Elementary Graph Algorithms

CSCI 3100

Graphs

Graph $G = (V, E)$

- V = set of vertices
- E = set of edges $\subseteq (V \times V)$

Types of graphs

- Undirected: edge $(u, v) = (v, u)$; for all v, $(v, v) \notin E$ (No self loops.)
- Directed: (u, v) is edge from u to v, denoted as $u \rightarrow v$. Self loops are allowed.
- Weighted: each edge has an associated weight, given by a weight function $w : E \rightarrow \mathbf{R}$.
- Dense: $|E| \approx |V|^2$.
- Sparse: $|E| << |V|^2$.

$|E| = O(|V|^2)$

Graphs

If $(u, v) \in E$, then vertex v is adjacent to vertex u.

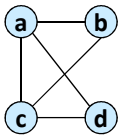
Adjacency relationship is:
- Symmetric if G is undirected.
- Not necessarily so if G is directed.

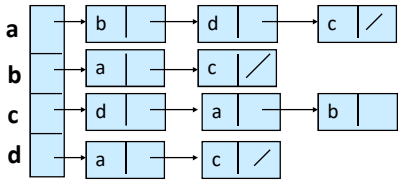
If G is connected:
- There is a path between every pair of vertices.
- $|E| \geq |V| - 1$.
- Furthermore, if $|E| = |V| - 1$, then G is a tree.

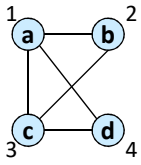
Other definitions in Appendix B (B.4 and B.5) as needed.

Representation of Graphs

Two standard ways.
- Adjacency Lists.
- Adjacency Matrix.







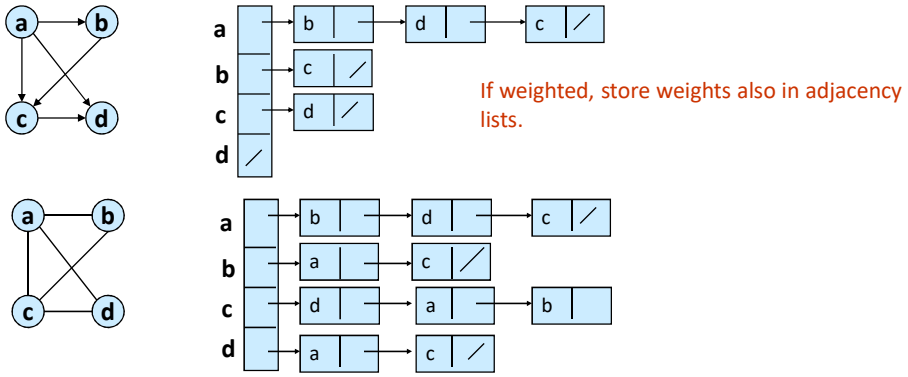
	1	2	3	4
1	0	1	1	1
2	1	0	1	0
3	1	1	0	1
4	1	0	1	0

Adjacency Lists

Consists of an array *Adj* of $|V|$ lists.

One list per vertex.

For $u \in V$, $Adj[u]$ consists of all vertices adjacent to u.



Storage Requirement

For directed graphs:

- Sum of lengths of all adj. lists is

$$\sum_{v \in V} \text{out-degree}(v) = |E|$$

No. of edges leaving v

- Total storage: $\Theta(V+E)$

For undirected graphs:

- Sum of lengths of all adj. lists is

$$\sum_{v \in V} \text{degree}(v) = 2|E|$$

No. of edges incident on v. Edge (u,v) is incident on vertices u and v.

- Total storage: $\Theta(V+E)$

Pros and Cons: adjacency list

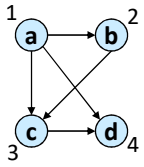
Pros

- Space-efficient, when a graph is sparse.
- Can be modified to support many graph variants.

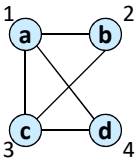
Cons

- Determining if an edge $(u,v) \in G$ is not efficient.
  - Have to search in u's adjacency list. Time complexity: $\Theta(degree(u))$
  - Worst case time complexity to search u's adjacency list: $\Theta(|V|)$

Adjacency Matrix



	1	2	3	4
1	0	1	1	1
2	0	0	1	0
3	0	0	0	1
4	0	0	0	0



	1	2	3	4
1	0	1	1	1
2	1	0	1	0
3	1	1	0	1
4	1	0	1	0

$|V| \times |V|$ matrix A.

