Phase transition for the threshold contact process, an "annealed approximation" of heterogeneous random Boolean networks



Shirshendu Chatterjee

City University of New York, City College

S. Chatterjee (CUNY) Random Boolean Networks

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

S. Chatterjee (CUNY) Random Boolean Networks



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

Boolean network models

- Protein and RNA concentrations in networks are often modeled using system of differential equations. But
 - for large networks the number of parameters (e.g. production rates, interaction strengths, decay rates) can be huge.
 - Boolean network models sometimes do better than Diff. Eq. models:
 - Albert et. al., J.Theor.Bio. (2003)
 - Chaves et. al., J.Theor.Bio. (2005)

show this for segment polarity network in Drosophila.

- Random boolean networks have been recently used to model
 - yeast transcriptional network (Kauffman et. al., PNAS (2003)),
 - yeast cell-cycle network (Li et. al., PNAS (2004)) .

Basic framework

- Genes = nodes of a directed graph G_n .
- Vertex set $V_n = \{1, 2, \dots, n\}$, n = number of genes.
- Edge set $E_n := \{ \langle u, v \rangle : u \text{ influences } v \}.$
- Once chosen, G_n remains fixed through time.

Basic framework

- Genes = nodes of a directed graph G_n .
- Vertex set $V_n = \{1, 2, \dots, n\}$, n = number of genes.
- Edge set $E_n := \{ \langle u, v \rangle : u \text{ influences } v \}.$
- Once chosen, G_n remains fixed through time.
- The state of x at time t = 0, 1, 2, ... is $\eta_t(x) \in \{1(\text{`on'}), 0(\text{`off'})\}.$
- $\langle u, v \rangle \in E_n$ means u is an input for v. $Y^x =$ input set for x.
- Update rule: For time-independent Boolean functions $\phi_x: \{0,1\}^{|Y^x|} \to \{0,1\}$,

$$\eta_{t+1}(x) = \phi_x((\eta_t(y), y \in Y^x)), \qquad x \in V_n.$$

Random Boolean network model

To understand general properties of such dynamical systems,

various random Boolean network models have been formulated.

- Originally developed by Kauffman (1969).
 - Each node has r inputs, which are uniformly chosen
 - The values $\{\phi_x(\mathbf{y})\}$ are independent coin flips.

Local update rules: An example



w

Many studies using simulation and heuristics.

Derrida and Pomeau (1986) studied the model under additional assumption:

Network changes at every step. The input sets $\{Y^x\}_{x\in[n]}$ and the values $\{f_x(\mathbf{v}) : x \in [n], \mathbf{v} \in \{0, 1\}^r\}$ are resampled at each time step.

Then, if $\eta_0, \tilde{\eta}_0 \in \{0, 1\}^n$ are two initial configurations and $n^{-1}d_{Ham}(\eta_0, \tilde{\eta}_0) = x \in [0, 1]$, then after one time step

$$d_{\mathit{Ham}}(\eta_1, ilde\eta_1)\sim \mathit{Binomial}\left(n,2p(1-p)[1-(1-x)^r]
ight).$$

One can iterate the above to have

$$d_{Ham}(oldsymbol{\eta}_{t+1}, oldsymbol{ ilde\eta}_{t+1}) \sim Binomial\left(n, 2p(1-p)[1-(1-y_t)^r]
ight),$$
here $y_t = n^{-1} d_{Ham}(oldsymbol{\eta}_t, oldsymbol{ ilde\eta}_t).$

This in turn implies

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{d_{Ham}(\boldsymbol{\eta}_t,\tilde{\boldsymbol{\eta}}_t)}{n}=y^*,$$

where y^* is a fixed point (with x in its basin of attraction) of the map $y \mapsto \varphi(y) := 2p(1-p)(1-(1-y)^r)$. If 2p(1-p)r < 1, then 0 is the only fixed point of φ . It is *stable*.

If 2p(1-p)r > 1, then φ has another fixed point $y^* \in (0,1)$ which is stable, and 0 is an unstable fixed point.

