Greedy Method

- The greedy method is perhaps the most straight forward design technique.
- we can be applied to a wide variety of problems. Most, though not all, of these problems have n inputs and require us to obtain a subset that satisfies some constraints.
- Any subset that satisfies those constraints is called a feasible solution.
- We need to find a feasible solution that either maximizes or minimizes a given objective function.
- A feasible solution that does this is called an optimal solution.

```
Algorithm Greedy(a, n)
    // a[1:n] contains the n inputs.
         solution := \emptyset; // Initialize the solution.
         for i := 1 to n do
6
              x := \mathsf{Select}(a);
              if Feasible(solution, x) then
                   solution := Union(solution, x);
9
10
         return solution;
12
```

Greedy method control abstraction for the subset paradigm

Advantages of the Greedy Approach:

- The greedy approach is easy to implement.
- Typically have less time complexity.
- Greedy algorithms can be used for optimization purposes or finding close to optimization in case of Hard problems.

Disadvantages of the Greedy Approach:

The local optimal solution may not always be globally optimal.

Applications of Greedy Algorithm

- This algorithm is also used to solve the fractional knapsack problem.
- It is used in a job sequencing with a deadline.
- It is used to find the minimum spanning tree using the prim's algorithm or the Kruskal's algorithm.
- It is used to Choose Optimal Merge Patterns.
- It is used in finding the single sourse shortest path.

1.KNAPSACK PROBLEM

- Apply the greedy method to solve the knapsack problem.
- We are given n objects and a knapsack or bag.
- Objects has a weight w_i and the knapsack has a capacity m.
- If a fraction x_i , $0 <= x_i <= 1$, of object i is placed into the knapsack, then a profit of $p_i x_i$ is earned.
- The objective is to obtain a filling of the knapsack that maximizes the total profit earned.
- Since the Knapsack capacity is m, we require the total weight of all chosen objects to be at most m.

Formally, the problem can be stated as

$$\underset{1 \le i \le n}{\text{maximize}} \sum_{1 \le i \le n} p_i x_i \tag{4.1}$$

subject to
$$\sum_{1 \le i \le n} w_i x_i \le m$$
 (4.2)

and
$$0 \le x_i \le 1, \quad 1 \le i \le n$$

$$\tag{4.3}$$

The profits and weights are positive numbers. A feasible solution(or filling) is any set (x_1, x_n) satisfying (4.2) and (4.3) above.

An optimal solution is a feasible solution for which (4.1) is maximized.

Example 4.1 Consider the following instance of the knapsack problem: $n = 3, m = 20, (p_1, p_2, p_3) = (25, 24, 15), \text{ and } (w_1, w_2, w_3) = (18, 15, 10).$

Four feasible solutions are:

	(x_1, x_2, x_3)	$\sum w_i x_i$	$\sum p_i x_i$
1.	(1/2, 1/3, 1/4)	$\overline{16.5}$	24.25
2.	(1, 2/15, 0)	20	28.2
3.	(0, 2/3, 1)	20	31
4.	(0, 1, 1/2)	20	31.5

Of these four feasible solutions, solution 4 yields the maximum profit. we shall soon see, this solution is optimal for the given problem instance.

2. Consider the following instance of the knapsack problem: n = 7, m = 15, (p1, p2, P3,, p7) = (10, 5, 15, 7, 6, 18, 3) and, (w1, w2, w3,, w7) = (2, 3, 5, 7, 1, 4, 1).

3.Objects: 1 2 3 4 5 6 7 Profit (P): 5 10 15 7 8 9 4 Weight(w): 1 3 5 4 1 3 2

W (Weight of the knapsack): 15

n (no of items): 7

4.For the given set of items and knapsack capacity = 60 kg, find the optimal solution for the fractional knapsack problem making use of greedy approach.

ltem	Weight	Value
1	5	30
2	10	40
3	15	45
4	22	77
5	25	90

```
Algorithm GreedyKnapsack(m, n)
//p[1:n] and w[1:n] contain the profits and weights respectively
// of the n objects ordered such that p[i]/w[i] \geq p[i+1]/w[i+1].
//m is the knapsack size and x[1:n] is the solution vector.
    for i := 1 to n do x[i] := 0.0; // Initialize x
    U:=m;
    for i := 1 to n do
        if (w[i] > U) then break;
        x[i] := 1.0; \ U := U - w[i];
    if (i \le n) then x[i] := U/w[i];
```

