In this lecture, we introduce the method of coupling, which will allow us to obtain bounds on the mixing time of several Markov chains. The idea of coupling is to consider two random walks on a Markov chain such that the walks viewed in isolation have the correct distribution but make dependent transitions at each stage. By starting one of the random walks from the stationary distribution and bounding the time for the two chains to collide, we obtain bounds on the mixing time of the random walk.

Example. Consider the following random walk on the n-dimensional hypercube. At any state Y_t,

1. pick $i \in [n], r \in \{0, 1\}$ uniformly at random

2. set the ith bit of Y_t to r

In a coupling of walks $X = X_0, X_1, X_2 \ldots$ and $Y = Y_0, Y_1, Y_2 \ldots$, we allow X_t to depend on $X_0, X_1, \ldots X_{t-1}$ and $Y_0, Y_1, \ldots Y_t$, as long as it remains faithful to the Markov Chain. Consider the following coupling for the random walk on the hypercube. At state X_t, Y_t,

1. pick $i \in [n], r \in \{0, 1\}$ uniformly at random

2. set the ith bit of X_t and the ith bit of Y_t to r

Note that if Y_0 is chosen from the stationary distribution, then Y_i will be uniform for all i, and if the ith coordinate of X_t and Y_t agree for some t, this coordinate will agree thereafter. Let D_t denote the number of positions in which X_t and Y_t differ. We would like to bound the expected value of D_t. Since at time t, the probability of choosing one of the D_t positions in which X_t and Y_t differ is $\frac{D_t}{n}$, we have

$$E[D_{t+1}] = D_t - \frac{D_t}{n},$$

and therefore

$$E[D_t] \leq \left(1 - \frac{1}{n}\right)^t D_0 \leq \left(1 - \frac{1}{n}\right)^t n \leq n e^{-\frac{1}{n}t}$$

We say X and Y are *coupled* at time t if $X_t = Y_t$. The following lemma shows that a bound on the coupling time of a random walk provides a bound on the mixing rate.

Lemma 1. Coupling Lemma $|P_x^t - P_y^t| \leq 2Pr[X_t \neq Y_t]$

Proof. Let $A = \{v : P_X^t(v) \geq P_Y^t\}$. Then

$$
\begin{aligned}
\frac{1}{2}|P_X^t - P_Y^t| &= \sum_{v \in A} P_X^t(v) - P_Y^t(v) = P_X^t(A) - P_Y^t(A) = Pr(X_t \in A) - Pr(Y_t \in A) \\
&= Pr(X_t \in A, Y_t \in A) + Pr(X_t \in A, Y_t \notin A) - Pr(X_t \in A, Y_t \in A) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - Pr(X_t \in A, Y_t \in A) \\
&\leq Pr(X_t \in A, Y_t \notin A) \\
&\leq Pr(X_t \neq Y_t).
\end{aligned}
$$

As a corollary, the random walk on the hypercube has mixing time is $O(n \log n)$.

Examples.

1. Consider the random walk on the line $[0, n-1]$ with

$$
\begin{aligned}
Pr[X_{t+1} = X_t + 1] = Pr[X_{t+1} = X_t - 1] &= \frac{1}{2} \quad \text{for} \quad x \neq 0, n-1 \\
Pr[X_{t+1} = X_t] = Pr[X_{t+1} = X_t + 1] &= \frac{1}{2} \quad \text{for} \quad x = 0 \\
Pr[X_{t+1} = X_t] = Pr[X_{t+1} = X_t - 1] &= \frac{1}{2} \quad \text{for} \quad x = n-1.
\end{aligned}
$$

Let Y_0 be chosen from the stationary distribution and couple X and Y by setting $X_{t+1} = X_t + (Y_{t+1} - Y_t)$ if possible, and $X_{t+1} = X_t$ otherwise. Without loss of generality, assume $X_0 < Y_0$. Then the expected coupling time for X and Y is bounded above by the expected hitting time for a random walk starting at X_0 to reach $n-1$. Since this is at most n^2, the coupling time (and by the Coupling Lemma, the mixing time) is at most n^2.

2

2. Consider the random walk on a cycle, in which transitions are unit steps clockwise or counterclockwise and stationary moves, each with probability $\frac{1}{3}$. Note that if X_0 and Y_0 have different parity and we attempt to use the identity coupling for X and Y, then X_t and Y_t will have different parity for all t and the two chains will never couple. Instead, we must adjust the coupling according to the distance between the states.

Case I. X_t, Y_t have the same parity. In this case, X choses the same move as Y.

Case II. X_t, Y_t have different parity. In this case, couple X and Y as follows

$$
\begin{array}{ccc}
Y & & X \\
\text{counterclockwise} & \rightarrow & \text{counterclockwise} \\
\text{clockwise} & \rightarrow & \text{stationary} \\
\text{stationary} & \rightarrow & \text{clockwise}
\end{array}
$$

Now, if X_t and Y_t have different parity, then X_{t+1} and Y_{t+1} have the same parity with probability $\frac{2}{3}$. Thus, in expected $\frac{3}{2}$ steps, we will move to Case I.

Sampling Proper Colorings

Recall that a proper coloring of a graph $G = (V, E)$ is a coloring of the vertices such that any pair of adjacent vertices receives different colors. If the maximum degree of G is Δ, then the greedy algorithm gives a proper coloring of G with $k = \Delta + 1$ colors. Our goal is to sample uniformly from the set of proper colorings using the following Markov Chain.

