

CO453: Network Design – Winter 2007

Instructor: Chaitanya Swamy

Solutions to Assignment 4

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: The dual to the given LP relaxation for the weighted set cover problem is

$$\max \sum_e y_e \quad \text{subject to} \quad \sum_{e \in S} y_e \leq w_S \quad \forall S; \quad y_e \geq 0 \quad \forall e.$$

The complementary slackness conditions for this pair of primal and dual programs is: $x_S > 0 \implies \sum_{e \in S} y_e = w_S$, and $y_e > 0 \implies \sum_{S: e \in S} x_S = 1$. Let \mathcal{S}' denote the collection $\{S : x_S^* > 0\}$, where x^* is an optimal primal solution. Let y^* be an optimal dual solution. First, note that \mathcal{S}' is a set cover since of every element e , we have $\sum_{S: e \in S} x_S^* \geq 1$ which implies that there is some set S such that $e \in S$ and $x_S^* > 0$ (so $S \in \mathcal{S}'$). The cost of \mathcal{S}' is

$$\sum_{S: x_S^* > 0} w_S = \sum_{S: x_S^* > 0} \left(\sum_{e \in S} y_e^* \right) = \sum_e y_e^* \left(\sum_{S: e \in S} 1 \right) \leq B \sum_e y_e^* \leq B \cdot OPT.$$

The first equality follows from complementary slackness, whereas the penultimate inequality follows from the definition of B . Thus, we have shown that the algorithm is a B -approximation algorithm.

Q2(a) Clearly if $G' = (V \setminus V', E[V \setminus V'])$ is acyclic then it contains no directed triangles. Now suppose that G' has no directed triangles, and assume for a contradiction that G' contains a cycle. Among all such cycles choose the cycle $C = u_0, u_1, \dots, u_{k-1}, u_k = u_0$ with fewest edges (or nodes). Consider the pair u_0, u_2 . Since G is a tournament it contains either the edge (u_0, u_2) or (u_2, u_0) , and G' must also contain this edge since both u_0, u_2 are in G' . If $(u_0, u_2) \in G'$, then G' contains the smaller cycle $u_0, u_2, u_3, \dots, u_k = u_0$, which contradicts the choice of C ; if $(u_2, u_0) \in G'$, then G' contains a directed triangle, which contradicts our assumption. Thus, if G' does not contain any directed triangles, it must be acyclic.

(b) By part (a), we know that to render G' acyclic, it is necessary and sufficient to choose a set V' that contains a node from every directed triangle of G (so that after removing V' , the resulting graph G' contains no directed triangles). This is precisely a set cover problem, where we want to cover all the directed triangles of G using the vertices of G . More precisely, the ground set or universe U is the set of all directed triangles of G , and for every vertex v of G , we have a set S_v of cost c_v consisting of all the directed triangles of G (i.e., elements of U) that contain v . (For example, suppose $V = \{a, b, c, d\}$, $E = \{(a, b), (b, c), (c, a), (b, d), (d, a)\}$. Then G has two triangles $C_1 = \{(a, b), (b, c), (c, a)\}$ and $C_2 = \{(a, b), (b, d), (d, a)\}$. The resulting set-cover instance (U, \mathcal{S}) has $U = \{C_1, C_2\}$, and $\mathcal{S} = \{S_a = \{C_1, C_2\}, S_b = \{C_1, C_2\}, S_c = \{C_1\}, S_d = \{C_2\}\}$.)

Notice that in this set-cover instance, every element of U , which is a directed triangle C , lies in exactly three sets, namely the sets corresponding to the nodes of C . Thus, we can use any of the B -approximation algorithms for set-cover (the primal-dual algorithm, the LP-rounding algorithm

covered in class, or the LP-rounding algorithm of Q1) to obtain a 3-approximation algorithm for feedback vertex set on tournament graphs.