Number vertices from 1 to $|V|$ in some arbitrary manner.

A is then given by:

$$A[i,j] = a_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

$A = A^T$ for undirected graphs.

Space and Time

Space: $\Theta(|V|^2)$.

- Not memory efficient for large graphs.

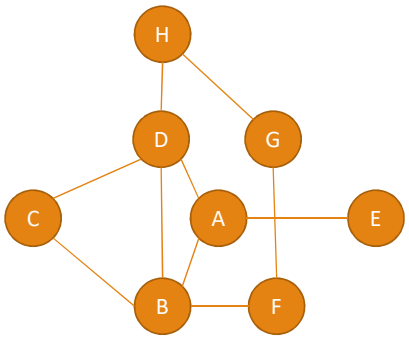
Time: to list all vertices adjacent to u: $\Theta(|V|)$.

Time: to determine if $(u, v) \in E$: $\Theta(1)$.

Can store weights instead of bits for weighted graph.

Exercise

Draw adjacency list and adjacency matrix for the following graph:



Graph-searching Algorithms

Searching a graph:

- Systematically follow the edges of a graph to visit the vertices of the graph.

Used to discover the structure of a graph.

Standard graph-searching algorithms.

- Breadth-first Search (BFS).
- Depth-first Search (DFS).

Breadth-first Search

Input: Graph $G = (V, E)$, either directed or undirected, and ***source vertex*** $s \in V$.

Output:

- $d[v]$ = distance (smallest # of edges, or shortest path) from s to v, for all $v \in V$. $d[v] = \infty$ if v is not reachable from s.
- $\pi[v] = u$ such that (u, v) is last edge on shortest path $s \rightsquigarrow v$.
  - u is v's predecessor.
- Builds breadth-first tree with root s that contains all reachable vertices.

Definitions:
Path between vertices u and v: Sequence of vertices $(v_1, v_2, \ldots, v_k)$ such that $u=v_1$ and $v=v_k$, and $(v_i, v_{i+1}) \in E$, for all $1 \le i \le k-1$.
Length of the path: Number of edges in the path.
Path is simple if no vertex is repeated.

Breadth-first Search

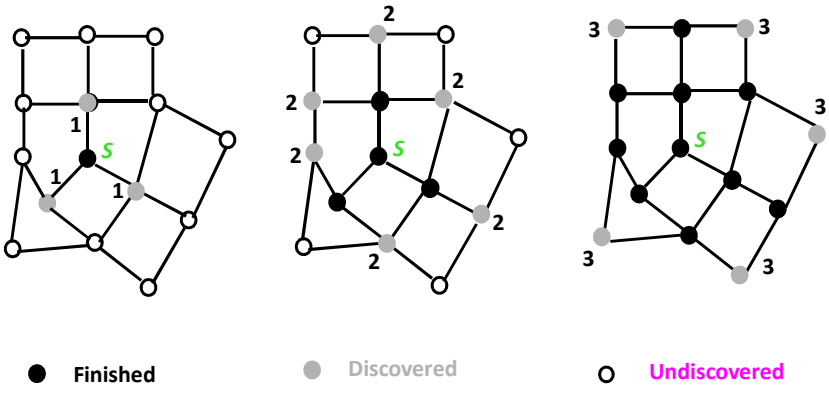
Expands the frontier between discovered and undiscovered vertices uniformly across the breadth of the frontier.

- A vertex is "discovered" the first time it is encountered during the search.
- A vertex is "finished" if all vertices adjacent to it have been discovered.

Colors the vertices to keep track of progress.

- White – Undiscovered.
- Gray – Discovered but not finished.
- Black – Finished.
  - Colors are required only to reason about the algorithm. Can be implemented without colors.