Based on this observation, Derrida and Pomeau argued that the *order-chaos* phase transition curve for Kauffman's Random Boolean Network is given by

$$2p(1-p)\cdot r=1.$$

Extension of DP('86) to complex networks

Empirical evidence suggests that gene network may have

- heterogeneous degree distribution
- correlation between in-degree and out-degree,
- community structure

Extension of DP('86) to complex networks

Empirical evidence suggests that gene network may have

- heterogeneous degree distribution
- correlation between in-degree and out-degree,
- community structure

In the physics and biology literature people have considered similar evolution of Boolean functions on complex network models with

prescribed in-degree distribution pⁱⁿ = (pⁱⁿ_k)_{k≥1}: Here the order-chaos phase transition curve for has been claimed to be

$$2p(1-p) \cdot r^{in} = 1$$
, where r^{in} is average in-degree.

Extension of DP('86) to complex networks

Empirical evidence suggests that gene network may have

- heterogeneous degree distribution
- correlation between in-degree and out-degree,
- community structure

In the physics and biology literature people have considered similar evolution of Boolean functions on complex network models with

prescribed in-degree distribution pⁱⁿ = (pⁱⁿ_k)_{k≥1}: Here the order-chaos phase transition curve for has been claimed to be

$$2p(1-p) \cdot r^{in} = 1$$
, where r^{in} is average in-degree.

prescribed joint distribution of in-degree and out-degree
 p^{in,out} = (p^{in,out}_{k,l})_{k,l≥1}: Here the order-chaos phase transition curve for has been claimed to be

$$2p(1-p) \cdot \frac{r^{in,out}}{r^{in}} = 1$$
, where r^{in} is average in-degree

and $r^{in,out}$ is the average in-degree × out-degree.

Construction of *G_n* with in-degree distribution

• In-degree distribution: $\mathbf{p}^{in} = \{p_k^{in}\}_{k \ge 1}.$

- In-degree distribution: $\mathbf{p}^{in} = \{p_k^{in}\}_{k \ge 1}.$
- Choose I_1, I_2, \ldots, I_n *i.i.d.* from \mathbf{p}^{in} .

Construction of G_n with in-degree distribution



- In-degree distribution: $\mathbf{p}^{in} = \{p_k^{in}\}_{k \ge 1}.$
- Choose I_1, I_2, \ldots, I_n *i.i.d.* from \mathbf{p}^{in} .
- Allocate I_x oriented half-edges to node x pointing to it.



• In-degree distribution:

$$\mathbf{p}^{in} = \{p_k^{in}\}_{k \ge 1}$$
.
• Choose l_1, l_2, \dots, l_n *i.i.d.* from \mathbf{p}^{in} .
• Allocate l_x oriented half-edges to
node x pointing to it.
• Choose the input set
 $Y^x = \{Y_1(x), \dots, Y_{l_x}(x)\}$ uniformly
at random from $V_n \setminus \{x\}$



- In-degree distribution: $\mathbf{p}^{in} = \{p_k^{in}\}_{k \ge 1}$.
- Choose I_1, I_2, \ldots, I_n *i.i.d.* from \mathbf{p}^{in} .
- Allocate *I_x* oriented half-edges to node *x* pointing to it.
- Choose the input set $Y^{x} = \{Y_{1}(x), \dots, Y_{l_{x}}(x)\}$ uniformly at random from $V_{n} \setminus \{x\}$.





- In-degree distribution: $\mathbf{p}^{in} = \{p_k^{in}\}_{k \ge 1}.$
- Choose I_1, I_2, \ldots, I_n *i.i.d.* from \mathbf{p}^{in} .
- Allocate I_x oriented half-edges to node x pointing to it.
- Choose the input set $Y^x = \{Y_1(x), \dots, Y_{I_x}(x)\}$ uniformly at random from $V_n \setminus \{x\}$.





