Problem Set Two

Due date: Tuesday, 19 February 2008

1. **(25pt)** The following problem is a variant of the set cover problem: Given a ground set V and subsets $S_1, \ldots, S_m$ and a parameter k, find k sets $S_{i_1}, \ldots, S_{i_k}$ such that the size of their union is maximized. Show that the greedy algorithm which stops after picking k sets achieves a constant approximation ratio (In fact the constant is $\frac{e-1}{e} \simeq 0.632$, but any constant factor will receive full credit.). Using this claim, give an alternate proof that the greedy algorithm gives a factor of $O(\log n)$ for the original set cover problem. [Hint: Show that at every stage greedy picks a set which covers at least $1/k$ fraction of the uncovered elements.]

2. **(30pt)** Consider the following variant of the wire routing problem. We are given a set of pairs s_i, t_i and each pair has an associated profit 1. If our solution has a wire from s_i to t_i then we gain profit 1. Our goal is to place wires without exceeding the channel capacities such that we connect *some* of the s_i, t_i pairs and maximize the total profit we obtain.

   (a) Write an integer program similar to the one given in class for wire routing to describe the above problem. Assume you have a black box that can solve the LP in polynomial time.

   (b) Give a randomized rounding scheme that with probability greater than some constant fraction achieves a valid solution (i.e. does not overflow capacities) that has profit at least a constant fraction of the optimum profit. You may assume that all capacities are larger than $c \ln m$ where c is a large enough constant. You may also assume that the optimum profit is larger than a large enough constant $\bar{c}$ and the number of edges is larger than some large enough constant $\hat{c}$. (When we state "large enough", we mean that you can choose it.)

   You might want to use the following variants of the Chernoff Bound. Given n independent random variables $X_1, X_2, \ldots, X_n$, where $X_i = 1$ with probability p_i and zero otherwise, and $\mu =$

$\sum_i p_i$, we have:

$$\Pr[\sum_i X_i > (1+\delta)\mu] < \exp(\frac{-\mu\delta^2}{4})$$

where $2e - 1 > \delta$. When $2e - 1 \leq \delta$,

$$\Pr[\sum_i X_i > (1+\delta)\mu] < 2^{(-\mu\delta)}$$

Similarly,

$$\Pr[\sum_i X_i < (1-\delta)\mu] < \exp(\frac{-\mu\delta^2}{2})$$

where $0 < \delta \leq 1$.

3. **(15pt)** Consider a problem where we have m sets $S_1, \ldots, S_m$, where $S_i \subseteq V = [1, \ldots, n]$. The goal is to choose a set of points $I \subseteq V$ such that for each S_i, we have $S_i \cap I \neq \emptyset$. Give an algorithm that provides an $O(\log m)$ approximation to the smallest such set I.

4. **(15pt)** One of the polynomial max flow algorithms that we will discuss shortly relies on the existence of an efficient way to find a path from s to t which has the largest *bottleneck capacity*, whereby bottleneck capacity we mean the smallest residual capacity on the path. Describe an efficient algorithm to find such a path and prove its correctness. What is the running time of your algorithm? [Hint: start with Dijkstra's shortest path algorithm, and modify appropriately.]

5. **(15pt)** Suppose we are given a graph and an s-t flow. Prove that for any two s-t cuts $(A, V-A)$ and $(B, V-B)$, the flow across these cuts has the same value, (i.e. $f(A, V-A) = f(B, V-B)$).