Kelle Dhein

SES Rough Draft

*C. Elegans* GRN Network

[This course really took me out of my comfort zone, and I think that will be evident as you read my rough draft. I do not have much experience reading/writing physics articles or talking about network models, so any feedback, even (especially) basic-seeming stuff, will be much appreciated]

**Introduction**

Scientists have been analyzing the informational properties of living systems

to gain a better understanding of what distinguishes those systems from other physical systems. The guiding assumption of such a research program is that living systems process information in a way that other physical systems do not. An interesting and closely related assumption is that the way living systems process information is intrinsically tied to the set of novel phenomena surrounding living systems. If those assumptions are true, then there may be signatures in the informational dynamics of living systems that distinguish those systems from other physical systems.

In what follows, we describe a series of experiments meant to gather data about the informational dynamics of single biological system in the hopes that the data will be useful in future attempts to test the theory that the informational dynamics of living systems distinguish them from other physical systems. More specifically, we provide quantitative measurements of the informational dynamics of the gene and protein network regulating the early embryonic cell cycle of the nematode *Caenorhabditis elegans*. In measuring those dynamics, we follow previous researchers [7, 23-26, and Kim et al] in distinguishing abstract informational patterns from ‘informational architecture.’

A measure of information must satisfy two conditions to be a measure of the informational architecture of a living system: (1) the measure must correspond to a specific physically instantiated object or process in a living system, and (2) that physically instantiated object or process must play some causal role in the functioning of the living system. We do not claim that information comprising the informational architecture of a living system is itself an essential agent in the functioning of that living system. Rather, we make the weaker claim that information comprising the informational architecture of a living system is processed in the functioning of a living system.

In this paper, we focus our informational analyses on the network of proteins and genes regulating the early cell divisions of a developing *C. Elegans* embryo. Simulating the dynamics of that system as a network graph, we measure the information being transmitted between genes and proteins as the system functions. To measure that information, we utilize two informational measures, Schreiber’s transfer entropy (TE) and active information (AI). Both metrics are dynamic in that they determine how much our uncertainty about the present state of a system is reduced by considering past states of that system. [Should I describe the difference between TE and AI? Include the equations?]

In addition to simulating and analyzing the biologically functional path of the *C. Elegans* gene regulatory network, we also explored the state space of our network by altering our simulation to analyze non-biologically functional paths of the system. By comparing the way the gene/protein network regulating the *C. Elegans* early embryonic cell cycle processes information when it is functioning properly with the way the network processes information when it is not functioning properly, we sought to determine whether any signatures in the informational data distinguished the biologically functional pathway of the network.

**Model Description**

Multi-cellular organisms originate from a single cell, which divides itself into two daughter cells. Those daughter cells go onto divide themselves into daughter cells so that the organism gains cells at an exponential rate. While dividing, a cell goes through a series of phases. Though those phases can vary across organisms and across the developmental timeline of a single organism, there are generally four phases in the division of cell: gap 1, synthesis, gap 2, and mitosis [6]. Gap 1 and gap 2 are transition phases that help to correct potential errors in the cell division process and ensure that the cell enters the synthesis and mitosis phases correctly. In the synthesis stage, the nuclear DNA of the dividing cell replicates itself so that both daughter cells receive the genetic information of the dividing cell. Finally, in the mitosis phase, the dividing cell physically splits into two distinct daughter cells.

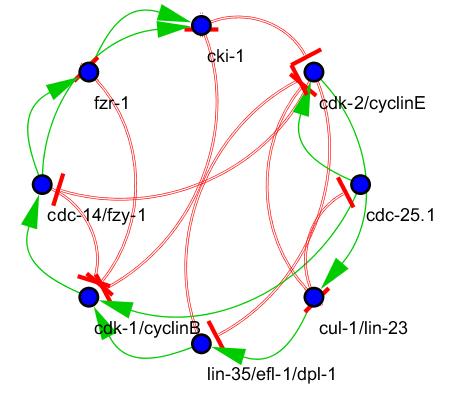
In the later stages of *C. Elegans* development, its dividing cells go through the four phases just mentioned [6]. In the very beginning stages of its development, however, *C. Elegans* cells skip the gap 1 and gap 2 phases and divide rapidly as successive generations of cells alternate between the synthesis and mitosis phases.

Based on previous research [main paper, 2, 10, 11, 12, 13, 14, 16, 17, 18], we developed a Boolean network model of the genes and proteins responsible for regulating cell division in early embryonic *C. Elegans*. In other words, our model simulates the dynamics of a set of proteins and genes as they regulate the transition of a developing C. Elegan’s cell from synthesis to mitosis. Boolean network models have also been used to simulate the cell cycles of budding yeast [2], fission yeast [1], mammals [4], and mammalian cancer cells [5].

