Sub-linear Algorithms	December 14, 2015
	Homework 4
Lecturer: Ronitt Rubinfeld	*Due Date: December 28, 2015*

Turn in your solution to each *problem on a separate sheet of paper, with your name on each one.*

1. A *vertex cover* V' of a set of edges E' is a set of nodes such that every edge of E' is adjacent to one of the nodes in V'.

   For graph $G = (V, E)$, let the *transitive closure graph* $TC(G)$ be the graph $G^{(tc)}(V, E^{(tc)})$ where $(u, v) \in E^{(tc)}$ if there is a directed path from u to v in G.

   Let $f : V \to \{0, 1\}$ be a labeling of the vertices of a known directed acyclic graph G by 0 and 1. For any pair of nodes x and y, we say that $x \leq_G y$ if there is a path from x to y in G. We say that f is *monotone* if for all $x \leq_G y$, $f(x) \leq f(y)$. The *minimum distance of f to monotone* is the minimum number of nodes that must be relabeled in order to turn f into a monotone function.

   Let E' be the set of violating edges in $TC(G)$ according to f. Show that the minimum distance of f to monotone is equal to the minimum size of a vertex cover of E'.

2. This problem is about testing monotonicity of functions defined over a directed graph G. The function maps nodes into binary values (i.e., $f : V \to \{0, 1\}$), and we say that it is *monotone* if for all directed edges (u, v), we have that $f(u) \leq f(v)$. We say that f is ϵ-close to monotone if there is a monotone function g such that g and f differ on at most $\epsilon|V|$ entries.

   (a) Let $V = \{v_1, \ldots, v_n\}$. For each directed graph $G = (V, E)$, let $B_G = (V', E')$ be the bipartite graph where $V' = \{v_1, \ldots, v_n\} \bigcup \{v'_1, \ldots, v'_n\}$, and $(v_i, v'_j) \in E'$ iff v_j is reachable from v_i in G.
   Show that a q-query testing algorithm for B_G with distance parameter $\epsilon/2$ yields a q-query testing algorithm for G with distance parameter ϵ.

   (b) Let f be a function on V which is ϵ-far from monotone over graph G. Then $TC(G)$ has a matching of violated edges of size at least $(\epsilon/2)|V|$. (Recall previous problem).

   (c) Show that if f is a function over bipartite graph G, there is a test for monotonicity of f with query complexity at most $O(\sqrt{|V|/\epsilon})$.

3. In the Run Length Encoding (RLE) compression scheme, the data is encoded as follows: each run, or a sequence of consecutive occurrences of the same character, is stored as a pair containing the character in the first location and the length of the run in the second location.[1] For example, the string 11111101000 would be stored as $(1, 6)(0, 1)(1, 1)(0, 3)$. The *cost of the run-length encoding*, denoted by $C(w)$, is the sum over all runs of $\log(\ell + 1) + \log|\Sigma|$.

   Assume that the alphabet characters are all in the set $\{0, 1\}$, i.e., that the alphabet Σ is of size 2.

[1]Run-length encoding is used to compress black and white images, faxes, and other simple graphic images, such as icons and line drawings, which usually contain many long runs.

(a) Give an algorithm that, given a parameter ϵ, outputs an ϵn-additive estimate[2] to $C(w)$ with high probability and makes $\text{poly}(1/\epsilon, \log n)$ queries.

(b) Show that there is a distribution on inputs such any that any deterministic approximation algorithm for $C(w)$ making an expected number of queries that is $o(\frac{n}{A^2 \log n})$ must fail to output an A-multiplicative approximation with probability at least 1/3. (Here the expectation in the number of queries is over the choice of an input from the distribution). (It's also ok to give a lower bound for deterministic algorithms by showing that for each algorithm there is an input that causes it to fail).

[2]For a function f, algorithm A outputs an ϵn-additive estimate if on any input x, $f(x) - \epsilon n \leq A(x) \leq f(x) + \epsilon n$.