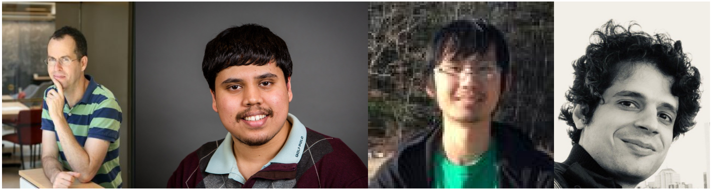
Information and inference on trees

Yury Polyanskiy

EECS
Massachusetts Institute of Technology

Feb. 5, 2020 CSoI Workshop, Honolulu, HI

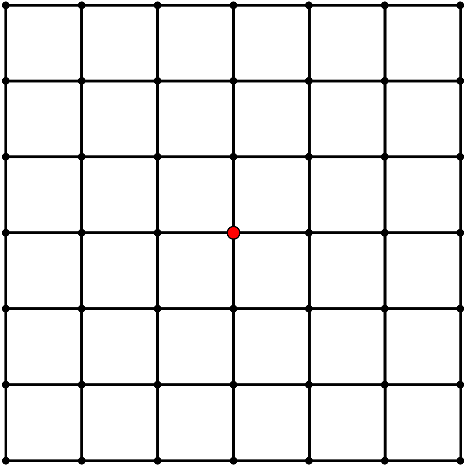


with: E. Mossel, A. Makur, Yuzhou Gu, H. Roozbehani

Motivation: Information Propagation in 2D Grid

- How does information spread in time?

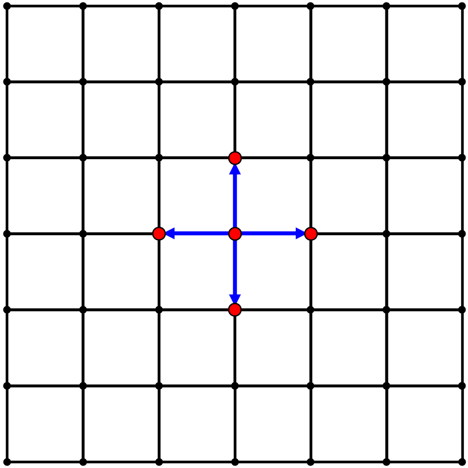
$\boxed{\leftarrow\uparrow\downarrow\rightarrow \text{ are } \mathrm{BSC}(\delta)}$



Motivation: Information Propagation in 2D Grid

- How does information spread in time?

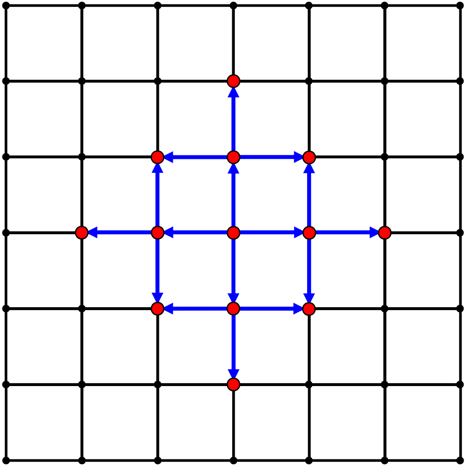
$\boxed{\leftarrow\uparrow\downarrow\rightarrow \text{ are BSC}(\delta)}$



Motivation: Information Propagation in 2D Grid

- How does information spread in time?

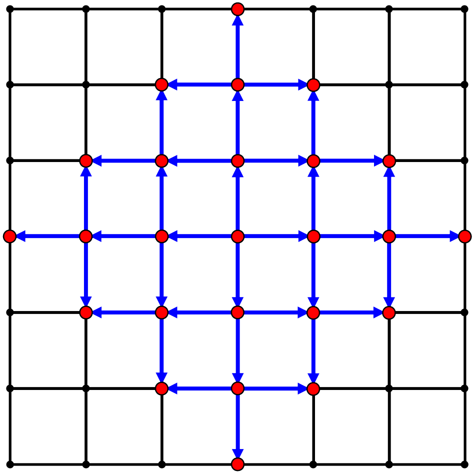
$\boxed{\leftarrow\uparrow\downarrow\rightarrow \text{ are BSC}(\delta)}$



Motivation: Information Propagation in 2D Grid

- How does information spread in time?

$\boxed{\leftarrow\uparrow\downarrow\rightarrow \text{ are BSC}(\delta)}$

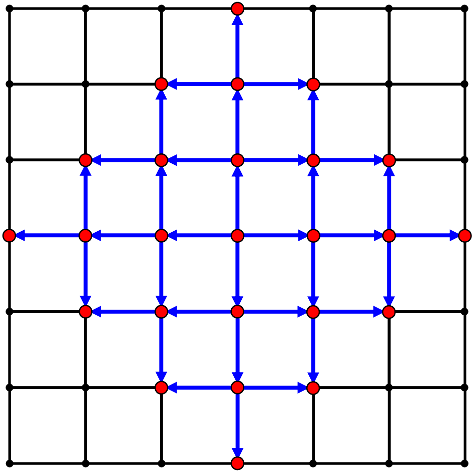


Motivation: Information Propagation in 2D Grid

- How does information spread in time?

$\boxed{\leftarrow\uparrow\downarrow\rightarrow \text{ are BSC}(\delta)}$

- Can we invent relay functions so that far boundary contains non-trivial information about the original bit?

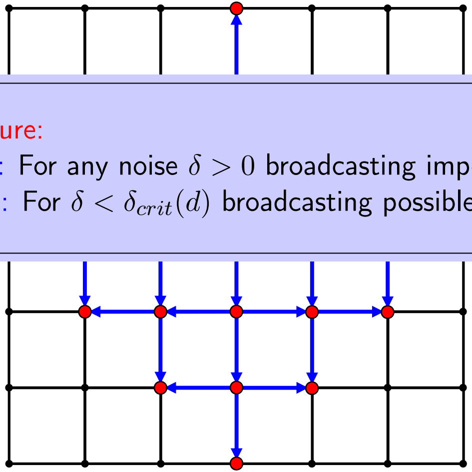


Motivation: Information Propagation in 2D Grid

- How does information spread in time?

$\leftarrow\uparrow\downarrow\rightarrow$ are BSC(δ)

- Can we invent relay functions so that far boundary contains non-trivial information about the original bit?



Main conjecture:

2 dimensions: For any noise $\delta > 0$ broadcasting impossible

$d \geq 3$ dimen.: For $\delta < \delta_{crit}(d)$ broadcasting possible

Related Models in the Literature

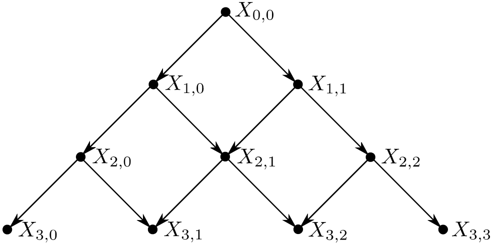
- **Communication Networks:**
  Sender broadcasts single bit through network.

Related Models in the Literature

- **Communication Networks:**
  Sender broadcasts single bit through network.
- **Reliable Computation and Storage:** [von56, HW91, ES03, Ung07]
  Broadcasting model is noisy circuit to remember a bit using perfect gates and faulty wires.

Related Models in the Literature

- **Communication Networks:**
  Sender broadcasts single bit through network.
- **Reliable Computation and Storage:**
  Broadcasting model is noisy circuit to remember a bit using perfect gates and faulty wires.
- **Probabilistic Cellular Automata:**
  Impossibility of broadcasting on 2D regular grid parallels ergodicity of 1D probabilistic cellular automata.



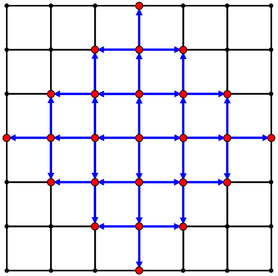
Related Models in the Literature

- **Communication Networks:**
  Sender broadcasts single bit through network.
- **Reliable Computation and Storage:**
  Broadcasting model is noisy circuit to remember a bit using perfect gates and faulty wires.
- **Probabilistic Cellular Automata:**
  Broadcasting on 2D regular grid parallels 1D probabilistic cellular automata.
- **Ancestral Data Reconstruction:**
  Reconstruction on *trees* $\Leftrightarrow$ Infer trait of ancestor from observed population.

