Mixing Times

Swastika Dey

December 2025

Contents

What is Mixing Time of a Markov Chain?

A Markov chain is a process which moves among the elements of a set χ in the following manner: when at $x \in \chi$, the next position is chosen according to a fixed probability distribution $P(x, .)$ depending only on x. More precisely, a sequence of random variables $(X_0, X_1, ...)$ is a Markov chain with state-space χ and transition matrix P if for all $x, y \in \chi, \forall t \geq 1$ and all events $H_{t-1} = \cap_{s=0}^{(t-1)} \{X_s = x_s\}$ satisfying $P(H_{t-1}) \cap \{X_t = x\}) > 0$, we have $P(X_{t+1} = y | H_{t-1} \cap \{X_t = x\}) = P(X_{t+1} = y | X_t = x) = P(x, y)$.

The xth row of P is the distribution $P(x, .)$. We call P is stochastic if all entries in matrix P are non-negative and $\sum_{y \in \chi} P(x, y) = 1, \forall x \in \chi$.

A chain P is called irreducible if for any two states $x, y \in \chi$, there exists an integer $t \ni P^t(x, y) > 0$.

Let $\tau(x) = \{t \geq 1 : P^t(x, x) > 0\}$ be the set of times when it is possible for the chain to return to starting position x. The period of state x is defined as $\gcd(\tau(x))$. A chain will be called aperiodic if all states have period 1.

For a Markov chain on χ with transition matrix P, a probability distribution π satisfying $\pi P = \pi$ is called stationary distribution.

For a finite irreducible aperiodic Markov chain, the t-step transition probability of Markov chain converges to stationary distribution after sufficiently large time t and coincides.

For $\epsilon > 0$, we define **mixing time** as $t_{mix}(\epsilon) = \inf\{t \geq 0 | d(t) < \epsilon\}$, where we define $d(t) = \frac{1}{2} \sup_{x \in \chi} \sum_{y \in \chi} |P^t(x, y) - \pi(y)|$ (the total variation distance between t-step transition probability and stationary distribution).

It is not always possible to observe the exact mixing time $t_{mix}(\epsilon)$ for Markov chains, so we give upper and lower bounds to the mixing time.

Cutoff Phenomenon:

Now this $d(t)$ is decreasing with respect to t and $0 \leq d(t) \leq 1$.

For all t,

$$\begin{aligned}
d(t+1) &= \frac{1}{2}\sup_{x\in\chi}\sum_{y\in\chi}|P^{t+1}(x,y)-\pi(y)| \\
&= \frac{1}{2}\sup_{x\in\chi}\sum_{y\in\chi}|\sum_{z\in\chi}P(z,y)(P^{t}(x,z)-\pi(z))| \\
&\leq \frac{1}{2}\sup_{x\in\chi}\sum_{y\in\chi}P(z,y)\sum_{z\in\chi}|(P^{t}(x,z)-\pi(z))| \\
&= d(t) \ \ (\text{As } \sum_{y\in\chi}P(z,y)=1, \text{ irrespective of z}).
\end{aligned}$$

Equivalently,, $\forall \epsilon > \epsilon'$ we have $t_{mix}(\epsilon) > t_{mix}(\epsilon')$.

Define $t_{mix} := t_{mix}(\frac{1}{4})$, then observe $t_{mix}(\epsilon) \leq t_{mix}\lceil \log_2(\frac{1}{\epsilon})\rceil \ \forall \epsilon \in (0,1)$. Hence $\forall \epsilon \in (0,\frac{1}{4}), t_{mix} \leq t_{mix}(\epsilon) \leq t_{mix}\lceil \log_2(\frac{1}{\epsilon})\rceil$, which implies $t_{mix} \asymp t_{mix}(\epsilon)$, $\forall \epsilon \in (0,\frac{1}{4})$.

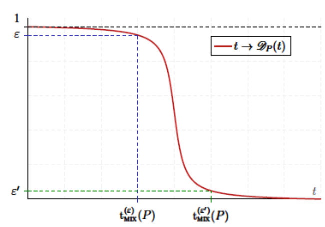

Figure 1: Graph of d(t) vs t

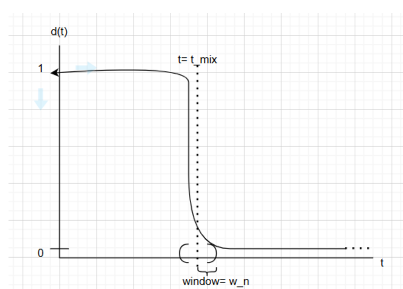

Figure 2: Cutoff phenomenon

In some cases, this $d(t)$ abruptly falls from 1 to 0 at time t_{mix}, which means $d^{(n)}(t) \overset{n\to\infty}{\rightrightarrows} \begin{cases} 1 & \text{if } t < t_{mix} \\ 0 & \text{if } t > t_{mix} \end{cases}$. This is called **cutoff-phenomenon**. Equivalently, we can say, $\forall \epsilon \in (0,1)$, $t^{(n)}_{mix}(\epsilon) \sim t^{(n)}_{mix}$ i.e; $\lim_{n \longrightarrow \infty} \frac{t^{(n)}_{mix}(\epsilon)}{t^{(n)}_{mix}} = 1$.

We say a sequence of Markov chains have a cutoff at $t^{(n)}_{mix}$ with a window of size $\mathcal{O}(w_n)$ if $w_n = o(t^{(n)}_{mix})$ and equivalently, $\lim_{\alpha \longrightarrow -\infty} \liminf_{n \longrightarrow \infty} d_n(t^{(n)}_{mix} + \alpha w_n) = 1$ and $\lim_{\alpha \longrightarrow \infty} \limsup_{n \longrightarrow \infty} d_n(t^{(n)}_{mix} + \alpha w_n) = 0$. Figure-2 describes the cutoff phenomenon.

Below we shall discuss some examples of Markov chains and different upper and lower bounds for their mixing time.

1 Lazy Random Walk on Hypercube:

Consider lazy random walk on hypercube $\{0,1\}^n$. Here the vertices are considered as the states. So, state-space χ is all possible n-length binary representations and $|\chi| = 2^n < \infty$.

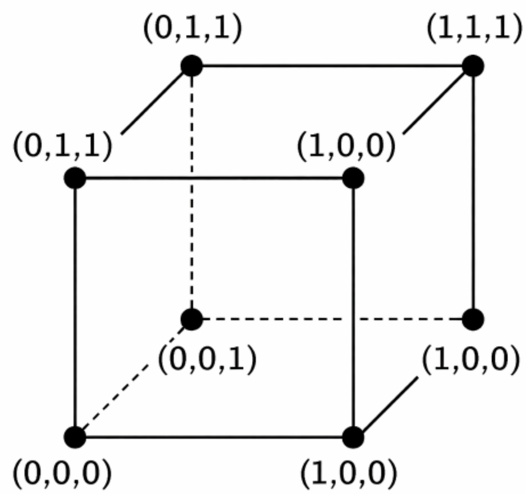

Figure 3: Hypercube $\{0,1\}^3$

The Markov chain runs as follows: Choose ith co-ordinate of present state uniformly with probability $\frac{1}{n}$ and then replace the bit present there with uniformly chosen bit among 0, 1 each with probability $\frac{1}{2}$. Thus the transition matrix is:

$$P(x,y) = \begin{cases} \frac{1}{2n} \times n \quad \text{if } y = x \text{ (Happens when same bit is chosen for chosen co-ordinate)} \\ \frac{1}{2n} \quad \text{if } x, y \text{ differ by exactly one co-ordinate} \\ 0 \quad \text{otherwise} \end{cases}$$

Notice that, to reach any vertex y from any vertex x, we always have a positive probability making the Markov chain irreducible. The Markov chain is also aperiodic as $P(x,x) = \frac{1}{2} > 0$. So, it converges to it's stationary distribution after sufficiently large time t and has mixing time.

Notice $P(x, y) = P(y, x)$, so P is symmetric and column sum of P is 1, hence $\pi = \frac{1}{2^n}(1, 1, \ldots, 1)$ satisfies detailed balance equation $\pi(x)P(x, y) = \pi(y)P(y, x)$ and so $\pi = \frac{1}{2^n}(1, 1, \ldots, 1)$ is a stationary distribution. As the chain in irreducible and finite it is the unique stationary distribution.

1.1 Upper Bound using Spectral Gap

Now here transition matrix P is symmetric, so eigenvalues of P are real and hence we are able to order the eigenvalues. For any eigenfunction f with eigenvalue λ, we have

$$\begin{aligned}
||Pf||_\infty &= \max_{x\in\chi} |\sum_{y\in\chi} P(x, y)f(y)| \\
&\leq \max_{x\in\chi} \sum_{y\in\chi} P(x, y)|f(y)| \\
&\leq \max_{y\in\chi} |f(y)| \max_{x\in\chi} \sum_{y\in\chi} P(x, y) \\
&= ||f||_\infty
\end{aligned}$$

$$\Rightarrow |\lambda|.||f||_\infty \leq ||f||_\infty \Rightarrow |\lambda| \leq 1.$$

Now as row sum of transition matrix is 1, $P\mathbf{1} = 1\mathbf{1}$ which implies that 1 is an eigenvalue. By Perron-Frobenius them maximum eigenvalue of P is unique as all entries of P is non-negative. So, $1 > \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_{|\chi|}$. Also the eigenvector corresponding to eigenvalue 1 is unique upto scaling.

P being symmetric we can write by spectral decomposition that $P = U\Lambda U^T$, where $U = [\mathbf{v}_1 : \ldots : \mathbf{v}_{|\chi|}]$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_{|\chi|})$ where $\mathbf{v}_i$ is the eigenvector corresponding to eigenvalue λ_i for all $i = 1, \ldots, |\chi|$ and hence $\mathbf{v}_i$'s are orthonormal to each other implying $U^T U = I$.

So,

$$
\begin{aligned}
&P^t = U\Lambda^t U^T = \lambda_1^t \mathbf{v}_1 \mathbf{v}_1^T + \cdots + \lambda_{|\chi|}^t \mathbf{v}_{|\chi|} \mathbf{v}_{|\chi|}^T \\
&\Rightarrow P^t(x,y) = \lambda_1^t v_1(x) v_1(y) + \cdots + \lambda_{|\chi|}^t v_{|\chi|}(x) v_{|\chi|}(y) \\
&\Rightarrow P^t(x,y) = \frac{1}{|\chi|} + \lambda_2^t v_2(x) v_2(y) + \cdots + \lambda_{|\chi|}^t v_{|\chi|}(x) v_{|\chi|}(y) \\
&\Rightarrow P^t(x,y) - \frac{1}{|\chi|} = \lambda_2^t v_2(x) v_2(y) + \cdots + \lambda_{|\chi|}^t v_{|\chi|}(x) v_{|\chi|}(y) \\
&\Rightarrow |P^t(x,y) - \frac{1}{|\chi|}| = |\lambda_2^t v_2(x) v_2(y) + \cdots + \lambda_{|\chi|}^t v_{|\chi|}(x) v_{|\chi|}(y)| \\
&\Rightarrow |P^t(x,y) - \frac{1}{|\chi|}| \leq \sum_{i=2}^{|\chi|} |\lambda_i|^t \leq |\lambda_2|^t |\chi| \\
&\Rightarrow d(t) = \frac{1}{2} \sup_{x \in \chi} \sum_{y \in \chi} |P^t(x,y) - \frac{1}{|\chi|}| \leq |\lambda_2|^t |\chi|^2.
\end{aligned}
$$

Using Product chain we get second largest eigenvalue as $(1 - \frac{1}{n})$. Thus

$$
\begin{aligned}
t_{mix}(\epsilon) &= \inf\{t \geq 0 | |\lambda_2|^t |\chi|^2 < \epsilon\} \\
&= \inf\{t \geq 0 | (1 - \frac{1}{n})^t 2^{2n} < \epsilon \\
&= \mathcal{O}(n^2).
\end{aligned}
$$

This is not a good upper bound, there exists other tools using which we can give it tighter bound.

