CMPUT 675 - Assignment #4

Fall 2014, University of Alberta
Due November 14 in class.

Pages: 4

This assignment is to be completed individually. I understand that you may want to discuss the assignment with other students, a good guide for understanding my expectations is that you should not take notes or work out precise details in your discussions (keep it high-level). Mention any discussions and cite any resources you used on the writeup you hand in.

To be clear, whenever you are asked to give an approximation algorithm for a problem, it is expected that you will both describe the algorithm and prove the claimed approximation guarantee. If you can only think of an algorithm with a worse approximation guarantee than I am asking for, then describe it anyway. You may get partial marks (though, it cannot be entirely trivial). The same goes with lower bounds.

Problem 1)

Marks: 2

Say that an instance of MAX 2SAT is balanced if every clause has exactly two literals (over different variables) and the weight of all clauses in which the literal x_i appears equals the weight of all clauses in which the literal $\overline{x_i}$ appears for each variable x_i. Show that the standard hyperplane rounding algorithm we used in the lecture for MAX 2SAT in fact gives us a β-approximation for balanced instances where β is defined below. [2 marks]

$$\beta := \min_{\theta \in [0,\pi]} \frac{\frac{1}{2} + \frac{1}{2\pi} \cdot \theta}{\frac{3}{4} - \frac{1}{4} \cdot \cos\theta} \approx 0.94394.$$

You do not have to prove the numerical estimate.

Problem 2)

Marks: 2

Recall the GROUP STEINER TREE problem from the lectures. Here, we have a graph $G = (V, E)$ with edge lengths $d(e) \geq 0$, "groups" of nodes $X_1, X_2, \ldots, X_k \subseteq V$, and a root node r. The goal is to find the cheapest subset of edges F such that for each group X_i, there is a path from r to some node in X_i in the graph (V, F). The LP relaxation we considered is the following.

$$
\begin{aligned}
\text{minimize :} \quad & \sum_{e \in E} d(e) \cdot x_e \\
\text{subject to :} \quad & \sum_{e \in \delta(S)} x_e \;\geq\; 1 \quad \text{for each } S \subseteq V - \{r\} \text{ such that } X_i \subseteq X \text{ for some } 1 \leq i \leq k \\
& \qquad\quad \mathbf{x} \;\geq\; 0
\end{aligned}
\tag{\textbf{LP-Q2}}
$$

We saw that the integrality gap of (**LP-Q2**) is $O(\log n \cdot \log k)$ (where $n = |V|$) if G is a tree. Using tree embeddings, this lead to a randomized $O(\log^2 n \cdot \log k)$-approximation in general graphs.

Your job is to show that the integrality gap of (**LP-Q2**) is $O(\log^2 n \cdot \log k)$ for any instance of GROUP STEINER TREE. [2 marks]

Problem 3)

Marks: 4

In the *Capacitated Dial-a-Ride* problem, we are given a metric (V, d), a single vehicle with a given integer capacity $C \geq 0$ located at a given depot node $r \in V$, and k different clients. Each client i is currently located at a vertex $s_i \in V$ and needs to be transported to a location $t_i \in V$.

The goal is to find the shortest tour that starts and ends at the depot r that transports clients between their start locations and their end locations. We are allowed to drop clients off at intermediate locations and then pick them up later in the tour. The only additional constraint is that the vehicle can hold at most C clients at any time

- Suppose the metric is in fact a tree metric. Give a constant-factor approximation for this case.
  Hint: For each edge e, let $\ell(e)$ denote the number of clients s_i, t_i such that e lies on the path between s_i and t_i. Describe an algorithm that crosses each edge at most $O(1) \cdot \ell(e)/C$ times and argue why this is a good approximation. [2 marks]
- Recall the random construction of a tree metric for (V, d) that we saw in class. It associated every node u of the tree to a subset $S_u \subseteq V$. Keeping this fact in mind (when dropping off clients at intermediate nodes in your 2-approximation for trees), describe a randomized $O(\log n)$-approximation in general metrics. You may want to review other specific properties of the tree metrics that are constructed in the randomized algorithm we saw in class. [2 marks]

Problem 4)

Marks: 4

Consider the following generalization of the local search heuristic for k-MEDIAN we saw in the lectures. Recall that we let $f(S) = \sum_{j \in C} d(j, S)$ for any subset S of facilities.

Let p be a constant. Initialize S to any set of k facilities. While there are some $A \subseteq S$ and $B \subseteq V - S$ with $|A| = |B| \leq p$ such that $f((S - A) \cup B) < f(S)$, update $S \leftarrow (S - A) \cup B$. Repeat until no such improvements are possible and return this final set S.

