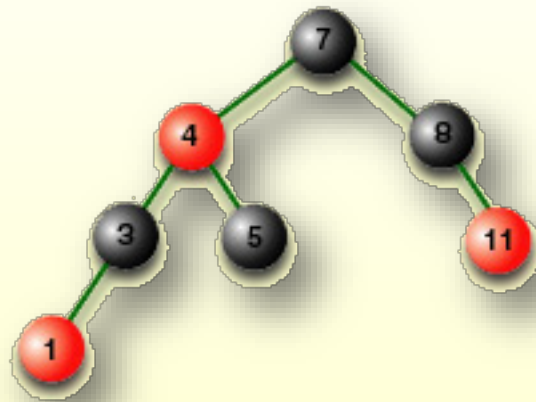


CS161:

Design and Analysis of Algorithms



Lecture 13

Leonidas Guibas

Outline

- ◆ Last lecture: **Augmenting data structures; Amortized analysis**
- ◆ Today new topic: **Graph algorithms**
 - ◆ breadth-first search
 - ◆ depth-first search
 - ◆ topological sorting

Slides modified from

- <http://www.cs.virginia.edu/~luebke/cs332/>
- <http://www.cs.unc.edu/~plaisted/>

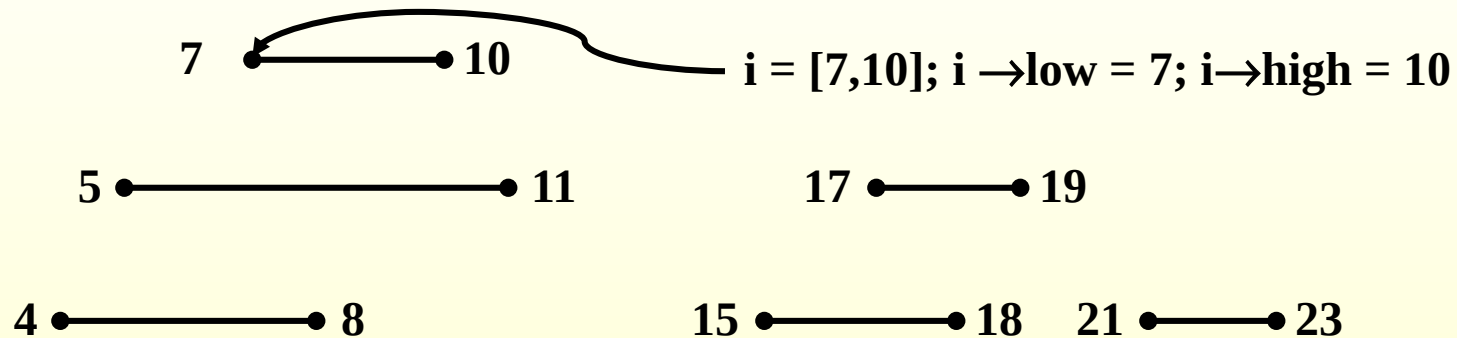
Dynamic Order Statistics

- ◆ *OS-Trees*: a structure to support finding the i -th element of a **dynamic** set in $O(\lg n)$ time
 - ◆ Support standard dynamic set operations (**Insert()**, **Delete()**, **Min()**, **Max()**, **Succ()**, **Pred()**)
 - ◆ Also support these order statistic operations:
 - void OS-Select(root, i); -- find element of given rank**
 - int OS-Rank(x); -- find rank of given element**

Interval Trees

- ◆ The problem: maintain a set of intervals

- ◆ E.g., time intervals for a scheduling program:



- ◆ Query: find an interval in the set that overlaps a given query interval (conflict detection)

- ◆ $[14, 16] \rightarrow [15, 18]$
- ◆ $[16, 19] \rightarrow [15, 18]$ or $[17, 19]$
- ◆ $[12, 14] \rightarrow \text{NULL}$

The Methodology

- ◆ Following these steps:
 - ◆ Pick underlying data structure
 - ◆ Decide what additional information to store
 - ◆ Figure out how to maintain the information
 - ◆ Develop the desired new operations

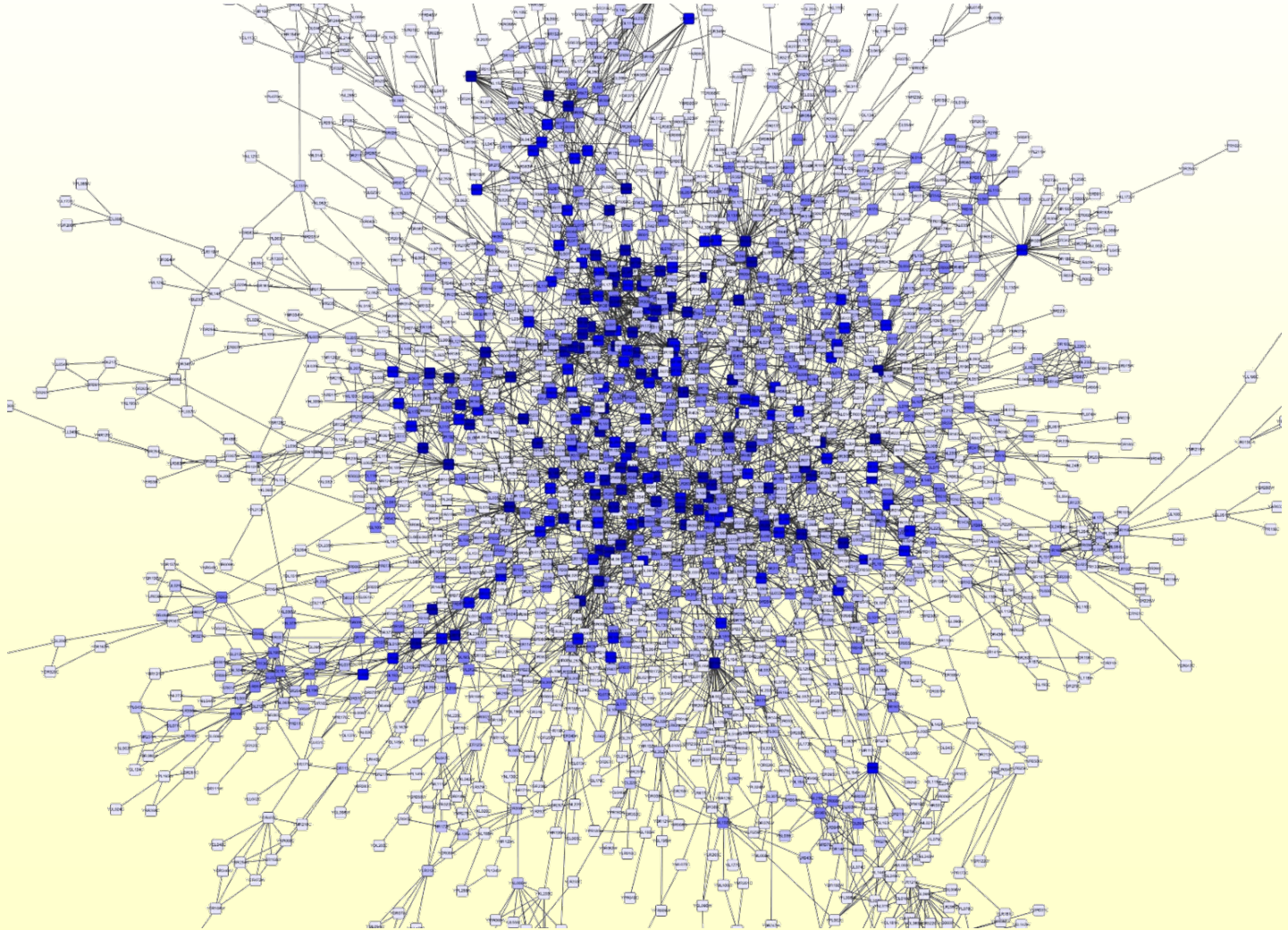
OS-Trees

- ◆ Choose underlying data structure
 - ◆ E.g., red-black trees
- ◆ Determine additional information to maintain
 - ◆ E.g., subtree sizes
- ◆ Verify that information can be maintained for operations that modify the structure
 - ◆ E.g., Insert(), Delete() (don't forget rotations!)
- ◆ Develop new operations
 - ◆ E.g., OS-Rank(), OS-Select()

Interval Trees

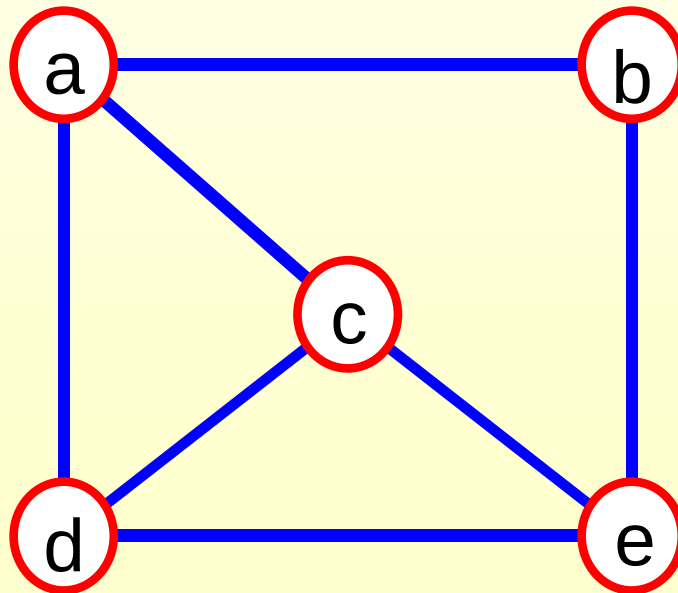
- ◆ Following the methodology:
 - ◆ Pick underlying data structure
 - ◆ Red-black trees will store intervals, keyed on $i \rightarrow low$ (the left endpoint)
 - ◆ Decide what additional information to store
 - ◆ Store the maximum right endpoint in the subtree rooted at i
 - ◆ *Figure out how to maintain the information*
 - ◆ *How would we maintain max field for a BST?*
 - ◆ *What's different?*
 - ◆ Develop the desired new operations

Data Structures and Algorithms for Graphs



What is a Graph?

