

Problem Set 9 Solutions

This problem set is due **at 9:00pm on Wednesday, May 9, 2012.**

Problem 9-1. Plumbing Problem

The Mario brothers are hired to outfit the Stata Center with a plumbing system. There is a set of n rooms in the building that ideally would all have plumbing. However, due to a reduced budget compounded by a fear of leaks, the administrators ask the brothers to provide the minimum piping needed to service at least the rooms in a mission-critical subset S of all rooms.

The Mario brothers have asked you to help find the piping of minimum overall length that serves all rooms in S , and may serve some other rooms in the building as well. Keep in mind that, for engineering reasons, pipes can only be joined inside rooms.

- (a) Luigi, one of the Mario brothers, assumes that any pair of rooms can be connected by pipes and suggests finding the Minimum Pipe Spanning Tree (MPST) for the rooms in the mission-critical set S . Let G be a complete graph where vertices represent rooms in the building and the weight of each edge is the length of the pipes required to connect the corresponding rooms. The MPST for S is simply the Minimum Spanning Tree for the subgraph of G induced by S .

Provide an example of a room graph where finding the MPST for S would not give the piping of minimum overall length.

Solution:

See figure below.

- (b) Luigi has taken a closer look at the blueprint, which shows the distance $d(x, y)$ for each pair of rooms x, y in the building. He notes that the distances between any two rooms satisfy the triangle inequality. So, for any three rooms x, y and z , $d(x, z) \leq d(x, y) + d(y, z)$.

Assume any pair of rooms x, y can be connected using a pipe of length $d(x, y)$. Devise an efficient polynomial-time approximation algorithm for the plumbing problem with an approximation ratio $\alpha \leq 2$. **Hint:** Consider building an MPST on S .

Solution: This is the metric Steiner tree problem. Construct an MST T on the subgraph induced by set S . The claim is that the cost of the MST is $2 \cdot OPT$.

Proof Take a Steiner tree of cost OPT . Double the edges (as in the 2-approximation algorithm for TSP). Find an Euler tour in the resulting Eulerian graph, by traversing the edges with a DFS. The cost of the Euler tour is $2 \cdot OPT$. Use shortcutting trick to obtain a Hamiltonian cycle on vertices of S . By triangle inequality the shortcuts don't increase the cost of tour. If an arbitrary edge of this Hamiltonian cycle is deleted we obtain a path that spans S with a cost at most $2 \cdot OPT$.

- (c) Mario notes that due to the topology of the building, pipes can only run between some pairs of rooms and not others. What's more, Alyssa P. Hacker sets up a few strategic wormholes, shortcuts through spacetime, so that the pipe-based space-time displacements no longer satisfy the triangle inequality.

You are provided with an updated blueprint of possible direct pipe connections and effective pipe lengths for each of them. Devise a polynomial-time approximation algorithm with approximation ratio $\alpha \leq 2$, using what you learned in part (b).

Solution: We produce a polynomial reduction from the Steiner tree problem to the metric Steiner problem in (b) that also preserves the approximation factor. Use Johnson's or Floyd Warshall algorithm to compute all pairs shortest paths for the input graph, yielding a complete graph that encodes all of the shortest path weights between all pairs of vertices in original graph. By construction it satisfies the triangle inequality. Since all edges in E' that were also in E have a cost smaller than or equal to those in E , the cost of the optimal solution to the transformed graph will not exceed that of the optimal solution to original problem.

Now we prove the converse. Given a Steiner tree T' we obtain a Steiner tree T in polynomial time of at most the same cost. Replace each of T' by the corresponding shortest path to obtain a subgraph of G . All the vertices are connected, but the subgraph may contain cycles. Remove edges to obtain tree T . By construction the reduction preserves the approximation factor.

Problem 9-2. Huffman Coding

- (a) You are working with a Huffman encoding scheme over n symbols. How long could a codeword be in the worst case? Provide an example set of frequencies f_1, f_2, \dots, f_n (as a closed formula of n) that would give rise to this longest codeword.

