Randomized Algorithms 2015A – Problem Set 2

Robert Krauthgamer and Moni Naor

Due: December 30, 2014

1. Analyze the construction below of a stretch 3 distance oracle for a graph G, and show that its storage requirement almost matches that shown in class. (It is not really a distance oracle because it's query time is not fast enough.) Analyze also the accuracy (stretch bound). Explain whether your bounds (storage and accuracy) hold in the worst-case, in expectation, or with high probability.

   Preprocess(G): Choose $L \subset V$ as a random set of $l = \sqrt{n}\,\mathrm{polylog}\, n$ "landmark" vertices (with or without repetitions). For every vertex $v \in V$, store its distances (i) to the $\sqrt{n}$ vertices closest to it (denoted $B_v \subset V$); and (ii) to all the landmark vertices.

   Query(u,v): If $u \in B_v$, i.e., u is among the $\sqrt{n}$ closest to v, report the distance. Otherwise, report $\min_{w \in L}[d(u,w) + d(w,v)]$.

   Hint: in the "otherwise" case, show that $L \cap B_v \neq \emptyset$.

2. Let B be a randomized algorithm that approximates some function $f(x)$ as follows:

   $$\forall x, \quad \Pr\Big[B(x) \in (1 \pm \varepsilon) f(x)\Big] \geq 2/3.$$

   Let algorithm C output the median of $O(\log \frac{1}{\delta})$ independent executions of algorithm B on the same input. Prove that

   $$\forall x, \quad \Pr\Big[C(x) \in (1 \pm \varepsilon) f(x)\Big] \geq 1 - \delta.$$

3. Design a streaming algorithm for the ℓ_1-point query problem, i.e., producing an estimate $\tilde{x}_i \in x_i \pm \varepsilon \|x\|_1$.

   For simplicity, ignore the issue of storing the random bits.

   Hint: Show a linear sketch by extending the count-min sketch seen in class (so as to remove the restriction $x_i \geq 0$).

Extra credit:

4. Let φ be an arbitrary 2-SAT formula on n variables $x_1, \ldots, x_n$. Assume that every clause c has weight $w_c \geq 0$, and the total weight is $\sum_c w_c = 1$ (by normalization). A 2-SAT formula φ' will be called a *sparsifier* of φ if it contains a subset of the clauses of φ, with arbitrary new weights w'_c.

Show that φ admits a sparsifier φ' with $O(n/\varepsilon^2)$ clauses, such that for *every* truth assignment A to the n variables, the value of φ (i.e., total weight of clauses satisfied by A) differs from that of φ' by at most ε (additively).

Hint: Sample exactly $t = O(n/\varepsilon^2)$ clauses from φ with repetitions, and give each of them weight $1/t$, and analyze any fixed truth assignment using Hoeffding's inequality (not Chernoff).