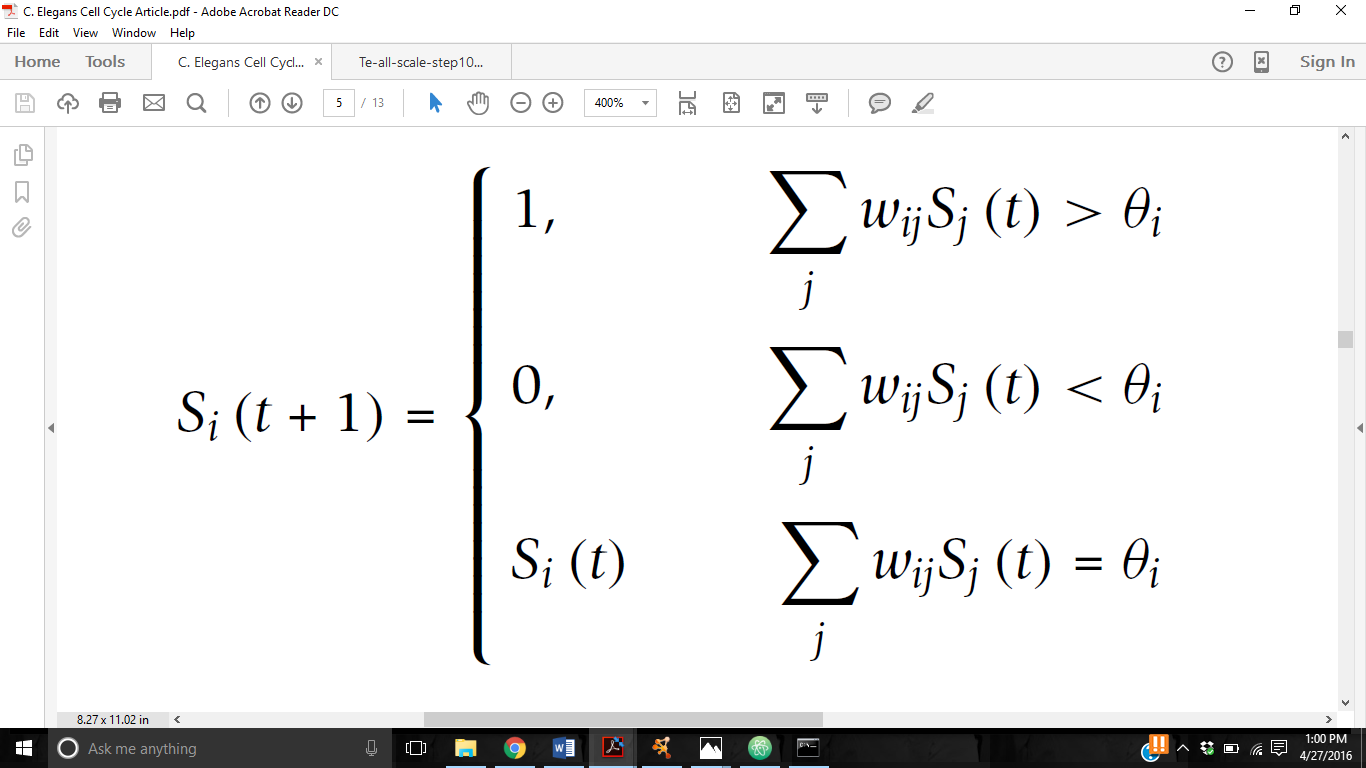
In creating our Boolean network model of cell division in early embryonic *C. Elegans*, we built on the model created by Huang et al’s []. Our model is a network model in that it represents proteins, genes, and groups of functionally unified proteins and genes as nodes. Our network model is Boolean in that each of those nodes can have one of two mutually exclusive values, 1 or 0 (on or off, respectively). The edges between nodes are meant to capture the ways in which the genes, proteins, and functionally unified groups of genes and proteins interact with each other as they regulate the early embryonic cell cycle of *C. Elegans*. [Go further into how previous experimental biologists figured out how all the genes and proteins interact during this process?]

The network topology is visualized below in Figure 1. The edges are directed and have a weight value of 1 or -1. Edges with weight of -1 are red and edges with a weight of 1 are green. The network topology consists of eight nodes and twenty-one edges.

