Trees, Heaps and Graphs

By 106119018, 106119094, 106119140

Trees

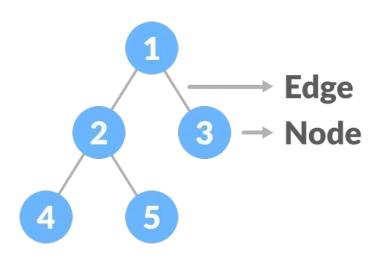
Terminologies

Node

- A node is an entity that contains a key/value/pointer pointing to it
- Nodes are of two types:
- Leaf/external nodes, which are nodes that do not have any child nodes and are the last node in the respective branch of the tree
- Internal nodes, which are nodes that have at least one child node

Edge

 An edge is a link that connects two nodes (parent node to child node)



Terminologies

Root

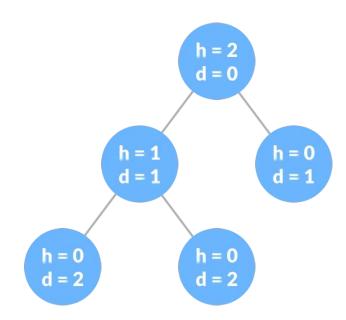
 Root if a node that is the topmost node of the tree, from which the tree starts and originates (i.e., it is the parent most node of the tree)

Height of a node

 Height of a node is the number of from the node to the deepest leaf, i.e., longest path from the node to any leaf node

Depth of a node

 Depth of a node is the number of edges from the root node to the current node



Terminologies

Height of a tree

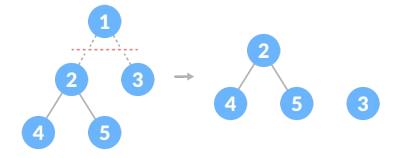
 Height of the tree is the maximum number of edges between the root node to any leaf node

Degree of a node

 It is the total number of edges connected connecting to the node's children (i.e., the number of children)

Forest

A forest is a collection of disjoint trees



Tree traversal

Traversing a tree means to traverse through every node of a tree in a particular order

There are 4 common ways of tree traversal each with a different order of traversal

- Pre-order traversal
- Inorder traversal
- Post-order traversal
- Level order traversal

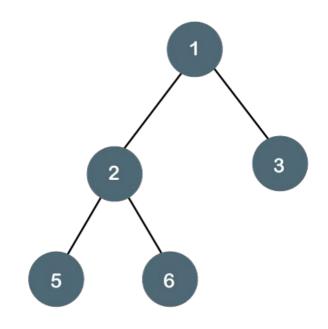
These methods are implemented with the help of different data structures like stacks, queues and arrays

Pre-order tree traversal

In pre-order traversal, the parent node is visited first, then the left child and then the right child

Order: PLR

If we were to print every node in the following tree based on pre-order traversal, the order will be:

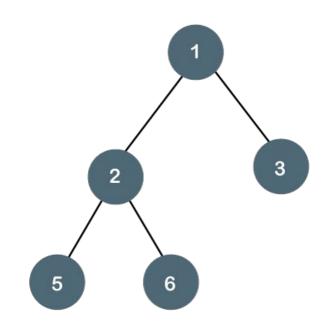


Inorder tree traversal

In inorder traversal, the left child is visited first, then the parent node and finally the right child

Order: LPR

If we were to print every node in the following tree based on inorder traversal, the order will be:

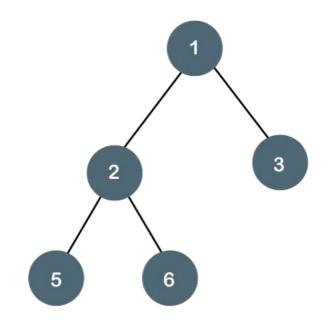


Post-order tree traversal

In post-order traversal, the left child is visited first, then the right child and then at the end the parent node

Order: LRP

If we were to print every node in the following tree based on post-order traversal, the order will be:

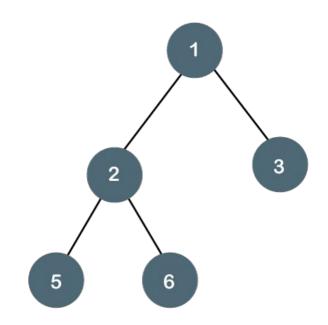


Level-order tree traversal

In level-order traversal, we traverse the node based on the depth

Nodes in the same depth are visited together, with the leftmost node being visited first

If we were to print every node in the following tree based on level-order traversal, the order will be:



Different types of trees

Here are the following types of trees which will be explained more in the later slides

- Binary trees
- Binary search trees
- Self-balancing binary trees

Ideally a tree can have any number of children but we will be focusing on trees with nodes having at most 2 children

Binary tree

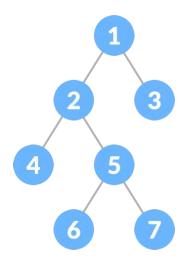
These trees are the type of trees that have at max 2 children

There are different types of binary trees

- Full binary tree
- Perfect binary tree
- Complete binary tree
- Skewed binary tree
- Balanced binary tree

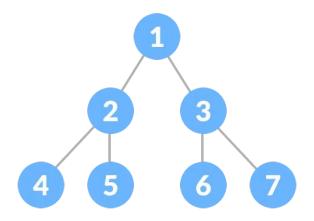
Full binary tree

Full binary trees are trees in which every node has either 0 or 2 child nodes



Perfect binary tree

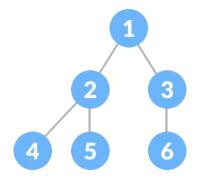
It is a type of binary tree where every internal node has exactly 2 child nodes and all the leaf nodes are at the same level/depth



Complete binary tree

A complete binary tree is one which:

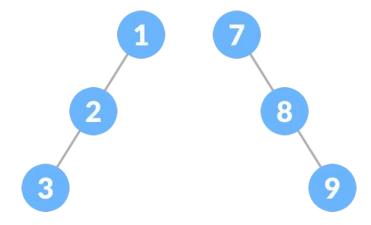
- Every level apart from the leaf node's level must be filled
- All leaf nodes must be filled from the left.
- Last leaf node may not have a right sibling (it need not be full binary tree)



Skewed binary tree

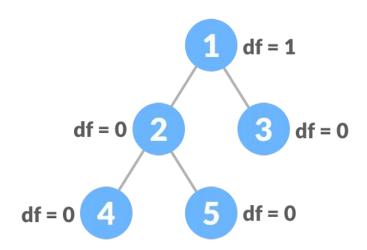
A skewed binary tree is one in which every node of tree has only left child nodes or right child nodes

There can be of two types, left skewed and right skewed



Balanced binary tree

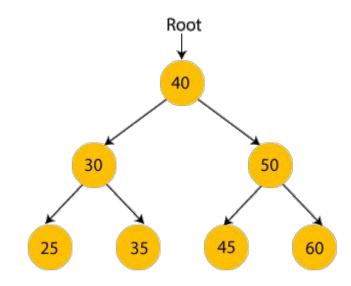
A balanced binary tree is one in which every node's left child (left subtree) and right child (right subtree) has a height difference of not more than 1



Binary search tree

Binary search tree is a specific instance of a binary tree which maintains the tree in a particular order, such that the nodes on the left subtree contain a value lesser than the parent node's value and the right subtree contains a value more than the parent node's value.