- A graph $G = (V, E)$ is composed of:
 - V : set of **vertices**
 - E : set of **edges** connecting the **vertices** in V
- An **edge** $e = (u, v)$ is a pair of **vertices**
- Example:

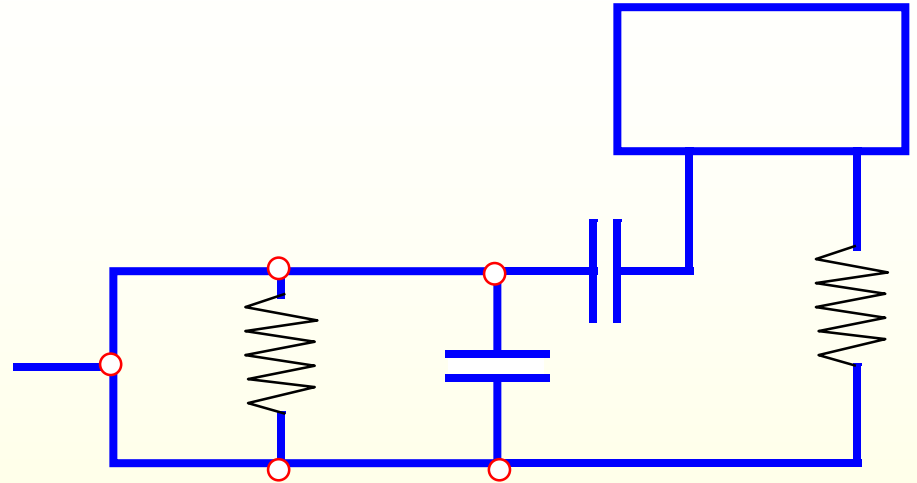


$V = \{a, b, c, d, e\}$

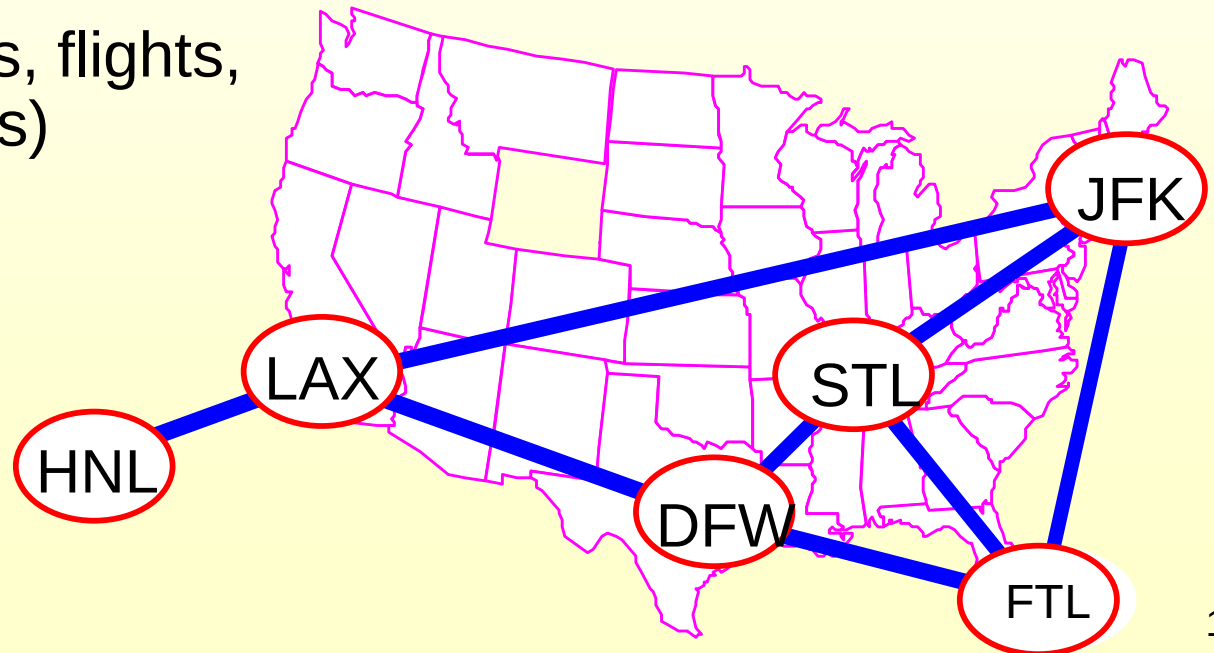
$E = \{(a, b), (a, c), (a, d), (b, e), (c, d), (c, e), (d, e)\}$

Many Applications

- ♦ electronic circuits



- ♦ **networks** (roads, flights, communications)



Formally: 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**: $|E| \ll |V|^2$.

- ◆ $|E| = O(|V|^2)$

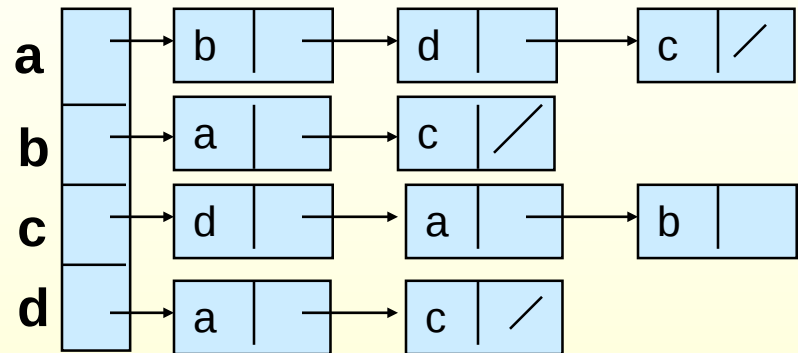
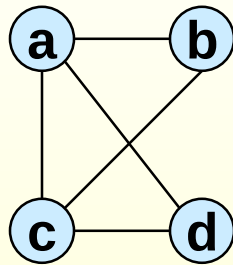
Formally: Graphs

- ◆ 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.

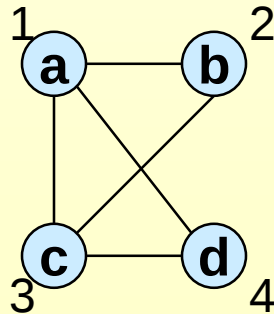
Representation of Graphs

◆ Two standard ways

◆ Adjacency Lists



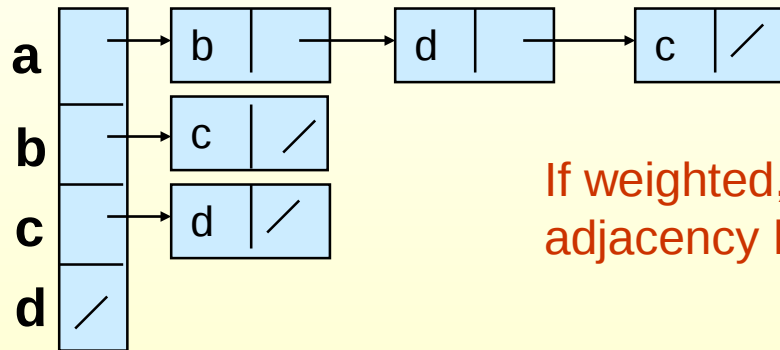
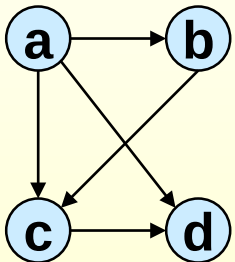
◆ Adjacency Matrix



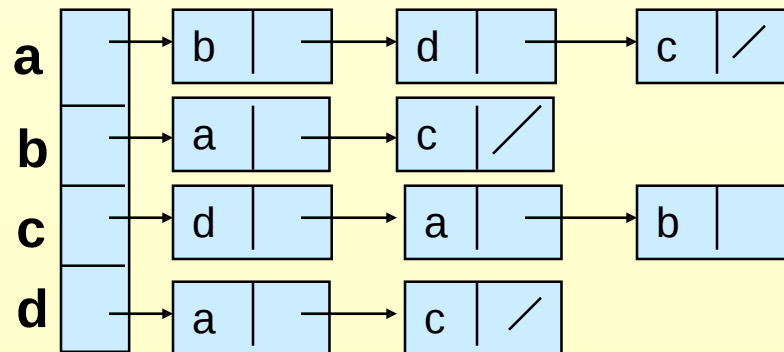
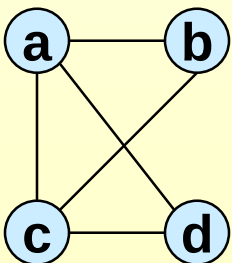
	1	2	3	4
1	0	1	1	1
2	1	0	1	0
3	1	1	0	1
4	1	0	1	0

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 .



If weighted, store weights also in adjacency lists.



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: Adj List

◆ Pros

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

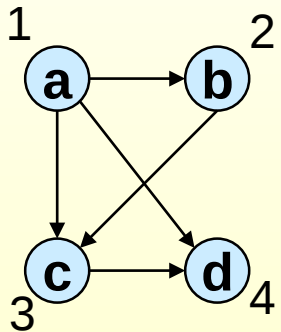
◆ Cons

- ◆ **Determining if an edge $(u,v) \in G$ is not efficient.**
 - ◆ Have to search in u 's adjacency list. $\Theta(\text{degree}(u))$ time.
 - ◆ $\Theta(V)$ in the worst case.

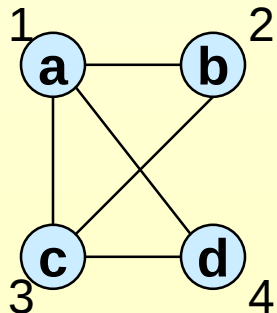
Adjacency Matrix

- ♦ $|V| \times |V|$ matrix A .
- ♦ Number vertices from 1 to $|V|$ in some arbitrary manner.
- ♦ A is then given by:

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



	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

$A = A^T$ 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.