Figure 1

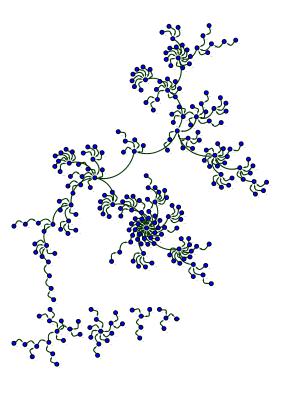


The Boolean value of any target node (on/off) is determined at each time step by the edge values of all the upstream source nodes that were active in the previous time step. More formally, each node’s Boolean value is determined by the following function:



Where *S*(subscript)*i* represents the state of target node *i* at some time *t*. *j* represents a source node upstream of *i*, and *w*(subscript)*i*(subscript)*j* represents the weight value of the edge originating from source node *j* and terminating in target node *i*. The threshold for target node *i* is denoted by θ(subscript)*i*. All nodes in our network have a threshold value of 0. At each time step, every node in our network is updated simultaneously based on the node values of the previous time step.

In order to determine the attractor landscape for our network topology (Figure 2), we simulated the trajectory of every possible value configuration for our network topology. Because each node can have two values and there are eight nodes in our topology, there are 2^8, or 256, possible value configurations of our topology. We ran simulations lasting for 20 time steps on each of the 256 possible initial network configurations. The resulting attractor landscape showed that the trajectories of the 256 initial network configurations converged to five different fixed states, or attractors. The largest attractor, or the fixed state that most of the initial configurations fell into, was the biologically functional attractor. That is, of the 256 unique initial configurations, 219 of those configurations converged on a fixed state that would lead to the proper regulation of the early embryonic cell cycle of *C. Elegans*.

Figure 2 [I realize it’s difficult to actually see the 5 distinct attractors in this visualization. Thinking about using more colors to distinguish attractors in final version] 

In measuring the informational dynamics of the *C. Elegans* early embryonic cell cycle, we measured the TE between every pair of nodes and the AI of each node for seven distinct network trajectories. Those trajectories included the biologically functional trajectory and the single path trajectories of four more initial network configurations, each of which converged on a different, non-biologically functional attractor. So, five of the network trajectories we analyzed were single path trajectories corresponding to the five fixed attractor states of our network. The sixth network trajectory we analyzed was not a single path trajectory, but the averaged trajectories of all 219 initial network configurations that converged on the biologically functional attractor. The sixth trajectory, in other words, encompassed the entire basin of the biologically functional attractor. The seventh network trajectory we analyzed encompassed all 256 initial network configurations.

We calculated the TE of every node pair and the AI of every node by looking at a one time step history of the node(s) being measured. We chose to use a common history length in calculating the TE and AI of our network’s node pairs and nodes to better compare the results across the seven network trajectories. We decided to use the relatively short history length of one time step because the four time step trajectories that lead to the four respective non-biological attractors had relatively short trajectory lengths. For three of the non-biological attractor basins, the longest trajectory is 2 time steps. For the fourth non-biological attractor basin, the longest trajectory is 5 time steps. For the biologically functional trajectory in the biologically functional attractor basin, the trajectory is 8 time steps long.

**Results**

While analyzing the data produced by measuring the TE and AI of the seven network trajectories just mentioned, we paid special attention to node pairs sharing a causal edge and the nodes comprising the control kernel for our network’s biologically functional attractor. A node pair shares a causal edge in our network so long as they share an edge in the network topology (Figure 1).

We borrow the notion of a control kernel outlined by Kim et al [source number], which they define as the smallest set of nodes that, when pinned to a constant value at every time step, force the network into a certain attractor state, regardless of the network’s initial configuration. A consequence of that definition of a control kernel is that the set of nodes constituting a control kernel must be pinned to whatever Boolean value they hold in the desired attractor state.

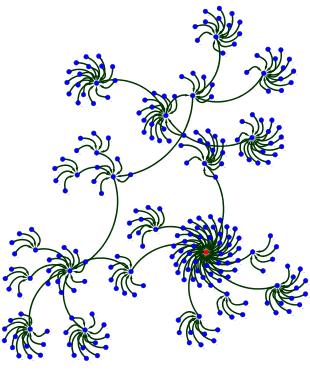
We focused on the biologically functional control kernel, or the control kernel capable of forcing any of the 256 possible initial conditions of our network into the biologically functional attractor state. We determined the control kernel for the biologically functional attractor by comparing the node values of our network’s five fixed point attractors. We found the node values that made the biologically functional attractor unique, the node values that the biologically functional attractor did not share with any of the other four attractor states. Those node values were:

lin-35/efl-1/dpl-1 = 1 (on)

cdc-25.1 = 0 (off)

To test whether those node values acted as a control kernel for the biologically functional attractor, we held them constant at every time step for a 20 time step simulation of all 256 possible trajectories. The result was that the attractor landscape went from containing five fixed attractor states to a single fixed attractor state (Figure 3). That remaining fixed attractor state was the biologically functional attractor (red).