Q3(a): Let X be a random variable denoting the weight of the covered elements. Observe that we can write $X = \sum_{e \in U} w_e X_e$, where X_e is an indicator random variable that is 1 if e is covered and 0 otherwise. Thus, $E[X] = \sum_{e \in U} w_e E[X_e]$, and we will show that $E[X_e] = \Pr[X_e = 1] \geq 0.5$. Suppose that e lies in $k \geq 1$ sets. We know that each of these sets belongs to a distinct pair of \mathcal{P} , so each of these sets is picked *independently* with probability 0.5. Therefore, the probability that none of these sets is picked is 2^{-k} , so $\Pr[X_e = 1] = 1 - 2^{-k} \geq 0.5$. Hence, the expected weight of covered elements, $E[X] \geq 0.5 \sum_{e \in U} w_e$. Since $\sum_{e \in U} w_e$ is a trivial upper bound on the optimum, this shows that the algorithm is a 0.5-approximation algorithm.

(b) Suppose that e lies in the k sets S_1, \dots, S_k . For each set S_i , we set $\hat{x}_{S_i} = 1$ independently with probability $x_{S_i}^*$. Thus,

$$\begin{aligned} \Pr[\hat{z}_e = 1] &= \Pr[\hat{x}_{S_i} = 1 \text{ for some set } S_i, i = 1, \dots, k] \\ &= 1 - (1 - x_{S_1}^*) \dots (1 - x_{S_k}^*) \geq 1 - \left(1 - \frac{\sum_{i=1}^k x_{S_i}^*}{k}\right)^k \geq 1 - \left(1 - \frac{z_e^*}{k}\right)^k \quad (1) \\ &\geq \left(1 - \left(1 - \frac{1}{k}\right)^k\right) z_e^*. \end{aligned}$$

Here the first and last inequalities follow from the inequalities given in the question, and (1) follows since (x^*, z^*) is a feasible solution to the LP, so $\sum_{S:e \in S} x_S^* = \sum_{i=1}^k x_{S_i}^* \geq z_e^*$.

Remarks: The two inequalities given in the question are proved as follows. The first follows from the well-known inequality that the geometric mean of k numbers is at most the arithmetic mean of the k numbers. So $[(1 - y_1)(1 - y_2) \dots (1 - y_k)]^{1/k} \leq \frac{\sum_{i=1}^k (1 - y_i)}{k}$, or equivalently, $\prod_{i=1}^k (1 - y_i) \leq \left(1 - \frac{\sum_{i=1}^k y_i}{k}\right)^k$, and the first inequality follows. To prove the second inequality, consider the function $f(x) = 1 - \left(1 - \frac{x}{k}\right)^k$. This function is *concave* since $f''(x) = -\frac{k-1}{k} \left(1 - \frac{x}{k}\right)^{k-2} \leq 0$. Thus by definition of a concave function, for any $x \in [0, 1]$, we have $f(x) = f((1-x) \cdot 0 + x \cdot 1) \geq (1-x) \cdot f(0) + x \cdot f(1) = \left(1 - \left(1 - \frac{1}{k}\right)^k\right)x$.

(c) Again, let $X = \sum_{e \in U} w_e X_e$ denote the weight of the covered elements, with $X_e = 1$ if element e is covered (i.e., $\sum_{S:e \in S} \hat{x}_S \geq 1$) and 0 otherwise. $\Pr[X_e = 1]$ is simply $\Pr[\hat{z}_e = 1]$, which as shown in part (b) is at least $\left(1 - \left(1 - \frac{1}{k}\right)^k\right) z_e^* \geq \left(1 - \frac{1}{e}\right) z_e^*$. Thus, $E[X] \geq \left(1 - \frac{1}{e}\right) \sum_{e \in U} w_e z_e^* = \left(1 - \frac{1}{e}\right) \cdot OPT$.

(d) In the hybrid algorithm, if element e is covered by k sets, then by parts (a) and (b), we have

$$\Pr[X_e = 1] \geq 0.5(1 - 2^{-k}) + 0.5 \left(1 - \left(1 - \frac{1}{k}\right)^k\right) z_e^* \geq 0.5 \left(1 - 2^{-k} + 1 - \left(1 - \frac{1}{k}\right)^k\right) z_e^*,$$

where the second inequality follows since $z_e^* \leq 1$. Let $c(k) = 0.5(1 - 2^{-k} + 1 - \left(1 - \frac{1}{k}\right)^k)$. If we prove that $c(k) \geq 0.75$, then this would show that the expected weight of covered elements is at least $0.75 \sum_{e \in U} w_e z_e^* \geq 0.75 \cdot OPT$.