1.2 Upper Bound using Coupling

A coupling of 2 probability distributions μ and ν is a pair of random variables (X, Y) defined on same probability space such that the marginal distribution of X is μ and the marginal distribution of Y is ν.

A coupling of Markov chains with transition matrix P is a process $(X_t, Y_t)_{t=0}^{\infty}$ with the property that both (X_t) and (Y_t) are Markov chains with transition matrix P. We modify 2 Markov chains such that after their simultaneous first visit to a state, they stay together. We denote such time as $\tau_{couple} := \min\{t \geq 0 | X_s = Y_s \forall s \geq t\}$. Now it can be shown that $d(t) \leq \max_{x,y \in \chi} P_{x,y}(\tau_{couple} > t)$.

For lazy random walk on hypercube $\{0,1\}^n$, we couple the 2 walks as follows: First we pick one among the n co-ordinates uniformly at random. Suppose ith co-ordinate is picked, then in both walks we replace the ith co-ordinate bit with the same random fair bit. So, this time onwards both walks agree in the ith co-ordinate. If τ be first time when all co-ordinates have been selected at least once, then the 2 walkers agree with each other after time τ onwards. Observe that solving this is same as solving coupon-collector problem.

Now,

$$\begin{aligned} P(\tau > t) &= P(\cup_{i=1}^n [i\text{-th co-ordinate doesn't appear before time } t] \\ &\leq \sum_{i=1}^n P(i\text{-th co-ordinate doesn't appear before time } t) \\ &= \sum_{i=1}^n (1 - \frac{1}{n})^t \\ &\leq ne^{-\frac{t}{n}}. \end{aligned}$$

$$\Rightarrow d(t) \leq ne^{-\frac{t}{n}}.$$

Hence we get $t_{mix}(\epsilon) = \inf\{t \geq 0 | ne^{-\frac{t}{n}} \leq \epsilon\} = \lceil n \log n + n \log(\frac{1}{\epsilon}) \rceil$.

1.3 Upper Bound using Strong Stationary Time

A move of this walk can be constructed using the following random mapping representation: An element (j, B) from $\{1, 2, \ldots, n\} \times \{0, 1\}$ is selected uniformly at random and coordinate j of the current state is updated with the bit B. In this construction, the chain is determined by the i.i.d. sequence $(Z_t)_{t\geq 0}$, where $Z_t = (j_t, B_t)$ is the coordinate-bit pair used to update at step t.

Define $\tau_{\text{refresh}} := \min\{t \geq 0 : \{j_1, \ldots, j_t\} = \{1, 2, \ldots, n\}\}$, to be the first time when all coordinates have been selected at least once for updating. Now at time τ_{refresh} all of the coordinates have been replaced with independent fair bits, so the distribution of the chain at this time is uniform on $\{0,1\}^n$. This τ_{refresh} is a stopping time with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$, since $\{\tau_{\text{refresh}} \leq t\} = \Big\{\{j_1, \ldots, j_t\} = \{1, 2, \ldots, n\}\Big\} \in \mathcal{F}_t$.

Given a starting state x, a state y is called a **halting state** for a stopping time τ if $X_t = y \quad \Rightarrow \quad \tau \leq t$. For this lazy random walk on $\{0,1\}^n$, starting from $(0, \ldots, 0)$, the state $(1, \ldots, 1)$ is a halting state for the stopping time τ_{refresh}.

We call a stopping time τ to be **strong stationary time**, if for all times t and all states y, $P_x\{\tau = t,\ X_\tau = y\} = P_x\{\tau = t\}\,\pi(y)$ which means X_τ is independent of τ and follows stationary distribution π. Hence for strong stationary time τ, we have $P_x\{\tau \le t,\ X_\tau = y\} = P_x\{\tau \le t\}\,\pi(y)$.

We define $s_x(t) := \max_{y\in\chi}(1 - \frac{P^t(x,y)}{\pi(y)})$. Now fix a state x. Then,

$$\begin{aligned}
s_x(t) &= \max_{y\in\chi}(1 - \frac{P^t(x,y)}{\pi(y)}) \\
&= \max_{y\in\chi}(1 - \frac{P_x(X_t = y)}{\pi(y)}) \\
&\le \max_{y\in\chi}(1 - \frac{P_x(X_t = y, \tau \le t)}{\pi(y)}) \\
&= \max_{y\in\chi}(1 - \frac{P_x(\tau \le t)\pi(y)}{\pi(y)}) \\
&= P_x(\tau > t).
\end{aligned}$$

Observe that equality holds in above if $P_x(X_t = y) = P_x(X_t = y, \tau \le t)$, i.e; $P_x(X_t = y, \tau > t) = 0$ which means y is reached before time τ. So, this equality holds for halting state and the time τ is called **optimal strong stationary time**. Hence $\tau_{refresh}$ is a optimal strong stationary time and $s_x(t) = P_x(\tau_{refresh} > t)$. Now

$$\begin{aligned}
d(t) &= \frac{1}{2}\sup_{x\in\chi}\sum_{y\in\chi}|P^t(x,y) - \pi(y)| \\
&= \sup_{x\in\chi}\sum_{y\in\chi,\, P^t(x,y)<\pi(y)}(\pi(y) - P^t(x,y)) \\
&= \sup_{x\in\chi}\sum_{y\in\chi,\, P^t(x,y)<\pi(y)}\pi(y)(1 - \frac{P^t(x,y)}{\pi(y)}) \\
&\le \sup_{x\in\chi}\max_{y\in\chi}(1 - \frac{P^t(x,y)}{\pi(y)}) \\
&\le \sup_{x\in\chi} P_x(\tau_{refresh} > t).
\end{aligned}$$

Now observe that this $\tau_{refresh}$ is again equivalent to coupon-collector time and hence using 1.2 we get $d(t) \le ne^{-\frac{t}{n}}$ and thus $t_{mix}(\epsilon) = \inf\{t \ge 0 | ne^{-\frac{t}{n}} \le \epsilon\} = \lceil n\log n + n\log(\frac{1}{\epsilon})\rceil$.

1.4 Upper Bound using All Eigenvalues

Lazy random walk on $\{0,1\}^n$ is the n-product chain of lazy random walk on cycle $\{0,1\}$. Using product chain formula we get, the eigenvalues of lazy random walk on $\{0,1\}^n$ are $(1-\frac{k}{n})$ with algebraic multiplicity $\binom{n}{k}$ for all $k = 0,\dots,n$.

Now observe that

$$
\begin{aligned}
(\sum_{y\in\chi} |P^t(x,y) - \pi(y)|)^2 &\leq \sum_{y\in\chi} (\frac{P^t(x,y)}{\pi(y)} - 1)^2 \pi(y). \sum_{y\in\chi} \pi(y) \\
&= \sum_{y\in\chi} (\frac{P^t(x,y)}{\pi(y)} - 1)^2 \pi(y) \\
&= \sum_{y\in\chi} \frac{(P^t(x,y) - \pi(y))^2}{\pi(y)} \\
&= \sum_{y\in\chi} \frac{(\sum_{j=2}^{|\chi|} \lambda_j^t v_j(x) v_j(y))^2}{\pi(y)} \\
&= \sum_{y\in\chi} \frac{\sum_{j=2}^{|\chi|} (\lambda_j^t v_j(x) v_j(y))^2}{\pi(y)} \quad \text{(As } v_j\text{'s are orthonormal, other terms} = 0) \\
&= \sum_{j=2}^{|\chi|} \lambda_j^{2t} v_j^2(x) \sum_{y\in\chi} \frac{v_j^2(y)}{\pi(y)} \\
&= |\chi| \sum_{j=2}^{|\chi|} \lambda_j^{2t} v_j^2(x) \\
&= \sum_{j=2}^{|\chi|} \frac{\lambda_j^{2t} v_j^2(x)}{\pi(x)}.
\end{aligned}
$$

So, we get

$$\sum_{x\in\chi}(\sum_{y\in\chi}|P^t(x,y)-\pi(y)|)^2\pi(x) \leq \sum_{x\in\chi}\sum_{j=2}^{|\chi|}\frac{\lambda_j^{2t}v_j^2(x)}{\pi(x)}\pi(x)$$
$$= \sum_{j=2}^{|\chi|}\lambda_j^{2t}\sum_{x\in\chi}v_j^2(x)$$
$$= \sum_{j=2}^{|\chi|}\lambda_j^{2t} \quad (\text{As } \sum_{x\in\chi}v_j^2(x)=1).$$
$$\Rightarrow \sup_{x\in\chi}(\sum_{y\in\chi}|P^t(x,y)-\pi(y)|)^2\sum_{x\in\chi}\pi(x) \leq \sum_{j=2}^{|\chi|}\lambda_j^{2t}$$
$$\Rightarrow \sup_{x\in\chi}(\sum_{y\in\chi}|P^t(x,y)-\pi(y)|)^2 \leq \sum_{j=2}^{|\chi|}\lambda_j^{2t}$$
$$\Rightarrow d(t) \leq \frac{1}{2}.\sqrt{\sum_{j=2}^{|\chi|}\lambda_j^{2t}}.$$

Now putting all eigenvalues we get that
$d(t) \leq \frac{1}{2}.\sqrt{\sum_{j=2}^{n}\binom{n}{k}(1-\frac{k}{n})^{2t}} \leq \frac{1}{2}.\sqrt{\sum_{j=1}^{n}\binom{n}{k}e^{-\frac{2tk}{n}}} = \sqrt{(1+e^{-\frac{2t}{n}})^n-1}.$

Thus for $t=\frac{n\log n}{2}+cn$, with $c>1$ we get $d(\frac{n\log n}{2}+cn) \leq \sqrt{(1+\frac{e^{-2c}}{n})^n-1} < e^{-2c}$.

Thus $t_{mix}(\epsilon) = \lceil\frac{n\log n}{2}+n\log(\frac{1}{\epsilon})\rceil$.

