

Information Dynamics of the Gene Regulatory Network Underlying Mammalian Cortical Area Development

Harrison B. Smith

Arizona, United States

Abstract

Information measures have been proposed as a method to quantitatively distinguish biotic networks from abiotic or random networks [1]. Additionally, there has been increased interest in understanding the underlying control kernels that can map biologically unfavorable systems into biologically favorable ones [2]. Here we use the boolean network representation of mammalian cortical area development to show that information measures can be used to distinguish biologically relevant gene expression trajectories from all possible mammalian cortical area gene expression trajectories. Using dynamic information measures to distinguish non-biologically viable from biologically viable cortical gene expression networks, and the mechanisms to control each has implications for understanding the development of all mammalian brains.

1. Introduction

2 The information dynamics of yeast and budding yeast cell cycle networks
3 reveal key differences between biological networks and their randomly gener-
4 ated counterparts [1]. Here I am interested in investigating if these differences
5 are present between categories of trajectories of a network interpreted to be
6 biologically relevant. In order to investigate this, I use a boolean model of
7 the gene regulatory network underlying mammalian cortical area develop-
8 ment [3].

9 This particular system is studied because the healthy development of
10 a mammalian adult brain (specifically a mouse brain) is recognized by a
11 particular expression gradient of five genes. Additionally, this healthy adult
12 expression gradient will only be realized if, on embryonic day 8, there is a
13 particular (but different) expression gradient of these same five genes. Thus,

14 while the initial and desired biological network states are specified by these
15 empirically measured gradients, the rules governing the transition from the
16 initial to final states must be determined through dynamical models.

17 Previous research has determined the optimal interaction rules necessary
18 to reach the biologically desired final state from the specified initial biological
19 state [3]. I use these interaction rules to generate the state transition map
20 of the gene regulatory network underlying mammalian cortical area devel-
21 opment, and analyze how the transfer entropy (TE) and active information
22 (AI) of nodes in the network differs between the biologically relevant tra-
23 jectory, trajectories within the biologically relevant primary attractor, and
24 trajectories within any attractor of the state transition network as a whole.
25 I also determine the presence of control kernels in the network, compare to
26 previous results on this networks control kernels, and attempt to identify
27 how they could play a role in shaping the information measures we observe
28 in these networks [2].

29 2. Model Description

30 The boolean model of the gene regulatory network underlying mammalian
31 cortical area development is described by ten nodes, Fgf8, Emx2, Pax6, Coup-tfi,
32 Sp8, Fgf8, Emx2, Pax6, Coup-tfi Sp8. The italicized nodes represent
33 genes, and the non-italicized nodes represent proteins (for clarity, I
34 will herein denote the ten above nodes as gF, gE, gP, gC, gS, pF, pE, pP,
35 pC, pS). Each node can either be in state 0 (inactive) or 1 (active). Nodes
36 can be connected by an activation link or inhibition link. The interactions
37 between nodes was determined in Giacomantonio and Goodhill, 2010 [3] by
38 iterating through all possible connections of nodes, where the best interaction
39 network was deemed to be the one that ended up in the biologically desired
40 state with the greatest probability. The rule for updating a nodes state is as
41 follows: a node is activated if in the previous time step all nodes regulating
42 it through activation links are activated, and all nodes regulating it through
43 inhibition links are inactive. In other words, all conditions must be met for
44 a node to become activated. In the original work used to determine the
45 most likely interaction network [3], the network update rules are determin-
46 istic, but were implemented stochastically. This was to avoid synchronous
47 node updating, seen by the authors as unrealistic. Additionally, research
48 on mammalian cortical area development actually recognizes different initial

49 and desired expression gradients in the anterior and posterior areas of the
 50 cortex.

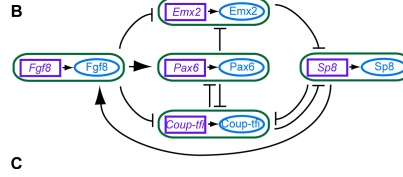


Figure 1: The "best" anterior network, used for all analyses in this work. Arrows indicate activation edges, and bars indicate inhibition edges. Gene nodes (purple) and protein nodes (blue) are encapsulated for clarity (figure adopted from [3])

51 When determining system dynamics for my information analyses, I fol-
 52 lowed strictly deterministic update rules using the best interaction network
 53 as determined by previous work [3] (See Fig. 1). As a consequence of follow-
 54 ing deterministic update rules, it is impossible to reach the desired posterior
 55 steady state from the posterior initial state 3, as they are in different at-
 56 tractor basins. Thus, my control kernel analyses and information dynamic
 57 analyses reference only on the anterior network.