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 search tree with root s that contains all reachable vertices.

Definitions:

Path between vertices u and v : Sequence of vertices $(v_1, v_2, \dots, v_k)$ such that $u=v_1$ and $v=v_k$, and $(v_i, v_{i+1}) \in E$, for all $1 \leq i \leq k-1$.

Length of the path: Number of edges in the path.

Path is **simple** if no vertex is repeated.

Breadth-First Search (BFS)

- ◆ Expands the frontier between discovered and undiscovered vertices **uniformly** across the length of the frontier by using a queue.
 - ◆ 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(G,s)

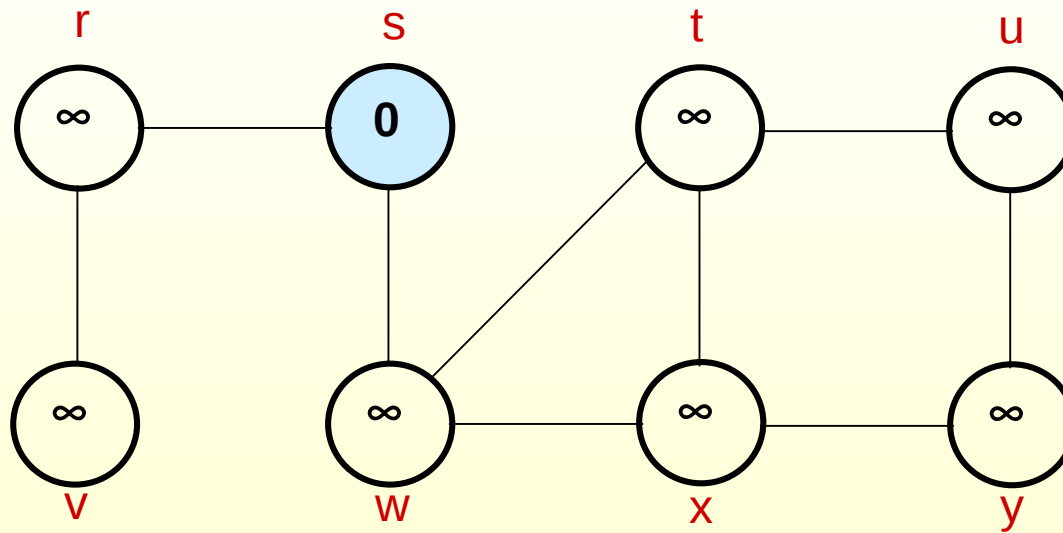
```
1. for each vertex  $u$  in  $V[G] - \{s\}$ 
2     do  $color[u] \leftarrow \text{white}$ 
3      $d[u] \leftarrow \infty$ 
4      $\pi[u] \leftarrow \text{nil}$ 
5  $color[s] \leftarrow \text{gray}$ 
6  $d[s] \leftarrow 0$ 
7  $\pi[s] \leftarrow \text{nil}$ 
8  $Q \leftarrow \Phi$ 
9 enqueue( $Q, s$ )
10 while  $Q \neq \Phi$ 
11     do  $u \leftarrow \text{dequeue}(Q)$ 
12         for each  $v$  in  $\text{Adj}[u]$ 
13             do if  $color[v] = \text{white}$ 
14                 then  $color[v] \leftarrow \text{gray}$ 
15                      $d[v] \leftarrow d[u] + 1$ 
16                      $\pi[v] \leftarrow u$ 
17                     enqueue( $Q, v$ )
18      $color[u] \leftarrow \text{black}$ 
```

white: undiscovered
gray: discovered
black: finished

Q : a queue of discovered
vertices

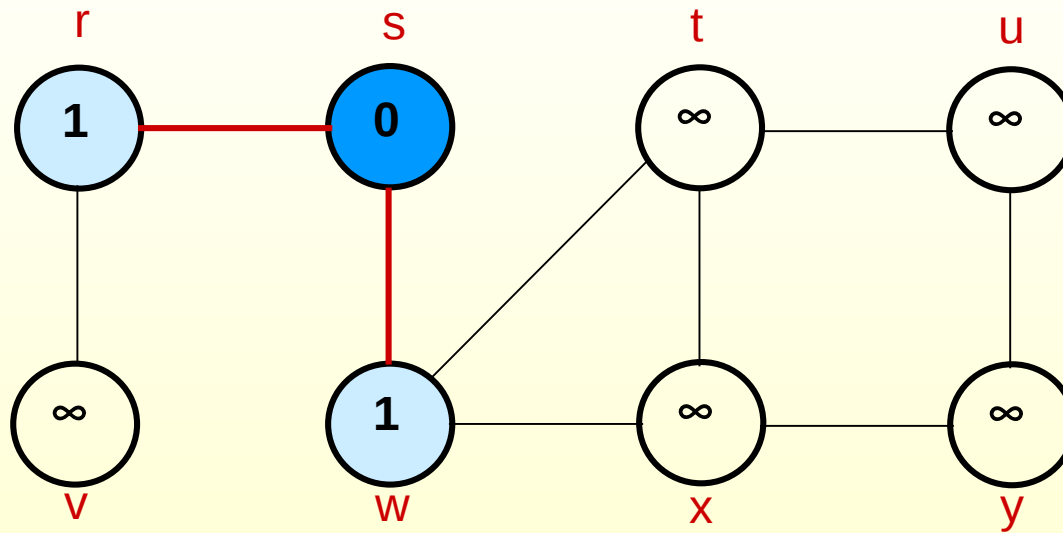
$color[v]$: color of v
 $d[v]$: distance from s to v
 $\pi[v]$: predecessor of v

Example (BFS)

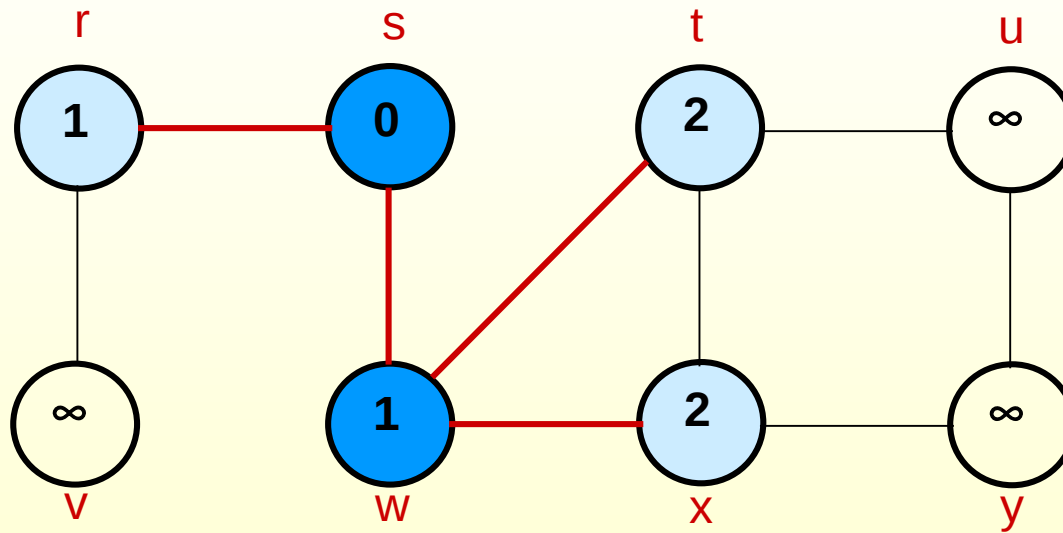


Q: s
0

Example (BFS)

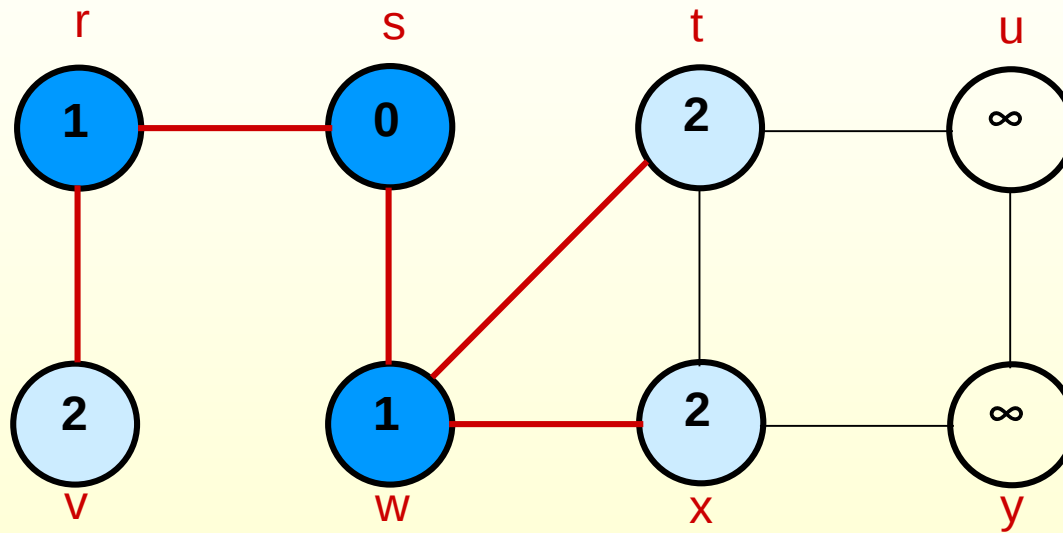


Example (BFS)