Figure 3

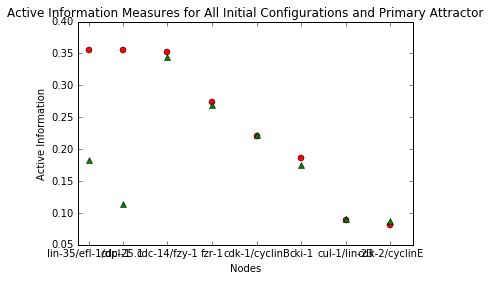


Thus, we concluded that the node values ‘lin-35/efl-1/dpl-1’ = 1 and ‘cdc-25.1’ = 0 comprised the control kernel for the biologically functional attractor.

Figures 4 through 7 contain the results of our informational analyses of the seven network trajectories outlined earlier. Figure 4 compares the active information for each node averaged over the trajectories of all the initial conditions with the active information for each node averaged over the trajectories of all the initial conditions that lead to the biologically functional attractor.

[I had never done any kind of coding before this semester. I have never used matplotlib. Hence this series of sloppy graphs. If you have any tips and tricks or a favorite tutorial, let me know. Going to try and get these looking nicer for the final paper next week.]

Figure 4



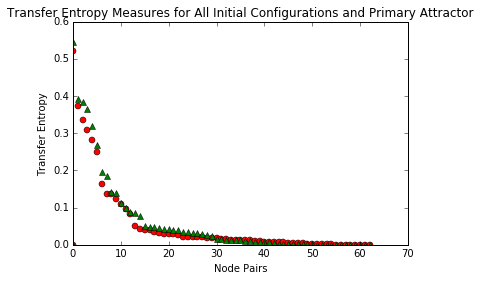
Green Triangle= Primary Attractor

Red Circle= All Initial Configurations

[Two notches furthest left in x-axis are control kernel nodes]

Figure 5 compares the transfer entropy for each node pair averaged over the trajectories of all the initial conditions with the transfer entropy for each node pair averaged over the trajectories of all the initial conditions that lead to the biologically functional attractor.

Figure 5

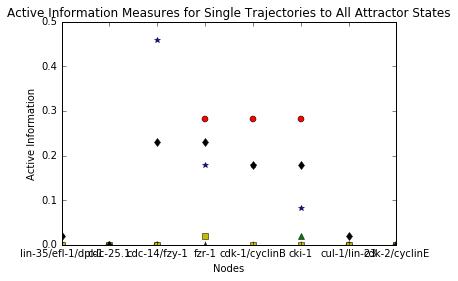


Green Triangle= Primary Attractor

Red Circle= All Initial Configurations

Figure 6 compares the active information measures for every node across 5 single trajectories, each of which start from unique initial configurations and each of which lead to one of the 5 attractor states of our network.

