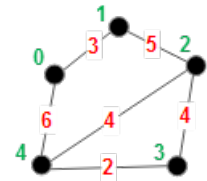


Minimum Spanning Trees

We consider connected undirected graphs that have a weight on each edge. Each weight is a positive integer. In the graph to the right, the nodes are numbered (in green) 0, 1, ..., 4. The edge weights are given in red.



A *minimum-cost spanning tree*, or *minimum spanning tree*, is a spanning tree whose sum of the weights on its edges is a minimum over all spanning trees of the graph. The graph to the right has two minimum spanning trees, with cost 14. They use edges $\{3, 4\}$, $\{0, 1\}$, one of $\{2, 3\}$ and $\{2, 4\}$, and $\{1, 2\}$.

There are many applications of minimum spanning trees. Here's one. A company has to lay cable in a neighborhood, reaching all houses (the nodes of the graph). The cables have to be laid along certain roads (the edges). The cost on an edge could depend on the length of the edge, how deep the cable has to be buried, and other factors. A spanning tree would describe where the cable should be laid. A minimum spanning tree would minimize the cost.

Here are other applications. Telecommunications networks, including the internet, have loops, or cycles, of course. There are "routing protocols" for sending packets from one node on the network to another. These protocols require each router to maintain a spanning tree, and a minimum-cost spanning tree is best.

Two greedy algorithms

We have already discussed the basic additive algorithm for constructing a spanning tree for a graph G , which is:

Start with the nodes of G and no edges;

Repeat until no longer possible:

Add an edge (of G) that connects two unconnected components, i.e. does not form a cycle.

We show two abstract algorithms for constructing a minimum-cost spanning tree that are refinements of this basic algorithm. They are both *greedy* algorithms—if you don't know this term, visit JavaHyperText entry "greedy algorithm" and study it before reading further.

Kruskal's algorithm

Kruskal published his minimum spanning tree algorithm in 1956. In it, he makes the above abstract algorithm into a greedy one by always adding an edge of minimum weight. Here is his algorithm at an abstract level:

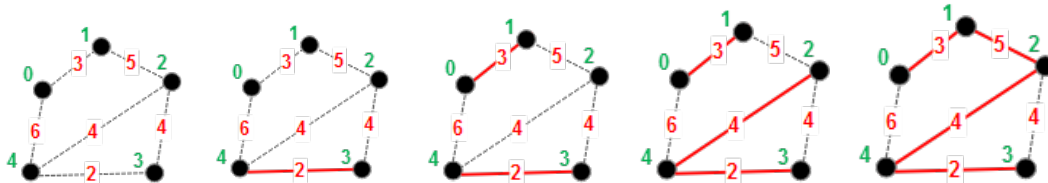
Start with the nodes of G and no edges;

Repeat until no longer possible:

Add an edge (of G) *of minimum weight* that connects two unconnected components.

We step through this algorithm using the graph given at the top of the page. Below, the leftmost graph shows the initial graph being constructed. It has the nodes of G and that is all. Edges are dotted just to indicate the edges of G .

The first iteration adds edge $\{3, 4\}$ with minimum weight 2. The second iteration adds edge $\{0, 1\}$ with minimum weight 3. In the third iteration, there is a choice because two edges have weight 4. We arbitrarily chose edge $\{2, 4\}$. For the fourth iteration, edge $\{2, 3\}$ with weight 4 cannot be chosen because its endpoints are in the same component (and a cycle would be formed), so edge $\{1, 2\}$ with weight 5 is chosen. That's it.



The Jarnik/Prim/Dijkstra (JDP) algorithm

Wikipedia says that this algorithm was first developed in 1930 by Czech mathematician Vojtěch Jarník. It was rediscovered and republished by Robert Prim in 1957 and Edsger W. Dijkstra in 1959, although Dijkstra may have developed it when he developed his shortest path algorithm in 1956. It is usually known as Prim's algorithm, but in light of its history, we call it the JDP algorithm.

The JDP algorithm uses the same greedy choice as Kruskal's algorithm, but there is a major difference. In Kruskal's algorithm, the added edges can belong to different connected components. For example, after edges $\{0, 1\}$ and $\{3, 4\}$ are added, they belong to two different connected components. The JDP algorithm requires that all added edges belong to the same component. Here's the algorithm.

Start with the nodes of G and no edges;

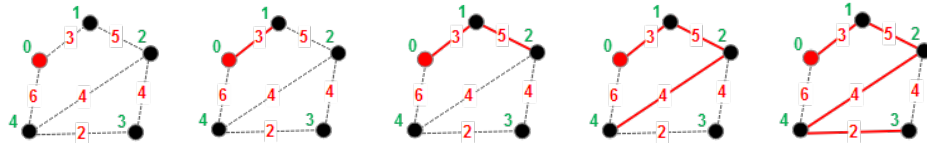
Arbitrarily choose a node w ;

Repeat until no longer possible:

Add an edge (of G) *of minimum weight that joins a connected component with the component that contains w .*

We step through this algorithm using the graph given at the top of the page. Below, the leftmost graph shows the initial graph being constructed, with node w being red node 0. This initial graph contains the nodes of G and no edges. The dotted edges indicate the edges of G .

Because of the requirement that all added edges must be in the component containing node w , the first iteration can add only edge $\{0, 1\}$ or $\{0, 4\}$; it greedily adds the one with minimum weight: $\{0, 1\}$. In the same way, iteration 2 must add either $\{1, 2\}$ or $\{0, 4\}$; it greedily adds the one with minimum weight: $\{1, 2\}$. Iteration 3 can add one of $\{2, 4\}$, $\{2, 3\}$, and $\{0, 4\}$. Two have minimum weight 4, and we have arbitrarily chosen edge $\{2, 4\}$. Finally, iteration 4 greedily adds edge $\{3, 4\}$ —the edge with minimum weight over all edges turns out to be the last one added!



It has been proven that both greedy algorithms —Kruskal and JDP— construct a minimal-cost spanning tree. Further, if the edge weights are all different, they construct the same spanning tree.