Introduction to Algorithms	Problem Set 8
CS 482 Spring 2004	Due April 9, 2004

Please hand in each problem on a separate sheet with your name and NetID on each.

Reading: Chapter 6.

(1) Your friends' pre-school-age daughter Madison has recently learned to spell some simple words. To help encourage this, her parents got her a colorful set of refrigerator magnets featuring the letters of the alphabet (some number of copies of the letter 'A,' some number of copies of the letter 'B,' and so on), and the last time you saw her the two of you spent a while arranging the magnets to spell out words that she knows.

Somehow with you and Madison, things always end up getting more elaborate than originally planned, and soon the two of you were trying to spell out words so as to use up all the magnets in the full set – that is, picking words that she knows how to spell, so that once they were all spelled out, each magnet was participating in the spelling of exactly one of the words. (Multiple copies of words are okay here; so for example, if the set of refrigerator magnets includes two copies of each of 'C,' 'A,' and 'T,' it would be okay to spell out "CAT" twice.)

This turned out to be pretty difficult, and it was only later that you realized a plausible reason for this. Suppose we consider a general version of the problem of *Using Up All the Refrigerator Magnets*, where we replace the English alphabet by an arbitrary collection of symbols, and we model Madison's vocabulary as an arbitrary set of strings over this collection of symbols. The goal is the same as in the previous paragraph.

Prove that *Using Up All the Refrigerator Magnets* is NP-complete.

(2) There are many different ways to formalize the intuitive problem of *clustering*, where the goal is to divide up a collection of objects into groups that are "similar" to one another.

First, a natural way to express the the input to a clustering problem is via a set of objects $p_1, p_2, \ldots, p_n$, with a numerical distance $d(p_i, p_j)$ defined on each pair. (We require only that $d(p_i, p_i) = 0$; that $d(p_i, p_j) > 0$ for distinct p_i and p_j; and that distances are symmetric: $d(p_i, p_j) = d(p_j, p_i)$.)

Section 2.5 of the book (which we didn't cover in lecture) considers one reasonable formulation of the clustering problem: divide the objects into k sets so as to *maximize* the minimum distance between any pair of objects in distinct clusters. This turns out to be solvable by a nice application of the minimum spanning tree problem.

A different but seemingly related way to formalize the clustering problem would be as follows: divide the objects into k sets so as to *minimize* the maximum distance between any pair of objects in the same cluster. Note the change: where the formulation in the previous paragraph sought clusters so that no two were "close together," this new formulation seeks clusters so that none of them is too "wide" — i.e. no cluster contains two points at a large distance from one another.

Given the similarities, it's perhaps surprising that this new formulation is computationally hard to solve optimally. To be able to think about this in terms of NP-completeness, let's write it first as a yes/no decision problem: given n objects $p_1, p_2, \ldots, p_n$ with distances on them as above, and a bound B, we define the *Low-Diameter Clustering* problem as follows: can the objects be partitioned into k sets, so that no two points in the same set are at distance greater than B from one another?

Prove that *Low-Diameter Clustering* is NP-complete.

(3) In a barter economy, people trade goods and services directly, without money as an intermediate step in the process. Trades happen when each party views the set of goods they're getting as more valuable than the set of goods they're giving in return. Historically, societies tend to move from barter-based to money-based economies, and so various on-line systems that have been experimenting with barter can be viewed as an intentional attempt at regression back to this earlier form of economic interaction. In doing this, they've re-discovered some of the inherent difficulties with barter relative to money-based systems; one such difficulty is the complexity of identifying opportunities for trading, even when these opportunities exist.

To model this complexity, we need a notion that each person assigns a *value* to each object in the world, indicating how much this object would be worth to them. Thus, we assume there is a set of n people $p_1, \ldots, p_n$, and a set of m distinct objects $a_1, \ldots, a_m$. Each object is owned by one of the people. Now, each person p_i has a *valuation function* v_i, defined so that $v_i(a_j)$ is a non-negative number that specifies how much object a_j is worth to p_i — the larger the number, the more valuable the object is to them. Note that everyone assigns a valuation to each object, including the ones they don't currently possess, and different people can assign very different valuations to the same object.

A two-person trade is possible in a system like this when there are people p_i and p_j, and subsets of objects A_i and A_j possessed by p_i and p_j respectively, so that each person would prefer the objects in the subset they don't currently have. More precisely,

- p_i's total valuation for the objects in A_j exceeds his or her total valuation for the objects in A_i, and
- p_j's total valuation for the objects in A_i exceeds his or her total valuation for the objects in A_j.

(Note that A_i doesn't have to be *all* the objects possessed by p_i (and likewise for A_j); A_i and A_j can be arbitrary subsets of their possessions that meet these criteria.)

Suppose you are given an instance of a barter economy, specified by the above data on people's valuations for objects. (To prevent problems with representing real numbers, we'll assume that each person's valuation for each object is a natural number.) Prove that the problem of determining whether a two-person trade is possible is NP-complete.