Chapter 3

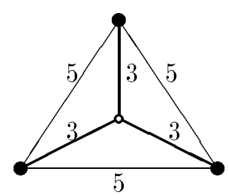
Metric Steiner tree and TSP

The origin of the Steiner tree problem goes back to Gauss, who posed it in a letter to Schumacher. This problem and its generalizations will be studied extensively in this monograph.

Problem 3.1 (Metric Steiner tree) Given a graph $G = (V, E)$ whose edge costs satisfy triangle inequality and whose vertices are partitioned into two sets, *required* and *Steiner*, find a minimum cost tree containing all the required vertices and any subset of the Steiner vertices.

Remark: There is no loss of generality in requiring that the edge costs satisfy triangle inequality: if they don't satisfy triangle inequality, construct the *metric closure* of G, say G', which has the same vertex set as G and edge costs given by shortest distances in G. Clearly, the cost of the optimal Steiner tree in both graphs must be the same. Now, obtaining a Steiner tree in G', and replacing edges by paths wherever needed, gives a Steiner tree in G of at most the same cost.

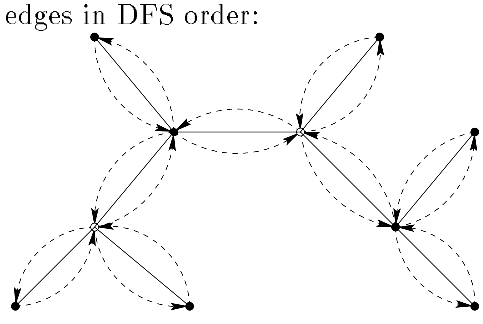
Let R denote the set of required vertices. Clearly, a minimum spanning tree (MST) on R is a feasible solution for this problem. Since the problem of finding an MST is in **P**and the metric Steiner tree problem is **NP**-hard, we cannot expect the MST on R to always give an optimal Steiner tree; below is an example in which the MST is strictly costlier.



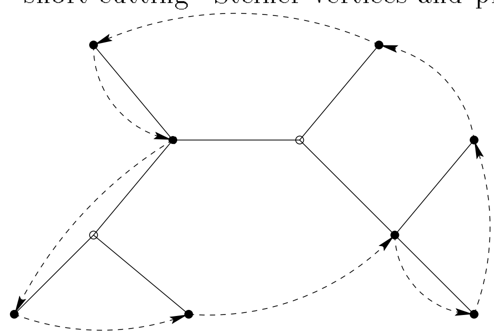
Even so, an MST on R is not much more costly than an optimal Steiner tree:

Theorem 3.2 *The cost of an MST on R is within* $2 \cdot \mathrm{OPT}$.

Proof : Consider a Steiner tree of cost OPT. By doubling its edges we obtain an Eulerian graph connecting all vertices of R and, possibly, some Steiner vertices. Find an Euler tour of this graph, for example by traversing the edges in DFS order:



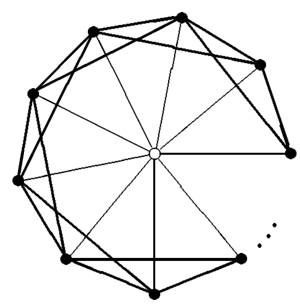
The cost of this Euler tour is $2 \cdot \text{OPT}$. Next obtain a Hamilton tour on the vertices of R by traversing the Euler tour and "short-cutting" Steiner vertices and previously visited vertices of R:



Because of triangle inequality, the shortcuts do not not increase the cost of the tour. If we delete one edge of this Hamilton tour, we obtain a path that spans R and has cost at most $2 \cdot \text{OPT}$. This path is also a spanning tree on R. Hence, the MST on R has cost at most $2 \cdot \text{OPT}$. □

Theorem 3.2 gives a straightforward factor 2 algorithm for the metric Steiner tree problem: simply find an MST on the set of required vertices. As in the case of set cover, the "correct" way of viewing this algorithm is in the setting of LP-duality theory. This will provide the lower bound on which this algorithm is based, and will also help solve generalizations of this problem.

