Sublinear Time and Space Algorithms 2024A – Problem Set 3

Robert Krauthgamer

Due: February 7, 2024

General instructions: Please keep your answers short and easy to read. You can use results, calculations or notation seen in class without repeating them, unless asked explicitly to redo them.

1. Let $y \in \mathbb{R}^n$ be the frequency vector of an input stream in the turnstile model (i.e., allowing insertions and deletions), and suppose its coordinates are integers in the range $[-n^2, n^2]$.

   Design a linear sketch that detects whether $|\operatorname{supp}(y)| = 1$ using storage requirement of $O(1)$ words (i.e., $O(\log n)$ bits), not counting storage of the algorithm's random coins. Its success probability should be at least $1 - 1/n$.

   Hint: Use a variant of the AMS sketch with large random coefficients.

2. In the Steiner Forest problem on a graph $G = (V, E)$, given k vertex-pairs $(u_1, v_1), \ldots, (u_k, v_k)$, the goal is to find a subset of the edges $E' \subset E$ of minimum size, such that in the subgraph $G' = (V, E')$, each u_i is connected to its respective v_i. Notice that an optimal G' can have between 1 and k connected components.

   Design a streaming algorithm for the following restricted setting: The graph G is known in advance and fixed to be a complete binary tree $T = (V, E)$ with n leaves, hence $|V| = O(n)$, and the input stream contains k vertex-pairs $(u_1, v_1), \ldots, (u_k, v_k)$. The algorithm should $(1+\epsilon)$-approximate the optimal size $|E'|$ (no need to report E' itself), with storage requirement of $(\epsilon^{-1} \log n)^{O(1)}$ bits.

   Hint: Since T is a tree, there is a unique path connecting u_i with v_i.

   Remark: As done in class, do not count storage of the algorithm's random coins.

3. An unweighted graph G is called *k-connected* if every cut $(S, \bar{S})$ contains at least k edges.

   Design a streaming algorithm that determines whether a dynamic graph G on vertex set $V = [n]$ (i.e., a stream of edge insertions and deletions) is 2-connected, using storage $\tilde{O}(n)$.

   Hint: First verify that G is connected by constructing a spanning tree T. Then classify all possible cuts $(S, \bar{S})$ into those that contain two or more edges of the tree T and the rest, and finally use additional (independent) samples to verify whatever is still needed.