58 Here the biological initial state is,
 59 gF^* , gE , gP , gC , gS , pF^* , pE , pP , pC , pS ,
 60 and the desired steady state is,
 61 gF^* , gE , gP^* , gC , gS^* , pF^* , pE , pP^* , pC , pS^* ,
 62 where activated nodes are denoted with an asterisk (*).

63 3. Results

64 3.1. Attractor

65 The state transition diagram of a network under deterministic update
 66 rules shows how each network state maps to another network state. Because
 67 I have 10 nodes which can each take on two states, there are $2^{10} = 1024$
 68 possible network states in my state transition diagram. It consists of 3 basins
 69 of attractions. The largest basin consists of 841 nodes and contains the initial
 70 state as well as the biologically desired steady state, which is a fixed attractor.
 71 The second largest basin consists of 174 nodes, with a 2-node cyclic attractor.
 72 The smallest basin consists of 9 nodes, with a fixed attractor.

3.2. Control Nodes

Control kernels were identified using two different definitions (denoted here as type I and type II). In the first (type I), the control kernel is defined as the minimal set of network nodes needing to be regulated to drive the network state to converge to any desired attractor regardless of the systems initial state [2]. Using this definition, previous work [2] found there to be six control kernels (each denoted with an asterisk) of one node each:

gF^* , gE , gP^* , gC , gS^* , pF^* , pE , pP^* , pC , pS^* ,

Repeating the analysis, I also found there to be six control kernels of one node each (any of which must be active to reach the desired attractor state), but I found the Pax6 gene and protein to be control kernel nodes instead of the Emx2 gene and protein:

gF^* , gE , gP^* , gC , gS^* , pF^* , pE , pP^* , pC , pS^* ,

My result is consistent with the desired steady state of my system, where the Emx2 and Coup-tfi genes and protein nodes are all inactive, and all other nodes are active. I believe the previously identified type I control kernel by Kim et al, 2013 [2] to be erroneous, because by their own definition of the control kernel, the desired attractor state is impossible to obtain by pinning the Emx2 gene or protein to the active state, and here was found not to converge on the primary attractor when pinned to the inactive state.

Besides driving the network to the primary attractor basin, I found there to be type I control kernels that drive the network state to the smallest basin of attraction, but not the remaining intermediate-size basin of attraction. This intermediate-size basin differs from the others in that it has a cyclic attractor state of two nodes instead of a single fixed attractor state. Perhaps this difference affects the causal power of the control kernel, preventing it from arriving in this attractor basin. Further analysis is necessary to either support or reject this hypothesis.

I also investigated a second definition of a control kernel (type II), defined by the minimal set of nodes needed to be active in order to guarantee that, regardless of the activity of the other nodes, the state will be in the primary attractor and thus converge on the desired steady state. Unlike the type I control kernels, regulating a network to a specific attractor basin using these type II kernels does not require intervening every time step to update a node. Instead, these type II control kernels define the minimum set of nodes that must be regulated only once in order to reach a specific attractor. Abiding by this definition, I found nine control kernels of two nodes each. Any state which has any one of the following protein nodes active,

111 pF*, pP*, pS*,
 112 along with any one of the following gene nodes active,
 113 gF*, gP*, gS*,
 114 will lie within the primary attractor basin.

115 3.3. Information Dynamics

116 I used a time series length of 22 to calculate transfer entropy and active
 117 information because this is the maximum number of steps needed to drive any
 118 network state to its basins attractor. Because the biological initial network
 119 state converges to the desired steady state attractor in only 7 time steps, the
 120 history length was limited to 6 at the highest. History lengths between 2 and
 121 5 yielded visually similar distributions for rank ordered transfer entropy and
 122 active information among each of the datasets. A history length of 2 was
 123 chosen for all analyses because it is the shortest history length, relative to
 124 the biological trajectory length, that still appears visually similar to longer
 125 history lengths for the aforementioned distributions.