- In-degree distribution: $\mathbf{p}^{in} = \{p_k^{in}\}_{k \ge 1}.$
- Choose I_1, I_2, \ldots, I_n *i.i.d.* from \mathbf{p}^{in} .
- Allocate I_x oriented half-edges to node x pointing to it.
- Choose the input set $Y^x = \{Y_1(x), \dots, Y_{I_x}(x)\}$ uniformly at random from $V_n \setminus \{x\}$.





- In-degree distribution: $\mathbf{p}^{in} = \{p_k^{in}\}_{k \ge 1}.$
- Choose I_1, I_2, \ldots, I_n *i.i.d.* from \mathbf{p}^{in} .
- Allocate *I_x* oriented half-edges to node *x* pointing to it.
- Choose the input set $Y^x = \{Y_1(x), \dots, Y_{I_x}(x)\}$ uniformly at random from $V_n \setminus \{x\}$.



 Joint degree distribution: p^{in,out}, marginals have equal mean.

- Joint degree distribution: p^{in,out}, marginals have equal mean.
- Choose $\{(I_x, O_x)\}_{x=1}^n$ *i.i.d.* from $\mathbf{p}^{\text{in,out}}$.

Construction of *G_n* with (in/out)-degree distribution



- Joint degree distribution: p^{in,out}, marginals have equal mean.
- Choose $\{(I_x, O_x)\}_{x=1}^n$ *i.i.d.* from $\mathbf{p}^{\text{in,out}}$.
- Need total in-degree = total out-degree.

Construction of *G_n* with (in/out)-degree distribution



- Joint degree distribution: p^{in,out}, marginals have equal mean.
- Choose $\{(I_x, O_x)\}_{x=1}^n$ *i.i.d.* from $\mathbf{p}^{\text{in,out}}$.
- Need total in-degree = total out-degree.
- Allocate *I_x* and *O_x* oriented half-edges to *x* pointing inward and outward resp.

Construction of *G_n* with (in/out)-degree distribution



- Joint degree distribution: p^{in,out}, marginals have equal mean.
- Choose $\{(I_x, O_x)\}_{x=1}^n$ *i.i.d.* from $\mathbf{p}^{\text{in,out}}$.
- Need total in-degree = total out-degree.
- Allocate *I_x* and *O_x* oriented half-edges to *x* pointing inward and outward resp.
- Choose a matching uniformly at random between the sets of inward and outward half-edges.





Dynamics

Once G_n is chosen let Y^x be the input set of node x.

Recall: $\eta_{t+1}(x) = \phi_x((\eta_t(y), y \in Y^x)).$

Adopting Kauffman's model, the values $\phi_x(\mathbf{v})$ are 1 or 0 with probability p and 1 - p.

Proving rigorous results about the formulated discrete dynamical systems turns out to be quite hard.

We will consider a slight modification of the dynamical system, which has been called "Annealed Approximation" in the physics literature.

The special case of our results, where the in-degree is fixed (not random), was considered earlier – C. & Durrett (2011), Montford & Valesin (2013).

An Example

Fix node $x \in V_n$.

In LHS figure the states of $\{Y_i^x\}_{i=1}^4$ is the same in time t-1 and time t. Hence $\eta_t(x) = \eta_{t+1}(x)$.

But in RHS figure the state of Y_4^x is different in time t - 1 and time t. Hence $\eta_t(x) \neq \eta_{t+1}(x)$ with probability 2p(1-p), as both values of f_x are independent biased coin flips.



Consider another process
$$\{\zeta_t(x)\}_{t\geq 1}$$
, where $\zeta_t(x) = \mathbf{1}_{\{\eta_t(x)\neq\eta_{t-1}(x)\}}$.

Then approximate dynamics for ζ_t : A *threshold contact process* in which

$$\mathbb{P}\left(\left.\zeta_{t+1}(x)=1
ight|\left.\zeta_t(y)=1
ight.$$
 for at least one $y\in Y^{ imes}
ight)=2p(1-p),$

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

Let $\zeta_t, t \geq 1$, be the threshold contact process with parameter q = 2p(1-p).

Let $\zeta_t, t \ge 1$, be the threshold contact process with parameter q = 2p(1-p).

Persistence time for ζ_t is the time to reach all-zero configuration.