Solution:

Consider the following set of frequencies: $f_{n-i} = 2^{-i}$ for $i \in \{1, \dots, n-1\}$ for an alphabet of length n . The longest codeword will have length $n-1$. The cumulative sum of all frequencies up to and including the $j-1$ -th smallest frequency will be less than f_j : $\sum_{i=1}^{j-1} f_i < f_j$. So, by construction, the tree will be a spine with a leaf joining each intermediate subtree.

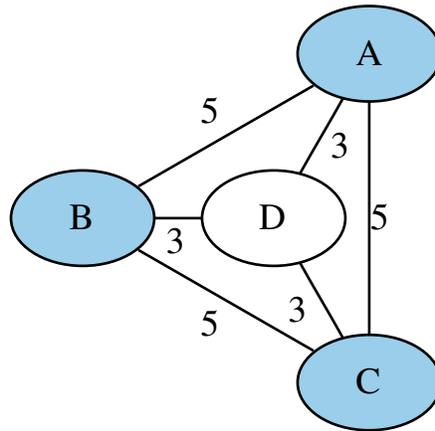


Figure 1: Illustration of a complete graph D_4 . The shaded, outlying nodes are part of the required set S . The MST for outlying nodes will have a higher cost than the tree that also spans the internal vertex.

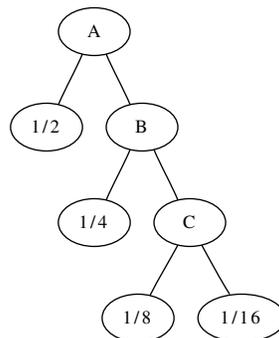


Figure 2: Illustration of a Huffman code tree with the longest codeword for $n = 4$.

- (b) Prove that if some character occurs with frequency more than $2/5$, then there is guaranteed to be a codeword of length 1.

Solution:

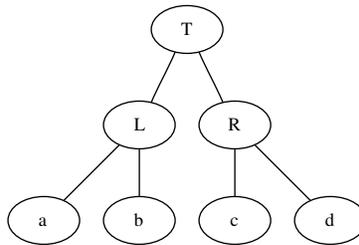


Figure 3: Illustration of a hypothetical tree with no codewords of length 1.

Assume for the purpose of obtaining a contradiction that there is no codeword of length 1. Then both children of the root can not be leaves. Let the left child of the root be L , right child R , and the two children of each subtree a , b , and c , d respectively.

Without loss of generality assume that the character with frequency greater than $2/5$ appears in the subtree a of the left child of the root. The right subtree R will have to have combined weight more than $2/5$, since otherwise R being smaller than a would be joined first with b . The combined weight on b will be less than $1/5$ since the total sum of frequencies has to add up to 1.

At the same time, the right subtree has to have cumulative weight less than $3/5$, since one of the elements in L has frequency in excess of $2/5$. One of the two subtrees of R has to have weight less than $3/10$, but then it would be combined with subtree b , since the total weight of R is greater than $2/5$ and so the other subtree of R would have weight greater than that of b .

This is a contradiction since, b was joined with a , so there is guaranteed to be a codeword of length 1.

- (c) Prove that if all characters occur with frequency less than $1/3$, then there is guaranteed to be no codeword of length 1.

Solution: There have to be at least 4 symbols in the alphabet. Assume by contradiction that there is a codeword of length 1. The codeword of length 1 will branch from the root of the tree. WLOG say it has frequency f_x . By construction of the Huffman code tree, it has to have been merged with a subtree, that is a result of two subtrees (possibly leaves), call them A and B . Both A and B have to have cumulative weight less than f_x , since this last element was pulled last from the priority queue. Then, since $f_x \leq 1/3$ by assumption that all characters occur with frequency less than $1/3$,

$w(A) + w(B) + f_x < 1$. This is a contradiction. Since the sum of all weights in the subtree has to add up to 1.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.046J / 18.410J Design and Analysis of Algorithms
Spring 2012

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.