Homework 10

Out: 11 Apr. Due: 18 Apr.

Submit your solutions in pdf format on Gradescope by **5pm on Friday, April 18**. *Solutions may be written either in LaTeX (with either machine-drawn or hand-drawn diagrams) or* **legibly** *by hand. (The LaTeX source for this homework is provided in case you want to use it as a template.) Please be sure to begin the solution for each problem on a new page, and to tag each of your solutions to the correct problem! Per course policy, no late solutions will be accepted. Take time to write* **clear** *and* **concise** *answers; confused and long-winded solutions may be penalized. You are encouraged to form small groups (two to four people) to work through the homework, but you* **must** *write up all your solutions on your own. Depending on grading resources, we reserve the right to grade a random subset of the problems and check off the rest; so you are advised to attempt all the problems.*

1. Let $G = (V, E)$ be a connected graph, and let $\mathcal{T}$ denote the set of all *spanning trees* of G. (Note that $|\mathcal{T}|$ is in general exponentially large as a function of the size of G.) Consider the Markov chain on state space $\mathcal{T}$ which makes transitions from its current state (a tree T) as follows:

- pick an edge $e \in E \setminus T$ u.a.r.
- add e to T, creating a unique cycle C
- remove one edge of C u.a.r., obtaining a new tree T'

(a) Show that the above Markov chain is ergodic (i.e., irreducible and aperiodic) and that its stationary distribution is uniform over $\mathcal{T}$.

(b) Explain how to modify the transition probabilities of the chain so that its stationary distribution is $\pi(T) \propto \prod_{e \in T} \lambda_e$, where $\{\lambda_e : e \in E\}$ are positive edge weights. Be sure to justify why this new distribution is indeed stationary.

2. Consider random walk on the n-dimensional hypercube $\{0,1\}^n$, where to avoid periodicity we introduce a self-loop probability of $\frac{1}{2}$ on every state. This process can be described more formally as follows, where the current state is $x = x_1 \dots x_n$:

- Pick a coordinate $i \in \{1, \dots, n\}$ u.a.r.
- Set $x_i = 0$ or 1, each with probability $\frac{1}{2}$.

[You should check that this process corresponds to the above random walk.]

(a) Verify that this random walk is ergodic and that its stationary distribution is uniform over $\{0,1\}^n$.

(b) Devise a coupling for the random walk and use it to show that the mixing time is $O(n \log n)$.

3. Recall the "random transpositions" card shuffle that we defined in class. Here the states, as usual, are all $n!$ permutations of an n-card deck, and at each step the shuffle proceeds as follows:

- pick two positions, $i, j \in \{1, \dots, n\}$ independently and u.a.r. (note that $i = j$ is possible)
- swap the cards at positions i and j

As we saw in class, this shuffle converges to the uniform distribution. (It is irreducible because any permutation can be written as the product of transpositions; it is aperiodic because there is a self-loop probability of $1/n$ at each state; and the stationary distribution is uniform because the transition probabilities are symmetric.)

In this problem you will show that $O(n^2)$ shuffles are enough to mix up the deck.

Here is a coupling (X_t, Y_t) for this process. At each step, we choose a position $i \in \{1, \ldots, n\}$ and a card c u.a.r. Then in both copies X_t, Y_t we swap card c with the card in position i. (Note that this is a valid coupling, because both copies, viewed separately, are in fact swapping the cards in two randomly chosen positions, as specified in the original process.)

To analyze this coupling, let $d_t = d(X_t, Y_t)$ be the *distance* between the two copies after t steps, i.e., the number of cards whose positions differ in X_t and Y_t.

(a) Explain *carefully* why d_t never increases with t.

(b) Show that d_t decreases by at least 1 with probability $\left(\frac{d_t}{n}\right)^2$.

(c) Deduce that, for any choice of initial states X_0, Y_0, the expected number of steps T until $X_T = Y_T$ is at most cn^2 for some constant c. [HINT: Recall the expected value of a geometric r.v. Recall also that $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$.]

