* Mixing Time of a Markov Chain.

Def. <u>ε-mixing time</u> of a Markov Chain Q

is the smallest time $T_{mix}(\varepsilon)$ such that for all $t > T_{mix}(\varepsilon)$

and for all initial state $q^{(0)}$ (a distribution over $\mathcal{X}^n$)

$$\Big| \big(q^{(0)}\big)^T \cdot \underset{Q\cdot Q \cdots Q}{\underset{\uparrow}{Q^t}} - \pi^T \Big|_{TV} \le \varepsilon$$

π^T: stationary distribution

where $|p-q|_{TV} = \frac{1}{2}\sum_{x \in \mathcal{X}^n} |p(x) - q(x)|$ is the <u>total variation distance</u>.

* Bounding mixing time via <u>Coupling</u>

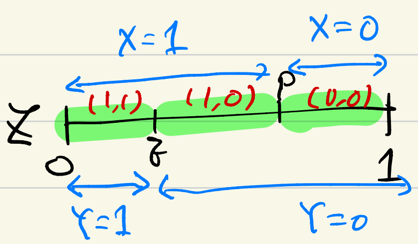
Def. a coupling of two random variables X and Y with distributions $P_X(x)$ and $P_Y(y)$, is a construction of a joint probability distribution over (X,Y), i.e., $P_{XY}(x,y)$ such that the marginals are preserved:

$$\begin{cases} \sum_y P_{XY}(x,y) = P_X(x) \\ \sum_x P_{XY}(x,y) = P_Y(y) \end{cases}$$

example> $P_X \sim Bern(p)$
$P_Y \sim Bern(q)$

$$P_{XY}: \quad \begin{array}{c|cc|cc} & 1-p & p & & \\ & X=0 & 1 & Y & \\ \hline & \left[\begin{array}{c|c} (1-p)q & p(1-q) \\ \hline (1-p)(1-q) & pq \end{array}\right] & & \begin{array}{c} =0 \\ 1 \end{array} & \begin{array}{c} 1-q \\ q \end{array} \end{array}$$

"optimal" coupling. $\underline{\underline{p > q}}$



$$\begin{array}{cc} \left[\begin{array}{c|c} 1-p & p-q \\ \hline 0 & q \end{array}\right] & \begin{array}{cc} Y & \\ 0 & 1-q \\ 1 & q \end{array} \\ \begin{array}{cc} X=0 & 1 \\ 1-p & p \end{array} & \end{array}$$

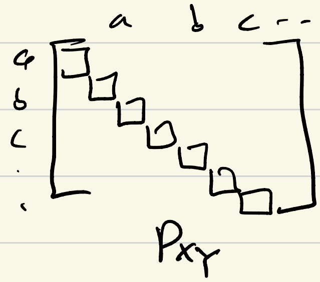
* How to sample from "optimal" coupling.

Step 1: draw $Z \sim U[0,1]$: shared randomness

Step 2: $X = \begin{cases} 1 & Z \le p \\ 0 & \text{otherwise} \end{cases}$, $Y = \begin{cases} 1 & Z \le q \\ 0 & \text{otherwise} \end{cases}$

Def. Optimal Coupling for $X, Y \in \mathcal{X}$, given P_X, P_Y

$$\min_{\substack{\text{all couplings} \\ \text{s.t. marginals being } P_X \text{ and } P_Y}} P_{XY}(X \neq Y)$$



Claim: Coupling Lemma.

for 2 r.v.s X, Y, we have

$$|P_X - P_Y|_{TV} = \min_{\text{couplings}} P_{XY}(X \neq Y)$$

proof>

$$P_{XY}(X \neq Y) = 1 - \sum_{a \in \mathcal{X}} \underbrace{P_{XY}(a,a)}_{\leq \min\{P_X(a), P_Y(a)\}}$$

