**Lab: Graphs – Dijkstra, Shortest Paths and MST**

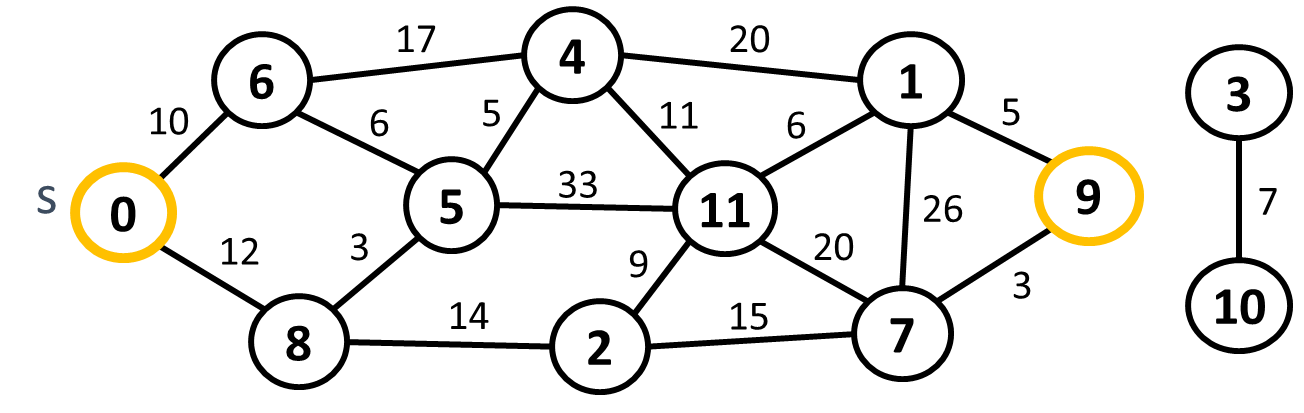
This document defines the lab for the ["Algorithms – Advanced (Java)" course @ Software University](https://softuni.bg/trainings/3611/algorithms-advanced-with-java-january-2022). Please submit your solutions (source code) of all below described problems in [Judge](https://judge.softuni.bg/Contests/2488/Graphs-DijkstraI-and-MST-Lab).

# Dijkstra’s Shortest Path Algorithm

Finding the **shortest path between two nodes** in an unweighted graph is done by applying simple BFS. When we’re working with weighted graphs though, things get more complicated. **Dijkstra’s algorithm** is one of the most famous ones that solve this task.

A classical application of the shortest path algorithm might be to find the shortest path between two towns on a map holding towns connected with roads where each road holds the distance between two towns.

Example: Find the shortest path between **node 0** and **node 9** in the following weighted undirected graph:



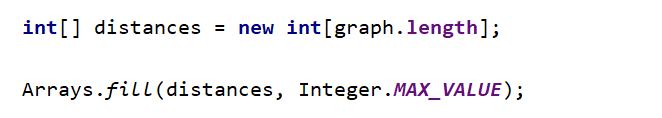
The result is: 0->8->5->4->11->1->9 (length 42).

## Provided Assets

The graph is given as a square matrix where the indices of the rows and columns represent nodes; each cell represents the weight of the edge between two nodes – the row and the column. If the cell has a value of 0 this means that there is no edge connecting the two nodes (or the row and column indices are the same).

## Initialize Distances

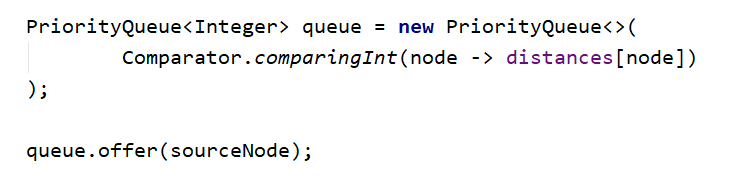
We need an array to hold the minimum distance to each node. Initially, the distance to the source node is 0 and the distance to all other nodes is set to infinity (or, in our case, the maximal value for the int type):



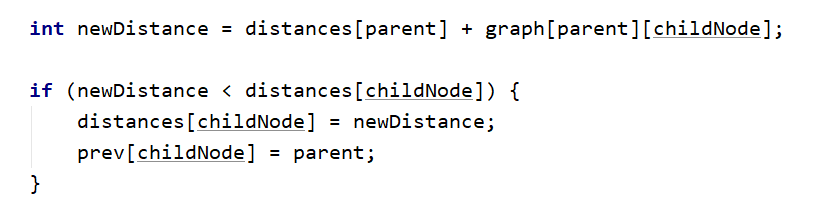
We also need to keep track of the nodes we’ve visited and, to reconstruct the path later, the previous node. The source node has no previous, the value for it will be **-1**.

