

Lab walk-through #7 – Part 2	
Topic	Advanced Graph Algorithms
Week	10 – Session 2
Week	11 – Session 1 & 2

1. Introduction

This is a continuation of Lab 7.1 from last week on graphs and their algorithms. Here, the focus will be on directed graphs, minimum-spanning trees and sorting and shortest-path algorithms for graphs.

2. Lab Objectives

By the end of this lab the student will learn how to implement and study various graphing algorithms and their applications.

3. Lab Setup

Before beginning this lab, you should have:

1. The Eclipse and Java Runtime Environment on your computer
2. Completed the previous searching algorithms lab and its practice questions
3. Read Chapter 4 in your textbook

4. Lab Exercise

Directed graphs have edges that are one-way: the pair of vertices that defines each edge are **ordered**. Many applications are naturally expressed as in terms of directed graphs, such as internet connections, phone calls, paper citations in academia, and stock transactions.

The **outdegree** is the number of edges going from a vertex. The **indegree** of a vertex is the number of edges leading to the vertex. The first vertex in a directed edge is called its **head**, the second is its **tail** and the notation $v \rightarrow w$ is used to represent this relationship. A **directed path** in a digraph is a sequence of vertices in which there is a (directed) edge pointing from each vertex in the sequence to its successor in the sequence. We say that a vertex w is **reachable** from a vertex v if there is a directed path from v to w . Each vertex is reachable from itself.

Part 1: An API for a Digraph

What follows is the API for a directed graph and is almost identical to the undirected graph you came across in walkthrough 7.1.

public class Digraph		
Digraph(int V)		<i>create a V-vertex digraph with no edges</i>
Digraph(In in)		<i>read a digraph from input stream in</i>
int V()		<i>number of vertices</i>
int E()		<i>number of edges</i>
void addEdge(int v, int w)		<i>add edge $v \rightarrow w$ to this digraph</i>
Iterable<Integer> adj(int v)		<i>vertices connected to v by edges pointing from v</i>
Digraph reverse()		<i>reverse of this digraph</i>
String toString()		<i>string representation</i>
API for a digraph		

As you can see, we still use the adjacency list representation we used for undirected graphs, but it is simpler since each edge appears just once. The linked list for each vertex contains all the nodes that can be reached from the vertex with one edge.

Task 1: Implement Digraph.java

Implement the API shown above. Create a Java project called “cas2xb3_lab7_2” and create a class called Digraph.java. Remember to use the Bag.java ADT provided to you on the textbook website. Add the following code to your Digraph.java:

```
public class Digraph{
    private final int V;
    private int E;
    private Bag<Integer>[] adj;

    public Digraph(int V){
        this.V = V;
        this.E = 0;
        adj = (Bag<Integer>[]) new Bag[V];

        for (int v = 0; v < V; v++)
            adj[v] = new Bag<Integer>();
    }

    public int V() { return V; }
    public int E() { return E; }

    public void addEdge(int v, int w){
        adj[v].add(w);
        E++;
    }

    public Iterable<Integer> adj(int v) { return adj[v]; }

    public Digraph reverse(){
        Digraph R = new Digraph(V);
        for (int v = 0; v < V; v++)
            for (int w : adj(v))
                R.addEdge(w, v);
        return R;
    }
}
```

Implement the other functions yourself, or add them in from the previous walkthrough.

As in our previous walkthrough, this is our base implementation, and all other algorithms or applications we develop should be put into separate classes.

Recall the reachability problem from the previous walkthrough. For digraphs, we ask the following question:

Single-source reachability: *Is there a directed path from s to a given target vertex v ?*

The solution to this is exactly the same as the reachability problem we tackled in the previous walkthrough and the API we will use to solve this is shown below.

<code>public class DirectedDFS</code>	
<code>DirectedDFS(Digraph G, int s)</code>	<i>find vertices in G that are reachable from s</i>
<code>DirectedDFS(Digraph G, Iterable<Integer> sources)</code>	<i>find vertices in G that are reachable from sources</i>
<code>boolean marked(int v)</code>	<i>is v reachable?</i>
<i>API for reachability in digraphs</i>	

Task 2: Reachability in a directed graph

Implement the above API to carry out DFS on a digraph. Create a new class called DirectedDFS and add the following code to it (as shown on page 571 of your textbook). You can use their main method to test this implementation. Convince yourself that this works as intended.

```
public class DirectedDFS{
    private boolean[] marked;

    public DirectedDFS(Digraph G, int s){
        marked = new boolean[G.V()];
        dfs(G, s);
    }

    public DirectedDFS(Digraph G, Iterable<Integer> sources){
        marked = new boolean[G.V()];
        for (int s : sources)
            if (!marked[s]) dfs(G, s);
    }

    private void dfs(Digraph G, int v){
        marked[v] = true;
        for (int w : G.adj(v))
            if (!marked[w]) dfs(G, w);
    }

    public boolean marked(int v){ return marked[v]; }
}
```

By adding a second constructor that takes a list of vertices, this API supports for clients the following generalization of the problem, that is, the multiple-source reachability problem. This problem arises in the classic string processing algorithm (section 5.4 in your textbook). This isn't exactly DFS as it just marks the vertices it visits, but this can be substituted for printing if needed.

