The Stability of Gene Network Dynamics

Edward Ott

Reference:

S. Squires
Overview

• Our main result is a simple, very general criterion for determining the stability of discrete state dynamical models of gene regulation.

• This result makes possible treatment of network topology as well as other real factors previously inaccessible to study. Importantly, the result can be applied to networks derived from experimental data.
Gene Expression and Regulation

- **Genes**: Functional segments of DNA.
- **Gene expression**: The function of a given gene may be to produce ("express") a specific protein.
- **Regulation**: A gene’s expression is regulated, in large part, by the binding of particular proteins to its promoter region.

![Diagram showing gene expression and regulation](3)
The resulting directed network of interactions is given by the network’s ‘adjacency matrix’ $A$

$A_{ij} = 1$ if there is a link or 0 otherwise.

Figure taken from http://rsif.royalsocietypublishing.org/content/5/Suppl_1/S85.full
Naïve Boolean Model

- The state of each gene (network node) is either on (symbol 1) or off (symbol 0).

- Time is discretized: \( t = 0, 1, 2, 3, \ldots \).

- At time \( t \), the states of each node \( i \), denoted \( \sigma_i^t (=0 \text{ or } 1) \), are simultaneously updated using a deterministic Boolean function (truth table) of the time \( t-1 \) states, \( \sigma_j^{t-1} \), of the \( K_i^{in} \) nodes that input to node \( i \). (\( K_i^{in} = \) the in-degree of node \( i \).)

\[
K_i^{in} = 3: \quad j_2 \rightarrow i \quad j_1 \rightarrow i \quad j_3 \rightarrow i
\]
Update Truth Table

<table>
<thead>
<tr>
<th>current state at time $t$ (input to gene $i$)</th>
<th>State of gene $i$ at time $t+1$ (output)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene $j_1$</td>
<td>Gene $j_2$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$K_{i}^{in} = 2$

$K_{i}^{in}$ is the number of inputs to gene $i$. The table shows the current state of genes $j_1$ and $j_2$ at time $t$, and the state of gene $i$ at time $t+1$. The diagram illustrates the input connections to gene $i$.
Properties of the Boolean Model

• **Finiteness**: Eventually the system must return to a previously visited state.

• **Determinism**: Upon this return, the subsequent dynamics will be the same as for the previous visit.

• **Attractors**: Every initial condition produces a trajectory that eventually goes to a periodic orbit, called the “attractor” of that initial condition, and different initial conditions can go to different periodic orbit attractors.
Assumed significance of attractors

- An attractor may be thought of as representing a specific pattern of protein expression that defines the character of cells [Kauffman (‘69)].
Kauffman’s $N$-$K$ Net Model


$K_i^{in} = K$ independent of $i$. $N = \# \text{ of nodes}$. 

The $K$ inputs to each node $i$ are randomly chosen from amongst all the $N-1$ other nodes.

The output entries for each node’s truth table are randomly chosen with probability of 0 being $\frac{1}{2}$ and probability of 1 being $\frac{1}{2}$. 
Stability

• (State) = \( \sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N) \)

• Distance between two states, \( \sigma \) and \( \tilde{\sigma} \):
  \[ H(\sigma, \tilde{\sigma}) = \sum_i |\sigma_i - \tilde{\sigma}_i| \] (the Hamming distance)

• Instability (stability): If at \( t=0 \), \( H \ll N \), then \( H \) initially grows (shrinks) with increasing time \( t \).
Chaotic and stable dynamics for different networks

Hamming Distance

\( \text{<Kin>} = 4.65 \)

\[ q_c = 0.22 \]

\[ q = 0.5 \]

\[ q = 0.3 \]

\[ q = 0.2 \]
Past Work on Stability


\[ p = \text{probability of a } 0 \text{ in the truth table output.} \]

Network and truth table are “annealed”. explain

Unstable if \( 2p(1-p) > 1/K \).

Others consider in-degree distribution \( P(K_{in}) \) with input nodes still chosen uniformly randomly [e.g., Aldana and Cluzel, PNAS (2003): \( P(K_{in}) \) scale-free].
Motivation for Our Work

Real networks are far from the model networks previously analyzed.

We would like to be able to analyze any fixed (i.e., not annealed) network, and we are interested in such effects as assortativity, community structure, different $p_i$'s at each node, nonsynchronous updates with heterogeneous link time delays.

We can treat all of these effects [Pomerance et al. PNAS (‘09)]!

