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

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

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

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

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

2. Model Description

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

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

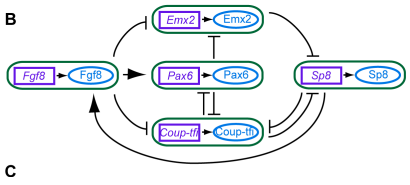

Figure 1: The "best" anterior network, used for all analyses in this work. Arrows indicate activation edges, and bars indicate inhibition edges. Gene nodes (purple) and protein nodes (blue) are encapsulated for clarity (figure adopted from [3])

When determining system dynamics for my information analyses, I followed strictly deterministic update rules using the best interaction network as determined by previous work [3] (See Fig. 1). As a consequence of following deterministic update rules, it is impossible to reach the desired posterior steady state from the posterior initial state 3, as they are in different attractor basins. Thus, my control kernel analyses and information dynamic analyses reference only on the anterior network.

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

3. Results

3.1. Attractor

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

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

3.3. Information Dynamics

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

When comparing rank ordered TE to AI across the different trajectory regimes, there are a few things that stand out 2. 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.

Looking at TE vs edge connections, it can be seen that in both the edge case and no edge case, the primary attractor and cumulative trajectories show more node pairs with transfer entropy than without 3. 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.

Comparing heatmaps across nodes for TE and AI, it can be seen that in the biological trajectory (first row, 4), 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− > pF, because pS has no active information. The data is more 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.

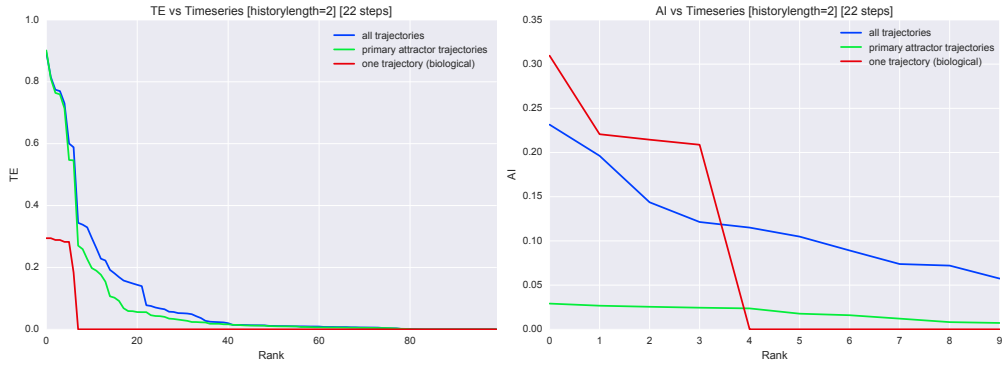

Figure 2: Rank ordered TE (left) and AI (right) across trajectory regimes. The biological trajectory has the lowest TE and highest AI, while TE is similar across trajectories in the primary attractor and cumulative across the network. Despite the similarity in TE, nodes in trajectories of the primary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

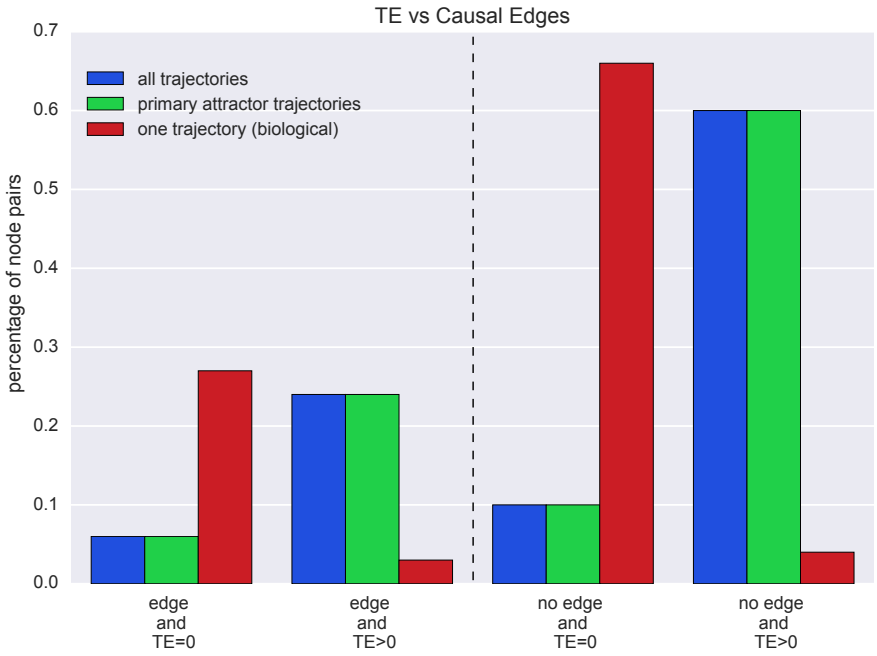

Figure 3: In both the edge case and no edge case, the primary attractor and cumulative trajectories show more node pairs with transfer entropy than without. The majority of nodes in these two regimes are correlated via information transfer but without causal connection (the percentage of nodes with TE > 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?