The Time Complexity of knapsack problem is O(n log n)

2. [0/1 Knapsack] Consider the knapsack problem discussed in this section. We add the requirement that $x_i = 1$ or $x_i = 0$, $1 \le i \le n$; that is, an object is either included or not included into the knapsack. We wish to solve the problem

$$\max \sum_{1}^{n} p_i x_i$$
 subject to $\sum_{1}^{n} w_i x_i \leq m$ and $x_i = 0$ or $1, \ 1 \leq i \leq n$

One greedy strategy is to consider the objects in order of nonincreasing density p_i/w_i and add the object into the knapsack if it fits. Show that this strategy doesn't necessarily yield an optimal solution.

2. Job Sequencing with Deadlines

- Given a set of n jobs. Associated with job i is an integer deadline $d_i > 0$ and a profit $p_i > 0$.
- For any job i the profit p_i is earned iff the job is completed by its deadline.
- To complete a job, one has to process the job on a machine for one unit of time.
- Only one machine is available for processing jobs.
- A feasible solution for this problem is a subset J of jobs such that each job in this subset can be completed by its deadline.
- The value of a feasible solution J is the sum of the profits of the jobs in J, or $\Sigma_{i \in I}$ p_i
- An optimal solution is a feasible solution with maximum value.
- Here again, since the problem involves the identification of a subset, it fits the subset paradigm.

Example 4.2 Let n = 4, $(p_1, p_2, p_3, p_4) = (100, 10, 15, 27)$ and $(d_1, d_2, d_3, d_4) = (2, 1, 2, 1)$. The feasible solutions and their values are:

	feasible	processing	
	${f solution}$	sequence	\mathbf{value}
1.	$(1, \ 2)$	2, 1	110
2.	(1, 3)	1, 3 or 3, 1	115
3.	(1, 4)	4, 1	127
4.	(2, 3)	2, 3	25
5.	(3, 4)	4, 3	42
6.	(1)	1	100
7.	(2)	2	10
8.	(3)	3	15
9.	(4)	4	27

Example 4.3 Let $n = 5, (p_1, \ldots, p_5) = (20, 15, 10, 5, 1)$ and $(d_1, \ldots, d_5) = (2, 2, 1, 3, 3)$. Using the above feasibility rule, we have

J	assigned slots	job considered	action	profit
Ø	none	1	assign to [1, 2]	0
$\{1\}$	$[1, \ 2]$	2	assign to $[0, 1]$	20
$\{1, 2\}$	[0, 1], [1, 2]	3	cannot fit; reject	35
$\{1, 2\}$	[0, 1], [1, 2]	4	assign to $[2, 3]$	35
$\{1, 2, 4\}$	[0, 1], [1, 2], [2, 3]	5	reject	40

The optimal solution is $J = \{1, 2, 4\}$ with a profit of 40.

Below is the greedy algorithm that is always supposed to give an optimal solution to the job sequencing problem.

Step-01:

Sorting of all the given jobs in the decreasing order of their profit.

Step-02:

- Checking the value of the maximum deadline.
- Drawing a Gantt chart such that the maximum time on the Gantt chart is the value of the maximum deadline.

Step-03:

- Picking up the jobs one after the other.
- Adding the jobs on the Gantt chart in such a way that they are as far as possible from 0. This ensures that the job gets completed before the given deadline.

Jobs	J1	Ј2	Ј3	J4	J5	Ј6
Deadlines	5	3	3	2	4	2
Profits	200	180	190	300	120	100

Step-01:

Firstly, we need to sort all the given jobs in decreasing order of their profit as follows.

Jobs	J4	J1	J3	J2	J5	J6
Deadlines	2	5	3	3	4	2
Profits	300	200	190	180	120	100

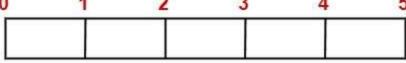
Step-02:

For each step, we calculate the value of the maximum deadline.