In our analysis, we anneal the truth tables but not the network, which we hold fixed. (We call our technique ‘semi-annealing’).
Results of Our Theory

• We predict the average of the time asymptotic Hamming distance between two orbits originating from two initially close states.

• We obtain a simple general condition for the stability of the network dynamics:

\[ q_i = 2 \, p_i \, (1-p_i) \]
\[ Q_{ij} = q_i \, A_{ij} \]
\[ \lambda_Q = \text{Largest eigenvalue of the matrix } Q \text{ (this eigenvalue is real and positive by the Perron-Frobenius theorem since } Q_{ij} \geq 0 \text{).} \]

\[ \lambda_Q > 1 \text{ for instability,} \]
\[ \lambda_Q < 1 \text{ for stability.} \]
We compare results of our semi-annealed theory with frozen simulations. Issues we study include:

- Nodal in/out degree correlation
- Assortativity / disassortativity
- $p_i$ correlation with nodal degrees
- Differing time delays along different links
- Community structure
- Motifs
- Finite size effects

Experimentally determined network for yeast
An Example

\[ N = 10,000 \]

\[ P(K) \propto K^{-2.1} \]

for \( K < 15 \)

\[ q_i = q \Rightarrow \lambda_Q = q \lambda_A \]

\[ \langle K^{in} K^{out} \rangle_{node} \]

Nodal in/out degree correlations can be a big effect.

\[ \lambda_A \cong \langle K^{in} K^{out} \rangle_{node} / \langle K \rangle_{node} \]

- Solid line and arrows = semi-annealed theory.
- Symbols = numerical results for frozen case.
- Open symbols: different delay times.
Another Example

• Same as in previous slide but for three networks with different assortativity / disassortativity.
• Starting with a randomly constructed non-assortative network, the assortativity coefficient,

\[
\rho = \frac{\langle K^\text{out}_i K^\text{in}_j \rangle_{\text{edges } j \rightarrow i}}{\langle K^\text{in}_i K^\text{out}_j \rangle_{\text{node}}}
\]

is tuned by swapping edges:

• \( \lambda_Q = q \lambda_A \)

\[
\lambda_A \approx \frac{\langle K^\text{in}_i K^\text{out}_j \rangle_{\text{node}}}{\langle K \rangle_{\text{node}}} \rho
\]

Restrepo, Ott, Hunt, PhysRevE '07

Assortativity can be a big effect.
Network Instability and Cancer

• Tumor dissections show that nearby cells in cancer tumors are very heterogeneous and have widely different gene expression patterns. (This contrasts with normal tissue, where, e.g., cells in muscle are all similar.)

• We conjecture that this observed variability of nearby cells in the same tumor might be due to mutation-induced breakdown of the dynamical stability of the gene network.

• Ongoing, rapid experimental advances in gene network reconstruction may soon make tests and applications of this conjecture feasible.
Summary

• We introduce a ‘semi-annealing’ technique for studying the stability of discrete state gene network models.

• Our technique greatly expands the range of factors that can be treated, allowing treatment of arbitrary network topology, node-specific bias $p_i$, etc.

• We conjecture that gene network instability may be an important causal factor in some cancers.

Review of Derrida-Pomeau Method

(Basis of all previous work on stability.)

$K^{in}$ and $p$ are the same at each node.

Annealing: At each time-step

1. Truth table outputs are randomly chosen.

2. At each node the $K^{in}$ other nodes that input to that node are randomly chosen.

But we are really interested in the frozen case.

However, the annealed situation can be analyzed, but not the frozen case.

Assumption: For $N \gg 1$, (annealed results) $\sim$ (frozen results).
Our Technique: ‘Semi-Annealing’

Pomerance, Ott, Girvan, Losert, PNAS (2009)

The network is held fixed.

Each node $i$ has its own $p_i$ (bias to be on or off).

Semi-annealed: At each time step, only the truth table is randomly chosen, and this is done at each node $i$ according to the fixed bias probability $p_i$ for that node.

For $N \gg 1$, our numerical simulations show:

(semi-annealed) $\sim$ (frozen)


**Semi-Annealed Analysis**

- Consider two node states, $\sigma(t)$ and $\tilde{\sigma}(t)$, where $\sigma$ and $\tilde{\sigma}$ evolve from slightly different initial conditions.

- Let $y_i(t) = \text{(the probability that } \sigma_i(t) \text{ and } \tilde{\sigma}_i(t) \text{ differ)}$.

