

**ALGORITHM AND DATA STRUCTURE PROJECT**

**DOCUMENTATION REPORT**

PROJECT 2  
ITCS 6114

**DEPARTMENT OF COMPUTER SCIENCE SUBMITTED TO  
Dewan T. Ahmed, Ph.D.**

**SUBMITTED BY:  
Harshitha Keshavaraju Vijayalakshmi Urma Haldar**

**801084151 801074045**

**Table of Contents:**

[***Data Structure & Algorithms - Spanning Tree……………………………………………………………………………….. 2***](%5Cl%20%22_Toc530779463%22)

[*General Properties of Spanning Tree………………………………………………………………………………… 2*](%5Cl%20%22_Toc530779464%22)

[*Mathematical Properties of Spanning Tree……………………………………………………………………….3*](%5Cl%20%22_Toc530779465%22)

[***Application of Spanning Tree………………………………………………………………………………………………………………………3***](%5Cl%20%22_Toc530779466%22)

[*Minimum Spanning Tree (MST)…………………………………………………………………………………………3*](%5Cl%20%22_Toc530779467%22)

[*Minimum Spanning-Tree Algorithm………………………………………………………………………………….3*](%5Cl%20%22_Toc530779468%22)

[***Prim's Spanning Tree Algorithm……………………………………………………………………………………………………………….4***](%5Cl%20%22_Toc530779469%22)

*Pseudo code …………………………………………………………………………………………………………………… 5*

*Source Code and output……………….……………………………………………………………………………………………………………….9*

***Dijkstra's Algorithm*** *……………………………………………………………………………………………………. 12*

*Relaxation………………………………………………………………………………………………………………………..13*

*Pseudo code and implementation……………………………………………………………………………………..14*

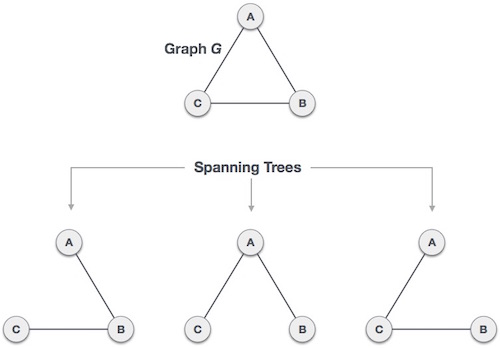
*Source code and output for directed and undirected. ……………………………………………………….18*

*Conclusion …..………………………………………………………………………………………………………………….. 28*

# 

# Data Structure & Algorithms - Spanning Tree

A spanning tree is a subset of Graph G, which has all the vertices secured with least conceivable number of edges. Thus, a traversing tree does not have cycles and it can't be disconnected. By this definition, we can reach a determination that each associated and undirected Graph G has something like one spanning tree. A disconnected diagram does not have any spanning tree, as it can't be spanned to all its vertices.



We found three spanning trees off one complete graph. A complete undirected graph can have maximum **nn-2** number of spanning trees, where **n** is the number of nodes. In the above addressed example, **n is 3,** hence **33−2 = 3**spanning trees are possible.

## General Properties of Spanning Tree

We now understand that one graph can have more than one spanning tree. Following are a few properties of the spanning tree connected to graph G −

* A connected graph G can have more than one spanning tree.
* All possible spanning trees of graph G, have the same number of edges and vertices.
* The spanning tree does not have any cycle (loops).
* Removing one edge from the spanning tree will make the graph disconnected, i.e. the spanning tree is **minimally connected**.
* Adding one edge to the spanning tree will create a circuit or loop, i.e. the spanning tree is **maximally acyclic**.

## Mathematical Properties of Spanning Tree

* Spanning tree has **n-1** edges, where **n** is the number of nodes (vertices).
* From a complete graph, by removing maximum **e - n + 1** edges, we can construct a spanning tree.
* A complete graph can have maximum **nn-2** number of spanning trees.

Thus, we can conclude that spanning trees are a subset of connected Graph G and disconnected graphs do not have spanning tree.