BFS for Shortest Paths

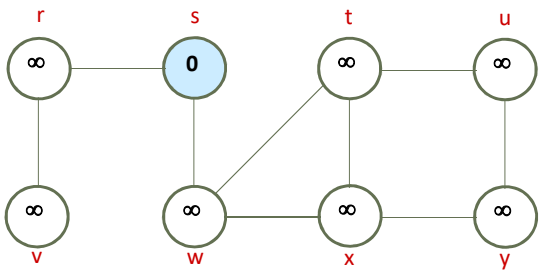


```
BFS(G, s)
 1  for each vertex u ∈ G.V − {s}
 2      u.color = WHITE
 3      u.d = ∞
 4      u.π = NIL
 5  s.color = GRAY
 6  s.d = 0
 7  s.π = NIL
 8  Q = ∅
 9  ENQUEUE(Q, s)
10  while Q ≠ ∅
11      u = DEQUEUE(Q)
12      for each v ∈ G.Adj[u]
13          if v.color == WHITE
14              v.color = GRAY
15              v.d = u.d + 1
16              v.π = u
17              ENQUEUE(Q, v)
18      u.color = BLACK
```

white: undiscovered
gray: discovered
black: finished

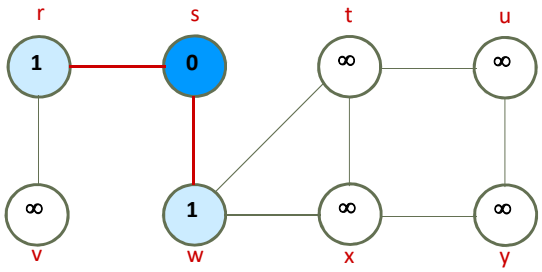
Q: a queue of discovered vertices
v.color: color of v
v.d: distance from s to v
v.π: predecessor of v

Example (BFS)



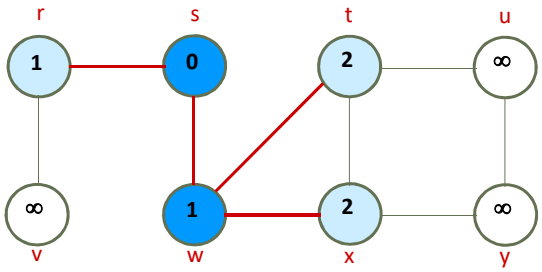
Q: s
0

Example (BFS)



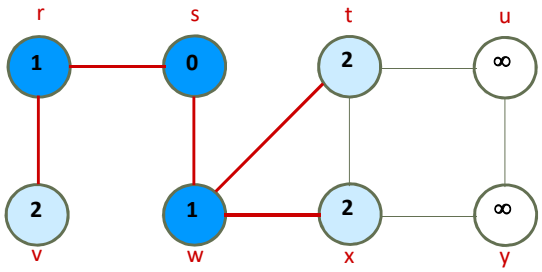
Q: w r
1 1

Example (BFS)



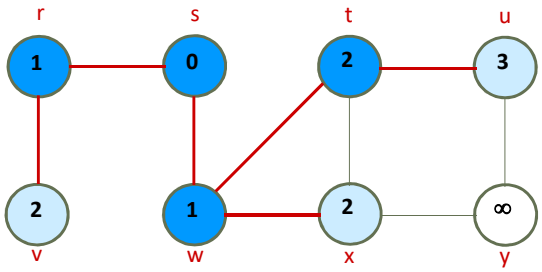
Q: r t x
1 2 2

Example (BFS)



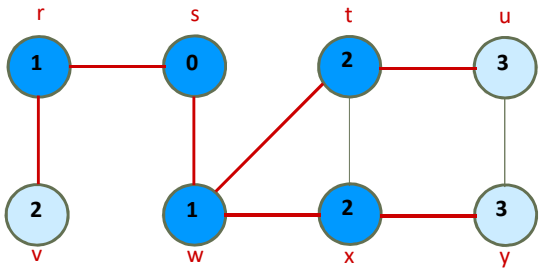
Q: t x v
2 2 2

Example (BFS)



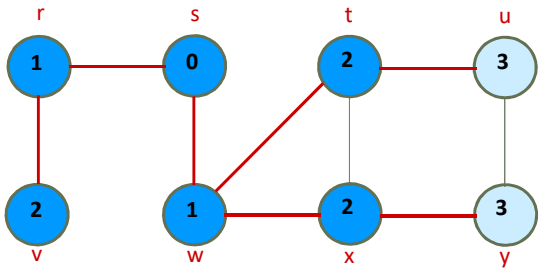
Q: x v u
2 2 3

Example (BFS)