An inorder traversal of a BST always prints the values of a tree in ascending order



BST - search operation

Search operation is such that, given a value X:

```
if (X == node.value) return true;
else if (X < node.value) searchInLeftSubtree();
else searchInRightSubtree();
```

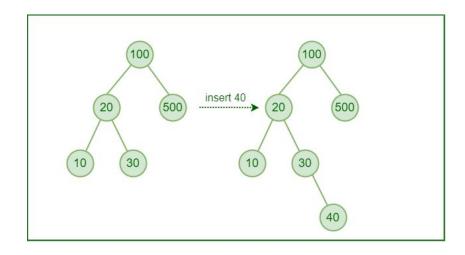
On average, this takes a o(log(n)) time where n is the number of nodes

At worst case it can still take o(n) time if the BST is skewed

BST - insertion

Given a node containing value **X**, we have to insert it in a BST tree such that we maintain the property of a binary search tree

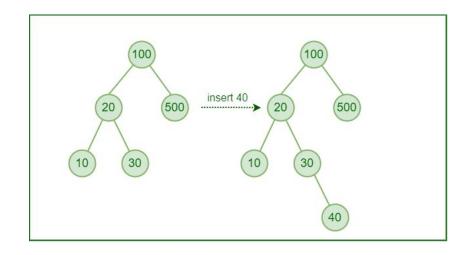
While inserting **X**, we compare the value with the current node's value, if it is more, we move it to the right subtree else move it to the left subtree if it is lesser



BST - insertion

We will finally come to a leaf node (let's say **A**), where we once again compare the value of the node with **A**. We insert the new node as a left child to **A** if the value of our node is lesser than **A**'s and right if it's value is more than **A**'s.

This on average takes o(logn) and at worst case can take o(n) if the BST is skewed.



BST - deletion

Deletion of BST is a two step operation, one is performing the search operation to find the node and the other is the actual deletion of the node

 Case 1: node is a leaf node (no. of children = 0): in this case, deletion is simple, the leaf node simply gets deleted

BST - deletion

Case 2: no. of children of the node = 1

In this case the child node replaces the current node as a child to the current node's parent

```
50 50

/ \ delete(30) / \

30 70 ---> 40 70

\ \ / \ / \

40 60 80 60 80
```

BST - deletion

 Case 3: no. of children of the node = 2

In this case the we take the current node's inorder successor and copy its contents to the node to be deleted, and delete the selected node

On average, this operation takes o(logn) but on the worst case it can take o(n) if the tree is skewed

```
50 60

/ \ delete(50) / \

40 70 ---> 40 70

/\ \

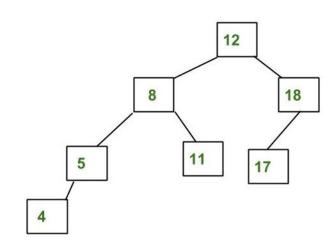
60 80 80
```

Self-balancing trees (AVL)

As seen before, operations on a simple BST can take o(n), rendering the main property of a BST useless

We will now look at self balancing trees and focus on AVL tree

Self balancing trees are trees in which, on insertion and deletion, the tree will always remain a balanced binary search tree



Rotation in AVL trees

```
y x
/ \ Right Rotation / \
x T3 ----> T1 y
/ \ <----- / \
T1 T2 Left Rotation T2 T3
```

Here we see that the property of a BST is not violated

AVL - insertion

- First we perform insertion like BST, let this be w
- We then go up from w and find the first unbalanced node z
- Let y be the child of z that comes in the way till y
- We arrive at 4 cases to rebalance the tree

LL case

 ${\bf y}$ is the left subchild of ${\bf z}$ and ${\bf x}$ is the left subchild of ${\bf y}$

```
T1, T2, T3 and T4 are subtrees.

z
y
/\
y T4 Right Rotate (z) x z
/\
x T3 T1 T2 T3 T4
/\
T1 T2
```

LR case

y is the left subchild of z and x is the right subchild of y

```
Left Rotate (y) x
                                 T4 Right Rotate(z)
Z
                               T3
                                                  T1 T2
T3
   T4
     T3
                            T2
 T2
```

RR case

 \mathbf{y} is the right subchild of \mathbf{z} and \mathbf{x} is the right subchild of \mathbf{y}

```
T1
           Left Rotate(z)
   T2
                                  T2 T3
     T3
        T4
```

RL case

y is the right subchild of z and x is the left subchild of y

```
z z x /\
/\ T1 y Right Rotate (y) T1 x Left Rotate(z) z y
/\ -----> /\ ----> /\ /\
x T4 T2 y T1 T2 T3

T4
/\ T2 T3 T3 T4
```

AVL - deletion

- Likewise, we perform normal BST deletion operation first
- We then try to find the first unbalanced node by getting balancing factor (height of left subtree - height of right subtree)
- If balancing factor > 1, its a LL or LR case
 - If balancing factor of left subtree of unbalanced node is >= 0, then its a LL case, or else it is LR
- If balancing factor < -1, its a RR or RL case
 - If balancing factor of right subtree of unbalanced node is <= 0, then its a RR case, else it is RL

AVL - search

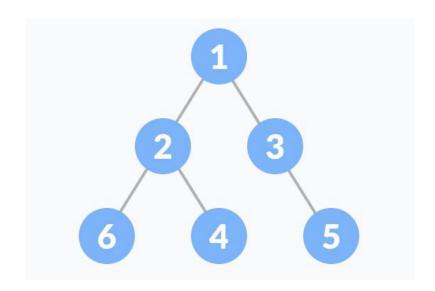
The search operation for an AVL tree is similar to a BST

The insertion, deletion and search operation in an AVL tree will always have a worst case time complexity of o(logn)

Gate question

Which of the following matches the given tree?

- a) Not a full BT, not a complete BT
- b) Full BT, not a complete BT
- c) Not a full BT, is a complete BT
- d) Is a full BT and complete Bt

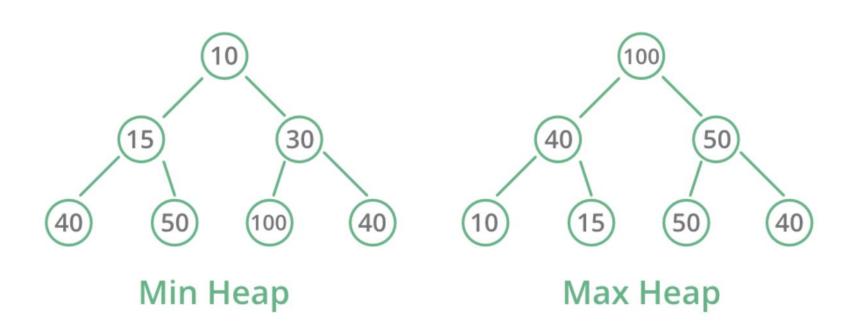


Binary Heaps

WHAT IS A BINARY HEAP?