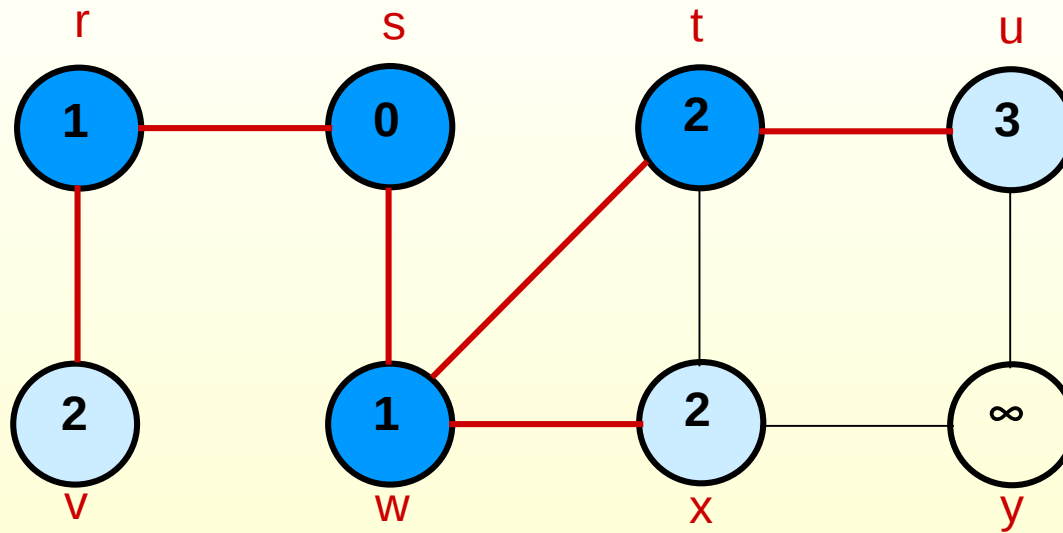
Q: r t x
1 2 2

Example (BFS)



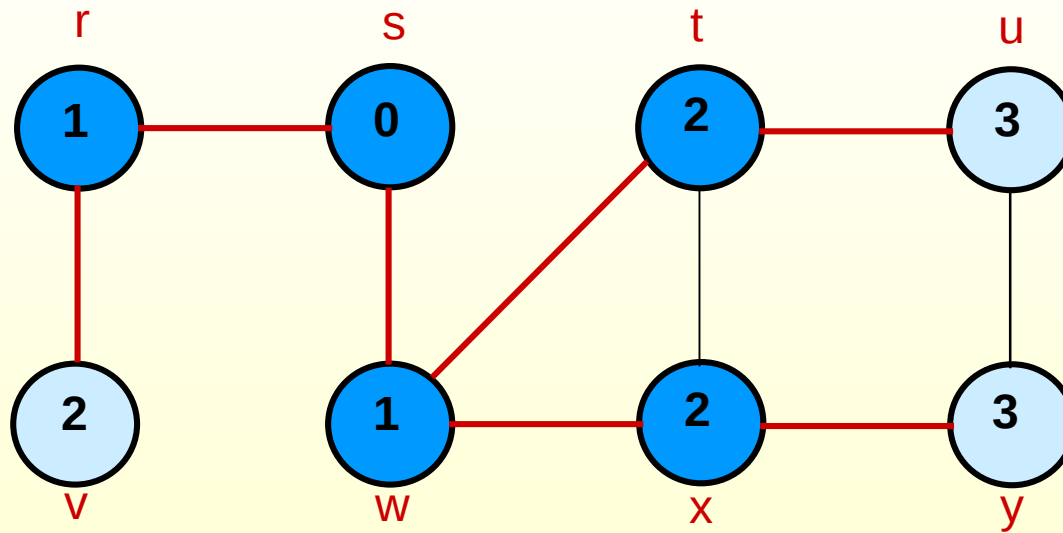
Q: t x v
2 2 2

Example (BFS)



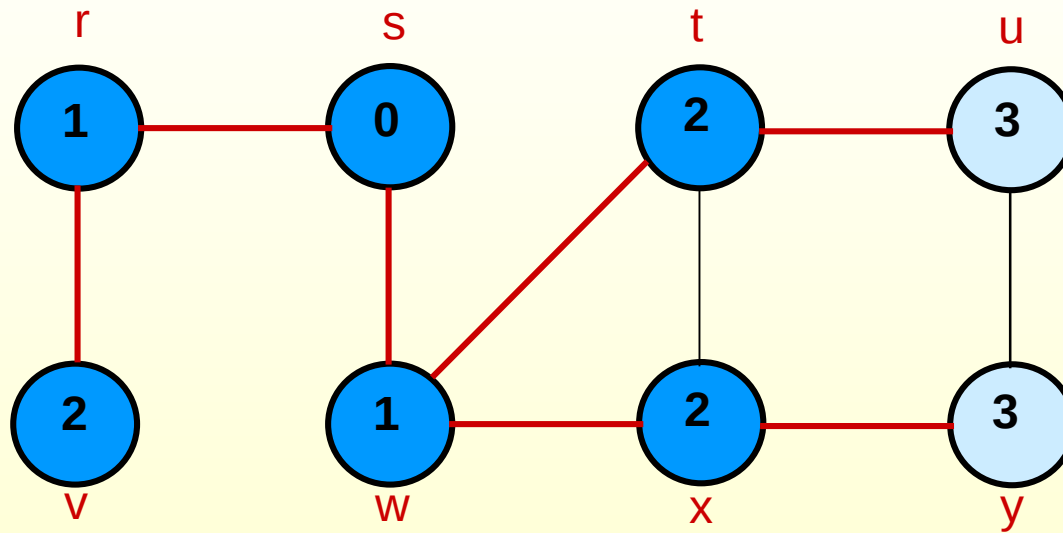
Q:	x	v	u
	2	2	3

Example (BFS)



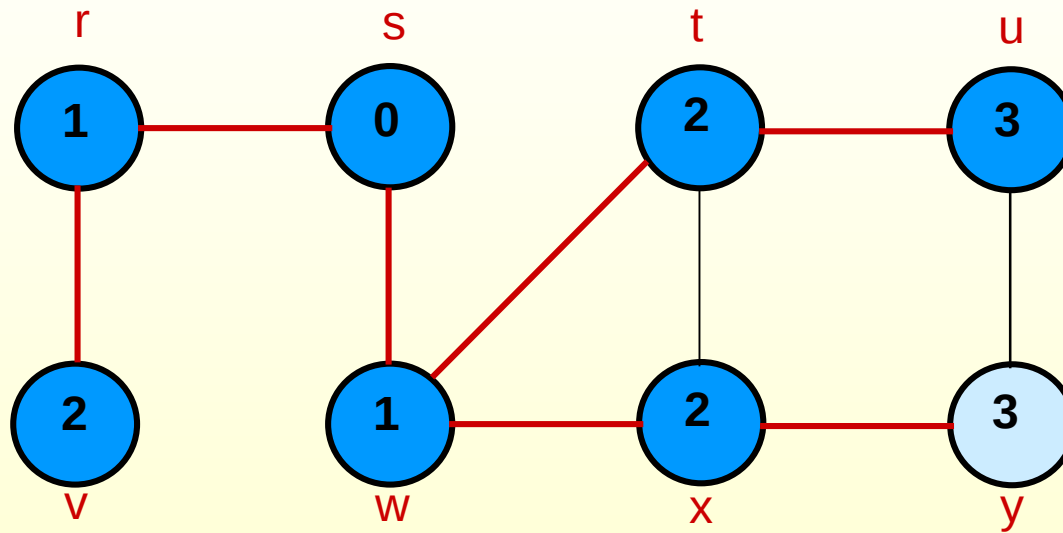
Q:	v	u	y
	2	3	3

Example (BFS)



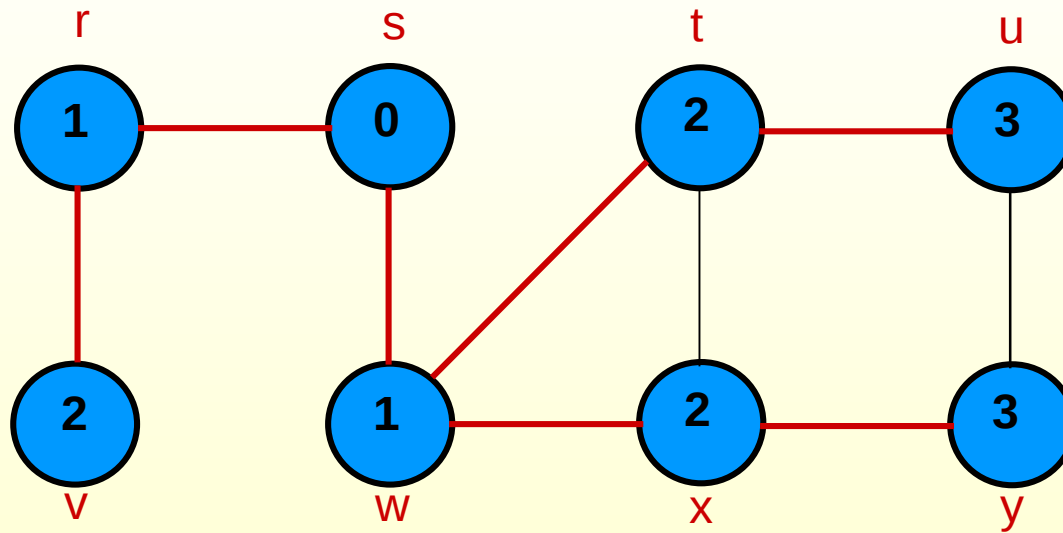
Q: u y
3 3

Example (BFS)



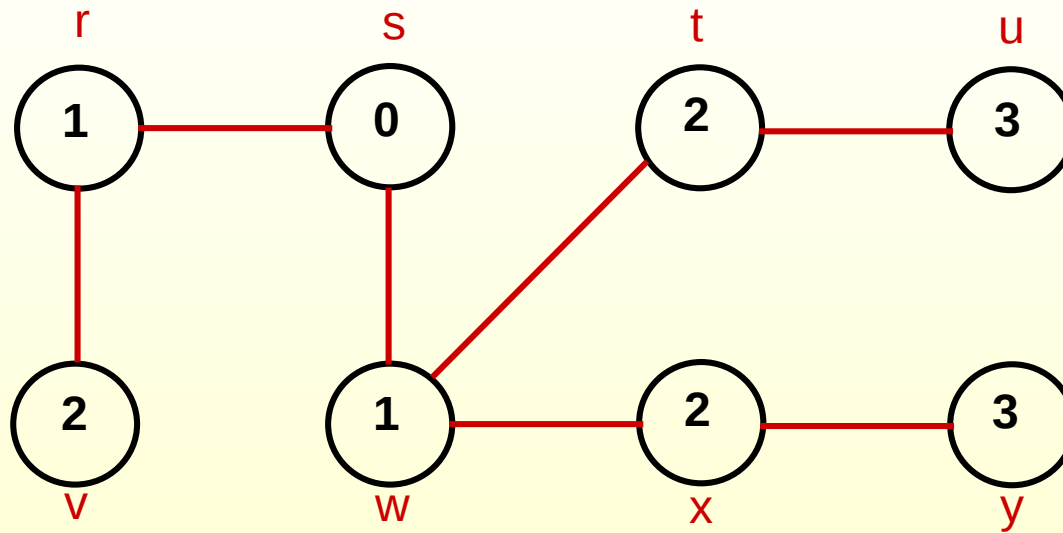
Q: y
3

Example (BFS)



Q: $\emptyset$

Example (BFS)



BFS 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 $\rightarrow$ total running time of BFS is $O(V+E)$, linear in the size of the adjacency list representation of graph.
- ◆ Correctness Proof
 - ◆ Straightforward

Breadth-First Search 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**:
 - ◆ 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**.
 $|E_\pi| = |V_\pi| - 1$.