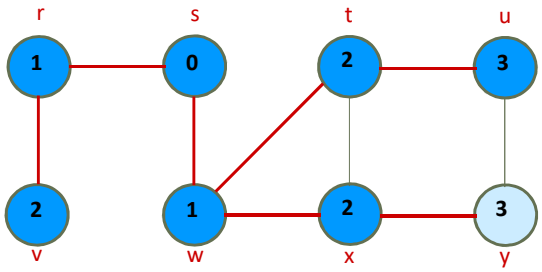
Q: v u y
2 3 3

Example (BFS)



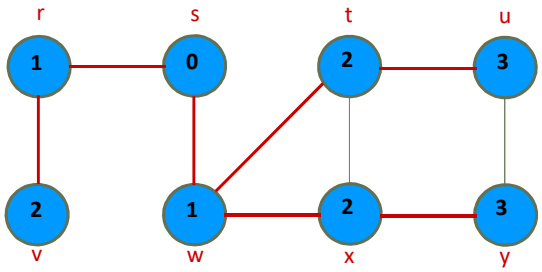
Q: u y
3 3

Example (BFS)



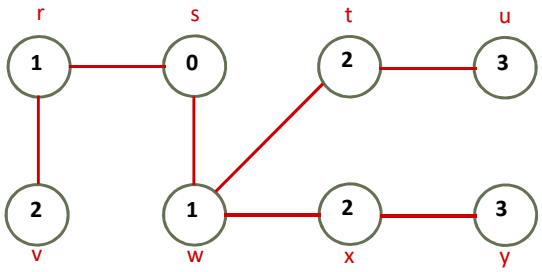
Q: y
3

Example (BFS)



Q: ∅

Example (BFS)



BF Tree

Analysis of BFS

Initialization takes $O(V)$.

Traversal Loop

- After initialization, each vertex is enqueued and dequeued at most once, and each operation takes $O(1)$. So, total time for queuing is $O(V)$.
- The adjacency list of each vertex is scanned at most once. The sum of lengths of all adjacency lists is $\Theta(E)$.

Summing up over all vertices => total running time of BFS is $O(V+E)$, linear in the size of the adjacency list representation of graph.

```
BFS(G, s)
 1  for each vertex u ∈ G.V − {s}
 2      u.color = WHITE
 3      u.d = ∞
 4      u.π = NIL
 5  s.color = GRAY
 6  s.d = 0
 7  s.π = NIL
 8  Q = ∅
 9  ENQUEUE(Q, s)
10  while Q ≠ ∅
11      u = DEQUEUE(Q)
12      for each v ∈ G.Adj[u]
13          if v.color == WHITE
14              v.color = GRAY
15              v.d = u.d + 1
16              v.π = u
17              ENQUEUE(Q, v)
18      u.color = BLACK
```

Breadth-first Tree

For a graph $G = (V, E)$ with source s, the **predecessor subgraph** of G is $G_\pi = (V_\pi, E_\pi)$ where

- $V_\pi = \{v \in V : \pi[v] \neq \text{NIL}\} \cup \{s\}$
- $E_\pi = \{(\pi[v], v) \in E : v \in V_\pi - \{s\}\}$

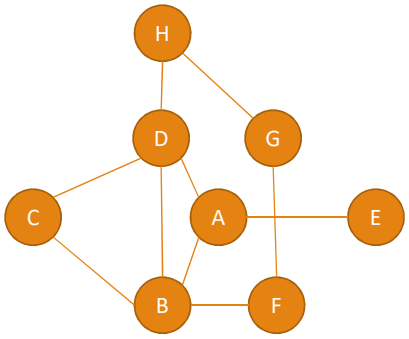
The predecessor subgraph G_π is a **breadth-first tree** if:

- V_π consists of the vertices reachable from s and
- for all $v \in V_\pi$, there is a unique simple path from s to v in G_π that is also a shortest path from s to v in G.

The edges in E_π are called **tree edges**. $|E_\pi| = |V_\pi| - 1$.