1.5 Lower Bound Using Bottleneck Ratio Got through Cheggar's Inequality

Lazy random walk on $\{0,1\}^n$ is the n-product chain of lazy random walk on cycle $\{0,1\}$. Using product chain formula we get, the eigenvalues of lazy random walk on $\{0,1\}^n$ are $(1-\frac{k}{n})$ with algebraic multiplicity $\binom{n}{k}$ for all $k=0,\ldots,n$.

Now by Cheggar's inequality we know $\frac{(\Phi^*)^2}{2} \leq \gamma \leq 2\Phi^*$ where Φ^* is the bottle-neck ratio, i.e; $\Phi^* = \min_{S|\pi(S)\leq\frac{1}{2}}\frac{\sum_{x\in S,y\in S^c}\pi(x)P(x,y)}{\sum_{x\in S}\pi(x)}$ and γ is spectral gap, i.e; $\gamma=(1-\lambda_2)$.

Here $\gamma=\frac{1}{n}$ and $\Phi^* = \min_{S|\pi(S)\leq\frac{1}{2}}\frac{\frac{1}{2^n}\sum_{x\in S,y\in S^c}P(x,y)}{\sum_{x\in S}\frac{1}{2^n}}$.

Take $S = \{x | x^1 = 0\}$, so $|S| = 2^{n-1}$ and hence $\Phi^* = 2^{-n+1} \sum_{x \in S, y \in S^c} P(x, y) = 2^{-n+1} 2^{n-1} \frac{1}{2n}$.

Now, as $\gamma \leq 2\Phi^*$ and $\Phi^* = \frac{1}{2n} = \frac{\gamma}{2}$, hence here the bound is sharp.

Now it can be shown that $t_{mix} \geq \frac{1}{4\Phi^*} = \frac{n}{2}$.

But this is a lighter lower bound, we will now give a tighter lower bound

1.6 Lower Bound Using Distinguished Statistics

Now we shall introduce a tighter lower bound on mixing time using a distinguished statistic f on χ such that the equilibrium distance can be bounded from below.

A result tells that for a distinguished statistic $f : \chi \longrightarrow \mathbb{R}$ for a Markov chain with transition matrix P and stationary distribution π and starting state x, define $\sigma_*^2 = \max\{Var_x(f(X_t)), Var_\pi(f)\}$ (where $Var_\pi(f)$ denotes variance of f under stationary distribution π), if $|\mathbb{E}_x(f(X_t)) - \mathbb{E}_\pi(f)| > r\sigma_*$ (for r ¿ 0), then $||P^t(x, .) - \pi||_{TV} \geq (1 - \frac{8}{r^2})$.

For lazy random walk on $\{0, 1\}^n$, we take the Hamming weight $W(X_t) = \sum_{i=1}^n X_t^i$ of the $\{0, 1\}^n$ as $f(X_t)$.

We know stationary distribution is $\pi(x) = \frac{1}{2^n}$ which is uniform distribution over all possible states.

Now probability that $W(X_t) = x$ is equivalent to the proportion of vertices with $W(X_t) = x$ which equals $\frac{\binom{n}{x}}{2^n}$, so $W(X_t)$ follows $Bin(n, \frac{1}{2})$ distribution. So $\mathbb{E}_\pi(f) = \frac{n}{2}$ and $Var_\pi(f) = \frac{n}{4}$.

Now we shall compute $\mathbb{E}_x(f(X_t))$ and $Var_x(f(X_t))$ conditioning on $R_t =$ no. of non-updated co-ordinate in X_t starting from $\mathbf{x} = (1, \ldots, 1)$. Now observe that $W(X_t) \geq R_t$, $\forall t$.

Now, starting from $(1, \dots, 1)$,

$$\begin{aligned}
&P(W(X_t) = x | R_t = r) \\
&= \frac{\text{no. of ways to choose a previously updated co-ordinate with 1}}{\text{no. of ways to choose any previously updated co-ordinate}} \\
&= \frac{\binom{n-r}{x-r}}{2^{n-r}} \\
&= \text{distribution of } r + B \text{ where } B \sim Bin(n-r, \frac{1}{2}).
\end{aligned}$$

So, we get $\mathbb{E}_1(W(X_t)|R_t) = R_t + \frac{(n-R_t)}{2} = \frac{(n+R_t)}{2}$ and $Var_1(W(X_t)|R_t) = 0 + \frac{(n-R_t)}{4} = \frac{(n-R_t)}{4}$.

Now $\mathbb{E}_1(W(X_t)) = \frac{(n+\mathbb{E}(R_t))}{2}$ where $R_t = \sum_{i=1}^n \mathbb{1}(\text{co-ordinate i not updated till time t}) =: I_i^t$, so we get

$$\begin{aligned}
\mathbb{E}(R_t) &= \sum_{i=1}^n \mathbb{E}(\mathbb{1}(\text{co-ordinate i not updated till time t})) \\
&= \sum_{i=1}^n P(\text{co-ordinate i not updated till time t}) \\
&= \sum_{i=1}^n (1 - P(\text{co-ordinate i chosen for update before time t})) \\
&= \sum_{i=1}^n \prod_{j=1}^t (1 - P(\text{co-ordinate chosen for update in jth step})) \\
&= \sum_{i=1}^n (1 - \frac{1}{n})^t \\
&= n(1 - \frac{1}{n})^t.
\end{aligned}$$

Now

$$\begin{aligned}
Var_1(W(X_t)) &= \mathbb{E}(Var_1((W(X_t)|R_t)) + Var_1(\mathbb{E}(W(X_t)|R_t)) \\
&= \mathbb{E}(\frac{(n-R_t)}{4}) + Var_1(\frac{(n+R_t)}{2}) \\
&= \frac{n}{4} - \frac{\mathbb{E}(R_t)}{4} + \frac{Var_1(R_t)}{4}.
\end{aligned}$$

So, now we are left to find $Var_1(R_t) = \sum_{i=1}^n Var_1(I_i^t) + \sum_{i=1}^n \sum_{j=1}^n Cov_1(I_i^t, I_j^t)$.

Now observe

$$\begin{aligned}
Cov_1(I_i^t, I_j^t) &= \mathbb{E}_1(I_i^t I_j^t) - \mathbb{E}_1(I_i^t)\mathbb{E}_1(I_j^t) \\
&= P(\text{co-ordinate i and j not updated till time t}) - (1 - \frac{1}{n})^{2t} \\
&= (1 - \frac{2}{n})^t - (1 - \frac{1}{n})^{2t} \\
&< 0
\end{aligned}$$

Hence

$$\begin{aligned}
Var_1(R_t) &< \sum_{i=1}^n Var_1(I_i^t) \\
&= \sum_{i=1}^n \mathbb{E}_1(I_i^t) - (\mathbb{E}_1(I_i^t))^2 \quad (\text{As } (I_i^t)^2 = I_i^t) \\
&\leq \sum_{i=1}^n \mathbb{E}_1(I_i^t) \\
&= \mathbb{E}(R_t)
\end{aligned}$$

So, $Var_1(W(X_t)) < \frac{n}{4} - \frac{\mathbb{E}(R_t)}{4} + \frac{\mathbb{E}(R_t)}{4} = \frac{n}{4}$ which implies $\sigma_* = \frac{\sqrt{n}}{2}$.

Now $|\mathbb{E}_\pi(W) - \mathbb{E}_1(W(X_t))| = \frac{n}{2}(1 - \frac{1}{n})^t = \sigma_*\sqrt{n}(1 - \frac{1}{n})^t > \sigma_*\sqrt{n}e^{-\frac{t}{n-1}}$.

Taking $t_n = \frac{(n-1)\log n}{2} - (\alpha - 1)n > \frac{n \log n}{2} - \alpha n$ (for $\alpha > 1$), we get $|\mathbb{E}_\pi(W) - \mathbb{E}_1(W(X_t))| > \sigma_* e^{(\alpha-1)}$ which by the theorem gives $||P^{t_n}(\mathbf{1}, .) - \pi||_{TV} \geq (1 - 8e^{2-2\alpha})$ and hence we get $d(\frac{n \log n}{2} - \alpha n) \geq (1 - 8e^{2-2\alpha})$.

So, the mixing time $t_{mix}(1 - \epsilon) = \lfloor \frac{n \log n}{2} - \log(\frac{1}{\epsilon})n \rfloor$.

2 Proper q-colorings:

The proper colorings of a graph $G = (V, E)$ are elements of $x \in \chi = \{1, 2, \ldots, q\}^V$ such that $x(v) \neq x(w)$, $\forall \{v, w\} \in E$, i.e; an assignment of colors to the vertices V, subject to the constraint that neighboring vertices do not receive the same color. A vertex v is chosen uniformly at random, a color is selected uniformly at random among all q colors, and the vertex v is recolored with the chosen color.

2.1 Upper Bound using Grand-Coupling to Metropolis Chain

Applying the **Metropolis rule** to this chain, where the stationary distribution π is the probability measure that is uniform over the space of all proper q-colorings. When at a proper coloring, if the color k is proposed to update a vertex, then the Metropolis rule accepts the proposed recoloring with probability 1 if it yields a proper coloring and rejects it otherwise. So, for the Metropolis chain, if vertex v is selected, the chance of remaining in the current coloring is $\frac{(1+q-a)}{q}$, where a is the no. of allowable colors to vertex v.

Using **Grand Couplings** we shall give an upper bound of $\frac{1}{(1-\frac{3\Delta}{q})} n \lceil \log n + \log(\frac{1}{\epsilon}) \rceil$ to it's mixing time for $q > 3\Delta$, where Δ is the maximum degree.

We want to construct a collection of random variables $\{X_t^x : x \in \mathcal{X},\ t = 0, 1, 2, \ldots\}$ such that for each $x \in \mathcal{X}$, the sequence $(X_t^x)_{t \geq 0}$ is a Markov chain with transition matrix P started from x which is called a **grand coupling**.

Let $f : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ be a function and let Z be a $\mathcal{Z}$-valued random variable such that $P(x, y) = P\Big(f(x, Z) = y\Big)$. Let $Z_1, Z_2, \ldots$ be an i.i.d. sequence, each with the same distribution as Z, and define $X_0^x = x$, $X_t^x = f\Big(X_{t-1}^x, Z_t\Big), t \geq 1$. Since each of the processes $(X_t^x)_{t \geq 0}$ is a Markov chain started from x with transition matrix P, this construction yields a grand coupling. We emphasize that the chains $(X_t^x)_{t \geq 0}$ all live on the same probability space, each being determined by the same sequence of random variables $(Z_t)_{t \geq 1}$.

In the grand coupling, at each move a single vertex-color pair (v, k) is generated, uniformly at random from $V \times \{1, \ldots, q\}$ and independently of the past. For each $x \in \widetilde{\mathcal{X}}$, the coloring X_t^x is updated by attempting to recolor vertex v with color k, accepting the update if and only if the proposed new color is different from the colors at vertices neighboring v. (If a recoloring is not accepted, the chain X_t^x remains in its current state.) The essential point is that the same vertex and color are used for all the chains $\left(X_t^x\right)$.