Graph Search Algorithms

- ◆ BFS, DFS often used for their “by-products”
- ◆ BFS provides shortest path distances to the source, and the BFS tree is a shortest path tree
- ◆ BFS selects a set of graph edges with useful properties

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” – using a stack.
- ◆ 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 $2|V|$.
 - ◆ $d[v] = \textit{discovery time}$ (v turns from white to gray)
 - ◆ $f[v] = \textit{finishing time}$ (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)

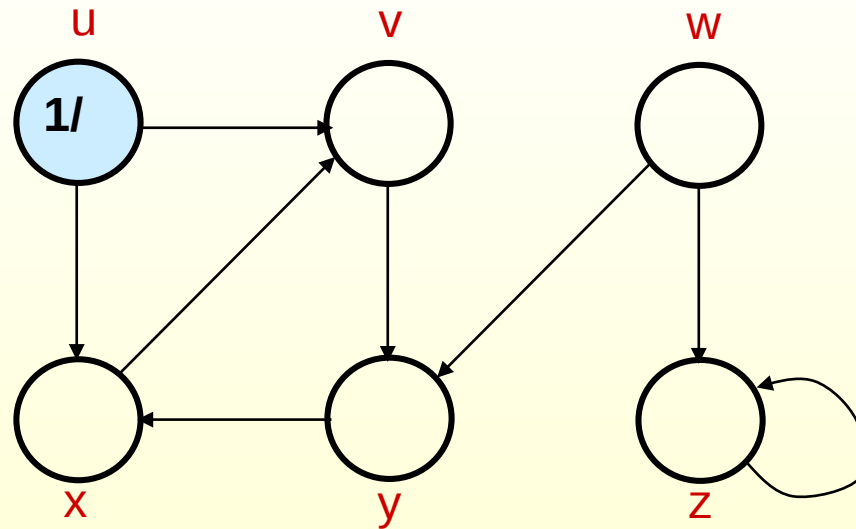
1. **for** each vertex $u \in V[G]$
2. **do** $color[u] \leftarrow \text{white}$
3. $\pi[u] \leftarrow \text{NIL}$
4. $time \leftarrow 0$
5. **for** each vertex $u \in V[G]$
6. **do if** $color[u] = \text{white}$
7. **then** DFS-Visit(u)

Uses a global timestamp *time*.

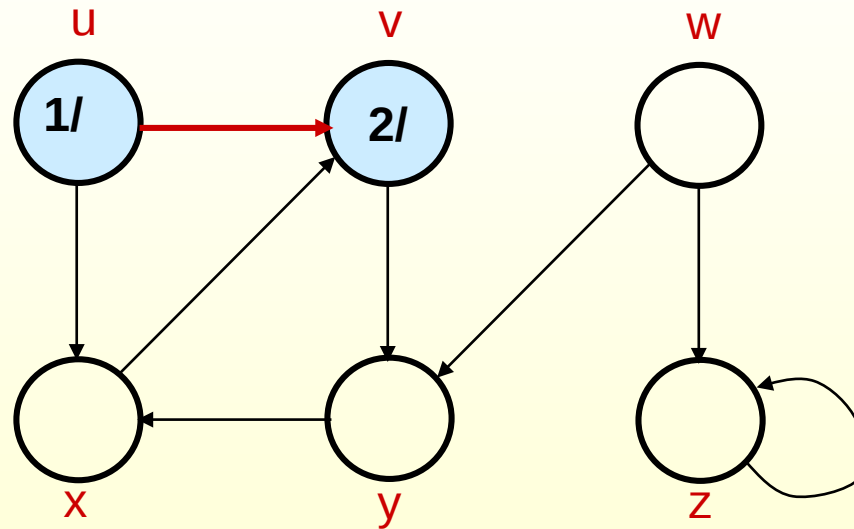
DFS-Visit(u)

1. $color[u] \leftarrow \text{GRAY} \quad \nabla$ White vertex u
 has been discovered
2. $time \leftarrow time + 1$
3. $d[u] \leftarrow time$
4. **for** each $v \in Adj[u]$
5. **do if** $color[v] = \text{WHITE}$
6. **then** $\pi[v] \leftarrow u$
7. DFS-Visit(v)
8. $color[u] \leftarrow \text{BLACK} \quad \nabla$ Blacken u ;
 it is finished.
9. $f[u] \leftarrow time \leftarrow time + 1$

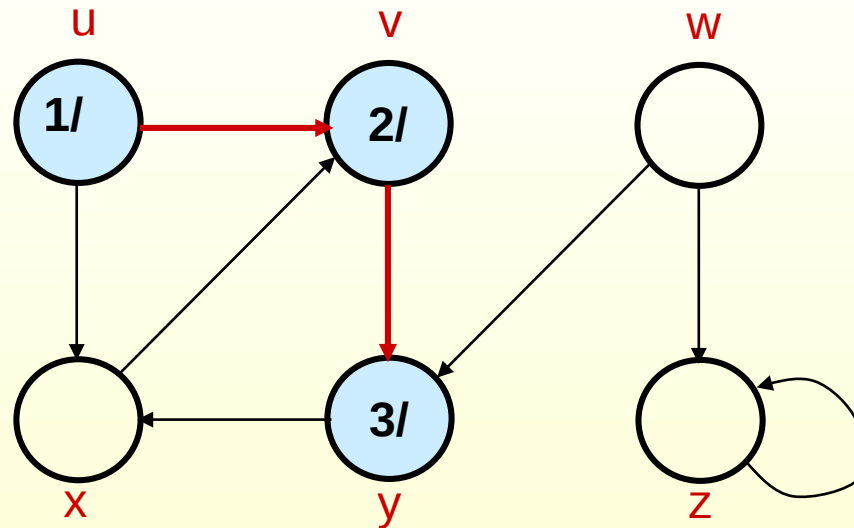
Example (DFS)



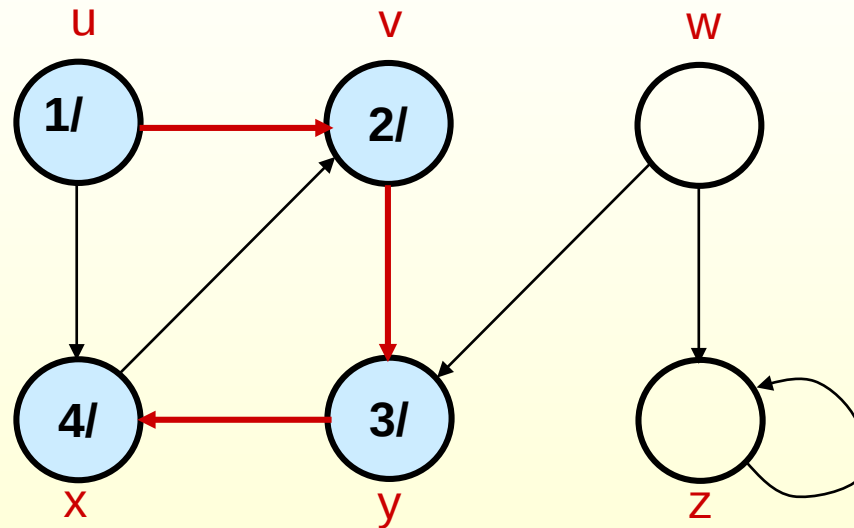
Example (DFS)



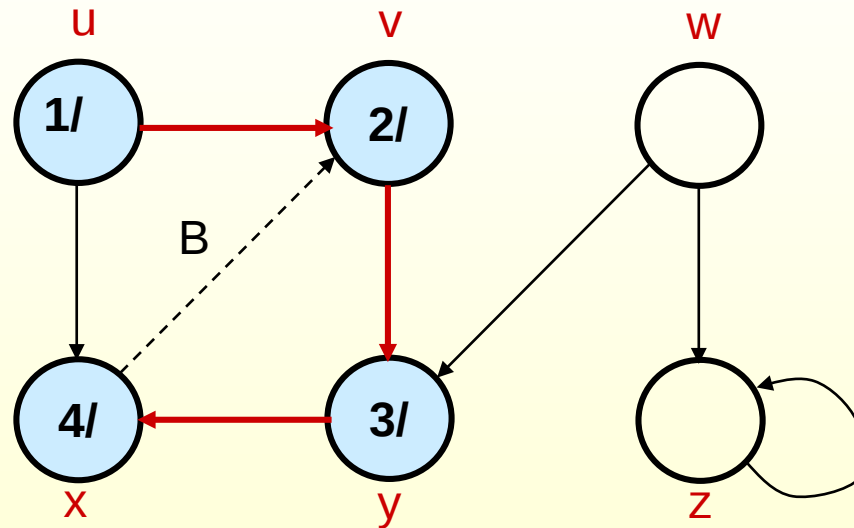
Example (DFS)



