Persistence of activity in random Boolean networks

Shirshendu Chatterjee

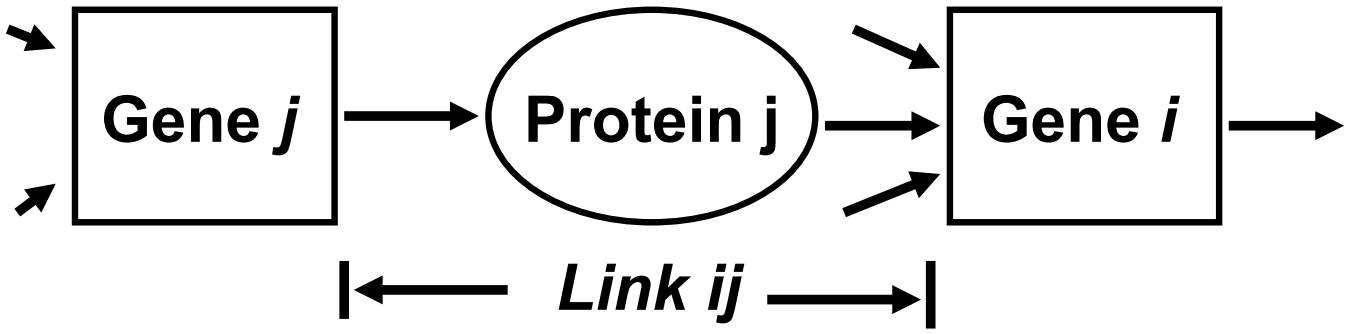
Courant Institute, NYU

Joint work with **Rick Durrett**

October 2012

The process of gene regulation

Transcriptional regulation: Proteins called transcription factors bind to specific sequences of the DNA to help or hinder the transcription of individual genes



The Result:
A complex web of interactions

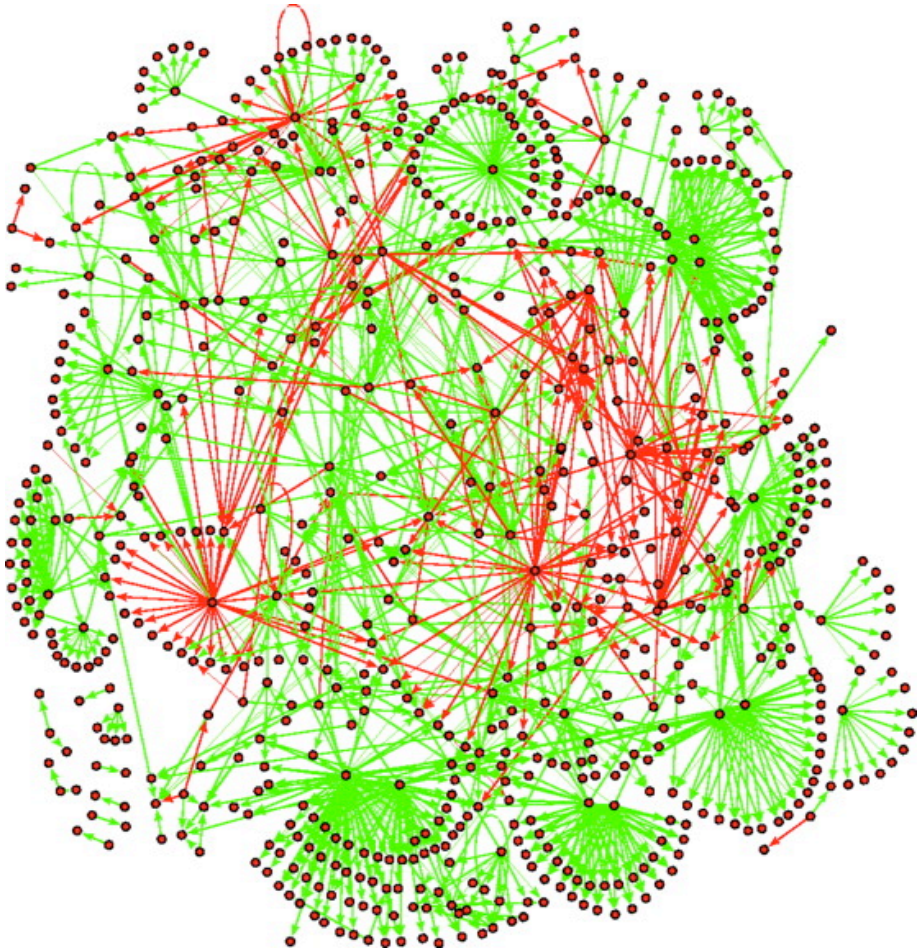
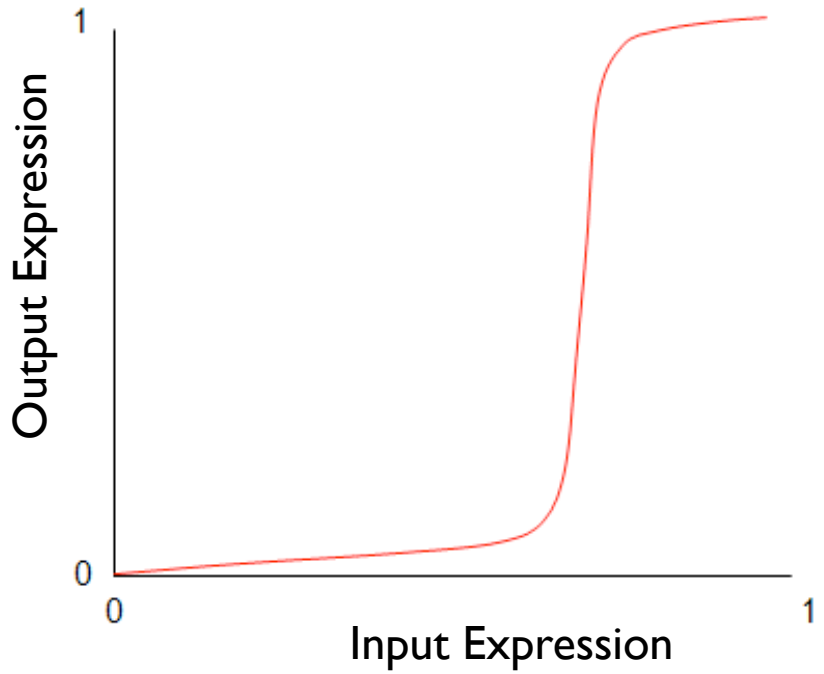