- A binary heap is a **complete binary tree** which satisfies the heap ordering property.
- A complete binary tree is a binary tree in which every level, except possibly the last, is completely filled, and all nodes in the last level are as far left as possible.
- This heap ordering property is either a max or min-heap property
- Since a heap is a complete binary tree, it has a smallest possible height a heap with N nodes always has O(log N) height.

Heap Data Structure



GATE 2011 QUESTION

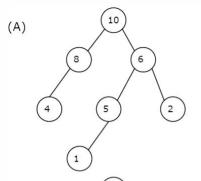
Q. A max-heap is a heap where the value of each parent is greater than or equal to the values of its children. Which of the following is a max-heap?

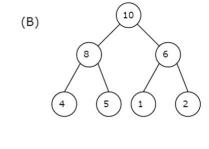


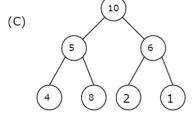
(B) B

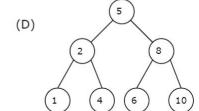
(C) C

(D) D









GATE 2011 QUESTION

Q. A max-heap is a heap where the value of each parent is greater than or equal to the values of its children. Which of the following is a max-heap?

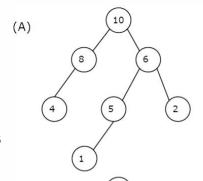
(A) A

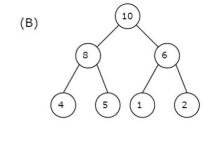
(B) B

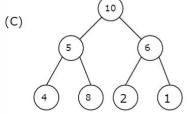
(C) C

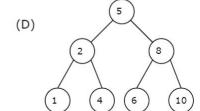
(D) D

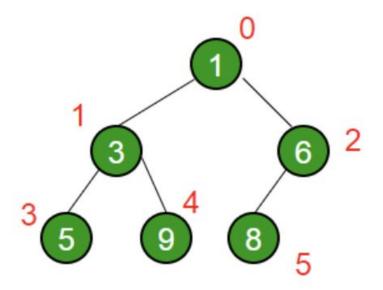
☐ Answer (B)

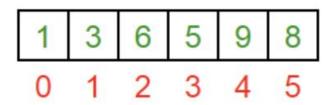












REPRESENTING A BINARY HEAP

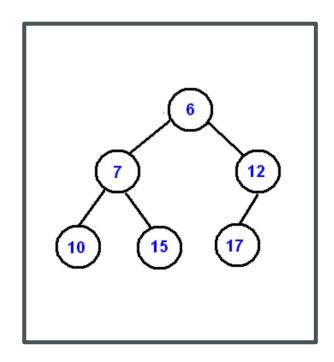
- It is represented as an array by uniquely storing its level order traversal
- Relation between nodes:
- Left child: (2*i)+1
- Right child : (2*i)+2
- Parent : (i-1)/2

WEAKLY ORDERED

- We know how a <u>binary</u> <u>search</u> <u>tree</u> is developed with lesser keys to the left; greater keys to the right as we descend
- But for nodes in a heap, we don't have this strong ordering and this can cause us some difficulty.
- Hence No Convenient Search Mechanism
- Makes it less efficient for deletions

WEAKLY ORDERED

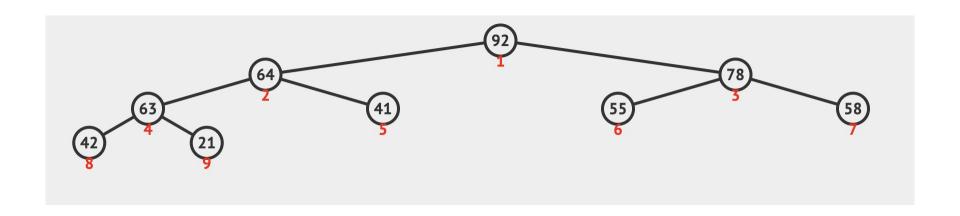
- As you see from the picture, there is no particular relationship among nodes on any given level, even among the siblings.
- ► Sufficient Ordering: Yet, there is 'sufficient ordering' to allow
 - quick removal (yes, a delete) of the <u>maximum</u> node and
 - **fast insertion** of new nodes.
- These are the operations that are mostly used when dealing with this data structure



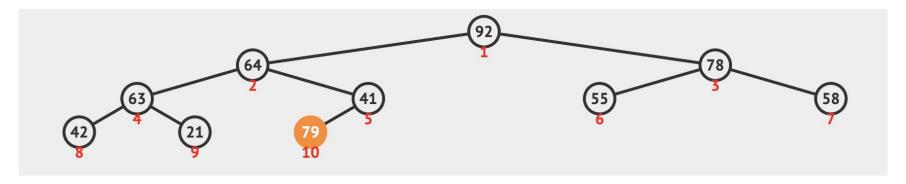
INSERT OPERATION

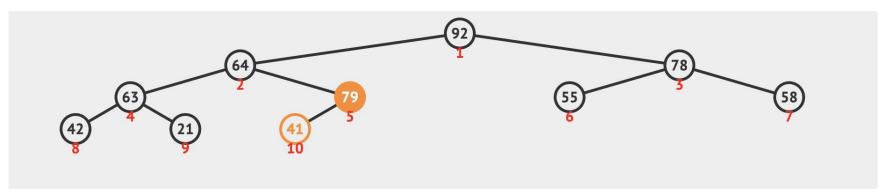
- First increase the heap size by 1, so that it can store the new element.
- Insert the new element at the end of the Heap.
- This newly inserted element may distort the properties of Heap for its parents. So, in order to keep the properties of Heap, **heapify** this newly inserted element following a bottom-up approach.
- This process is called "percolation up".
- The comparison is repeated until the parent satisfies the heap property with percolating element.

