Spring 2010 **CMSC 858Y: Homework 5** Samir Khuller

Due in class: Complete by Apr 30.

(1) Consider an instance of the Set Cover problem over the set of elements $X = \{x_1, x_2, \ldots x_n\}$. Let $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ be the collection of sets. We are also told that **each** element x_i belongs to *at least* $\frac{m}{2}$ sets in $\mathcal{S}$. Suppose we run the greedy algorithm to find a "small" set-cover (i.e., we pick a set that is covering the largest number of uncovered elements). Prove that the number of sets picked is $O(\log n)$.

(2) Consider the following "prim like" algorithm for the Steiner tree problem. Given a graph $G = (V, E)$ with positive weights on the edges, and a special subset S of the vertices. We wish to obtain a steiner tree that spans all the vertices in S. Let $T_1 = \{s_1\}$. At each step T_{i+1} is computed from T_i as follows: attach the vertex from $S - T_i$ that is the "closest" to T_i by a path to T_i and call the newly added special vertex s_{i+1}. (Thus T_i always contains the vertices $s_1, s_2, \ldots, s_i$.) Prove that this algorithm finds a steiner tree whose cost is at most twice the minimum weight steiner tree.

(3) In class we studied the vertex cover problem in which our goal was to cover *all* the edges using minimum weighted set of vertices. Consider the problem of covering a *fixed* number of the edges instead of all the edges. Let the parameter p which is part of the input denote the number of edges to be covered. The integer program for this problem is given below. The LP can be obtained by relaxing the integrality constraints.

$$\begin{aligned}
\min \quad & \sum_{j=1}^{n} w_j \cdot x_j \\
\text{subject to} \quad & \\
& x_i + x_j + y_{ij} \geq 1, \forall (v_i, v_j) \in E \\
& - \sum_{(v_i, v_j) \in E} y_{ij} \geq -(|E| - p) \\
& x_j \in \{0, 1\}, j = 1, 2, \ldots, n \\
& y_{ij} \in \{0, 1\}, \forall (v_i, v_j) \in E
\end{aligned}$$

Design a primal dual approximation algorithm for this problem with a constant approximation factor.

(4) Suppose you are a restaurant critic. There are n restaurants you need to compare. You sample a few individuals and ask for their input on "comparing" pairs of restaurants. If a person says that j is better than i, you add a directed edge from i to j. You now obtain a graph. Clearly, this graph may have cycles since different people have different recommendations. Your goal is to produce a DAG (Directed Acyclic Graph) that provides a consistent set of recommendations (or a partial order among the restaurants). How would you find a DAG that maximizes the number of edges in it? This solution might be one that represents the collective opinions in the best way. Develop a polynomial time algorithm to give a $\frac{1}{2}$ approximation.

(5) Give an example showing that there is a "gap" between the optimal integral solution for the Facility location IP, and the corresponding LP relaxation. Prove that this gap is some constant $c > 1$. In other words, argue that the optimal integral solution has cost more than the fractional solution by a factor of at least $c > 1$.