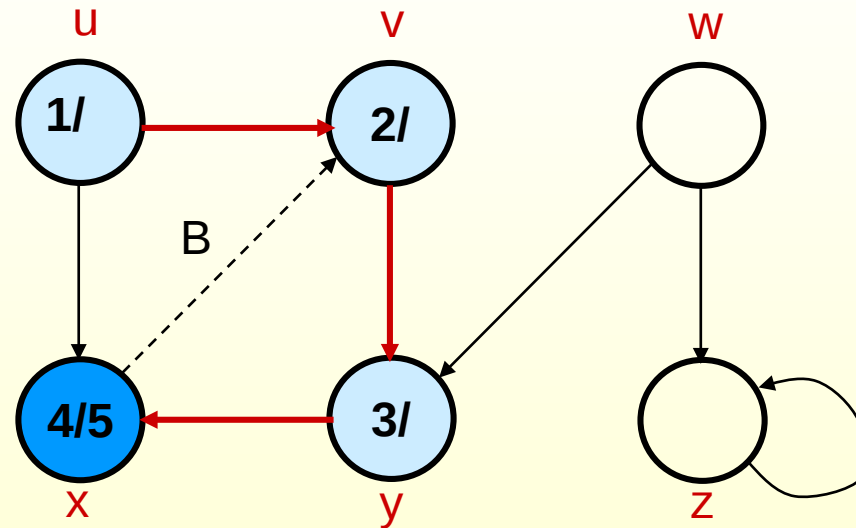
Example (DFS)



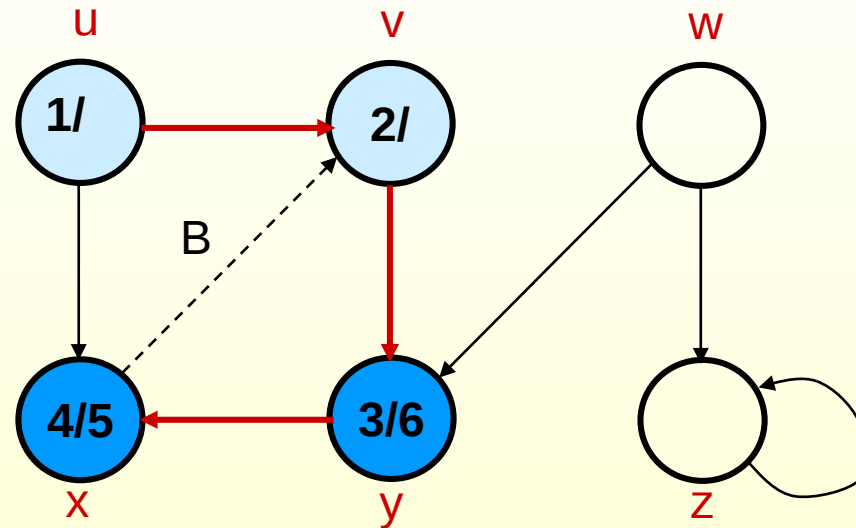
Example (DFS)



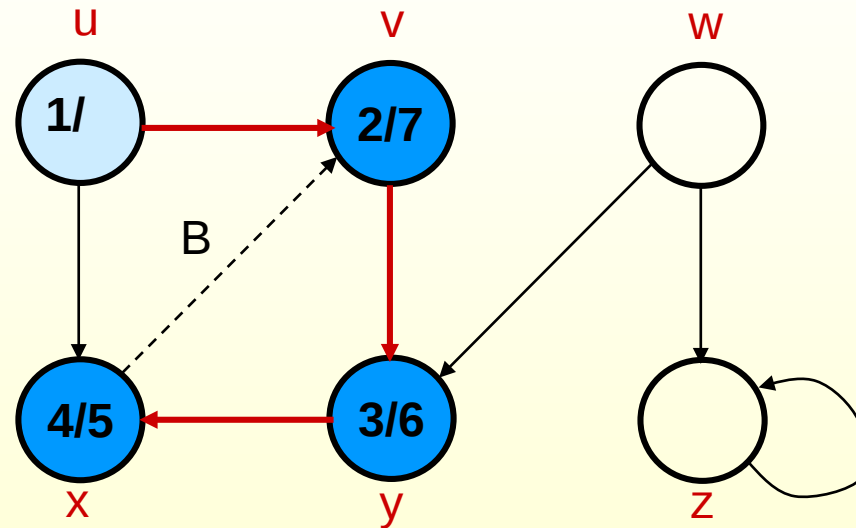
Example (DFS)



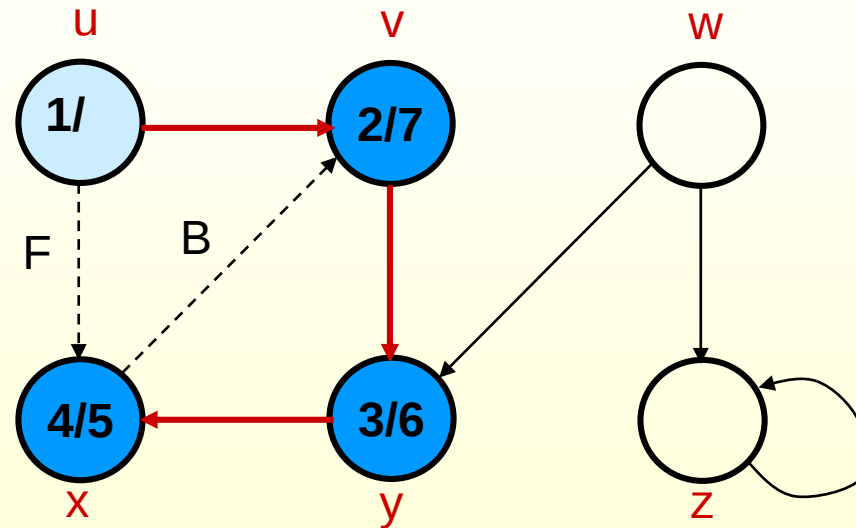
Example (DFS)



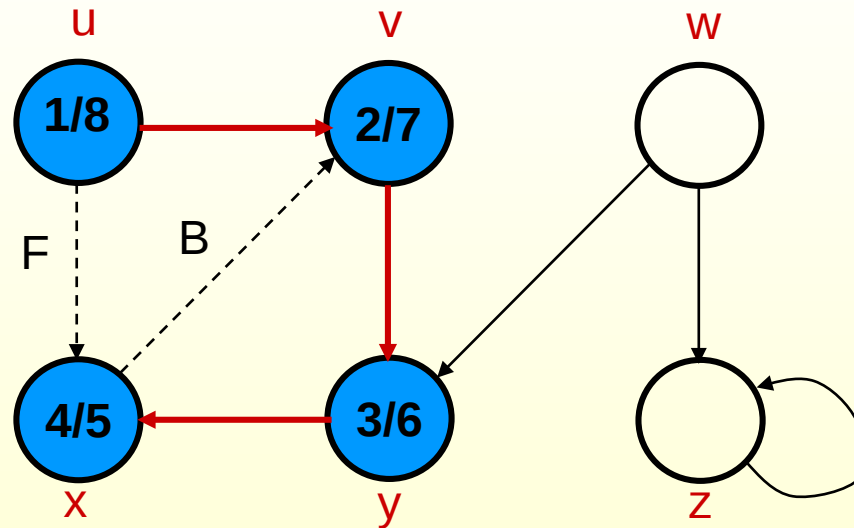
Example (DFS)



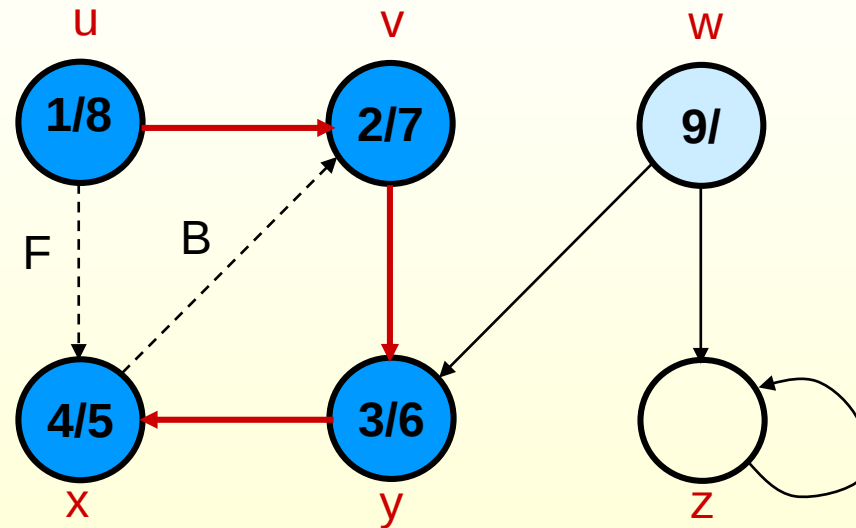
Example (DFS)



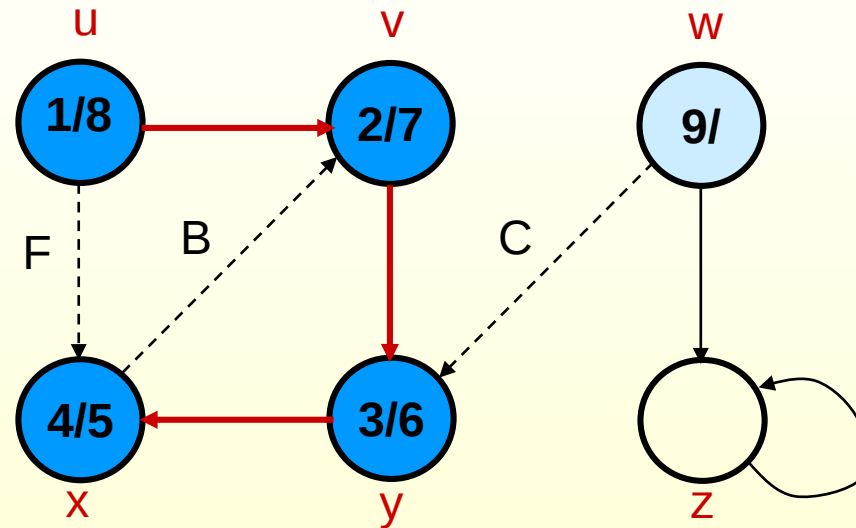
Example (DFS)