(i) Pick vertex v at random

(ii) Pick color c at random

(iii) Recolor v with c if possible, otherwise do nothing

Notice that because the loop probabilities are nonzero, the chain is aperiodic. For what values of k is the chain irreducible? The complete graph $K_{\Delta+2}$ shows that the chain is not irreducible for $k = \Delta + 1$, since no vertex can be recolored at any step.

For $k = \Delta + 2$, consider two colorings σ and τ and an ordering $v_1, v_2, \ldots v_n$ on the vertices of G. Starting from coloring σ, recolor the vertices v_i in order as follows: switch v_i from color $\sigma(v_i)$ to $\tau(v_i)$ if such a switch results in a proper coloring. Otherwise, v_i has adjacent vertices $v_{j_1}, v_{j_2} \ldots v_{j_l}$ of color $\tau(v_i)$, with $i < j_k$ for $1 \le k \le l$ (since τ is a proper coloring). Now, for each v_{j_k}, recolor v_{j_k} to a color not in the neighborhood of v_{j_k} and not equal to $\tau(v_i)$. Since this forbids at most $\Delta + 1$ colors, there is at least one legal color for each v_{j_k}.

3

After recoloring all the v_{j_k}, giving v_i color $\tau(v_i)$ will result in a proper coloring. This gives a path in the Markov chain between any two colorings and therefore, the chain is irreducible.

Theorem 2. *The Markov chain is rapidly mixing for $k \geq 2\Delta$.*

Proof. Let A_t denote the set of vertices which agree in the two colorings at time t, let D_t denote the vertices which disagree, and define

$$d'(v) = \begin{cases} \text{number of neighbors of } v \text{ in } A_t \text{ for } v \in D_t \\ \text{number of neighbors of } v \text{ in } D_t \text{ for } v \in A_t \end{cases}$$

Then $\sum_{v \in A_t} d'(v) = \sum_{v \in D_t} d'(v) = m'$, the number of edges crossing D_t and A_t. We will define a bijection g on the colors and onsider the following coupling: if Y picks color c and recolors vertex v with c, then X recolors v with color $g(c)$. Let C_X (C_Y) denote the colors X (Y) has in the neighborhood of v that Y (X) does not have. C_X corresponds to the set of colors Y (but not X) can use to recolor v.

If $v \in D_t$, let g be the identity. Otherwise, we can assume $|C_X| \leq |C_Y|$ without loss of generality. Define g in such a way that it is a bijection from C_X onto a subset S of colors in C_Y with $|C_X| = |S|$ and leaves all other colors fixed.

If v moves from the set of agreeing vertices to the set of disagreeing vertices after recoloring, $|D_{t+1}| = |D_t| + 1$. In order for this to occur, the color c must be an element of C_Y (if c is an element of C_X, then v remains the same color in both). Since $|C_Y| \leq d'(v)$, the probablity of this is

$$Pr\left[|D_{t+1}| = |D_t| + 1\right] \leq \frac{1}{n} \sum_{v \in A_t} \frac{d'(v)}{k} = \frac{m'}{kn}.$$

Now, if v moves from the set of disagreeing vertices to agreeing vertices, then $|D_{t+1}| = |D_t| - 1$. Since $v \in D_t$, g is the identity and in order for v to be in A_{t+1}, it suffices for the color c to be different from all the colors X_t and Y_t assigned to neighbors of v. The number of such colors is at least $k - 2\Delta + d'(v)$, so we have

$$Pr\left[|D_{t+1}| < |D_t|\right] \geq \frac{1}{n} \sum_{v \in D_t} \frac{k - 2\Delta + d'(v)}{k} = \frac{k - 2\Delta}{kn}|D_t| + \frac{m'}{kn}$$

Then

4

$$
\begin{aligned}
E(|D_{t+1}|) \;&\leq\; D_t + \frac{m'}{kn} - \frac{k-2\Delta}{kn}|D_t| - \frac{m'}{kn} \\
&=\; D_t - \frac{k-2\Delta}{kn}D_t,
\end{aligned}
$$

implying

$$
E(D_t) \leq \left(1 - \frac{k-2\Delta}{kn}\right)^t n.
$$

In particular, for $t \sim \frac{kn}{k-2\Delta}\log\frac{n}{e}$, we have $E(D_t) \leq \epsilon$, so $|P_X^t - P_Y^t| \leq 2\epsilon$ by the coupling lemma. $\qquad\square$

In the next class, we will see a different chain for sampling proper colorings and show that it is rapidly mixing for any $k > \frac{11}{6}\Delta$ using the method of path coupling. Furthermore, rapid mixing of the new chain will also imply rapid mixing for the chain we have considered here with $k > \frac{11}{6}\Delta$. The problem of whether this chain is rapidly mixing for smaller values of k remains open.

Although coupling is often a useful tool in proving rapid mixing results, it can often be difficult to design appropriate couplings to specific Markov Chains. When the coupling technique works, it usually establishes better bounds on mixing times than known through methods of conductance and canonical paths. However, in 1999, Kumar and Ramesh showed that there are chains for which *no* Markovian coupling argument can prove rapid mixing. In particular, they show that for the Jerrum-Sinclair chain for sampling perfect and near-perfect matchings, there is a notion of "distance" between states, under which most transitions under any coupling are distance increasing. In particular, any coupling requires an exponential number of steps t before states X_t, Y_t become equal, giving an exponential lower bound on the coupling time for any such method.