Exercise

Use BFS algorithm to determine the breadth-first tree of the following graph starting with node A:



Shortest Path

Breadth First Search computes the shortest path between two nodes of a graph

Proof is in Ch 22.2 – READ IT 

Depth-first Search (DFS)

Explore edges out of the most recently discovered vertex *v*.

When all edges of *v* have been explored, backtrack to explore other edges leaving the vertex from which *v* was discovered (its *predecessor*).

"Search as deep as possible first."

Continue until all vertices reachable from the original source are discovered.

If any undiscovered vertices remain, then one of them is chosen as a new source and search is repeated from that source.

Depth-first Search

Input: $G = (V, E)$, directed or undirected. No source vertex given!

Output:

- 2 **timestamps** on each vertex. Integers between 1 and 2|V|.
  - $d[v]$ = ***discovery time*** (v turns from white to gray)
  - $f[v]$ = ***finishing time*** (v turns from gray to black)
- $\pi[v]$: predecessor of $v = u$, such that v was discovered during the scan of u's adjacency list.

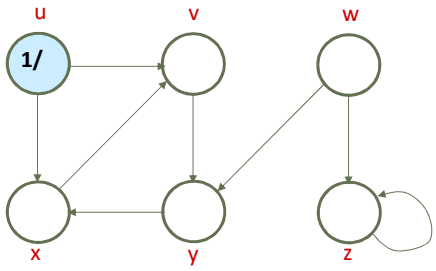
Uses the same coloring scheme for vertices as BFS.

Pseudo-code

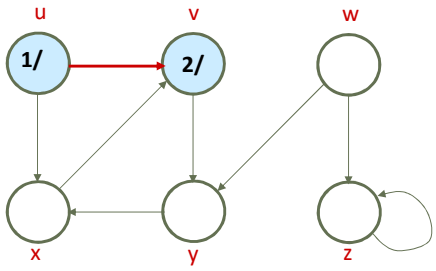
```
DFS(G)
1  for each vertex u ∈ G.V
2      u.color = WHITE
3      u.π = NIL
4  time = 0
5  for each vertex u ∈ G.V
6      if u.color == WHITE
7          DFS-VISIT(G, u)
```

```
DFS-VISIT(G, u)
1  time = time + 1
2  u.d = time
3  u.color = GRAY
4  for each v ∈ G.Adj[u]
5      if v.color == WHITE
6          v.π = u
7          DFS-VISIT(G, v)
8  u.color = BLACK
9  time = time + 1
10 u.f = time
```

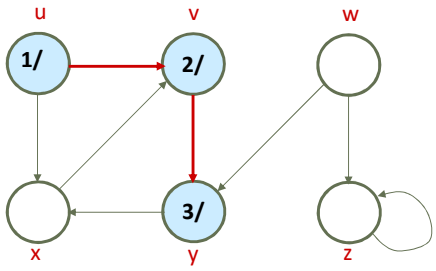
Example (DFS)



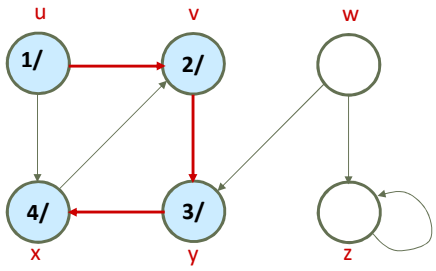
Example (DFS)



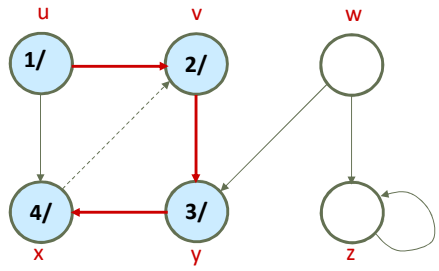
Example (DFS)



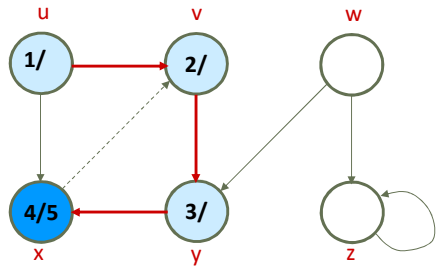
Example (DFS)