INSERT OPERATION – INITIAL HEAP

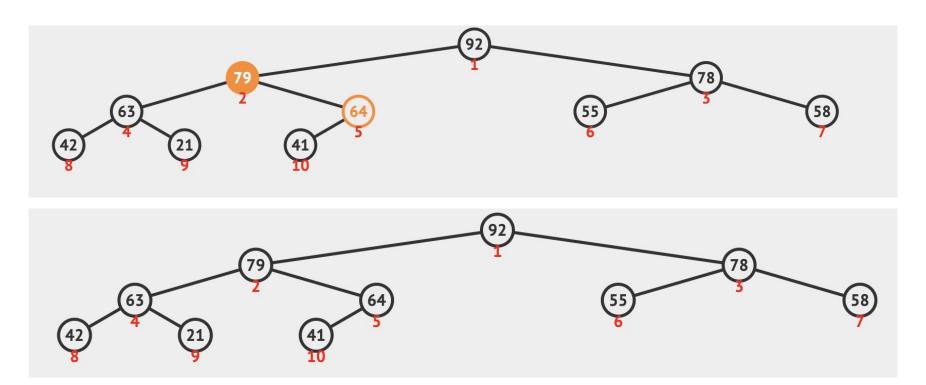


INSERT OPERATION – INSERTING 79 TO LAST POSITION





INSERT OPERATION – INSERTING 79 TO LAST POSITION



INSERT OPERATION CODE – O(LOGN) TIME

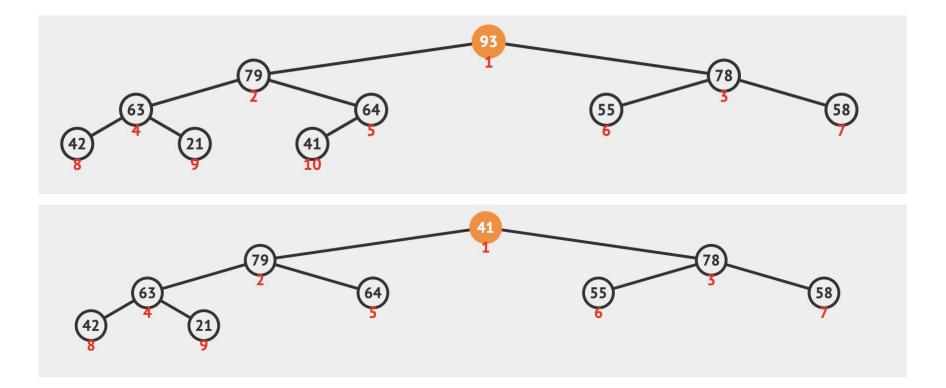
```
void insertNode(int arr[], int& n, int Key)
    n = n + 1;
    arr[n-1] = Key;
    heapify(arr, n, n - 1);
```

```
void heapify(int arr[], int n, int i)
16
         int parent = (i - 1) / 2;
         if (arr[parent] > 0) {
19
             if (arr[i] > arr[parent]) {
                  swap(arr[i], arr[parent]);
                  heapify(arr, n, parent);
28
```

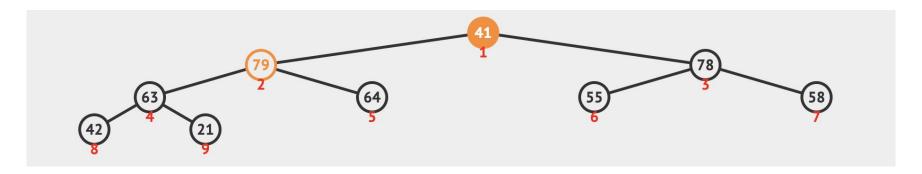
DELETE OPERATION

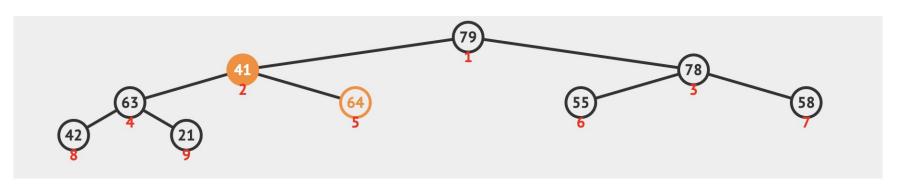
- If we delete the element from the heap, we consider the case where it deletes the root element of the tree as this is an important operation
- Since we delete the root element from the heap it will distort the properties of the heap so we need to perform heapify operations so that it maintains the property of the heap.
- We remove the root and replace it with the last element of the heap and then restore the heap property by *percolating down*. Similar to insertion, the worst-case runtime is O{log n).

DELETE OPERATION – EXTRACTING MAX NODE

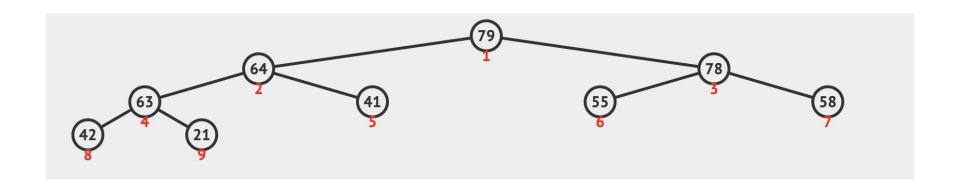


DELETE OPERATION – EXTRACTING MAX NODE





DELETE OPERATION – UPDATED HEAP



DELETE OPERATION CODE – O(LOGN) TIME

```
∨ void deleteRoot(int arr[], int& n)
         int lastElement = arr[n - 1];
         arr[0] = lastElement;
42
         n = n - 1;
         heapify(arr, n, 0);
```

```
void heapify(int arr[], int n, int i)
    int largest = i; // Initialize largest as root
    int l = 2 * i + 1; // left = 2*i + 1
    int r = 2 * i + 2; // right = 2*i + 2
    if (l < n && arr[l] > arr[largest])
        largest = l;
    if (r < n && arr[r] > arr[largest])
        largest = r;
    if (largest != i) {
        swap(arr[i], arr[largest]);
        heapify(arr, n, largest);
```

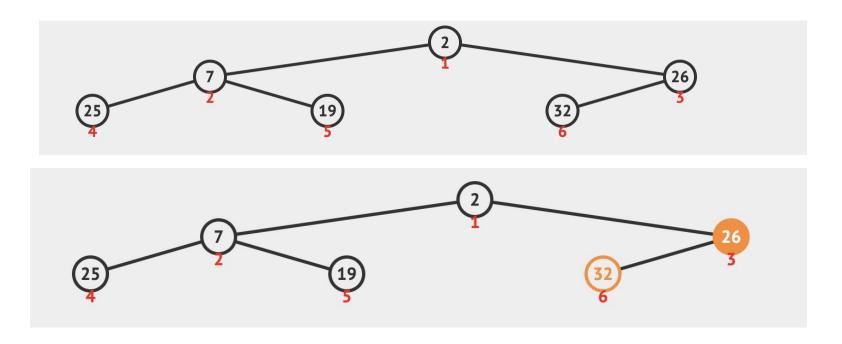
BUILD HEAP – USING HEAPIFY UP

- Start at the top of the heap (the beginning of the array) and heapify bottom-up.
- At each step, the previously sifted items (the items before the current item in the array) form a valid heap, and sifting the next item up places it into a valid position in the heap.
- After sifting up each node, all items satisfy the heap property.
- This approach is more expensive as deeper leveled nodes require more swaps moving upwards and no. of nodes increase at deeper levels
- Time complexity is O(N * logN) similar to repeatedly inserting elements