126 When comparing rank ordered TE to AI across the different trajectory
 127 regimes, there are a few things that stand out 2. The biological trajectory
 128 has the lowest TE and highest AI, while TE is similar across trajectories
 129 in the primary attractor and cumulative across the network. Despite the
 130 similarity in TE, nodes in trajectories of the primary attractor show much
 131 lower AI.

132 Looking at TE vs edge connections, it can be seen that in both the edge
 133 case and no edge case, the primary attractor and cumulative trajectories
 134 show more node pairs with transfer entropy than without 3. The majority
 135 of nodes in these two regimes are correlated via information transfer but
 136 without causal connection (the percentage of nodes with TE $\neq 0$ is higher
 137 in the no edge case than the edge case). The majority of node pairs in the
 138 biological state trajectory have no transfer entropy. Of those that do, more
 139 are correlated via information transfer without causal connection than with
 140 causal connection.

141 Comparing heatmaps across nodes for TE and AI, it can be seen that in
 142 the biological trajectory (first row, 4), there is no transfer entropy either to or
 143 from any of the non-control kernel nodes. There is also no transfer entropy
 144 to or from gS. Nor to gF and pS. Comparing a-1 to b-1, the nodes pairs
 145 which had highest TE contained nodes which had the highest AI, except
 146 for pS \rightarrow pF, because pS has no active information. The data is more
 147 difficult to interpret across the primary attractor trajectories and all network

trajectories, although it does appear that node pairs with relatively high TE are the same causal node pairs in both regimes.

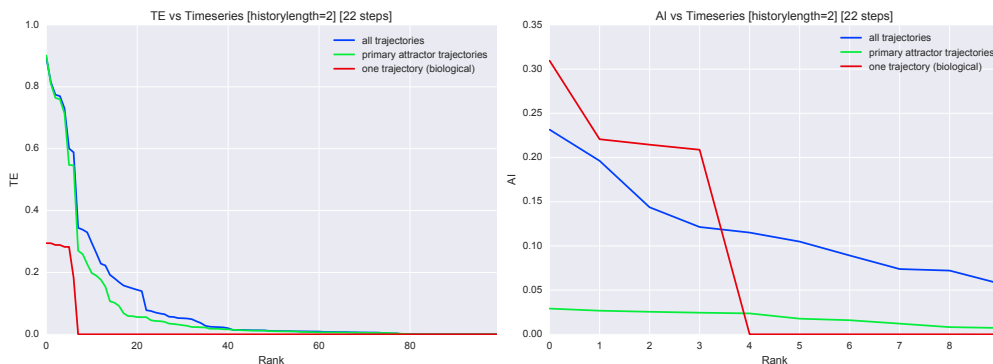


Figure 2: Rank ordered TE (left) and AI (right) across trajectory regimes. The biological trajectory has the lowest TE and highest AI, while TE is similar across trajectories in the primary attractor and cumulative across the network. Despite the similarity in TE, nodes in trajectories of the primary attractor show much lower AI.

4. Summary/Discussion

In studying the dynamics of the boolean model of the gene regulatory network of mammalian cortical area development, I found six control kernel nodes, two of which vary from existing literature. I also laid out the definition for a second type of control kernel, characterized by the idea that there is a way to control the attractor basin for a network without causually intervening in every step of the network dynamics.

I also investigated how, in this network, the biologically favored trajectory differs from other trajectories which converge to the biologically favored attractor, and ones that comprise the entire state transition network. The biological trajectory appears to contain nodes which are very predictive of their states by their own history (high AI), and not by the history of other nodes (low TE). Additionally it appears that when looking at node pairs which have non-zero transfer entropy across any trajectory regime, the information transfer is facilitated without casual connection (between nodes not directly connected by an edge).