For two colorings $x, y \in \widetilde{\mathcal{X}}$, define the metric $\rho(x,y) := \sum_{v \in V} \mathbf{1}_{\{x(v) \neq y(v)\}}$, which is the number of vertices at which x and y disagree.

Suppose $\rho(x,y) = 1$, so that x and y agree everywhere except at a single vertex v_0. Let $\mathcal{N}$ denote the set of colors appearing among the neighbors of v_0 in x, which is the same as the set of colors appearing among the neighbors of v_0 in y. As v represents a random sample from V, and k a random sample from $\{1, 2, \ldots, q\}$, independent of v, we consider the distance after updating x and y in one step of the grand coupling, $\rho\left(X_1^x, X_1^y\right)$. Now in one move the distance can at most increase by 1 or at most decrease by 1 or remain same.

Case-1: This distance is zero, i.e; decrease by 1 in one step, if and only if the vertex v_0 is selected for updating and the proposed color is not in N. This occurs with probability $P\left\{\rho(X_1^x, X_1^y) = 0\right\} = \left(\frac{1}{n}\right)\left(\frac{q - |\mathcal{N}|}{q}\right) \geq \frac{q - \Delta}{nq}$.

Case-2: Suppose now a vertex w which is a neighbor of v_0 is selected for updating. If the proposed color is $x(v_0)$ or $y(v_0)$, the number of disagreements may possibly increase by at most 1. If neither the color $x(v_0)$ or $y(v_0)$ is proposed, the new color will be accepted in x if and only if it accepted in y. Thus, the only way a new disagreement can possibly be introduced is if a neighbor of v_0 is selected for updating (which has probability $\frac{\Delta}{n}$), and either $x(v_0)$ or $y(v_0)$ is proposed (which has probability $\leq \frac{2}{q}$). Thus the distance increase by 1 in one step with probability $\mathbb{P}\left\{\rho(X_1^x, X_1^y) = 2\right\} \leq \left(\frac{\Delta}{n}\right)\left(\frac{2}{q}\right)$.

Thus $\mathbb{E}(\rho(X_1^x, X_1^y) - 1) \leq \frac{2\Delta}{qn} - \frac{q-\Delta}{qn}$ which implies $\mathbb{E}(\rho(X_1^x, X_1^y)) \leq 1 - \frac{q - 3\Delta}{qn}$.

Now, for $q > 3\Delta$, we have $1 - (3\Delta/q) > 0$, and hence $\mathbb{E}\left[\rho(X_1^x, X_1^y)\right] \leq 1 - \frac{(1 - \frac{3\Delta}{q})}{n} < 1$. Hence

$$\begin{aligned}
\mathbb{E}\left[\rho(X_1^x, X_1^y)\right] &\leq \sum_{k=1}^{r} \mathbb{E}\left[\rho(X_1^{x_k}, X_1^{x_{k-1}})\right] \\
&\leq r\left(1 - \frac{(1 - \frac{3\Delta}{q})}{n}\right) \\
&= \rho(x,y)\left(1 - \frac{}{n}\right).
\end{aligned}$$

Now observe that Conditional on the event $X_{t-1}^x = x_{t-1}$ and $X_{t-1}^y = y_{t-1}$, $\left(X_t^x,\, X_t^y\right)$

has the same distribution as $\left(X_1^{x_{t-1}},\ X_1^{y_{t-1}}\right)$. Hence $\mathbb{E}\Big[\rho(X_t^x, X_t^y) \mid X_{t-1}^x = x_{t-1}, X_{t-1}^y = y_{t-1}\Big] \leq \rho(x_{t-1}, y_{t-1})\left(1 - \frac{(1-\frac{3\Delta}{q})}{n}\right)$. Now taking expectations and iterating over time we get $\mathbb{E}\Big[\rho(X_t^x, X_t^y)\Big] \leq \rho(x,y)\left(1 - \frac{(1-\frac{3\Delta}{q})}{n}\right)^t$. Now by Markov's Inequality, since $\rho(x,y) \geq 1$ whenever $x \neq y$, we get $\mathbb{P}\{X_t^x \neq X_t^y\} = \mathbb{P}\{\rho(X_t^x, X_t^y) \geq 1\} \leq \rho(x,y)\left(1 - \frac{(1-\frac{3\Delta}{q})}{n}\right)^t \leq n\, e^{-t\frac{(1-\frac{3\Delta}{q})}{n}}$ which implies $d(t) \leq ne^{-t\,(c_{\mathrm{met}}(\Delta,q)/n)}$.

So, if $t \geq \frac{1}{(1-\frac{3\Delta}{q})} n\Big[\log n + \log(1/\epsilon)\Big]$, then $d(t) \leq \epsilon$, which gives us $t_{mix}(\epsilon) \leq \frac{1}{(1-\frac{3\Delta}{q})} n\lceil \log n + \log(\frac{1}{\epsilon})\rceil$ for $q > 3\Delta$.

2.2 Upper Bound using Path-Coupling to Glauber Dynamics

For a given configuration x and a vertex v, call a color j **allowable** at v if j is different from all colors assigned to neighbors of v, i.e; a color j is allowable at v if $j \notin \{x(w) : w \sim v\}$. Given a proper q-coloring x, we can generate a new coloring by

- selecting a vertex $v \in V$ uniformly at random,
- selecting a color j uniformly at random from the allowable colors at v,
- re-coloring vertex v with color j.

We claim that the resulting chain has the uniform stationary distribution. To see why, note that transitions are permitted only between colorings that differ at a single vertex. If x and y agree everywhere except at vertex v, then the probability of moving from x to y is $P(x,y) = \frac{1}{|V|}\frac{1}{|A_v(x)|}$, where $A_v(x)$ denotes the set of allowable colors at v in the configuration x. Since $A_v(x) = A_v(y)$, this probability is equal to the probability of moving from y to x. Thus $P(x,y) = P(y,x)$, and the detailed balance equations are satisfied by the uniform distribution. Consequently, the uniform distribution is stationary for this chain. This chain is called the **Glauber dynamics** for proper q-colorings. Note that when a vertex v is updated in a coloring x, the new coloring is chosen from π conditioned on the set of colorings that agree with x at all vertices other than v.

Using **Path Coupling** we shall give an upper bound of $\lceil \frac{(q-\Delta)}{(q-2\Delta)} n(\log n + \log(\frac{1}{\epsilon}))\rceil$ to it's mixing time for $q > 2\Delta$, where Δ is the maximum degree.

Suppose that the state space $\mathcal{X}$ of a Markov chain (X_t) is the vertex set of a connected graph $G = (\mathcal{X}, E_0)$, and ℓ is a length function defined on E_0,i.e; ℓ assigns a length $\ell(x, y)$ to each edge $\{x, y\} \in E_0$. We assume that $\ell(x, y) \geq 1$ for all edges $\{x, y\}$. If $x_0, x_1, \ldots, x_r$ is a path in G, we define its length to be $\sum_{i=1}^{r} \ell(x_{i-1}, x_i)$. then The **path metric** on $\mathcal{X}$ is defined by $\rho(x, y) := \min\{\text{length of } \gamma : \gamma \text{ is a path from } x \text{ to } y\}$. Since we have assumed that $\ell(x, y) \geq 1$, it follows that $\rho(x, y) \geq \mathbf{1}_{\{x \neq y\}}$. Hence, for any pair of random variables (X, Y), we get $\mathbb{P}\{X \neq Y\} = \mathbb{E}\Big[\mathbf{1}_{\{X \neq Y\}}\Big] \leq \mathbb{E}\, \rho(X, Y)$. Minimizing over all couplings (X, Y) of μ and ν shows that, $||\mu - \nu||_{TV} \leq \inf \Big\{\mathbb{E}\Big[\rho(X, Y)\Big] : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\Big\}$.

Suppose the state space $\mathcal{X}$ of a Markov chain is the vertex set of a graph with length function ℓ defined on edges. Let ρ be the corresponding path metric. Suppose that for each edge $\{x, y\}$ there exists a coupling (X_1, Y_1) of the distributions $P(x, \cdot)$ and $P(y, \cdot)$ such that $\mathbb{E}_{x,y}[\rho(X_1, Y_1)] \leq \rho(x, y)\, e^{-\alpha}$. Then $d(t) \leq e^{-\alpha t} diam(\chi)$ where $diam(\chi) = \max_{x,y \in \chi} \rho(x, y)$. Consequently, $t_{mix}(\epsilon) \leq \lceil \frac{\log(diam(\chi)) + \log(\frac{1}{\epsilon})}{\alpha} \rceil$.

Take the path metric $\rho(x, y) = \sum_{v \in V} \mathbb{1}(x(v) \neq y(v))$ which is the number of sites at which x and y differ. Two colorings are neighbors if and only if they differ at a single vertex. Note that this neighboring rule defines a graph different from the graph defined by the transitions of the chain, since the chain moves only among proper colorings. $A_v(x)$ be the set of allowable colors at v in configuration x. Let x and y be two configurations which agree everywhere except at vertex v.

Now we describe how to simultaneously evolve two chains, one started at x and the other started at y, such that each chain viewed alone is a Glauber chain:

- First, we pick a vertex w uniformly at random from the vertex set of the graph.
- We will update the color of w in both the chain started from x and the chain started from y:
  - (1) If v is not a neighbor of w, then we can update the two chains with the same color. Each chain is updated with the correct distribution because $A_w(x) = A_w(y)$.
  - (2) Suppose now one of the neighbors of w is v. Without loss of generality, we assume that $|A_w(x)| \leq |A_w(y)|$. Generate a random color U from $A_w(y)$ and use this to update y at w:
    - (i) If $U \neq x(v)$, then update the configuration x at w to U.
    - (ii) If $U = x(v)$ and $|A_w(x)| = |A_w(y)|$, set $x(w) = y(v)$.

(iii) If $U = x(v)$ and $|A_w(x)| < |A_w(y)|$, draw a random color from $A_w(x)$.

These updates x at w to a color chosen uniformly from $A_w(x)$.

The probability that the 2 configurations don't update to the same color is $\frac{1}{|A_w(y)|}$ which happens when the chosen color is a neighboring color of the chosen vertex of exactly one of the configurations x or y, not the other. This probability is bounded by $\frac{1}{(q-\Delta)}$. Given two states x and y with $\rho(x, y) = 1$, we have constructed a coupling (X_1, Y_1) of $P(x, .)$ and $P(y, .)$. The distance $\rho(X_1, Y_1)$ increases by at most 1 only in the case where a neighbor of v is updated and the updates are different in the 2 configurations, so increase happens with probability $(1 - \frac{1}{n})$. Also, the distance decreases by at most 1, when v is selected to be updated, so decrease happens with probability $\leq \frac{\deg(v)}{n(q-\Delta)}$. In all other cases, the distance stays same. Hence $\mathbb{E}_{x,y}(\rho(X_1, Y_1)) \leq 1 - \frac{1}{n} + \frac{\deg(v)}{n(q-\Delta)} \leq 1 - \frac{1}{n}(1 - \frac{\Delta}{q-\Delta})$.