Take a good look at the trace on page 572 to make sure you understand how DFS works on a digraph.

If job x must be completed before job y, job y before job z, and job z before job x, then someone has made a mistake, because those three constraints cannot all be satisfied. In general, if a precedence-constrained scheduling problem has a directed cycle, then there is no feasible solution. To check for such errors, we need to be able to solve the following problem:

A **directed acyclic graph** (DAG) is a digraph with no directed cycles. We ask the question that is a particular digraph a DAG?

Task 3: Check for cycles in a digraph

```
public class DirectedCycle{
    private boolean[] marked;
    private int[] edgeTo;
    private Stack<Integer> cycle; //vertices on a cycle (if it exists)
    private boolean[] onStack; // vertices on recursive call stack

    public DirectedCycle(Digraph G){
        onStack = new boolean[G.V()];
        edgeTo = new int[G.V()];
    }
}
```

```

        marked = new boolean[G.V()];
        for (int v = 0; v < G.V(); v++)
            if (!marked[v]) dfs(G, v);
    }

    private void dfs(Digraph G, int v){...} ;//explained below

    public boolean hasCycle(){ return cycle != null; }
    public Iterable<Integer> cycle(){ return cycle; }
}

```

This class adds to our standard recursive dfs() a boolean array onStack[] to keep track of the vertices for which the recursive call has not completed. When it finds an edge v->w to a vertex w that is on the stack, it has discovered a directed cycle, which it can recover by following edgeTo[] links.

When executing dfs(G, v), we have followed a directed path from the source to v. To keep track of this path, DirectedCycle maintains a vertex-indexed array onStack[] that marks the vertices on the recursive call stack (by setting onStack[v] to true on entry to dfs(G, v) and to false on exit).

DirectedCycle also maintains an edgeTo[] array so that it can return the cycle when it is detected, in the same way as DepthFirstPaths (page 536) and BreadthFirstPaths (page 540) return paths.

```

private void dfs(Digraph G, int v){
    onStack[v] = true;
    marked[v] = true;

    for (int w : G.adj(v))
        if (this.hasCycle()) return;
        else if (!marked[w]){ edgeTo[w] = v; dfs(G, w); }
        else if (onStack[w]){
            cycle = new Stack<Integer>();
            for (int x = v; x != w; x = edgeTo[x])
                cycle.push(x);
            cycle.push(w);
            cycle.push(v);
        }
    onStack[v] = false;
}

```

Note that this algorithm returns the first cycle it finds rather than all the cycles. In principle there could be an exponential number of them, so we look for only one.

As in the previous walkthrough, the importance of graphs comes in traversing the vertices in a systematic order. The next step is to actually do an appropriate DFS in a particular order, and depending on the application you're interested in, there are three ways you can do this:

1. **Preorder** : Put the vertex on a queue before the recursive calls.
2. **Postorder** : Put the vertex on a queue after the recursive calls.
3. **Reverse postorder** : Put the vertex on a stack after the recursive calls.

Task 4: Implement Depth First Orderings on a given graph

Let us create a new class for this purpose. Call it DepthFirstOrder. It will contain two queues and a stack to carry out these three traversals.

```
private boolean[] marked;  
private Queue<Integer> pre; // vertices in preorder  
private Queue<Integer> post; // vertices in postorder  
private Stack<Integer> reversePost; // vertices in reverse postorder
```

Instantiate these variables in the constructor and call dfs() if there's a vertex that hasn't been visited (marked). Then the implementation of dfs is trivial.

```
private void dfs(Digraph G, int v){  
    pre.add(v);  
    marked[v] = true;  
    for (int w : G.adj(v))  
        if (!marked[w])  
            dfs(G, w);  
    post.add(v);  
    reversePost.push(v);  
}
```

Convince yourself that this is correct by printing out the results. Do this by creating public helper methods as shown below.

```
public Iterable<Integer> pre(){ return pre; }  
public Iterable<Integer> post(){ return post; }  
public Iterable<Integer> reversePost(){ return reversePost; }
```

Task 5: Topological sort

What is meant by Topological sorting?