Related Models in the Literature

- **Communication Networks:**
  Sender broadcasts single bit through network.
- **Reliable Computation and Storage:**
  Broadcasting model is noisy circuit to remember a bit using perfect gates and faulty wires.
- **Probabilistic Cellular Automata:**
  Broadcasting on 2D regular grid parallels 1D probabilistic cellular automata.
- **Ancestral Data Reconstruction:**
  Reconstruction on *trees* $\Leftrightarrow$ Infer trait of ancestor from observed population.
- **Ferromagnetic Ising Models:** [BRZ95, EKPS00]
  Reconstruction impossible on *tree* $\Leftrightarrow$ Free boundary Gibbs state of Ising model on tree is extremal.

How to assess information decay in networks?



Information percolation:

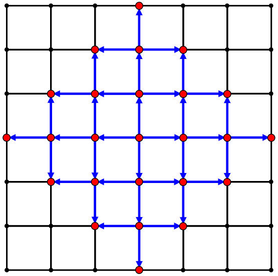
In Graphical Models

$$I(X_a; X_b) \leq \mathrm{perc}(a, b)$$

$$\mathrm{perc}(a, b) = \mathbb{P}[\exists \text{ open path } a \to b]$$

each edge/vertex open w.p. η_{KL}

How to assess information decay in networks?



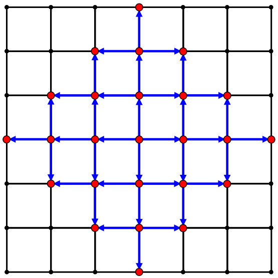
Information percolation:
In Graphical Models
$$I(X_a; X_b) \le \mathrm{perc}(a, b)$$

$$\mathrm{perc}(a, b) = \mathbb{P}[\exists \text{ open path } a \to b]$$
each edge/vertex open w.p. η_{KL}

Established in a sequence of papers:

1. [P.-Wu'16]: "Dissipation of information in channels with input constraints"
2. [P.-Wu'17]: "Strong data-processing inequalities for channels and Bayesian networks"
3. [P.-Wu'18]: "Application of information-percolation method to reconstruction problems on graphs"

How to assess information decay in networks?



Information percolation:

In Graphical Models

$$I(X_a; X_b) \le \mathrm{perc}(a,b)$$

$$\mathrm{perc}(a,b) = \mathbb{P}[\exists \text{ open path } a \to b]$$

each edge/vertex open w.p. η_{KL}

Established in a sequence of papers:

1. [P.-Wu'16]: "Dissi with input constraints"
2. [P.-Wu'17]: "Stro channels and Bayesian networks"
3. [P.-Wu'18]: "Application of information-percolation method to reconstruction problems on graphs"

What is η_{KL}?

Data processing inequality

$$\begin{tikzcd}
P_X \arrow[dr] & & & P_Y \\
& X \arrow[r, "P_{Y|X}"] & Y \arrow[ur] \arrow[dr] & \\
Q_X \arrow[ur] & & & Q_Y
\end{tikzcd}$$

- For any channel $P_{Y|X}$ we always have:

$$D(Q_Y \| P_Y) \leq D(Q_X \| P_X)$$

i.e. channels contract divergence (in fact, any f-divergence)

Data processing inequality

$$\begin{tikzcd}
P_X \arrow[dr] & & & P_Y \\
& X \arrow[r, "P_{Y|X}"] & Y \arrow[ur] \arrow[dr] & \\
Q_X \arrow[ur] & & & Q_Y
\end{tikzcd}$$

- For any channel $P_{Y|X}$ we always have:

$$D(Q_Y \| P_Y) \le D(Q_X \| P_X)$$

  i.e. channels contract divergence (in fact, any f-divergence)

- Equivalently, for any Markov chain $U \to X \to Y$ we have

$$I(U;Y) \le I(U;X)$$

Data processing inequality

$$\begin{tikzcd}
P_X \arrow[dr] & & & P_Y \\
& X \arrow[r, "P_{Y|X}"] & Y \arrow[ur] \arrow[dr] & \\
Q_X \arrow[ur] & & & Q_Y
\end{tikzcd}$$

- For any channel $P_{Y|X}$ we always have:

$$D(Q_Y \| P_Y) \le D(Q_X \| P_X)$$

  i.e. channels contract divergence (in fact, any f-divergence)
- Equivalently, for any Markov chain $U \to X \to Y$ we have

$$I(U;Y) \le I(U;X)$$

- In most cases, inequality is strict...

Strong data-processing inequality (SDPI)

$$\begin{tikzcd}
P_X \arrow[dr] & & & P_Y \\
& X \arrow[r, "P_{Y|X}"] & Y \arrow[ur] \arrow[dr] & \\
Q_X \arrow[ur] & & & Q_Y
\end{tikzcd}$$

Definition (Two types of SDPI constants)

- [Input-free η_{KL}] Fix channel $P_{Y|X}$ then

$$\eta_{\mathsf{KL}}(P_{Y|X}) \triangleq \sup_{Q_X, P_X} \frac{D(Q_Y||P_Y)}{D(Q_X||P_X)}$$

Strong data-processing inequality (SDPI)

$$
\begin{tikzcd}
P_X \arrow[dr] & & & P_Y \\
& X \arrow[r, "P_{Y|X}"] & Y \arrow[ur] \arrow[dr] & \\
Q_X \arrow[ur] & & & Q_Y
\end{tikzcd}
$$

Definition (Two types of SDPI constants)

- [Input-free η_{KL}] Fix channel $P_{Y|X}$ then

$$\eta_{\mathsf{KL}}(P_{Y|X}) \triangleq \sup_{Q_X, P_X} \frac{D(Q_Y || P_Y)}{D(Q_X || P_X)} = \sup_{U \to X \to Y} \frac{I(U;Y)}{I(U;X)}$$

Strong data-processing inequality (SDPI)

$$\begin{tikzcd} P_X \arrow[dr] & & & P_Y \\ & X \arrow[r, "P_{Y|X}"] & Y \arrow[ur] \arrow[dr] & \\ Q_X \arrow[ur] & & & Q_Y \end{tikzcd}$$

Definition (Two types of SDPI constants)

- [Input-free η_{KL}] Fix channel $P_{Y|X}$ then

$$\eta_{\mathsf{KL}}(P_{Y|X}) \triangleq \sup_{Q_X, P_X} \frac{D(Q_Y||P_Y)}{D(Q_X||P_X)} = \sup_{U \to X \to Y} \frac{I(U;Y)}{I(U;X)}$$

- [Fixed-input η_{KL}] Fix channel $P_{Y|X}$ and input distribution P_X then

$$\eta_{\mathsf{KL}}(P_X, P_{Y|X}) \triangleq \sup_{Q_X} \frac{D(Q_Y||P_Y)}{D(Q_X||P_X)}$$

Strong data-processing inequality (SDPI)

$$\begin{tikzcd}
P_X \arrow[dr] & & & P_Y \\
& X \arrow[r, "P_{Y|X}"] & Y \arrow[ur] \arrow[dr] & \\
Q_X \arrow[ur] & & & Q_Y
\end{tikzcd}$$

Definition (Two types of SDPI constants)

- [Input-free η_{KL}] Fix channel $P_{Y|X}$ then

$$\eta_{\mathsf{KL}}(P_{Y|X}) \triangleq \sup_{Q_X, P_X} \frac{D(Q_Y||P_Y)}{D(Q_X||P_X)} = \sup_{U \to X \to Y} \frac{I(U;Y)}{I(U;X)}$$

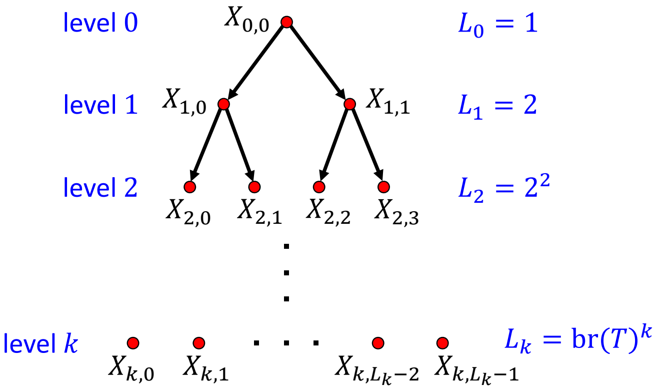
- [Fixed-input η_{KL}] Fix channel $P_{Y|X}$ and input distribution P_X then

$$\eta_{\mathsf{KL}}(P_X, P_{Y|X}) \triangleq \sup_{Q_X} \frac{D(Q_Y||P_Y)}{D(Q_X||P_X)} = \sup_{U \to X \to Y} \frac{I(U;Y)}{I(U;X)}$$

Next: Special case of broadcasting problem

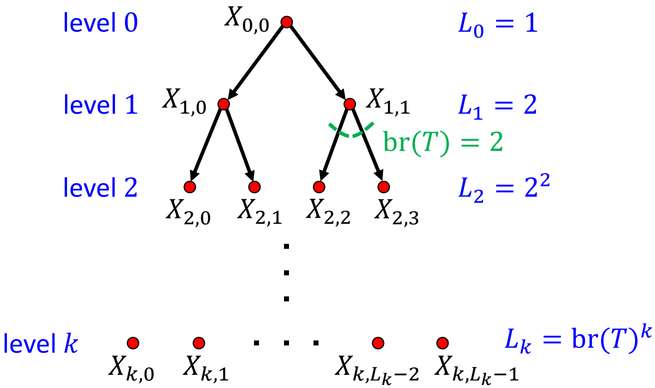
Broadcasting on Trees

- Fix infinite tree T with branching number $\mathrm{br}(T)$.