Figure taken from http://rsif.royalsocietypublishing.org/content/5/Suppl_1/S85.full

Building a simple model for gene regulation: Why Boolean?



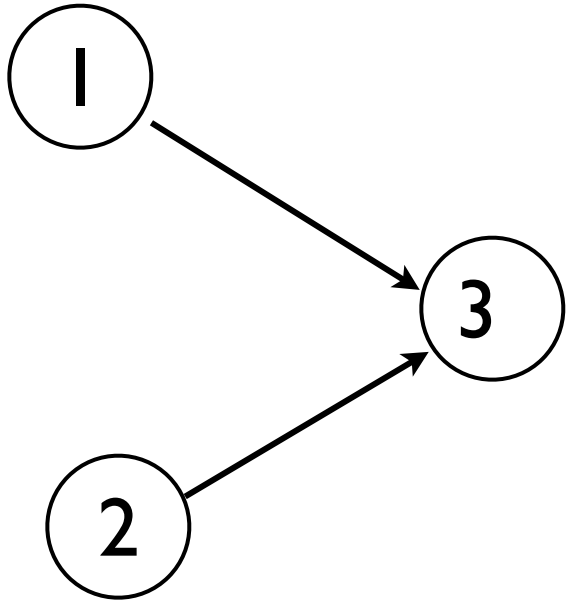
Input/output regulatory relationships between genes are observed to be strongly sigmoidal and well approximated by step functions.

Boolean network models

- Originally developed by Kauffman (1969) to model genetic regulatory networks.
- Protein and RNA concentrations in networks are often modeled using system of differential equations. But
  - the number of parameters can be huge.
  - Boolean networks sometimes outperform the diff. eq. models.
- Random boolean networks have been recently used to model
  - yeast transcriptional network,
  - yeast cell-cycle network.