While it is hard to say anything confidently about what the implications of the other intricacies of the TE and AI mean for these systems, there

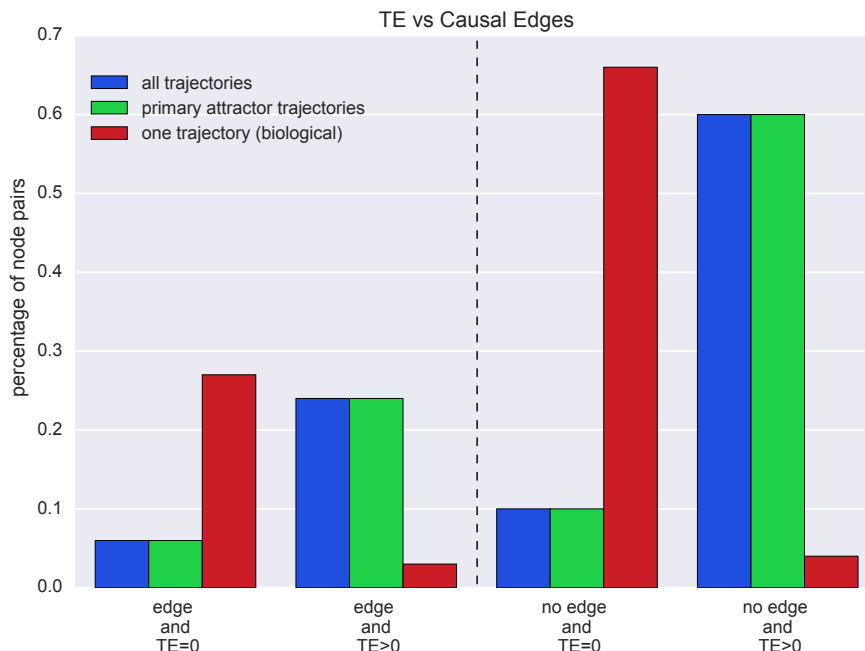


Figure 3: In both the edge case and no edge case, the primary attractor and cumulative trajectories show more node pairs with transfer entropy than without. The majority of nodes in these two regimes are correlated via information transfer but without causal connection (the percentage of nodes with $TE > 0$ is higher in the no edge case than the edge case). The majority of node pairs in the biological state trajectory have no transfer entropy. Of those that do, more are correlated via information transfer without causal connection than with causal connection.

are many more interesting questions to ask about this network, especially when taking its stochasticity into account. How would TE and AI across this deterministic implementation of the gene regulatory network underlying mammalian cortical area development differ from trajectories sampled from a stochastic implementation? How does the AI and TE of trajectories sampled from a stochastic implementation which successfully reaches the desired biological attractor differ from those trajectories which do not reach the desired biological attractor? How do the information dynamics differ between anterior and posterior stochastic trajectories?

177 **5. References**

- 178 [1] H. Kim, P. Davies, S. I. Walker, New scaling relation for information
179 transfer in biological networks, Journal of The Royal Society Interface
180 12 (2015) 20150944.
- 181 [2] J. Kim, S.-M. Park, K.-H. Cho, Discovery of a kernel for controlling
182 biomolecular regulatory networks, Scientific reports 3 (2013).
- 183 [3] C. E. Giacomantonio, G. J. Goodhill, A boolean model of the gene regu-
184 latory network underlying mammalian cortical area development, PLoS
185 Comput Biol 6 (2010) e1000936.

The author has calculated two different information measures in a boolean network representation of a mammalian gene regulatory network. He also showed identified control kernels in this network representation. He provide a new definition of control for these types of systems and compared that with previous results.

While I think the results presented are interesting, they are difficult to evaluate for a number of reasons. First and foremost, the author did not describe or define the information measures used. In the absence of any description, the biological relevance is difficult to determine. Second, the author did not describe the results of Kim et al 2015, which they used both for motivation for the study and comparison of their results. Without a description of the key findings in that study, the results of the present work are not clear. With regard to the new type of control described, the author should motivate why this type of control is either qualitatively different, and/or superior to the previous measures of control.

While I cannot recommend this article for publication in its current form, I would be happy to recommend it for publication if the above revisions are made.

