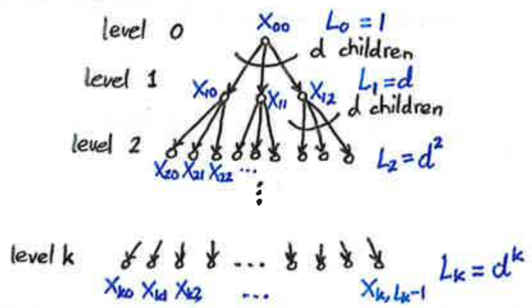
Anuran Makur
2 April 2018

Broadcasting on Bounded Degree DAGs: [Makur-Mossel-Polyanskiy 2018]

★ Broadcasting on Trees:

- Long line of work: [Bleher-Ruiz-Zagrebnov 1995], ..., [Evans-Kenyon-Peres-Schulman 2000], ...
  - ↳ statistical physics result for regular trees
  - ↳ broadcasting on general trees
  - ↳ many extensions
- Applications: phylogenetic reconstruction, random constraint satisfaction problems, etc.
- Model & Notation: (simple case of [EKPS 2000])

We are given an infinite d-ary tree. Let X_{kj} = node at the jth position in level k,
L_k = no. of nodes at level k,
$X_k = (X_{k0}, \ldots, X_{k,L_k-1})$ (i.e. all nodes at level k).



Each X_{kj} is a Bernoulli random variable.
Each edge is an independent $\mathrm{BSC}(\delta)$ with $0 < \delta < \frac{1}{2}$.

$$\{0,1\} \ni X \rightarrow \boxed{\mathrm{BSC}(\delta)} \rightarrow Y = \begin{cases} X & \text{wp } 1-2\delta \ [\text{copy}] \\ \mathrm{Ber}(\frac{1}{2}) & \text{wp } 2\delta \ [\text{ind. bit}] \end{cases} = \begin{cases} X & \text{wp } 1-\delta \\ 1-X & \text{wp } \delta \end{cases}$$

Let $X_{00} \sim \mathrm{Ber}(\frac{1}{2})$. This defines joint distribution of $\{X_{kj} : k \geq 0, 0 \leq j < L_k\}$.

- Broadcasting Question: Can we decode X_0 from X_k as $k \to \infty$?

$$X_0 \sim \mathrm{Ber}(\tfrac{1}{2}) \begin{cases} \rightarrow X_0 = 1 : P_{X_k|X_0=1} \\ \rightarrow X_0 = 0 : P_{X_k|X_0=0} \end{cases}$$

Hypothesis Testing: Use ML decoder for min. prob. of error.
$\hat{X}^k_{ML}(X_k)$ is the ML decoder.

- Thm: (Phase Transition)
  1) If $\underbrace{(1-2\delta)^2 d > 1}_{\delta < \frac{1}{2} - \frac{1}{2\sqrt{d}}}$, then $\underbrace{\lim_{k\to\infty} \mathbb{P}(\hat{X}^k_{ML}(X_k) \neq X_0) < \frac{1}{2}}_{\triangleq \text{ broadcast/reconstruction possible}}$.
  2) If $\underbrace{(1-2\delta)^2 d < 1}_{\delta > \frac{1}{2} - \frac{1}{2\sqrt{d}}}$, then $\underbrace{\lim_{k\to\infty} \mathbb{P}(\hat{X}^k_{ML}(X_k) \neq X_0) = \frac{1}{2}}_{\triangleq \text{ broadcast/reconstruction impossible}}$.

  critical threshold

Intuition:
- $(1-2\delta)^2$ is the contraction of mutual information of a $\mathrm{BSC}(\delta)$
- d is the "repetition code" factor
- Competition between these forces

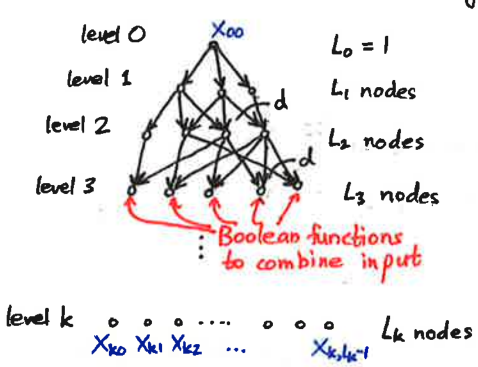
- Observation: If $d \leq 1$, then broadcast impossible. So, if L_k sub-exponential, then broadcast is impossible.

Can we have broadcast with sub-exponential L_k?

★ Existence of DAGs where Broadcasting is Possible:

- Model: (same notation as before)

We have an infinite DAG with a single source/root node in topological ordering.



As before $X_{kj} \sim$ Bernoulli and $X_{00} \sim \mathrm{Ber}(\frac{1}{2})$, and each edge is an independent $\mathrm{BSC}(\delta)$ with $0 < \delta < \frac{1}{2}$.