Local update rules: An example

current state time t		State of gene 3 at t+1
Gene 1	Gene 2	
0	0	0
0	1	0
1	0	1
1	1	0



Node with 2 inputs

Describing the Boolean Network

- Network topology:

$$A_{ij} = \begin{cases} 1 \text{ if link from } j \to i \\ 0 \text{ otherwise} \end{cases}$$

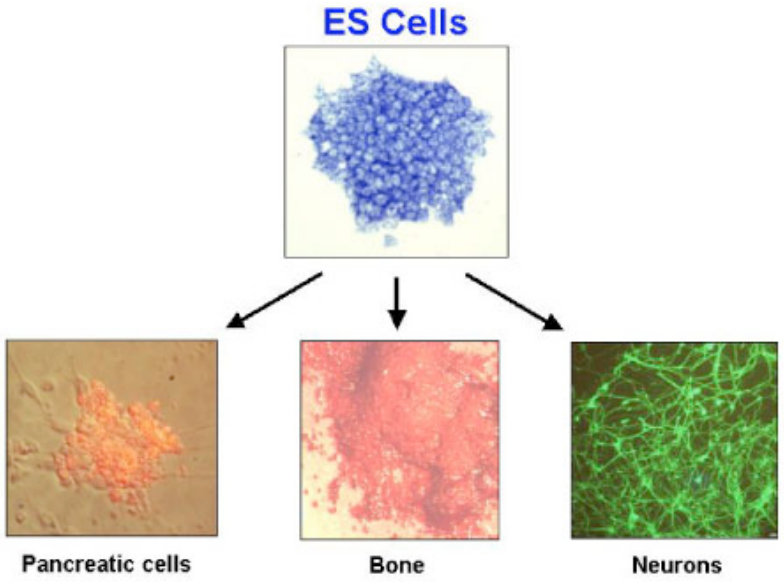
- Update functions:
  - Output rows randomly filled in
  - Bias - probability of a 0 appearing in the output row

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.

Significance of the attractors

- The attractors may represent a specific pattern of protein expression that define a cell's character
- In single celled organisms this could be different cell states: growing, dividing, starving, etc.
- In multicellular organisms these could correspond to different cell types.



Our random boolean network

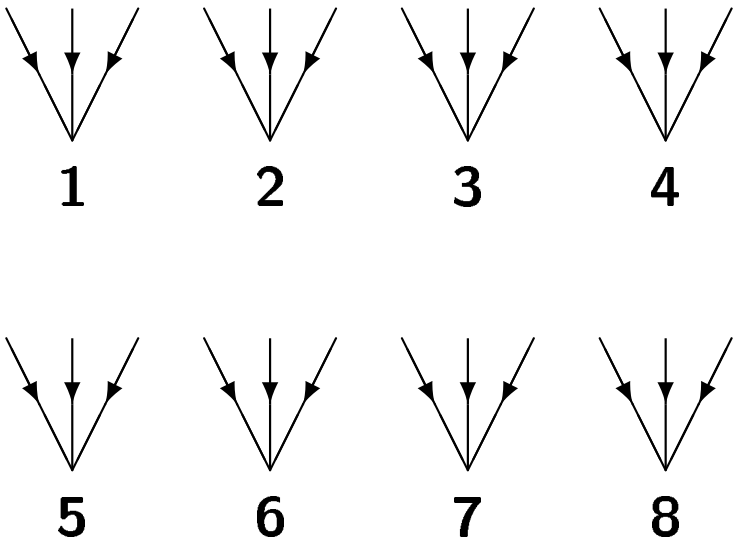
- n nodes, $V_n = \{1, 2, \ldots, n\}$.
- Each node x has r distinct input nodes $y_1(x), y_2(x), \ldots, y_r(x)$ chosen randomly from $V_n \setminus \{x\}$.
- Consider the directed graph $G_n = (V_n, E_n)$, where the edge set

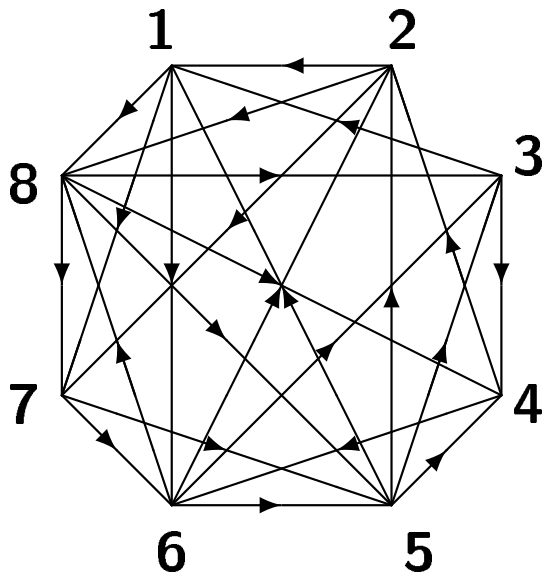
$$E_n = \{(y_i(x), x) : x \in V_n, 1 \leq i \leq r\}.$$