Given a digraph, put the vertices in order such that all its directed edges point from a vertex earlier in the order to a vertex later in the order (or report that doing so is not possible).

A digraph has a topological order if and only if it is a DAG.

Therefore, in order to do a topological sort, first check if a digraph is a DAG. This is where our previous implementation comes in handy. Go ahead and implement topological sort in a new class called Topological.java and check your answers with the implementation on page 581 in your textbook.

Part 2 – Minimum Spanning Tree

This walkthrough will concentrate on the application of the MST algorithm, and then principles of Minimum Spanning trees will be left as an exercise for you to read up on, starting with page 604 in your textbook. A short introduction is included here for the whole picture.

A **spanning tree** of a graph is a connected sub-graph with no cycles that includes all the vertices. A **minimum spanning tree (MST)** of an edge-weighted graph is a spanning tree whose weight (the sum of the weights of its edges) is no larger than the weight of any other spanning tree. In this section, we examine two classical algorithms for computing MSTs: **Prim's algorithm** and **Kruskal's algorithm**.

MST algorithms are now important in the design of many types of networks (communication, electrical, hydraulic, computer, road, rail, air, and many others) and also in the study of biological, chemical, and physical networks that are found in nature.

There are a few assumptions we make:

1. The graph is connected.
2. The edge weights are not necessarily distances.
3. The edge weights may be 0 or negative.
4. The edge weights are all different.

In summary, we assume throughout the presentation that our job is to find the MST of a connected edge-weighted graph with arbitrary (but distinct) weights.

A **cut** of a graph is a partition of its vertices into two non-empty disjoint sets. A crossing edge of a cut is an edge that connects a vertex in one set with a vertex in the other. Typically, we specify a cut by specifying a set of vertices, leaving implicit the assumption that the cut comprises the given vertex set and its complement, so that a crossing edge is an edge from a vertex in the set to a vertex not in the set.

Now we will provide a principle that allows Prim's and Kruskal's algorithm to work. Read in your textbook for why this works:

Given any cut in an edge-weighted graph, the crossing edge of minimum weight is in the MST of the graph.

We make one final observation before moving onto the implementation: MST algorithms are always greedy. If you don't know what this means, search for the **greedy algorithm paradigm** in any algorithms textbook.

In order to implement the algorithms, we will need an **edge-weighted graph**. We accomplish this by creating an ADT called Edge and adding onto our Graph implementation right in the beginning of walkthrough 7.1.

The API for an Edge ADT looks like this:

```

public class Edge implements Comparable<Edge>
    Edge(int v, int w, double weight)
    double weight()
    int either()
    int other(int v)
    int compareTo(Edge that)
    String toString()

```

initializing constructor
weight of this edge
either of this edge's vertices
the other vertex
compare this edge to e
string representation

API for a weighted edge

Task 6: Implement the Edge and EdgeWeightedGraph API

Both can be found on page 608 of your textbook and you can check your implementation with that on page 610.

Our choice to use explicit Edge objects leads to clear and compact client code (verify this on page 613). It carries a small price: each adjacency-list node has a reference to an Edge object, with redundant information (all the nodes on v's adjacency list have a v). We also pay object overhead cost. Although we have only one copy of each Edge, we do have two references to each Edge object. An alternative and widely used approach is to keep two list nodes corresponding to each edge, just as in Graph, each with a vertex and the edge weight in each list node. This alternative also carries a price—two nodes, including two copies of the weight for each edge.

Task 7: Prim's Algorithm - Lazy version

Our first MST method, known as Prim's algorithm, is to attach a new edge to a single growing tree at each step. Start with any vertex as a single-vertex tree; then add V-1 edges to it, always taking next (colouring black) the minimum weight edge that connects a vertex on the tree to a vertex not yet on the tree (a crossing edge for the cut defined by tree vertices). Remember to import the appropriate libraries when copying the code that follows.

We use three data structures to represent the graph:

1. **Vertices on the tree:** We use a vertex-indexed boolean array marked[], where marked[v] is true if v is on the tree.
2. **Edges on the tree:** We use one of two data structures: a queue mst to collect MST edges or a vertex-indexed array edgeTo[] of Edge objects, where edgeTo[v] is the Edge that connects v to the tree.
3. **Crossing edges:** We use a MinPQ<Edge> priority queue that compares edges by weight.