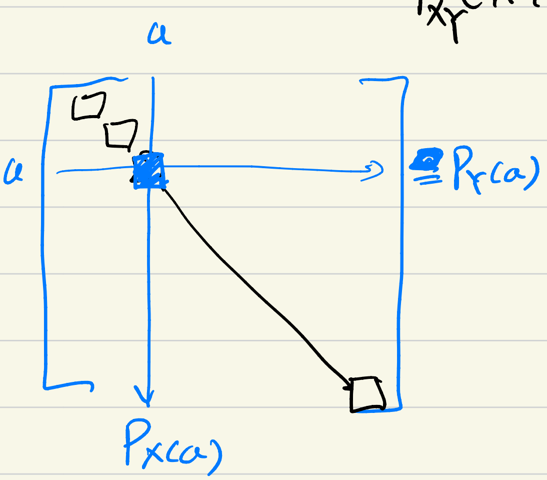
$$\sum_a P_X(a) = 1$$

$$\geq 1 - \sum_a \min\{P_X(a), P_Y(a)\}$$

$$= \sum_a \left\{ P_X(a) - \min\{P_X(a), P_Y(a)\} \right\}$$

$$= \sum_a \max\left\{ 0, P_X(a) - P_Y(a) \right\}$$

$$\stackrel{\Delta}{=} |P_X - P_Y|_{TV}$$



* Corollary:

$$|P_{X_t} - P_{Y_t}|_{TV} \leq P_{X_t Y_t}(X_t \neq Y_t), \quad \text{any coupling } P_{XY}.$$

* Coupling for Bounding $T_{mix}(\varepsilon)$ of Gibbs Sampling

Strategy: coupled Markov chains. $(X_0, Y_0) \dashrightarrow (X_t, Y_t)$

$$\| P_{X_t} - \pi \|_{TV} \le \max_{P_{X_0}, \underset{\underset{\pi}{\uparrow}}{P_{Y_0}}} \| P_{X_t} - \underset{\underset{\pi}{\uparrow}}{P_{Y_t}} \|_{TV}$$

Coupling lemma $\rightarrow$ $\le \max_{P_{X_0} \cdot P_{Y_0}} \underline{\mathbb{P}_{X_t Y_t}(X_t \neq Y_t)}$

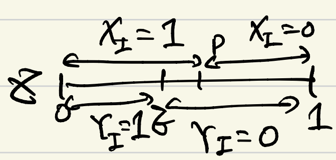
$\uparrow$ easier to analyze.

*Proposed Coupling. $\mathcal{X} = \{0, 1\}$.