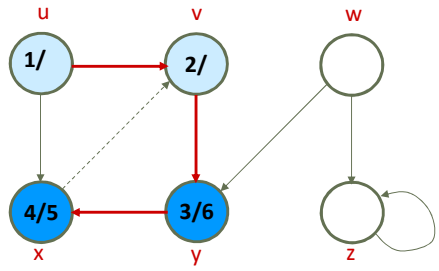
Example (DFS)



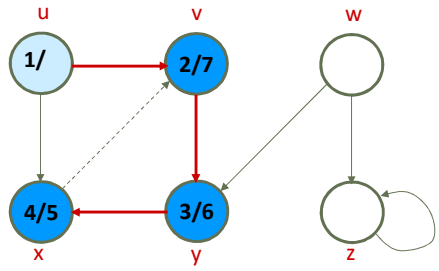
Example (DFS)



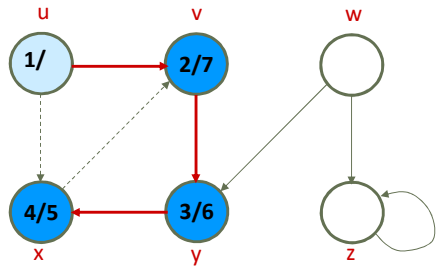
Example (DFS)



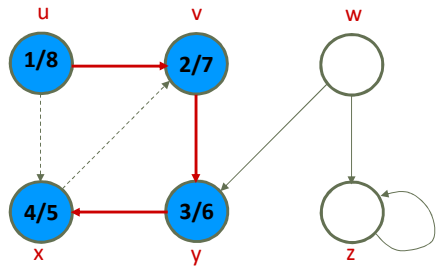
Example (DFS)



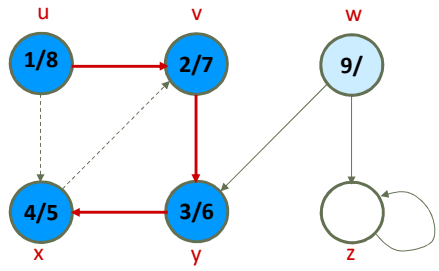
Example (DFS)



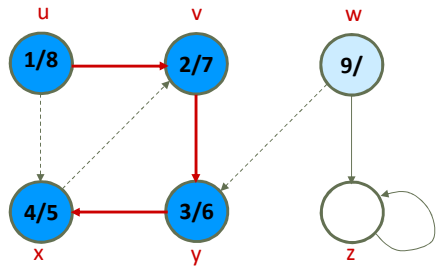
Example (DFS)



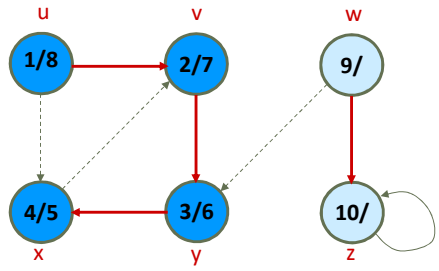
Example (DFS)



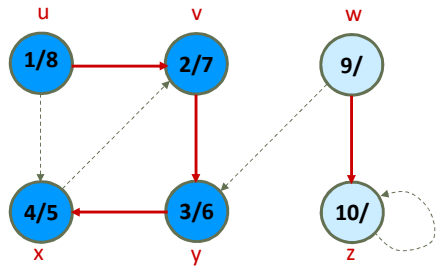
Example (DFS)



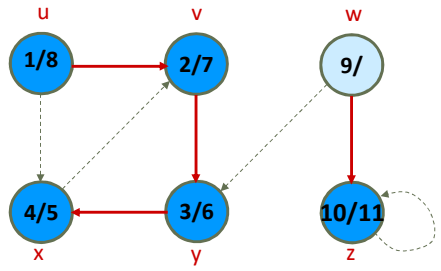
Example (DFS)



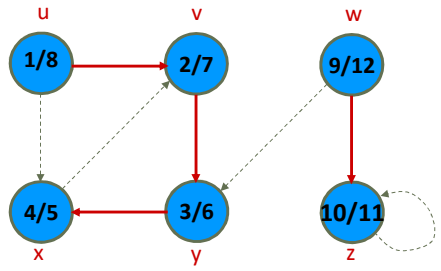
Example (DFS)



Example (DFS)



Example (DFS)



Analysis of DFS

Loops on lines 1-2 & 5-7 take $\Theta(V)$ time, excluding time to execute DFS-Visit.