For $q > 2\Delta$, we get $1 - \frac{1}{n}(1 - \frac{\Delta}{q-\Delta}) < 1$ and hence $\mathbb{E}_{x,y}(\rho(X_1, Y_1)) \leq e^{-\frac{1}{n}(1-\frac{\Delta}{q-\Delta})}$ which implies that $d(t) \leq ne^{-\frac{1}{n}(1-\frac{\Delta}{q-\Delta})t}$, where $diam(\chi) = n$.

Thus we get $t_{mix}(\epsilon) \leq \lceil \frac{(q-\Delta)}{(q-2\Delta)} n(\log n + \log(\frac{1}{\epsilon})) \rceil$ for $q > 2\Delta$.

3 Ising Model Fast Mixing in High Temperature:

A **spin system** is a probability distribution on $\mathcal{X} = \{-1, 1\}^V$, where V is the vertex set of a graph $G = (V, E)$. The value $\sigma(v)$ is called the **spin** at vertex v. The physical interpretation is that magnets, each having one of the two possible orientations represented by $+1$ and -1, are placed on the vertices of the graph; a configuration specifies the orientations of these magnets. The **nearest-neighbor Ising model** is the most widely studied spin system. In this system, the energy of a configuration σ is defined by $H(\sigma) = -\sum_{\substack{v,w \in V \\ v \sim w}} \sigma(v)\sigma(w)$ which is basically the sum of pairs of disagreeing spin neighbors minus the sum of pairs of agreeing spin neighbors. Clearly, the energy increases with the number of pairs of neighboring vertices whose spins disagree. The Gibbs distribution corresponding to the energy H is the probability distribution μ on $\mathcal{X}$ defined by $\mu(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)}$, where $Z(\beta) := \sum_{\sigma \in \mathcal{X}} e^{-\beta H(\sigma)}$.

The parameter $\beta \geq 0$ determines the influence of the energy function. In the physical interpretation, β is the **inverse temperature**. At infinite temperature ($\beta = 0$), the energy function H plays no role and μ becomes the uniform distribution on $\mathcal{X}$. In this case, there is no interaction between the spins at different vertices and the random variables $\{\sigma(v)\}_{v \in V}$ are independent. As $\beta > 0$ increases, the bias of μ toward low-energy configurations also increases.

The Glauber dynamics for the Gibbs distribution μ move from a starting configuration σ by picking a vertex w uniformly at random from V and then generating a new configuration according to μ conditioned on the set of configurations agreeing with σ on vertices different from w. The conditional μ-probability of spin $+1$ at a vertex w is $p(\sigma; w) := \frac{e^{\beta S(\sigma;w)}}{e^{\beta S(\sigma;w)} + e^{-\beta S(\sigma;w)}} = \frac{1+\tanh\left(\beta S(\sigma;w)\right)}{2}$, where $S(\sigma; w) := \sum_{u: u \sim w} \sigma(u)$. Note that $p(\sigma; w)$ depends only on the spins at vertices adjacent to w. Therefore, the transition matrix on $\mathcal{X}$ is given by $P(\sigma, \sigma') = \frac{1}{|V|} \sum_{w \in V} \frac{e^{\beta \sigma'(w) S(\sigma;w)}}{e^{\beta \sigma'(w) S(\sigma;w)} + e^{-\beta \sigma'(w) S(\sigma;w)}} \mathbf{1}\{\sigma(v) = \sigma'(v) \text{ for } v \neq w\}$. This chain has stationary distribution as the Gibbs distribution μ.

We will be particularly interested in how the mixing time varies with β. For small β, the chain will mix rapidly, while for large β, the chain will converge slowly. Understanding this phase transition between slow and fast mixing has been a topic of great interest and activity since the late 1980's; here we only scratch the surface.

3.1 Upper Bound Got through Path-Coupling

Using **path-coupling**, we shall to show that on any graph of bounded degree, for small values of β, the Glauber dynamics for the Ising model is mixing fast.

We shall show, for Glauber dynamics for the Ising model on a graph with n vertices and maximal degree Δ, we have $t_{rel} \leq \frac{n}{1-\Delta tanh(\beta)}$ where $\Delta tanh(\beta) < 1$. Also we will show $d(t) \leq n(1-\frac{1-\Delta tanh(\beta)}{n})^t$ which gives $t_{mix}(\epsilon) \leq \lceil \frac{n(\log n+\log(\frac{1}{\epsilon}))}{1-\Delta tanh(\beta)} \rceil$ for $\beta < \frac{1}{\Delta}$. If every vertex of the graph has even degree, we have $t_{rel} \leq \frac{n}{1-\frac{\Delta}{2}tanh(2\beta)}$ where $\frac{\Delta}{2}tanh(2\beta) < 1$, then $d(t) \leq n(1-\frac{1-\frac{\Delta}{2}tanh(2\beta)}{n})^t$ which gives $t_{mix}(\epsilon) \leq \lceil \frac{n(\log n+\log(\frac{1}{\epsilon}))}{1-\frac{\Delta}{2}tanh(2\beta)} \rceil$. Now we shall prove this.

Define the distance ρ on $\mathcal{X}$ by $\rho(\sigma,\tau) := \frac{1}{2}\sum_{u\in V}\Big|\sigma(u)-\tau(u)\Big|$, which is a path metric. Let σ and τ be two configurations with $\rho(\sigma,\tau)=1$. The spins of σ and τ agree everywhere except at a single vertex v. Assume that $\sigma(v)=-1$ and $\tau(v)=+1$. Define $\mathcal{N}(v) := \{u : u \sim v\}$ to be the set of vertices neighboring v. We now describe a coupling (X,Y) of one step of the chain started in configuration σ with one step of the chain started in configuration τ.

Pick a vertex w uniformly at random from V.

- If $w \notin N(v)$, then the neighbors of w agree in both configurations σ and τ. As the probability of updating the spin at w to $+1$, $p(\sigma;w) := \frac{e^{\beta S(\sigma;w)}}{e^{\beta S(\sigma;w)}+e^{-\beta S(\sigma;w)}}$ (with $S(\sigma;w) := \sum_{u:u\sim w}\sigma(u)$), depends only on the spins at the neighbors of w, it is the same for the chain started in σ as for the chain started in τ . Thus we can update both chains together.

- If $w \in N(v)$, the probabilities of updating to $+1$ at w are no longer the same for the two chains, so we cannot always update together. We do, however, use a single random variable as a common source of randomness so that the two chains agree as often as possible. Let U be a uniform random variable on $[0,1]$, and define $X(w) = \begin{cases} +1, & \text{if } U \leq p(\sigma;w) \\ -1, & \text{if } U > p(\sigma;w) \end{cases}$ and $Y(w) = \begin{cases} +1, & \text{if } U \leq p(\tau;w) \\ -1, & \text{if } U > p(\tau;w) \end{cases}$.

  For all vertices $u \neq w$, set $X(u)=\sigma(u), \qquad Y(u)=\tau(u)$. Note that since the function tanh is non-decreasing, and since $S(\sigma;w) \leq S(\tau;w)$ owing to $\sigma(v)=-1$ and $\tau(v)=+1$, we always have $p(\sigma;w) \leq p(\tau;w)$.

- If $w=v$, then the distance become 0 (decreasing by 1) by updating the spin of selected vertex with same spin for both configuration which is possible as for both chain the neighbors of that vertex can't be different because that vertex was the only difference between the 2 configurations.

Hence if $w = v$ then $\rho(X, Y) = 0$, if $w \notin N(v) \cup \{v\}$, then $\rho(X, Y) = 1$, if $w \in N(v)$ and $p(\sigma; w) < U \leq p(\tau; w)$ then $\rho(X, Y) = 2$.

Thus $\mathbb{E}_{\sigma,\tau}(\rho(X, Y) - 1) \leq \frac{1}{n} \sum_{w \in \mathcal{N}(v)} (p(\tau; w) - p(\sigma; w)) - \frac{1}{n}$, which gives $\mathbb{E}_{\sigma,\tau}(\rho(X, Y)) \leq 1 - \frac{1}{n} + \frac{1}{n} \sum_{w \in \mathcal{N}(v)} (p(\tau; w) - p(\sigma; w))$.

Define $s := S(w, \tau) - 1 = S(w, \sigma) + 1$.

Then $p(\tau; w) - p(\sigma; w) = \frac{1}{2}(\tanh(\beta(s + 1)) - \tanh(\beta(s - 1))) \leq \tanh(\beta)$ which gives us $\mathbb{E}_{\sigma,\tau}(\rho(X, Y)) \leq 1 - \frac{1}{n}(1 - |\mathcal{N}(v)| \tanh(\beta)) \leq 1 - \frac{1}{n}(1 - \Delta \tanh(\beta))$.

Now a result tells us that if there exists a constant $\theta < 1$ such that for each $x, y \in \chi$ there exists a coupling (X_1, Y_1) of $P(x, .)$ and $P(y, .)$ satisfying $\mathbb{E}(\rho(X_1, Y_1)) \leq \theta\rho(x, y)$, then spectral gap $\gamma \geq 1 - \theta$.

Here $\theta = 1 - \frac{1}{n}(1 - \Delta \tanh(\beta))$ which gives $\gamma \geq \frac{1}{n}(1 - \Delta \tanh(\beta))$ and hence $t_{rel} \leq \frac{n}{\Delta \tanh(\beta)}$.

Now here for path-coupling $diam(\chi) = n$ and $e^{-\alpha} = (1 - \frac{1 - \Delta \tanh(\beta)}{n})$, which gives $d(t) \leq n(1 - \frac{1 - \Delta \tanh(\beta)}{n})^t$ for $\Delta \tanh(\beta) < 1$ and consequently $t_{mix}(\epsilon) \leq \lceil \frac{n(\log n + \log(\frac{1}{\epsilon}))}{1 - \Delta \tanh(\beta)} \rceil$.

As $\tanh(x) \leq x$ implies $\Delta \tanh(\beta) < 1$ for $\beta < \frac{1}{\Delta}$, hence the above holds for $\beta < \frac{1}{\Delta}$.

Now for a graph with all even degree vertices s takes only odd values and hence $p(\tau; w) - p(\sigma; w) = \frac{1}{2}(\tanh(\beta(s + 1)) - \tanh(\beta(s - 1))) \leq \frac{1}{2} \tanh(2\beta)$ and similarly as previous part $\mathbb{E}_{\sigma,\tau}(\rho(X, Y)) \leq 1 - \frac{1}{n}(1 - \frac{\Delta}{2} \tanh(2\beta))$ which by path-coupling gives $d(t) \leq n(1 - \frac{1 - \frac{\Delta}{2} tanh(2\beta)}{n})^t$ for $\frac{\Delta}{2} \tanh(2\beta) < 1$ and consequently $t_{mix}(\epsilon) \leq \lceil \frac{n(\log n + \log(\frac{1}{\epsilon}))}{1 - \frac{\Delta}{2} tanh(2\beta)} \rceil$.