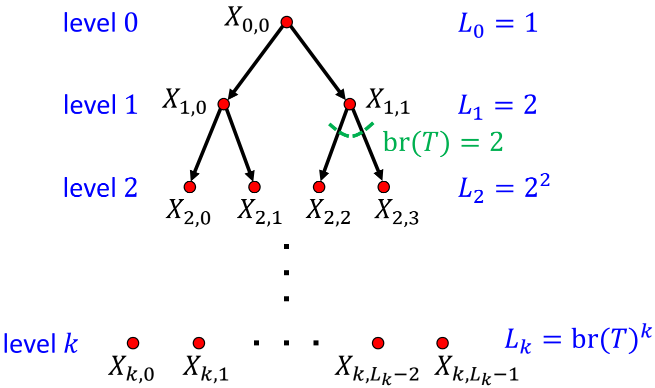
Broadcasting on Trees

- Fix infinite tree T with branching number $\mathrm{br}(T)$.



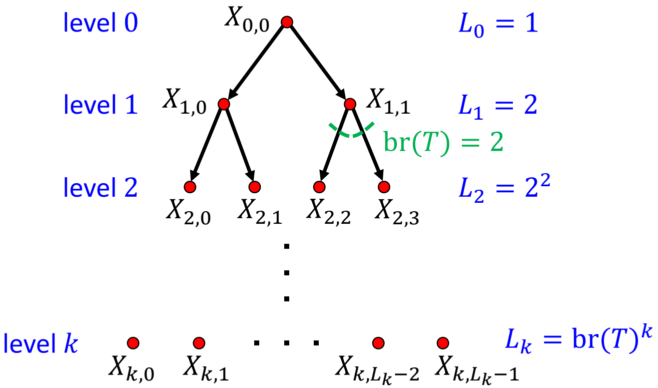
Broadcasting on Trees

- Fix infinite tree T with branching number $\mathrm{br}(T)$.
- Root $X_{0,0} \sim$ Bernoulli$\left(\frac{1}{2}\right)$



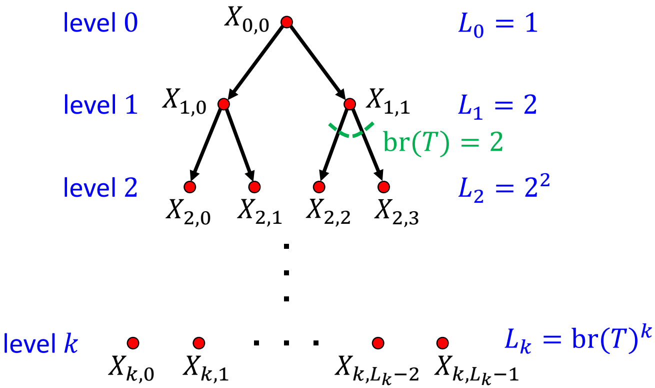
Broadcasting on Trees

- Fix infinite tree T with branching number $\mathrm{br}(T)$.
- Root $X_{0,0} \sim$ Bernoulli$\left(\frac{1}{2}\right)$
- Edges are independent BSCs with crossover probability $\delta \in \left(0, \frac{1}{2}\right)$.



Broadcasting on Trees

- Fix infinite tree T with branching number $\mathrm{br}(T)$.
- Root $X_{0,0} \sim$ Bernoulli$\left(\frac{1}{2}\right)$
- Edges are independent BSCs with crossover probability $\delta \in \left(0, \frac{1}{2}\right)$.
- Let $P_{\mathsf{ML}}^{(k)} = \mathbb{P}\left(\hat{X}_{\mathsf{ML}}^{k}(X_k) \neq X_{0,0}\right)$, where $X_k = \left(X_{k,0}, \ldots, X_{k,\mathrm{br}(T)^k-1}\right)$.



Broadcasting on Trees: Who cares?

To summarize:

- Root variable $X_{0,0}$ is the information
- It spreads along a tree of $BSC(\delta)$'s.
- Goal: Reconstruct $X_{0,0}$ from a vector of far-away leaves X_k
- If $P_e \to 1/2$ then we say problem is non-reconstructible

Broadcasting on Trees: Who cares?

To summarize:

- Root variable $X_{0,0}$ is the information
- It spreads along a tree of $BSC(\delta)$'s.
- Goal: Reconstruct $X_{0,0}$ from a vector of far-away leaves X_k
- If $P_e \to 1/2$ then we say problem is non-reconstructible

This (or similar) question is common:

- Coding: analysis of sparse-graph codes
- CS: Random constraint satisfaction (e.g. k-SAT)
- Stats/ML: Community detection

Direct proof: density evolution

How would a non-IT guy prove it?

- Algorithm: belief propagation from leaves to root
- ... optimal on trees **(!)**

Direct proof: density evolution

How would a non-IT guy prove it?

- Algorithm: belief propagation from leaves to root
- ... optimal on trees **(!)**
- Only need to analyze evolution of the (density of) messages. Easy?

Direct proof: density evolution

How would a non-IT guy prove it?

- Algorithm: belief propagation from leaves to root
- ... optimal on trees **(!)**
- Only need to analyze evolution of the (density of) messages. Easy?
- Evolution operator $T \circ S$: acts on prob. dist. μ on $[0, +\infty]$ via:

$$S(\mu) = \text{Law of } \ln \frac{\delta e^L + \bar{\delta}}{\bar{\delta} e^L + \delta}, \qquad L \sim \mu$$

$$T(\mu) = \text{Law of } L'(L_1, L_2)\,, \qquad L_1, L_2 \overset{iid}{\sim} \mu$$

and

$$L' = \begin{cases} L_1 + L_2, & \text{w.p. } p(L_1, L_2) + p(-L_1, -L_2) \\ |L_1 - L_2|, & \text{o/w} \end{cases}$$

and $p(L_1, L_2) = (1 + e^{L_1})^{-1}(1 + e^{L_2})^{-1}$.

Direct proof: density evolution

How would a non-IT guy prove it?

- Algorithm: belief propagation from leaves to root
- ... optimal on trees **(!)**
- Only need to analyze evolution of the (density of) messages. Easy?
- Evolution operator $T \circ S$: acts on prob. dist. μ on $[0, +\infty]$ via:

$$S(\mu) = \text{Law of } \ln \frac{\delta e^L + \bar{\delta}}{\bar{\delta} e^L + \delta}, \qquad L \sim \mu$$

ar

$$\boxed{\text{non-reconstruction} \iff T \circ S \circ T \circ \cdots \circ S(\delta_\infty) \approx \delta_0}$$

$$L' = \begin{cases} L_1 + L_2, & \text{w.p. } p(L_1, L_2) + p(-L_1, -L_2) \\ |L_1 - L_2|, & \text{o/w} \end{cases}$$

and $p(L_1, L_2) = (1 + e^{L_1})^{-1}(1 + e^{L_2})^{-1}$.

Direct proof: density evolution

How would a non-IT guy prove it?

- Algorithm: belief propagation from leaves to root
- ... optimal on trees **(!)**
- Only need to analyze evolution of the (density of) messages. Easy?
- Evolution operator $T \circ S$: acts on prob. dist. μ on $[0, +\infty]$ via:

$$S(\mu) = \text{Law of } \ln \frac{\delta e^L + \bar{\delta}}{\bar{\delta} e^L + \delta}, \qquad L \sim \mu$$

ar

non-reconstruction $\iff T \circ S \circ T \circ \cdots \circ S(\delta_\infty) \approx \delta_0$

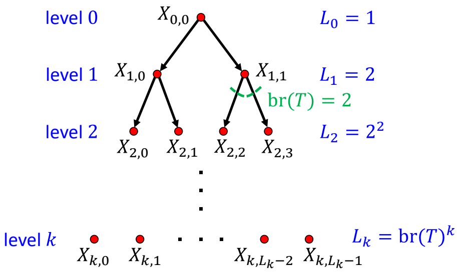
... pretty tough to work with (unless you are [BRZ95])

ar

Broadcasting on Trees

Theorem (Phase Transition for Trees [KS66, BRZ95, EKPS00])

- If $\delta < \frac{1}{2} - \frac{1}{2\sqrt{\mathrm{br}(T)}}$, then reconstruction possible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} < \frac{1}{2}$.
- If $\delta > \frac{1}{2} - \frac{1}{2\sqrt{\mathrm{br}(T)}}$, then reconstruction impossible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} = \frac{1}{2}$.

