Sublinear Time and Space Algorithms 2020B – Problem Set 5

Robert Krauthgamer

Due: July 20, 2020

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. Design a 1-coreset (i.e., exact and not approximate) for Minimum Enclosing Ball in $\mathbb{R}^d$ under ℓ_∞ norm, i.e., the cost function is

   $$C_P^\infty(x) = \max_{p \in P} \|p - x\|_\infty.$$

   What is the size of your coreset (as a function of d)? Does this cost function satisfy the Merge and Reduce properties? And the Disjoint Union property?

2. 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.

3. Analyze Algorithm 2 below for counting triangles in a graph given as a stream, and show that with constant high probability, the additive error is $|\tilde{T} - T| \le \varepsilon T$. Can this algorithm be applied also for dynamic graphs (i.e., a stream of edge insertions and deletions)? Explain how/why not.

   Notation (similar to class): Assume $t > 0$ is a known lower bound for the actual number of triangles T, and let x_S count the number of edges internal to the vertices $S \subset V$.

   Algorithm 2

   1. Init: pick $k = O(\frac{n^3}{\varepsilon^2 t})$ random subsets $S_1, \ldots, S_k \subset V$ each of size 3 (with replacement)
   2. Update: maintain $x_{S_1}, \ldots, x_{S_k}$ (explicitly)
   3. Output: compute $z = \sum_{i \in [k]} \mathbb{1}_{\{x_{S_i} = 3\}}$ and $N = \binom{n}{3}$, and report $\tilde{T} = \frac{N}{k} \cdot z$

   Hint: Use Chebyshev's inequality.

Extra credit:

4. Show how to improve Algorithm 2 above by choosing the random sets S_i only from those sets S that satisfy $x_S \geq 1$. The resulting **Algorithm 2'** should have space requirement $k' = O(\frac{mn}{\varepsilon^2 t})$ words, and work also for dynamic graphs.

   Hint: Use ℓ_0-samplers and an estimator for $N' = \|x\|_0$.