Initializations	X_0	Y_0
Repeat $t = 1, \dots$	$(X^{(t)}$,	$Y^{(t)})$

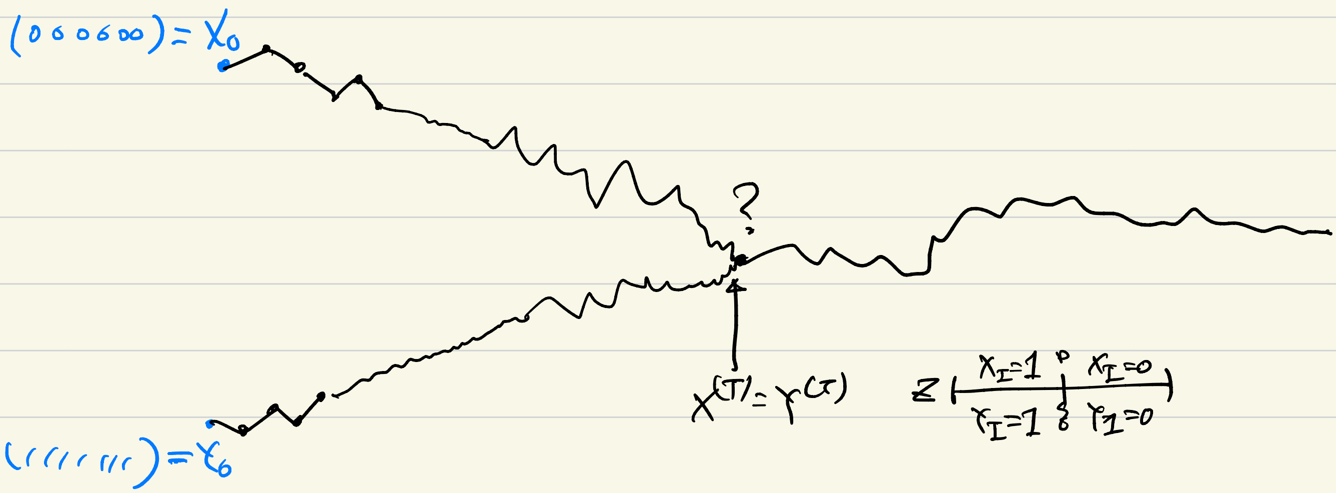
Step 1: sample $I \in \{1, \dots, n\}$

Step 2: draw $(X_I^{(t+1)}, Y_I^{(t+1)})$ from optimal coupling



$$P_{XY} \quad \begin{array}{cc|c} & & Y_I^{(t+1)} \\ \left[\begin{array}{c|c} \min\{1-p, 1-q\} & \max\{0, p-q\} \\ \hline \max\{0, q-p\} & \min\{p, q\} \end{array}\right] & & \begin{array}{c} 0 \\ 1 \Leftarrow \mathbb{P}(Y_I^{(t+1)} \mid \underset{\underset{q}{\parallel}}{Y_{-I}^{(t)}}) \end{array} \\ \begin{array}{ccc} X_I^{(t+1)} & 0 & 1 \end{array} & & \end{array}$$

$1 \Uparrow \mathbb{P}(X_I^{(t+1)} \mid X_{-I}^{(t)}) = p$

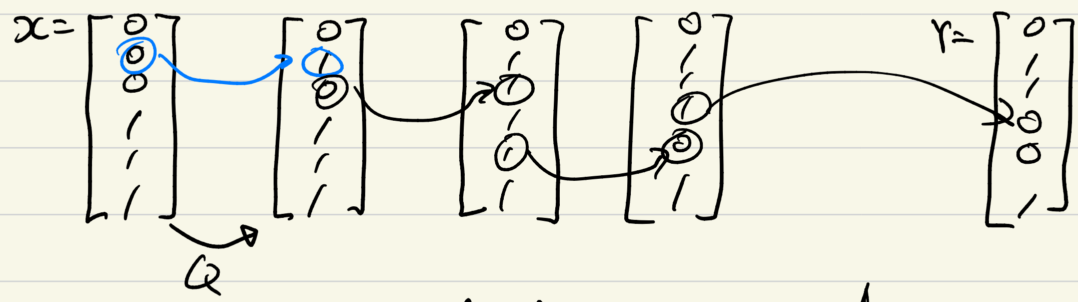


* Bounding $\underline{\underline{\mathbb{P}(X_t \neq Y_t)}}$ by Path Coupling. [Bubley, Dyer, 1997 FOCS]

Def. $D(X,Y)$ is the minimum # of transitions on M.C to move from x to Y.

For Gibbs Sampling. $D(X,Y)$ = Hamming Distance (X,Y)

$\triangleq$ # of different entries b.tw. X & Y.



4 transitions. $= d_H(x,y)$

* Strategy: ① to show

$$\mathbb{E}_{(X_{t+1},Y_{t+1})}\left[D(X_{t+1},Y_{t+1}) \mid X_t, Y_t\right] \leq \alpha \cdot D(X_t, Y_t)$$

for some $\alpha < 1$.