Example 3.3 A tight example is provided by a graph with n required vertices and one Steiner vertex. Each edge between the Steiner vertex and a required vertex has cost 1, and all other edges have cost $(2 - \epsilon)$, where $\epsilon > 0$ is a small number (not all edges of cost $(2 - \epsilon)$ are shown below). In this graph, an MST on R has cost $(2 - \epsilon)(n - 1)$, while $\text{OPT} = n$.



□

Exercise 3.4 Let $G = (V, E)$ be a graph with non-negative edge costs. The vertices of G are partitioned into two sets, *senders* and *receivers*. The problem is to find a minimum cost subgraph of G that has a path connecting each receiver to a sender. Give a good approximation algorithm for this NP-hard problem.

Approximation algorithms for TSP

The following is a well-studied problem in combinatorial optimization.

Problem 3.5 (Traveling salesman problem (TSP)) Given a complete graph with non-negative edge costs, find a minimum cost cycle visiting every vertex exactly once.

Not only is it **NP**-hard to solve this problem exactly, but also approximately:

Theorem 3.6 *For any polynomial time computable function $\alpha(n)$, TSP cannot be approximated within a factor of $\alpha(n)$, unless* $\mathbf{P} = \mathbf{NP}$.

Proof : Assume for a contradiction that for any graph on n vertices, we can find in polynomial time a salesman tour whose cost is within a factor of $\alpha(n)$ from the optimum. We show that this implies a polynomial time algorithm for deciding whether a given graph has a Hamiltonian cycle.

Let G be a graph on n vertices. We extend G to the complete graph on n vertices, assigning unit cost to edges of G, and a cost of $n\alpha(n)$ to edges not in G. Clearly, the optimal salesman tour in the new graph has a cost of n if and only if G has a Hamiltonian cycle. Moreover, any tour that contains a new edge costs more than $n\alpha(n)$. So, an $\alpha(n)$ approximation algorithm finds a tour of cost n whenever one exists. $\square$

Notice that in order to obtain such a strong non-approximability result, we had to assign edge costs that violate triangle inequality. If we restrict ourselves to graphs in which edge costs satisfy triangle inequality, i.e., the *metric traveling salesman problem*, the problem remains **NP**-complete, but it is no longer hard to approximate.

We will first present a simple factor 2 algorithm. The lower bound we will use for obtaining this factor is the cost of an MST in G. This is a lower bound because deleting any edge from an optimal solution to TSP we get a spanning tree of G.

Algorithm 3.7 (Metric TSP – factor 2)

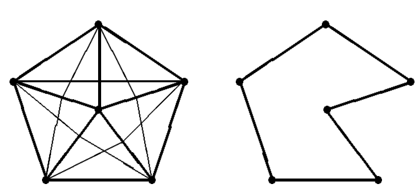
1. Find an MST, T, of G.
2. Double every edge of the MST to obtain an Eulerian graph.
3. Find an Euler tour, $\mathcal{T}$, on this graph.
4. Output the tour that visits vertices of G in order of their first appearence in $\mathcal{T}$. Let $\mathcal{C}$ be this tour.

Notice that Step 4 is similar to the "short-cutting" step in Theorem 3.2.

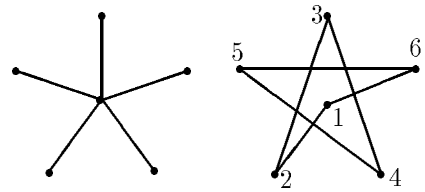
Theorem 3.8 *Algorithm 3.7 is a factor 2 algorithm for metric TSP.*

Proof : As noted above, $\text{cost}(T) \leq \text{OPT}$. Since $\mathcal{T}$ contains each edge of T twice, $\text{cost}(\mathcal{T}) = 2 \cdot \text{cost}(T)$. Because of triangle inequality, after the "short-cutting" step, $\text{cost}(\mathcal{C}) \leq \text{cost}(\mathcal{T})$. Combining these inequalities we get that $\text{cost}(\mathcal{C}) \leq 2 \cdot \text{OPT}$. $\square$

Example 3.9 A tight example for this algorithm is given by the complete graph on n vertices with edges of cost 1 and 2. We present the graph for $n = 6$ below. The thick edges have cost 1 and the remaining edges have cost 2. On n vertices, the graph will have $2n-2$ edges of cost 2, and the remaining edges of cost 1, with the cost 2 edges forming the union of a star and an $n-1$ cycle. The optimal TSP tour has cost n as shown below.



