CS602 Applied Algorithms **Jul 2023**

Assignment 2

Total Marks: 40 *Deadline: Nov 10, 14:00*

Note: The assignment needs to be done individually. Any kind of discussion with other students is not allowed. If you feel the need to discuss anything, you can reach out to the instructor. You can directly use any result proved in the class. You can use other sources (book/internet), but you should give a reference and specify for which part it has been used.

Que 1 [10 marks]. Recall that in the Steiner tree problem we are given a edge weighted graph G with a set of terminal vertices and we have to compute a minimum weight subgraph where all terminal vertices are connected with each other. Consider the following algorithm for minimum weight Steiner tree problem.

Compute shortest path for each pair of terminals. Construct a complete graph H on the terminal vertices where edge weights are the computed shortest path lengths between pairs of terminals. Construct a minimum spanning tree T for the graph H. Take the union of the paths in G corresponding to the MST edges. Remove any redundant edges in any order.

Assume that the number of terminals given is 3. Prove that the above algorithm gives a 4/3-approximate Steiner tree.

Hint: consider the optimal Steiner tree on 3 terminals. Try to upper bound this in terms of sum of paths between two pairs of terminals (appropriately chosen), with a multiplicative factor. Upper bound the sum by weight of a tree in the graph H.

Que 2 [10 marks]. Consider the minimum weight Steiner forest problem with terminal pairs $(s_1, t_1), (s_2, t_2)$. That is, we want a Steiner forest that connects s_1 with t_1 and s_2 with t_2. Recall the primal dual algorithm for the Steiner forest problem discussed in the class. Show that the algorithm gives a solution with approximation factor 3/2.

Hint: look closely into the analysis done for the approximation factor

Que 3 [8+2 marks]. Consider a variant of the vertex cover problem. Here, we have weights on both edges and vertices. We want to maximize the weight of edges covered and minimize the weight of selected vertices. You can think of this as getting a reward for covering an edge and also getting a reward for not selecting a vertex.

Formally, let $G(V, E)$ be a graph. Let $P_1, P_2, \ldots, P_n$ be the weights of the vertices. Let $\{Q_{i,j} : (i, j) \in E\}$ be the weights on the edges. The goal is select a subset $S \subseteq V$ of vertices so as to maximize the below expression.

$$\sum_{i \notin S} P_i + \sum_{\substack{(i,j)\in E: \\ i \in S \\ \text{or } j \in S}} Q_{i,j}.$$

We write the following integer program.

$$\max \sum_{i=1}^{n} (1 - y_i) P_i + \sum_{(i,j)\in E} z_{i,j} Q_{i,j} \text{ subject to.}$$

$$y_i \in \{0, 1\} \text{ for } 1 \le i \le n$$

$$z_{i,j} \in \{0, 1\} \text{ for } (i, j) \in E$$

$$z_{i,j} \le y_i + y_j \text{ for } (i, j) \in E$$

First relax the integer constraint and write a linear program. We find an optimal solution for that LP using some LP solver. Let the optimal solution be (y^*, z^*). The optimal solution can be fractional. We convert it to a Boolean solution with the following rounding scheme: For each $1 \leq i \leq n$, select vertex i with probability $(1-\lambda)+\lambda y_i^*$.

Show that the approximation factor (i.e., rounded solution cost divided by LP optimal) is at least $\min\{\lambda, 1-\lambda^2/4\}$.

What should be the choice of λ that gives the best approximation factor. What is the approximation factor for this choice.

Que 4 [5+5 marks]. Suppose we are given n objects, their pairwise dissimilarities $\{d_{i,j} : 1 \leq i < j \leq n\}$ and also their pairwise similarities $\{s_{i,j} : 1 \leq i < j \leq n\}$. Similarities and dissimilarities are positive real numbers, and not necessarily related to each other. We want to partition the objects into **two** clusters so as to maximize the total dissimilarity of pairs lying in different clusters and the total similarity of pairs lying in same clusters. That is, maximize the following objective function

$$\sum_{\substack{i,j \\ \text{which are in} \\ \text{different clusters}}} d_{i,j} + \sum_{\substack{i,j \\ \text{which are in} \\ \text{same clusters}}} s_{i,j}.$$

We can write the following integer program for the problem.

$$\max \sum_{i,j} d_{i,j}(1-z_i z_j)/2 + \sum_{i,j} s_{i,j}(1+z_i z_j)/2 \text{ subject to}$$
$$z_i \in \{-1,1\} \text{ for } 1 \leq i \leq n$$

Relax the program to a vector program. Write the corresponding semidefinite program. Suppose we round the SDP optimal solution by the same scheme as in Max-cut problem. Show that we get the same approximation factor (roughly 0.878).