② $$\left| P_{X_{t+1}} - P_{Y_{t+1}} \right|_{TV} \overset{\text{coupling Lemma}}{\leq} P_{X_{t+1},Y_{t+1}}(X_{t+1} \neq Y_{t+1})$$

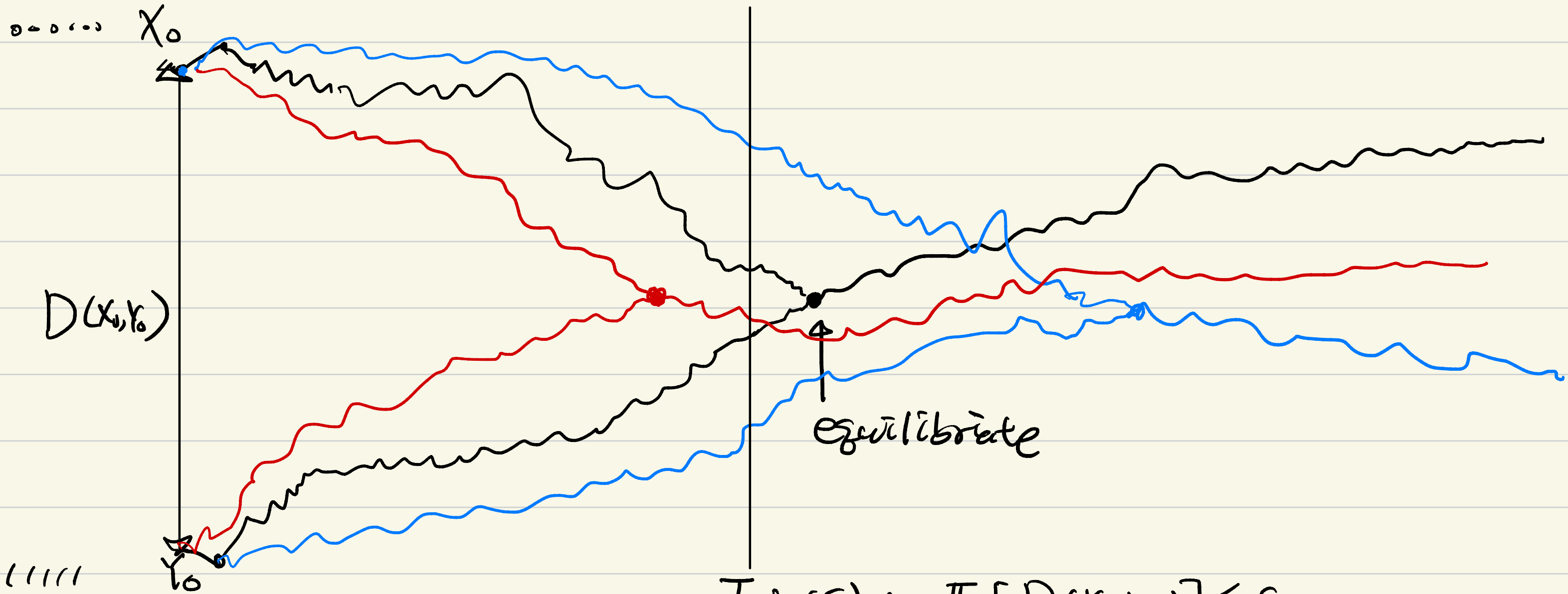
$D(x,y) = \begin{cases} 0 & x=y \\ \geq 1 & x\neq y \end{cases} \longrightarrow$ $$\leq \mathbb{E}_{(X_{t+1},Y_{t+1})}\left[D(X_{t+1},Y_{t+1})\right]$$

Towering $\longrightarrow$ $$= \mathbb{E}_{(X_1,Y_1)}\Big[\cdots \mathbb{E}_{(X_t,Y_t)}\Big[\underbrace{\mathbb{E}_{X_{t+1},Y_{t+1}}\left[D(X_{t+1},Y_{t+1}) \mid (X_t,Y_t)\right]}_{① \to\ \leq D(X_t,Y_t)\times\alpha}\Big] \Big| (X_{t-1},Y_{t-1})\Big] \cdots \Big| X_0,Y_0\Big]$$

$$\leq \alpha^{t+1} \cdot D(X_0,Y_0) \leq \varepsilon$$

$$t \geq \frac{\log \frac{\cancelto{n}{D(X_0,Y_0)}}{\varepsilon}}{\log 1/\alpha} = T_{mix}(\varepsilon), \longrightarrow |P_{X_t} - P_{Y_t}|_{TV} \leq \varepsilon$$