Let d = no. of incoming edges at each node.
↑ bounded indegree

Inputs at each node are combined using Boolean processing functions.

Question: The processing functions allow information fusion. How small does this allow us to make L_k?

- Impossibility of Reconstruction:

Prop: If $L_k \leq \dfrac{\log(k)}{d \log(\frac{1}{2\delta})}$, then $\lim_{k\to\infty} \mathbb{P}(\hat{X}^k_{ML}(X_k) \neq X_0) = \frac{1}{2}$ regardless of our choice of processing functions.

So, the best L_k we can hope for is $L_k \geq C(\delta, d)\log(k)$ for some constant $C(\delta, d)$.

<u>Proof:</u> Let $A_k = \{$all dL_k edges from level $k-1$ to level k generate independent bits$\}$.
$\{A_k\}_{k\geq 1}$ are mutually independent and $\mathbb{P}(A_k) = (2\delta)^{dL_k}$.

$$L_k \leq \frac{\log(k)}{d\log\left(\frac{1}{2\delta}\right)} \Leftrightarrow (2\delta)^{dL_k} \geq \frac{1}{k}$$
$$\Rightarrow \sum_{k\geq 1}\mathbb{P}(A_k) \geq \sum_{k\geq 1}\frac{1}{k} = \infty$$

So, by Borel-Cantelli lemma, $\{A_k\}_{k\geq 1}$ occur i.o. (i.e. $\mathbb{P}(\bigcap_{m=1}^{\infty}\bigcup_{k\geq m} A_k) = 1$) almost surely.
Once an A_k occurs, all subsequent levels are independent of X_0 and the prob. of error in ML decoding $= \frac{1}{2}$. ∎

• **Random DAG Model:**

We prove existence of DAGs where broadcast is possible for $L_k \geq C(\delta,d)\log(k)$ using probabilistic method.

DAG Model:
- Fix $L_0, L_1, L_2, \ldots$, i.e. the no. of nodes at each level, and d.
- For each node X_{kj}, select d parents in level $k-1$ independently and uniformly (with repetition).
- This defines a random DAG G. → This is strictly speaking a multigraph.

Let $\sigma_k \triangleq \frac{1}{L_k}\sum_{j=0}^{L_k-1} X_{kj}$ be the proportion of 1's in level k.

$\{\sigma_k : k\in\mathbb{N}\}$ forms a Markov chain, and σ_k is a sufficient statistic of X_k for performing inference about $\underset{X_0}{\overset{}{\sigma_0}}$ ($X_0 = \sigma_0$).
↳ Intuition: order of X_{kj} in X_k does not matter.

• **Thm:** Let $d=3$, all processing functions be majority, and $L_k \geq C(\delta)\log(k)$. ← some constant exists
1) If $0<\delta<\frac{1}{6}$, then $\limsup_{k\to\infty}\mathbb{P}\big(\underbrace{\mathbb{1}\{\sigma_k \geq \tfrac{1}{2}\}}_{\text{majority decoder}} \neq \sigma_0\big) < \frac{1}{2}$. ($\Rightarrow$ broadcast possible with ML decoder)
2) If $\underbrace{\frac{1}{6}}_{\text{critical threshold}}<\delta<\frac{1}{2}$, then $\lim_{k\to\infty}\|P_{\sigma_k}^{+} - P_{\sigma_k}^{-}\|_{TV} = 0$. ($\Rightarrow \mathbb{P}(\hat{\sigma}_{ML}^{k}(\sigma_k)\neq\sigma_0) = \frac{1}{2}(1-\|P_{\sigma_k}^{+}-P_{\sigma_k}^{-}\|_{TV}) \xrightarrow{k\to\infty}\frac{1}{2}$, i.e. broadcast impossible)
where $P_{\sigma_k}^{+} = P_{\sigma_k|\sigma_0=1}$ and $P_{\sigma_k}^{-} = P_{\sigma_k|\sigma_0=0}$.

Part 2 holds for $L_k = o\left(\left(\frac{2}{3(1-2\delta)}\right)^k\right)$.

