Albert-Ludwigs-Universität, Inst. für Informatik
Prof. Dr. Fabian Kuhn
M. Ahmadi, A. R. Molla, O. Saukh February 4, 2016

Algorithm Theory, Winter Term 2015/16
Problem Set 12 - Sample Solution

Exercise 1: Covering as many Elements as Possible (8 points)

We consider the following variant of the set cover problem discussed in the lecture. We are given a set of elements X and a collection $\mathcal{S} \subseteq 2^X$ of subsets of X such that $\bigcup_{S \in \mathcal{S}} S = X$. In addition, we are given an integer parameter $k \geq 2$.

Instead of finding a collection $\mathcal{C} \subseteq \mathcal{S}$ of the sets which covers all elements, the goal is to find a set of (at most) k set $S_1, \dots, S_k \in \mathcal{S}$ such that the number of covered elements $|S_1 \cup \dots \cup S_k|$ is maximized.

We consider the greedy set cover algorithm from the lecture, but we stop the algorithm after adding k sets.

(a) (2 points) Show that for $k = 2$, the described greedy algorithm has approximation ratio at most 4/3.

(b) (4 points) Let us now consider a general parameter $k \geq 2$. Show that if an optimal choice of k sets $S_1, \dots, S_k$ covers ℓ elements, after adding t sets, the greedy algorithm covers at least $\frac{\ell}{k} \cdot \sum_{i=1}^{t} \left(1 - \frac{1}{k}\right)^{i-1}$ elements.

(c) (2 points) Prove that the approximation ratio of the greedy algorithm is at most $\frac{e}{e-1}$. You can use that $(1 - 1/k)^k < e^{-1}$.

Solution

(a) Let S denote the optimal solution which covers x elements. For the case $k = 2$, the optimal solution consists of one or two non empty sets in $\mathcal{S}$. Let us consider the two cases as follows:

- Assume that the optimal solution consists of only one non empty set, that is, $S = \{S_1\}$. Then S_1 is a set in $\mathcal{S}$ with cardinality x, and all other sets in $\mathcal{S}$ are subset of S_1. Since the greedy approach picks S_1 as the set with maximum cardinality in $\mathcal{S}$, the greedy solution is optimal.
- Assume the optimal solution consists of more than one set, that is $S = \{S_1, S_2\}$. Moreover, let $F = \{F_1, F_2\}$ denote the greedy solution. Since $|S_1 \cup S_2| = x$, then either $|S_1| \geq x/2$ or $|S_2| \geq x/2$. Hence, we can claim that the set with maximum cardinality in $\mathcal{S}$ has cardinality $\geq x/2$, which is picked by the greedy algorithm for the first choice (i.e., $|F_1| \geq x/2$).

  After the first choice of the greedy solution, $\{S_1 \cup S_2\} \setminus F_1$ is the set of uncovered elements in the optimal solution. These elements are covered by S_1 and S_2. Hence, either S_1 or S_2 covers at least half of these elements. Therefore, we can claim that the set in $\mathcal{S}$ which covers the maximum number of uncovered elements covers at least $|\{S_1 \cup S_2\} \setminus F_1|/2$ elements.

  Therefore the greedy solution covers at least $|F_1| + \frac{x-|F_1|}{2} = \frac{x+|F_1|}{2}$ elements. Considering $|F_1| \geq x/2$, the greedy solution covers at least $\frac{x+x/2}{2} = \frac{3x}{4}$ elements, which proves the claim on greedy solution's approximation factor.

(b) The core part of the proof is the following claim.

Claim 1. Considering the optimal solution $S = \{S_1, S_2, \ldots, S_k\}$, for any $A \subseteq \bigcup_{i=1}^{k} S_i$, there exists at least one set U in S such that $|U \cap A| \geq \frac{|A|}{k}$. That is, U covers at least $\frac{|A|}{k}$ elements in A.

