Homework #6

Due: 2025-1-9 23:59 | 3 Problems, 50 Pts

Name: XXX, ID: XXX

Problem 1 (14'). Consider the following algorithm to estimate the frequency of any number in data streams. The numbers in the data stream are in $[n] := \{1, 2, \cdots, n\}$.

Algorithm 1: Estimate the frequency of numbers in data streams

$\boldsymbol{C} \leftarrow 0$ ▷ $\boldsymbol{C}$ is a $t \times k$-dimension matrix

Choose t independent hash functions $h_1, \cdots, h_t : [n] \to [k]$ from a 2-universal hash family

for *j in data streams* **do**

  for $i = 1, 2, \cdots, t$ **do**

    $\boldsymbol{C}[i][h_i(j)] \leftarrow \boldsymbol{C}[i][h_i(j)] + 1$

  end

end

for *a in queries* **do**

  output $\hat{f}(a) = \min_{1 \le i \le t} \boldsymbol{C}[i][h_i(a)]$

end

For any element a, suppose the real frequency of a is $f(a)$. When $k = \left\lceil \frac{2}{\epsilon} \right\rceil$, $t = \left\lceil \log_2 \frac{1}{\delta} \right\rceil$, prove that for any given a, with probability at least $1 - \delta$, $f(a) \le \hat{f}(a) \le f(a) + \epsilon L$, where L is the length of the data stream. ◀

Problem 2 (18'). A bipartite graph is a graph whose vertices can be divided into two disjoint and independent sets U and V such that every edge connects a vertex in U to one in V. Find out and prove the threshold for $\mathcal{G}(n, p)$ to be bipartite.

[Hint: The definition of bipartite graph is equivalent to a graph that does not contain any odd-length cycles.] ◀

Problem 3 (18'). A vertex is called an isolated vertex if it does not have any edges. Prove that, the threshold for $\mathcal{G}(n, p)$ of the existence of isolated vertex is $p = \frac{\ln n}{n}$. ◀