• **Intuition:**

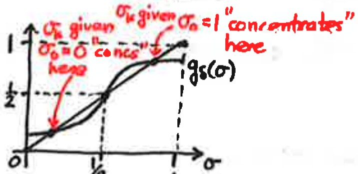
Given $\sigma_{k-1}=\sigma$, $X_{kj} \overset{iid}{\sim} \mathrm{maj}(\mathrm{Ber}(\sigma*\delta), \mathrm{Ber}(\sigma*\delta), \mathrm{Ber}(\sigma*\delta))$. ← indep. Ber. vars; $\sigma*\delta = \sigma(1-\delta)+\delta(1-\sigma)$

$$\Rightarrow \mathbb{P}(X_{kj}=1|\sigma_{k-1}=\sigma) = (\sigma*\delta)^3 + 3(\sigma*\delta)^2(1-\sigma*\delta) \triangleq g_\delta(\sigma) \leftarrow \text{cubic poly. in } \sigma$$

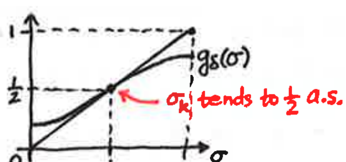
Since $L_k\sigma_k \sim \mathrm{binomial}(L_k, g_\delta(\sigma))\,|\,\sigma_{k-1}=\sigma$, $\mathbb{E}[\sigma_k|\sigma_{k-1}=\sigma] = g_\delta(\sigma)$.

For large k, given $\sigma_{k-1}=\sigma$, $\sigma_k \approx \mathbb{E}[\sigma_k|\sigma_{k-1}=\sigma] = g_\delta(\sigma)$.

<u>Fixed point analysis:</u>



<u>Case:</u> $\delta<\frac{1}{6}$ (3 fixed points)



<u>Case:</u> $\delta\geq\frac{1}{6}$ (1 fixed point)

We require concentration inequalities to rigorize this intuition.

• **Proof of Part 2:** ($\delta>\frac{1}{6}$)

First, construct monotone coupling of $\{\sigma_k^{+}:k\in\mathbb{N}\}$ ($\{\sigma_k\}$ started at $\sigma_0=1$) and $\{\sigma_k^{-}:k\in\mathbb{N}\}$ ($\{\sigma_k\}$ started at $\sigma_0=1$). So, we have $\{(\sigma_k^{+},\sigma_k^{-}):k\in\mathbb{N}\}$ ← Markovian coupling s.t. $\sigma_k^{+}\geq\sigma_k^{-}$ a.s., $\forall k$.

$$\|P_{\sigma_k}^{+}-P_{\sigma_k}^{-}\|_{TV} \underset{\text{Dobrushin}}{\leq} \mathbb{P}(\sigma_k^{+}\neq\sigma_k^{-}) = \mathbb{P}(\underbrace{\sigma_k^{+}-\sigma_k^{-}}_{\geq 0 \text{ a.s.}}\geq\tfrac{1}{L_k}) \underset{\text{Markov}}{\leq} L_k\,\mathbb{E}[\sigma_k^{+}-\sigma_k^{-}].$$

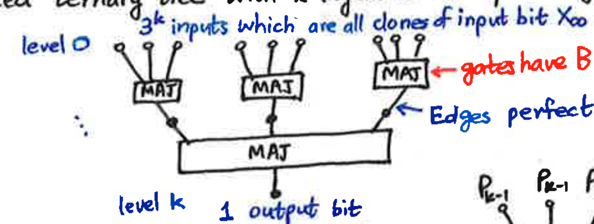
$$\mathbb{E}[\sigma_k^{+}-\sigma_k^{-}] = \mathbb{E}[\mathbb{E}[\sigma_k^{+}-\sigma_k^{-}|\sigma_{k-1}^{+},\sigma_{k-1}^{-}]] = \mathbb{E}[\underbrace{g(\sigma_{k-1}^{+})-g(\sigma_{k-1}^{-})}_{\geq 0\text{ a.s.}}] \leq \underbrace{\tfrac{3}{2}(1-2\delta)}_{\text{Lip. constant}<1\text{ for }\delta>\frac{1}{6}}\mathbb{E}[\overbrace{\sigma_{k-1}^{+}-\sigma_{k-1}^{-}}^{\text{a.s.}}] \underset{\text{recursion}}{\leq} \left(\tfrac{3}{2}(1-2\delta)\right)^k\underbrace{\mathbb{E}[\sigma_0^{+}-\sigma_0^{-}]}_{=1} = \left(\tfrac{3}{2}(1-2\delta)\right)^k$$

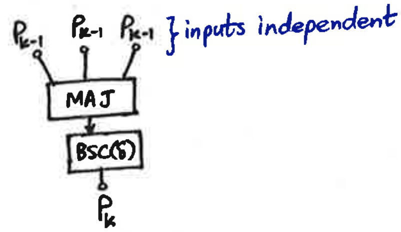
$$\Rightarrow \|P_{\sigma_k}^{+}-P_{\sigma_k}^{-}\|_{TV} \leq L_k\,\mathbb{E}[\sigma_k^{+}-\sigma_k^{-}] \leq L_k\left(\tfrac{3}{2}(1-2\delta)\right)^k \xrightarrow{k\to\infty} 0 \text{ when } L_k = o\left(\left(\tfrac{2}{3(1-2\delta)}\right)^k\right).$$ ∎

- Remarks:

① **Von Neumann Model of Storing a Bit:** [Von Neumann 1956] (cf. [Hajek-Weller 1991])

Balanced ternary tree with k layers of 3-input majority gates.



