**The Informational Dynamics of the *Caenorhabditis elegans* Early Embryonic Cell Cycle Gene Regulatory Network**

Kelle Dhein, Hyunju Kim, and Sara Walker

*Arizona State University*

In exploring the transitions between physics, chemistry, and biology, scientists have been searching for a reliable signature that distinguishes living systems from non-living systems. The goal is not merely to build a more rigorous taxonomy of empirical phenomena, but to gain deeper insights into the set of astounding phenomena surrounding life. One candidate for such a signature are metrics of information. Here we show that the informational dynamics of a biological system (a gene regulatory network in *Caenorhabditis elegans*) show interesting variations between biologically functional outcomes and non-biologically functional outcomes. In analyzing the information dynamics of a biological system, we apply the concept of a control kernel to our network model.

1. **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 [4]. 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* (*C. elegans*). In measuring those dynamics, we follow previous researchers [4] 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. *C. Elegans*, a species of nematode, is a mutlicelluar eukaryotic organism with a nervous system , but is relatively simple compared to other organisms that fulfill the same criteria. Researchers have been able to fully sequence the C*. Elegans* genome, create a network model of all 302 neurons in the *C. Elegans’* nervous system [7], and map the development of every somatic cell in a developing *C. Elegans* [6, 8]. Our project builds on past research that determined the genes and proteins responsible for regulating the early embryogenesis of *C. Elegans* [1, 9].

By simulating the dynamics of those genes and proteins 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. AI measures how much our uncertainty about the present state of a single node is reduced by looking into the past states of that same node. TE measures how much our uncertainty about the present state of a single node is reduced by looking into the states of another node.

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.

1. **Model Description**

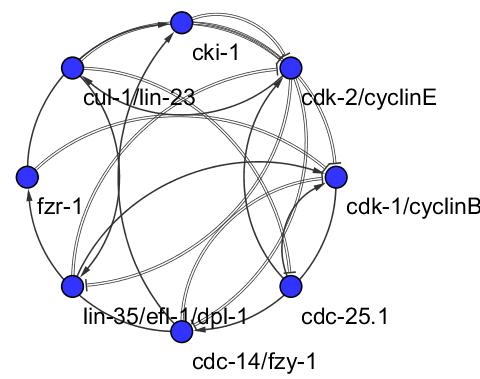
Multi-cellular organisms originate from a single cell, which divides itself into two daughter cells. Those daughter cells go onto divide themselves into daughter cells so that the organism gains cells at an exponential rate. While dividing, a cell goes through a series of phases. Though those phases 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 [9]. 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. 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 [9].

Based on previous research [4, 9], 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 yeast [2] and mammals [3].