(d) Finally, deduce that the mixing time satisfies $\tau(\varepsilon) \leq \frac{cn^2}{\varepsilon}$. [HINT: Use Markov's inequality and the coupling lemma. In fact, the mixing time for this process satisfies $\tau(\varepsilon) \leq cn^2 \log(\frac{1}{\varepsilon})$, but you are not required to prove this.]

4. In this problem we will show that we can simplify the definition of mixing time to make it depend on just a specific constant variation distance, in analogous fashion to the way we use a standard value of $\frac{1}{2}$ or $\frac{1}{4}$ for the error probability in randomized algorithms. In particular, we'll show that, for an ergodic Markov chain, $\tau(\varepsilon) \leq \tau(1/2e) \cdot \ln(1/\varepsilon)$ for any $\varepsilon > 0$. This means we can define $\tau_{\text{mix}} := \tau(1/2e)$ (the time until the variation distance drops down to $1/2e$) as the mixing time, knowing that we only pay a logarithmic penalty on top of τ_{mix} for achieving any desired variation distance ε. Our analysis will make extensive use of coupling. In particular, you will need to recall that, for any r.v.'s X, Y with distributions μ, ν respectively, any coupling of X, Y satisfies $\Pr[X \neq Y] \geq \|\mu - \nu\|$, and there is an "optimal" coupling for which equality holds, i.e., for which $\Pr[X \neq Y] = \|\mu - \nu\|$.

For a finite, ergodic Markov chain with transition matrix P and stationary distribution π, we will use the following definitions (the first two of which we've seen in class):

$$\Delta_x(t) := \|P_x^t - \pi\|; \qquad \Delta(t) := \max_x \Delta_x(t); \qquad D_{xy}(t) := \|P_x^t - P_y^t\|; \qquad D(t) := \max_{x,y} D_{xy}(t).$$

(a) Prove that $\Delta_x(t)$ is monotonically decreasing in t. [HINT: Couple two copies $(X_t), (Y_t)$ of the chain, with $X_0 = x$ and $Y_0 \sim \pi$. Start with an optimal coupling, so that $\Pr[X_t \neq Y_t] = \Delta_x(t)$. Now couple the next move of X_t, Y_t so that $\Pr[X_{t+1} \neq Y_{t+1}] \leq \Pr[X_t \neq Y_t]$.]

(b) Prove that $D(s + t) \leq D(s)D(t)$. [HINT: Let $X_0 = x, Y_0 = y$ and again start with an optimal coupling so that $\Pr[X_t \neq Y_t] = D_{xy}(t)$. Now couple the next s steps so that $\Pr[X_{t+s} \neq Y_{t+s} \mid X_t, Y_t] \leq D(s)$.]

(c) Use part (b) to prove that $\tau(\varepsilon) \leq \tau_{\text{mix}} \cdot \lceil \ln(1/\varepsilon) \rceil$, where as above $\tau_{\text{mix}} := \tau(1/2e)$. [HINT: Start by showing that $\Delta(k\tau_{\text{mix}}) \leq D(\tau_{\text{mix}})^k$ for any non-negative integer k. You will need the fact that $\Delta(t) \leq D(t) \leq 2\Delta(t)$, which follows from the triangle inequality.]

5. [Extra Credit: You may use points obtained on this problem to make up for points you drop on other problems. You can get full credit on the HW without doing this problem.]

Here is an unusual card trick. I take a shuffled deck and turn up the cards one by one. I ask you to select one of the first ten cards, without telling me which one; let $c_1 \in \{1, 2, \ldots, 13\}$ be the numerical value of your card. You then count c_1 cards from the one you selected, and note that card; call its value c_2. You then count a further c_2 cards and note that card, and so on until the deck is exhausted. At that point, I am able to identify the last card you noted (at least most of the time).

Describe how I perform this amazing feat, and give a qualitative explanation for why it works. [HINT: think about coupling. You are not expected to perform any calculations to justify why the method works. You are encouraged to try it on a friend a few times and estimate the success probability—it should certainly be enough to win comfortably in a gambling situation. Or if you are really interested you could simulate the trick with a program and get a much better estimate of the success probability.]