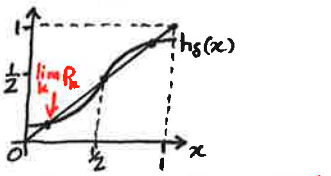


Consider a single noisy majority gate:
Let P_k = prob. of error (i.e. value at node $\neq$ value of input)

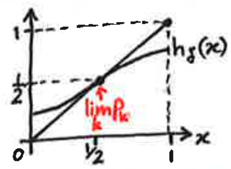
recursion
$$\begin{cases} P_k = \underset{\substack{\uparrow \\ \text{generate wrong bit indep'ly in BSC}}}{\delta} + \underset{\substack{\uparrow \\ \text{copy in BSC}}}{(1-2\delta)}\underbrace{(3P_{k-1}^2 - 2P_{k-1}^3)}_{\substack{\text{prob. at least 2 inputs of MAJ are wrong} \\ \text{(all bits at input are supposed to be equal)}}} \\ P_0 = 0 \end{cases}$$

Define $h_\delta(x) = \delta + (1-2\delta)(3x^2 - 2x^3)$. Then, $P_k = h_\delta^{(k)}(0)$. ← k-fold composition

Fixed Point Analysis:



Case: $\delta < \frac{1}{6}$ (3 fixed points)



Case: $\delta \geq \frac{1}{6}$ (1 fixed point)
↑ THRESHOLD $\frac{1}{6}$ is the SAME!

$$\boxed{\begin{array}{l} \text{Hence, if } \delta < \frac{1}{6}, \text{ then } \lim_{k\to\infty} P_k < \frac{1}{2}, \quad \text{(storage possible)} \\ \quad \& \text{ if } \delta \geq \frac{1}{6}, \text{ then } \lim_{k\to\infty} P_k = \frac{1}{2}. \quad \text{(storage impossible)} \end{array}}$$

② **Comparison to our model:**
- We have noise on edges, not vertices.
- We need concentration of measure on top of the fixed point analysis.

③ **Open Question:** If $d=3$ and $L_k = O(\log(k))$ ← random DAG model, reconstruction is impossible for all choices of Boolean processing functions when $\delta > \frac{1}{6}$. (Equivalently, majority processing functions are "optimal.")

④ **Evans-Schulman Estimate:** [Evans-Schulman 1999] For deterministic DAGs, $I(X_0;X_k) \leq L_k((1-2\delta)^2 d)^k$.
Hence, if $L_k = o\left(\left(\frac{1}{(1-2\delta)^2 d}\right)^k\right)$ and $(1-2\delta)^2 d < 1$, then $\lim_{k\to\infty} I(X_0;X_k) = 0 \underset{\text{Pinsker}}{\Rightarrow} \lim_{k\to\infty} \|P_{X_k}^+ - P_{X_k}^-\|_{TV} = 0$.

This means $(1-2\delta)^2 3 < 1$ ($\Leftrightarrow \delta > \frac{1}{2} - \frac{1}{2\sqrt{3}} = 0.211\ldots$) implies that reconstruction is impossible for all deterministic DAGs for any choice of processing functions.
↳ as well as random DAGs

- **Cor:** (Existence) For any $\delta < \frac{1}{6}$, there exists a DAG ← deterministic with $d=3$, $L_k = O(\log(k))$, and majority processing functions such that $\lim_{k\to\infty} \mathbb{P}(\underset{\substack{\uparrow \\ \text{ML decoder}}}{\hat{X}_{ML}^k}(X_k) \neq X_0) < \frac{1}{2}$.

Proof: Fix $d=3$, $L_k = O(\log(k))$, and $\delta < \frac{1}{6}$.
From theorem, $\exists \underset{\substack{\uparrow \\ \text{dep.s on } \delta}}{\varepsilon} > 0$ s.t. $\mathbb{P}(\mathbb{1}\{\sigma_k \geq \frac{1}{2}\} \neq X_0) \leq \frac{1}{2} - 2\varepsilon$ for all suff. large k.
Let $P_k(G) \triangleq \mathbb{P}(\underset{\substack{\uparrow \\ \text{ML decoder knows } G}}{\hat{X}_{ML}^k}(X_k|G) \neq X_0 | G)$ ← random DAG realization be the prob. of error in ML decoding given G.
Let $E_k = \{\text{all DAGs } \mathcal{G} \text{ s.t. } P_k(\mathcal{G}) \leq \frac{1}{2} - \varepsilon\}$.
$$\frac{1}{2} - 2\varepsilon \geq \mathbb{P}(\mathbb{1}\{\sigma_k \geq \tfrac{1}{2}\} \neq X_0) \underset{\substack{\uparrow \\ \text{ML decoder optimal}}}{\geq} \mathbb{E}[P_k(G)] \geq \mathbb{E}[P_k(G)|G \notin E_k]\,\mathbb{P}(G \notin E_k) \overset{\text{def. of } E_k}{\geq} \left(\tfrac{1}{2} - \varepsilon\right)\mathbb{P}(G \notin E_k) \Rightarrow \underline{\mathbb{P}(G \in E_k) \geq \frac{2\varepsilon}{1-2\varepsilon} > 0} \text{ for all } k$$
Since $\{E_k\}_{k\geq 1}$ is a decreasing sequence (as $P_k(\cdot)$ is increasing in k), $\mathbb{P}(G \in \bigcap_{k\geq 1} E_k) = \lim_{k\to\infty} \mathbb{P}(G \in E_k) \geq \frac{2\varepsilon}{1-2\varepsilon} > 0$. ▨