Implementation-wise, they can be represented using the following code:

```

private boolean[] marked; // MST vertices
private Queue<Edge> mst; // MST edges
private MinPQ<Edge> pq; // crossing (and ineligible) edges

```

Next we have to **maintain the set of crossing edges**. Each time that we add an edge to the tree, we also add a vertex to the tree. To maintain the set of crossing edges, we need to add to the priority queue all edges from that vertex to any non-tree vertex (using marked[] to identify such edges). But we must do more: any edge connecting the vertex just added to a tree vertex that is already on the priority queue now becomes ineligible (it is no longer a crossing edge because it connects two tree vertices).

An eager implementation of Prim's algorithm would remove such edges from the priority queue; we first consider a simpler lazy implementation of the algorithm where we leave such edges on the priority queue, deferring the eligibility test to when we remove them.

However, the question still remains: how do we (efficiently) find the crossing edge of minimal weight?

There are a number of ways of doing this, and we will start with a simplified one first. Create a new class called `LazyPrimMST` and implement Prim's algorithm in the constructor of the class so that the properties of the MST can be explored using other functions.

```
public LazyPrimMST(EdgeWeightedGraph G){
    pq = new MinPQ<Edge>();
    marked = new boolean[G.V()];
    mst = new Queue<Edge>();
    visit(G, 0); // assumes G is connected

    while (!pq.isEmpty()){
        Edge e = pq.delMin(); // Get lowest-weight
        int v = e.either(), w = e.other(v); // edge from pq.

        if (marked[v] && marked[w]) continue; // Skip if ineligible.
        mst.enqueue(e); // Add edge to tree.

        if (!marked[v]) visit(G, v); // Add vertex to tree
        if (!marked[w]) visit(G, w); // (either v or w).
    }
}
```

The inner loop takes an edge from the priority queue and (if it is not ineligible) add it to the tree, and also add to the tree the new vertex that it leads to, updating the set of crossing edges by calling `visit()` with that vertex as argument. The `weight()` method requires iterating through the tree edges to add up the edge weights (lazy approach) or keeping a running total in an instance variable (eager approach).

We use a private method `visit()` that puts a vertex on the tree, by marking it as visited and then putting all of its incident edges that are not ineligible onto the priority queue, thus ensuring that the priority queue contains the crossing edges from tree vertices to non-tree vertices (perhaps also some ineligible edges).

```
private void visit(EdgeWeightedGraph G, int v){
    // Mark v and add to pq all edges from v to unmarked vertices.
    marked[v] = true;
    for (Edge e : G.adj(v))
        if (!marked[e.other(v)]) pq.insert(e);
}
```

Task 8: Prim's Algorithm - Eager version

To improve the LazyPrimMST(call the new class PrimMST.java), we might try to delete ineligible edges from the priority queue, so that the priority queue contains only the crossing edges between tree vertices and non-tree vertices. But we can eliminate even more edges.

The key is to note that our only interest is in the **minimal edge** from each non-tree vertex to a tree vertex. When we add a vertex v to the tree, the only possible change with respect to each non-tree vertex w is that adding v brings w closer than before to the tree. In short, we do not need to keep on the priority queue all of the edges from w to tree vertices—we just need to keep track of the minimum-weight edge and check whether the addition of v to the tree necessitates that we update that minimum (because of an edge v - w that has lower weight), which we can do as we process each edge in v 's adjacency list.

In other words, we maintain on the priority queue just one edge for each non-tree vertex w : the shortest edge that connects it to the tree. Any longer edge connecting w to the tree will become ineligible at some point, so there is no need to keep it on the priority queue.

We now replace our previous three variables with the following:

```
private Edge[] edgeTo; // shortest edge from tree vertex
private double[] distTo; // distTo[w] = edgeTo[w].weight()
private boolean[] marked; // true if v on tree
private IndexMinPQ<Double> pq; // eligible crossing edges
```

So if v is not on the tree but has at least one edge connecting it to the tree, then $\text{edgeTo}[v]$ is the shortest edge connecting v to the tree, and $\text{distTo}[v]$ is the weight of that edge. Also, all such vertices v are maintained on the index priority queue, as an index v associated with the weight of $\text{edgeTo}[v]$. This implies that **the minimum key on the priority queue is the weight of the minimal-weight crossing edge, and its associated vertex v is the next to add to the tree.**