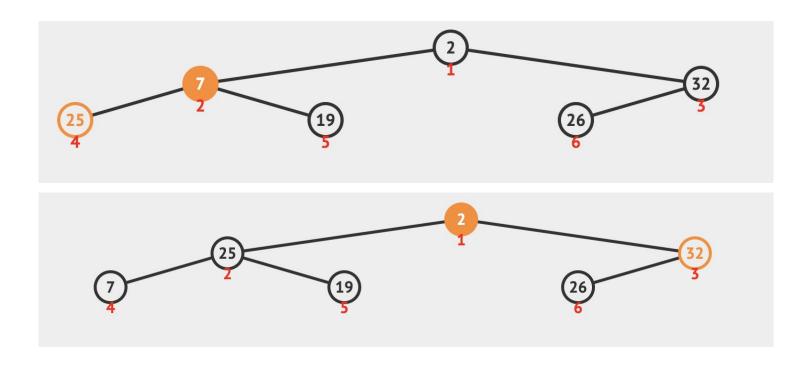
BUILD HEAP - EFFICIENT USING HEAPIFY DOWN

- This approach is optimized by observing the fact that the leaf nodes need not to be heapified as they already follow the heap property.
- Also, the array representation of the complete binary tree contains the level order traversal of the tree.
- So the idea is to find the position of the last non-leaf node and perform the heapify operation of each non-leaf node in reverse level order.
- Here each heapify operation swaps the element downwards till it satisfies the heap property
- This operation hence is more expensive at higher levels and is more efficient as at higher levels the no. of nodes greatly reduces
- Time complexity is O(N)

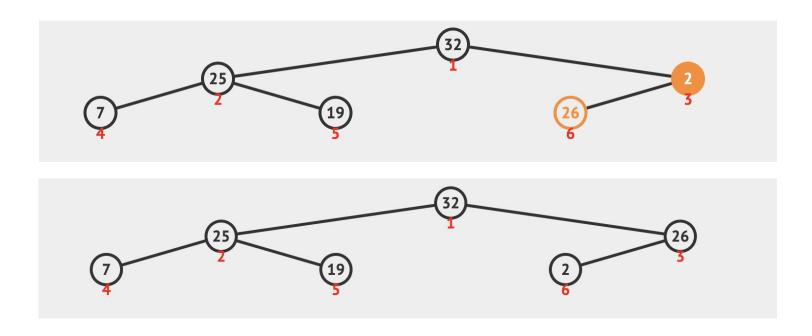
BUILD HEAP - 2,7,26,25,19,32



BUILD HEAP - 2,7,26,25,19,32



BUILD HEAP - 2,7,26,25,19,32

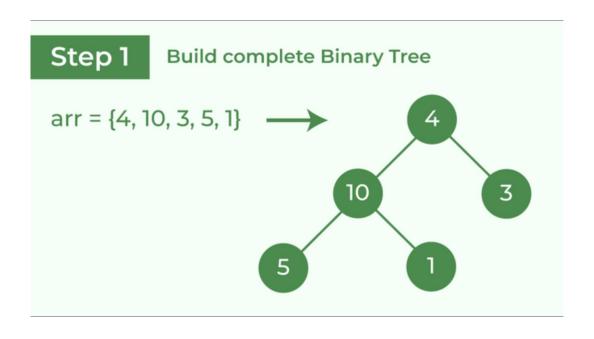


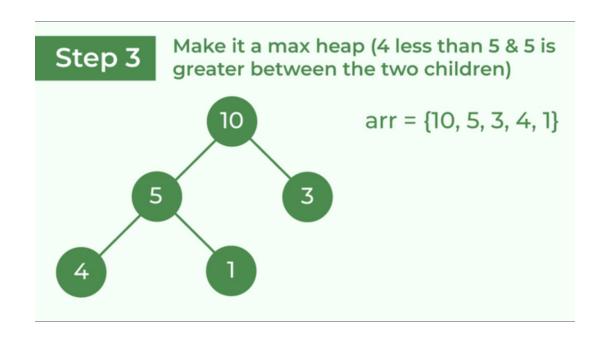
BUILD HEAP – CODE O(N)

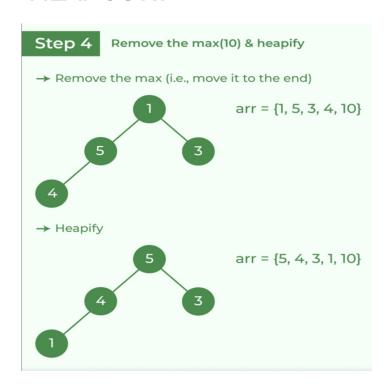
```
void buildHeap(int arr[], int N)
{
    // Index of last non-leaf node
    int startIdx = (N / 2) - 1;

    // Perform reverse level order traversal
    // from last non-leaf node and heapify
    // each node
    for (int i = startIdx; i >= 0; i--) {
        heapify(arr, N, i);
    }
}
```

```
/oid heapify(int arr[], int N, int i)
   int largest = i; // Initialize largest as root
   int l = 2 * i + 1; // left = 2*i + 1
   int r = 2 * i + 2; // right = 2*i + 2
   if (l < N && arr[l] > arr[largest])
       largest = l;
   if (r < N && arr[r] > arr[largest])
       largest = r;
   if (largest != i) {
       swap(arr[i], arr[largest]);
       heapify(arr, N, largest);
```



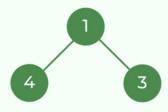




Step 5

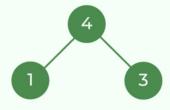
Remove the current max(5) & heapify

→ Remove the max (i.e., move it to the end)



 $arr = \{1, 4, 3, 5, 10\}$

→ Heapify

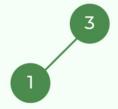


 $arr = \{4, 1, 3, 5, 10\}$

Step 6

Remove the current max(4) & heapify

→ Remove the max (i.e., move it to the end)



 $arr = {3, 1, 4, 5, 10}$

It is already in max heap form

Step 7

Remove the max(3)

 $arr = \{1, 3, 4, 5, 10\}$

The array is now sorted

HEAP SORT - CODE O(NLOGN)

```
void heapSort(int arr[], int N)
    for (int i = N / 2 - 1; i >= 0; i--)
        heapify(arr, N, i);
    for (int i = N - 1; i > 0; i--) {
        swap(arr[0], arr[i]);
        heapify(arr, i, 0);
```

```
void heapify(int arr[], int N, int i)
    int largest = i;
    int l = 2 * i + 1;
    int r = 2 * i + 2;
    if (l < N && arr[l] > arr[largest])
        largest = l;
    if (r < N && arr[r] > arr[largest])
        largest = r;
    if (largest != i) {
        swap(arr[i], arr[largest]);
        heapify(arr, N, largest);
```

GATE 2021 Question

Let H be a binary min-heap consisting of n elements implemented as an array. What is the worst case time complexity of an optimal algorithm to find the maximum element in H?

- A. $\Theta(1)$
- B. $\Theta(\log n)$
- C. $\Theta(n)$
- D. $\Theta(n \log n)$

Answer

In a min heap, maximum element is present in one of the leaf nodes.