Figure 6



Red Circle= Ending in Fixed Attractor State 112

Green Triangle= Ending in Fixed Attractor State 0

Blue Star= Ending in Fixed Attractor State 117

Yellow Square= Ending in Fixed Attractor State 113

Black Diamond= Biologically Functional Trajectory

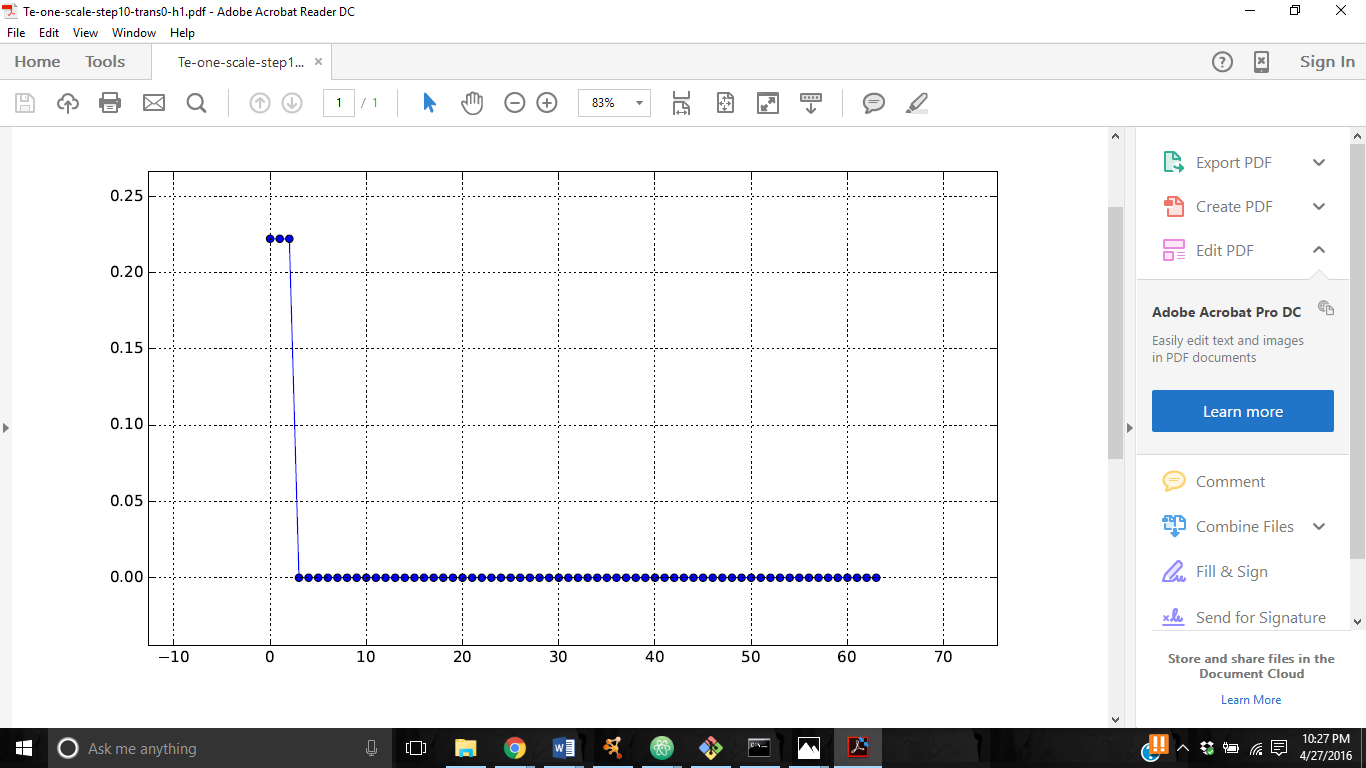
[Two notches furthest left in x-axis are control kernel nodes]

Figure 7 compares the transfer entropy for each node pair for every node pair across 5 single trajectories, each of which start from unique initial configurations and each of which lead to one of the 5 attractor states of our network.