Now add the following code to your class PrimMST.java:

```
public PrimMST(EdgeWeightedGraph G){
    edgeTo = new Edge[G.V()];
    distTo = new double[G.V()];
    marked = new boolean[G.V()];

    for (int v = 0; v < G.V(); v++)
        distTo[v] = Double.POSITIVE_INFINITY;

    pq = new IndexMinPQ<Double>(G.V());
    distTo[0] = 0.0;
    pq.insert(0, 0.0); // Initialize pq with 0, weight 0.

    while (!pq.isEmpty())
        visit(G, pq.delMin()); // Add closest vertex to tree.
}
```

```

private void visit(EdgeWeightedGraph G, int v){
    // Add v to tree; update data structures.
    marked[v] = true;
    for (Edge e : G.adj(v)){
        int w = e.other(v);
        if (marked[w]) continue; // v-w is ineligible.
        if (e.weight() < distTo[w]){
            // Edge e is new best connection from tree to w.
            edgeTo[w] = e;
            distTo[w] = e.weight();

            if (pq.contains(w)) pq.change(w, distTo[w]);
            else pq.insert(w, distTo[w]);
        }
    }
}

```

There are two methods missing here – edges() and weight(). Implement them as an exercise. Can you test this algorithm with client code? Do this now.

Task 8: Kruskal's Algorithm

Prim's algorithm builds the MST one edge at a time, finding a new edge to attach to a single growing tree at each step. Kruskal's algorithm also builds the MST one edge at a time; but, by contrast, it finds an edge that connects two trees in a forest of growing trees. We start with a degenerate forest of V single-vertex trees and perform the operation of combining two trees (using the shortest edge possible) until there is just one tree left: the MST.

To implement this algorithm, we use a priority queue to consider the edges in order by weight, a union-find data structure to identify those that cause cycles, and a queue to collect the MST edges. The weight() method requires iterating through the queue to add the edge weights. Implement this in your class called KruskalMST.java, as shown below. Remember to import the UnionFind and MinPQ libraries found on the textbook website.

```

public class KruskalMST{
    private Queue<Edge> mst;

    public KruskalMST(EdgeWeightedGraph G){
        mst = new Queue<Edge>();
        MinPQ<Edge> pq = new MinPQ<Edge>(G.edges());
        UF uf = new UF(G.V());

        while (!pq.isEmpty() && mst.size() < G.V()-1){
            Edge e = pq.delMin(); // Get min weight edge on pq
            int v = e.either(), w = e.other(v); //and vertices.
            if (uf.connected(v, w))
                continue; //Ignore ineligible edges.
        }
    }
}

```

```

        uf.union(v, w); // Merge components.
        mst.enqueue(e); // Add edge to mst.
    }
}
public Iterable<Edge> edges(){ return mst; }
public double weight() // Complete this.
}

```

This implementation of Kruskal's algorithm uses a queue to hold MST edges, a priority queue to hold edges not yet examined, and a union-find data structure for identifying ineligible edges. The MST edges are returned to the client in increasing order of their weights.

Task 9: Orders of Growth

Think about the order of growth of these algorithms you have implemented. Which one is "better"? Well, it depends on the structure of the graph.

For most graphs, the cost of finding the MST is only slightly higher than the cost of extracting the graph's edges. This rule holds except for huge graphs that are extremely sparse, but the available performance improvement over the best-known algorithms even in this case is a small constant factor, perhaps a factor of 10 at best.

Part 3 – Shortest Path

All the algorithms we have studied so far are small sub-sections of problems in the real world. Or algorithms that give us more insight into how a graph is structured. However, the largest use of graphs in almost any application is to answer the following question:

Find the lowest-cost way to get from one vertex to another.

A **shortest path** from vertex s to vertex t in an edge-weighted digraph is a directed path from s to t with the property that no other such path has a lower weight.

The data structures that we need to represent shortest paths are straightforward:

1. **Edges on the shortest-paths tree:** As for DFS, BFS, and Prim's algorithm, we use a parent-edge representation in the form of a vertex-indexed array `edgeTo[]` of `DirectedEdge` objects, where `edgeTo[v]` is edge that connects v to its parent in the tree (the last edge on a shortest path from s to v).
2. **Distance to the source:** We use a vertex-indexed array `distTo[]` such that `distTo[v]` is the length of the shortest known path from s to v .