# Application of Spanning Tree

Spanning tree is basically used to find a minimum path to connect all nodes in a graph. Common application of spanning trees are −

* **Civil Network Planning**
* **Computer Network Routing Protocol**
* **Cluster Analysis**

Let us understand this through a small example. Consider, city network as a huge graph and now plans to deploy telephone lines in such a way that in minimum lines we can connect to all city nodes. This is where the spanning tree comes into picture.

## Minimum Spanning Tree (MST)

In a weighted graph, a minimum spanning tree is a spanning tree that has minimum weight than all other spanning trees of the same graph. In real-world situations, this weight can be measured as distance, congestion, traffic load or any arbitrary value denoted to the edges.

## Minimum Spanning-Tree Algorithm

We shall learn about two most important spanning tree algorithms here −

* [Kruskal's Algorithm](https://www.tutorialspoint.com/data_structures_algorithms/kruskals_spanning_tree_algorithm.htm)
* [Prim's Algorithm](https://www.tutorialspoint.com/data_structures_algorithms/prims_spanning_tree_algorithm.htm)

Both are greedy algorithms.

# 

# Prim's Spanning Tree Algorithm

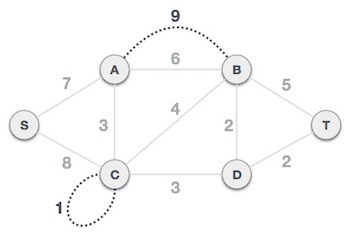
Prim's algorithm to find minimum cost spanning tree (as Kruskal's algorithm) uses the greedy approach. Prim's algorithm shares a similarity with the **shortest path first** algorithms.

Prim's algorithm, in contrast with Kruskal's algorithm, treats the nodes as a single tree and keeps on adding new nodes to the spanning tree from the given graph.

To contrast with Kruskal's algorithm and to understand Prim's algorithm better, we shall use the same example −



## Step 1 - Remove all loops and parallel edges



Remove all loops and parallel edges from the given graph. In case of parallel edges, keep the one which has the least cost associated and remove all others.

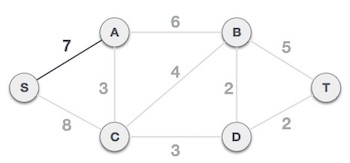


## Step 2 - Choose any arbitrary node as root node

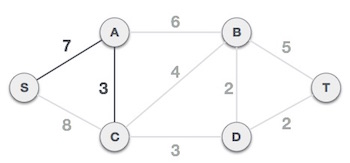
In this case, we choose **S** node as the root node of Prim's spanning tree. This node is arbitrarily chosen, so any node can be the root node. One may wonder why any video can be a root node. So the answer is, in the spanning tree all the nodes of a graph are included and because it is connected then there must be at least one edge, which will join it to the rest of the tree.

## Step 3 - Check outgoing edges and select the one with less cost

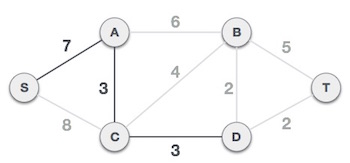
After choosing the root node **S**, we see that S,A and S,C are two edges with weight 7 and 8, respectively. We choose the edge S,A as it is lesser than the other.



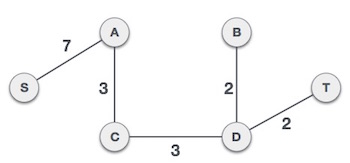
Now, the tree S-7-A is treated as one node and we check for all edges going out from it. We select the one which has the lowest cost and include it in the tree.



After this step, S-7-A-3-C tree is formed. Now we'll again treat it as a node and will check all the edges again. However, we will choose only the least cost edge. In this case, C-3-D is the new edge, which is less than other edges' cost 8, 6, 4, etc.



After adding node **D** to the spanning tree, we now have two edges going out of it having the same cost, i.e. D-2-T and D-2-B. Thus, we can add either one. But the next step will again yield edge 2 as the least cost. Hence, we are showing a spanning tree with both edges included.



We may find that the output spanning tree of the same graph using two different algorithms is same.