## Find Nearest Unvisited Node

The next steps take place in a loop – we find the nearest unvisited node and start from there. When all possible nodes are traversed, we’ll exit the loop. This can be done using a **PriorityQueue** where all **nodes** are **ordered** by their **current** **distance** from the **source.**



Finding the nearest unvisited node is simple – loop through all nodes; for each node check whether it’s been visited (this info is kept in the visited[] array), check if the distance between the source node and the current one is smaller than the current shortest distance:



When we’re done, if newDistance is still equal to Integer.MaxValue, this means all possible nodes have been traversed and we need to exit the loop; if not, mark the newNode as visited.

## Improve Shortest Distances

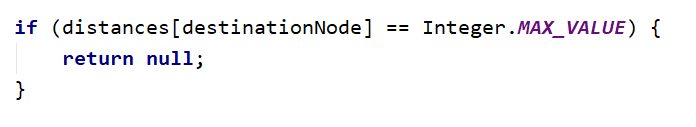
Using the node we just found, we need to go through all nodes connected to it and improve the shortest distances. A node is connected to another node if there is an edge between them; in the context of our matrix, this means a cell with a value greater than 0.

You need to do the following:

1. Loop through each node
2. Check the ones that are connected to the minNode
3. Calculate the distance from the source node to the current node – just add the shortest distance to minNode and the distance between minNode and the node
4. If the calculated distance is shorter than the current shortest distance for the current node – update the shortest distance and make minNode the previous element

## Reconstruct Shortest Path

First thing’s first – if the shortest distance to the destination is still infinity then we haven’t found a path:



To reconstruct the path, we start from the destination node and move back (using the info kept in the previous[] array) until we reach the source – the source has no previous, so the value will be **-1**. At each step, add the node to a list. Reverse the list in the end and return it.

Congratulations! You’ve completed one of the most famous graph algorithms – Dijkstra’s!

# Minimum Spanning Tree (MST) – Kruskal’s Algorithm

If we have a weighted undirected graph we can extract a sub-graph where all nodes (vertices) of the original graph are connected by edges without any cycles. This is referred to as a **spanning tree**. A **minimum spanning tree (MST)** is the spanning tree with the smallest weight (several MST could exist in some graphs).

For example, a cable operator wants to connect its customers to a **cable network**. The homes of the customers are connected by streets (edges) with different lengths (weights). To find out how to connect all homes to its network most efficiently (least distance covered) you need to find the **MST**.

One simple algorithm to find the MST of a given graph is [**Kruskal’s algorithm**](https://en.wikipedia.org/wiki/Kruskal%27s_algorithm). Example:

|  |  |  |
| --- | --- | --- |
| **Graph** | **Output** | **MST** |
|  | Minimum spanning forest weight: 49  (3 5) -> 2  (0 8) -> 5  (1 7) -> 7  (6 8) -> 7  (1 4) -> 8  (3 6) -> 8  (2 6) -> 12 |  |

## Class Edge

In a weighted graph, an edge holds a lot of information – it connects two nodes, one being the source and the other – the destination, and also has a weight. To simplify our work, we need a class Edge to hold this info. We can implement Comparable<Edge> to compare edges by weight. We can also override toString() to be able to print an edge in a human-readable way. Assuming all weights are integers, we’ve added an implementation of the Edge class to the skeleton.

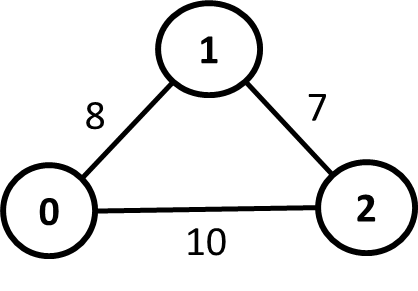
## Sort Edges

Kruskal’s algorithm works by taking all edges in turn, each time the one with the smallest weight is picked. Having implemented Comparable<Edge>, we can simply sort the list of edges, calling the sort() method of the list.