Here, the value of the maximum deadline is 5.

So, we draw a Gantt chart as follows and assign it with a maximum time on the Gantt chart with 5 units as shown below.

1 2 3 4 5



Gantt Chart

Now We will be considering each job one by one in the same order as they appear in the Step-01.

We are then supported to place the jobs on the Gantt chart as far as possible from 0.

Step-03:

- We now consider job4.
- Since the deadline for job4 is 2, we will be placing it in the first empty cell before deadline 2 as follows.

Step-04:

Now, we go with job1.

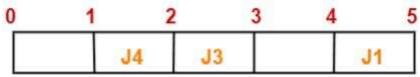
Since the deadline for job1 is 5, we will be placing it in the first empty cell before deadline 5 as shown below

0 1	2	! 3	4	. 5
	J4			J1

J4

Step-05:

- We now consider job3.
- Since the deadline for job3 is 3, we will be placing it in the first empty cell before deadline 3 as shown in the following figure.



J3

Step-06:

- Next, we go with job2.
- Since the deadline for job2 is 3, we will be placing it in the first empty cell before deadline 3.
- Since the second cell and third cell are already filled, so we place job2 in the first cell as shown below.
 0
 1
 2
 3
 4
 5

Step-07:

- Now, we consider job5.
- Since the deadline for job5 is 4, we will be placing it in the first empty cell before deadline 4 as shown in the following figure.

0	1	2	2 3	3 4	5
	J2	J4	J3	J5	J1

- Now, We can observe that the only job left is job6 whose deadline is 2.
- Since all the slots before deadline 2 are already occupied, job6 cannot be completed.

Part-01:

- The optimal schedule is-Job2, Job4, Job3, Job5, Job1
- In order to obtain the maximum profit this is the required order in which the jobs must be completed.

Part-02:

- As we can observe, all jobs are not completed on the optimal schedule.
- This is because job6 was not completed within the given deadline.

Part-03:

Maximum earned profit = Sum of the profit of all the jobs from the optimal schedule

- = Profit of job2 + Profit of job4 + Profit of job3 + Profit of job5 + Profit of job1
- = 180 + 300 + 190 + 120 + 200
- = 990 units

```
Algorithm JS(d, j, n)
\begin{array}{c}2\\3\\4\\5\\6\\7\\8\end{array}
     //d[i] \ge 1, 1 \le i \le n are the deadlines, n \ge 1. The jobs
     // are ordered such that p[1] \geq p[2] \geq \cdots \geq p[n]. J[i]
     // is the ith job in the optimal solution, 1 \le i \le k.
     // Also, at termination d[J[i]] \leq d[J[i+1]], 1 \leq i < k.
          d[0] := J[0] := 0; // Initialize.
          J[1] := 1; // Include job 1.
9
          k := 1:
          for i := 2 to n do
10
11
               // Consider jobs in nonincreasing order of p[i]. Find
12
               // position for i and check feasibility of insertion.
13
14
               r := k:
               while ((d[J[r]] > d[i]) and (d[J[r]] \neq r)) do r := r - 1;
15
               if ((d[J[r]] \leq d[i]) and (d[i] > r)) then
16
17
18
                    // Insert i into J[].
                    for q := k to (r+1) step -1 do J[q+1] := J[q];
19
                    J[r+1] := i; k := k+1;
20
21
22
23
          return k;
24
```

Greedy algorithm for sequencing unit time jobs with deadlines and profits

Analysis of the algorithm:

In the job sequencing with deadlines algorithm, we make use of two loops, one loop within another. Hence, the complexity of this algorithm would be O(n2).

3. Minimum cost Spanning Tree

A **spanning tree** is a subset of an undirected Graph that has all the vertices connected by minimum number of edges. For a graph, there may exist more than one spanning tree.