Intuitively,



$T_{mix}(\varepsilon)$: $\mathbb{E}[D(X_t,Y_t)] \le \varepsilon$

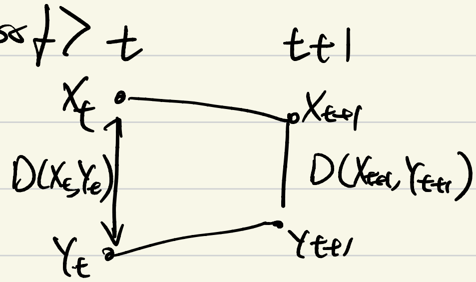
$\Rightarrow \|P_{X_t} - P_{Y_t}\|_{TV} \le \varepsilon$

$\Rightarrow \|P_{X_t} - \pi\|_{TV} \le \varepsilon$

* Claim: If $\mathbb{E}[D(X_{t+1},Y_{t+1}) \mid D(X_t,Y_t)=1] \le \alpha$

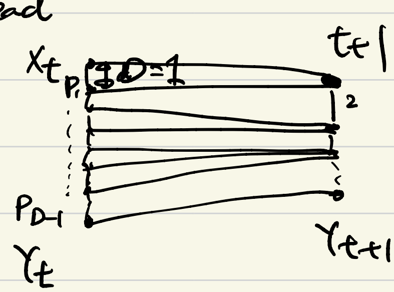
then $\mathbb{E}[D(X_{t+1},Y_{t+1}) \mid X_t, Y_t] \le \alpha \cdot D(X_t,Y_t) \Leftrightarrow$ ①

proof)



complicated.

Instead



Consider: a path $P = (X_t, P_1, P_2, \text{-----}, P_{D(X_t,Y_t)-1}, Y_t)$

Triangular Ineq.

let $P' = (X_{t+1}, P_1', P_2', \text{---}, Y_{t+1})$

$$\mathbb{E}[D(X_{t+1},Y_{t+1}) \mid X_t,Y_t] \le \mathbb{E}[\underbrace{D(X_{t+1},P_1')}_{\le\alpha} + \underbrace{D(P_1',P_2')}_{\le\alpha} + \cdots + \underbrace{D(P_{last}',Y_{t+1})}_{\le\alpha}]$$

by supposition

$$\le \alpha \cdot D(X_t,Y_t).$$

* We are left to find α s.t.

$$\mathbb{E}[D(X_{t+1}, Y_{t+1}) \mid D(X_t, Y_t)=1] \leq \alpha.$$

* Concrete Example > Ising Model : $(G, \{\theta_{ij}\})$

$$P(x) = \frac{1}{Z} \prod_{(i,j)\in E} e^{\theta_{ij} x_i x_j}, \quad x_i \in \{+1, -1\}$$

* claim: $\mathbb{E}[D(X_{t+1}, Y_{t+1}) \mid D(X_t, Y_t)=1] \leq 1 - \frac{1 - d_{max} \cdot |\tanh(\theta_{max})|}{n}$

$d_{max} \cdot |\tanh(\theta_{max})| < 1$, then $\alpha < 1$, fast mixing.