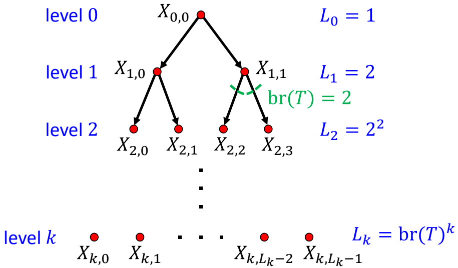


Broadcasting on Trees

Theorem (Phase Transition for Trees [KS66, BRZ95, EKPS00])

- If $(1-2\delta)^2\ \mathrm{br}(T) > 1$, then reconstruction possible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} < \frac{1}{2}$.
- If $(1-2\delta)^2\ \mathrm{br}(T) < 1$, then reconstruction impossible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} = \frac{1}{2}$.

Proof Idea: Strong data processing inequality [AG76, ES99]

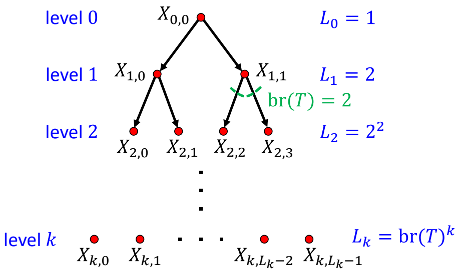


Theorem (Phase Transition for Trees [KS66, BRZ95, EKPS00])

- If $(1-2\delta)^2 \, \mathrm{br}(T) > 1$, then reconstruction possible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} < \frac{1}{2}$.
- If $(1-2\delta)^2 \, \mathrm{br}(T) < 1$, then reconstruction impossible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} = \frac{1}{2}$.

Proof Idea: Strong data processing inequality [AG76, ES99]

- If $P_{Y|X} = \mathsf{BSC}(\delta)$, then for any $U \to X \to Y$: $I(U;Y) \le (1-2\delta)^2 I(U;X)$.



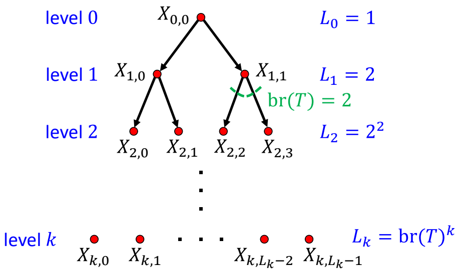
Broadcasting on Trees

Theorem (Phase Transition for Trees [KS66, BRZ95, EKPS00])

- If $(1-2\delta)^2\,\mathrm{br}(T) > 1$, then reconstruction possible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} < \frac{1}{2}$.
- If $(1-2\delta)^2\,\mathrm{br}(T) < 1$, then reconstruction impossible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} = \frac{1}{2}$.

Proof Idea: Strong data processing inequality [AG76, ES99]

- If $P_{Y|X} = \mathsf{BSC}(\delta)$, then for any $U \to X \to Y$: $I(U;Y) \le (1-2\delta)^2 I(U;X)$.
- For any $0 \le j < \mathrm{br}(T)^k$, $I(X_{0,0}; X_{k,j}) \le (1-2\delta)^{2k}$.



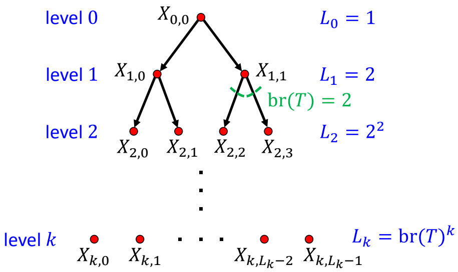
Broadcasting on Trees

Theorem (Phase Transition for Trees [KS66, BRZ95, EKPS00])

- If $(1-2\delta)^2\ \mathrm{br}(T) > 1$, then reconstruction possible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} < \frac{1}{2}$.
- If $(1-2\delta)^2\ \mathrm{br}(T) < 1$, then reconstruction impossible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} = \frac{1}{2}$.

Proof Idea: Strong data processing inequality [AG76, ES99]

- If $P_{Y|X} = \mathsf{BSC}(\delta)$, then for any $U \to X \to Y$: $I(U;Y) \le (1-2\delta)^2 I(U;X)$.
- For any $0 \le j < \mathrm{br}(T)^k$, $I(X_{0,0}; X_{k,j}) \le (1-2\delta)^{2k}$.
- $\mathrm{br}(T)^k$ paths from X_0 to X_k: $I(X_0;X_k) \le \left(\mathrm{br}(T)(1-2\delta)^2\right)^k$.



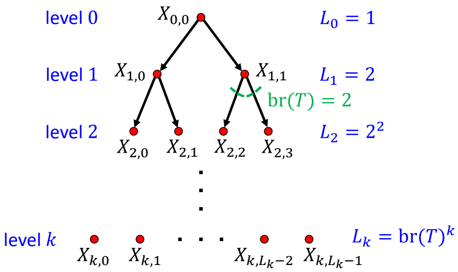
Broadcasting on Trees

Theorem (Phase Transition for Trees [KS66, BRZ95, EKPS00])

- If $(1-2\delta)^2 \, \mathrm{br}(T) > 1$, then reconstruction possible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} < \frac{1}{2}$.
- If $(1-2\delta)^2 \, \mathrm{br}(T) < 1$, then reconstruction impossible: $\lim_{k\to\infty} P_{\mathsf{ML}}^{(k)} = \frac{1}{2}$.

Proof Idea: Strong data processing inequality [AG76, ES99]

Layers grow by $\mathrm{br}(T)$ and information contracts by $(1-2\delta)^2$. So, whichever effect wins determines reconstruction.



Broadcasting on trees: lower bound

- The IT intuition above is awesome.
- Annoyance: lower bound is shown in a very different way!

Broadcasting on trees: lower bound

- The IT intuition above is awesome.
- Annoyance: lower bound is shown in a very different way!
- Kesten-Stigum bound [KS66]: Let

$$S = \sum_{v \in L_k} f(X_v)$$

  where $f =$ second eigenfunction of the noisy channel.
- For BSC: $f(\sigma) = \sigma, \sigma = \pm 1$.
- Analysis: $\mathbb{E}[S|X_0 = \pm + 1]$ and $\mathrm{Var}[S]$ can be computed easily due to choice of f.
- ... It shows the $X_0 = \pm 1$ can be separated if $\lambda_2^2 \, \mathrm{br}(T) > 1$.

Broadcasting on trees: lower bound

- The IT intuition above is awesome.
- Annoyance: lower bound is shown in a very different way!
- Kesten-Stigum bound [KS66]: Let

$$S = \sum_{v \in L_k} f(X_v)$$

  where $f =$ second eigenfunction of the noisy channel.
- For BSC: $f(\sigma) = \sigma, \sigma = \pm 1$.
- Analysis: $\mathbb{E}[S|X_0 = \pm + 1]$ and $\mathrm{Var}[S]$ can be computed easily due to choice of f.
- ... It shows the $X_0 = \pm 1$ can be separated if $\lambda_2^2 \operatorname{br}(T) > 1$.
- In other words, KS corresponds to a suboptimal majority-vote decoder.
- ... and thus results in a suboptimal P_e.
- ... but surprisingly recovers the right threshold for BSC (but not in general, e.g. for Potts with $q = 5$).

Broadcasting on trees: lower bound

- The IT intuition above is awesome.
- Annoyance: lower bound is shown in a very different way!
- Kesten-Stigum bound [KS66]: Let

$$S = \sum_{v \in L_k} f(X_v)$$

  where $f =$ second eigenfunction of the noisy channel.
- For BSC: $f(\sigma) = \sigma, \sigma = \pm 1$.
- Analysis: $\mathbb{E}[S|X_0 = \pm + 1]$ and $\mathrm{Var}[S]$ can be computed easily due to choice of f.
- ... It shows the $X_0 = \pm 1$ can be separated if $\lambda_2^2 \,\mathrm{br}(T) > 1$.
- In other words, KS corresponds to a suboptimal majority-vote decoder.
- ... and thus results in a suboptimal P_e.
- ... but surprisingly recovers the right threshold for BSC (but not in general, e.g. for Potts with $q = 5$).
- Can we analyze the optimal decoder? (without studying $T \circ S$)
  This is one goal of my talk

Some background first...