Properties

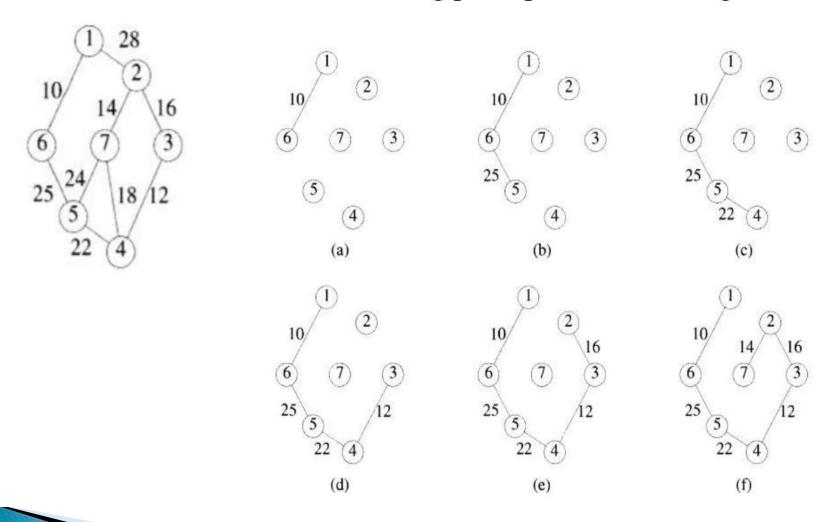
- A spanning tree does not have any cycle.
- · Any vertex can be reached from any other vertex.
- Let G = (V, E) be an undirected connected graph. A subgraph t = (V, E')of G is a spanning tree of G iff t is a tree.
- If there are n number of vertices, the spanning tree should have n - 1 number of edges.
- A Minimum Spanning Tree (MST) is a subset of edges that connects all the vertices together with the minimum possible total edge weight. To derive an MST, Prim's algorithm or Kruskal's algorithm can be used.

a.Prim's algorithm

- Prim's algorithm is a greedy algorithm that starts from one vertex and continue to add the edges with the smallest weight until the goal is reached. The steps to implement the prim's algorithm are given as follows –
- First, we have to initialize an MST with the randomly chosen vertex.
- Now, we have to find all the edges that connect the tree in the above step with the new vertices. From the edges found, select the minimum edge and add it to the tree.
- Repeat step 2 until the minimum spanning tree is formed.

Prim's algorithm

working principle of Prim's Algorithm



```
Algorithm Prim(E, cost, n, t)
23456789
    //E is the set of edges in G. cost[1:n,1:n] is the cost
    // adjacency matrix of an n vertex graph such that cost[i, j] is
    // either a positive real number or \infty if no edge (i, j) exists.
    // A minimum spanning tree is computed and stored as a set of
    // edges in the array t[1:n-1,1:2]. (t[i,1],t[i,2]) is an edge in
        the minimum-cost spanning tree. The final cost is returned.
         Let (k, l) be an edge of minimum cost in E;
10
         mincost := cost[k, l];
         t[1,1] := k; t[1,2] := l;
11
12
         for i := 1 to n do // Initialize near.
              if (cost[i, l] < cost[i, k]) then near[i] := l;
13
14
              else near[i] := k;
         near[k] := near[l] := 0;
15
16
         for i := 2 to n - 1 do
17
         \{ // \text{ Find } n-2 \text{ additional edges for } t. \}
              Let j be an index such that near[j] \neq 0 and
18
              cost[j, near[j]] is minimum;
19
20
              t[i,1] := j; \ t[i,2] := near[j];
              mincost := mincost + cost[j, near[j]];
21
22
              near[j] := 0;
23
              for k := 1 to n do // Update near[].
24
                  if ((near[k] \neq 0) and (cost[k, near[k]] > cost[k, j]))
25
                       then near[k] := j;
26
27
         return mincost;
28
```

The time complexity of the prim's algorithm is $O(V^2)$, where V is the no. of vertices.

b.Kruskal's algorithm

In Kruskal's algorithm, we start from edges with the lowest weight and keep adding the edges until the goal is reached.