* Proof > Gibbs Sampling: (X, Y), assume $X_{-i} = Y_{-i}$, $X_i \neq Y_i$

step 1: draw $I \in \{1, \dots, n\}$

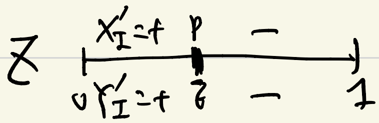
step 2: $Z \sim U[0,1]$

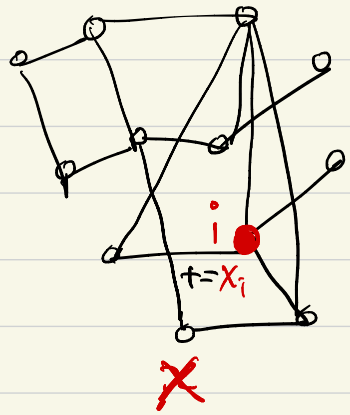
step 3. $X'_I = \begin{cases} + & \text{if } Z \leq \mathbb{P}(X'_I = +1 \mid X'_{-I} = X_{-I}) \\ - & \end{cases}$

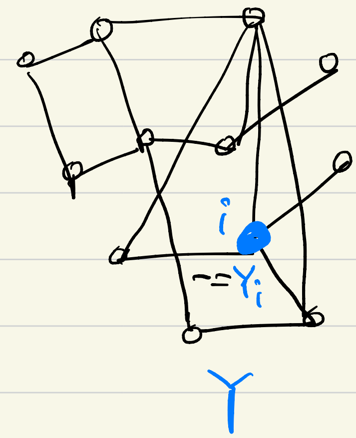
$Y'_I = \begin{cases} + & \text{if } Z \leq \mathbb{P}(Y'_I = +1 \mid Y'_{-I} = Y_{-I}) \\ - & \end{cases}$

upper bound $\mathbb{E}[D(X', Y')] \leq$

Case 1: $I = i$ (which happens w.p $\frac{1}{n}$)

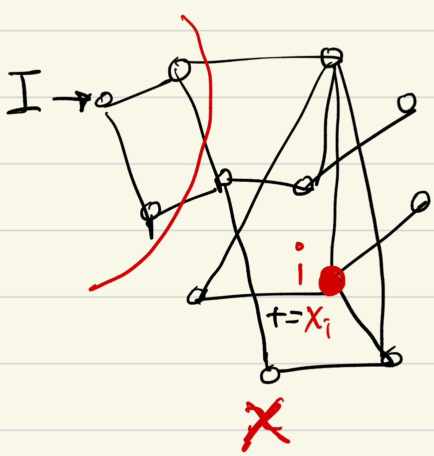


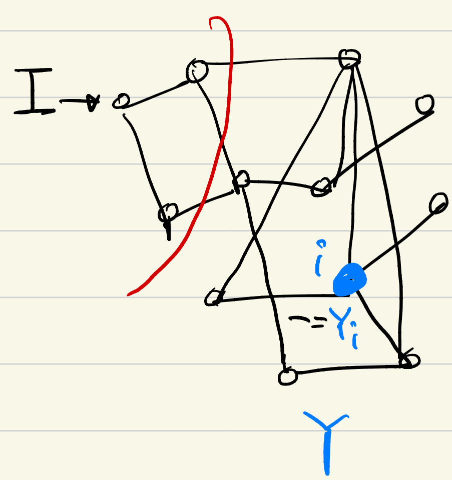


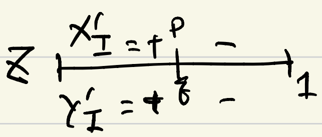


$$X' = Y' \longrightarrow \mathbb{E}[D(X', Y') \mid \underbrace{I = i}_{\mathbb{P}(I=i) = \frac{1}{n}}, X_{-i} = Y_{-i}] = 0$$