## Preventing Cycles – Concept

Kruskal’s algorithm is simple – having the sorted edges, we take each one, in turn, check whether it causes a cycle if added to the current MST and if not – we add it to the MST. How do we check for cycles though?

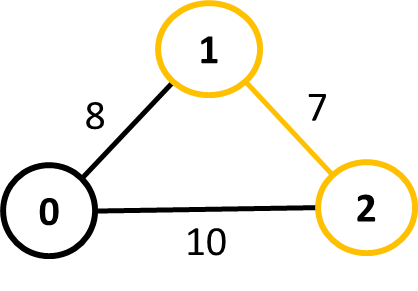
The algorithm works by taking the sorted edges and building the tree from scratch. In a tree, we have a root node and descendants; each descendant has a parent. Consider the following graph:



We have three edges. Following the format described in the Edge class, **(startNode endNode) -> weight**, these are:

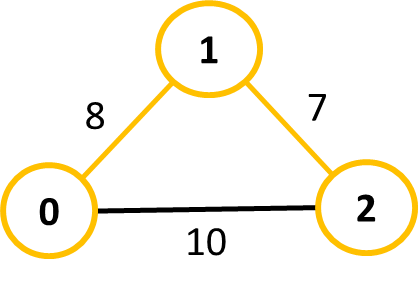
* (1 2) -> 7
* (1 0) -> 8
* (0 2) -> 10

At first, the tree is empty, so none of the nodes have parents. We begin by taking the first node, (1 2) -> 7.



There are no cycles when adding the first edge. However, we now have a tree with two nodes, we need to mark one of the nodes as a parent and the other as a child. It doesn’t matter which will be which, so let’s say that 1 is the parent of 2.

Next, we take the edge (1 0) -> 8.



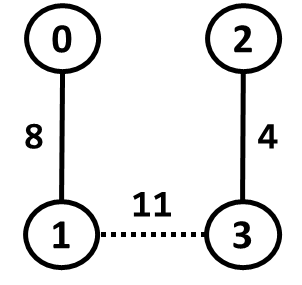
0 has no parent and neither has 1. We can add the edge to the MST. Again, we have to mark which node is the parent. Let’s say 1 is the parent again.

In the final step, we take (0 2) -> 10. This will cause a cycle, but how do we tell? We compare parents for the two nodes of the edge, if they are the same, this means there is a cycle. The parent of 0 is 1; the parent of 2 is also 1, therefore we have a cycle and skip this edge.

What if we chose 0 to be the parent of 1 during step 2? When considering the final edge, 0 has no parent. The parent of 2 is 1, but the parent of 1 is 0, therefore 2 is a descendant of 0.

In essence, we go up the tree following the parents of each node until we reach the **root**. We **compare the roots for both nodes of an edge and if they have the same root then we have a cycle**.

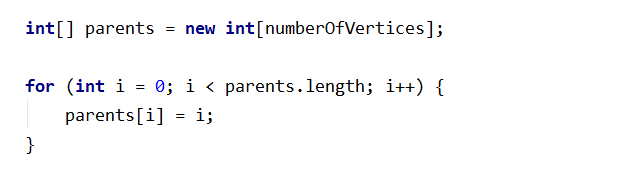
If we have two trees and add a node that connects them, how do we merge the trees? Consider the following example where we have added to the MST two edges (0 1) -> 8 and (2 3) -> 4. We’ve marked that 0 is the parent (in this case also root node) of 1 and 2 is the parent (root node) of 3. When we add (1 3) -> 11 which doesn’t cause a cycle, we need to set parents in such a way that the tree will have a single root.