3.2 On Complete Graph

Consider the complete graph G on n vertices which is the graph which includes all $\binom{n}{2}$ possible edges. Observe that here the interaction term $\sigma(v) \sum_{w|w \sim v} \sigma(w)$ is $\mathcal{O}(n)$. We take $\beta = \frac{\alpha}{n}$ with $\alpha = \mathcal{O}(1)$ such that $\beta \sum \sigma(v)\sigma(w)$ is $\mathcal{O}(1)$.

We shall show that considering the Glauber dynamics for Ising Model on complete graph G with inverse temperature $\beta = \frac{\alpha}{n}$ we get an upper bound of $\lceil \frac{n(\log n + \log(\frac{1}{\epsilon}))}{1 - \alpha} \rceil$ for it's mixing time if $\alpha < 1$ and a lower bound of $C_0 e^{nr(\alpha)}$ (for some universal constant $C_0 > 0$) where $r(\alpha) > 0$ if $\alpha > 1$.

3.2.1 Upper Bound Using Fast Mixing Upper Bound

Here $\Delta = (n-1)$ for all vertices as the graph is complete on n vertices. Now $\Delta\tanh(\beta) = (n-1)\tanh(\frac{\alpha}{n}) \leq \alpha$. So for $\alpha < 1$, we have $\Delta\tanh(\beta) < 1$ and hence by the upper bound of fast mixing in previous part, we get $t_{mix} \leq \lceil \frac{n(\log n + \log(\frac{1}{\epsilon}))}{1-\alpha} \rceil$.

3.2.2 Lower Bound Using Bottleneck-Ratio

Consider the set $A_k := \{\sigma : |\{v : \sigma(v) = 1\}| = k\}$, i.e; set of configurations of Ising model with exactly k no. of $(+1)$'s.

Now by counting observe that for $\sigma \in A_k$, we have

$$\begin{aligned} -\beta H(\sigma) &= \frac{\alpha}{n} \sum_{\substack{v,w \in V \\ v \sim w}} \sigma(v)\sigma(w) \\ &= \frac{\alpha}{n}((+1)^2 \binom{k}{2} + (-1)^2 \binom{n-k}{2} + (+1)(-1)k(n-k)) \\ &= \frac{\alpha}{n}(\binom{k}{2} + \binom{n-k}{2} - k(n-k)). \end{aligned}$$

Thus the stationary distribution (which is the Gibbs distribution) becomes $\pi(A_k) = \frac{\binom{n}{k} e^{\frac{\alpha}{n}(\binom{k}{2} + \binom{n-k}{2} - k(n-k))}}{\sum_{k=0}^{n} \binom{n}{k} e^{\frac{\alpha}{n}(\binom{k}{2} + \binom{n-k}{2} - k(n-k))}} = \frac{\binom{n}{k} e^{\frac{\alpha}{n}(\binom{k}{2} + \binom{n-k}{2} - k(n-k))}}{Z(\alpha)}$.

Define $a_k := \binom{n}{k} e^{\frac{\alpha}{n}((\binom{k}{2} + \binom{n-k}{2} - k(n-k))}$, taking logarithm and using Stirling's formula, we get that $\log(a_{\lfloor cn \rfloor}) = n(-c\log(c) - (1-c)\log(1-c) + \frac{\alpha(1-2c)^2}{2})(1 + o(1))$.

Define $\phi_\alpha(c) := -c\log(c) - (1-c)\log(1-c) + \frac{\alpha(1-2c)^2}{2}$.

Observe $\phi'_\alpha(\frac{1}{2}) = 0$ and $\phi''_\alpha(\frac{1}{2}) = -4(1-\alpha)$. Hence $c = \frac{1}{2}$ is a critical point of ϕ_α, and in particular it is a local maximum or minimum depending on the value of α. Now for $\alpha > 1$, ϕ_α has a local minimum at $\frac{1}{2}$.

Define $S = \{\sigma | \sum_{u \in V} \sigma(u) > 0\}$ and by symmetry $\pi(S) \leq \frac{1}{2}$. Observe that only way to get from S^c to S is through $A_{\lfloor \frac{n}{2} \rfloor}$ as at most one spin can change in 1 step. Therefore we get $\pi(S) = \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \pi(A_j)$ and $Q(S, S^c) \leq \pi(A_{\lfloor \frac{n}{2} \rfloor})$.

Let c_α be the value of c maximizing ϕ_α over $[0, 1/2]$. Since $1/2$ is a strict local minimum, $c_\alpha < 1/2$. Therefore, $\Phi(S) \leq \frac{e^{\phi_\alpha(1/2)\, n(1+o(1))}}{Z(\alpha)\, \pi(A^c_{\lfloor c_\alpha n \rfloor})} = \frac{e^{\phi_\alpha(1/2)\, n(1+o(1))}}{e^{\phi_\alpha(c_\alpha)\, n(1+o(1))}}$.

Since $\phi_\alpha(c_\alpha) > \phi_\alpha(1/2)$, there exist constants $r(\alpha) > 0$ and $b > 0$ such that $\Phi^* \leq be^{-nr(\alpha)}$. Thus $t_{mix} \geq \frac{1}{4b} e^{nr(\alpha)}$.

3.3 On Cycle

Consider a n-cycle which is the graph with n edges, every vertex having exactly 2 neighbors and which becomes a tree if we remove either of the edges. We shall show that considering the Glauber dynamics for Ising Model on n-cycle with inverse temperature $\beta > 0$ the relaxation time becomes $\frac{n}{1-\tanh(2\beta)}$ and for fixed $\epsilon > 0$ we have $\frac{n \log n(1+o(1))}{2(1-\tanh(2\beta))} \leq t_{mix}(\epsilon) \leq \frac{n \log n(1+o(1))}{(1-\tanh(2\beta))}$.

3.3.1 Upper Bound Using Fast Mixing Upper Bound

Here $\Delta = 2$ for all vertices. So $\frac{\Delta}{2} \tanh 2\beta = \tanh 2\beta < 1$. Now by the upper bound of fast mixing of Ising model on graph with all even degree vertices in previous part, we can say $t_{mix}(\epsilon) \leq \frac{n \log n(1+o(1))}{(1-\tanh(2\beta))}$.

3.3.2 Lower Bound Using Wilson's Theorem

Here to give the lower bound we shall use **the following theorem by Wilson**: For an irreducible aperiodic Markov chain with state-space χ and transition matrix P, let Φ be an eigenfunction with real eigenvalue $\lambda \in (\frac{1}{2}, 1)$. Fix $\epsilon \in (0, 1)$ and let $R > 0$ satisfy $\mathbb{E}_x(|\Phi(X_1) - \Phi(x)|^2) \leq R$, $\forall x \in \chi$, then $t_{mix}(\epsilon) \geq \frac{1}{2\log(\frac{1}{\lambda})}(\log(\frac{(1-\lambda)\Phi(x)^2}{2R}) + \log(\frac{1-\epsilon}{\epsilon}))$.

First we shall show that $\Phi(\sigma) := \sum_{i=1}^n \sigma(i)$ is an eigenvector corresponding to eigenvalue $\lambda = 1 - \frac{1}{n}(1 - \tanh 2\beta)$.

If vertex i is selected for updating, a positive spin is placed at i with probability $\frac{1+\tanh\Big(\beta\Big(\sigma(i-1)+\sigma(i+1)\Big)\Big)}{2}$ and if vertex i is selected for updating, a negative spin is placed at i with probability $\frac{1-\tanh\Big(\beta\Big(\sigma(i-1)+\sigma(i+1)\Big)\Big)}{2}$ and a spin i is chosen uniformly with probability $\frac{1}{n}$.

Define $\Phi_i(\sigma) = \sigma(i)$.

Then

$$
\begin{aligned}
&(P\Phi_i)(\sigma) \\
&= (+1)\frac{1+\tanh\Big(\beta\Big(\sigma(i-1)+\sigma(i+1)\Big)\Big)}{2n} + (-1)\frac{1-\tanh\Big(\beta\Big(\sigma(i-1)+\sigma(i+1)\Big)\Big)}{2n} + (1-\frac{1}{n})\sigma(i) \\
&= \frac{\tanh\left(\beta(\sigma(i-1)+\sigma(i+1))\right)}{n} + (1-\frac{1}{n})\sigma(i).
\end{aligned}
$$

Now observe $(\sigma(i-1)+\sigma(i+1)\Big) \in \{-2,0,2\}$ and as $\tanh(.)$ is linear on $\{-2,0,2\}$ and as $\tanh(\beta x) = \frac{\tanh(2\beta)}{2}x$ for $x \in \{-2,0,2\}$, we say that $(P\Phi_i)(\sigma) = \frac{\tanh(2\beta)}{2n}(\sigma(i-1)+\sigma(i+1))) + (1-\frac{1}{n})\sigma(i)$.

Now summing over $i = 1,\dots,n$, we get $(P\Phi)(\sigma) = \frac{\tanh(2\beta)}{2n}(\Phi(\sigma)) + (1-\frac{1}{n})\Phi(\sigma)$ which implies $(P\Phi)(\sigma) = (1-\frac{1}{n}+\frac{\tanh(2\beta)}{2n})\Phi(\sigma) = (1-\frac{1}{n}(1-\tanh 2\beta))\Phi(\sigma)$ which proves $\Phi(\sigma) := \sum_{i=1}^{n}\sigma(i)$ is an eigenvector corresponding to eigenvalue $\lambda = 1-\frac{1}{n}(1-\tanh 2\beta)$.

Now from the upper bound of fast mixing of Ising model on graph with all even degree vertices, we can say $t_{rel} \leq \frac{n}{(1-\tanh 2\beta)}$ and what we just showed tells $\lambda_2 \geq 1-\frac{1}{n}(1-\tanh 2\beta)$ and hence spectral gap $\gamma \leq \frac{1}{n}(1-\tanh 2\beta)$ which implies $t_{rel} \geq \frac{n}{(1-\tanh 2\beta)}$. Hence we get $t_{rel} = \frac{n}{(1-\tanh 2\beta)}$.

Now observe that if $\tilde{\sigma}$ is the state obtained after updating σ according to the Glauber dynamics, then $|\Phi(\tilde{\sigma}) - \Phi(\sigma)| \leq 2$, as by definition $\Phi(\sigma) = \sum_{i=1}^{n}\sigma(i)$ and in one step at most the value can increase by 2 if a (-1) spin turns (+1) and at most the value can decrease by 2 if a (+1) spin turns (-1).

Now applying the Wilson's theorem we get $t_{mix}(\epsilon) \geq (1+o(1))(\frac{n\log(\frac{n^2\frac{(1-\tanh 2\beta)}{n}}{8}+\log(\frac{1}{2\epsilon})}{(1-\tanh 2\beta)})$ which gives $\frac{n\log n(1+o(1))}{2(1-\tanh(2\beta))} \leq t_{mix}(\epsilon)$.