Case 2: $I \notin \{i\} \cup \partial i$



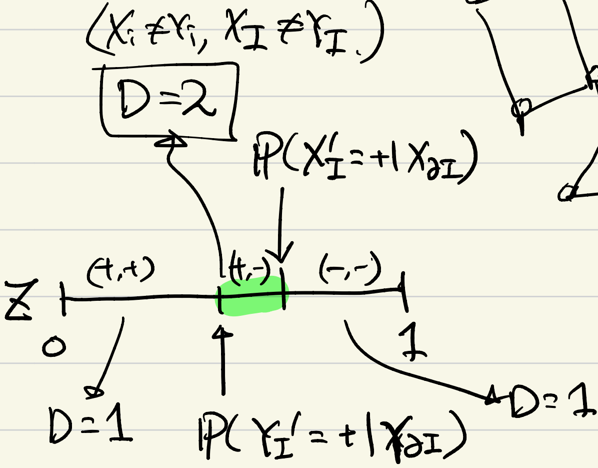


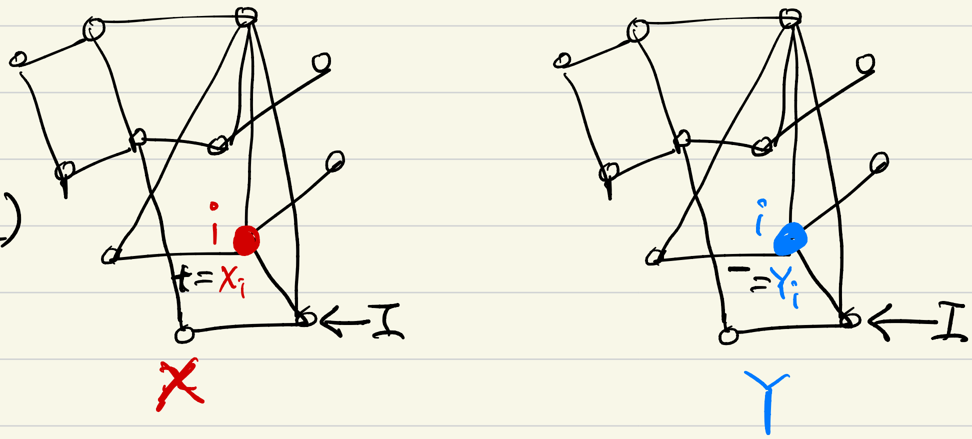


$X'_I = Y'_I$.

$$\mathbb{E}[D(X', Y') \mid X_{-i} = Y_{-i}, \underbrace{I \notin (\{i\} \cup \partial i)}_{\mathbb{P}(\quad) = 1 - \frac{1 + |\partial i|}{n}}] = 1$$

Case 3: $I \in \partial i$





$$\left| \mathbb{P}(X_I'=+|X_{\partial I}) - \mathbb{P}(Y_I'=+|X_{\partial I}) \right| \qquad \begin{cases} X_i=+ \\ Y_i=- \\ X_{-i}=Y_{-i} \end{cases}$$

Bayes $\rightarrow$ $\|$

$$\left| \frac{A^{(+)} \cdot f_{iI}(\overset{X_i}{\overset{\|}{+}},+)}{A^{(+)}\cdot f_{iI}(+,+) + A^{(-)}\cdot f_{iI}(+,-)} - \frac{A^{(+)} f_{iI}(\overset{Y_i}{\overset{\|}{-}},+)}{A^{(+)} f_{iI}(-,+) + A^{(-)} f_{iI}(-,-)} \right| \quad (**)$$

$$A^{(+)} = \prod_{j\in\partial I\setminus\{i\}} f_{jI}(\underset{\underset{Y_j}{\|}}{X_j}, \overset{X_I}{\overset{\|}{+}}), \quad A^{(-)} = \prod_{j\in\partial I\setminus\{i\}} f_{jI}(\underset{\underset{Y_j}{\|}}{X_j}, \overset{X_I}{\overset{\|}{-}})$$