5. References

[1] H. Kim, P. Davies, S. I. Walker, New scaling relation for information transfer in biological networks, Journal of The Royal Society Interface 12 (2015) 20150944.

[2] J. Kim, S.-M. Park, K.-H. Cho, Discovery of a kernel for controlling biomolecular regulatory networks, Scientific reports 3 (2013).

[3] C. E. Giacomantonio, G. J. Goodhill, A boolean model of the gene regulatory network underlying mammalian cortical area development, PLoS Comput Biol 6 (2010) e1000936.

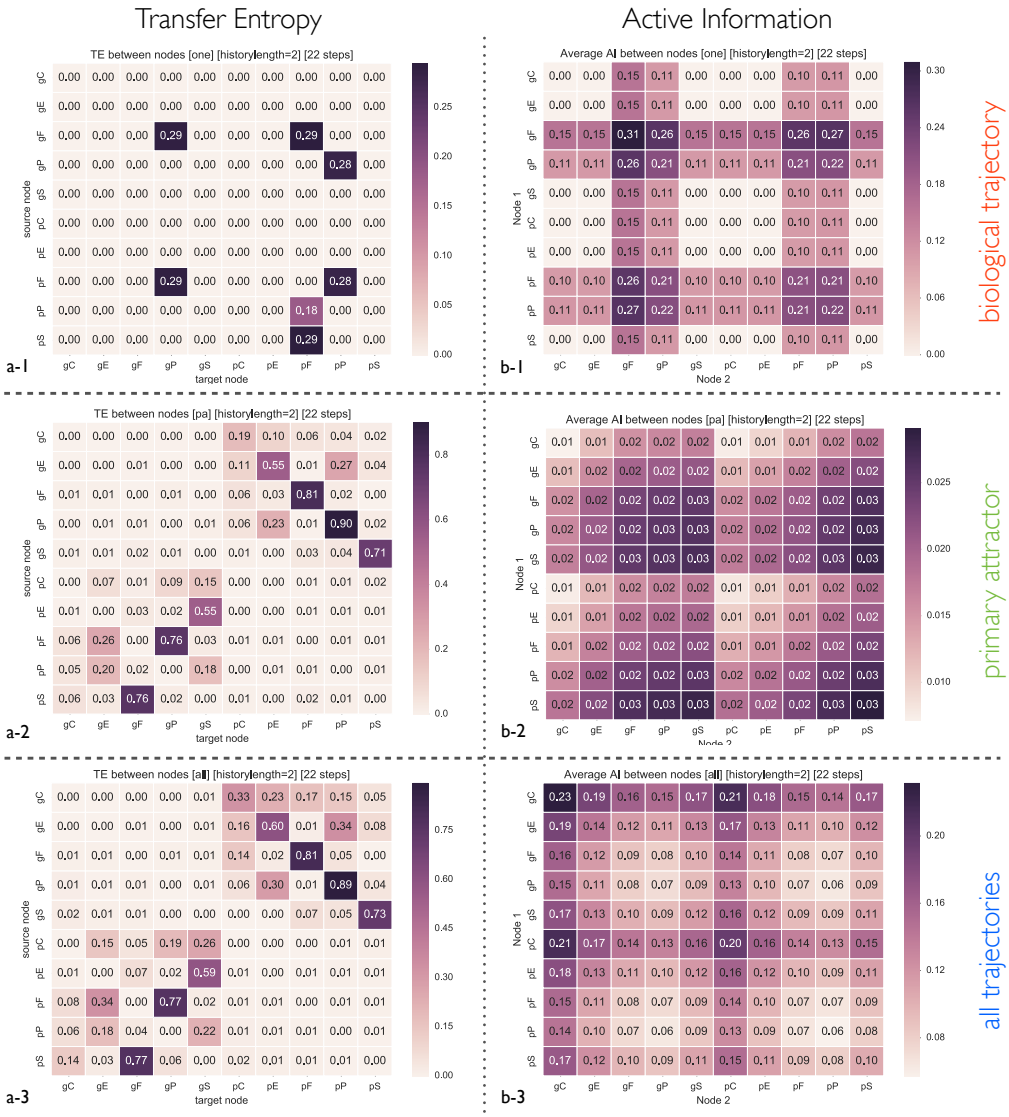

Figure 4: Heatmaps displaying TE from source to target nodes in column a (left) and the average AI between 2 nodes in column b (right). Column b heatmaps are symmetric, and the AI of a node can be obtained by look at the diagonal. Heatmaps of the biological trajectory are in row 1 (top), heatmaps of 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$->$pF, because pS has no active information.