- What about $d=2$?

(For $d > 3$, we can neglect inputs and existence of DAGs follows from $d=3$ case.)

Thm: Let $d=2$, all processing functions at even levels be AND, all processing functions at odd levels be OR, and $L_k \geq C(\delta)\log(k)$.
   - different from $C(\delta)$ earlier
   - random DAG model

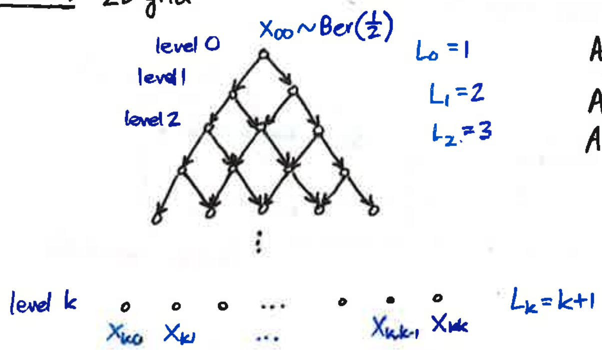
1) If $0 < \delta < \frac{3-\sqrt{7}}{4} = 0.088\ldots$, then $\limsup_{k\to\infty} \mathbb{P}\big(\underbrace{\mathbb{1}\{\sigma_{2k} \geq t(\delta)\}}_{\text{biased majority decoder}} \neq \sigma_0\big) < \frac{1}{2}$.

2) If $\underbrace{\frac{3-\sqrt{7}}{4}}_{\text{critical threshold}} < \delta < \frac{1}{2}$, then $\lim_{k\to\infty} \|P^+_{\sigma_{2k}} - P^-_{\sigma_{2k}}\|_{TV} = 0$.

Part 2 holds for $L_k = o(\tilde{C}(\delta)^k)$.
   - deps on Lip. constant as before

★ Regular 2D Grids – Impossibility of Broadcasting:

- Model: 2D grid



A 2D grid is a specific DAG (deterministic).
All processing functions with 2 inputs are the same.
All processing functions with 1 input are the identity.

- Conjecture: Broadcasting is impossible for 2D grids (as defined above) regardless of the noise level δ.

- Why?

  ① Random DAG view $\Rightarrow$ one fixed point for all $\delta \in (0,\frac{1}{2})$ when $d=2$, $L_k = k+1$. (naïve)
  ↳ Proof of deterministic case is much more difficult.

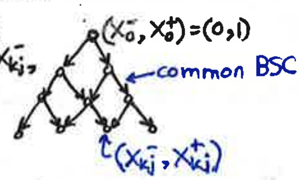
  ② PCA view $\Rightarrow$ If reconstruction is possible in 2D grid, then 2D grid is not ergodic.
  This suggests existence of 1D PCA with binary state space that is not ergodic.
  (Known constructions require a lot more states [Gács 2001].)
  So, 2D grid should be ergodic.

  Note: PCA different from 2D grid because:
  a) PCA uses weak convergence, while we use TV convergence,
  b) 2D grid is a PCA with boundary conditions. (PCA has stronger separation between all zeros and all ones initial configs.)

- Thm: If all processing functions are AND, or all are XOR, then $\underbrace{\lim_{k\to\infty} \|P^+_{X_k} - P^-_{X_k}\|_{TV} = 0}_{\text{broadcast impossible}}$ regardless of $\delta \in (0,\frac{1}{2})$.

- Proof of AND case: (Sketch)

  ① Monotone Markovian Coupling:
  Let $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$ be the Markov chains started at $X_0^+ = 1$ and $X_0^- = 0$, respectively.
  We "run" these chains on the same grid:
  → Each BSC(δ) either copies both X_{kj}^+ and X_{kj}^-, or generates the same independent bit for both chains.



  Then, $\forall k,j,\ X_{kj}^+ \geq X_{kj}^-$ a.s.. [Check this]

[continued.]

