CSE 6140 / CX 4140
Computational Science and Engineering Algorithms

Homework 5

1. (25 points)

A *combinatorial auction* is a particular mechanism developed by economists for selling a collection of items to a collection of potential buyers. (The Federal Communications Commission has studied this type of auction for assigning stations on the radio spectrum to broadcasting companies.)

Here's a simple type of combinatorial auction. There are n items for sale, labeled $I_1, \ldots, I_n$. Each item is indivisible and can only be sold to one person. Now, m different people place *bids*: The i^{th} bid specifies a subset S_i of the items, and an *offering price* x_i that the bidder is willing to pay for the items in the set S_i, as a single unit. (We'll represent this bid as the pair (S_i, x_i).)

An auctioneer now looks at the set of all m bids; she chooses to *accept* some of these bids and to *reject* the others. Each person whose bid i is accepted gets to take all the items in the corresponding set S_i. Thus the rule is that no two accepted bids can specify sets that contain a common item, since this would involve giving the same item to two different people.

The auctioneer collects the sum of the offering prices of all accepted bids. (Note that this is a "one-shot" auction; there is no opportunity to place further bids.) The auctioneer's goal is to collect as much money as possible.

Thus, the problem of *Winner Determination for Combinatorial Auctions* asks: Given items $I_1, \ldots, I_n$, bids $(S_1, x_1), \ldots, (S_m, x_m)$, and a bound B, is there a collection of bids that the auctioneer can accept so as to collect an amount of money that is at least B?

Example. Suppose an auctioneer decides to use this method to sell some excess computer equipment. There are four items labeled "`PC`," "`monitor`," "`printer`", and "`scanner`"; and three people place bids. Define

$$S_1 = \{\texttt{PC}, \texttt{monitor}\}, S_2 = \{\texttt{PC}, \texttt{printer}\}, S_3 = \{\texttt{monitor}, \texttt{printer}, \texttt{scanner}\}$$

and

$$x_1 = x_2 = x_3 = 1.$$

The bids are $(S_1, x_1), (S_2, x_2), (S_3, x_3)$, and the bound B is equal to 2.

Then the answer to this instance is no: The auctioneer can accept at most one of the bids (since any two bids have a desired item in common), and this results in a total monetary value of only 1.

Prove that the problem of Winner Determination in Combinatorial Auctions is NP-complete. Reduce from Independent Set.

2. (25 points)

In the 1970s, researchers including Mark Granovetter and Thomas Schelling in the mathematical social sciences began trying to develop models of certain kinds of collective human behaviors: Why do particular fads catch on while others die out? Why do particular technological innovations achieve widespread adoption, while others remain focused on a small group of users? What are the dynamics by which rioting and looting behavior sometimes (but only rarely) emerges from a crowd of angry people? They proposed that these are all examples of *cascade processes*, in which an individual's behavior is highly influenced by the behaviors of his or her friends, and so if a few individuals instigate the process, it can spread to more and more people and eventually have a very wide impact. We can think of this process as being like the spread of an illness, or a rumor, jumping from person to person.

The most basic version of their models is the following. There is some underlying *behavior* (e.g., playing ice hockey, owning a cell phone, taking part in a riot), and at any point in time each person is either an *adopter* of the behavior or a *nonadopter*. We represent the population by a directed graph $G = (V, E)$ in which the nodes correspond to people and there is an edge (v, w) if person v has influence over the behavior of person w: If person v adopts the behavior, then this helps induce person w to adopt it as well. Each person w also has a given *threshold* $\theta(w) \in [0, 1]$, and this has the following meaning: At any time when at least a $\theta(w)$ fraction of the nodes with edges to w are adopters of the behavior, the node w will become an adopter as well.

Note that nodes with lower thresholds are more easily convinced to adopt the behavior, while nodes with higher thresholds are more conservative. A node w with threshold $\theta(w) = 0$ will adopt the behavior immediately, with no inducement from friends. Finally, we need a convention about nodes with no incoming edges: We will say that they become adopters if $\theta(w) = 0$, and cannot become adopters if they have any larger threshold.

Given an instance of this model, we can simulate the spread of the behavior as follows.

```
Initially, set all nodes w with θ(w) = 0 to be adopters
   (All other nodes start out as nonadopters)
Until there is no change in the set of adopters:
   For each nonadopter w simultaneously:
      If at least a θ(w) fraction of nodes with edges to w are
         adopters then
           w becomes an adopter
      Endif
   Endfor
End
Output the final set of adopters
```

Note that this process terminates, since there are only n individuals total, and at least one new person becomes an adopter in each iteration.