- BMS channels
- Channel comparison orders: degraded, more capable, less noisy

BMS channels

Definition

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)| - x) \qquad \forall x, y$$

- For example, BSC, BEC, BI-AWGN, but also...

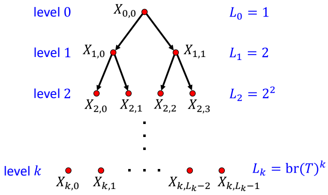
BMS channels

Definition

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)| - x) \qquad \forall x, y$$

- For example, BSC, BEC, BI-AWGN, but also...the $X_0 \to X_k$ channel!



BMS channels

Definition

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)| - x) \qquad \forall x, y$$

- For example, BSC, BEC, BI-AWGN, but also...the $X_0 \to X_k$ channel!

BMS channels

Definition

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)|-x) \qquad \forall x, y$$

- For example, BSC, BEC, BI-AWGN, but also...the $X_0 \to X_k$ channel!
- Let $X \sim \mathrm{Uniform}\{\pm 1\}$ and define
  1. $P_e = \mathbb{P}[X \neq \hat{X}_{ML}(Y)]$
  2. $C = I(X;Y)$
  3. $C_{\chi^2} = \chi^2(P_{Y|X=+1} \| P_Y)$

BMS channels

Definition

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)| - x) \qquad \forall x, y$$

- For example, BSC, BEC, BI-AWGN, but also...the $X_0 \to X_k$ channel!
- Let $X \sim \mathrm{Uniform}\{\pm 1\}$ and define
  1. $P_e = \mathbb{P}[X \neq \hat{X}_{ML}(Y)]$
  2. $C = I(X; Y)$
  3. $C_{\chi^2} = \chi^2(P_{Y|X=+1} \| P_Y)$
- Every BMS has a BSC-mixture representation:

$$Y = (\Delta, \mathsf{BSC}_\Delta(X)), \quad \Delta \sim P_\Delta \perp\!\!\!\perp X$$

BMS channels

Definition

$P_{Y|X} : \{\pm 1\} \to \mathcal{Y}$ called BMS if there is a bijection $h : \mathcal{Y} \to \mathcal{Y}$ s.t.

$$P_{Y|X}(y|x) = P_{Y|X}(h(y)| - x) \qquad \forall x, y$$

- For example, BSC, BEC, BI-AWGN, but also...the $X_0 \to X_k$ channel!
- Let $X \sim \mathrm{Uniform}\{\pm 1\}$ and define
  1. $P_e = \mathbb{P}[X \neq \hat{X}_{ML}(Y)] \qquad = \mathbb{E}[|1 - 2\Delta|]$
  2. $C = I(X; Y) \qquad = \log 2 - \mathbb{E}[h(\Delta)]$
  3. $C_{\chi^2} = \chi^2(P_{Y|X=+1} \| P_Y) \qquad = \mathbb{E}[(1 - 2\Delta)^2]$
- Every BMS has a BSC-mixture representation:

$$Y = (\Delta, \mathsf{BSC}_\Delta(X))\,, \quad \Delta \sim P_\Delta \perp\!\!\!\perp X$$

- The evolution operator $T \circ S$ described dist. of $\log \frac{1-\Delta}{\Delta}$.

Less noisy comparison

Definition

We say $P_{Y|X} \le_{ln} P_{Z|X}$ if for every $P_{U,X}$ we have

$$\begin{tikzcd}[row sep=small] & & Y \\ U \arrow[r] & X \arrow[ur] \arrow[dr] & \\ & & Z \end{tikzcd} \qquad \implies I(U;Y) \le I(U;Z)$$

Less noisy comparison

Definition

We say $P_{Y|X} \le_{ln} P_{Z|X}$ if for every $P_{U,X}$ we have

$$\begin{tikzcd}[row sep=small] & & Y \\ U \arrow[r] & X \arrow[ur] \arrow[dr] & \\ & & Z \end{tikzcd} \qquad \implies I(U;Y) \le I(U;Z)$$

- The meaning is that $P_{Z|X}$ is a better channel (in the sense above)
- Other partial orders exist: $P_{Y|X} \le_{deg} P_{Z|X}$, $P_{Y|X} \le_{mc} P_{Z|X}$ (degradation, more capable)
- ... we won't need them

Application of channel comparisons

- Consider a system processing X_0 into Y_1, Y_2 as follows (arrows are noisy channels):

$$\begin{tikzcd} & X_1 \arrow[r] & Y_1 \\ X_0 \arrow[ur] \arrow[dr] & & \\ & X_2 \arrow[r] & Y_2 \end{tikzcd}$$

Application of channel comparisons

- Consider a system processing X_0 into Y_1, Y_2 as follows (arrows are noisy channels):

$$\begin{tikzcd} & X_1 \arrow[r] & Y_1 \\ X_0 \arrow[ur] \arrow[dr] & & \\ & X_2 \arrow[r] & Y_2 \end{tikzcd}$$

- Suppose we replaced $X_1 \to Y_1$ with a less noisy channel $X_1 \to Z_1$

$$\begin{tikzcd} & X_1 \arrow[r, dashed] & \color{red}{Z_1} \\ X_0 \arrow[ur] \arrow[dr] & & \\ & X_2 \arrow[r] & Y_2 \end{tikzcd}$$

Question: Is the channel $X_0 \to (Z_1, Y_2)$ less noisy than $X_0 \to (Y_1, Y_2)$?

Application of channel comparisons

- Consider a system processing X_0 into Y_1, Y_2 as follows (arrows are noisy channels):

$$\begin{tikzcd} & X_1 \arrow[r] & Y_1 \\ X_0 \arrow[ur] \arrow[dr] & & \\ & X_2 \arrow[r] & Y_2 \end{tikzcd}$$

- Suppose we replaced $X_1 \to Y_1$ with a less noisy channel $X_1 \to Z_1$

$$\begin{tikzcd} & X_1 \arrow[r, dashed] & \color{red}{Z_1} \\ X_0 \arrow[ur] \arrow[dr] & & \\ & X_2 \arrow[r] & Y_2 \end{tikzcd}$$

Question: Is the channel $X_0 \to (Z_1, Y_2)$ less noisy than $X_0 \to (Y_1, Y_2)$? Yes!

$$\begin{tikzcd} & & & \color{red}{Z_1} \\ & & X_1 \arrow[ur, dashed] \arrow[r] & Y_1 \\ U \arrow[r] & X_0 \arrow[ur] \arrow[dr] & & \\ & & X_2 \arrow[r] & Y_2 \end{tikzcd}$$

$$\begin{aligned} I(U; Y_1, Y_2) &= I(U; Y_2) + I(U; Y_1 | Y_2) \\ &\le I(U; Y_2) + I(U; Z_1 | Y_2) \qquad \text{by def. of } \le_{ln} \\ &= I(U; Z_1, Y_2) \end{aligned}$$

Comparison method for analyzing networks

$$
\begin{tikzcd}[row sep=small]
 & X_1 \arrow[r] & Y_1 \\
X_0 \arrow[ur] \arrow[dr] & & \\
 & X_2 \arrow[r] & Y_2
\end{tikzcd}
\quad \Longrightarrow \quad
\begin{tikzcd}[row sep=small]
 & X_1 \arrow[r, dashed] & Z_1 \\
X_0 \arrow[ur] \arrow[dr] & & \\
 & X_2 \arrow[r, dashed] & Z_2
\end{tikzcd}
$$

- Meta-principle: Given a network, replace channels with less/more noisy.
- If this preserves less noisy relation, then get bounds on $I(X_0; Y_1, Y_2)$ etc.
- This is only useful if we can find simple channels $P_{Z|X}$

Comparison method for analyzing networks

$$\begin{tikzcd} & X_1 \arrow[r] & Y_1 \\ X_0 \arrow[ur] \arrow[dr] & & \\ & X_2 \arrow[r] & Y_2 \end{tikzcd} \quad \Longrightarrow \quad \begin{tikzcd} & X_1 \arrow[r, dashed] & Z_1 \\ X_0 \arrow[ur] \arrow[dr] & & \\ & X_2 \arrow[r, dashed] & Z_2 \end{tikzcd}$$

- Meta-principle: Given a network, replace channels with less/more noisy.
- If this preserves less noisy relation, then get bounds on $I(X_0; Y_1, Y_2)$ etc.
- This is only useful if we can find simple channels $P_{Z|X}$
- Alas, it is very hard to prove $\leq_{ln}$ relation...

