CS 787: Advanced Algorithms **Homework 2**

Out: 10/12/09 **Due: 11/02/09**

Guidelines

- This homework consists of a few exercises followed by some problems. The exercises are meant for *your practice only*, and do not have to be turned in. You are required to turn in the problems. We will provide solutions to all the questions.
- Collaboration is not allowed. Some of the problems are difficult, so please get started early. Late submissions do not get any credit.

Exercises

1. Consider the following variant of 3-SAT called **Not-All-Equal-SAT**: Given a 3-CNF formula, decide whether there exists an assignment to the variables of the formula such that every clause contains at least one literal that is true and at least one literal that is false. Show that this problem is NP-complete.
2. Recall that we showed in class that the natural LP relaxation for the vertex cover problem has an integrality gap of at most 2. Give an example where the integrality gap of the LP is $2 - o(1)$.

Problems

1. Recall that in the set cover problem our goal is to cover a given set of elements using a collection of subsets of the least total cost. In the **maximum coverage problem** our goal is complementary—we want to cover the most number of elements using a collection of subsets with total cost no larger than a given bound B.
   (a) Give a constant factor approximation for the maximum coverage problem. Try to obtain as good an approximation as possible.
   (b) Prove that if maximum coverage can be approximated within a factor of $1-1/e+\epsilon$ in polynomial time for some constant $\epsilon > 0$, then there exists a constant $\epsilon' > 0$ such that set cover can be approximated within a factor of $(1-\epsilon')\log n$ in polynomial time. You may assume that you know the cost of the optimal solution for the set cover problem (but not the actual solution).
2. A **legal** k**-coloring** of a graph is an assignment of colors $1, 2, \cdots, k$ to the vertices of the graph such that no two adjacent vertices receive the same color. A graph is k**-colorable** if there exists a legal k-coloring of its vertices. The problem of finding a legal k-coloring of a k-colorable graph is NP-complete for $k \geq 3$.
   (a) Prove that graphs with maximum degree Δ are $(\Delta + 1)$-colorable. Also give a polynomial time algorithm for finding a $(\Delta + 1)$-coloring.
   (b) Give a polynomial time algorithm for 2-coloring a bipartite graph.
   (c) Using parts (a) and (b) above, give a polynomial time algorithm for finding an $O(\sqrt{n})$-coloring of a 3-colorable graph.
   (Hint: Prove that the neighborhood of any vertex in a 3*-colorable graph is* 2*-colorable. Use this fact to color the neighborhood of high-degree nodes.)*
   (d) Extend the algorithm from part (c) to obtain an $O(n^{2/3})$-coloring for a 4-colorable graph in polynomial time.

(Aside: The best known algorithm for coloring 3-colorable graphs uses $O(n^{0.2111})$ colors.)

3. Recall the LP developed in class for the facility location problem.

   (a) Find the dual of this LP, and give an interpretation to the dual problem (that is, explain what the dual feasible solutions represent).

   (b) Give an instance of the facility location problem for which the integrality gap of the LP is strictly larger than 1. Aim for an integrality gap lying between 1 and 2, and try to obtain as large a factor as you can.

4. We will now develop an approximation for the *non-metric* facility location problem based on the LP relaxation discussed in class. Let LPC(x, y) denote the LP's cost for a feasible solution (x, y). Consider the optimal LP solution $(x^*, y^*) = \text{argmin}_{(x,y)}$ LPC(x, y), and let $A_i = \sum_j c_{ij} x^*_{ij}$ denote the "average" routing cost of customer i in this solution. Recall from class that we can find a feasible solution $(\tilde{x}, \tilde{y})$ by setting to zero all x^*_{ij} for which $c_{ij} > 2A_i$, and modifying the y-values appropriately.

   We will now round $(\tilde{x}, \tilde{y})$ to an integral solution. For every customer i, let $F(i)$ denote the set of facilities j for which $\tilde{x}_{i,j} > 0$. Give an algorithm for picking a subset of facilities that contains at least one facility from each set $F(i)$, and approximately minimizes the total facility opening cost of this subset. Prove that the total cost of your solution is at most $O(\log n)$ times LPC$(\tilde{x}, \tilde{y})$, where n denotes the number of customers. Conclude that you obtain an $O(\log n)$ approximation.

   Hint: Interpret this problem of picking a subset of facilities as a set cover problem. Note that you should not have to use the triangle inequality anywhere in your argument.

5. Approximating the TSP on directed graphs is much harder than on undirected graphs. Until recently the best known algorithm for the former achieved an $O(\log n)$-approximation. A few months back a new $O(\log n/ \log\log n)$ approximation was developed. In this question we will consider the TSP in special directed graphs that we will call $\{1, 2\}$-graphs. These are complete graphs with each directed edge of length 1 or 2. (Convince yourself that these lengths satisfy the triangle inequality.)

   (a) First we will look at a problem closely related to the TSP. In the cycle-cover problem our goal is to find a minimum weight collection of simple (directed) cycles in the graph such that each vertex in the graph is contained in exactly one cycle. Prove that the minimum weight cycle cover in a graph can be found in polynomial time. (Cycles of length 2 are allowed.)

       Hint: Reduce the problem to min-cost perfect matching.

   (b) Use the algorithm from part (a) to give a $3/2$-approximation for the TSP on $\{1, 2\}$-graphs.