4 Ising Model on b-ary tree of depth k

The coupling of Glauber dynamics for the Ising model that was used in the upper bound of fast mixing of Ising model contracts the Hamming distance, provided $\theta\Delta < 1$. Therefore, the Glauber dynamics for the Ising model on a b-ary tree mixes in $O(n \log n)$ steps, provided $\theta < \frac{1}{b+1}$. We now improve this result, showing that the same coupling contracts a weighted path metric whenever $\theta < \frac{1}{2\sqrt{b}}$. Let T be a finite, rooted b-ary tree of depth k. Fix $0 < \alpha < 1$. We define a graph with vertex set $\{-1, 1\}^T$ by placing an edge between configurations σ and τ if they agree everywhere except at a single vertex v. The length of this edge is defined to be $\alpha^{|v|-k}$, where $|v|$ denotes the depth of vertex v. The shortest path between arbitrary configurations σ and τ has length $\rho(\sigma, \tau) = \sum_{v \in T} \alpha^{|v|-k} \mathbf{1}\{\sigma(v) \neq \tau(v)\}$.

4.1 Upper Bound for Mixing Time

Theorem 4.1. *Let $\theta := \tanh(\beta)$. Consider the Glauber dynamics for the Ising model on finite rooted b-ary tree T of depth k, which has $n = \Theta(b^k)$ vertices. If $\alpha = 1/\sqrt{b}$, then for any pair of neighboring configurations σ and τ, there exists a coupling (X_1, Y_1) of the Glauber dynamics started from σ and τ such that the metric $\rho(\sigma, \tau) = \sum_{v \in T} \alpha^{|v|-k} \mathbf{1}\{\sigma(v) \neq \tau(v)\}$ contracts whenever $\theta < 1/(2\sqrt{b})$. Specifically, letting $c_\theta := 1 - 2\theta\sqrt{b}$, we have $\mathbb{E}_{\sigma,\tau}\Big[\rho(X_1, Y_1)\Big] \leq \Big(1 - \frac{c_\theta}{n}\Big)\rho(\sigma, \tau)$. Therefore, if $\theta < 1/(2\sqrt{b})$, then $t_{\mathrm{mix}}(\varepsilon) \leq \frac{n}{c_\theta}\lceil \frac{3}{2}\log n + \log(1/\varepsilon)\rceil$.*

Now we shall prove the above theorem. Suppose σ and τ are configurations which agree everywhere except v where $(-1) = \sigma(v) = -\tau(v)$. Therefore $\rho(\sigma, \tau) = \alpha^{|v|-k}$.

Let (X_1, Y_1) be one step of the coupling used for fast mixing in high temperature. A vertex v is selected uniformly with probability $\frac{1}{n}$. We say the coupling fails if a neighbor w of v is selected and the coupling doesn't update the spin at w identically in both σ and τ. Given a neighbor of v is selected for updating, the coupling fails with probability $(p(\tau; w) - p(\sigma; w)) \leq \tanh(\beta) = \theta$.

If a child w of v is selected for updating and the coupling fails, then the distance increased by $(\rho(X_1, Y_1) - \rho(\sigma, \tau)) = \alpha^{|v|-k+1} = \alpha\rho(\sigma, \tau)$.

If the parent of v is selected for updating and the coupling fails, then the distance increased by $\rho(X_1, Y_1) - \rho(\sigma, \tau) = \alpha^{|v|-k+1} = \alpha^{-1}\rho(\sigma, \tau)$.

Therefore $\mathbb{E}_{\sigma,\tau}(\rho(X_1, Y_1)) \leq \frac{\theta b}{n}\alpha\rho(\sigma,\tau) + \frac{\theta}{n}\alpha^{-1}\rho(\sigma,\tau) + (1-\frac{1}{n})\rho(\sigma,\tau)$ which gives us $\frac{\mathbb{E}_{\sigma,\tau}(\rho(X_1,Y_1))}{\rho(\sigma,\tau)} \leq (1 - \frac{1}{n} + \frac{\alpha^{-1}+b\alpha}{n})$.

The function $f(\alpha) = \alpha^{-1} + b\alpha$ is minimized over $[0,1]$ at $\alpha = \frac{1}{\sqrt{b}}$ where it has value $2\sqrt{b}$.

Thus $\frac{\mathbb{E}_{\sigma,\tau}(\rho(X_1,Y_1))}{\rho(\sigma,\tau)} \leq (1 - \frac{1}{n} + \frac{2\sqrt{b}\theta}{n})$.

So for $\theta < \frac{1}{2\sqrt{b}}$, the contraction is obtained. The diameter of the tree in the metric ρ isn't more than $\alpha^{-k}n = b^{\frac{k}{2}}n$, the diameter is at most $n^{\frac{3}{2}}$. Hence $d(t) \leq e^{-\frac{c_\theta t}{n}} diam(\chi) \leq n\sqrt{n}e^{-\frac{c_\theta t}{n}}$. Consequently, $t_{mix}(\epsilon) \leq \frac{n}{c_\theta}\lceil \frac{3}{2}\log n + \log(1/\varepsilon)\rceil$.

4.2 Upper Bound for Relaxation Time Using Compression Ratio

Theorem 4.2. *The Glauber dynamics for the Ising model on the finite, rooted, b-ary tree of depth k satisfies* $t_{\mathrm{rel}} \leq n_k^{c_T(\beta;b)}$, *where* $c_T(\beta;b) := \frac{2\beta(3b+1)}{\log b} + 1$ *and* n_k *is the number of vertices in the tree.*

Now we shall prove this with the following lemma: Let $G = (V,E)$ have maximum degree Δ, where $|V| = n$, and let $\widetilde{G} = (V, \widetilde{E})$, where $\widetilde{E} \subset E$. Let $r = |E \setminus \widetilde{E}|$. If γ is the spectral gap for the Glauber dynamics for the Ising model on G and $\widetilde{\gamma}$ is the spectral gap for the dynamics on $\widetilde{G}$, then $\frac{1}{\gamma} \leq \frac{e^{2\beta(\Delta+2r)}}{\widetilde{\gamma}}$.

First we shall prove the above result.

For any $\sigma \in \{-1,1\}^V$, we have

$$\begin{aligned}
\pi(\sigma) &= \frac{e^{\beta\sum_{(v,w)\in E}\sigma(v)\sigma(w)+\beta\sum_{(v,w)\in\tilde{E}\setminus E}\sigma(v)\sigma(w)}}{\sum_\tau e^{\beta\sum_{(v,w)\in E}\tau(v)\tau(w)+\beta\sum_{(v,w)\in\tilde{E}\setminus E}\tau(v)\tau(w)}} \\
&\geq \frac{e^{\beta r}e^{\beta\sum_{(v,w)\in\tilde{E}}\sigma(v)\sigma(w)}}{e^{-\beta r}\sum_\tau e^{\beta\sum_{(v,w)\in\tilde{E}}\tau(v)\tau(w)}} \\
&= e^{-2\beta r}\tilde{\pi}(\sigma),
\end{aligned}$$

where $\tilde{\pi}(\sigma) := \frac{e^{\beta\sum_{(v,w)\in\tilde{E}}\sigma(v)\sigma(w)}}{\sum_\tau e^{\beta\sum_{(v,w)\in\tilde{E}}\tau(v)\tau(w)}}$.

So, $\tilde{\pi}(\sigma) \leq e^{2\beta r}\pi(\sigma)$.

Hence for any configurations σ and τ, we have $P(\sigma,\tau) \geq \frac{1}{n}\frac{\mathbb{1}(P(\sigma,\tau)>0)}{(1+e^{2\beta\Delta})}$ and also $\tilde{P}(\sigma,\tau) \leq \frac{1}{n}\frac{\mathbb{1}(P(\sigma,\tau)>0)e^{2\beta\Delta}}{(1+e^{2\beta\Delta})} \leq \frac{1}{n}\mathbb{1}(P(\sigma,\tau)>0)e^{2\beta\Delta}$ (As $(1+e^{2\beta\Delta})>1$).

Hence we have $\tilde{\pi}(\sigma)\tilde{P}(\sigma,\tau) \leq e^{2\beta(\Delta+r)}\pi(\sigma)P(\sigma,\tau)$.

So for any function

$$\begin{aligned}\tilde{\varepsilon}(\{) &:= \sum_{\sigma,\tau\in\chi}(f(\sigma)-f(\tau))^2\tilde{\pi}(\sigma)\tilde{P}(\sigma,\tau)\\ &\leq e^{2\beta(\Delta+r)}\sum_{\sigma,\tau\in\chi}(f(\sigma)-f(\tau))^2\pi(\sigma)P(\sigma,\tau)\\ &= e^{2\beta(\Delta+r)}\varepsilon(\{).\end{aligned}$$

Now a result tells for two reversible transition matrices $\tilde{P}$ and P with stationary distributions $\tilde{\pi}$ and π respectively, if for any function f, we get $\tilde{\varepsilon}(\{) \leq \alpha\varepsilon(\{)$, then it can be shown that $\tilde{\gamma} \leq \max_{x\in\chi}\frac{\pi(x)}{\tilde{\pi}(x)}\alpha\gamma$.

Thus for our present problem we have $\tilde{\varepsilon}(\{) \leq e^{2\beta(\Delta+r)}\varepsilon(\{)$ for any function f which implies $\tilde{\gamma} \leq \max_{x\in\chi}\frac{\pi(x)}{\tilde{\pi}(x)}e^{2\beta(\Delta+r)}\gamma$ and as we already got $\tilde{\pi}(\sigma) \leq e^{2\beta r}\pi(\sigma)$, we can say $\tilde{\gamma} \leq e^{2\beta(\Delta+2r)}\gamma$.

Thus the stated lemma is proved. Now we prove the theorem.

Consider $\tilde{T}_{b,k}$ to be the graph obtained by removing all edges incident to the root. Now applying the previous lemma we get $t_{rel}(T_{k+1}) \leq e^{2\beta(\Delta+2r)}t_{rel}(\tilde{T}_{b,k+1})$ which implies $\frac{t_{rel}(T_{k+1})}{n_{k+1}} \leq e^{2\beta(\Delta+2r)}\frac{t_{rel}(\tilde{T}_{b,k+1})}{n_{k+1}}$.

As $\Delta = b+1$ and $r = b$, we get $\frac{t_{rel}(T_{k+1})}{n_{k+1}} \leq e^{2\beta(3b+1)}\frac{t_{rel}(\tilde{T}_{b,k+1})}{n_{k+1}}$.