Proof of AND case cont'd:

Let each node of the grid be $Y_{kj} \triangleq (X_{kj}^-, X_{kj}^+) \in \{\underset{(0,0)}{0_c}, \overset{(0,1)}{1_u}, \underset{(1,1)}{1_c}\}$. ← (1,0) is not required in the alphabet

(BSC has matrix $W = \begin{array}{c|ccc} & 0_c & 1_u & 1_c \\ \hline 0_c & 1-\delta & 0 & \delta \\ 1_u & \delta & 1-2\delta & \delta \\ 1_c & \delta & 0 & 1-\delta \end{array}$, and AND operates entrywise.)

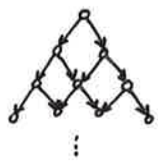
② Reduction to Coupled Grid:

$$\|P_{X_k^+} - P_{X_k^-}\|_{TV} \underset{\text{Dobrushin}}{\leq} \mathbb{P}(X_k^+ \neq X_k^-) = 1 - \mathbb{P}(X_k^+ = X_k^-)$$

Since $(\{X_k^+ = X_k^-\})_{k \geq 1}$ is increasing, $\lim_{k\to\infty} \|P_{X_k^+} - P_{X_k^-}\|_{TV} \leq 1 - \mathbb{P}(\exists k, X_k^+ = X_k^-)$.

So, it suffices to prove that: $\mathbb{P}(A) = 1$ for $A \triangleq \{\exists k$, there are no 1_u's in level $k\} = \{\exists k, X_k^+ = X_k^-\}$.

③ Oriented Bond Percolation: [Durrett 1984]



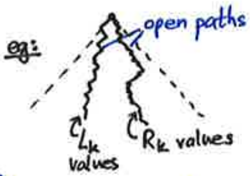
Remove ("closed") each edge indep'ly wp $1-p$, or keep it ("open") wp $p \in [0,1]$.

Let $B = \{\exists$ infinite open path starting at root$\}$,

R_k = rightmost node index at level k that is connected to root,

L_k = leftmost ——"——.

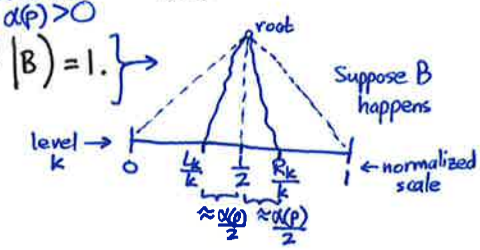
eg:



Thm: For phase transition threshold $\delta_{perc} \in (\frac{1}{2}, 1)$:

1) If $p > \delta_{perc}$, then $\mathbb{P}_p(B) > 0$ and $\mathbb{P}_p\left(\lim_{k\to\infty} \frac{R_k}{k} = \frac{1+\alpha(p)}{2} \text{ and } \lim_{k\to\infty} \frac{L_k}{k} = \frac{1-\alpha(p)}{2} \middle| B\right) = 1$. (for some $\alpha(p) > 0$; ← measure depends on p)

2) If $p < \delta_{perc}$, then $\mathbb{P}_p(B) = 0$.



④ Case I: $p = 1-2\delta < \delta_{perc}$ $(\Leftrightarrow \delta > \frac{1-\delta_{perc}}{2})$

Edge open $\Leftrightarrow$ BSC copies.

By Thm part 2, $\mathbb{P}(B) = 0 \Leftrightarrow \mathbb{P}(\underbrace{\{\text{no. of nodes connected to root with copies is finite}\}}_{B^c}) = 1$.

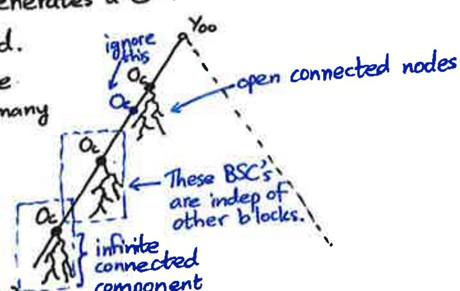
Since B^c occurs a.s., there is a level where no BSCs copy $\Rightarrow$ there is a level k with no 1_u's. $(B^c \subseteq A)$

Hence, $\mathbb{P}(A) = 1$.

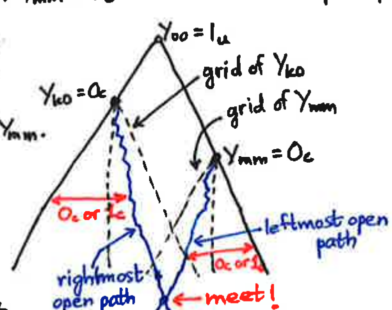
⑤ Case II: $p = 1-\delta > \delta_{perc}$ $(\Leftrightarrow \delta < 1-\delta_{perc})$

- Edge open $\Leftrightarrow$ BSC copies or generates a 0 as the new bit.
- Consider "left edge" of 2D grid.
- Since BSC's on this side generate indep 0 wp δ, there are infinitely many $Y_{k0} = 0_c$ on the left side.

