15-451 Algorithms, Spring 2019
Practice Problems

Weighted Multiplicative Spanners. We saw a greedy algorithm for finding a multiplicative spanner of an unweighted graph in lecture. Recall a k-multiplicative spanner $H = (V, E')$ of a given unweighted graph $G = (V, E)$ on n nodes, is a subgraph (so $E' \subseteq E$) for which for all pairs u, v of vertices in V, $d_G(u, v) \leq d_H(u, v) \leq k \cdot d_G(u, v)$. In this problem we will find a multiplicative spanner $H = (V, E')$ in a *weighted* graph $G = (V, E)$ on n nodes, where each edge $e \in E$ has a positive edge weight w_e. Consider the following algorithm:

1. Initialize E' to $\emptyset$
2. Let $E = \{e_1 = \{u_1, v_1\}, e_2 = \{u_2, v_2\}, \ldots, e_m = \{u_m, v_m\}\}$ be such that
$$w_{e_1} \leq w_{e_2} \leq w_{e_3} \leq \cdots \leq w_{e_m}.$$
3. For $i = 1, 2, \ldots, m$,
   (a) If the distance between u_i and v_i in $H = (V, E')$ is more than $k \cdot w_e$, then add the edge e_i to E', otherwise discard the edge.
4. Output $H = (V, E')$.

1. Argue that H is a k-multiplicative spanner.

2. Argue that for any choices of the weights w_e, the girth (minimum cycle length) of H is at least $k + 2$.

3. What is an upper bound on the number of edges in H?

The Variance of CountSketch. Recall in lecture we introduced the COUNTSKETCH, which is a random linear map S from $\mathbb{R}^n$ to $\mathbb{R}^k$, for $k = \Theta(1/\epsilon^2)$, defined as follows. Let $h : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, k\}$ be a 2-wise independent hash function, and $\sigma : \{1, 2, \ldots, n\} \to \{-1, 1\}$ be a 4-wise independent hash function. Then for $i = 1, 2, \ldots, k$, we have $(Sx)_i = \sum_{j \text{ s.t. } h(j)=i} \sigma(j)x_j$, where x is the n-dimensional input vector.

In lecture, we showed $\mathbf{E}[\|Sx\|^2] = \|x\|^2$, and claimed that $\mathbf{Var}[\|Sx\|^2] = O(\|x\|^4/k)$. We saw that these statements, by Chebyshev's inequality, imply $\mathbf{Pr}[|\|Sx\|^2 - \|x\|^2| > \epsilon\|x\|^2] \leq \frac{1}{10}$.

Prove that $\mathbf{Var}[\|Sx\|^2] \leq \frac{2}{k}\|x\|_2^4$, where $\|x\|_2^2 = \sum_{i=1}^n x_i^2$.

(1)

Locality Sensitive Hashing (LSH) for Jaccard Similarity In lecture we looked at LSH for Hamming distance on the Hamming cube. Here we look at the Jaccard measure: choose a random permutation π on the universe U. For a set $S \subseteq U$, the LSH for Jaccard measure is simply $h(S)$ =First element in S according to permutation π. Consider two sets S_1 and S_2. The Jaccard measure between them is $J(S_1, S_2) = |S_1 \cap S_2|/|S_1 \cup S_2|$.

1. Argue that $\mathbf{Pr}[h(S_1) = h(S_2)] = J(S_1, S_2)$.

   Suppose we define distance as $D(S_1, S_2) = 1 - J(S_1, S_2)$.

2. Show that for any $r > 0$, if $D(S_1, S_2) < r$, then $\mathbf{Pr}[h(S_1) = h(S_2)] \geq 1 - r$.

3. Show that for any $r > 0$ and $c > 1$, if $D(S_1, S_2) \geq cr$, then $\mathbf{Pr}[h(S_1) = h(S_2) \leq 1 - cr$.

4. What is the expected query time and the space if you have n sets, as a function of c?