Collision time for Markov chains

Lemma 0.1. *For a given reversible n-state Markov chain P and its stationary distribution π ($\pi_i = O(1/n), \forall i$), denote X_t ($t < \sqrt{n}$)be the t^{th} state of the random walk according to P. Consider a fixed starting state $X_0 = s$, then*

$$Pr[\bigcap_{0<i<t} X_i \neq s] \leq 1 - \Theta(t\pi_s).$$

Proof. At first, claim

$$\Pr[X_{2t} = s] \geq \pi_s$$

. It is because

$$\begin{aligned}
\Pr[X_{2t} = s] &= \sum_i Pr[X_t = i] Pr[X_{2t} = s | X_t = i] \\
&= \sum_i Pr[X_t = i] Pr[X_t = s | X_0 = i] \\
&= \sum_i P^t_{si} P^t_{is} \\
&= \sum_i P^t_{si} P^t_{si} \frac{\pi_s}{\pi_i} \\
&= \pi_s \sum_i (P^t_{si})^2 \frac{1}{\pi_i} \\
&= \pi_s \left(\sum_i (P^t_{si})^2 \frac{1}{\pi_i} \right) \left(\sum_i \pi_i \right) \\
&\geq \pi_s \left(\sum_i P^t_{si} \right)^2 \\
&= \pi_s
\end{aligned}$$

$Pr[\bigcap_{0<i<t} X_i \neq s] = 1 - Pr[\bigcup_{0<i<t} X_i = X_0]$, and we divide two cases.

Case 1. $1/2 < \sum_{0<i<t} Pr[X_0 = X_i]$
It means $\exists j, Pr[X_0 = X_j] > \frac{1}{2t}$. Hence,

$$1 - Pr[\bigcup_{0<i<t} X_0 = X_i] < 1 - Pr[X_0 = X_j] < 1 - \frac{1}{2t} < 1 - \frac{t}{2n} < 1 - \Theta(t\pi_s).$$

The third inequality is because $t < \sqrt{n}$.

Case 2. $1/2 > \sum_{0<i<t} Pr[X_0 = X_i] = T$

$$\begin{aligned}
&1 - Pr[\bigcup_{0<i<t} X_0 = X_t] \\
&< 1 - \sum_{0<i<t} Pr[X_0 = X_i] + \sum_{0<i<j<t} Pr[X_0 = X_i \& X_0 = X_j] \\
&< 1 - \sum_{0<i<t} Pr[X_0 = X_i] + \sum_{0<i<j<t} Pr[X_0 = X_i \& X_i = X_j] \\
&< 1 - \sum_{0<i<t} Pr[X_0 = X_i] + \sum_{0<i<j<t} Pr[X_0 = X_i] Pr[X_0 = X_{j-i}] \\
&< 1 - T + T^2 \\
&< 1 - T/2
\end{aligned}$$

The last inequality is because $T < 1/2$. By the above claim, we know $T > \Theta(t\pi_s)$, which implies $1 - T/2 < 1 - \Theta(t\pi_s)$ and complete the proof.

□

Theorem 0.2. *The collision time of the random walk induced by n-state reversible Markov chain P that has the uniform stationary distribution is $O(\sqrt{n})$.*

Proof. Let X_t be the state after t steps. Define the following event A_i, B_i, C_i:

$$A_i = \bigcap_{0 \le i < j \le \sqrt{n}} X_i \neq X_j$$

$$B_i = \bigcap_{0 \le j < k \le i} X_j \neq X_k$$

$$C_i = \bigcap_{0 \le j \le i < k \le \sqrt{n}} X_j \neq X_k$$

Our goal is to show the following probability Q is bounded by a constant < 1:

$$\begin{aligned}
Q &= \Pr\left[\bigcap_{0 \le i < j \le \sqrt{n}} X_i \neq X_j\right] \\
&= \prod_{i=0}^{\sqrt{n}} \Pr\left[\bigcap_{j \le i} X_j \neq X_{i+1} \middle| B_i\right] \\
&= \prod_{i=0}^{\sqrt{n}} E\left[\sum_{k \neq X(t), \forall t \le i} P_{X_i k} \middle| B_i\right]
\end{aligned}$$

Now suppose $\forall i, E\left[\sum_{k \neq X(t), \forall t \le i} P_{X_i k} \middle| B_i\right] > 1/2$, otherwise there is nothing to prove.

$$\begin{aligned}
Q &= \Pr[\bigcap_{1 \le i < \sqrt{n}} A_i] \\
&= \prod_{1 \le i < \sqrt{n}} Pr[A_i | A_1, \ldots, A_{i-1}]
\end{aligned}$$

Also, $Pr[A_i|A_1,\ldots,A_{i-1}]$ is same as $Pr[A_i]$ in the Markov chain $P^{(i)} = P - \{X_0,\ldots,X_{i-1}\}$, which is also reversible. Let $\pi^{(i)}$ be the stationary distribution of $P^{(i)}$. By Lemma 0.1,

$$Pr[A_i|A_1,\ldots,A_{i-1}] \le E\left[1-\Theta\left((\sqrt{n}-i)\pi^{(i)}_{X_i}\right)|A_1,\ldots,A_{i-1}\right]$$

Note that

$$\pi^{(i)}_{X_i} \approx \frac{\sum_{k\neq X(t),\forall t\le i} P_{X_i k}}{n}.$$

Because $\bigcap_{j<i} A_j = B_i \bigcap C_i$, $E[\pi^{(i)}_{X_i}|A_1,\ldots,A_{i-1}] = E[\pi^{(i)}_{X_i}|B_i,C_i]$.

$$\begin{aligned}
E[\pi^{(i)}_{X_i}|B_i] &= \Pr(C_i|B_i)E[\pi^{(i)}_{X_i}|B_i,C_i] + \Pr(\bar{C}_i|B_i)E[\pi^{(i)}_{X_i}|B_i,\bar{C}_i] \\
&\le \Pr(C_i|B_i)E[\pi^{(i)}_{X_i}|B_i,C_i] + \Pr(\bar{C}_i|B_i)\frac{1}{n} \\
&\le E[\pi^{(i)}_{X_i}|B_i,C_i] + \frac{1}{100n}
\end{aligned}$$

The last inequality is because $\Pr(C_i|B_i) \ge 99/100$, otherwise there is nothing to prove because $Q < \Pr(B_i,C_i) = \Pr(B_i)\Pr(C_i|B_i) < 99/100$.

$$\begin{aligned}
\sum_{i\le\sqrt{n}} (\sqrt{n}-i)E[\pi^{(i)}_{X_i}|B_i,C_i] &\ge \sum_{i\le\sqrt{n}} (\sqrt{n}-i)\left(E[\pi^{(i)}_{X_i}|B_i] - \frac{1}{100n}\right) \\
&\ge \sum_{i\le\sqrt{n}} (\sqrt{n}-i)\left(\frac{1}{2n} - \frac{1}{100n}\right) \\
&\ge 0.1
\end{aligned}$$

Therefore,

$$\begin{aligned}
Q &= \prod_{1\le i<\sqrt{n}} Pr[A_i|A_1,\ldots,A_{i-1}] \\
&\le \prod_{1\le i<\sqrt{n}} E\left[1-\Theta\left((\sqrt{n}-i)\pi^{(i)}_{X_i}\right)|A_1,\ldots,A_{i-1}\right] \\
&\approx Exp\left[-\sum_{i\le\sqrt{n}} (\sqrt{n}-i)E[\pi^{(i)}_{X_i}|B_i,C_i]\right] \\
&\le e^{-0.1}
\end{aligned}$$

□