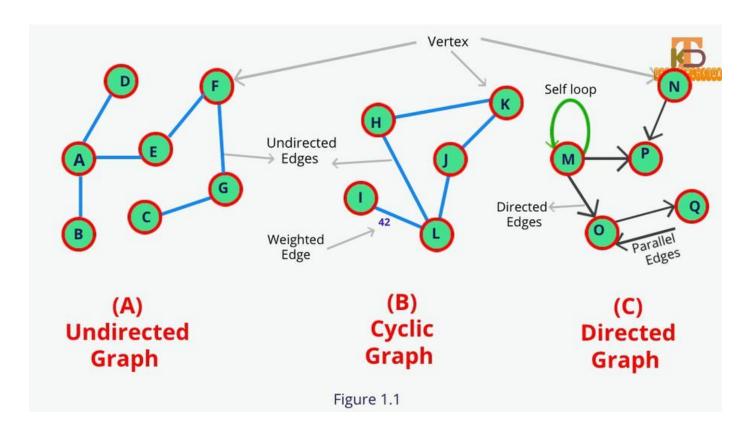
If index of heap starts with 1, indices of leaves are $\lfloor n/2 \rfloor + 1, \lfloor n/2 \rfloor + 2, \lfloor n/2 \rfloor + 3 \ldots n$.

So, we have to perform linear search on at most n/2+1 elements to find the maximum element. (we cannot perform binary search since it is not guaranteed that leaves are in sorted order) and that multiplied by some constant c will be the time complexity of the optimal algorithm. (Here, c includes the cost of all operations which includes comparison, index shift etc. for a single maximum element compare)

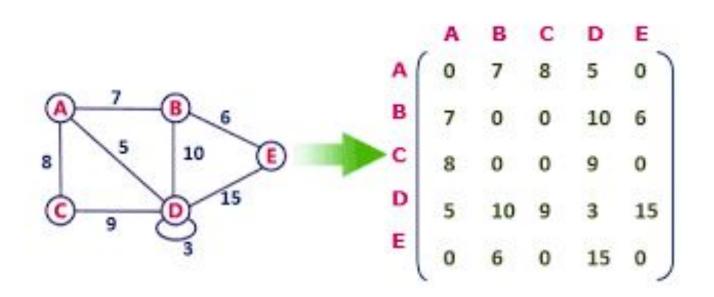
In asymptotic terms, c*(n/2+1) is $\Theta(n)$.

GRAPHS

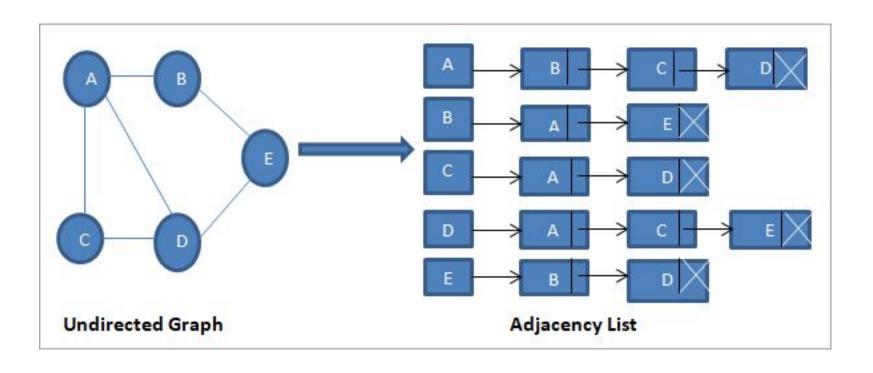
COMMON GRAPH DATA STRUCTURES



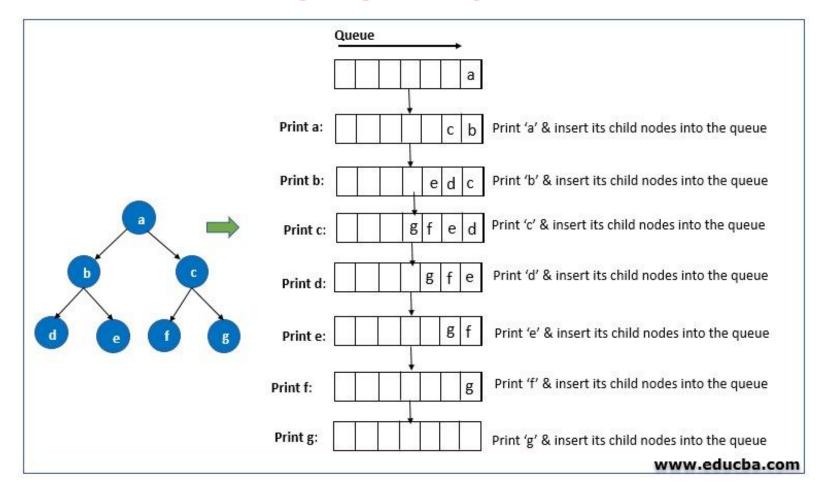
REPRESENTATION OF GRAPHS



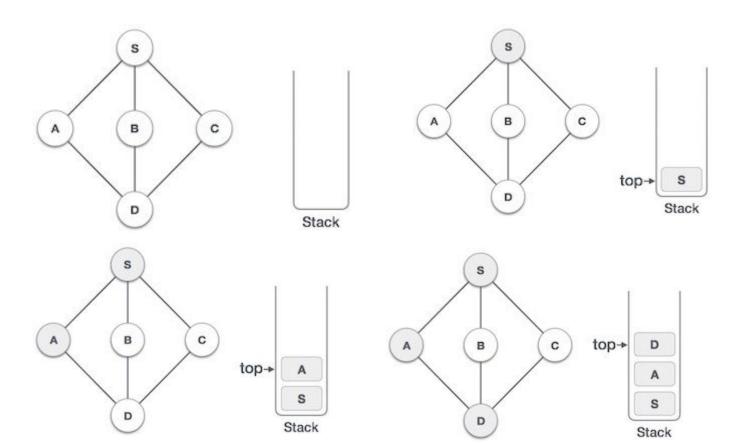
REPRESENTATION OF GRAPHS



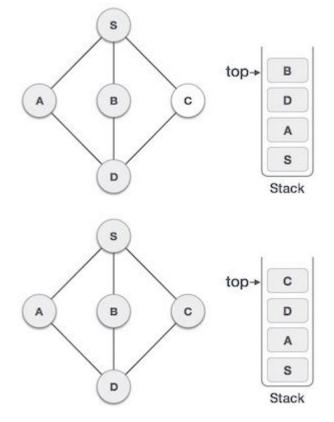
BREADTH FIRST SEARCH

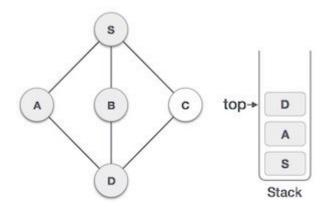


DEPTH FIRST SEARCH

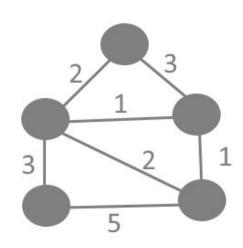


DEPTH FIRST SEARCH

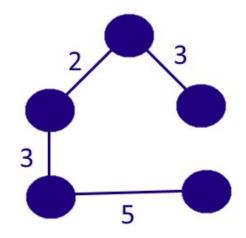




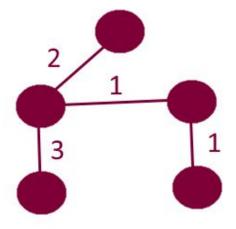
MINIMUM SPANNING TREE (MST)



