Approximation Algorithms **Homework 5**
Date: Nov 2, 2005 **Due:** Nov 9, 2005

Note: In all these problems: if we ask you to show an approximation ratio of ρ, you should tell us about the best approximation ratio you did prove, be it better or worse than ρ. Also, please refrain from consulting any external sources (papers, lecture notes, books, surveys) in attempting to solve these problems.

1. Write an ILP formulation for the following problems. Try to be as succinct as possible, i.e., use as few variables and constraints as you possibly can. Write out in $O(\cdot)$-notation, the number of variables and constraints in your solutions (E.g., the formulation for Vertex Cover we saw in class had $O(|V|)$ variables and $O(|E| + |V|)$ constraints in a graph $G = (V, E)$.

   - **Maximum s-t flow:** Given an undirected graph $G = (V, E)$ with a source node s and sink node t, and nonnegative capacities $c : E \to \mathbb{R}_+$ on the edges, find a "flow" of maximum value from s to t. Informally, an $s - t$ flow is a collection of paths from s to t such that the number of paths using any edge e is at most the edge's capacity. Formally, a flow is a function that emits a value at the source s and absorbs it at t while keeping it conserved at all other nodes (think of Kirchoff's law for electricity).
   - **Minimum s-t cut:** Given an undirected graph $G = (V, E)$ with a source node s and sink node t, and nonnegative capacities $c : E \to \mathbb{R}_+$ on the edges, find a "cut" of minimum total edge capacity separating s and t. A cut is a subset of edges whose deletion disconnects s from t. Without loss of generality, it can be defined by a subset $S \subset V$ of the nodes containing s (i.e., $s \in S$) where the cut is the set of edges with exactly one endpoint in S.
   - **Minimum s-t path:** Given an undirected graph $G = (V, E)$ with a source node s and sink node t, and nonnegative distances $d : E \to \mathbb{R}_+$ on the edges, find a path of minimum total distance connecting s and t.

2. Recall the K-median problem we defined earlier; we give a definition for the non-metric version here. Given an unweighted graph $G = (V, E)$ with nonnegative (nonmetric) distances $d : E \to \mathbb{R}_+$ and a number K, find a set C of size K that minimizes $\sum_{v \in V} \min_{c \in C} d(x, c)$.

   - Formulate the K-median problem as an ILP. Let the optimal value of its LP relaxation be C.
   - For the general (nonmetric) case, show how to round this LP solution to an integer solution with at most $O((1 + \epsilon) \log |V| \cdot K)$ medians of objective value $O((1 + \frac{1}{\epsilon}) \cdot C))$ for any $\epsilon > 0$. (Hint: Use the Filtering method of Lin-Vitter from class to transform to a set covering problem.)
   - For the metric case (as in class, when distances obey the triangle inequality), show how to round this LP solution to an integer solution with at most $O((1 + \epsilon) \cdot K)$ medians of objective value $O((1 + \frac{1}{\epsilon}) \cdot C))$ for any $\epsilon > 0$.

3. In class, we saw an algorithm for the facility location problem that rounded the natural LP and obtained a 4-approximation. Consider the following modified rounding procedure:

   - **(Sampling Phase)** For each potential facility i, independently add it to a set F_1 with probability y_i. For a demand j, if some facility in $\{i \mid x_{ij} > 0\}$ is picked in F_0, assign it to the closest such facility. Note that all demands may have not been assigned to facilities in this stage.
   - **(Second Phase)** Use the filtering+clustering procedure from class to open more facilities, and assign the remaining demands to these facilities.

Argue that the expected cost of F_0 plus the expected cost of the assigned demands is not much larger than Z_{LP}. Now since the second phase is performed only on a smaller set of nodes, can you show that the expected cost of the entire procedure is better than $4Z_{LP}$?

4. We saw the $1|r_j|\sum_j C_j$ problem in class, and argued that given a preemptive schedule for the problem (with completion times C_j^P), we could change it into a non-preemptive schedule with completion times $C_j^N \leq 2C_j^P$.

   Now suppose you are given m parallel identical machines (i.e., the problem called $P|r_j|\sum_j C_j$), and you have a preemptive schedule for this problem. (Note that a preemptive schedule for parallel machines must still ensure that any job can run on at most one machine at any given point in time; however, it can be stopped and continued later on *any* machine.)

   Given a preemptive schedule for m parallel identical machines, how do you into a non-preemptive schedule with completion times $C_j^N \leq 3C_j^P$?

   This immediately implies that $\sum_j C_j^N \leq 3\sum_j C_j^P$, and hence a 3-approximation.

   Note: In the single machine case with release dates, the best preemptive schedule is given by SRPT. This is not the case with parallel machines, where the problem is NP-hard even on two machines.