COS 521: Advanced Algorithm Design
Final Exam Fall 2006
Due: 48 hours after first reading,
no later than midnight on Monday, Jan 22

Instructions: Finish the test within **48 hours** after first reading it. You can consult any notes/handouts from the class and feel free to quote, without proof, any results from there. You are not allowed to consult any other source in any way.

DO NOT READ THE TEST BEFORE YOU ARE READY TO WORK ON IT.

Send email to Moses to let him know you have started the test (in case any clarifications need to be made).

Write and sign the honor code pledge on your exam. (The pledge is "I pledge my honor that have not violated the honor code during this examination.")

1. (10 points) Determine whether each of these statements is true or false and provide brief justification for your answer.

   (a) Suppose we toss a fair coin n times. Then with high probability, we will see a contiguous sequence of at least $\log_2 n - 2\log_2 \log_2 n$ heads.

   (b) Consider a bipartite graph $G(U, V, E)$ such that $|U| = |V|$. If every vertex has degree at least 4, then the graph has a perfect matching.

2. (15 points) Give brief answers to the following two questions.

   (a) Consider the following fault-tolerant network design problem: The goal is to design a graph on a given set of nodes by selecting a subset of possible edges such that every cut has at least 3 edges crossing it. For every pair (i, j) we are given the cost c_{ij} for installing the edge (i, j); the goal is to pick a minimum cost set of edges satisfying the connectivity requirements.

   Consider the following LP relaxation for the problem with variables x_{ij} for every unordered pair $\{i, j\}$ of vertices.

$$\begin{array}{rc} & \min \sum_{i,j} c_{ij} x_{ij} \\ \forall S \subset V, S \neq \emptyset & \sum_{(i,j) \in (S, \bar{S})} x_{ij} \geq 3 \\ \forall i, j \in V & 0 \leq x_{ij} \leq 1 \end{array}$$

   Explain how this linear program can be solved in polynomial time.

   (b) Given a graph $G(V, E)$, a dominating set is a subset of vertices $D \subseteq V$ such that for all $i \in V$, either $i \in D$ or there exists some $j \in D$ such that $(i, j) \in E$. Consider the problem of finding a dominating set of minimum size. Describe an $O(\log n)$ approximation for this problem.

3. (15 points) Suppose you are standing in a long linear parking lot and have forgotten where you parked your car. You decide to put your newly acquired skills from COS 521 to use in designing a sytematic strategy to search for your car. Let us model the problem as follows: Imagine that you start out at the origin of the real line. The car is located at some integer position on the line. Your goal is to design a strategy to traverse the real line in search of your car. Think of this as an online problem where you stop as soon as you reach the location where your car is parked. Assume that you have no prior information about the location of your car. Design a deterministic algorithm with constant competitive ratio for this problem.

4. (15 points) Given an undirected graph $G(V, E)$, consider the problem of directing every edge in G so as to minimize the maximum indegree in the resulting directed graph. Let $\delta^*(G)$ denote the optimum value for this problem.

(a) Write down an LP relaxation for this problem and its dual.

(b) Let $d^*(G) = \max_{S \subseteq V} \frac{|E(S)|}{|S|}$. Here $E(S)$ is the set of edges in the subgraph induced by S. Prove that $\delta^*(G) \geq d^*(G)$.

(c) Consider the following greedy heuristic for approximating $\delta^*(G)$: Let v be the minimum degree vertex in G (break ties arbitrarily). Orient all edges incident on v towards v. Remove v and all incident edges from G. Repeat on the remaining graph until the graph is empty. Let $\delta'(G)$ be the maximum indegree in the directed graph obtained by the edge orientations produced by this greedy algorithm. Show that $\delta'(G) \leq 2\delta^*(G)$.

5. (15 points) Given a 3-colorable graph $G(V, E)$, $|V| = n$, give a polynomial time algorithm to find a subset $S \subseteq V$, $|S| \geq n/2$ such that the induced subgraph on S does not contain a triangle. (A triangle is a complete graph on three vertices).

6. (15 points) Consider the following algorithm for finding a minimum cut in an undirected unweighted graph $G(V, E)$. First pick a random weight $w(e)$ independently for every edge e. Then compute a minimum spanning tree for graph G with these weights. Now delete the maximum weight edge in the spanning tree. This divides G into two parts. Show that this procedure produces a min cut in G with probability $\Omega(1/n^2)$.

7. (15 points) Google would like to design an efficient algorithm to measure the correlation between their query streams for two different days. The problem is modeled as follows: Suppose that there are n distinct queries. You can think of the query stream as a sequence of numbers $\in \{1, \ldots, n\}$. Let a_i be the total number of occurences of query i on day 1 and b_i be the total number of occurences of query i on day 2. The correlation between the query streams is defined to be

$$\frac{\sum_i a_i b_i}{\sqrt{(\sum_i a_i^2)(\sum_i b_i^2)}}$$

Your job is to design an efficient algorithm to estimate the correlation. Your algorithm should produce a compact sketch of the query stream for each day in one pass using a small amount of space. The correlation for two given days should be estimated by an appropriately designed function of the sketches for the two days. You may assume that the correlation is at least α, a parameter given to you. Describe an algorithm to estimate the correlation within factor $1 + \epsilon$ with probability $1 - \delta$. Analyze the size of your sketches as a function of n,α,ϵ and δ.