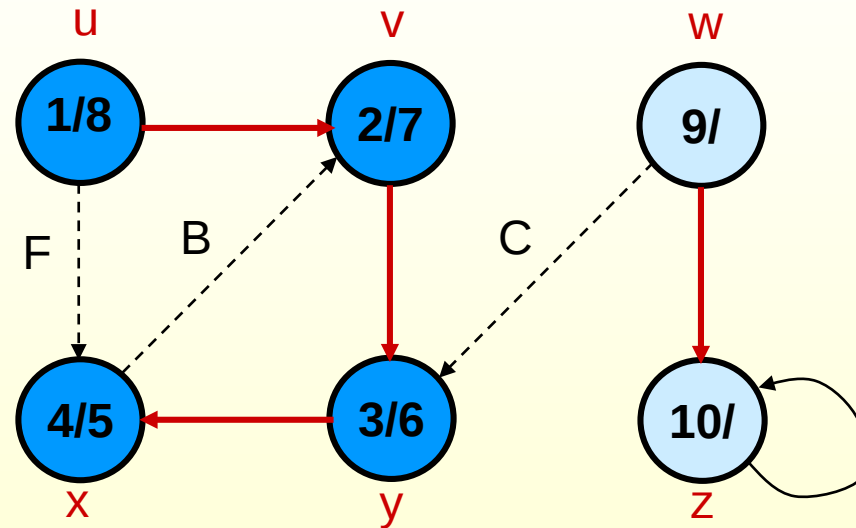
Example (DFS)



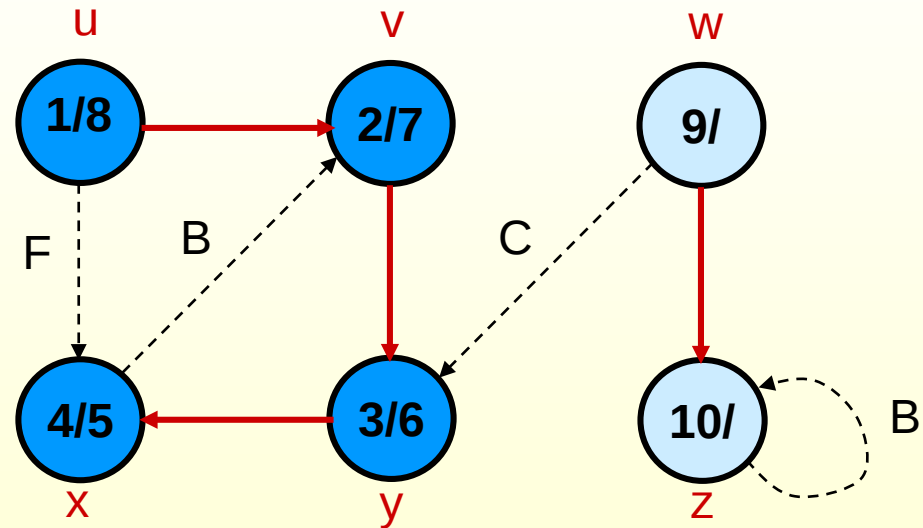
Example (DFS)



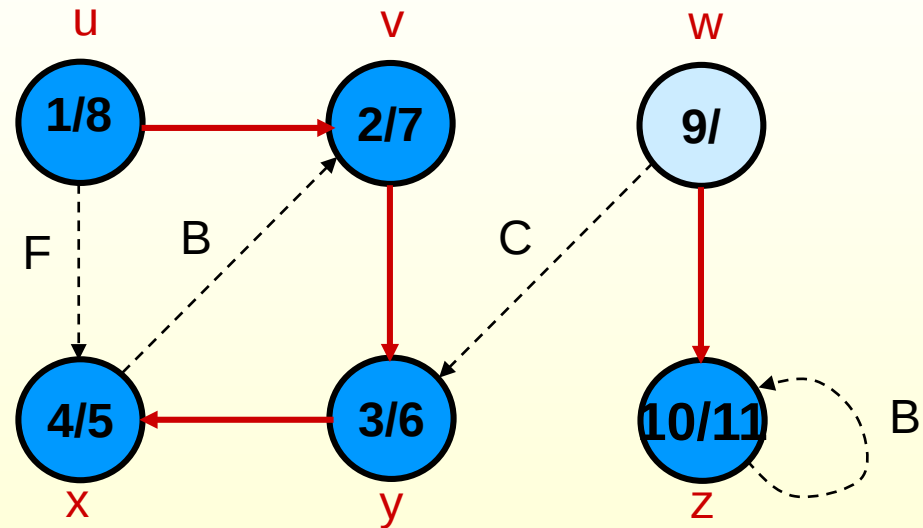
Example (DFS)



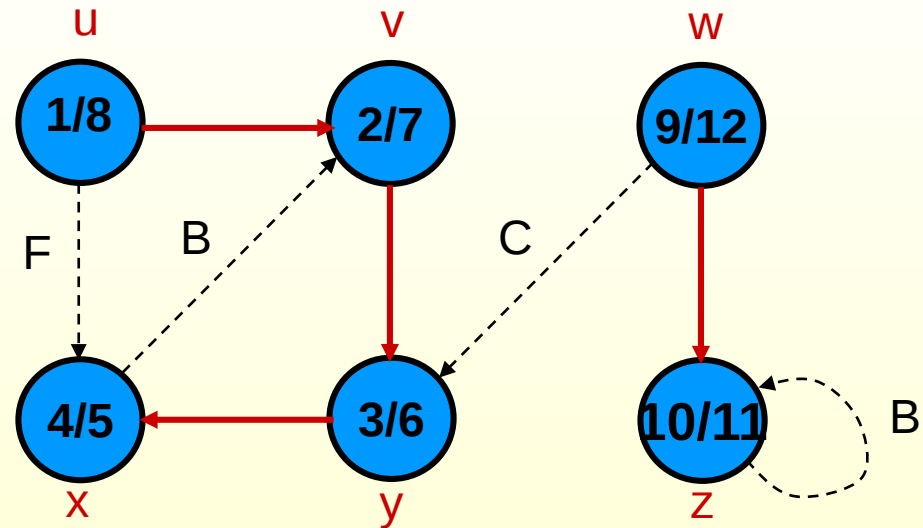
Example (DFS)



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 is painted gray the first time. Lines 3-6 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 $\Theta(V+E)$.

Parenthesis Theorem

Theorem 22.7

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

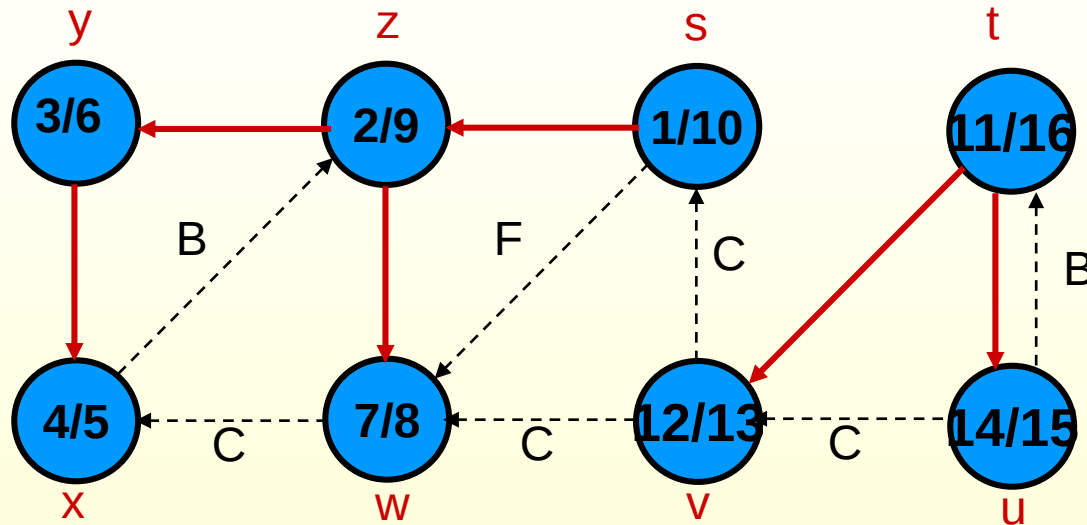
1. $d[u] < f[u] < d[v] < f[v]$ or $d[v] < f[v] < d[u] < f[u]$ and neither u nor v is a descendant of the other.
2. $d[u] < d[v] < f[v] < f[u]$ and v is a descendant of u .
3. $d[v] < d[u] < f[u] < f[v]$ and u is a descendant of v .

- ♦ 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.

White-Path Theorem

Theorem 22.9

v is a descendant of u if and only if at time $d[u]$, there is a path $u \rightsquigarrow v$ 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.