Use Thm part 1 {
- Each 0_c has a set of open connected nodes below it, and the set is ∞ wp $\mathbb{P}(B) > 0$.



- Since blocks of (0_c, connected nodes) are independent, we almost surely have $Y_{kj} = 0_c$ with an infinite open path connected to it. (Borel-Cantelli)
- Similarly, $\exists m$ s.t. $Y_{mm} = 0_c$ with an infinite open path connected to it a.s..
- By Thm part 1, the rightmost path from Y_{k0} meets the leftmost path from Y_{mm}.
- All nodes on these paths are 0_c, and all nodes enclosed by these paths are 0_c or 1_c!



$\therefore$ When the paths meet, all nodes at that level are not 1_u's. Hence, $\mathbb{P}(A) = 1$.

▨

★ Miscellaneous Notes:

① Evans-Schulman Estimate: [Evans-Schulman 1999] (c.f. [Polyanskiy-Wu 2017])

Consider a Bayesian network on a DAG with one source node X.
For any node W, we identify W with the random variable at W.
Let $\eta_W \triangleq \eta_{KL}(P_{W|pa(W)})$, and for any path $\pi = (V_0, \dots, V_k)$, $\eta_\pi \triangleq \prod_{i=1}^{k} \eta_{V_i}$.

contraction coefficient — parents of W — start — end

Note: For channel $P_{Y|X}$, $\eta_{KL}(P_{Y|X}) \triangleq \sup_{P_{U,X}:\, U \to X \to Y} \dfrac{I(U;Y)}{I(U;X)}$.

→ **Thm:** $\eta_{KL}(P_{V|X}) \leq \sum_{\pi: X \to V} \eta_\pi$ } path from X to V, for every set of nodes V.

Proof: Order all nodes in the DAG (so that $ord(X) = 0$) and the ordering is consistent with the topological ordering. For a set of nodes V, let $ord(V) = \sup\{ord(W) : W \in V\}$.

index in order

Suppose $W > V$ for a node W and set of nodes V, i.e. $ord(W) > ord(V)$.

$$\boxed{\textbf{Claim: } \eta_{KL}(P_{W,V|X}) \leq \eta_W \, \eta_{KL}(P_{V,pa(W)|X}) + (1-\eta_W)\, \eta_{KL}(P_{V|X}).}$$

Pf: Consider Markov chain $U \to X \to (V,A) \to W$ for arbitrary U and $A = pa(W) - V$.
Given V, we still have $U \to X \to A \to W$.
Hence, $I(U;W|V=v) \leq \underbrace{\eta_{KL}(P_{W|A,V=v})}_{\leq\, \eta_{KL}(P_{W|pa(W)}) \text{ [by definition]}} I(U;A|V=v) \Rightarrow I(U;W|V) \leq \eta_W \, I(U;A|V)$.

Adding $I(U;V)$ to both sides gives:

$$I(U;W,V) \leq \eta_W \, I(U;V,A) + (1-\eta_W)\, I(U;V)$$

$$\Rightarrow \frac{I(U;W,V)}{I(U;X)} \leq \eta_W \frac{I(U;V,A)}{I(U;X)} + (1-\eta_W) \frac{I(U;V)}{I(U;X)}$$

$$\Rightarrow \eta_{KL}(P_{W,V|X}) \leq \eta_W \, \eta_{KL}(P_{V,pa(W)|X}) + (1-\eta_W)\, \eta_{KL}(P_{V|X}). \quad ▨$$

The rest of proof follows by strong induction on the $ord(\cdot)$ of sets of nodes.

Inductive hypothesis → Assume $\eta_{KL}(P_{V|X}) \leq \sum_{\pi: X \to V} \eta_\pi$ for every set of nodes V with $ord(V) \leq k$.

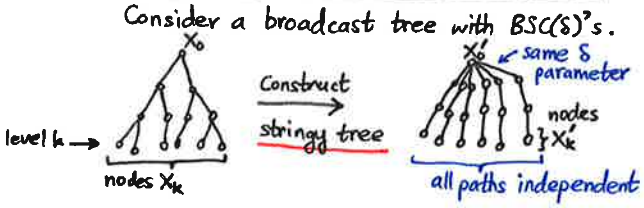
Let $ord(W) = k+1$ for node W. Then for any V with $ord(V) \leq k$, we have:

$$\begin{aligned}
\underbrace{\eta_{KL}(P_{W,V|X})}_{ord(W,V) = k+1} &\leq \eta_W \, \underbrace{\eta_{KL}(P_{V,pa(W)|X})}_{ord(V,pa(W)) \leq k} + (1-\eta_W)\, \eta_{KL}(P_{V|X}) \\
&\leq \eta_W \sum_{\pi: X \to V, pa(W)} \eta_\pi + (1-\eta_W) \sum_{\pi: X \to V} \eta_\pi \\
&= \eta_W \sum_{\pi: X \to A} \eta_\pi + \sum_{\pi: X \to V} \eta_\pi \\
&\leq \eta_W \sum_{\pi: X \to pa(W)} \eta_\pi + \sum_{\pi: X \to V} \eta_\pi \quad [\text{as } pa(W) \supseteq A] \\
&= \sum_{\pi: X \to W} \eta_\pi + \sum_{\pi: X \to V} \eta_\pi \\
&= \sum_{\pi: X \to W, V} \eta_\pi .
\end{aligned}$$