For $k = 1, 2$, we have $c(k) = 0.75$. For $k \geq 3$, we have $1 - 2^{-k} \geq \frac{7}{8}$ and $1 - \left(1 - \frac{1}{k}\right)^k \geq \left(1 - \frac{1}{e}\right) \geq 0.632$, so $c(k) \geq 0.75$.

(e) Following the hint, consider the function $g(x) = \frac{1}{4} + \frac{x}{2}$. Observe that in the hybrid algorithm, set S is picked with probability exactly $g(x_S^*)$. Suppose an element e is covered by k sets S_1, \dots, S_k . Proceeding as in part (b), we have

$$\begin{aligned} \Pr[X_e = 1] &= 1 - (1 - g(x_{S_1}^*)) \dots (1 - g(x_{S_k}^*)) \geq 1 - \left(1 - \frac{\sum_{i=1}^k g(x_{S_i}^*)}{k}\right)^k \\ &= 1 - \left(\frac{3}{4} - \frac{\sum_{i=1}^k x_{S_i}^*}{2k}\right)^k \geq 1 - \left(\frac{3}{4} - \frac{z_e^*}{2k}\right)^k \\ &\geq \left(1 - \left(\frac{3}{4} - \frac{1}{2k}\right)^k\right) z_e^*. \end{aligned}$$

The last inequality follows since the function $f(x) = 1 - \left(\frac{3}{4} - \frac{x}{2k}\right)^k$ is concave, as $f''(x) = -\frac{k-1}{4k} \left(\frac{3}{4} - \frac{x}{2k}\right)^{k-2} \leq 0$; hence, as argued in the remarks after part (b), we have $f(x) \geq (1-x) \cdot f(0) + x \cdot f(1) \geq \left(1 - \left(\frac{3}{4} - \frac{1}{2k}\right)^k\right)x$.

As in part (d), defining $c(k) = 1 - \left(\frac{3}{4} - \frac{1}{2k}\right)^k$, it suffices to show that $c(k) \geq 0.75$ for all integer $k \geq 1$. We have $c(1) = c(2) = 0.75$. Calculating $c'(k)$ yields $c'(k) = -\left(\frac{3}{4} - \frac{1}{2k}\right)^k \left[\ln(0.75 - 0.5/k) + \frac{1}{1.5k-1}\right]$. $\ln(0.75 - 0.5/k) \leq \ln(0.75) < -0.287$ for all $k \geq 1$, and $\frac{1}{1.5k-1} \leq \frac{1}{3.5} \leq 0.286$ for all $k \geq 3$. So for $k \geq 3$, we have $c'(k) > 0$. For $k \in [2, 3]$, we have $\ln(0.75 - 0.5/k) < -0.538$ and $\frac{1}{1.5k-1} \leq 0.5$, so again $c'(k) > 0$. Thus, $c'(k) > 0$ for $k \geq 2$, and hence $c(k) \geq c(2) = 0.75$ for all $k \geq 2$. Thus, the expected weight of covered elements is at least $0.75 \cdot OPT$.

Remarks: Although the probability of picking a set S is identical in both the hybrid algorithm and in the LP-rounding algorithm with function $g(\cdot)$ as defined above, there is a subtle distinction between the two algorithms. In the latter, sets S and S' belonging to different pairs are picked *independently* with probabilities $\frac{1}{4} + \frac{x_S^*}{2}$ and $\frac{1}{4} + \frac{x_{S'}^*}{2}$ respectively, whereas in the hybrid algorithm, the two sets are *not* picked independently. To see this, note that in the hybrid algorithm, we have $\Pr[S \text{ and } S' \text{ are picked}] = 0.5 \cdot \frac{1}{4} + 0.5 \cdot x_S^* x_{S'}^*$, which is not equal to $\Pr[S \text{ is picked}] \cdot \Pr[S' \text{ is picked}] = \left(\frac{1}{4} + \frac{x_S^*}{2}\right) \left(\frac{1}{4} + \frac{x_{S'}^*}{2}\right)$.