CO453: Network Design – Winter 2007

Instructor: Chaitanya Swamy

Assignment 4

Due: Monday, March 12, 2007 after class

You must give a proof of correctness of any algorithm you design, and argue briefly why it runs in polynomial time. You may use any proof or algorithm covered in class directly.

Q1: Recall the LP relaxation for the weighted set cover problem.

$$\min \sum_S w_S x_S \quad \text{subject to} \quad \sum_{S: e \in S} x_S \geq 1 \quad \forall e; \qquad x_S \geq 0 \quad \forall S.$$

Consider the following algorithm for the set cover problem. Solve the above LP relaxation to get an optimal solution x^*. Now pick every set S such that $x^*_S > 0$, that is, return the collection of sets $\{S : x^*_S > 0\}$ as the solution. Prove that this is a B-approximation algorithm, where $B = \max_e |\{S : e \in S\}|$ is the maximum number of sets an element lies in. (Remember to also argue that the solution returned is feasible, that is, it is a set cover.) **(10 marks)**

(Hint: Use Complementary Slackness.)

Q2 [Vazirani]:

Given a directed graph $G = (V, E)$ with costs c_v on the *vertices*, the *feedback vertex set* problem on directed graphs is to find a minimum-cost set of vertices whose removal makes the graph acyclic. That is, we want to find a set $V' \subseteq V$ of minimum cost such that the graph $G' = (V \setminus V', E[V \setminus V'])$ is acyclic, where $E[S] = \{(u, v) \in E : u, v \in S\}$. The set V' is called a feedback vertex set.

A *tournament* is a directed graph $G = (V, E)$ where for every pair $u, v \in V$, exactly one of the edges (u, v) or (v, u) is in E. (Think of G representing a tournament between players represented by the vertices, where every pair of players play against each other and each match results in a win for one of the players.) We will design a 3-approximation algorithm for the feedback vertex set problem on tournament graphs. In the following, $G = (V, E)$ will denote a tournament.

(a) Show that V' is a feedback vertex set iff the graph $G' = (V \setminus V', E[V \setminus V'])$ contains no directed triangles (cycles of length 3). **(4 marks)**

(b) Using part (a), formulate the feedback vertex set problem on tournaments as a set cover problem and argue that one of the approximation algorithms for set cover covered in class yields a 3-approximation for this set cover instance. **(6 marks)**

Q3: Consider the following variant of the set cover problem. We are given a universe U of n elements, and a collection $\mathcal{S}$ of subsets over U. We are also given a collection of pairs $\mathcal{P} = \{(S, T) : S, T \in \mathcal{S}\}$ that partition $\mathcal{S}$ (that is, each set $S \in \mathcal{S}$ appears in exactly one pair (S, T) of $\mathcal{P}$) with the property that for every element $e \in U$ and every pair $(S, T) \in \mathcal{P}$ at most one of the sets S and T contains element e. Each *element* $e \in U$ has a weight w_e. The goal is to pick a collection $\mathcal{S}'$ of sets choosing exactly one set from every pair of $\mathcal{P}$, so as to maximize the weight of the covered elements, i.e., $\sum_{e \in \bigcup_{S \in \mathcal{S}'}} w_e$. This problem is *NP*-hard, and we will design various approximation algorithms for the problem. Notice that this is a maximization problem, so an α-approximation algorithm is an algorithm that returns a solution of value *at least* α times the optimum, where α lies in $[0, 1]$. Thus, in order to prove an approximation guarantee, we now need find a good *upper bound* on the optimum value against which one can compare the value of the solution returned by our algorithm.

(a) Consider the following simple algorithm. For every pair $(S, T) \in \mathcal{P}$, pick set S with probability 0.5 and set T with probability 0.5 (the two events are mutually exclusive, i.e., exactly one of S and T is picked). Prove that the *expected* weight of the covered elements is at least $0.5 \sum_{e \in U} w_e$, and hence the algorithm is a *randomized* 0.5-approximation algorithm. What is the upper bound that is being used here? **(6 marks)**

We can give an improved $\left(1-\frac{1}{e}\right)$-approximation algorithm using LP-rounding. Consider the following IP formulation of the problem:

$$
\begin{aligned}
\max \quad & \sum_{e \in U} w_e z_e & & \text{(IP)} \\
\text{s.t.} \quad & \sum_{S \in \mathcal{S}: e \in S} x_S \geq z_e & \forall e \in U, & \quad (1) \\
& x_S + x_T = 1 & \forall (S, T) \in \mathcal{P}, & \quad (2) \\
& x_S, z_e \in \{0, 1\} & \forall S \in \mathcal{S}, e \in U. & \quad (3)
\end{aligned}
$$

