College of Computer & Information Science | Spring 2011
Northeastern University | Handout 20
CS5800: Algorithms | 4 April 2011

Problem Set 6 (due Wednesday, April 13)

1. (10 points) Selecting online advertisements

Major online portals like Google and Yahoo have considerable information about the individual users based on their past interactions. This allows them to post targeted advertisements to the users. Suppose a set U of n users, labeled 1 through n, visit the portal on a particular day. The portal has a set A of m ads, labeled 1 through m, to choose from. The analysis of the users has revealed k different groups (from a marketing standpoint), the ith group consisting of subset S_i of users from U. A user may be part of several groups; i.e., a user may be an element of several different S_i's. The jth ad is targeted to a subset $G_j \subseteq \{1, \ldots, k\}$ of the groups.

The portal needs to decide whether there exists a way of assigning advertisements to users such that the following conditions hold: (a) each user is shown exactly one ad; (b) ad j is shown to user i only if i is in a group k in G_j; (c) the number of times the ad j is shown is at least r_j, where r_j is a given integer.

Give a polynomial-time algorithm that takes the above input – U, A, the sets S_i's, the groups G_j's, the r_j's – and determines whether the portal can assign an ad to each user so that the above three conditions are satisfied, and if so, then returns such an assignment.

2. (10 points) Privacy of survey data

Last year was the year of the Census. The results of a survey such as the Census are often disclosed in aggregate form to ensure the confidentiality of the information. Revealing nothing other than the aggregates also has a downside since that may provide too little information.

Suppose a survey has produced a $m \times n$ table D, in which each entry d_{ij} is a nonnegative integer. Let $r(i)$ be the sum of the elements in the ith row of D, and let $c(j)$ denote the sum of he elements in the jth column of D. The surveyor would like to disclose all the m row sums and the n column sums. Furthermore, the surveyor wants to disclose a subset, say S, of the mn matrix elements, and yet suppress the remaining matrix elements to ensure the confidentiality of the supressed information. Unless care is exercised, the surveyor may permit someone to deduce the exact value of one or more of the suppressed elements. We say that an entry d_{ij} of the matrix is *unprotected*, if it is possible to deduce the exact value of d_{ij} from the row and column sums and the elements of S.

Given m, n, the row and column sums, and the elements of S, our problem is to identify all the unprotected elements of the matrix and their values.

Using the maximum flow problem, give a polynomial-time algorithm for the above problem. Justify the correctness of your algorithm. You need not worry about the complexity of your algorithm, as long as it is polynomial-time. For instance, you may make multiple (polynomially many) calls to maximum flow subroutines, if needed.

3. (10 points) Design of a salad

Problem 7.16 of text.

4. (10 points) Optimal strategies in pizza business

Problem 7.14 of text.