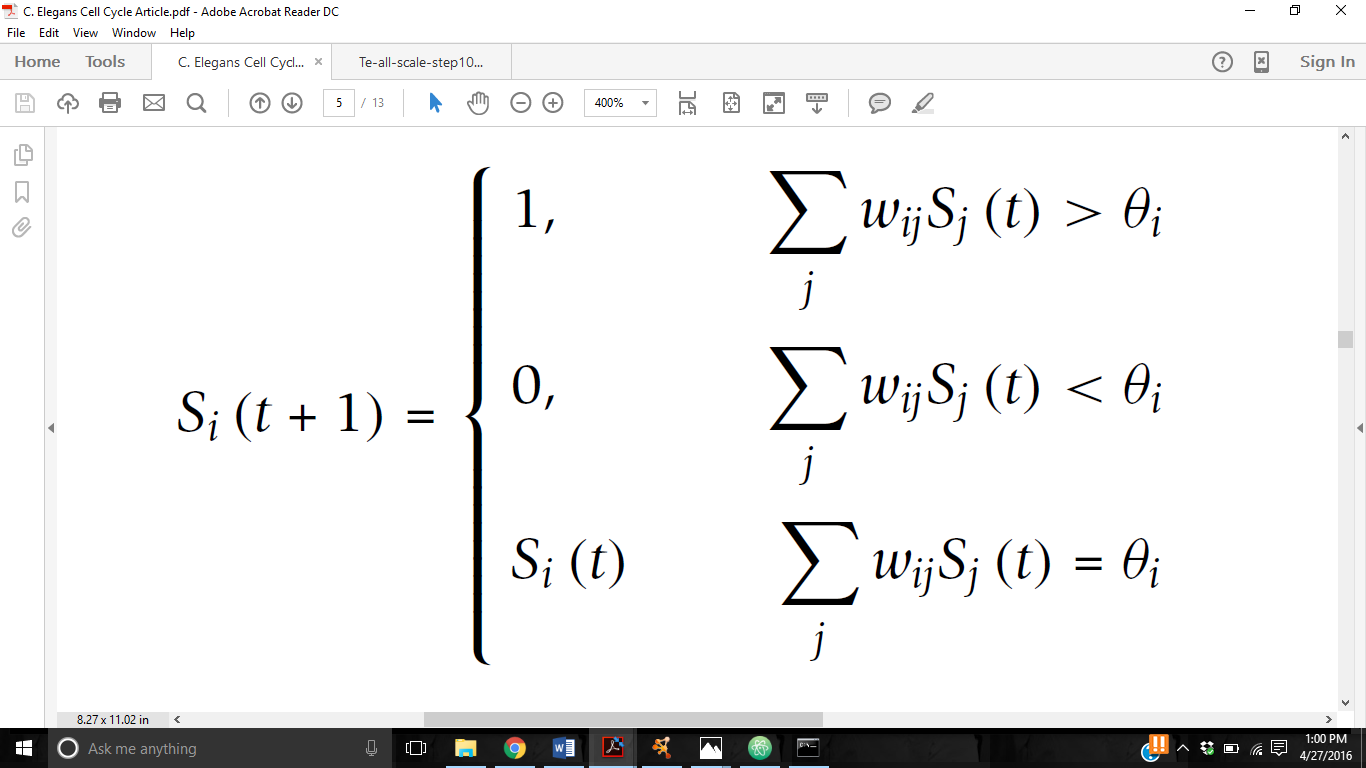
In creating our Boolean network model of cell division in early embryonic *C. Elegans*, we built on the model created by Xiaotai et al’s [9]. Our model is a network model in that it represents proteins, genes, and groups of functionally unified proteins and genes as nodes. Our network model is Boolean 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*. Boolean network models are an effective way of modeling genetic systems because genes and proteins are considered to be on or off depending on whether they are playing a functional role in some genetic 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 originate from a source node as two parallel lines and terminate at a target node with an flat bar. Genes and/or protein nodes that target other nodes with a -1 edge are deactivating or inhibiting some other set of genes and/or proteins. Edges with a weight of 1 originate from a source as a single solid line and terminate at a target node with an arrow. Nodes targeting other nodes with a 1 edge represent sets of genes and/or proteins that activate other genes or proteins by triggering them to play some role in regulating the early embryonic cell cycle. The network topology consists of eight nodes and twenty-one edges.



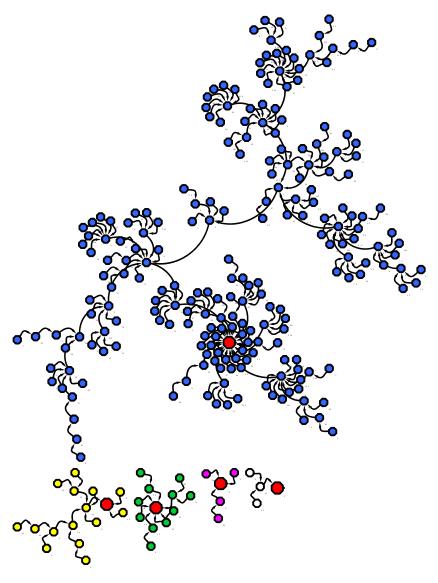
**Figure 1.** The network topology of the *C. elegans* early embryonic cell cycle.

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 *Si* represents the state of target node *i* at some time *t*. *j* represents a source node upstream of *i*, and *wij*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*i*. All nodes in our network have a threshold value of 0. At each time step, every node in our network updates simultaneously based on the network state in the previous time step.