Here z_e is a variable that indicates if element e is covered, and x_S indicates if set S is picked. Constraint (1) says that if element e is covered, we must pick a set S that contains e; constraint (2) says that exactly one set S of every pair must be selected. Relaxing the integrality constraints (3) to $0 \leq x_S, z_e \leq 1$ yields an LP relaxation (LP) (note that we need to impose that $z_e \leq 1$).

(b) Let (x^*, z^*) be an optimal LP solution, and let $OPT = \sum_e w_e z_e^*$. We round this to an integer solution $(\hat{x}, \hat{z})$ as follows. For a pair $(S, T) \in \mathcal{P}$, we set exactly one of $\hat{x}_S$ and $\hat{x}_T$ to 1, setting $\hat{x}_S = 1$ with probability x_S^* and $\hat{x}_T = 1$ with probability x_T^*. We do this independently for every pair in $\mathcal{P}$. Set $\hat{z}_e = \min(1, \sum_{S \in \mathcal{S}: e \in \mathcal{S}} \hat{x}_S)$. It is clear that $(\hat{x}, \hat{z})$ is a feasible solution to (IP). Prove for every element e, if e lies in k sets, then $\Pr[\hat{z}_e = 1] = \Pr[\hat{x}_S = 1 \text{ for some } S \in \mathcal{S} \text{ s.t. } e \in S] \geq \left(1 - (1 - \frac{1}{k})^k\right) z_e$. **(6 marks)**

You may use the following inequality: given numbers $y_1, \ldots, y_k \in [0, 1]$, we have

$$1 - (1 - y_1)(1 - y_2) \ldots (1 - y_k) \geq 1 - \left(1 - \frac{\sum_{i=1}^k y_i}{k}\right)^k \geq \left(1 - \left(1 - \tfrac{1}{k}\right)^k\right) \left(\sum_{i=1}^k y_i\right),$$

where the second inequality holds when $\sum_{i=1}^k y_i \leq 1$.

(c) Using part (b) and the fact that $\left(1 - \left(1 - \frac{1}{k}\right)^k\right) \geq \left(1 - \frac{1}{e}\right)$, prove that the expected weight of the covered elements is at least $\left(1 - \frac{1}{e}\right) \cdot OPT$. **(3 marks)**

(d) (Bonus part) One can improve the approximation ratio further to 0.75 by combining the algorithms in parts (a) and (b). First, prove that if an element e is contained in k sets of $\mathcal{S}$, then the probability that e is covered by the algorithm in part (a) is $1 - 2^{-k}$. Now consider the following hybrid algorithm: pick exactly one of the above two algorithms, choosing each exclusively with probability 0.5, and run the chosen algorithm. Prove that in the solution returned by the hybrid-algorithm, the probability that an element e is covered is at least $0.75 z_e^*$, and hence, that the hybrid-algorithm is a 0.75-approximation algorithm. **(7 marks)**

(e) (Bonus part) One can also give a "pure" LP-rounding algorithm that attains an approximation ratio of 0.75. Consider a generalization of the rounding procedure in part (b), where we set $\hat{x}_S$ to 1 with probability $g(x_S^*)$, and $\hat{x}_T = 1$ with probability $g(x_T^*)$ for a pair (S, T), where g is a function such that $g(y) + g(1 - y) = 1$ for all $y \in [0, 1]$. (Thus, the earlier rounding procedure corresponds to the function $g(y) = y$.) Give a function g for which this algorithm returns a solution of expected value at least $0.75 \cdot OPT$. **(8 marks)**

(Hint: Consider linear functions g first, and think about the probability that set S is picked in the hybrid algorithm. The following facts may be useful in the analysis. Given a twice differentiable function $f : \mathbb{R} \mapsto \mathbb{R}$, (i) f is called *convex* if $f''(x) \geq 0$ for all x. (ii) f is convex iff for all $x, y, \lambda \in [0, 1]$, $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y)$, (iii) if h is a linear function and $f(a) \leq h(a)$, $f(b) \leq h(b)$, then $f(x) \leq h(x)$ for all $x \in [a, b]$. f is called *concave* if $-f$ is convex.)