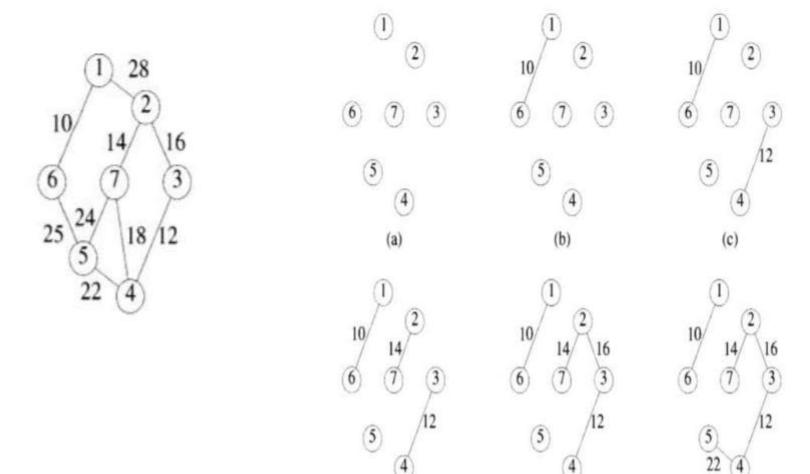
The steps to implement Kruskal's algorithm are listed as follows -

- First, sort all the edges from low weight to high.
- Now, take the edge with the lowest weight and add it to the spanning tree. If the edge to be added creates a cycle, then reject the edge.
- Continue to add the edges until we reach all vertices, and a minimum spanning tree is created.

Kruskal's algorithm

working principle of Kruskal's Algorithm

(f)



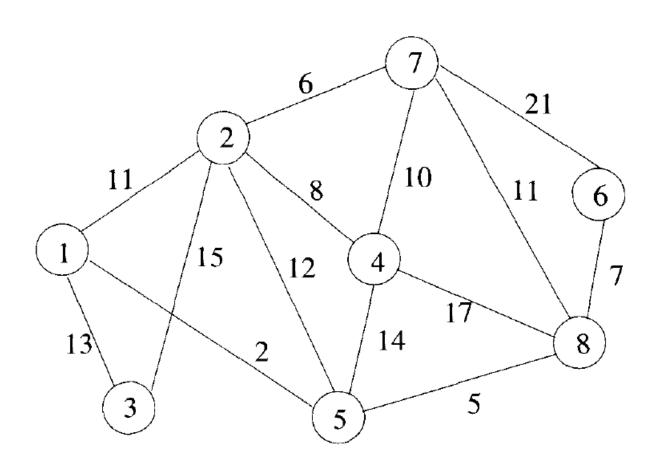
(d)

```
Algorithm Kruskal(E, cost, n, t)
    //E is the set of edges in G. G has n vertices. cost[u,v] is the
2
3
4
5
   // cost of edge (u, v). t is the set of edges in the minimum-cost
    // spanning tree. The final cost is returned.
\mathbf{6}
         Construct a heap out of the edge costs using Heapify;
7
         for i := 1 to n do parent[i] := -1;
8
         // Each vertex is in a different set.
9
         i := 0; mincost := 0.0;
         while ((i < n - 1) and (heap not empty)) do
10
         {
11
             Delete a minimum cost edge (u, v) from the heap
12
             and reheapify using Adjust;
13
             j := Find(u); k := Find(v);
14
15
             if (i \neq k) then
16
                  i := i + 1;
17
                  t[i,1] := u; t[i,2] := v;
18
                  mincost := mincost + cost[u, v];
19
20
                  Union(j,k);
21
22
         if (i \neq n-1) then write ("No spanning tree");
23
24
         else return mincost;
25
```

Time Complexity

The time complexity of Kruskal's algorithm is O(E logV) where E is the no. of edges, and V is the no. of vertices.