Comparison method for analyzing networks

$$\begin{tikzcd}[row sep=small] & X_1 \arrow[r] & Y_1 \\ X_0 \arrow[ur] \arrow[dr] & & \\ & X_2 \arrow[r] & Y_2 \end{tikzcd} \quad \Longrightarrow \quad \begin{tikzcd}[row sep=small] & X_1 \arrow[r, dashed] & Z_1 \\ X_0 \arrow[ur] \arrow[dr] & & \\ & X_2 \arrow[r, dashed] & Z_2 \end{tikzcd}$$

- Meta-principle: Given a network, replace channels with less/more noisy.
- If this preserves less noisy relation, then get bounds on $I(X_0; Y_1, Y_2)$ etc.
- This is only useful if we can find simple channels $P_{Z|X}$
- Alas, it is very hard to prove $\leq_{ln}$ relation... Or is it?

Extremality of BSC and BEC

Theorem (Roozbehani-P.'2019)

1. *Among all BMS channels W with fixed P_e the BSC and BEC are extremal w.r.t. degradation:*

$$\mathsf{BSC}_{P_e} \le_{deg} W \le_{deg} \mathsf{BEC}_{2P_e}\,.$$

Extremality of BSC and BEC

Theorem (Roozbehani-P.'2019)

1. *Among all BMS channels W with fixed P_e the BSC and BEC are extremal w.r.t. degradation:*

$$\mathsf{BSC}_{P_e} \le_{deg} W \le_{deg} \mathsf{BEC}_{2P_e}\,.$$

2. *Among all BMS channels W with fixed C the BSC and BEC are extremal w.r.t. more capable:*

$$\mathsf{BSC}_{h^{-1}(1-C)} \le_{mc} W \le_{mc} \mathsf{BEC}_{1-C}\,.$$

Extremality of BSC and BEC

Theorem (Roozbehani-P.'2019)

1. *Among all BMS channels W with fixed P_e the BSC and BEC are extremal w.r.t. degradation:*

$$\mathrm{BSC}_{P_e} \le_{deg} W \le_{deg} \mathrm{BEC}_{2P_e}\,.$$

2. *Among all BMS channels W with fixed C the BSC and BEC are extremal w.r.t. more capable:*

$$\mathrm{BSC}_{h^{-1}(1-C)} \le_{mc} W \le_{mc} \mathrm{BEC}_{1-C}\,.$$

3. *Among all BMS channels W with fixed C_{χ^2} the BSC and BEC are extremal w.r.t. less noisy:*

$$\mathrm{BSC}_{1/2-\sqrt{C_{\chi^2}}} \le_{ln} W \le_{ln} \mathrm{BEC}_{1-C_{\chi^2}}\,.$$

Note: We only care about No.3 here, which is new!

Extremality of BSC and BEC

Theorem (Roozbehani-P.'2019)

1. *... P_e ... degradation ...*
2. *... C ... more capable ...*
3. *Among all BMS channels W with fixed C_{χ^2} the BSC and BEC are extremal w.r.t. less noisy:*

$$\mathrm{BSC}_{1/2-\sqrt{C_{\chi^2}}} \le_{ln} W \le_{ln} \mathrm{BEC}_{1-C_{\chi^2}}\,.$$

- In [RP19] we used this to analyze new non-linear sparse-graph codes (LDMCs).
- The proof in fact shows a version of Mrs. Gerbers Lemma:

  divergence $d(p * \delta \| q * \delta)$ is convex in $C_{\chi^2} = (1-2\delta)^2\ \forall p, q \in [0,1]$

  (Usual MGL: $q = 1/2$ and C_{χ^2} replaced with C)

Extremality of BSC and BEC

Theorem (Roozbehani-P.'2019)

1. *... P_e ... degradation ...*
2. *... C ... more capable ...*
3. *Among all BMS channels W with fixed C_{χ^2} the BSC and BEC are extremal w.r.t. less noisy:*

$$\mathrm{BSC}_{1/2-\sqrt{C_{\chi^2}}} \le_{ln} W \le_{ln} \mathrm{BEC}_{1-C_{\chi^2}} .$$

- In [RP19] we used this to analyze new non-linear sparse-graph codes (LDMCs).
- The proof in fact shows a version of Mrs. Gerbers Lemma:

  divergence $d(p * \delta \| q * \delta)$ is convex in $C_{\chi^2} = (1-2\delta)^2 \; \forall p, q \in [0,1]$

  (Usual MGL: $q = 1/2$ and C_{χ^2} replaced with C)
- We are ready to get rid of Kesten-Stigum

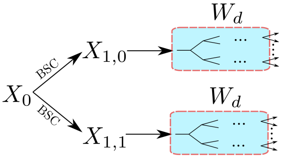
Theorem (sharp reconstruction lower bound)

Consider b-ary tree. If $b(1-2\delta)^2 > 1$ *then* $\liminf_{d\to\infty} I(X_0; X_{L_d}) > 0$

Theorem (sharp reconstruction lower bound)

Consider b-ary tree. If $b(1-2\delta)^2 > 1$ then $\liminf_{d\to\infty} I(X_0; X_{L_d}) > 0$

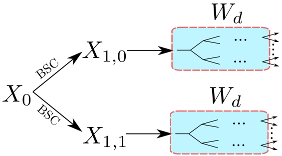
- Let $b = 2$ and $W_d =$ the BMS channel from X_0 to X_{L_d}. Note the recursive decomposition for W_{d+1} in terms of two W_d and BSC$_\delta$'s:



Theorem (sharp reconstruction lower bound)

Consider b-ary tree. If $b(1-2\delta)^2 > 1$ then $\liminf_{d\to\infty} I(X_0; X_{L_d}) > 0$

- Let $b = 2$ and W_d = the BMS channel from X_0 to X_{L_d}. Note the recursive decomposition for W_{d+1} in terms of two W_d and BSC_δ's:

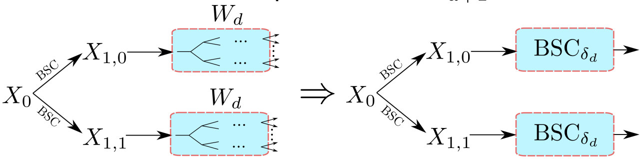


- Suppose (by induction) that $W_d \ge_{ln} \mathrm{BSC}_{\delta_d}$ for some δ_d. Then apply channel comparison to get:

Theorem (sharp reconstruction lower bound)

Consider b-ary tree. If $b(1-2\delta)^2 > 1$ then $\liminf_{d\to\infty} I(X_0; X_{L_d}) > 0$

- Let $b = 2$ and W_d = the BMS channel from X_0 to X_{L_d}. Note the recursive decomposition for W_{d+1} in terms of two W_d and BSC$_\delta$'s:



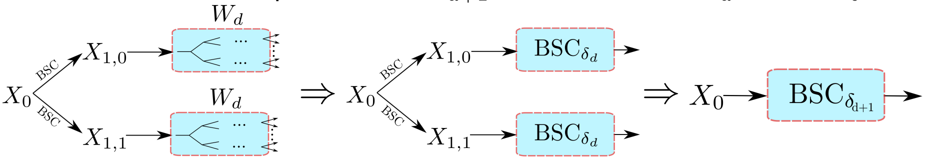
- Suppose (by induction) that $W_d \geq_{ln}$ BSC$_{\delta_d}$ for some δ_d. Then apply channel comparison to get:

$$W_{d+1} \geq_{ln} \text{two parallel } \mathrm{BSC}_{\delta_d * \delta}$$

Theorem (sharp reconstruction lower bound)

Consider b-ary tree. If $b(1-2\delta)^2 > 1$ then $\liminf_{d\to\infty} I(X_0; X_{L_d}) > 0$

- Let $b = 2$ and $W_d =$ the BMS channel from X_0 to X_{L_d}. Note the recursive decomposition for W_{d+1} in terms of two W_d and BSC_δ's:



- Suppose (by induction) that $W_d \geq_{ln} \mathrm{BSC}_{\delta_d}$ for some δ_d. Then apply channel comparison to get:

$$W_{d+1} \geq_{ln} \text{two parallel } \mathrm{BSC}_{\delta_d * \delta} \geq_{ln} \mathrm{BSC}_{\delta_{d+1}},$$

  where $\delta_{d+1} \triangleq J(\delta_d)$ for some explicit $J(x)$.