Graph



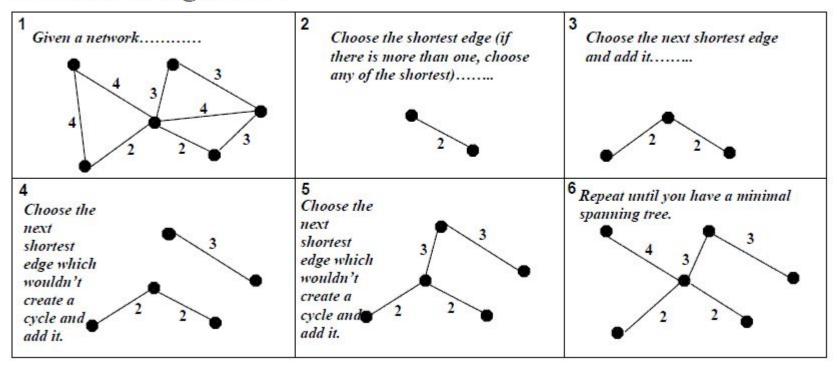
Spanning Tree Cost = 13



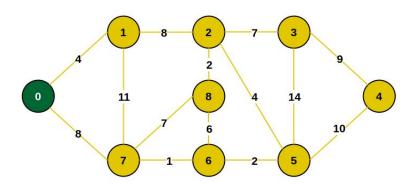
Minimum Spanning Tree, Cost = 7

KRUSKAL'S ALGORITHM

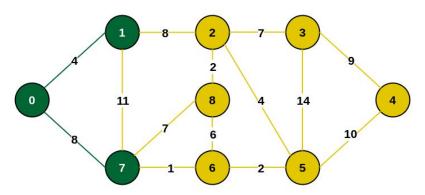
Kruskal's Algorithm

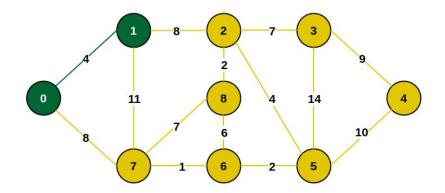


PRIM'S ALGORITHM

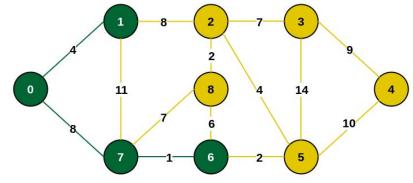


Select an arbitrary starting vertex. Here we have selected 0



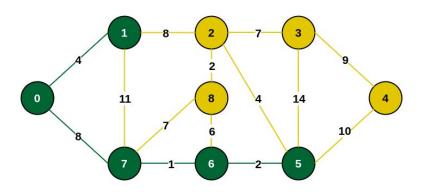


Minimum weighted edge from MST to other vertices is 0-1 with weight 4

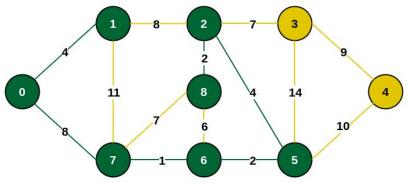


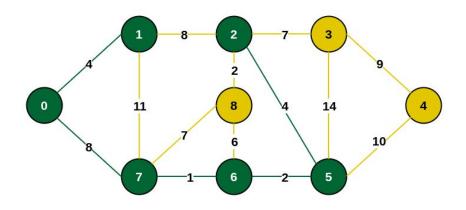
Minimum weighted edge from MST to other vertices is 7-6 with weight 1

PRIM'S ALGORITHM



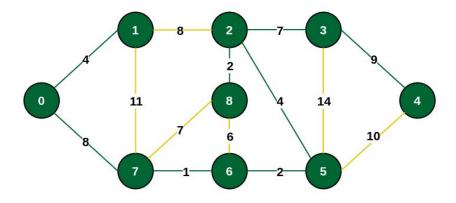
Minimum weighted edge from MST to other vertices is 6-5 with weight 2

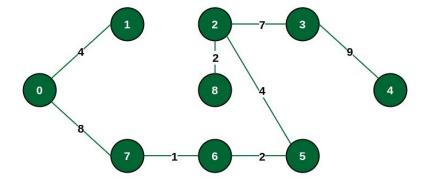




Minimum weighted edge from MST to other vertices is 5-2 with weight 4

PRIM'S ALGORITHM



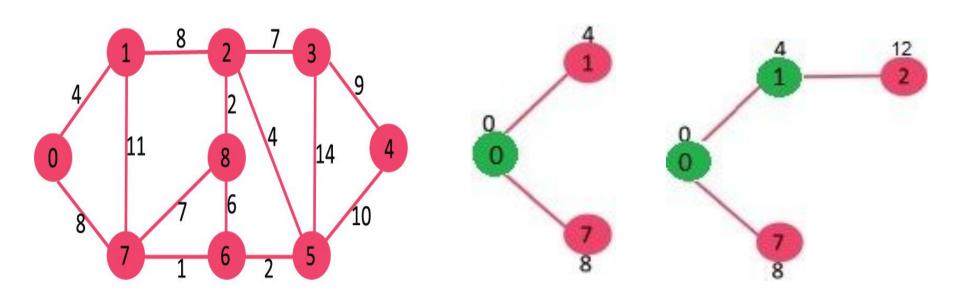


Minimum weighted edge from MST to other vertices is 3-4 with weight 9

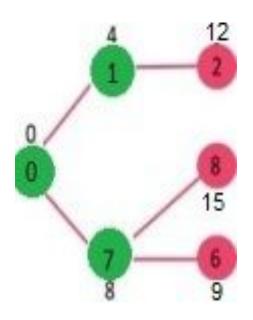
The final structure of MST

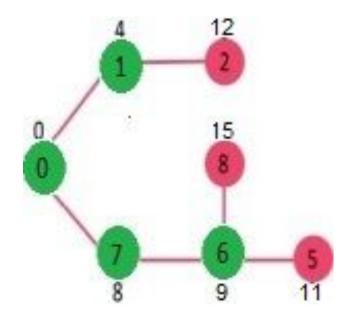
DIJKSTRA'S ALGORITHM

Given a graph and a source vertex in the graph, find the shortest paths from the source to all vertices in the given graph. (In the below example lets take 0 as source node)