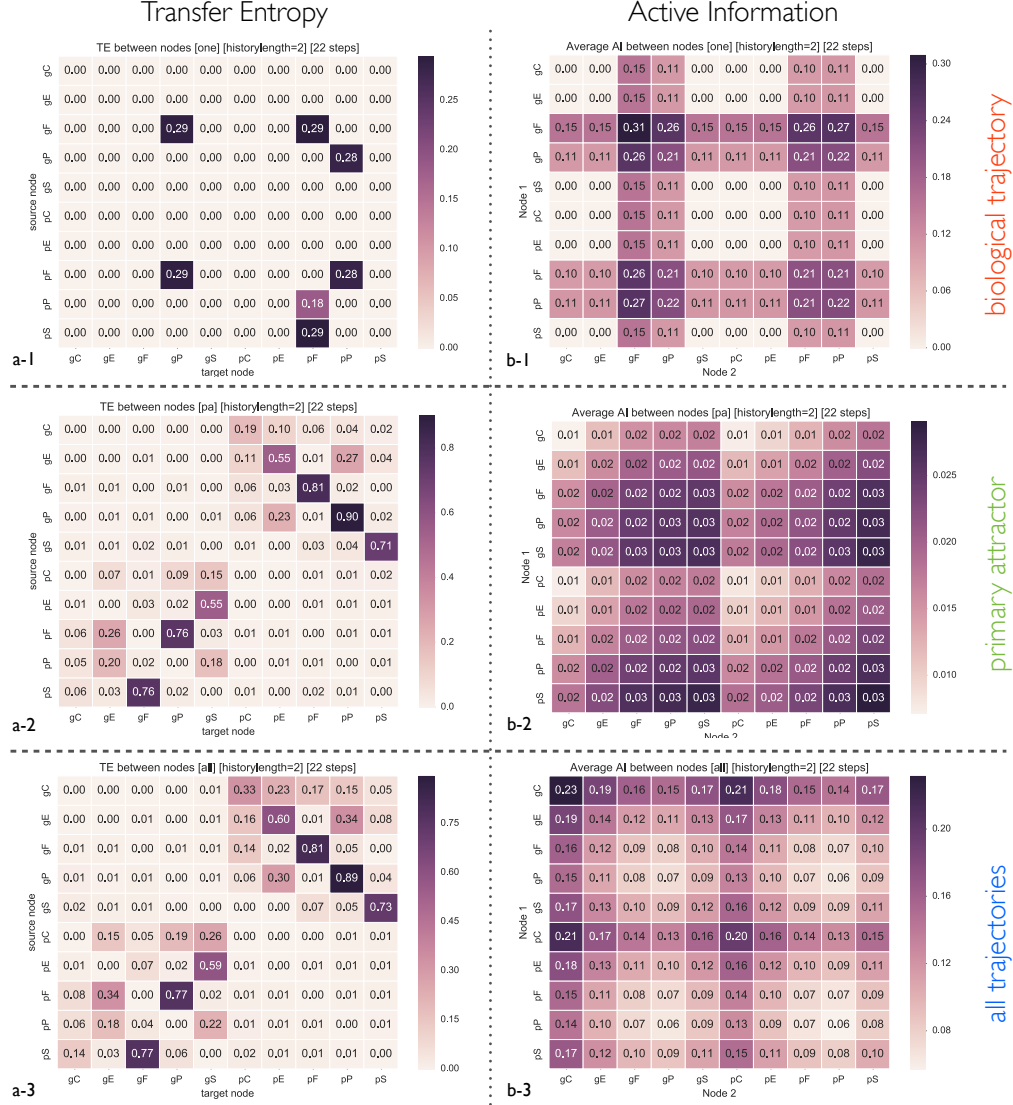


Figure 4: Heatmaps displaying TE from source to target nodes in column a (left) and the average AI between 2 nodes in column b (right). Column b heatmaps are symmetric, and the AI of a node can be obtained by look at the diagonal. Heatmaps of the biological trajectory are in row 1 (top), heatmaps of the primary attractor trajectories are in row 2 (middle) and heatmaps from all trajectories are in row 3 (bottom). Darker values means higher TE or higher AI, respectively. Each colorbar is scaled for its local heatmap. In a-1, there is no transfer entropy either to or from any of the non-control kernel nodes. There is also no transfer entropy to or from gS. Nor to gF and pS. Comparing a-1 to b-1, the nodes pairs which had highest TE contained nodes which had the highest AI, except for pS \rightarrow pF, because pS has no active information.