Let $\zeta_t, t \ge 1$, be the threshold contact process with parameter q = 2p(1-p).

Persistence time for ζ_t is the time to reach all-zero configuration.

For given in-degree distribution \mathbf{p}^{in} with finite second moment and $p_0^{\text{in}} = p_1^{\text{in}} = 0$, let r^{in} be the mean of \mathbf{p}^{in} , and \mathbb{P} be the law of G_n .

Let $\zeta_t, t \ge 1$, be the threshold contact process with parameter q = 2p(1-p).

Persistence time for ζ_t is the time to reach all-zero configuration.

For given in-degree distribution \mathbf{p}^{in} with finite second moment and $p_0^{\text{in}} = p_1^{\text{in}} = 0$, let r^{in} be the mean of \mathbf{p}^{in} , and \mathbb{P} be the law of G_n .

Theorem (C., PTRF 2015)

There is a set of 'good graphs' \mathcal{G}_n with $\mathbb{P}(\mathcal{G}_n) = 1 - o(1)$ such that for any $G_n \in \mathcal{G}_n$

- if $q \cdot r^{in} < 1$, then ζ_t persists for time $O(\log n)$.
- if $q \cdot r^{in} > 1$, then
 - ζ_t persists for time $O(e^{cn})$.
 - the density of 1's in $\zeta_t,$ starting from all-one at the beginning, stays close to

 $\pi = \pi(q, \mathbf{p}^{in})$, the survival probability for a branching process

with offspring distribution $(1-q)\delta_0 + q {f p}^{in}.$

For given in-degree/out-degree joint distribution $\mathbf{p}^{\text{in,out}}$ with finite second moment and $p_0^{\text{in}} = p_1^{\text{in}} = 0$, let r^{in} be the mean of \mathbf{p}^{in} , $r^{\text{in,out}}$ be the mean of in-degree \times out-degree, and \mathbb{P} be the law of G_n .

For given in-degree/out-degree joint distribution $\mathbf{p}^{\text{in,out}}$ with finite second moment and $p_0^{\text{in}} = p_1^{\text{in}} = 0$, let r^{in} be the mean of \mathbf{p}^{in} , $r^{\text{in,out}}$ be the mean of in-degree \times out-degree, and \mathbb{P} be the law of G_n .

Theorem (C., PTRF 2015)

There is a set of 'good graphs' \mathcal{G}_n with $\mathbb{P}(\mathcal{G}_n) = 1 - o(1)$ such that for any $G_n \in \mathcal{G}_n$

• if $q \cdot \frac{r^{in,out}}{r^{in}} < 1$, then ζ_t persists for time $O(\log n)$.

• if
$$q \cdot rac{r^{in,out}}{r^{in}} > 1$$
, then

- ζ_t persists for time $O(e^{cn})$.
- the density of 1's in $\zeta_t,$ starting from all-one at the beginning, stays close to

 $\pi = \pi(q, \mathbf{p}^{\textit{in,out}}),$ the survival probability for a branching process

with offspring distribution $(1-q)\delta_0 + q \tilde{\mathbf{p}}^{in}$,

where $\tilde{\boldsymbol{p}}^{\text{in}}$ is the size biased version of $\boldsymbol{p}^{\text{in}}.$

Intuition behind the results

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.
- The time-dual process (say $\hat{\xi}_t$) on \hat{G}_n satisfies
 - $x \in \hat{\xi}_t$ implies

$$P[Y^{ imes} \subset \hat{\xi}_{t+1}] = q = 1 - P[Y^{ imes} \subset (\hat{\xi}_{t+1})^c].$$

• Prolonged persistence of the two process are equivalent.

- The dual $\hat{\xi}_t$ behaves locally like a *branching process* on \hat{G}_n .
 - Let r be the average out-degree in a 'small' neighborhood of a typical vertex in \hat{G}_n .
 - Positive probability of survival when mean offspring number qr > 1.
Graphical representation. Let $\xi_t^A := \{x : \zeta_t(x) = 1\}$ if $A = \{y : \zeta_0(y) = 1\}.$



















Look at the process going backward in time.