Now, in the last few years, researchers in marketing and data mining have looked at how a model like this could be used to investigate "word-of-mouth" effects in the success of new products (the so-called *viral marketing* phenomenon). The idea here is that the behavior we're concerned with is the use of a new product; we may be able to convince a few key people in the population to try out this product, and hope to trigger as large a cascade as possible.

Concretely, suppose we choose a set of nodes $S \subseteq V$ and we reset the threshold of each node in S to 0. (By convincing them to try the product, we've ensured that they're adopters.) We then run the process described above, and see how large the final set of adopters is. Let's denote the size of this final set of adopters by $f(S)$ (note that we write it as a function of S, since it naturally depends on our choice of S). We could think of $f(S)$ as the *influence* of the set S, since it captures how widely the behavior spreads when "seeded" at S.

The goal, if we're marketing a product, is to find a small set S whose influence $f(S)$ is as large as possible. We thus define the *Influence Maximization Problem* as follows: Given a directed graph $G = (V, E)$, with a threshold value at each node, and parameters k and b, is there a set S of at most k nodes for which $f(S) \geq b$?

Prove that Influence Maximization is NP-complete.

Reduce from Vertex Cover.

Example. Suppose our graph $G = (V, E)$ has five nodes $\{a, b, c, d, e\}$, four edges (a, b), (b, c), (e, d), (d, c), and all node thresholds equal to $2/3$. Then the answer to the Influence Maximization instance defined by G, with $k = 2$ and $b = 5$, is yes: We can choose $S = \{a, e\}$, and this will cause the other three nodes to become adopters as well. (This is the only choice of S that will work here. For example, if we choose $S = \{a, d\}$, then b and c will become adopters, but e won't; if we choose $S = \{a, b\}$, then none of c, d, or e will become adopters.)

3. (25 points)

Madison is in preschool and has learned to spell some simple words. She has 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. The two of you tried 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 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 the problem of Using Up All the Refrigerator Magnets is NP-complete. Reduce from 3D matching.

4. (Next Page)

4. (25 points)

One thing that's not always apparent when thinking about traditional "continuous math" problems is the way discrete, combinatorial issues of the kind we're studying here can creep into what look like standard calculus questions.

Consider, for example, the traditional problem of minimizing a one-variable function like $f(x) = 3 + x - 3x^2$ over an interval like $x \in [0, 1]$. The derivative has a zero at $x = 1/6$, but this in fact is a maximum of the function, not a minimum; to get the minimum, one has to heed the standard warning to check the values on the boundary of the interval as well. (The minimum is in fact achieved on the boundary, at $x = 1$.)

Checking the boundary isn't such a problem when you have a function in one variable; but suppose we're now dealing with the problem of minimizing a function in n variables $x_1, x_2, \ldots, x_n$ over the unit cube, where each of $x_1, x_2, \ldots, x_n \in [0, 1]$. The minimum may be achieved on the interior of the cube, but it may be achieved on the boundary; and this latter prospect is rather daunting, since the boundary consists of 2^n "corners" (where each x_i is equal to either 0 or 1) as well as various pieces of other dimensions. Calculus books tend to get suspiciously vague around here, when trying to describe how to handle multivariable minimization problems in the face of this complexity.

It turns out there's a reason for this: Minimizing an n-variable function over the unit cube in n dimensions is as hard as an NP-complete problem. To make this concrete, let's consider the special case of polynomials with integer coefficients over n variables $x_1, x_2, \ldots, x_n$. To review some terminology, we say a ***monomial*** is a product of a real-number coefficient c and each variable x_i raised to some nonnegative integer power a_i; we can write this as $cx_1^{a_1}x_2^{a_2} \cdots x_n^{a_n}$. (For example, $2x_1^2x_2x_3^4$ is a monomial.) A ***polynomial*** is then a sum of a finite set of monomials. (For example, $2x_1^2x_2x_3^4 + x_1x_3 - 6x_2^2x_3^2$ is a polynomial.)

We define the *Multivariable Polynomial Minimization Problem* as follows: Given a polynomial in n variables with integer coefficients, and given an integer bound B, is there a choice of real numbers $x_1, x_2, \ldots, x_n \in [0, 1]$ that causes the polynomial to achieve a value that is $\leq B$?

Prove that *Multivariable Polynomial Minimization* is NP-complete.
Reduce from 3-SAT.