Task 10: Dijkstra's Algorithm implementation

Create a class called `DijkstraSP` and start with the following instance variables:

```

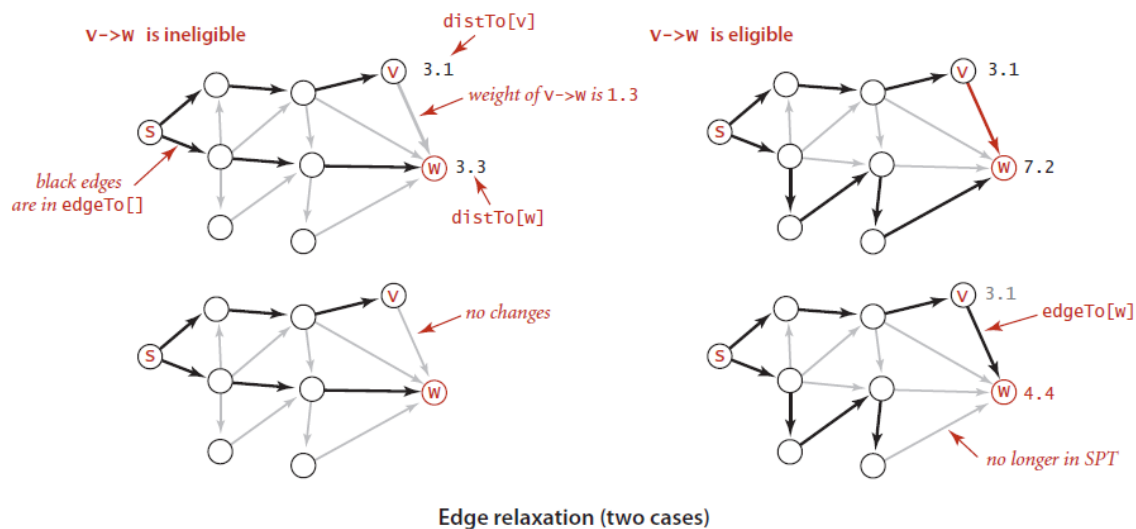
private DirectedEdge[] edgeTo;
private double[] distTo;
private IndexMinPQ<Double> pq;

```

By convention, `edgeTo[s]` is `null` and `distTo[s]` is `0`. We also adopt the convention that distances to vertices that are not reachable from the source are all `Double.POSITIVE_INFINITY`.

Our shortest-paths implementations are based on a simple operation known as **relaxation**. To relax an edge $v \rightarrow w$ means to test whether the best known way from s to w is to go from s to v , then take the edge from v to w , and, if so, update our data structures to indicate that to be the case.

The best known distance to w through v is the **sum** of $\text{distTo}[v]$ and $e.\text{weight}()$ — if that value is not smaller than $\text{distTo}[w]$, we say the edge is **ineligible**, and we ignore it; if it is smaller, we update the data structures.



However this is for a particular edge. We need to iterate through all of the vertices and make sure we relax all edges pointing from a given vertex, as shown below. Add this method to your DijkstraSP.java. Remember to add on the IndexMinPQ library from the textbook website.

```
private void relax(EdgeWeightedDigraph G, int v){
    for(DirectedEdge e : G.adj(v)){
        int w = e.to();
        if (distTo[w] > distTo[v] + e.weight()){
            distTo[w] = distTo[v] + e.weight();
            edgeTo[w] = e;
            if (pq.contains(w)) pq.changeKey(w, distTo[w]);
            else pq.insert(w, distTo[w]);
        }
    }
}
```

Note that any edge from a vertex whose $\text{distTo}[v]$ entry is finite to a vertex whose $\text{distTo}[]$ entry is infinite is eligible and will be added to $\text{edgeTo}[]$ if relaxed. In particular, some edge leaving the source is the first to be added to $\text{edgeTo}[]$.

Add a few more client code methods, as found on page 649. These allow you to make the best use of this class once implemented. Now we are ready to define the constructor of DijkstraSP.

```
public DijkstraSP(EdgeWeightedDigraph G, int s)
{
    edgeTo = new DirectedEdge[G.V()];
    distTo = new double[G.V()];
    pq = new IndexMinPQ<Double>(G.V());

    for (int v = 0; v < G.V(); v++)
        distTo[v] = Double.POSITIVE_INFINITY;
    distTo[s] = 0.0;
    pq.insert(s, 0.0);

    while (!pq.isEmpty())
        relax(G, pq.delMin());
}
```

This implementation of Dijkstra's algorithm grows the SPT by adding an edge at a time, always choosing the edge from a tree vertex to a non-tree vertex whose destination w is closest to s . Does this make sense? What is the order of growth of this algorithm?

This can now help us solve the more generalised problem called **all-pairs shortest paths**:

Given an edge-weighted digraph, support queries of the form Given a source vertex s and a target vertex t , is there a path from s to t ? If so, find a shortest such path (one whose total weight is minimal).

Can you work out how to do this? You just have to iterate over all the vertices in the graph and call the DijkstraSP constructor each time.

Can you work out how to solve the SP problem even faster? By relaxing vertices in topological order, we can solve the single source shortest-paths problem for edge-weighted DAGs in time proportional to $E + V$.

