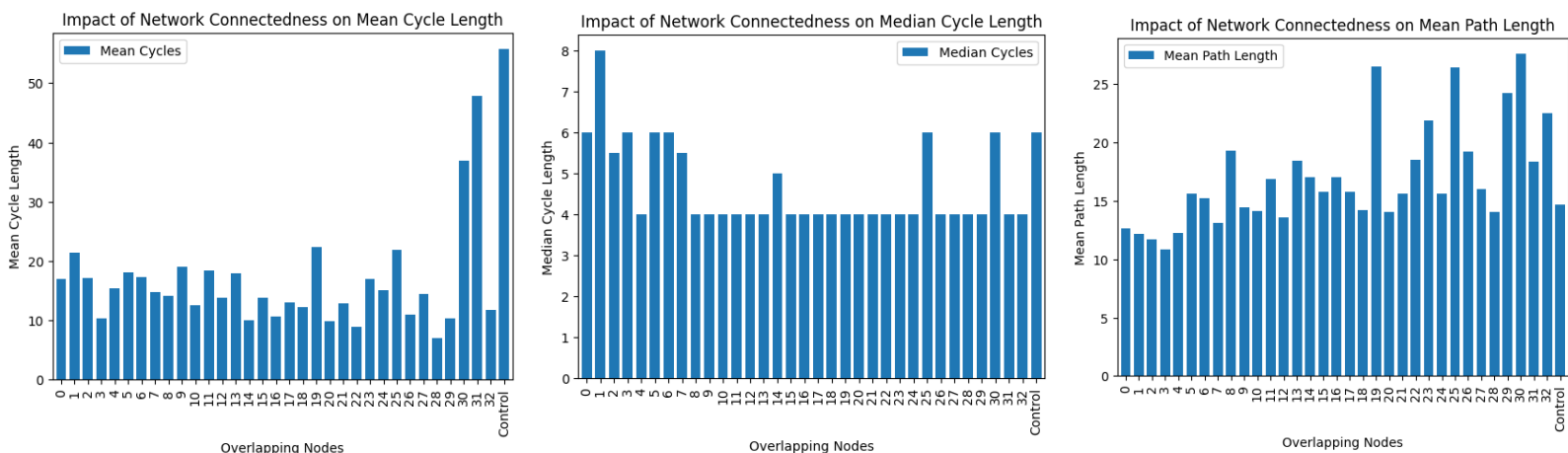
Results

Question 1



The above graph shows that results vary greatly across trials pertaining to randomly assigned start states for a given genetic network. This seems to suggest a strong impact between the start state and the dynamics of a genetic net; in other words, a randomly assigned start state has a drastic impact on the evolution of a genetic net. This goes against my intuition and takeaways from our in-class discussion, however. My initial hypothesis was that the start state of a genetic net can only have so strong an impact on the dynamics across different rounds of evolution, so the cycle length should not vary too much. However, this data seems to prove that theory wrong. In the future, I would be interested in using larger sizes of genetic nets and more trials to test the amount of variance across randomly assigned start states.

Question 2 (Figures 2a, 2b, and 2c respectively)



In Figure 2a, the plot measures the mean length of a cycle achieved across 200 rounds of evolution simulation. Based on the results of the experiment, there seems to be an inconclusive relationship between connectivity and average length of value cycles in evolutionary states. However, I would speculate a vague decrease among average cycle lengths as connectivity increases, barring values 30 and 31 as outliers. This seems to indicate that as connectivity increases, the length of the cycles achieved throughout evolution decreases. I will use this speculation in a later point.

In Figure 2b, the plot measures median cycle length across 200 rounds of evolution simulation. Similar to the previous plot, this data seems largely inconclusive. While one could speculate a vague trend of decreasing median cycle length based on connectivity, I would argue that the trend is not enough to be worth noting.

In figure 2c, the plot measures the average number of nodes necessary to reach the leading node in a cycle, tracked across 200 simulations of evolution. We can see a noticeable increase in path length based on connectivity, demonstrating that the number of nodes behind the leading node in a cycle increases as the number of overlapping nodes increases.

Based on this data, we can see that genetic nets have more volatile and diverse internal dynamics as their nodes are better connected. They begin to find less order over multiple iterations of evolution as the network and subnetworks get better connected; this is demonstrated by the fact that the length of cycles decrease, demonstrating less consistency in patterns among states. This is also backed by the path length increasing, demonstrating a longer amount of time in order to find the leading node in a cycle. The demonstration of fewer patterns suggests more diversity in cycles among better connected networks. In the future, it would be interesting to test for cycles in the subnetworks themselves and see the ways subnetworks develop their own patterns as a gene net experiences rounds of evolution.

Question 3

NetSize:	NumBoolChanges:	MCL:	MLL:	MedCL:	MedLL:
10	1	2.05	0.96	2	1
11	1	2.4	1.35	1	1
12	1	4.4	1.65	4	1
13	1	2.67	1.19	2	1
14	1	2.33	1.23	1	0
15	1	2.63	2.52	2	1
16	1	3.17	1.43	1	1
17	1	2.84	1.37	2	1
18	1	2.99	1.16	2	0
19	1	3.25	2.42	2	1
20	1	4.85	1.65	4	0
20	2	3.25	2.76	2	2
21	1	3.18	2.18	1	1
21	2	2.08	3.02	1	2
22	1	3.36	1.56	4	1
22	2	3.14	2.41	2	2
23	1	2.35	1.94	2	0
23	2	2.24	2.32	2	2
24	1	6.15	1.63	2	1
24	2	4.03	3.02	1	2
25	1	6.14	3.04	6	0
25	2	3.95	2.23	2	2
26	1	3.41	1.53	1	1
26	2	6.36	3.85	4	2
...					
49	1	9.31	4.3	5	1
49	2	12.95	4.63	2	2
49	3	3.78	4.19	2	3
49	4	3.96	4.93	2	3

While we do not yet have a chart for Question 3, we do have a table from the Jupyter notebook offered above. Although the results above the ellipses seem inconclusive, we could make an inference based on the 4 values at the bottom of the table, although it would be a poorly backed conclusion. For 4 identical networks of size 49, we can see both mean and median cycle length decrease as the number of changes implemented increases. In addition, mean and median leader length increase as the number of changes increases. Along the same lines as my analysis in Question 2, this would suggest that as we implement more changes to the network, we see more diversity in patterns and less consistency in evolution. However, note that this is poorly backed without an overall view of the data. It would be interesting in the future to add a control of 0 changes to see how the other networks, which implement changes, compare to the control

References

[Not Applicable]

Appendix (Individual Code • Shared Code)

See corresponding file attachments

For my on Question 2, you can check corresponding file attachments or [follow this link](#).