Now as $\tilde{T}_{b,k+1}$ is a partition of t_{k+1},

$$\begin{aligned}\frac{t_{rel}(\tilde{T}_{b,k+1})}{n_{k+1}} &= \frac{1}{\gamma n_{k+1}}\\ &= \max_{1\leq j\leq k}\frac{1}{n_j\gamma_j},\end{aligned}$$

where $n_1 = 1, n_2 = \ldots = n_k = n, \gamma_1 = 1, \gamma_2 = \ldots = \gamma_k$. (This comes from the following lemma: Consider a partition $\{V_i\}$ of a finite set V. For finite set S, π_i be a probability distribution on S^{V_i} and π be a probability distribution on S^V such that $\pi = \prod_{i=1}^d \pi_i$. Then if γ be spectral gap of the Glauber dynamics on S^V for π and γ_i be spectral gap of the Glauber dynamics on S^{V_i} for π_i we get $\frac{1}{n\gamma} = \max_{1\leq j\leq d} \frac{1}{n_i\gamma_i}$ where $|V| = n$ and $|V_i| = n_i$.)

Hence, $\frac{t_{rel}(\tilde{T}_{b,k+1})}{n_{k+1}} = \max\{1, \frac{t_{rel}(T_k)}{n_k}\}$. Therefore $\frac{t_{rel}(T_{k+1})}{n_{k+1}} \leq e^{2\beta(3b+1)} \max\{1, \frac{t_{rel}(T_k)}{n_k}\}$.

Now since $n_k \geq b^k$, we get $t_{rel}(T_k) \leq e^{2\beta(3b+1)} n_k \leq n_k^{1+\frac{2\beta(3b+1)}{\log b}}$.

5 Ising Model on Square

Consider the Glauber dynamics for the Ising model on the $n \times n$ box. Let $V = \{(j,k) \in \mathbb{Z}^2 : 0 \leq j,k \leq n-1\}$ and let $G = (V,E)$ be the graph where two vertices (j,k) and (j',k') are connected by an edge if and only if $\|(j,k)-(j',k')\|_2 = 1$.

A **theorem by Schonmann and Thomas** tells that the relaxation time $\frac{1}{(1-\lambda_*)}$ of the Glauber dynamics for the Ising model in an $n \times n$ square in 2 dimensions is at least $e^{\psi(\beta)n}$ where $\psi(\beta) > 0$ for sufficiently large β. More precisely, for $\alpha_\ell < 3^\ell$ (the no. of self-avoiding paths starting from origin of $\mathbb{Z}^2$ with length ℓ) and $\alpha := \lim_{\ell \longrightarrow \infty} \alpha_\ell^{\frac{1}{\ell}} \leq 3$ (the "connective constant" for the planar square lattice), if $\beta > \frac{1}{2}\log(\alpha)$ then we have $\psi(\beta) > 0$.

Here we shall prove this using the key idea presented in **Randall(2006)** for the hardcore lattice gas, where a topological obstruction was used rather than the usual cat by magnetization. An upper bound of $e^{C(\beta)n^{d-1}}$ for the relaxation time was proved by **Sinclair and Jerrum** for all dimensions, for all β.

For the proof of lower bound it will be useful to attach the spins to the lattice squares of the lattice rather than nodes.

5.1 Lower Bound Using Topological Structure

First we shall introduce a **fault-line**. A **fault line** (with at most k defects) is a self-avoiding lattice path in $\mathbb{Z}^2$ that connects either the left side to the right side, or the top side to the bottom of the box $[0,n]^2$, such that for each edge of the path, with at most k exceptions, the two faces adjacent to that edge carry opposite spins. In particular, no edge of the fault line lies on the boundary of $[0,n]^2$.

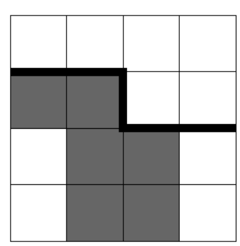

Figure 4: A fault line with 1 defect. (+1) spins are indicated by shaded squares and (-1) spins are indicated by white squares.

Lemma 5.1. *Denote by $\mathcal{F}_k$ the set of Ising configurations on $[0,n]^2$ that admit a fault line with at most k defects. Then $\pi(\mathcal{F}_k) \leq \sum_{\ell \geq n} 2(n+1)\, \alpha_\ell \, e^{2\beta(2k-\ell)}$. In particular, if k is fixed and $\beta > \frac{1}{2}\log\alpha$, then $\pi(\mathcal{F}_k)$ decays exponentially in n.*

Now we prove the above lemma.

For a self-avoiding lattice path ϕ of length ℓ from the left side to the right side (or from top to bottom) of $[0,n]^2$, let $\mathcal{F}_\phi$ be the set of Ising configurations on $[0,n]^2$ that has ϕ as a fault line with at most k defects.

Flipping all spins on one side of the fault line (say, the side containing the upper-left corner) defines a one-to-one mapping from $\mathcal{F}_\phi$ to its complement. This mapping increases the probability of a configuration by a factor of $e^{2\beta(\ell-2k)}$ and hence $\pi(\mathcal{F}_\phi) \leq e^{2\beta(2k-\ell)}$.

Now the no. of self avoiding lattice paths from left to right in $[0,n]^2$ is at most $(n+1)\alpha_l$ (for each starting point $(0,k)$ at most α_l for $k = 0,\dots,n$) and also from top to bottom at most $(n+1)\alpha_l$ (for each starting point (k,n) at most α_l for $k = 0,\dots,n$), no. of total self-avoiding paths ϕ in $[0,n]^2$ is at most $2(n+1)\alpha_l$ and for each ϕ, we have $\pi(\mathcal{F}_\phi) \leq e^{2\beta(2k-\ell)}$ where $l \leq n$, so $\pi(\mathcal{F}_k) \leq \sum_{\ell \geq n} 2(n+1)\, \alpha_\ell \, e^{2\beta(2k-\ell)}$.

Now we shall state another lemma and prove it which gives more idea on this fault line and it's defects.

Lemma 5.2. *(a) If, in a configuration σ, there is no all-plus crossing from the left side L of $[0,n]^2$ to the right side R, and there is also no all-minus crossing, then there exists a fault line with no defects connecting the top side to the bottom side of $[0,n]^2$.*

(b) Similarly, suppose γ^+ is a path of lattice squares, all labeled $+1$ in σ, from a square $q \subset [0,n]^2$ to the top side of $[0,n]^2$, and γ^- is a path of lattice squares, all labeled -1 in σ, from the same square q to the top side of $[0,n]^2$. Then there exists a lattice path ξ from the boundary of q to the top side of $[0,n]^2$ such that every edge of ξ is adjacent to two lattice squares with opposite labels in σ.

First we shall prove (a). Let A be the collection of lattice squares that can be reached from L (left boundary) by a path of lattice squares of the same label in configuration σ. Let $A^* = A \cup$ the set of squares that are separated from R by A. Then the boundary of A^* consists of part of the boundary of $[0,n]^2$ and a fault line which has 0 defects.

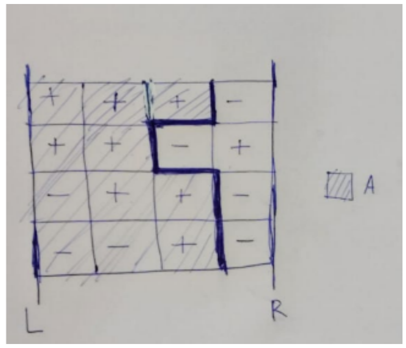

Figure 5: The black line from top to bottom is fault line without any defects. There is no all same label crossing from left to right, but from top to bottom there are all +1 crossings.

Now we shall prove (b). Suppose that the square q itself is labeled -1 in the configuration σ, and that γ^+ terminates at a square q^+ on the top side of $[0,n]^2$ which lies strictly to the left of the square q^- where γ^- terminates. Let A^+ be the collection of lattice squares that can be reached from γ^+ by a path of lattice squares labeled +1 in σ, and let A^*_+ denote the union of A^+ with the set of squares that are separated from the boundary of $[0,n]^2$ by A^+.

Let ξ_1 be a directed lattice edge having the square q on its right and a square belonging to γ^+ on its left. We extend ξ_1 to a directed lattice path ξ leading to the boundary of $[0,n]^2$ by inductively choosing each successive edge ξ_j so that it has a square (labeled +) in A^+ on its left and a square (labeled $-$) not belonging to A^*_+ on its right. Such a choice is always possible until ξ reaches the boundary of $[0,n]^2$. Moreover, the path ξ cannot form a cycle and must terminate on the top side of $[0,n]^2$ at a point lying between q^+ and q^-.

Now we prove the theorem for lower bound stated by Schonmann and Thomas.

Theorem 5.3. *The relaxation time $\frac{1}{(1-\lambda_*)}$ of the Glauber dynamics for the Ising model in an $n \times n$ square in 2 dimensions is at least $e^{\psi(\beta)n}$ where $\psi(\beta) > 0$ for sufficiently large β. More precisely, for $\alpha_\ell < 3^\ell$ (the no. of self-avoiding paths starting from origin of $\mathbb{Z}^2$ with length ℓ) and $\alpha := \lim_{\ell \longrightarrow \infty} \alpha_\ell^{\frac{1}{\ell}} \leq 3$ (the "connective constant" for the planar square lattice), if $\beta > \frac{1}{2}\log(\alpha)$ then we have $\psi(\beta) > 0$.*

To prove this we shall use the 2 previous lemmas. Let $\mathcal{S}_+$ be the set of configurations that admit both a top-to-bottom and a left-to-right crossing of $+1$ spins. Similarly, define $\mathcal{S}_-$ for -1 spins. Note that $\mathcal{S}_+ \cap \mathcal{S}_- = \varnothing$.

On the complement of $\mathcal{S}_+ \cup \mathcal{S}_-$, there is either no monochromatic left-to-right crossing (in which case there exists a top-to-bottom fault line by Lemma 5.2), or no monochromatic top-to-bottom crossing (in which case there exists a left-to-right fault line). Thus by lemma 5.1, $\pi(S_+) \longrightarrow \frac{1}{2}$ as $n \longrightarrow \infty$.

Let $\partial\mathcal{S}_+$ denote the external vertex boundary of $\mathcal{S}_+$, that is, the set of configurations outside $\mathcal{S}_+$ that are one flip away from $\mathcal{S}_+$. It suffices to show that $\pi(\partial\mathcal{S}_+)$ decays exponentially in n for $\beta > \frac{1}{2}\log\alpha$. By Lemma 5.1, it is enough to verify that every configuration $\sigma \in \partial\mathcal{S}_+$ admits a fault line with at most 3 defects. The case $\sigma \in \mathcal{S}_-$ is handled by Lemma 5.2.

Fix $\sigma \in \partial\mathcal{S}_+ \setminus \mathcal{S}_-$, and let q be a lattice square such that flipping the spin $\sigma(q)$ transforms σ into a configuration belonging to $\mathcal{S}_+$.

By Lemma 5.2, there exists a lattice path ξ from the boundary of q to the top side of $[0,n]^2$ such that every edge of ξ is adjacent to two lattice squares with opposite labels in σ; by symmetry, there is also such a path ξ' from the boundary of q to the bottom side of $[0,n]^2$. By adding at most 3 edges of q, we may concatenate these two paths to obtain a fault line with at most 3 defects.