Here, you will prove that $f(S) \leq (3 + \frac{2}{p}) \cdot OPT$ where S is any locally optimum solution with respect to these swaps.

Recall the notation we used in the lecture. Let S^* be an optimum solution, let $\sigma : S^* \to S$ map each facility in S^* to its nearest facility in S, let $\phi^* : C \to S^*$ map each client to its nearest facility in S^*, and let $\phi : C \to S$ map each client to its nearest facility in S.

This time, we partition S slightly differently into sets M, P, N. Let

- $M = \{i \in S : |\sigma^{-1}(i)| \geq p + 1\}$,
- $P = \{i \in S : 1 \leq |\sigma^{-1}(i)| \leq p\}$, and
- $N = \{i \in S : |\sigma^{-1}(i)| = 0\}$.

Partition S^* into sets M^*, P^* where

- $M^* = \{i \in S^* : \sigma(i) \in M\}$ and
- $P^* = \{i \in S^* : \sigma(i) \in P\}$.

This is sketched in Figure 1.

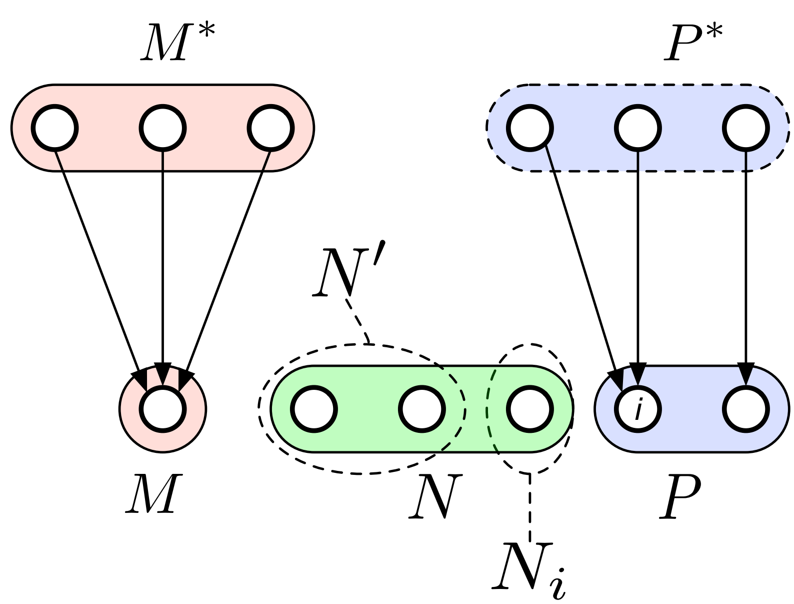

Figure 1: Illustration of the partition of S and S^* for $p = 2$. The downward pointing arrows indicate the mapping σ. The sets N' and N_i are described below.

- Show $|P| + |N| \geq |P^*|$. [0.5 marks]
- For each $i \in P$ let N_i be any subset of N of size $|\sigma^{-1}(i)| - 1$. Do this in way such that that $N_{i'} \cap N_{i''} = \emptyset$ for any two $i', i'' \in P$ (the previous part ensures this is possible).

  For each $i \in P$ let $A_i := \{i\} \cup N_i$ and $B_i := \sigma^{-1}(i)$, show the following holds. [1 mark]

$$0 \leq f((S - A_i) \cup B_i) - f(S) \leq \sum_{j:\phi^*(j) \in B_i} (d_j^* - d_j) + \sum_{j:\phi(j) \in A_i} 2d_j^*$$

- Let $N' = N - \cup_{i \in P} N_i$ be the set of facilities in N that were not involved in a swap in the previous part. Show $|M^*| \leq \left(\frac{p+1}{p}\right) \cdot |N'|$. [0.5 marks]
- Let $\mathcal{P} = M^* \times N'$ (i.e. all possible pairs). For every $(i, i') \in \mathcal{P}$ we have

$$0 \leq f(S - i' + i) - f(S) \leq \sum_{j:\phi^*(j) = i} (d_j^* - d_j) + \sum_{j:\phi(j) = i'} 2d_j^*.$$

  This was already proven in the lectures, you do not have to reprove it.

  Use these test swaps and the test swaps in the previous part for $i \in P$ to show $f(S) \leq (3 + \frac{2}{p}) \cdot OPT$. [2 marks]

  Hint: You will want to scale the inequalities generated by the $\mathcal{P}$ swaps down by some value when combining all of these inequalities.