1. Compute a minimum cost spanning tree for the graph of Figure using (a) Prim's algorithm and (b) Kruskal's algorithm.



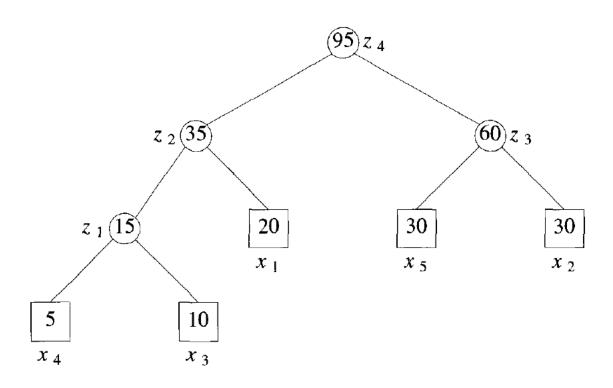
S.No.	Prim's Algorithm	Kruskal's Algorithm
1	This algorithm begins to construct the shortest spanning tree from any vertex in the graph.	This algorithm begins to construct the shortest spanning tree from the vertex having the lowest weight in the graph.
2	To obtain the minimum distance, it traverses one node more than one time.	It crosses one node only one time.
3	The time complexity of Prim's algorithm is $O(V^2)$.	The time complexity of Kruskal's algorithm is O(E log V).
4	In Prim's algorithm, all the graph elements must be connected.	Kruskal's algorithm may have disconnected graphs.
5	When it comes to dense graphs, the Prim's algorithm runs faster.	When it comes to sparse graphs, Kruskal's algorithm runs faster.
6	It prefers list data structure.	It prefers the heap data structure.

4. Optimal Merge Patterns

- Given n sorted files, there are many ways in which to pair wise merge them into a single sorted file. Different pairings require differing amounts of computing time.
- The problem we address ourselves to now is that of determining an optimal way(one requiring the fewest comparisons) pairwise merge n sorted files.
- Since this problem calls for an ordering among the pairs to be merged, it fits the ordering paradigm.

- Greedy attempt to obtain an optimal merge pattern is easy to formulate.
- Since merging an n-record file and an m-record file requires possibly n + m record moves, the obvious choice for a selection criterion is at each Step merge the two smallest size files together.
- The merge pattern such as the one just described will be referred to as a two-way merge pattern (each merge step involves the merging of two files).
- The two-way merge patterns can be represented by binary merge trees.
- The leaf nodes are drawn as squares and represent the given files. These nodes are called external nodes.
- The remaining nodes are drawn as circles and are called internal nodes. Each internal node has exactly two children, and it represents the file obtained by merging the files represented by its two children.
- The number in each node is the length(i.e. the number of records) of the file represented by that node.

we have five files (x1,x2,x3,x4,x5) with sizes(20,30,10,5,30), our greedy rule would generate the following merge pattern.



If d_i is the distance from the root to the external node for file xi and qi, the length of xi is then the total number of record moves for this binary merge tree i n

This sum is called the weighted external path length of the tree.

```
treenode = \mathbf{record} {
           treenode * lchild; treenode * rchild;
           integer weight;
     };
     Algorithm Tree(n)
     // list is a global list of n single node
     // binary trees as described above.
5
           for i := 1 to n - 1 do
6
7
                pt := \mathbf{new} \ treenode; // \text{ Get a new tree node.}
8
                (pt \rightarrow lchild) := \text{Least}(list); // \text{Merge two trees with}
                (pt \rightarrow rchild) := \text{Least}(list); // \text{ smallest lengths.}
9
                (pt \rightarrow weight) := ((pt \rightarrow lchild) \rightarrow weight)
10
                           +((pt \rightarrow rchild) \rightarrow weight);
11
                Insert(list, pt);
12
13
           }
14
           return Least(list); // Tree left in list is the merge tree.
15
```