Suppose that the MST found by the algorithm is the spanning star created by edges of cost 1. Moreover, suppose that the Euler tour constructed in Step 3 visits vertices in order shown below:



Then the tour obtained after short-cutting contains $n-2$ edges of cost 2, and has a total cost of $2n-2$. This is almost twice the cost of the optimal TSP tour. □

Essentially, this algorithm first finds a low cost Euler tour spanning the vertices of G, and then short-cuts this tour to find a travelling salesman tour. Is there a cheaper Euler tour than that found by doubling an MST? Notice that we only need to be concerned about the vertices of odd degree in the MST; let V' denote this set of vertices. $|V'|$ must be even since the sum of degrees of all vertices in the MST is even (it is $2n-2$). Now, if we add to the MST a minimum cost perfect matching on V', every vertex will have even degree, and we get an Eulerian graph. With this modification, the algorithm achieves an approximation guarantee of $\frac{3}{2}$.

Algorithm 3.10 (Metric TSP – factor $\frac{3}{2}$)

1. Find an MST of G, say T.
2. Compute a minimum cost perfect matching, M, on the set of odd vertices of T. Add M to T and obtain an Eulerian graph.
3. Find an Euler tour, $\mathcal{T}$, of this graph.
4. Output the tour that visits vertices of G in order of their first appearence in $\mathcal{T}$. Let $\mathcal{C}$ be this tour.

Interestingly, the proof of this algorithm is based on a second lower bound on OPT.

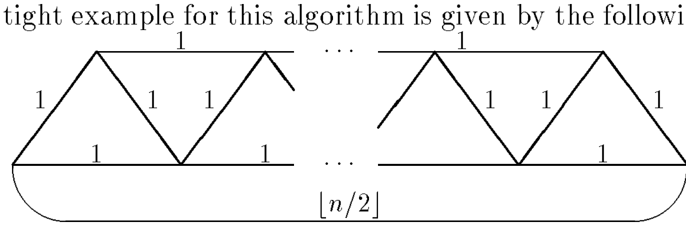
Lemma 3.11 *Let $V' \subseteq V$, such that $|V'|$ is even, and let M be a minimum cost perfect matching on V'. Then,* $\text{cost}(M) \leq \text{OPT}/2$.

Proof : Consider an optimal TSP tour of G, say τ. Let τ' be the tour on V' obtained by short-cutting τ. By triangle inequality, $\text{cost}(\tau') \leq \text{cost}(\tau)$. Now, τ' is the union of two perfect matchings on V', each consisting of alternate edges of τ. So, the cheaper of these matchings has cost $\leq \frac{\text{cost}(\tau')}{2} \leq \frac{\text{OPT}}{2}$. Hence the optimal matching also has cost $\leq \frac{\text{OPT}}{2}$. □

Theorem 3.12 *Algorithm 3.10 achieves an approximation guarantee of $\frac{3}{2}$ for metric TSP.*

Proof : The proof follows by putting together the two lower bounds on OPT. □

Example 3.13 A tight example for this algorithm is given by the following graph on n vertices:



Thick edges represent the MST found in step 1. This MST has only two odd vertices, and by adding the edge joining them we obtain a traveling salesman tour of cost $(n-1)+\lfloor n/2 \rfloor$. In contrast, the optimal tour has cost n. □

Finding a better approximation algorithm for metric TSP is currently one of the outstanding open problems in this area. Many researchers have conjectured that an approximation factor of 4/3 may be achievable.

Exercise 3.14 Consider the following variant of metric TSP: given vertices $u, v \in V$, find a minimum cost simple path from u to v that visits all vertices. First give a factor 2 approximation algorithm for this problem, and then improve it to factor $\frac{3}{2}$.

Exercise 3.15 Give a factor 2 approximation algorithm for: Given an undirected graph $G = (V, E)$, with non-negative edge costs, and a partitioning of V into two sets Senders and Recievers, find a minimum cost subgraph such that every Receiver vertex has a path to a Sender vertex.