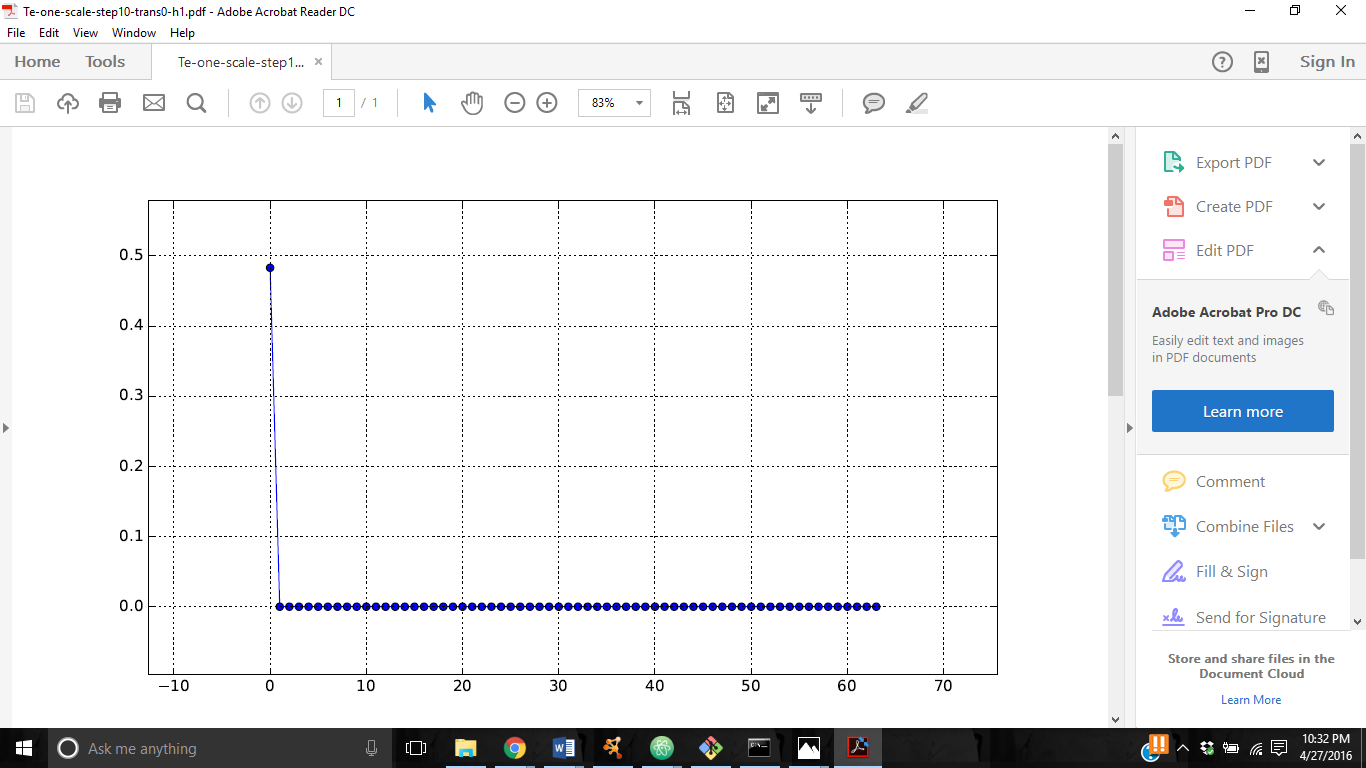
[I am planning on setting up this graph like Figure 5. Good idea? For now, here are the 5 TE graphs I am planning on combining. What is a good way to mark the causal edges on a data visualization that contains 5 different values (one for each trajectory) for each possible node pairing?]

Figure 7

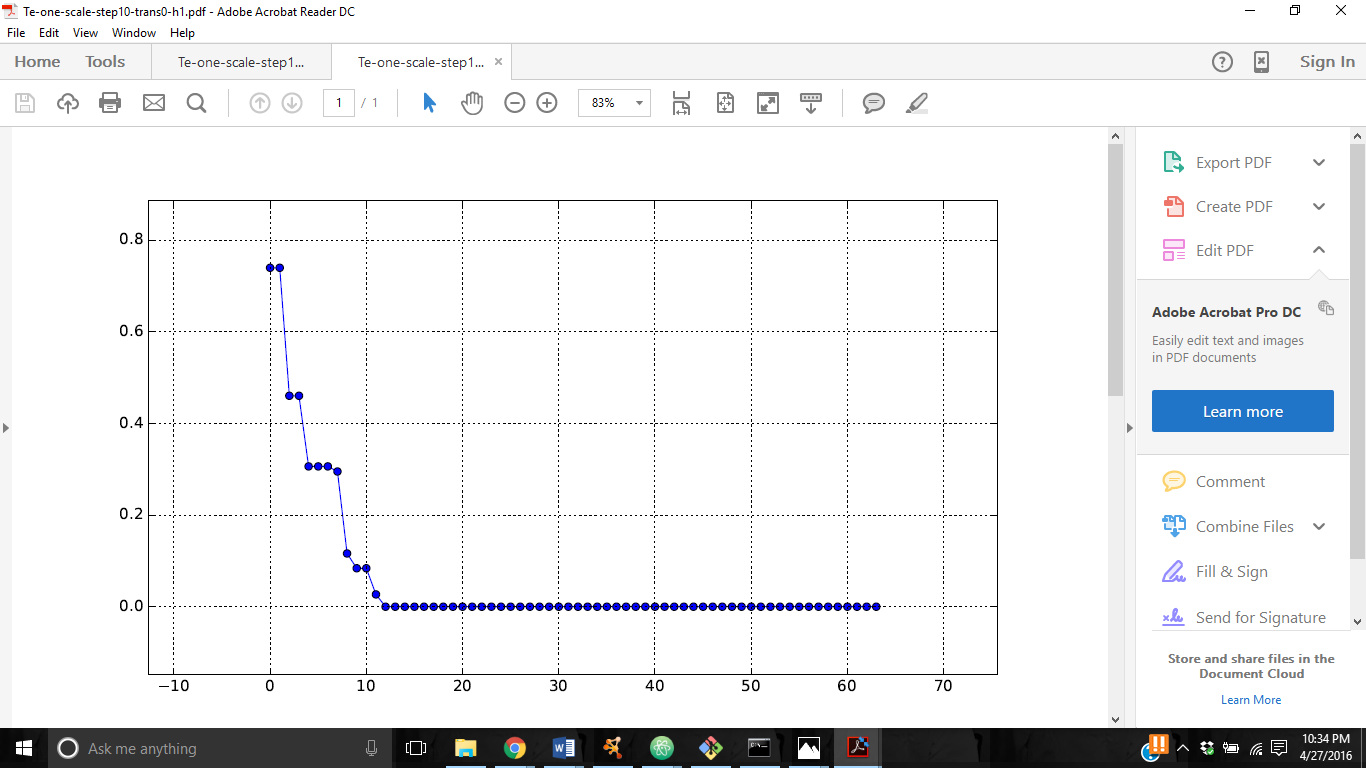
(Trajectory Ending in Attractor State 112)