Look at the process going backward in time.



Space

Look at the process going backward in time.

It is a *coalescing branching process* $\hat{\xi}_t$ on \hat{G}_n . Node x gives birth at time t to $|Y^x|$ many children at time t - 1 if the gadget is present.



Space

Look at the process going backward in time.

It is a coalescing branching process $\hat{\xi}_t$ on \hat{G}_n . Node x gives birth at time t to $|Y^x|$ many children at time t-1 if the gadget is present. Duality: Prolomged persistence of $\xi_t \Leftrightarrow$ Survival for $\hat{\xi}_t$.



Fix $A, B \subset V_n$.

Fix $A, B \subset V_n$.



Fix $A, B \subset V_n$. Duality relation between $\{\xi_t\}$ and $\{\hat{\xi}_t\}$: $P(\xi_t^A \cap B \neq \emptyset) = P(\hat{\xi}_t^B \cap A \neq \emptyset)$.



Fix $A, B \subset V_n$. Duality relation between $\{\xi_t\}$ and $\{\hat{\xi}_t\}$: $P(\xi_t^A \cap B \neq \emptyset) = P(\hat{\xi}_t^B \cap A \neq \emptyset)$. Putting $A = V_n, B = \{x\}, P(x \in \xi_t^{V_n}) = P(\hat{\xi}_t^{\{x\}} \neq \emptyset)$.



Fix $A, B \subset V_n$. Duality relation between $\{\xi_t\}$ and $\{\hat{\xi}_t\}$: $P(\xi_t^A \cap B \neq \emptyset) = P(\hat{\xi}_t^B \cap A \neq \emptyset)$. Putting $A = V_n, B = \{x\}, P(x \in \xi_t^{V_n}) = P(\hat{\xi}_t^{\{x\}} \neq \emptyset)$. So probability of prolonged persistence for ξ_t is same as the survival probability for $\hat{\xi}_t$.



Here, we get an upper bound for the size of the coalescing branching process on \hat{G}_n by coupling with a subcritical branching process.

This, together with soft argument, gives *logarithmic persistence* for the coalescing branching process.

The argument has mainly two parts:

- Survival of the coalescing branching process starting from singletons this gives information about the quasi-stationary density of 1's in the threshold contact process.
- Survival of the coalescing branching process starting from large sets this gives information about the persistence time for the threshold contact process.







Fix $x \in V_n$. If no collision occurs in the neighborhood of x, then $\hat{\xi}_t^{\{x\}}$ is a branching process.



Fix $x \in V_n$. If no collision occurs in the neighborhood of x, then $\hat{\xi}_t^{\{x\}}$ is a branching process.

If Z_t is a branching process with mean offspring μ and $B = \{Z_t \text{ survives }\}, \text{ then }$

$$P(Z_{t+1} < (\mu - \delta)Z_t|B) \le e^{-c(\delta)t}.$$



Fix $x \in V_n$. If no collision occurs in the neighborhood of x, then $\hat{\xi}_{\tau}^{\{x\}}$ is a branching process.

If Z_t is a branching process with mean offspring μ and $B = \{Z_t \text{ survives }\}$, then

$$P(Z_{t+1} < (\mu - \delta)Z_t|B) \le e^{-c(\delta)t}$$

$$P\left(\left|\hat{\xi}_{O(\log\log n)}^{\{x\}}\right| \geq (\log n)^a\right) \approx \pi.$$



Suppose $A \subset [n]$. Use a coupling between the directed *k*-neighborhood of *A* and a forest $\{Z_t^A, 0 \leq t \leq k\}$ along with partitions Z_t^A into C_t^A (closed), O_t^A (open), and R_t^A (removed). We choose $Z_0^A = A$ and $C_0^A = R_0^A = \emptyset$, so $O_0^A = A$. For each $t \geq 0$, every site of O_t^A mimics the corresponding vertex in G_n We scan the sites of Z_{t+1}^A in an increasing order, and define a site

to be

- removed if more than 2r children of the same ancestor has been scanned
- closed if collision occurs
- open otherwise