Construction:

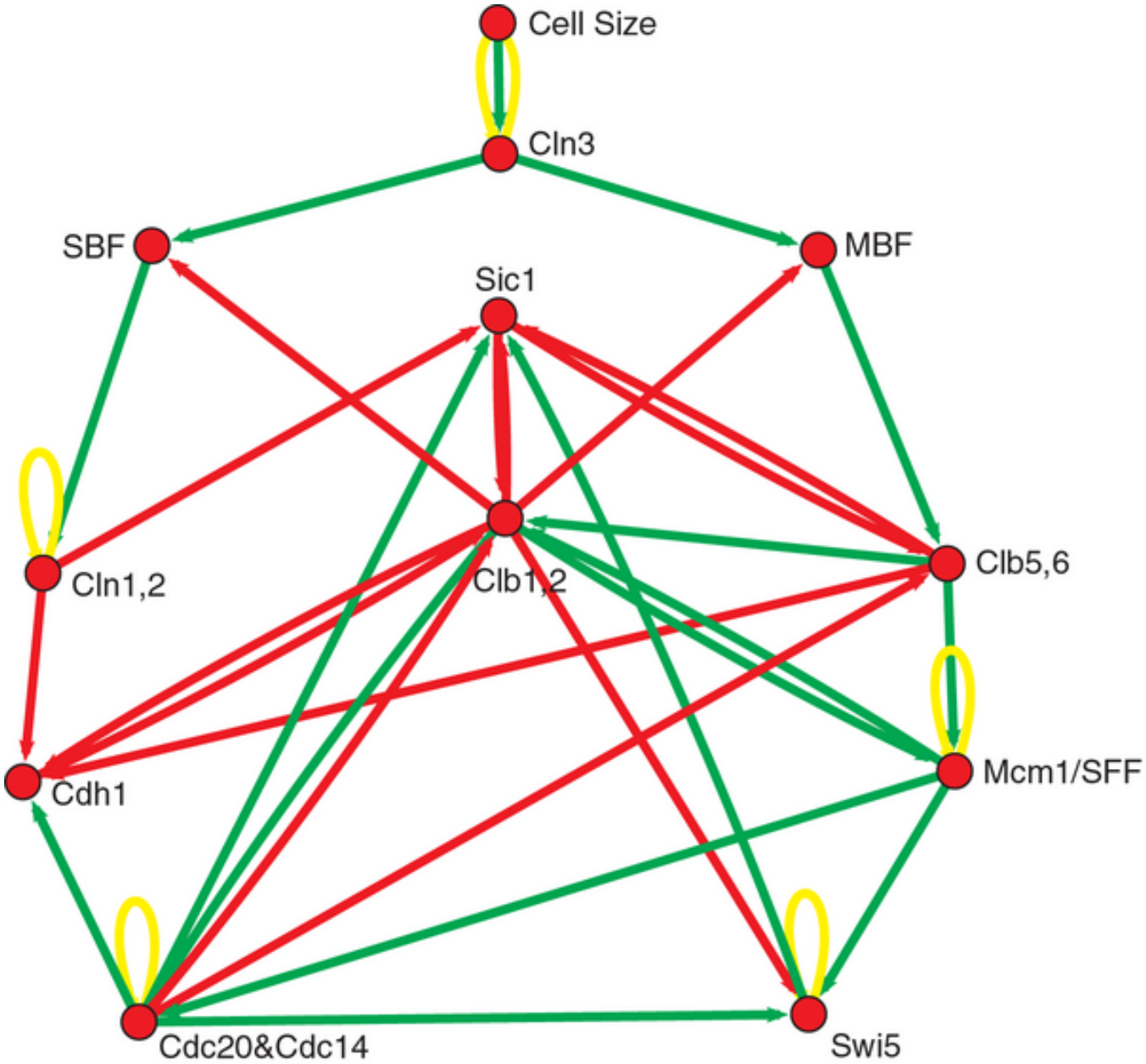
- Allocate r oriented half-edges to each node pointing to it.
- Pair the half-edges and the nodes uniformly at random.

Picture of the Construction





These are cartoons of regulatory networks



Dynamics

Discrete time system, $\eta_t(x) \in \{0,1\}$, $t = 0, 1, \ldots$.