Hence, the result is true for all sets of nodes V with $ord(V) \leq k+1$.
The proof is complete by induction. ▨

② Evans-Schulman Estimate for Trees: [Evans-Kenyon-Peres-Schulman 2000]

Consider a broadcast tree with BSC(δ)'s.



Thm: $\underbrace{I(X_0;X_k) \leq I(X_0';X_k')}_{\text{degradation construction!}} \leq \underbrace{\sum_i I(X_0';X_{ki}')}_{X_{ki}' \text{ cond. iid given } X_0'} \leq \underbrace{d^k}_{\text{no. of paths}} \underbrace{(1-2\delta)^{2k}}_{\text{contraction on path}} = \left((1-2\delta)^2 d\right)^k$.

So, if $(1-2\delta)^2 d < 1$, reconstruction is impossible.

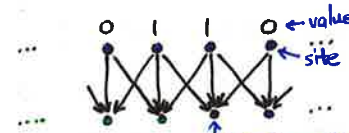
Anuran Makur
2 April 2018

3 Probabilistic Cellular Automata: (1D)

↙ abbrev. as PCA

- sites $\mathbb{Z}$
- state space S, $|S| < \infty$ (eg: $S = \{0,1\}$)
- configuration space $S^{\mathbb{Z}}$ (functions $\xi : \mathbb{Z} \to S$)
- deterministic function $f : S^{|\mathcal{N}|} \to S$ (e.g. $f = \text{MAJ}$)
- neighborhood $\mathcal{N}$ (eg: $\{-1,0,1\}$)
  ↑ $|\mathcal{N}| < +\infty$

eg:



↑ At time $k+1$, each site x simultaneously computes $f(\{\xi(x+i) : i \in \mathcal{N}\})$ and passes it through an indep. $\text{BSC}(\delta)$ to get $\eta(x)$. } Dynamics

Main Question: Is a PCA ergodic?

PCA defines a Markov process on $S^{\mathbb{Z}}$. For any initial config. $\xi \in S^{\mathbb{Z}}$, let ν_k^{ξ} be the prob. measure on $S^{\mathbb{Z}}$ at time k.

We say a PCA is ergodic $\Leftrightarrow$ $\exists !$ invariant measure ν_0 on $S^{\mathbb{Z}}$ s.t. $\forall$ init. config. $\xi \in S^{\mathbb{Z}}$, $\nu_k^{\xi} \xrightarrow{w} \nu_0$ as $k \to \infty$.
↑ weak convergence

Weak Convergence:

For $S^{\mathbb{Z}}$, the σ-algebra is defined as follows.

$$\mathcal{C} \triangleq \bigcup_{\substack{A \subseteq \mathbb{Z},\\ |A| < \infty}} \left\{ \{\xi \in S^{\mathbb{Z}} : \xi(A) = x_A\} \,\middle|\, x_A \in S^{|A|} \right\}$$ ← cylinder sets

$\sigma(\mathcal{C})$ is the σ-algebra on $S^{\mathbb{Z}}$.

By Daniell-Kolmogorov theorem, consistent finite-dim marginals defined on $\mathcal{C}$ uniquely determine measure on $S^{\mathbb{Z}}$.

Def: μ_n, μ measures on $(S^{\mathbb{Z}}, \sigma(\mathcal{C}))$.
$\mu_n \xrightarrow{w} \mu \Leftrightarrow \forall C \in \mathcal{C}, \lim_{n \to \infty} \mu_n(C) = \mu(C)$.

} ← weak convergence corresponds to convergence of finite-dim distributions

Note: $\mu_n \xrightarrow{w} \mu \Leftrightarrow \forall C \in \mathcal{C}, \lim_{n \to \infty} \mu_n(C) = \mu(C)$

$$\not\Rightarrow \lim_{n \to \infty} \sup_{A \in \sigma(\mathcal{C})} |\mu_n(A) - \mu(A)| \Leftrightarrow \lim_{n \to \infty} \|\mu_n - \mu\|_{TV}$$

i.e. weak convergence $\not\Rightarrow$ TV convergence.

For many PCA with special characteristics, ergodicity can be determined by convergence of dist.s over finite intervals (e.g. single sites).