In order to determine the attractor landscape for our network topology (Figure 2), we simulated the trajectory of every possible network state for our network topology. Because each node can have two values and there are eight nodes in our topology, there are 28, or 256, possible network states of our topology. We ran simulations lasting for 20 time steps on each of the 256 possible initial network states. We decided to run the simulation for 20 time steps after discovering that the longest running trajectory in our state space continued for 16 time steps. 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.** The attractor landscape of our network tracing the trajectory of every possible initial network state. Each of the red nodes represents an attractor state. The green nodes converge on attractor 1, the white nodes converge on attractor 2, the yellow nodes converge on the 3, the purple nodes converge on attractor 4, and the blue nodes converge on the biologically functional attractor.

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.

1. **Results**

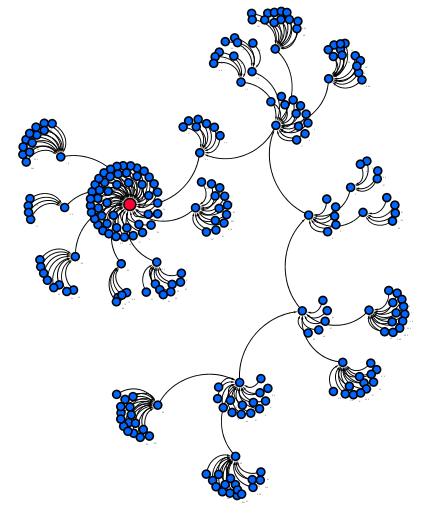
While analyzing the data produced by measuring the TE and AI of the seven network trajectories just mentioned, we paid special attention nodes comprising the control kernel for our network’s biologically functional attractor. We borrow the notion of a control kernel outlined by Kim et al [5], which they 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)

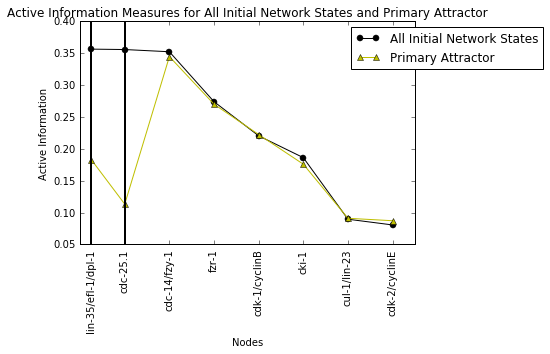
To test whether those node values acted as a control kernel for the biologically functional attractor, we held them constant at every time step for a 20 time step simulation of all 256 possible trajectories. The result was that the attractor landscape went from containing five fixed attractor states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



**Figure 3.** The attractor landscape of our network when the node representing lin-35/efl-1/dpl-1 is forced to remain active at every time step and the node representing cdc-25.1 is forced to remain inactive. The red node represents the biologically functional attractor state. All of the nodes converge to the biologically functional attractor state.

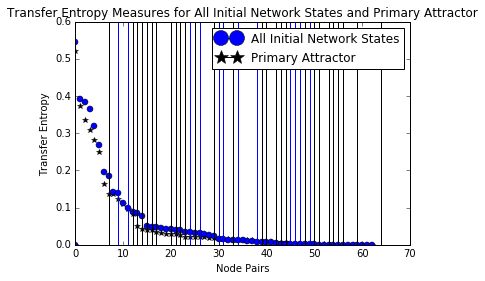
Thus, we concluded that the node values ‘lin-35/efl-1/dpl-1’ = 1 and ‘cdc-25.1’ = 0 comprised the control kernel for the biologically functional attractor.

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.



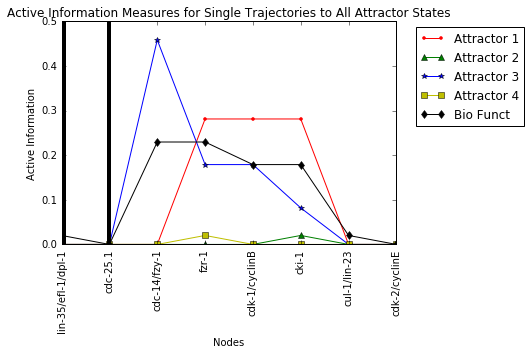
**Figure 4.** Active information measures for each node averaged over all of the network state trajectories and averaged over all of the primary attractor basin trajectories. Control kernel nodes are marked by black vertical lines.