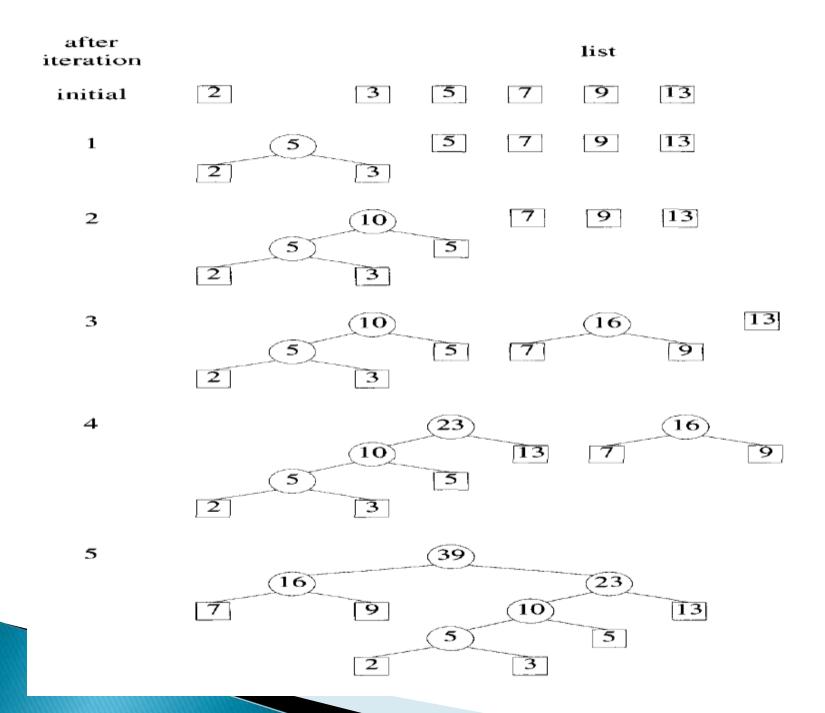
Algorithm to generate a two-way merge tree

- If list is kept in non decreasing order according to the weight value in the roots, then Least(list) requires only 0(1) time and insert(list), can be done in 0(n)time. Hence the total time taken is 0(n2).
- In case list is represented as a minheap in which the root value is less than or equal to the values of its children, then Least(list) and lnsert(list), can be done in O(logn) time. In this case the computing time for Tree is O(nlogn).

1. Find an optimal binary merge pattern for ten files whose lengths are 28, 32, 12, 5, 84, 53, 91, 35, 3, and 11.

Huffman Code

- Another application of binary trees with minimal weighted external path Length is to obtain an optimal set of codes for message M1,M2,..Mn+1.
- Each code is a binary string that is used for transmission of the corresponding message. At the receiving end the code is decoded using a decode tree.
- A decode tree is a binary tree in which external nodes represent messages.



- The binary bits in the code word for a message determine the branching
 - Needed at each level of the decode tree to reach the correct external node.
- For example, if we interpret a zero as a left branch and a one as a right branch, then the decode tree of Figure corresponds to codes 000,001, 01, and 1 for message M1, M2, M3 and M4 respectively.
- These codes are called Huffman codes. The cost of decoding a code word is proportional to the number of bits in the code. This number is equal to the distance of the corresponding external node from the root node. If qi is the relative frequency with which message Mi will be transmitted, then the expected decode time is

$$\sum_{1\leq i\leq n+1}q_id_i$$

where d is the distance of the external node for message Mi from the root node.

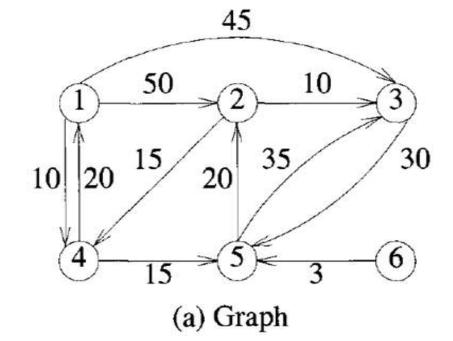
- The expected decode time is minimized by choosing code words resulting in a decode tree with minimal weighted external path length
- Hence the code that minimizes expected decode time also minimizes the expected length of a message.

Obtain a set of optimal Huffman codes for the messages (M_1, \ldots, M_7) with relative frequencies $(q_1, \ldots, q_7) = (4, 5, 7, 8, 10, 12, 20)$. Draw the decode tree for this set of codes.

5.Single Source Shortest Path Problem

- Graphs can be used to represent the highway structure of a state or country with vertices representing cities and edges representing sections of highway.
- The edges can then be assigned weights which may be either the distance between the two cities connected by the edge or the average time to drive along that section of highway.
- A motorist wishing to drive from city A to B would be interested in answers to the following questions:
- Is there a path from A to B
- If there is more than one path from A to B, which is the shortest path?

