Sublinear time algorithms	November 30, 2009
Homework 2	
Lecturer: Ronitt Rubinfeld	*Due Date: December 14, 2009*

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

1. In Alon's proof of the lower bound for testing the property of triangle-freeness, show that the number of edge-disjoint triangles is in the s-blow up of the graph is at least $\Omega(m|X|s^2)$ (see lecture notes for definitions of all variables).

2. (Canonical forms for graph property testers for the adjacency matrix model). Define a graph property to be a property that is preserved under graph isomorphism – i.e., if G has the property and G' is isomorphic to G, then G' must also have the property. Prove the theorem cited in class due to Goldreich and Trevisan, namely:

   Theorem 1 *Let P be any graph property in the adjacency matrix model. Let T be a property tester for P that on input G, ϵ, where G is a graph on n nodes, has query complexity $q(n, \epsilon)$. Then there is a (nonadaptive) tester T' that selects a random subset of $2q(n, \epsilon)$ nodes in G, queries the edges between every pair of the subset and decides whether to pass or fail based on the query answers and its own internal coin tosses. Furthermore, if T has 1-sided error, then so does T'.*

   It is interesting to make sure you can first solve the following, even though the solution will look nothing like your eventual solution (and don't bother to turn it in). Show that in the adjacency matrix model, any algorithm making q queries can be made into a *nonadaptive* (i.e., where the queries do not depend on the results of any previous queries) tester that uses only 2^q queries.

3. (bounds for testing triangle-freeness in the sparse graph model) Assume that an input graph of average degree d is represented via an adjacency list (note that there is no bound on the max degree, only the average degree). We say that graph G is ϵ-far from being triangle free, if at least ϵdn edges must be removed in order to make it triangle free. Assume that the property tester can make any of the following types of queries in one step:

   - *Degree queries:* For any vertex u, the algorithm can obtain the degree of u.
   - *Neighbor queries:* For any vertex u and index $1 \le i \le deg(u)$, the algorithm may obtain the ith neighbor of u.
   - *Vertex-pair queries:* For any pair of vertices (u, v), the algorithm can query whether there is an edge between u and v.

   (a) Show that any algorithm that tests triangle-freeness must perform $\Omega(\sqrt{n/d})$ queries. (Showing for two-sided error gets a check-plus!)

   (b) Assume that the maximum degree is $O(d)$. Show that one can test triangle-freeness using $O(d/\epsilon)$ queries.

4. (Property testing of the clusterability of a set of points.) Given a set X of points in any metric space. Assume that one can compute the distance between any pair of points in one step. Say that X is (k, b)-diameter clusterable if X can be partitioned into k subsets (clusters) such that the maximum distance between any pair of points in a cluster is b. Say that X is ϵ-far from (k, b)-diameter clusterable if at least $\epsilon|X|$ points must be deleted from X in order to make it (k, b)-diameter clusterable.

   Show how to distinguish the case when X is (k, b)-diameter clusterable from the case when X is ϵ-far from $(k, 2b)$-diameter clusterable. Your algorithm should run in time polynomial in $k, 1/\epsilon$. It is possible to get an algorithm which runs in time $O((k^2 \log k)/\epsilon)$.

   Note that the result presented in class works only for a specific metric space, whereas this algorithm should work in general.