15-852 Randomized Algorithms

Homework # 3 **due: Feb 26, 1997**

Exercises:

1. The NP-hard *set cover problem* is: given subsets $S_1, S_2, \ldots, S_m$ of $V = \{1, 2, \ldots, n\}$, cover V with as few of the S_i's as possible.

   Suppose we are given an instance of set-cover whose optimal solution has size p (it uses p of the sets S_i). Show that the simple greedy algorithm ("pick the largest subset; remove the nodes it covers from V; repeat.") has the property that after finding p subsets, it has covered at least a $1 - 1/e$ fraction of V.

Problems:

2. Consider the setup in exercise 1 (an instance of set-cover whose optimal solution has size p). However, suppose now that V contains some unknown ("hidden") set H of "important points", and what we would like is to find p subsets that cover a *constant fraction of* H, rather than just cover a constant fraction of V. Describe a randomized algorithm to do this. It should find p subsets, and have the property that for any $H \subseteq V$, the expected number of points in H covered is at least $|H|(1 - 1/e)$.

3. Assume we have a poly time transformation T from 3CNF formulas to 3CNF formulas with the following properties:

   – If formula ϕ is satisfiable, then $\phi' = T(\phi)$ is satisfiable.

   – If ϕ is not satisfiable, then not only is ϕ' not satisfiable, but in fact any assignment to the variables of ϕ' satisfies at most a $(1 - \varepsilon)$ fraction of the clauses of ϕ'. The quantity ε is a specific constant greater than 0, like 1/113.

   Show that for some $\varepsilon' > 0$, if there is a polynomial time algorithm that approximates the Clique problem on an N-node graph to a factor at most $N^{\varepsilon'}$, then P=NP (i.e., approximating Clique to better than $N^{\varepsilon'}$ is NP-hard.

   Hint: Use result on recycling random bits to derandomize the construction in Feb 17 class notes.

4. Problem 6.11 in the book: show that if G is *regular* (all nodes have the same degree) then the cover time of G is bounded by $O(n^2 \log n)$. Thus, for regular graphs of degree greater than $\log n$, we get a better bound than our result for general graphs.

   Hint: you may find it convenient to use Theorem 6.9.