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.



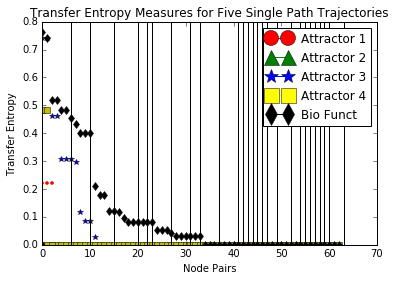
**Figure 5.** Transfer entropy measures for each node pair averaged over all of the network state trajectories and averaged over all of the primary attractor basin trajectories. Vertical lines mark node pairs containing at least one control kernel node. Blue vertical lines mark such pairs for the all initial network states plot. Black lines mark such pairs for the primary attractor plot.

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.** Active information measures for each node of each the five trajectories analyzed. Control kernel nodes are marked by black vertical lines..

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.



**Figure 7**. Transfer entropy measures for each of the five single trajectories analyzed. Black vertical lines mark node pairs within the biologically functional trajectory plot that contain at least one control kernel node.

**Summary/Discussion**

As shown by figures 4 and 5, the AI measures for each node and the TE 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.

An interesting exception to that trend was that every node except the control kernel nodes had similar AI values when averaged over all the trajectories in the primary attractor group vs. all possible trajectories (Figure 4). In the primary attractor group of trajectories, the two control kernel nodes have significantly lower AI values. A possible explanation for this anomaly is that the control kernel nodes in the primary attractor group maintain steady values, namely, those values that steer the network into the biologically functional attractor state. That homogeneity in node value may lead the control kernel nodes to have lower AI scores in the primary attractor group than they do in the all initial conditions group, where they may be alternating between active and inactive more frequently.

In comparing the TE measures of the 5 single trajectory network paths leading to distinct attractor states, we found that longer network trajectories processed more information. We also found that, within a single biologically functional trajectory, node pairs containing at least one control kernel node were slightly more likely to have a TE value of 0.

Future research into analyzing the informational dynamics of different network state trajectories within the same biological system may lead to a set of “best practices” for interrogating biological systems in general. Our preliminary results show the potential for computational models to articulate abstract aspects of biological function. Further research into whether the data provided by our project and projects like ours consistently point to specific kinds of biological phenomena are needed.

**Citations**

1. Bao Z, Zhao Z, Boyle TJ, Murray JI, Waterston RH. Control of cell cycle timing during C. elegans embryogenesis. Dev Biol 2008, 318: 65-72.
2. Davidich MI, Bornholdt S. Boolean network model predicts cell cycle sequence of fission yeast. PLoS One 2008, 3.
3. Faure A, Naldi A, Chaouiya C, Thieffry D. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics 2006, 22.
4. Kim, Hyunju, Paul Davis, Sara Imari Walker. New scaling relation for information transfer in biological networks. Journal of The Royal Society Interface. 2015, 12 (113).
5. Kim, Junil, Sang-Min Park & Kwang-Hyun Cho. Discovery of a Kernel for Controlling Biomolecular Regulatory Networks. *Scientific Reports* 2013, 3 : 2223
6. Kimble, J; Hirsh, D. "The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans". Developmental Biology 1979, 70 (2).
7. Kosinski, RA; Zaremba, M. Dynamics of the Model of the Caenorhabditis elegans Neural Network. Acta Physica Polonica B 38 2007, (6): 2201. Bibcode:2007AcPPB..38.2201K.
8. Sulston, JE; Horvitz, HR. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Developmental Biology 1977, 56 (1).
9. Xiaotai, Huang, Long Chen, Hung Chim, Leanne Lai Hang Chan, Zhongying Zhao, Hong Yan. Boolean genetic network model for the control of C. elegans early embryonic cell cycles. Biomedical Engineering OnLine. 2013, 12 (1).