**Pseudocode for Prim’s Algorithm:**

int \*MinimumSpanningTree( Graph g, int n, double \*\*costs ) {

Queue q;

int u, v;

int d[n], \*pi;

q = ConsEdgeQueue( g, costs );

pi = ConsPredList( n );

for(i=0;i<n;i++) {

d[i] = INFINITY;

}

**/\* Choose 0 as the "root" of the MST \*/**

d[0] = 0;

pi[0] = 0;

while ( !Empty( q ) ) {

u = Smallest( g );

for each v in g->adj[u] {

if ( (v in q) && costs[u][v] < d[v] ) {

pi[v] = u;

d[v] = costs[u][v];

}

}

}

return pi;

}

The time complexity is O(VlogV + ElogV) = O(ElogV)

Prim's algorithm grows a single tree until it becomes the minimum spanning tree and uses greedy approach.

**Source Code**

from collections import defaultdict

from heapq import \*

def prims( nodes, edges ):

curr = defaultdict( list )

for v1,v2,c in edges:

curr[ v1 ].append( (c, v1, v2) )

curr[ v2 ].append( (c, v2, v1) )

mst = []

used = set( nodes[ 0 ] )

available\_edges = curr[ nodes[0] ][:]

heapify( available\_edges )

while available\_edges:

cost, v1, v2 = heappop( available\_edges )

if v2 not in used:

used.add( v2 )

mst.append( ( v1, v2, cost ) )

for q in curr[ v2 ]:

if q[ 2 ] not in used:

heappush( available\_edges, q )

return mst

nodes =[]

varlist = "ABCDEFGHIJKLMNOPQRSTUVWXYZ"

varlist =list(varlist)

f1 = open(r'inp2\_undirected.txt','r')

a = []

for i in f1.readlines():

p=i.split()

a.append(p)

cost = 0

length = len(a)

no\_of\_vertices =int(a[0][0])

no\_of\_edges = int(a[0][1])

undirected = str(a[0][2])

edges = a[1:length-1]

if(undirected == 'U'):

print("The minimum spanning tree using Prim's Algorithm")

print("The number of Vertices is :", no\_of\_vertices)

print("The number of Edges is :", no\_of\_edges)

for i in range(0,no\_of\_vertices):

nodes.append(varlist[i])

out = prims( nodes, edges )

j= 0

for i in out:

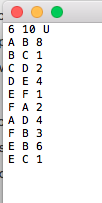
print (i)

cost += int(out[j][2])

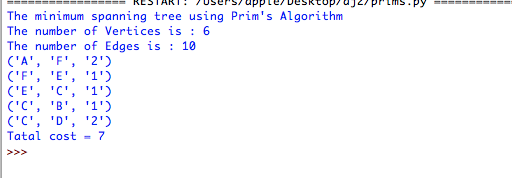
j +=1

print("Tatal cost =" ,cost)

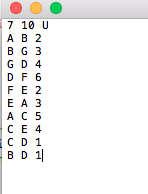
Input file 1:



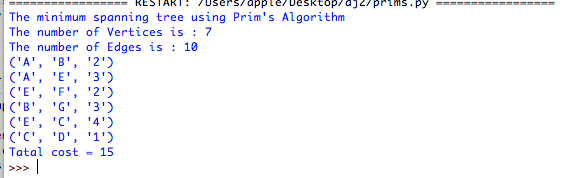
**Output 1:**

****

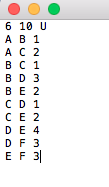
Input file 2:

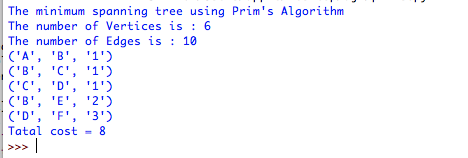


**Output 2 :**

****

File input 3:



**Output 3: **

**Dijkstra's Algorithm**

Djikstra's algorithm (named after its discover, [E.W. Dijkstra](https://www.cs.auckland.ac.nz/software/AlgAnim/e_w_dijkstra.html)) solves the problem of finding the shortest path from a point in a graph (the *source*) to a destination. It turns out that one can find the shortest paths from a given source to *all* points in a graph in the same time, hence this problem is sometimes called the **single-source shortest paths** problem.

