# CMPSCI 690RA SPRING 2020: HOMEWORK 2
# DUE 8PM, FRIDAY 3RD APRIL

- Homework may be completed in group of size with at most three students. You're not allowed to use material from the web or talk about the homework with anybody outside your collaboration group (aside from the lecturer or TA.)
- Solutions should be typed and uploaded as a pdf to `gradescope.com` (instructions to follow).
- To get full marks, answers must be sufficiently detailed, supported with rigorous proofs (for correctness and running time). Faster algorithms will typically get more marks than slower algorithms.

**Question 1.** Given a graph $G(V, E)$ on $n$ vertices, consider the following approach to generating an independent set: pick a random permutation $\pi$ of $[n]$ and define $S \subseteq V$ as follows:

$$S_\pi = \{v_{\pi(i)} : \forall j < i, (v_{\pi(i)}, v_{\pi(j)}) \notin E\}$$

(1) Prove that $S_\pi$ is an independent set.
(2) Compute $\mathbb{E}[|S_\pi|]$ and deduce that there exists an independent set of size $\sum_i 1/(d_i + 1)$ where $d_i$ is the degree of the $i$th node.

**Question 2.** A family of subsets $\mathcal{F}$ of $\{1, 2, \ldots, n\}$ is called an anti-chain if there is a no pair of sets $A$ and $B$ in $\mathcal{F}$ satisfying $A \subset B$.

(1) Give an example of $\mathcal{F}$ where $|\mathcal{F}| = \binom{n}{\lfloor n/2 \rfloor}$.
(2) Let $f_k$ be the number of sets in $\mathcal{F}$ with size $k$. Show that

$$\sum_{k=0}^{n} \frac{f_k}{\binom{n}{k}} \leq 1 \ .$$

Hint: Choose a random permutation of $\{1, 2, \ldots, n\}$ and let $X_k = 1$ if the first $k$ numbers in the permutation yield a set in $\mathcal{F}$. If $X = \sum_{k=0}^{n} X_k$, what can you say about $X$?

(3) Show that $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

**Question 3.** Let $G$ be a connected, non-bipartite graph with $m$ edges and $n$ nodes. Suppose you are currently at node $u$ and your friend is currently at node $v$. At each time step you both move to a random neighbor of your current node. Let $T$ be the random variable that equals the number of time steps until you are both at the same node at the same time. Prove that $\mathbb{E}[T] = O(m^2 n)$.

**Question 4.** In this problem we consider the problem of estimating how many solutions there are to the *knapsack problem*: Given items with sizes $0 < a_1 \leq a_2 \leq \ldots \leq a_n$ and an integer $b$, we want to design an efficient algorithm that estimates, up to a factor $1 + \epsilon$ with probability at least $1 - \delta$, the number of vectors $(x_1, \ldots, x_n) \in \{0, 1\}^n$ that satisfy $\sum_i a_i x_i \leq b$.

(1) An inefficient approach would be to sample vectors in $\{0, 1\}^n$ uniformly at random and estimate the fraction of these that satisfied $\sum_i a_i x_i \leq b$. How many samples would be required if $a_1 = \ldots = a_n = 1$ and $b = \sqrt{n}$.
(2) Consider a Markov chain $X_0, X_1, X_2 \ldots$, where the states are vectors in $\{0, 1\}^n$. The transitions are defined as follows: if the current state is $(x_1, x_2, \ldots, x_n)$ then:
    (a) Pick random $i \in [n]$
    (b) If $x_i = 1$, then set $x_i = 0$.
    (c) If $x_i = 0$, then set $x_i = 1$ if doing so still ensures $\sum_i a_i x_i \leq b$.

(d) Otherwise, remain in the current state.

Prove that the steady state of this Markov chain is uniform on the assumption $\sum_i a_i > b$.

(3) By running the Markov chain for a sufficiently long time we can generate samples (nearly) uniformly from $\Omega(b') = \{x \in \{0,1\}^n : \sum_i a_i x_i \le b'\}$ for any $b' < b$. You do not need to compute how long you would need to run the Markov chain to generate these samples. Instead, suppose you can generate samples uniformly from $\Omega(b')$ for any $b' < b$. How many samples do you need to estimate $|\Omega(b)|$ up to a factor $1+\epsilon$ with probability at least $1-\delta$.
**Hint:** Note that

$$|\Omega(b)| = \frac{|\Omega(b)|}{|\Omega(b_{k-1})|} \times \frac{|\Omega(b_{k-1})|}{|\Omega(b_{k-2})|} \times \ldots \times \frac{|\Omega(b_1)|}{|\Omega(b_0)|} \times |\Omega(b_0)|$$

where $b_0 = 0$, $b_i = \sum_{j=1}^{i} a_j$, and $k$ is the smallest integer such that $b_k \ge b$. You may want to consider estimating terms of the form $|\Omega(b_i)|/|\Omega(b_{i+1})|$ and will need to find a lower bound for such terms.

**Question 5.** Consider the random walk on a non-bipartite, connected graph on $n$ vertices, where each vertex has the same degree $d > n/2$. Show that

$$\tau(\epsilon) \le \frac{\ln \epsilon}{\ln(1-(2d-n)/d)} \ ,$$

where $\tau(\epsilon)$ is the mixing time of the chain. **Hint:** Find a coupling of the Markov chain corresponding to the random walk where at each step there is a probability of at least $(2d-n)/d$ that the components of the coupling become equal at each step, and stay equal if they are already equal. Then appeal to the Coupling Lemma.

**Question 6.** Let $P$ be the transition matrix for a finite, irreducible, aperiodic Markov chain. Let $m_j = \min_i(P_{i,j})$ be the smallest entry in the $j$th column of the matrix, i.e., no matter what state you are in, the probability you move to state $j$ at the next step is at least $m_j$. If $m = \sum_j m_j$, prove that for any $t, x$, the variational distance between $p_x^t$ and the stationary distribution is at most $(1-m)^t$ where $p_x^t$ be the distribution of the Markov chain after $t$ steps if the Markov chain started in state $x$. **Hint:** Find a coupling of the Markov chain corresponding to $P$ and then appeal to the Coupling Lemma.