DFS-Visit is called once for each white vertex $v \in V$ when it's painted gray the first time. Lines 4-7 of DFS-Visit is executed $|Adj[v]|$ times. The total cost of executing DFS-Visit is $\sum_{v \in V} |Adj[v]| = \Theta(E)$

Total running time of DFS is $\Theta(V+E)$.

```
DFS(G)
1  for each vertex u ∈ G.V
2      u.color = WHITE
3      u.π = NIL
4  time = 0
5  for each vertex u ∈ G.V
6      if u.color == WHITE
7          DFS-VISIT(G, u)
```

```
DFS-VISIT(G, u)
1  time = time + 1
2  u.d = time
3  u.color = GRAY
4  for each v ∈ G.Adj[u]
5      if v.color == WHITE
6          v.π = u
7          DFS-VISIT(G, v)
8  u.color = BLACK
9  time = time + 1
10 u.f = time
```

Theorem 22.7

For all u, v, exactly one of the following holds:

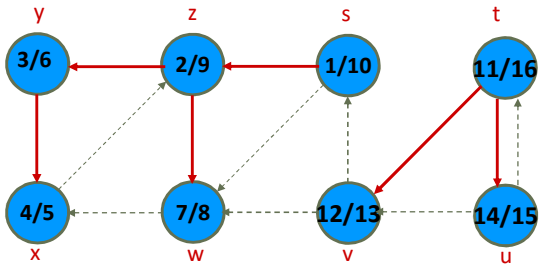
1. $d[u] < f[u] < d[v] < f[v]$ or $d[v] < f[v] < d[u] < f[u]$ and neither u nor v is a descendant of the other.
2. $d[u] < d[v] < f[v] < f[u]$ and v is a descendant of u.
3. $d[v] < d[u] < f[u] < f[v]$ and u is a descendant of v.

- So $d[u] < d[v] < f[u] < f[v]$ *cannot* happen.
- Like parentheses:
  - OK: ( ) [ ] ([ ]) [( )]
  - Not OK: ([ )] [( ])

Corollary

v is a proper descendant of u if and only if $d[u] < d[v] < f[v] < f[u]$.

Example (Parenthesis Theorem)



(s (z (y (x x) y) (w w) z) s) (t (v v) (u u) t)

Depth-First Trees

Predecessor subgraph defined slightly different from that of BFS.

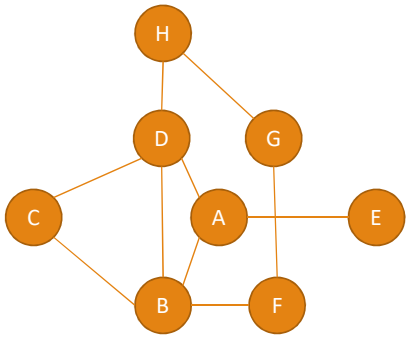
The predecessor subgraph of DFS is $G_\pi = (V, E_\pi)$ where $E_\pi = \{(\pi[v], v) : v \in V \text{ and } \pi[v] \neq \text{NIL}\}$.

- How does it differ from that of BFS?
- The predecessor subgraph G_π forms a *depth-first forest* composed of several *depth-first trees*. The edges in E_π are called *tree edges*.

Definition:
Forest: An acyclic graph G that may be disconnected.

Exercise

Apply DFS algorithm to the following graph and draw the depth first forest.



Theorem 22.9

v is a descendant of *u* if and only if at time *d*[*u*], there is a path from *u to v* consisting of only white vertices. (Except for *u*, which was *just* colored gray.)

Classification of Edges

Tree edge: in the depth-first forest. Found by exploring (u, v).

Back edge: (u, v), where u is a descendant of v (in the depth-first tree).

Forward edge: (u, v), where v is a descendant of u, but not a tree edge.

Cross edge: any other edge. Can go between vertices in same depth-first tree or in different depth-first trees.

Theorem:
In DFS of an undirected graph, we get only tree and back edges. No forward or cross edges.