The somewhat unexpected result that *all* the paths can be found as easily as one further demonstrates the value of reading the literature on algorithms.

This problem is related to the spanning tree one. The graph representing all the paths from one vertex to all the others must be a spanning tree - it must include all vertices. There will also be no cycles as a cycle would define more than one path from the selected vertex to at least one other vertex. For a graph,

|  |  |  |
| --- | --- | --- |
| **G** = **(V,E)** | where | * **V** is a set of vertices and * **E** is a set of edges. |

Dijkstra's algorithm keeps two sets of vertices:

|  |  |  |
| --- | --- | --- |
| **S** |  | the set of vertices whose shortest paths from the source have already been determined *and* |
| **V-** |  | The remaining vertices |

The other data structures needed are:

|  |  |
| --- | --- |
| **d** | array of best estimates of shortest path to each vertex |
| **pi** | an array of [predecessors](https://www.cs.auckland.ac.nz/software/AlgAnim/dij_pred.html) for each vertex |

The basic mode of operation is:

1. Initialise **d** and **pi**,
2. Set **S** to empty,
3. While there are still vertices in **V-S**,
   1. Sort the vertices in **V-S** according to the current best estimate of their distance from the source,
   2. Add **u**, the closest vertex in **V-S**, to **S**,
   3. **Relax** all the vertices still in **V-S** connected to **u**

#### Relaxation :

The **relaxation** process updates the costs of all the vertices, **v**, connected to a vertex, **u**, if we could improve the best estimate of the shortest path to **v** by including **(u,v)** in the path to **v**.

The relaxation procedure proceeds as follows:

initialise\_single\_source( Graph g, Node s )

for each vertex v in Vertices( g )

g.d[v] := infinity

g.pi[v] := nil

g.d[s] := 0;

This sets up the graph so that each node has no predecessor (**pi[v] = nil**) and the estimates of the cost (distance) of each node from the source (**d[v]**) are infinite, except for the source node itself (**d[s] = 0**).

**Logic for the code and Implementation:**

For all the programs we are taking graph inputs from a text file, as vertex, edges and weights. The data structures used are arrays, lists and heaps.

Runtime for djikstras is O(E lg V) . Every time the main loop executes, one vertex is extracted from the queue. Assuming that there are V vertices in the graph, the queue may contain O(V) vertices. Each pop operation takes O(lg V) time assuming the heap implementation of priority queues. So the total time required to execute the main loop itself is O(V lg V). In addition, we must consider the time spent in the function expand, which applies the function handle\_edge to each outgoing edge. Because expand is only called once per vertex, handle\_edge is only called once per edge. It might call push(v'), but there can be at most V such calls during the entire execution, so the total cost of that case arm is at most O(V lg V). The other case arm may be called O(E) times, however, and each call to increase\_priority takes O(lg V) time with the heap implementation. Therefore the total run time is O(V lg V + E lg V), which is O(E lg V) because V is O(E)assuming a connected graph.

**Pseudocode for Djikstra’s Algorithm:**

(\* Dijkstra's Algorithm \*)

let val q: queue = new\_queue()

val visited: vertexMap = create\_vertexMap()

fun expand(v: vertex) =

let val neighbors: vertex list = Graph.outgoing(v)

val dist: int = valOf(get(visited, v))

fun handle\_edge(v': vertex, weight: int) =

case get(visited, v') of

SOME(d') =>

if dist+weight < d'

then ( add(visited, v', dist+weight);

incr\_priority(q, v', dist+weight) )

else ()

| NONE => ( add(visited, v', dist+weight);

push(q, v', dist+weight) )

in

app handle\_edge neighbors

end

in

add(visited, v0, 0);

expand(v0);

while (not (empty\_queue(q)) do expand(pop(q))

end

**Source code :**

**For directed Dijkstra :**

from collections import deque, namedtuple

# initialize all nodes to infinity

inf = float('inf')