- Starting from $\delta_0 = 0$, analysis shows $\delta_\infty < 1/2$. Thus, for all d we have $W_d \geq_{ln} \mathrm{BSC}_{\delta_\infty}$

Define two quantities for $\delta < \delta_{crit}(b) = \frac{1}{2} - \sqrt{\frac{1}{4b}}$:

$$P_e(\delta) \triangleq \lim_{d\to\infty} \mathbb{P}[X_0 \neq \hat{X}_0(X_{L_d})]$$

$$I(\delta) \triangleq \lim_{d\to\infty} I(X_0; X_{L_d})$$

- In physics, behavior of quantities near the phase transition is often universal, e.g. *critical exponents*.
- So we ask: What are α, β, γ?

$$P_e(\delta_{crit} - \tau) = 1/2 - \Theta(\tau^\alpha)$$
$$I(\delta_{crit} - \tau) = (\gamma + o(1))\tau^\beta$$

Define two quantities for $\delta < \delta_{crit}(b) = \frac{1}{2} - \sqrt{\frac{1}{4b}}$:

$$P_e(\delta) \triangleq \lim_{d\to\infty} \mathbb{P}[X_0 \neq \hat{X}_0(X_{L_d})]$$

$$I(\delta) \triangleq \lim_{d\to\infty} I(X_0; X_{L_d})$$

- In physics, behavior of quantities near the phase transition is often universal, e.g. *critical exponents*.
- So we ask: What are α, β, γ?

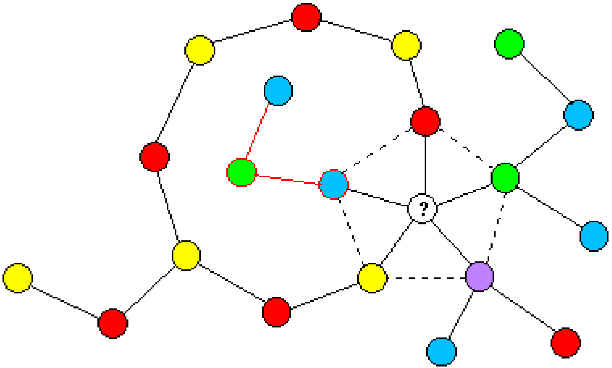
$$P_e(\delta_{crit} - \tau) = 1/2 - \Theta(\tau^\alpha)$$
$$I(\delta_{crit} - \tau) = (\gamma + o(1))\tau^\beta$$

- Previously: $\beta = 1$, $1/2 \le \alpha \le 1$, some loose bounds on γ.
- Our methods (rigorous, except for finite precision arithmetic):

$$\gamma \approx 8\sqrt{2}, \qquad 1/2 \le \alpha \le 0.504$$

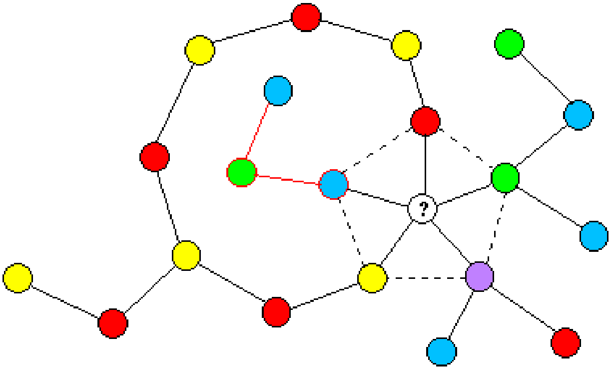
Next: Reconstruction on sparse graphs

Reconstructing random colorings



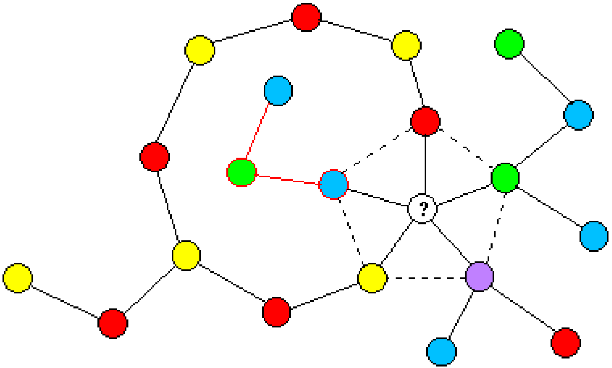
- Consider large sparse graph with randomly colored vertices
- Local rule: adjacent vertices have distinct colors
- Global question: Are there long-range dependencies?

Reconstructing random colorings



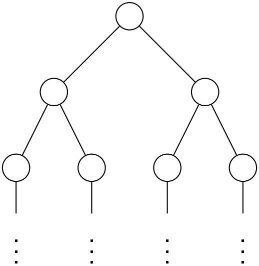
- Consider large sparse graph with randomly colored vertices
- Local rule: adjacent vertices have distinct colors
- Global question: Are there long-range dependencies?
- More exactly: Can we predict color of a vertex given colors of its far-away neighbors?

Reconstructing random colorings



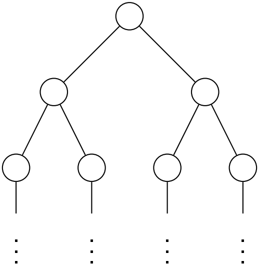
- Consider large sparse graph with randomly colored vertices
- Local rule: adjacent vertices have distinct colors
- Global question: Are there long-range dependencies?
- More exactly: Can we predict color of a vertex given colors of its far-away neighbors?
- ... if graph is locally tree-like we get a BoT question!

Reconstructing random colorings: achievability



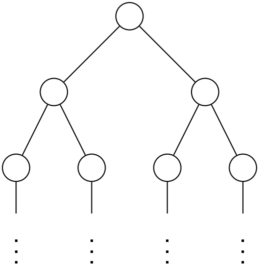
- Suppose we have k colors and regular graph of degree $d+1$.
- if $d \geq (1+o(1))k \log k$ then w.h.p. each node has among its descendants all colors except its own.

Reconstructing random colorings: achievability



- Suppose we have k colors and regular graph of degree $d + 1$.
- if $d \geq (1 + o(1))k \log k$ then w.h.p. each node has among its descendants all colors except its own.
- ... then can work backwards and reconstruct root color with certainty

Reconstructing random colorings: achievability

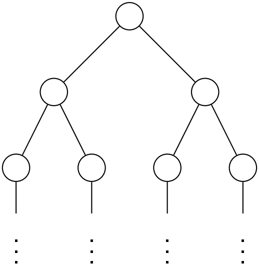


- Suppose we have k colors and regular graph of degree $d+1$.
- if $d \geq (1+o(1))k \log k$ then w.h.p. each node has among its descendants all colors except its own.
- ... then can work backwards and reconstruct root color with certainty
- ... i.e. BP message to the root has zero-entropy.
- So when

  $$d \geq (1+o(1))k \log k$$

  we can reconstruct! Is this tight?

Reconstructing random colorings: achievability



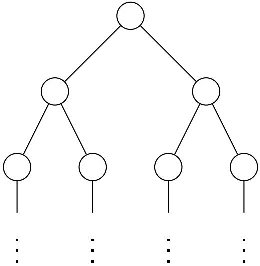
- Suppose we have k colors and regular graph of degree $d+1$.
- if $d \geq (1+o(1))k \log k$ then w.h.p. each node has among its descendants all colors except its own.
- ... then can work backwards and reconstruct root color with certainty
- ... i.e. BP message to the root has zero-entropy.
- So when
$$d \geq (1+o(1))k \log k$$
we can reconstruct! Is this tight?
- Yes! Two long papers: [Sly '09], [Bhatnagar-Vera-Vigoda-Weitz'11]

Broadcasting on trees: General edge channel

- Infinite tree $\mathcal{T}$ with marked root ρ.
- Reversible Markov kernel $W : [k] \to [k]$ with invariant distribution q^*.
- Each node has a color in $[k]$, where
  - Root color has distribution q^*.
  - Color of any non-root node is generated from color of its parent by applying W.
- We say the model has non-reconstruction if

$$\lim_{h \to \infty} I(\rho; L^h) = 0,$$

where L^h is the set of nodes on level h.

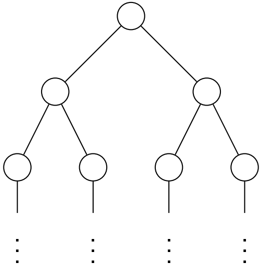


Broadcasting on trees: General edge channel

- Infinite tree $\mathcal{T}$ with marked root ρ.
- Reversible Markov kernel $W : [k] \to [k]$ with invariant distribution q^*.
- Each node has a color in $[k]$, where
  - Root color has distribution q^*.
  - Color of any non-root node is generated from color of its parent by applying W.
- We say the model has non-reconstruction if

$$\lim_{h\to\infty} I(\rho; L^h) = 0,$$

  where L^h is the set of nodes on level h.
- Note: For k-coloring channel $W(y|x) = \frac{1}{k-1}1\{y \neq x\}$



Broadcasting on trees: Result

Theorem (G.-Polyanskiy '19)

Let $\mathrm{br}(\mathcal{T})$ *be the branching number of the tree. Then we have non-reconstruction if*

$$\eta_{\mathsf{KL}}(q^*, W)\,\mathrm{br}(\mathcal{T}) < 1.$$

- If $\mathcal{T}$ is a d-regular tree or a Galton-Watson tree with expected offspring d, then $\mathrm{br}(\mathcal{T}) = d$.