Task 11: Parallel Job scheduling

Consider the following problem:

Parallel precedence-constrained scheduling. *Given a set of jobs of specified duration to be completed, with precedence constraints that specify that certain jobs have to be completed before certain other jobs are begun, how can we schedule the jobs on identical processors (as many as needed) such that they are all completed in the minimum amount of time while still respecting the constraints?*

Thousands or even millions of jobs might be involved, so we require an efficient algorithm. Remarkably, a **linear time** algorithm is available—an approach known as the critical path method demonstrates that the problem is equivalent to a longest paths problem in an edge-weighted DAG. This method has been used successfully in countless industrial applications.

The critical path method for parallel scheduling is to proceed as follows: Create an edge-weighted DAG with a source s , a sink t , and two vertices for each job (a start vertex and an end vertex). For each job, add an edge from its start vertex to its end vertex with weight equal to its duration. For each precedence constraint, $v \rightarrow w$, add a zero-weight edge from the end vertex corresponding to v to the beginning vertex corresponding to w . Also add zero-weight edges from the source to each job's start

vertex and from each job's end vertex to the sink. Now, schedule each job at the time given by the length of its longest path from the source.

We implement this using the Critical Path method for the application for parallel precedence-constrained job scheduling.

Try implementing this problem yourself, but from the code on page 665, this is the part you should focus on:

```
for (int i = 0; i < N; i++){
    String[] a = StdIn.readLine().split("\\s+");
    double duration = Double.parseDouble(a[0]);
    G.addEdge(new DirectedEdge(i, i+N, duration));
    G.addEdge(new DirectedEdge(s, i, 0.0));
    G.addEdge(new DirectedEdge(i+N, t, 0.0));

    for (int j = 1; j < a.length; j++){
        int successor = Integer.parseInt(a[j]);
        G.addEdge(new DirectedEdge(i+N, successor, 0.0));
    }
}
```

This implementation of the critical path method for job scheduling reduces the problem directly to the longest-paths problem in edge-weighted DAGs. It builds an edge-weighted digraph (which must be a DAG) from the job-scheduling problem specification, as prescribed by the critical path method.

The next part uses the implementation of an Acyclic Longest Path implementation (Page 660) to find the longest paths tree and to output the longest paths lengths **which are the start times for each jobs**.

It may not seem like it at first, but this is the best and most efficient solution to this problem. There are other variations that you can find, including the problem with deadlines (Page 667) which you can explore further.

Task 12: Bellman Ford

Returning to the single-source shortest path problem with no negative cycles, we consider the following solution:

Initialize distTo[s] to 0 and all other distTo[] values to infinity. Then, considering the digraph's edges in any order, relax all edges. Make V such passes.

It also solves the problem of finding negative cycles – if that is part of the problem. The implementation we concentrate on is a **queue-based Bellman-Ford** algorithm. This provides a number of advantages. Specifically, we can easily determine a priori that numerous edges are not going to lead to a successful relaxation in any given pass: the only edges that could lead to a change in distTo[] are those leaving a vertex whose distTo[] value changed in the previous pass. To keep track of such vertices, we use a FIFO queue.

With these considerations, the actual implementation is actually quite concise. It is based on two additional data structures:

1. A queue **queue** of vertices to be relaxed
2. A vertex-indexed boolean array **onQ[]** that indicates which vertices are on the queue, to avoid duplicates

Study the implementation below.

```
private double[] distTo; // length of path to v
private DirectedEdge[] edgeTo; // last edge on path to v
private boolean[] onQ; // Is this vertex on the queue?
private Queue<Integer> queue; // vertices being relaxed
private int cost; // number of calls to relax()
private Iterable<DirectedEdge> cycle; //negative cycle in edgeTo?

public BellmanFordSP(EdgeWeightedDigraph G, int s){
    distTo = new double[G.V()];
    edgeTo = new DirectedEdge[G.V()];
    onQ = new boolean[G.V()];
    queue = new Queue<Integer>();

    for (int v = 0; v < G.V(); v++){
        distTo[v] = Double.POSITIVE_INFINITY;

        distTo[s] = 0.0;
        queue.enqueue(s);
        onQ[s] = true;

        while (!queue.isEmpty() && !this.hasNegativeCycle()){
            int v = queue.dequeue();
            onQ[v] = false;
            relax(v);
        }
    }
}
```

We start by putting the source *s* on the queue, and then enter a loop where we take a vertex off the queue and relax it. To add vertices to the queue, we augment our `relax()` implementation from earlier to put the vertex pointed to by any edge that successfully relaxes onto the queue. It also periodically checks for a negative cycle in `edgeTo[]`.