Having the two roots, 0 and 2, of the nodes from the newly added edge, we can just mark one as a parent of the other, e.g. mark that 0 is a parent of 2. This means that, later on, when checking any of the four nodes, they will all have a single root - 0. 0, 1, and 2 all have 0 as a parent; 3 has 2 as a parent which has 0 as a parent, so the root of 3 is also 0. Therefore, **when adding a new edge to the MST, we need to mark one of the roots as a parent of the other.**

## Keeping Track of Parents - Setup

In a tree, a node can have at most one parent, therefore, we can keep information about each node’s parent in an array. To mark that a node has no parent, we can simply say that the node is its parent. Initializing the array is trivial:

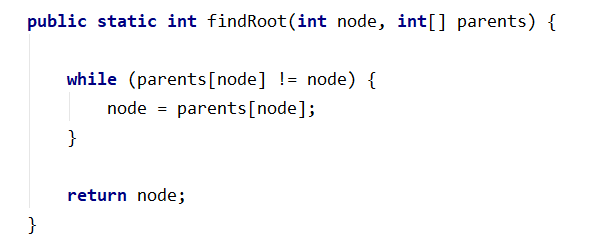


When adding an edge to the MST we’ll mark one of the nodes as a parent of the other node.

## Method - findRoot() - Implementation

We saw that when taking an edge, we need to find the roots for both nodes. We’ll need a method to find the root of a given node.

Finding the root is pretty easy. If the node is its parent, then it is the root. If the node has a different parent, we go to the parent and check again; we repeat until we find the root. We need to traverse the parent[] array until we reach an element that has equal index and value:

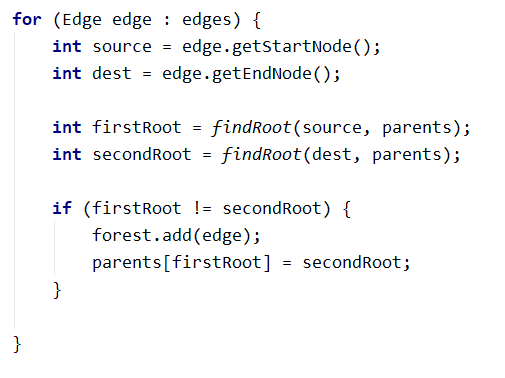


Note that if we have a forest to start with, we’ll have one root per connected component. The findRoot() method and the algorithm itself will still work correctly.

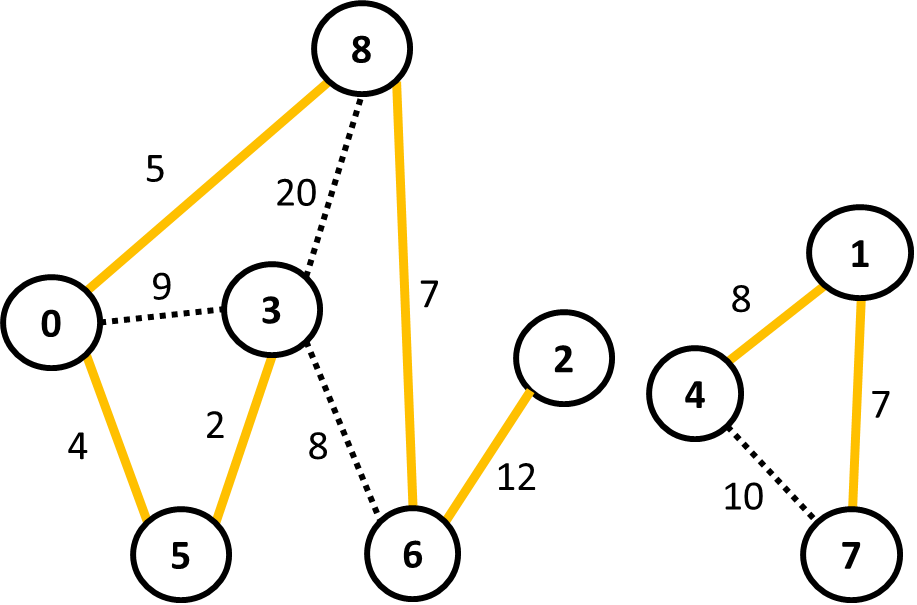
## Kruskal’s Algorithm – Implementation

Having the sorted edges and a way to prevent cycles, it’s time to implement the algorithm itself.

It’s pretty simple: 1) instantiate a list to hold the edges of the MST; 2) traverse all edges; 3) find the roots for both nodes in an edge; 4) if the roots are different – add the edge to the MST and mark one node as a parent of the other.



And this is the result visually:



You’ve successfully implemented Kruskal’s algorithm!