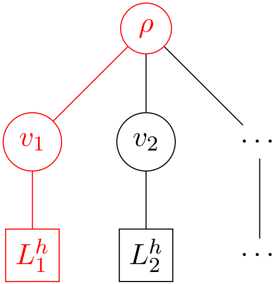
- Apply SDPI to the Markov Chain

$$L_i^h \to v_i \xrightarrow{W} \rho,$$

and get $I(\rho; L_i^h) \le \eta_{\mathsf{KL}}(q^*, W) I(v_i; L_i^h)$.

- Use conditional independence.

$$\begin{aligned} I(\rho; L^h) &\le \sum_i I(\rho; L_i^h) \\ &\le \sum_i \eta_{\mathsf{KL}}(q^*, W) I(v_i; L_i^h) \\ &= d\eta_{\mathsf{KL}}(q^*, W) I(\rho; L^{h-1}). \end{aligned}$$



- Apply induction.

Broadcasting on trees: Comparison

- For the coloring channel $W_{i,j} = \frac{1}{k-1}\mathbb{1}\{i \neq j\}$, we obtain non-reconstruction for

$$d < \frac{\log k}{\log k - \log(k-1)} = (1 - o(1))k \log k.$$

- ... Sharp!

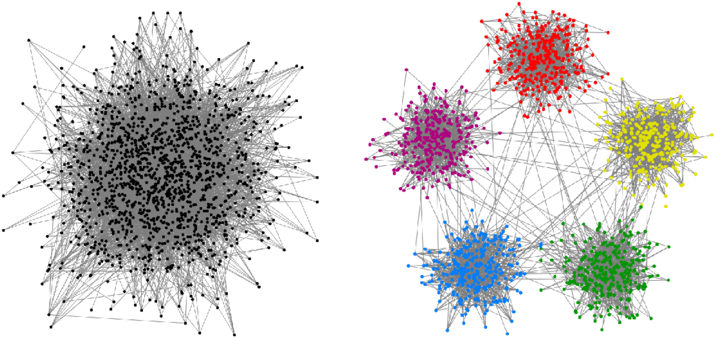
Broadcasting on trees: Comparison

- For the coloring channel $W_{i,j} = \frac{1}{k-1}\mathbb{1}\{i \neq j\}$, we obtain non-reconstruction for

$$d < \frac{\log k}{\log k - \log(k-1)} = (1 - o(1))k \log k.$$

- ... Sharp!
- For Potts channels and binary asymmetric channels, we obtain better numerical values for small d.
- Our results are non-asymptotic in k, d, and work for arbitrary trees.

Application: Community detection



- Unsupervised clustering problem
- See: $0/1$ similarity (i.e. graph)
- Want: Are there any clusters?

Stochastic block model

- A Model for community detection: symmetric k-SBM(a, b), $a, b > 0$

Stochastic block model

- A Model for community detection: symmetric k-SBM(a, b), $a, b > 0$
- n vertices, each assigned a uniformly random color in $[k]$.

Stochastic block model

- A Model for community detection: symmetric k-SBM(a, b), $a, b > 0$
- n vertices, each assigned a uniformly random color in $[k]$.
- A random graph $\mathbb{G}$ with indepedently selected edges

$$\mathbb{P}[(u,v) \in E(G)] = \begin{cases} \frac{a}{n}, & \text{if } u, v \text{ have same label} \\ \frac{b}{n}, & \text{o/w} \end{cases}$$

Stochastic block model

- A Model for community detection: symmetric k-SBM(a, b), $a, b > 0$
- n vertices, each assigned a uniformly random color in $[k]$.
- A random graph $\mathbb{G}$ with indepedently selected edges

$$\mathbb{P}[(u, v) \in E(G)] = \begin{cases} \frac{a}{n}, & \text{if } u, v \text{ have same label} \\ \frac{b}{n}, & \text{o/w} \end{cases}$$

- We say weak recovery is possible for parameters (k, a, b) if there exists $\epsilon > 0$ such that, with high probability, given the graph $\mathbb{G}$, we can construct a partition of the vertex set that is correct for at least ϵn vertices.

Stochastic block model: Result

Theorem (G.-Polyanskiy '19)

Let $d = \frac{a+(k-1)b}{k}$, $\lambda = \frac{a-b}{a+(k-1)b}$. *Weak recovery is impossible if*

$$d\eta_{\mathsf{KL}}(\mathrm{PC}_\lambda, q^*) < 1,$$

where PC_λ *is the Potts channel defined by*

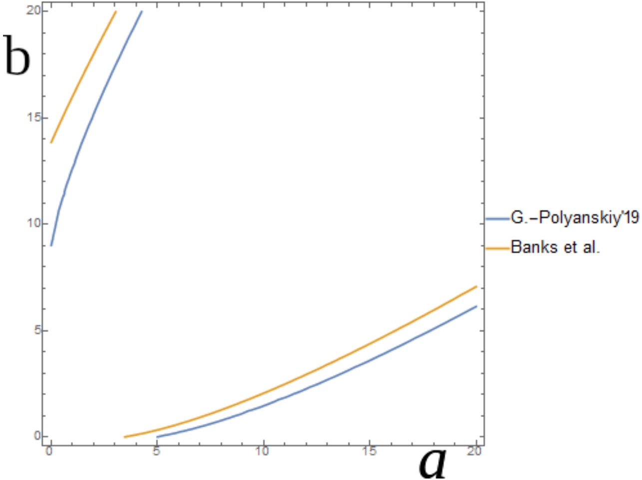
$$\mathrm{PC}_\lambda(x,y) = \frac{1-\lambda}{k}\mathbb{1}\{x \neq y\} + (\frac{1}{k} + \frac{k-1}{k}\lambda)\mathbb{1}\{x = y\}.$$

and q^* *is the uniform distribution.*

Proof.

Reduction from broadcasting on trees. ∎

Stochastic block model: Comparison



- For $k \geq 3$, $a > b$, we improve the state-of-the art [Banks et al. '16].
- Note: for $a < b$, exact threshold is known [Coja-Oghlan et al. '19.]

Conclusion

- Machine learning: exciting new **local to global** problems
- ... sometimes called combinatorial statistics
- Obvious connections with statistical physics

Conclusion

- Machine learning: exciting new **local to global** problems
- ... sometimes called combinatorial statistics
- Obvious connections with statistical physics
- Information theory: excellent tools for these problems
- Previously: only on the negative (impossibility) side and "easy" problems

Conclusion

- Machine learning: exciting new **local to global** problems
- ... sometimes called combinatorial statistics
- Obvious connections with statistical physics
- Information theory: excellent tools for these problems
- Previously: only on the negative (impossibility) side and "easy" problems
- New: channel comparison, SDPI, info-percolation
- ... positive results, sharp threshoulds, hard models

Thank You!

Rudolf Ahlswede and Péter Gács.
Spreading of sets in product spaces and hypercontraction of the Markov operator.
The Annals of Probability, 4(6):925–939, December 1976.

Pavel M. Bleher, Jean Ruiz, and Valentin A. Zagrebnov.
On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice.
Journal of Statistical Physics, 79(1-2):473–482, April 1995.

William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman.
Broadcasting on trees and the Ising model.
The Annals of Applied Probability, 10(2):410–433, May 2000.

William S. Evans and Leonard J. Schulman.
Signal propagation and noisy circuits.
IEEE Transactions on Information Theory, 45(7):2367–2373, November 1999.

William S. Evans and Leonard J. Schulman.
On the maximum tolerable noise of k-input gates for reliable computation by formulas.
IEEE Transactions on Information Theory, 49(11):3094–3098, November 2003.

Bruce Hajek and Timothy Weller.
On the maximum tolerable noise for reliable computation by formulas.
IEEE Transactions on Infomation Theory, 37(2):388–391, March 1991.

Harry Kesten and Bernt P. Stigum.
A limit theorem for multidimensional Galton-Watson processes.
The Annals of Mathematical Statistics, 37(5):1211–1223, October 1966.

Hajir Roozbehani and Yury Polyanskiy.
Low density majority codes and the problem of graceful degradation.
arXiv preprint arXiv:1911.12263, 2019.

Falk Unger.
Noise threshold for universality of 2-input gates.
In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pages 1901–1905, Nice, France, June 24-29 2007.

John von Neumann.
Probabilistic logics and the synthesis of reliable organisms from unreliable components.
In Claude E. Shannon and John McCarthy, editors, *Automata Studies*, volume 34 of *Annals of Mathematics Studies*, pages 43–98, Princeton, NJ, USA, 1956. Princeton University Press.