- Each site x has a random function $\phi_x : \{0,1\}^r \to \{0,1\}$ where the values are independent and each equals 0 with probability p.
- The updates are

$$\eta_{t+1}(x) = \phi_x(\eta_t(y_1(x)), \ldots \eta_t(y_r(x))).$$

Simulation Studies: Derrida and Pomeau (1986)

- Phase transition curve is $r \cdot 2p(1-p) = 1$.
- Below the curve 'ordered' behavior – rapid convergence to a fixed point
- Above the curve 'chaotic' behavior – exponentially prolonged persistence of changes.

Simplified Problem

Consider another process $\zeta_t(x)$ for $t \geq 1$. $\zeta_t(x) = 1$ if $\eta_t(x) \neq \eta_{t-1}(x)$.

Approximate dynamics: A threshold contact process in which $\zeta_{t+1}(x) = 1$ with probability $q = 2p(1-p)$ if

$$\max_{1 \leq i \leq r} \zeta_t(y_i(x)) = 1$$

and $\zeta_{t+1}(x) = 0$ otherwise.

Reason: If the status of at least one of the inputs has changed, then the new value will be different from the old with probability $q = 2p(1-p)$.

Conjecture. *If $r \geq 3$, prolonged persistence if $qr > 1$.*

Chaos is bad news for a gene regulatory network. Stuart Kauffman argues that they evolve to the edge of chaos.

Intuition behind the conjecture

Let ξ_t be the set valued process $\xi_t = \{x : \zeta_t(x) = 1\}$.

- Let the dual graph $\hat{G}_n = (V_n, \hat{E}_n)$, where $\hat{E}_n$ is obtained from E_n by reversing arrows.
- Look at the time-dual process $\hat{\xi}_t$ on $\hat{G}_n$, where
  - $x \in \hat{\xi}_t$ implies

$$P[y_i(x) \in \hat{\xi}_{t+1} \forall i] = q = 1 - P[y_i(x) \notin \hat{\xi}^c_{t+1} \forall i].$$

  - Duality: $P(\xi^A_t \cap B \neq \emptyset) = P(\hat{\xi}^B_t \cap A \neq \emptyset)$.
  - Prolonged persistence of the two process are equivalent.
- The dual $\hat{x}i_t$ is like a *branching process* on $\hat{G}_n$.
  - Within short distances ($< (1/4)$ the diameter) of a vertex $\hat{G}_n$ is essentially a directed $r-$tree.
  - Positive probability of survival when mean offspring number $qr > 1$.

Prolonged persistence when $q > 1/(r-1)$

Recall $q = 2p(1-p)$.

Let ρ be the survival probability for the branching process with offspring distribution $q\delta_r + (1-q)\delta_0$.

Theorem. (C. and Durrett (2010) *If $q(r-1) > 1$, then the threshold contact process on G_n persists for time $O(e^{\gamma n})$ with high probability. The fraction of 1's in the quasi-stationary density $\geq \rho - \delta$ for any $\delta > 0$ with high probability.*

Sketch of the proof: Consider $\hat{G}_n$.

- Let $A^* = \{y : x \to y \text{ for some } x \in A\}$ in $\hat{G}_n$. This is not the boundary since we may have $y \in A$.
- **Isoperimetric inequality:** If ϵ is small, then with high prob, $|A^*| \geq (r-1-\epsilon)|A|$ for all A with $|A| \leq n\epsilon$.
- If $q(r-1) > 1$ and ϵ is small, then whenever $|\hat{\xi}_t| \leq n\epsilon$, $|\hat{\xi}_{t+1}| > |\hat{\xi}_t|$ with probability $\geq 1 - e^{-\gamma n}$.

Remarks: $r-1$ is sharp.

If $q(r-1) < 1$ there are exponentially many bad small sets. If $\hat{\xi}_t$ is bad, then $|\hat{\xi}_{t+1}| < |\hat{\xi}_t|$ with high probability.

Prolonged persistence when $q > 1/r$

Theorem. *Same conclusion holds*
We proved a weaker bound for persistence time, which has recently been improved by Mountford and Valesin (arxiv).

Sketch of the proof:

- One needs to look at more than one generation.
- For suitable choices of $\epsilon > 0$ and $g \geq 1$, if $|\hat{\xi}_t| \leq \epsilon n$ for some $t > 0$, then $|\hat{\xi}_{t+g}| > \epsilon n$ with probability $\geq 1 - e^{-\gamma n}$.
- The above estimate gives exponential persistence.
- The argument for density of 1's in the quasi-stationary density is the same.

Work in progress : What happens i we consider more complex networks etc.

Thank you