- Let $q_i = \text{[the probability that } \sigma_i(t) \text{ and } \tilde{\sigma}_i(t) \text{ differ, given that the } \sigma(t-1) \text{ and } \tilde{\sigma}(t-1) \text{ inputs to } i \text{ differ]}
  
  \[ q_i = 1 - [p_i^2 + (1 - p_i)^2] = 2p_i(1 - p_i) \]
Update Truth Table

<table>
<thead>
<tr>
<th>current state at time $t$ (input to gene $i$)</th>
<th>State of gene $i$ at time $t+1$ (output)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene $j_1$</td>
<td>Gene $j_2$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$K_i^{\text{in}} = 2$

![Diagram](image-url)
Update of Nodal Probability $y_i(t)$

The probability that the inputs at $t-1$ to $i$ are not all the same is given by:

$$y_i(t) = q_i \left\{ 1 - \prod_{j, A_{ij}=1} \left[ 1 - y_j(t-1) \right] \right\}$$

The probability that inputs are all the same perturbs around $\sigma = \tilde{\sigma}$ (i.e., $y_i << 1$); linearization gives:

$$y_i(t) \approx q_i \sum_{j=1}^{N} A_{ij} y_j(t-1) = \sum_{j=1}^{N} Q_{ij} y_j(t-1), \text{ or } y(t) = \underline{Q} y(t-1)$$

where $Q_{ij} = q_i A_{ij} = "\text{modified adjacency matrix}"$. 19
Stability

\[ y_i(t) = \sum_{j} Q_{ij} y_j(t - 1), \quad Q_{ij} = q_i A_{ij} \]

\[ \lambda_Q = \text{largest eigenvalue of } Q, \text{ which, according to the Perron-Frobenius theorem, is real and positive } (Q_{ij} \geq 0). \]

\[ \lambda_Q < 1, \; y = 0 \text{ is stable,} \]

\[ \lambda_Q > 1, \; y = 0 \text{ is unstable,} \]

\[ \lambda_Q = 1, \; y = 0 \text{ is "edge of chaos".} \]
Numerical Tests

We show tests of our

(i) \( \lambda_Q \) stability criterion,

(ii) Predicted average saturated normalized Hamming distance between \( \sigma \) and \( \bar{\sigma} \):

\[
\bar{y} = \lim_{t \to \infty} \frac{1}{N} \sum_i y_i(t).
\]

(re - show slide #11)
Estimating the network and the node bias probabilities $p_i$ from data

- **Network**: Undirected links can be inferred from data by looking at co-expression patterns across a range of perturbation experiments, and other techniques can determine directed links.

- The nodal bias probabilities $p_i$ can be estimated from clinical expression data.
A Possible Strategy for Cancer Treatment

• Our cancer/network-stability hypothesis combined with our stability analysis suggests a possible cancer therapy strategy: Namely, design drugs that target those genes or links whose disabling would most reduce $\lambda_Q$. 
The Perron-Frobenius Eigenvalue

• Recall that, since $A_{ij}$ and $Q_{ij}$ are non-negative, their stability - determining eigenvalue of maximum magnitude is real and positive (Perron - Frobenius Theorem).

• "Markov" theory for $\lambda_A$, the Perron - Frobenius eigenvalue (Note: $\lambda_Q = q \lambda_A$ if $q_i = q$ is the same on all nodes.)


E.g., $\lambda_A \approx \frac{\langle K^{in} K^{out} \rangle_{nodes}}{\langle K \rangle_{nodes}} \rho$
Perron-Frobenius Eigenvalue (cont’d)

- Markov theory for $\lambda_Q$ with $Q_{ij} = q_i A_{ij}$:


  E.g., if $q_i$ is correlated with $K_i^{in} K_i^{out}$,

  $$\lambda_Q \approx \frac{\langle q K^{in} K^{out} \rangle_{\text{nodes}}}{\langle K \rangle_{\text{nodes}}}$$
Finite Size Effects

\[ \varepsilon = \text{fraction of initially flipped bits.} \]
\[ \text{no. of initially flipped bits} = \varepsilon N \]

Apparently we need \( \varepsilon N \geq 5 \).
Gene expression and regulation

Transcriptional regulation: Proteins called transcription factors bind to specific sequences of the DNA to help determine whether or not an individual gene produces its mRNA and the subsequent protein which, in turn, may then regulate another gene by binding to its associated DNA sequence.