Edge = namedtuple('Edge', 'start, end, cost')

def make\_edge(start, end, cost=1):

return Edge(start, end, cost)

class Graph:

def \_\_init\_\_(self, edges):

incorrect\_edges = [i for i in edges if len(i) not in [2, 3]]

if incorrect\_edges:

raise ValueError('Wrong edges data: {}'.format(incorrect\_edges))

self.edges = [make\_edge(\*edge) for edge in edges]

@property

def vertices(self):

return set(

sum(

([edge.start, edge.end] for edge in self.edges), []) )

def get\_node\_pairs(self, n1, n2, both\_ends=True):

if both\_ends:

node\_pairs = [[n1, n2], [n2, n1]]

else:

node\_pairs = [[n1, n2]]

return node\_pairs

def remove\_edge(self, n1, n2, both\_ends=True):

node\_pairs = self.get\_node\_pairs(n1, n2, both\_ends)

edges = self.edges[:]

for edge in edges:

if [edge.start, edge.end] in node\_pairs:

self.edges.remove(edge)

def add\_edge(self, n1, n2, cost=1, both\_ends=True):

node\_pairs = self.get\_node\_pairs(n1, n2, both\_ends)

for edge in self.edges:

if [edge.start, edge.end] in node\_pairs:

return ValueError('Edge {} {} already exists'.format(n1, n2))

self.edges.append(Edge(start=n1, end=n2, cost=cost))

if both\_ends:

self.edges.append(Edge(start=n2, end=n1, cost=cost))

@property

def neighbours(self):

neighbours = {vertex: set() for vertex in self.vertices}

for edge in self.edges:

neighbours[edge.start].add((edge.end, edge.cost))

return neighbours

def dijkstra(self, source, dest):

global distances

assert source in self.vertices, 'Such source node doesnt exist'

distances = {vertex: inf for vertex in self.vertices}

pre\_node = {

vertex: None for vertex in self.vertices

}

distances[source] = 0

vertices = self.vertices.copy()

while vertices:

present\_node = min(

vertices, key=lambda vertex: distances[vertex])

vertices.remove(present\_node)

if distances[present\_node] == inf:

break

for neighbour, cost in self.neighbours[present\_node]:

other\_path = distances[present\_node] + int(cost)

if other\_path < distances[neighbour]:

distances[neighbour] = other\_path

pre\_node[neighbour] = present\_node

path, present\_node = deque(), dest

while pre\_node[present\_node] is not None:

path.appendleft(present\_node)

present\_node = pre\_node[present\_node]

if path:

path.appendleft(present\_node)

return path,distances

f1 = open(r'inp2.txt','r')

distances = ()

a = []

var =[]

for i in f1.readlines():

p=i.split()

a.append(p)

inp =[]

length = len(a)

no\_of\_vertices =int( a[0][0])

no\_of\_edges = int(a[0][1])

directed = str(a[0][2])

graph = a[1:length-1]

for i in range(1,no\_of\_edges+1):

if(a[i][0] not in inp):

inp.append(a[i][0])

if(directed == 'D'):

print("The number of Vertices is :",no\_of\_vertices)

print("The number of Edges is :",no\_of\_edges)

graph = Graph(graph)

print("The path from the source to every other node in the directed \n")

for q in range(1,len(inp)):

out =graph.dijkstra("A",inp[q])

print(out[0])

print ("The shortest path code from source A to every other node is:\n ")

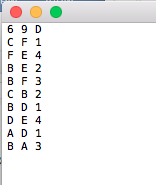
print(distances)

else:

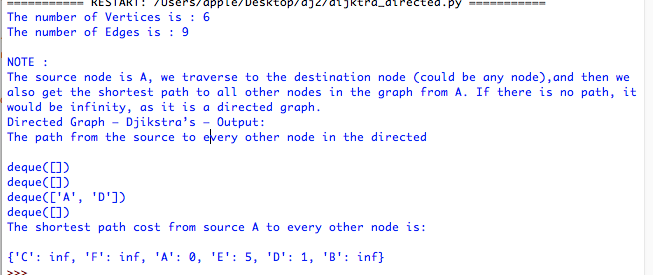
print(" This program is for directed graphs")

exit()

Input file 1:

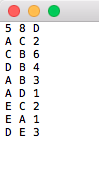


**Output 1:**

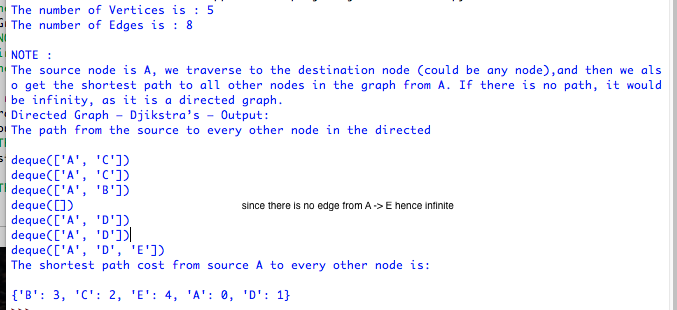


Here, the deque [] means there is no path from source to that particular node.

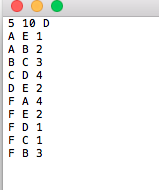
Input file 2:



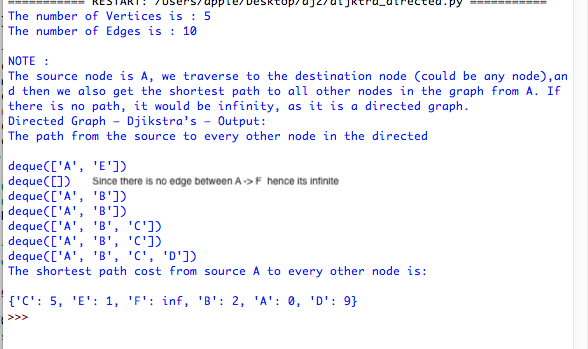
Output 2:



Input file 3:



Output 3:



**For undirected graph :**

from collections import defaultdict

from heapq import \*

def dijkstra(edges, f, t):

g = defaultdict(list)

for l,r,c in edges:

g[l].append((c,r))

q, seen, mins = [(0,f,())], set(), {f: 0}

while q:

(cost,v1,path) = heappop(q)

if v1 not in seen:

seen.add(v1)

path = (v1, path)

print(path,cost)

if v1 == t: return (cost, path)

for c, v2 in g.get(v1, ()):

if v2 in seen: continue

prev = mins.get(v2, None)

next = cost + int(c)

if prev is None or next < prev:

mins[v2] = next

heappush(q, (next, v2, path))

return (dijkstra(edges,t,f))

f1 = open(r'inp2\_undirected.txt','r')

a = []

for i in f1.readlines():

p=i.split()

a.append(p)

inp = []

length = len(a)

no\_of\_vertices =int( a[0][0])

no\_of\_edges = int(a[0][1])

undirected = str(a[0][2])

edges = a[1:length-1]

if(undirected == 'U'):

print("The number of Vertices is :",no\_of\_vertices)

print("The number of Edges is :",no\_of\_edges)

print ("=== Dijkstra ===")

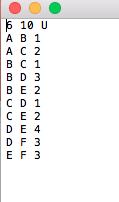
print("The path traversal is:")

out = (dijkstra(edges, "E", "A"))

#out2 = (dijkstra(edges, "A", "D"))

print(out)

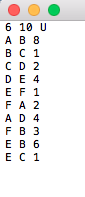
Input file 1:



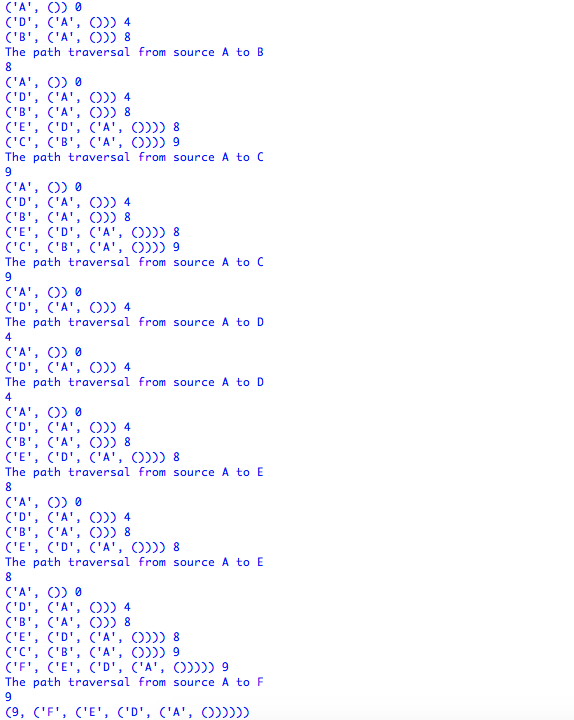
Output 1:



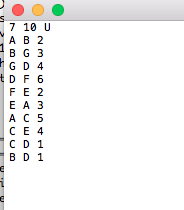
Input file 2:



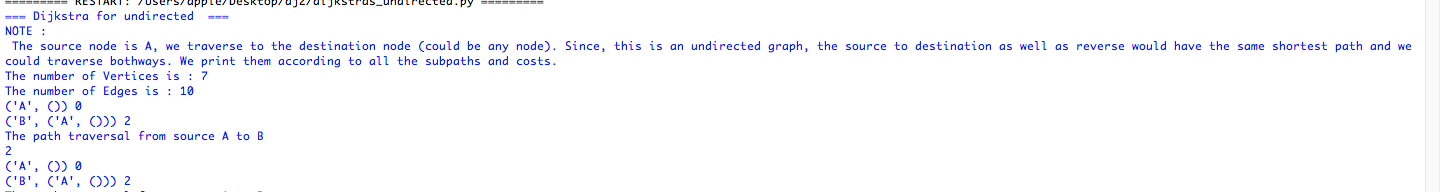
Output2 :

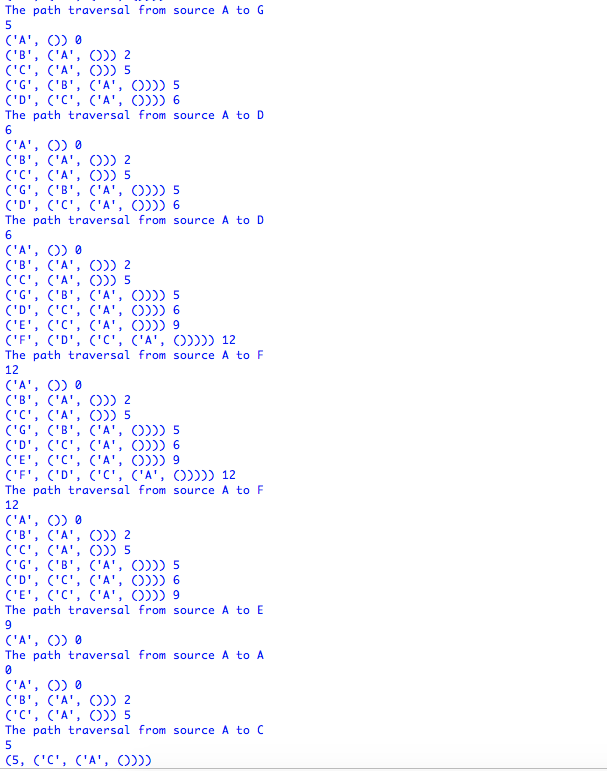


Input file 3:



OUTPUT 3:





CONCLUSION

The Dijkstra’s shortest path algorithm is the most commonly used to solve the single

source shortest path problem today. For a graph G(V, E), where V is the set of vertices

and E is the set of edges, the running time for finding a path between two vertices varies

when different data structure are used. This project uses binary heap to implement

Dijkstra’s algorithm although there are some data structures that may slightly improve the

time complexity.

Prim's (also known as Jarník's) algorithm is a greedy algorithm that finds a minimum spanning tree for a weighted undirected graph. This means it finds a subset of the edges that forms a tree that includes every vertex, where the total weight of all the edges in the tree is minimized. The algorithm operates by building this tree one vertex at a time, from an arbitrary starting vertex, at each step adding the cheapest possible connection from the tree to another vertex.

REFERENCES

1.Geeksforgeeks.com

2. stackoverflow.com

3. Wikipedia.org