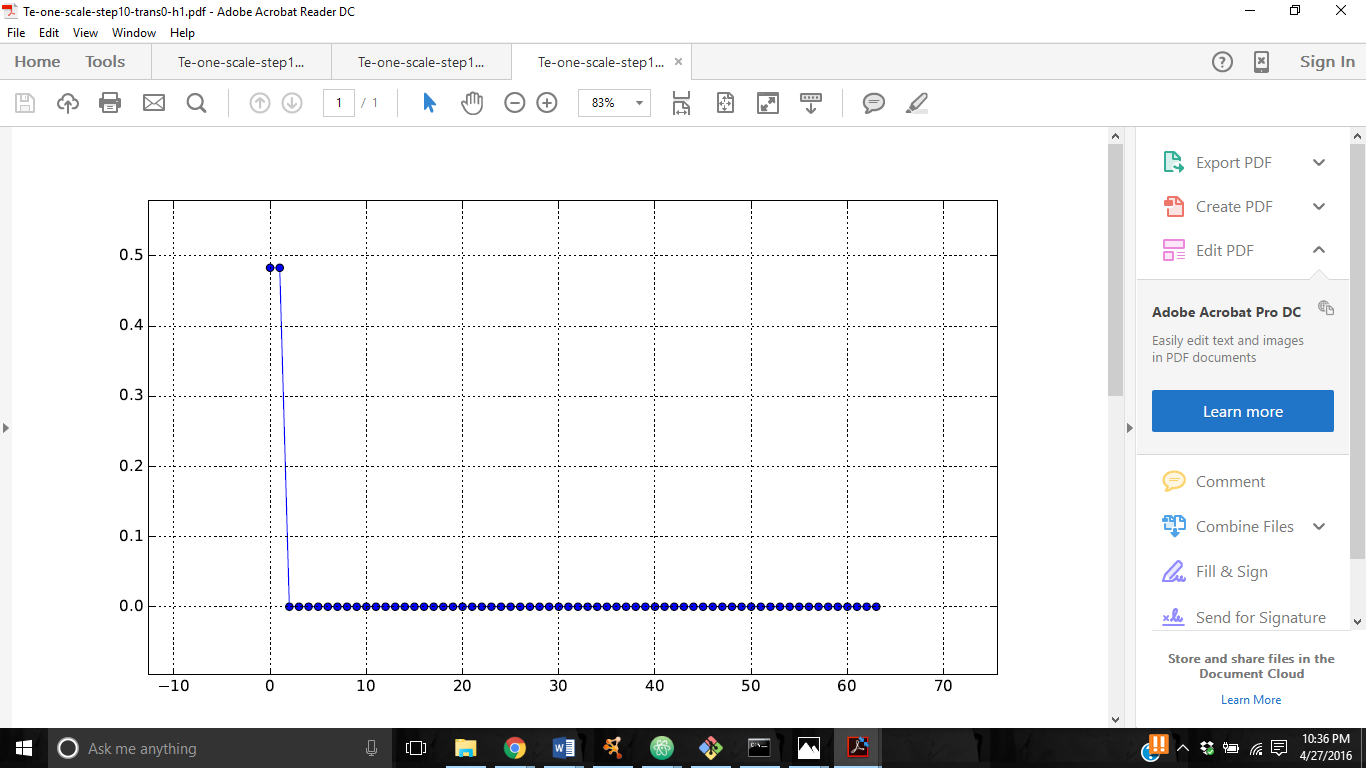
(Trajectory Ending in Attractor State 0)