Claim: $(**) \leq |\tanh(\theta_{iI})|$

$$(**) \quad \left| \frac{A^{(+)} e^{\overset{\theta}{\overset{\|}{\theta_{iI}}}}}{A^{(+)}e^{\theta} + A^{(-)}e^{-\theta}} - \frac{A^{(+)}e^{-\theta}}{A^{(+)}e^{-\theta} + A^{(-)}e^{\theta}} \right|$$

$$= \left| \frac{\overset{1}{\boxed{A^{(+)}A^{(-)}}}\,(e^{2\theta} - e^{-2\theta})}{(A^{(+)})^2 + (A^{(-)})^2 + \underset{1}{\boxed{A^{(+)}A^{(-)}}}\cdot(e^{2\theta}+e^{-2\theta})} \right|$$

$$\leq \left| \frac{e^{2\theta} - e^{-2\theta}}{2 + e^{2\theta} + e^{-2\theta}} \right|$$

$$A^{(+)}\cdot A^{(-)} = \prod_{j\in\partial I\setminus\{i\}} e^{\theta_{ij}} \cdot e^{-\theta_{ij}} = 1$$

$$= \left|\tanh(\underset{\theta_{iI}}{\theta})\right|$$

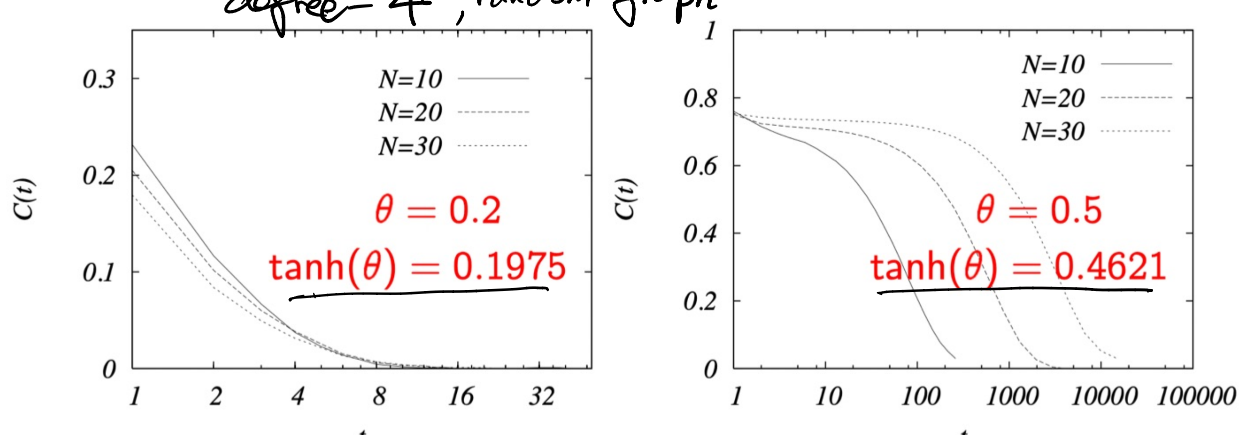
$$(A^{(+)})^2 + (A^{(-)})^2$$
$$= (A^{(+)})^2 + \left(\frac{1}{A^{(+)}}\right)^2$$
$$= \left(A^{(+)} - \frac{1}{A^{(+)}}\right)^2 + 2 \geq 2$$

* Putting 3 cases together.

$$\mathbb{E}[\, D(X', Y') \mid X_{-i} = Y_{-i} \,]$$
$$\leq \frac{1}{n} \cdot 0 + \left(1 - \frac{1 + \deg(i)}{n}\right) \cdot 1$$
$$+ \frac{1}{n} \sum_{j \in \partial i} \left\{ 2 \cdot |\tanh(\theta_{ij})| + 1\left(1 - |\tanh(\theta_{ij})|\right)\right\}$$
$$\leq \boxed{1 - \frac{1}{n} + \frac{1}{n} \deg(i) \cdot |\tanh(\theta_{\max})| \triangleq \alpha}$$

$$1 < \deg_{\max} \cdot \tanh(\theta_{\max})$$

degree = 4, random graph



$$C(t) = \frac{1}{|V|} \sum_{i \in V} x_i(0) x_i(t), \qquad t = \frac{1}{|V|}\,[\text{number of steps}]$$