Relax now looks like this after changes:

```
private void relax(EdgeWeightedDigraph G, int v)
{
    for (DirectedEdge e : G.adj(v)){
        int w = e.to();
        if (distTo[w] > distTo[v] + e.weight()){
```



```

        distTo[w] = distTo[v] + e.weight();
        edgeTo[w] = e;

        if (!onQ[w]){
            q.enqueue(w);
            onQ[w] = true;
        }
    }
    if (cost++ % G.V() == 0)
        findNegativeCycle();
}

```

The data structures ensure that: only one copy of each vertex appears on the queue, and that every vertex whose `edgeTo[]` and `distTo[]` values change in some pass is processed in the next pass.

But the question still remains: how do we check for negative cycles? In the following manner:

```

private void findNegativeCycle()
{
    int V = edgeTo.length;
    EdgeWeightedDigraph spt;
    spt = new EdgeWeightedDigraph(V);

    for (int v = 0; v < V; v++)
        if (edgeTo[v] != null)
            spt.addEdge(edgeTo[v]);

    EdgeWeightedCycleFinder cf;
    cf = new EdgeWeightedCycleFinder(spt);
    cycle = cf.cycle();
}

public boolean hasNegativeCycle(){ return cycle != null; }
public Iterable<Edge> negativeCycle(){ return cycle; }

```

To our implementation of Bellman-Ford, we:

1. Add an instance variable `cycle` and a private method `findNegativeCycle()` that sets `cycle` to an iterator for the edges of a negative cycle if one is found (and to null if none is found)
2. Calling `findNegativeCycle()` every V th call to `relax()`

Can you work out how to find a negative cycle?

Task 12: Arbitrage in currency exchange

Lastly, we look at one more example from the textbook as an application of the Bellman-Ford Algorithm.

The problem is as follows: Consider a market for financial transactions that is based on trading commodities.

```
% more rates.txt
5
USD 1 0.741 0.657 1.061 1.005
EUR 1.349 1 0.888 1.433 1.366
GBP 1.521 1.126 1 1.614 1.538
CHF 0.942 0.698 0.619 1 0.953
CAD 0.995 0.732 0.650 1.049 1
```

This table is equivalent to a complete edge-weighted digraph with a vertex corresponding to each currency and an edge corresponding to each conversion rate. An edge $s \rightarrow t$ with weight x corresponds to a conversion from s to t at exchange rate x . Paths in the digraph specify multistep conversions. For example, combining the conversion just mentioned with an edge $t \rightarrow u$ with weight y gives a path $s \rightarrow t \rightarrow u$ that represents a way to convert 1 unit of currency s into xy units of currency u .

For example, we might buy $1,012.206 = 741 \times 1.366$ Canadian dollars with our euros. Note that this gives a better rate than directly converting from U.S. dollars to Canadian dollars. You might expect xy to be equal to the weight of $s \rightarrow u$ in all such cases, but such tables represent a complex financial system where such consistency cannot be guaranteed.

Thus, finding the path from s to u such that the product of the weights is maximal is certainly of interest. Even more interesting is a case where the product of the edge weights is smaller than the weight of the edge from the last vertex back to the first. Can you find this case in the file above?

In our example, suppose that the weight of $u \rightarrow s$ is z and $xyz > 1$. Then cycle $s \rightarrow t \rightarrow u \rightarrow s$ gives a way to convert 1 unit of currency s into more than 1 unit (xyz) of currency s . In other words, we can make a $100(xyz - 1)$ percent profit by converting from s to t to u back to s .

For example, if we convert our 1,012.206 Canadian dollars back to US dollars, we get $1,012.206 \cdot 0.995 = 1,007.14497$ dollars, a 7.14497-dollar profit. That might not seem like much, but a currency trader might have 1 million dollars and be able to execute these transactions every minute, which would lead to profits of over \$7,000 per minute, or \$420,000 per hour!

This situation is an example of an **arbitrage** opportunity that would allow traders to make unlimited profits were it not for forces outside the model, such as transaction fees or limitations on the size of transactions. Even with these forces, arbitrage is plenty profitable in the real world.

But how is this modelled as a shortest-path problem?

The answer is that the arbitrage problem is a negative-cycle-detection problem in edge-weighted digraphs. You can find a proof of this on page 681 of your textbook.

What would the implementation of a solution to this problem look like? Try to implement the algorithm found on page 680. This BellmanFordSP client finds an ***arbitrage opportunity*** in a currency exchange table by constructing a complete-graph representation of the exchange table and then using the Bellman-Ford algorithm to find a negative cycle in the graph.