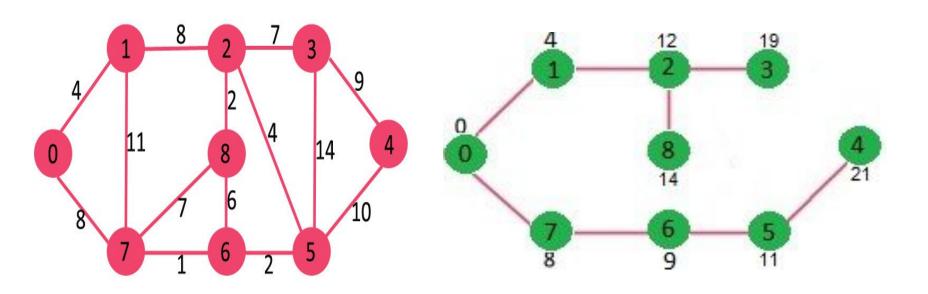


DIJKSTRA'S ALGORITHM

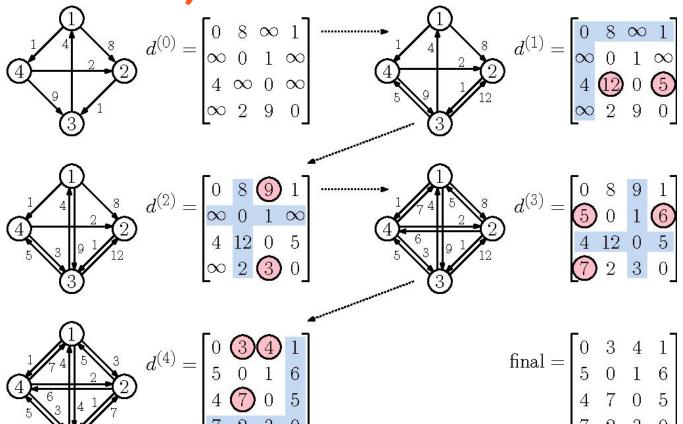




DIJKSTRA'S ALGORITHM

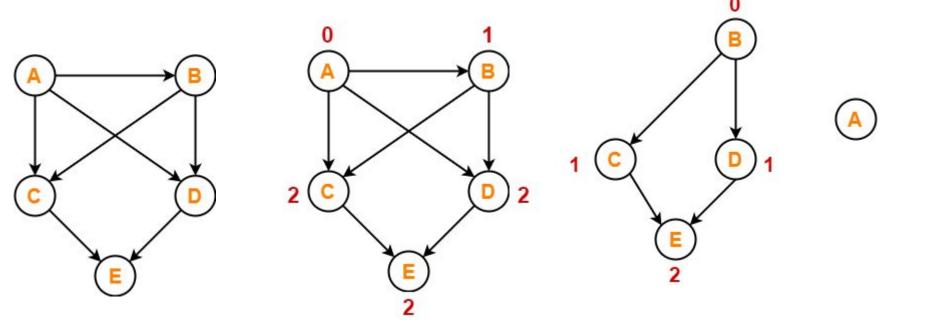


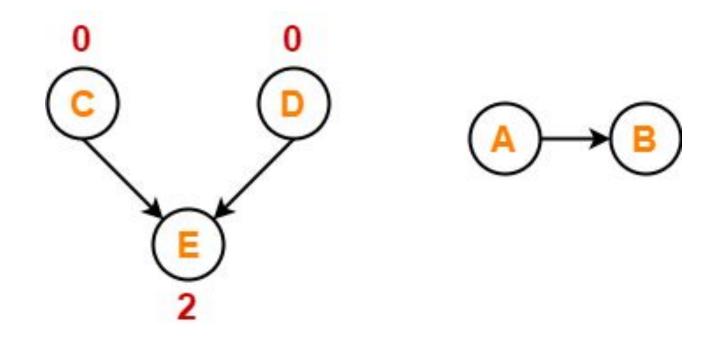
FLOYD WARSHALL ALGORITHM (All Pair Shortest Path)

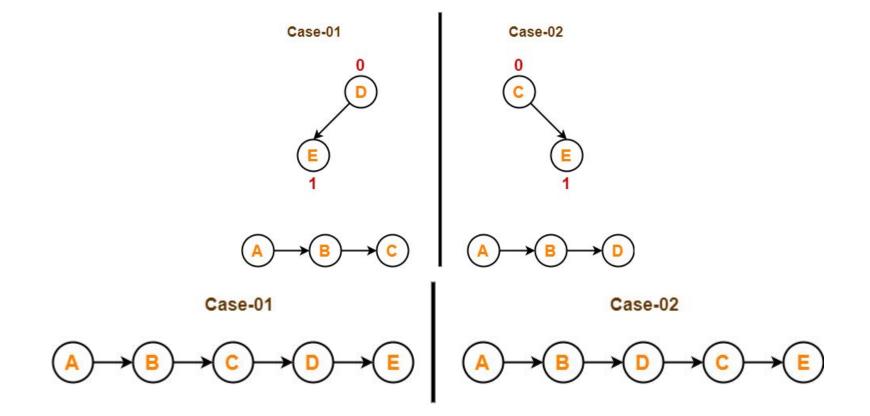


Topological Sort is a linear ordering of the vertices in such a way that if there is an edge in the DAG going from vertex 'u' to vertex 'v', then 'u' comes before 'v' in the ordering.

- Topological Sorting is possible if and only if the graph is a <u>Directed Acyclic Graph</u>.
- There may exist multiple different topological orderings for a given directed acyclic graph.





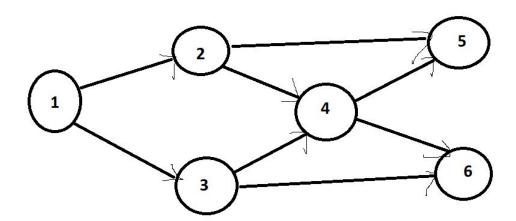


GATE QUESTION - 1

Consider the DAG with Consider $V = \{1, 2, 3, 4, 5, 6\}$, shown below. Which of the following options is NOT a topological ordering?

- (A) 123456
- **(B)** 132456
- (C) 132465
- **(D)** 3 2 4 1 6 5

ANS: (D)

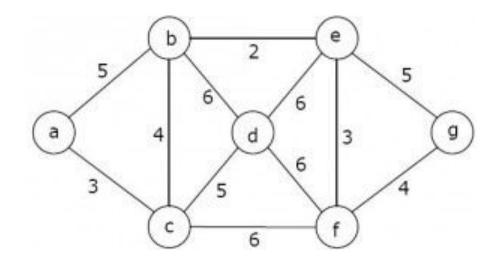


GATE QUESTION - 2

Which one of the following is NOT the sequence of edges added to the minimum spanning tree using Kruskal's algorithm?

- **(A)** (b,e)(e,f)(a,c)(b,c)(f,g)(c,d)
- **(B)** (b,e)(e,f)(a,c)(f,g)(b,c)(c,d)
- (C) (b,e)(a,c)(e,f)(b,c)(f,g)(c,d)
- **(D)** (b,e)(e,f)(b,c)(a,c)(f,g)(c,d)

Answer: (D)



GATE QUESTION - 3

Using Prim's algorithm to construct a minimum spanning tree starting with node A, which one of the following sequences of edges represents a possible order in which the edges would be added to construct the minimum spanning tree?

Answer: (D)