For $u \in Z_t^A$, let $\overleftarrow{u}_t^A \in A$ denote the unique root-ancestor of u. For any subset $B \subset A$ and $t \ge 1$ let $Z_0^{A,B} = O_0^{A,B} = B$ and

$$Z_t^{A,B} := \{ u \in Z_t^A : \overleftarrow{u}_t^A \in B \}$$
$$C_t^{A,B} := Z_t^{A,B} \cap C_t^A, \quad O_t^{A,B} := Z_t^{A,B} \cap O_t^A.$$





S. Chatterjee (CUNY) Random Boolean Networks

Ingredients

We will use the following LD estimate for the number of open nodes.

Let $X \ge 0$ be any random variable such that $2(EX)^2 \le EX^2 < \infty$. Then $\log Ee^{-tX} \le var(X)t^2/2 - E(X)t$ for any t > 0.

Let F be a forest with m roots having $(f_i)_{i=1}^m$ children at level k.



Then, using above LD estimate, if $\sum_{i \in [m]} f_i \ge 2Cmq^{-k}$, then

$$\mathbb{P}(|\hat{\xi}_k^{F_0}| \leq Cm) \leq \exp\left(-cq^k(\sum_{i:[m]} f_i)^2 / \sum_{i:[m]} f_i^2\right)$$

Proof of exponential persistence

Suppose $B \subset A \subset [n]$ such that |A| = m and $|B| \ge (1 - \delta)|A|$. Let

$$k_m := \log_2[\Delta \log(n/m)], \Delta$$
 large, and $ho, \sigma > 0$

There is a collection of good graphs, such that if G_n is good, then

• for any vertex set A with $(\log n)^a \le |A| \le \varepsilon n$, there is $D \subset A$ satisfying

$$|D| \geq \delta |A|, \quad |O_{
ho\sigma k_m}^{A,\{x\}}| \geq (4/\delta)q^{-
ho\sigma k_m} \ \forall \ x \in D.$$

• and hence letting F be the sub-forest of $\bigcup_{k=0}^{\rho\sigma k_m} O_k^{A,D}$ such that each root has $(4/\delta)q^{-\rho\sigma k_m}$ leaves, and applying the previous forest estimate

$$\mathbb{P}(|\hat{\xi}^{\mathcal{A}}_{\rho\sigma k_m}| > |\mathcal{A}|) \geq \mathbb{P}(|\hat{\xi}^{\mathcal{F}_0}_{\rho\sigma k_m}| > \frac{1}{\delta}|D|) \geq 1 - \exp(-cm/\log^b(n/m)).$$

Applying this estimate repeatedly, we get exponential persistence.

Choice of good graph

Let
$$E_{A,B} := \{|O_{\rho\sigma k_m}^{A,B}| \ge |B|(4/\delta)q^{-\rho\sigma k_m}\}$$
. $\eta > 0$ is small, so
 $\tilde{r} := r(1-\eta) > q^{-1}$.
 $H_k^{A,B} := \left\{\sum_{i=1}^{\rho} |C_{\rho(k-1)+i}^{A,B}| \le (1+5\delta)|O_{\rho(k-1)}^{A,B}|\right\}$,
 $L_{k,j}^{A,B} := \left\{|Z_{\rho(k-1)+j}^{A,B}| \ge \tilde{r}|O_{\rho(k-1)+j-1}^{A,B}|\right\}$ and $L_k^{A,B} := \bigcap_{j=1}^{\rho} L_{k,j}^{A,B}$.