- The length of a path is now defined to be the sum of the weights of the edges on that path.
- The starting vertex of the path is referred to as the source, and the last vertex the destination.
- The graphs are digraphs to allow for one-way streets.
- In the problem we consider, we are given a directed graph G = (V,E), a weighting function cost for the edges of G, and a source vertex v_o . The problem is to determine the shortest paths from v_o to all the remaining vertices of G.
- It is assumed that all the weights are positive. The shortest path between v_o and some other node v_o is an ordering among a subset of the edges. Hence this problem fits the ordering paradigm.

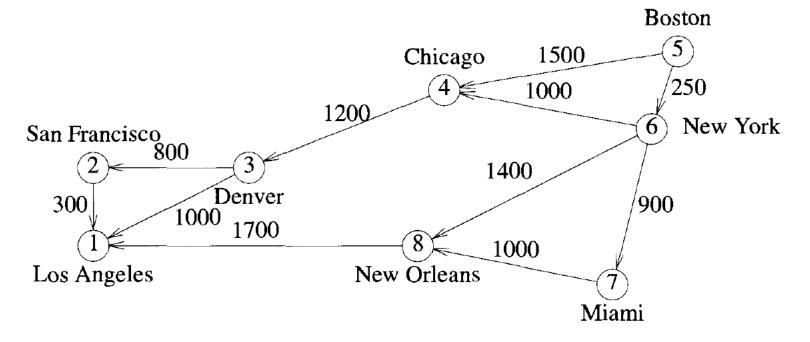


Path	Length			
1) 1,4	10			
2) 1, 4, 5	25			
3) 1, 4, 5, 2	45			
4) 1, 3	45			
(b) Shortest path	s from 1			

Sources	Destinations.							
1	2	3 4		S	6			
T .	50	45	(6)	20	2			
1,4	50	45	10	(25)	00			
1,4,5	45)	45	16	25	00			
1,452	45	45	100	25	P			
1 - 3	45	32.	10	25	20			

```
Algorithm ShortestPaths(v, cost, dist, n)
1
     // dist[j], 1 \le j \le n, is set to the length of the shortest
3
     // path from vertex v to vertex j in a digraph G with n
     // vertices. dist[v] is set to zero. G is represented by its
\begin{array}{c} 5 \\ 6 \\ 7 \\ 8 \end{array}
     // cost adjacency matrix cost[1:n,1:n].
          for i := 1 to n do
          \{ \ // \ \text{Initialize } S.
9
               S[i] := \mathbf{false}; \ dist[i] := cost[v, i];
10
          S[v] := \mathbf{true}; \ dist[v] := 0.0; \ // \ \mathrm{Put} \ v \ \mathrm{in} \ S.
11
12
          for num := 2 to n-1 do
13
14
                // Determine n-1 paths from v.
15
                Choose u from among those vertices not
                in S such that dist[u] is minimum;
16
                S[u] := \mathbf{true}; // \mathrm{Put} \ u \ \mathrm{in} \ S.
17
                for (each w adjacent to u with S[w] = false) do
18
19
                     // Update distances.
                     if (dist[w] > dist[u] + cost[u, w]) then
20
21
                                dist[w] := dist[u] + cost[u, w];
22
23
```

- The time taken by the algorithm on a graph with n vertices is O(n²)
- The running time of Dijkstra's algorithm using Binary min- heap method is O(ElogV)



Iteration S			Distance							
	S	Vertex selected	[1]	SF [2]	DEN [3]	CHI [4]	BOST [5]	NY [6]	MIA [7]	NO [8]
Initial			+∞	+∞	+∞	1500	0	250	+∞	+∞
1	{5}	6	+∞	+∞	+∞	1250	0	250	1150	1650
2	{5,6}	7	+∞	+∞	+∞	1250	0	250	1150	1650
3	{5,6,7}	4	+∞	+∞	2450	1250	0	250	1150	1650
4	{5,6,7,4}	8	3350	+∞	2450	1250	0	250	1150	1650
5	{5,6,7,4,8}	3	3350	3250	2450	1250	0	250	1150	1650
6	{5,6,7,4,8,3}	2	3350	3250	2450	1250	0	250	1150	1650
	{5,6,7,4,8,3,2}	j								

Use algorithm Shortest Paths to obtain in non decreasing order the lengths of the shortest paths from vertex1to all remaining vertices in the digraph