Proof. Let us assume that there does not exist such a set in S. Then we can say that all of the sets in S covers less than $\frac{|A|}{k}$ elements in A. Since there are at most k sets in S, the union of all the sets in S cannot cover all the elements in A. This contradicts the fact that the the sets in S cover all elements in $X \supseteq A$. □

Let $E = \{e_1, e_2, \ldots, e_\ell\}$ be the set of all elements that are covered by the optimal solution. Moreover, let y_t denote the number of elements that the greedy solution covers after choosing t sets. Hence, the number of uncovered elements by greedy approach in E before it chooses the t^{th} set is $\ell - y_{t-1}$. Based on Claim 1, for the set of uncovered elements in E there exist at least one set in the optimal solution that covers at least $\frac{\ell - y_{t-1}}{k}$ new elements in the optimal solution. Since in step t the greedy algorithm chooses a set in $\mathcal{S}$ which covers the maximum number of uncovered elements, $y_t - y_{t-1} \geq \frac{\ell - y_{t-1}}{k}$. Therefore, we have the following recurrence relation for the number of elements that the greedy solution covers.

$$y_t \geq \frac{\ell}{k} + y_{t-1}\left(1 - \frac{1}{k}\right)$$

Considering the fact that $y_0 = 0$ (before greedy chooses any set, the number of covered elements is zero), by repeated replacement, the above recurrence relation leads to the claim stated in the question.

(c) Considering the recurrence relation achieved in question (b), the number of elements that are covered by greedy solution with k chosen sets is at least

$$\begin{aligned}
\frac{\ell}{k} \cdot \sum_{i=1}^{k} \left(1 - \frac{1}{k}\right)^{i-1} &= \frac{\ell}{k} \cdot \frac{1 - \left(1 - \frac{1}{k}\right)^k}{1 - \left(1 - \frac{1}{k}\right)} \\
&\overset{**}{\geq} \frac{\ell}{k} \cdot \frac{1 - \frac{1}{e}}{1 - \left(1 - \frac{1}{k}\right)} \\
&= \ell \cdot \frac{e-1}{e}.
\end{aligned}$$

** $\left(1 - \frac{1}{k}\right)^k < \frac{1}{e}$

As a result we can calculate the approximation factor as follows.

$$\frac{Optimal}{Greedy} \leq \frac{\ell}{\ell \cdot \frac{e-1}{e}} = \frac{e}{e-1}$$

Exercise 2: TSP in Graphs with Edge Weights 1 and 2 (4 points)

Consider the family of complete undirected graph G in which all edges have length either 1 or 2. Give a 4/3-approximation for the TSP problem for this family of graphs. Note that G satisfies the triangle inequality.

Hint: Start with a minimum 2-matching in G. A 2-matching is a subset M of edges so that every vertex in G is incident to exactly two edges in M. You can assume that a minimum 2-matching can be computed in polynomial time.

Solution

First we prove the following claim.

Claim 2. The cost of any TSP tour in G is lower bounded by the minimum 2-matching in G.

Proof. The answer for TSP problem is a tour which visits each vertex in G exactly once. Therefore, this tour enters and exits each vertex via different edges. Hence, a TSP tour with minimum length in G is also a 2-matching for G. Hence, the minimum 2-matching is a lower bound for the length of any TSP tour. $\square$

Here we show that a 2-matching solution can be transformed to a TSP tour. A 2-matching consists of one or several components such that in each component each vertex is incident to exactly two edges. Due to this property, each of these components consists of at least 3 vertices (If a components has less than three vertices then each vertex cannot be incident to exactly 2 edges). There exist, therefore, at most $\frac{n}{3}$ such components. Let us denote these components by $C_1, C_2, \ldots, C_k$, where $k \leq \frac{n}{3}$. In each component C_i choose an arbitrary edge $\{u_i, v_i\}$. Remove all such edges in all the components and connect v_j to u_{j+1} for all $1 \leq j \leq k-1$. Moreover, connect v_k to u_1. The result of this transformation is a TSP tour. In this transformation we remove k edges and add k edges to the 2-matching.

If we transform the minimum 2-matching in G with weight W to a TSP tour, then by considering the worst case that all the k removed edges have weight 1 and all the added edges have weight 2, the length of the TSP tour is at most

$$W + 2k - k = W + k \leq W + \frac{n}{3}$$

Therefore, the total length of the transformed TSP tour is $\leq W + \frac{n}{3}$. Hence we get,

$$\begin{aligned}
\text{Length of the transformed TSP tour} &\leq W + \frac{n}{3} \\
&\leq W + \frac{W}{3} \\
&[n \leq W, \text{ since the minimum 2-matching consists of } n \text{ edges each of weights at least 1.}] \\
&= \left(1 + \frac{1}{3}\right) W \\
&= \frac{4}{3} \times W \\
&\leq \frac{4}{3} \times \text{Length of the minimum TSP tour} \\
&[\text{It follows from Claim 2; } W \leq \text{length of the minimum TSP tour.}]
\end{aligned}$$

This proves that there exists a TSP tour such that the ratio of its weight to the minimum TSP tour is at most $\frac{4}{3}$.