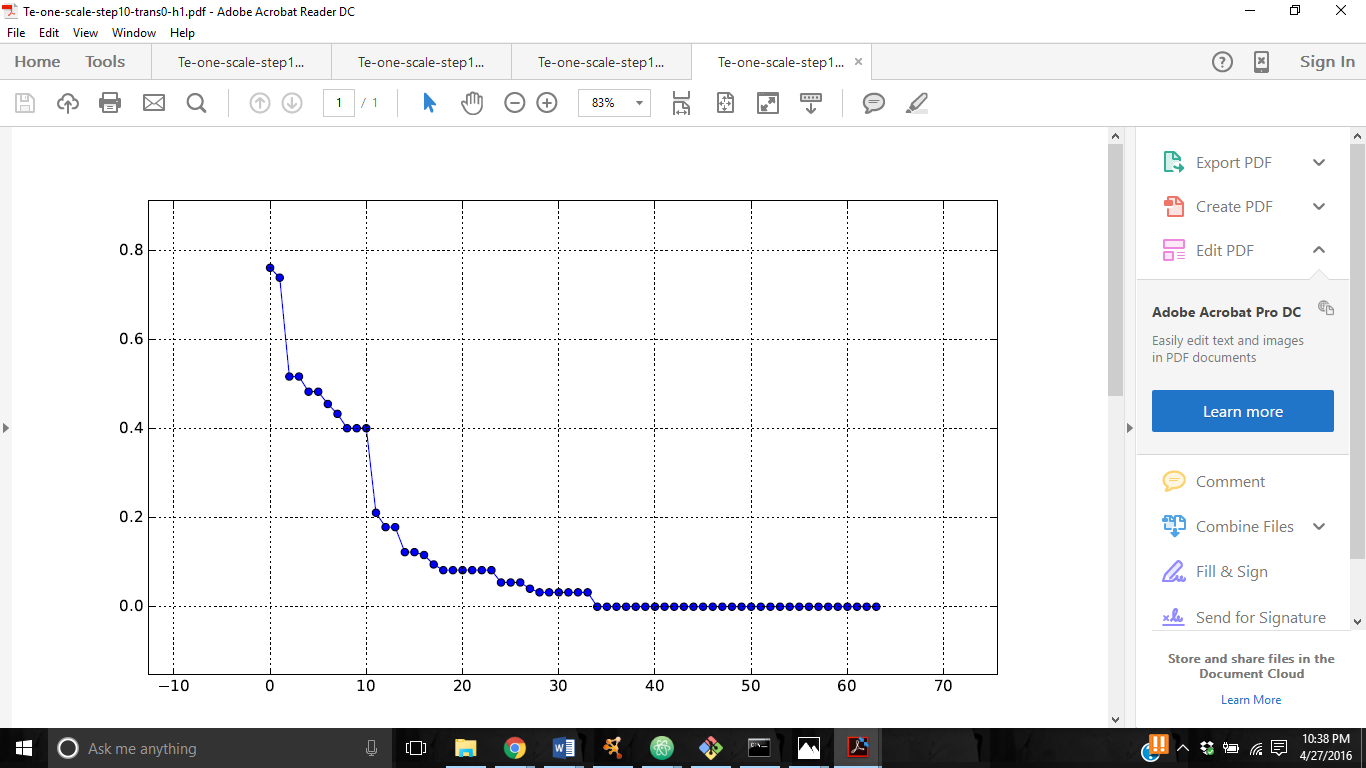
(Trajectory Ending in Attractor State 117)



(Trajectory Ending in Attractor State 113)



(Biologically Functional Trajectory)



**Summary/Discussion**

As shown by figures 4 and 5, the active information measures for each node and the transfer entropy measures for each pair of nodes averaged across all possible trajectories and all trajectories within the biologically functional attractor basin are similar. That similarity in results between all of the possible trajectories and all of the trajectories within the biologically functional attractor basin is probably due to how large the biologically functional attractor basin is. There are 256 distinct trajectories in our network’s attractor landscape, and 219 of those, or 85.5%, also fall within the biologically functional attractor basin.

In comparing the transfer entropy measures of the 5 single trajectory network paths leading to distinct attractor states, we found that longer network trajectories processed more information.

[I haven’t marked the causal edge pairs in my figure 7 graphs yet, so that may provide me with something cool. Otherwise, I have about 500 words I could spend discussing non-obvious interpretations of my data. Any ideas appreciated.]

**Citations [not yet compiled]**

Kim, Junil, Sang-Min Park & Kwang-Hyun Cho (2013). “Discovery of a Kernel for Controlling Biomolecular Regulatory Networks.” *Scientific Reports* 3 : 2223