 $\begin{array}{l} E_{A,B} \text{ occurs on } \cap_{k=1}^{\sigma k_m} H_k^{A,B} \cap_{k=k_m+1}^{\sigma k_m} L_k^{A,B}, \\ \text{ if } (m/n) \leq \varepsilon \text{ (small) and } \rho, \sigma \text{ are large.} \end{array} .$

$$\begin{split} \mathbb{P}\left(\left(H_{k}^{A,B}\right)^{c} \middle| |O_{\rho(k-1)}^{A,B}|\right) &\leq \exp\left(-(1+5\delta/2)|O_{\rho(k-1)}^{A,B}|\log(n/m)\right)\\ \mathbb{P}\left(\left(H_{k}^{A,B}\right)^{c} \cap \bigcap_{j=1}^{k-1} H_{j}^{A,B}\right) &\leq \exp\left(-(1+5\delta/2)|B|\log(n/m)\right)\\ &\leq \exp\left(-(1+\delta)m\log(n/m)\right).\\ |O_{\rho k_{m}}^{A,B}| &\geq 2^{k_{m}}|B| \text{ on } \bigcap_{1}^{k_{m}} H_{j}^{A,B}. \text{ For } k > k_{m},\\ \mathbb{P}\left(\left(L_{k,i}^{A,B}\right)^{c} \middle| |O_{\rho(k-1)+i-1}^{A,B}|\right) &\leq \exp\left(-|O_{\rho(k-1)+i-1}^{A,B}|I(\eta)\right),\\ \text{ where } I(\cdot) \text{ is the LD rate function.}\\ \mathbb{P}\left(\left(L_{k,i}^{A,B}\right)^{c} \cap_{j=1}^{i-1} L_{k,j}^{A,B} \cap_{j=1}^{k} H_{j}^{A,B}\right) &\leq \exp\left(-(\tilde{r}-(1+5\delta))2^{k_{m}}|B|I_{n}(\eta)\right)\\ &\leq \exp(-(1+3\delta)|B|\log(n/m)) &\leq \exp(-(1+\delta)m\log(n/m)).\\ \text{ Using union bound} \end{split}$$

$$\mathbb{P}\left(E_{A,B}^{c}\right) \leq Ck_{m} \exp(-(1+\delta)m\log(n/m))$$

$$\leq \exp(-(1+3\delta/4)m\log(n/m)).$$
S. Chatterjee (CUNY)
Random Boolean Networks

Choice of good graph (continued)

$$\mathbb{P}\left(E_{A,B}^{c}\right) \leq \exp(-(1+3\delta/4)m\log(n/m)).$$
$$E_{A} := \bigcap_{B \subset A: |B| \geq (1-\delta)|A|} \{|O_{\rho\sigma k_{m}}^{A,B}| \geq |B|(4/\delta)q^{-\rho\sigma k_{m}}\}.$$

Using union bound, if $(m/n) \leq \varepsilon$ (small), then

$$\mathbb{P}\left(\exists A \subset [n] : |A| = m, E_A^c \text{ occurs }\right) \leq \exp(-3\delta/8m\log(n/m)),$$

$$\mathbb{P}\left(\exists A \subset [n] : (\log n)^a \leq |A| \leq \varepsilon n, E_A^c \text{ occurs }\right)$$

$$\leq n \exp\left(-3\delta/8(\log n)^a \log \frac{n}{(\log n)^a}\right) = o(n^{-1/2}).$$

We choose our good graphs to be

$$\mathcal{G}_n := \{ E_A \text{ occurs } \forall A \subset [n] \text{ s.t. } (\log n)^a \le |A| \le \varepsilon n \}.$$

Let \hat{G}_n be the edge-reversed graph with out-degree distribution **p** and mean r.

Theorem (C., PTRF 2015)

If $2p(1-p) \cdot r > 1$, then the threshold contact process on G_n is supercritical. It persists for exponentially long. The density of 1's is $\approx \pi = \pi(p, \mathbf{p})$, the survival probability for a certain supercritical "branching process". If $2p(1-p) \cdot r < 1$, persistence time is logarithmic.
- rigorous results about the actual dynamical system
- Properties of the final configuration in the subcritical regime.
- Critical random Boolean networks: persistence time?
- Effect of node-dependent values of p?
- Other models of control, e.g. the "threshold model", where the ground graph is weighted and a gene is in state 1 at time t + 1 if the (signed) linear combination of its input weights at time t cross a threshold θ.

Thank you

S. Chatterjee (CUNY) Random Boolean Networks