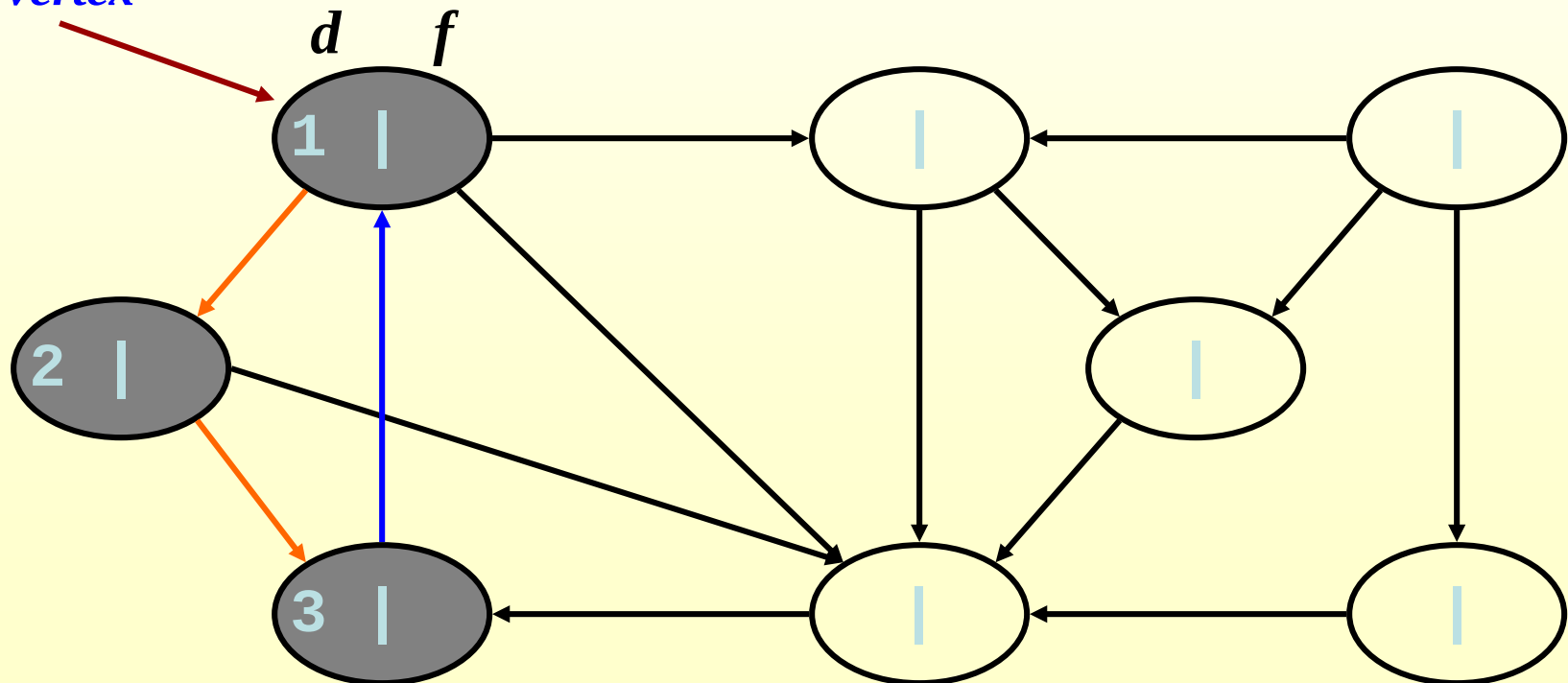
Theorem:

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

DFS and Cycles in a Graph

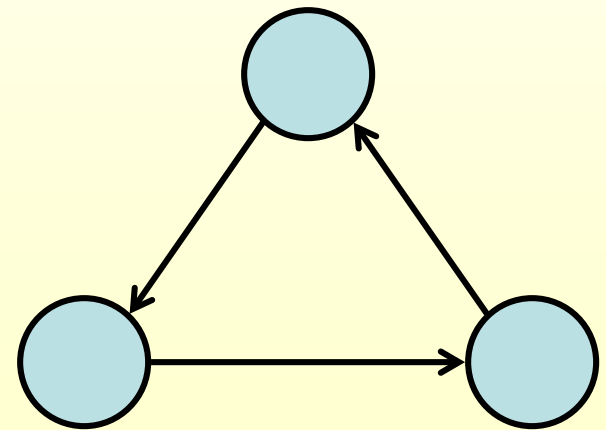
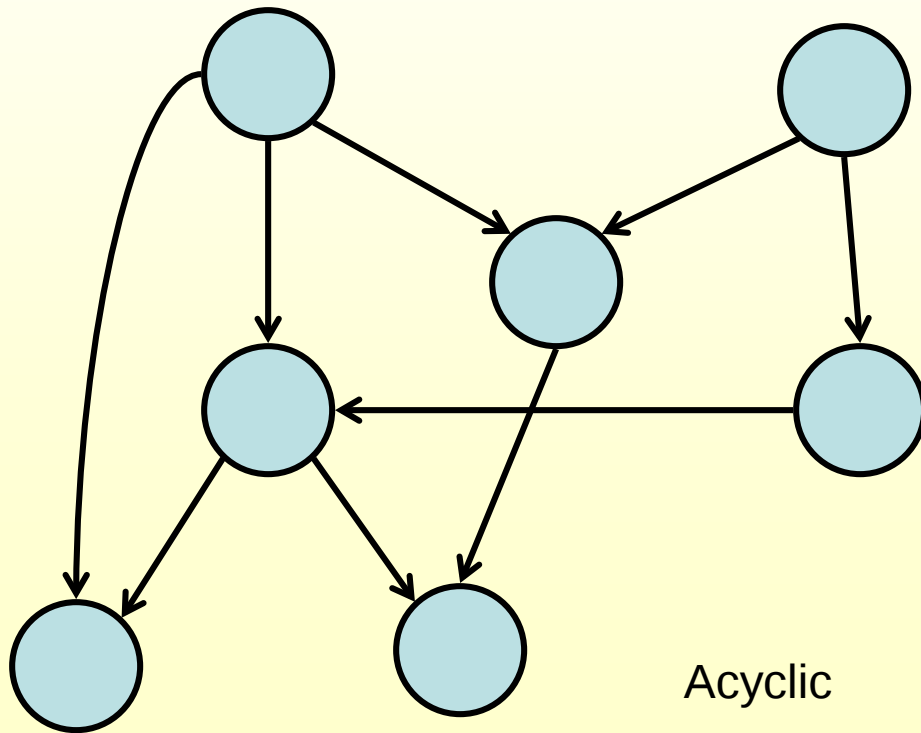
- ◆ A graph G is acyclic iff a DFS of G yields no back edges

source
vertex



Directed Acyclic Graphs

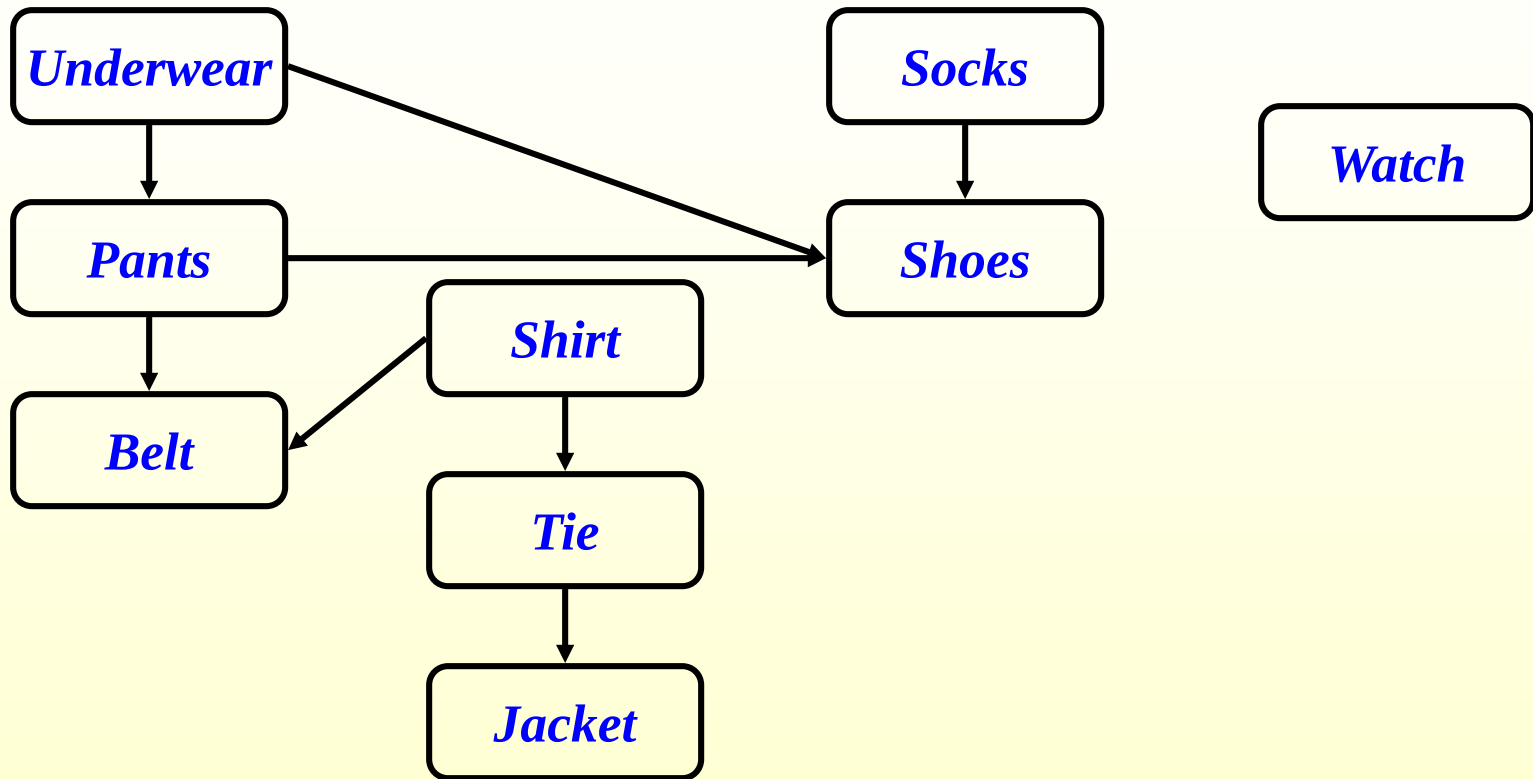
- ✦ A *directed acyclic graph* or *DAG* is a directed graph with no directed cycles:



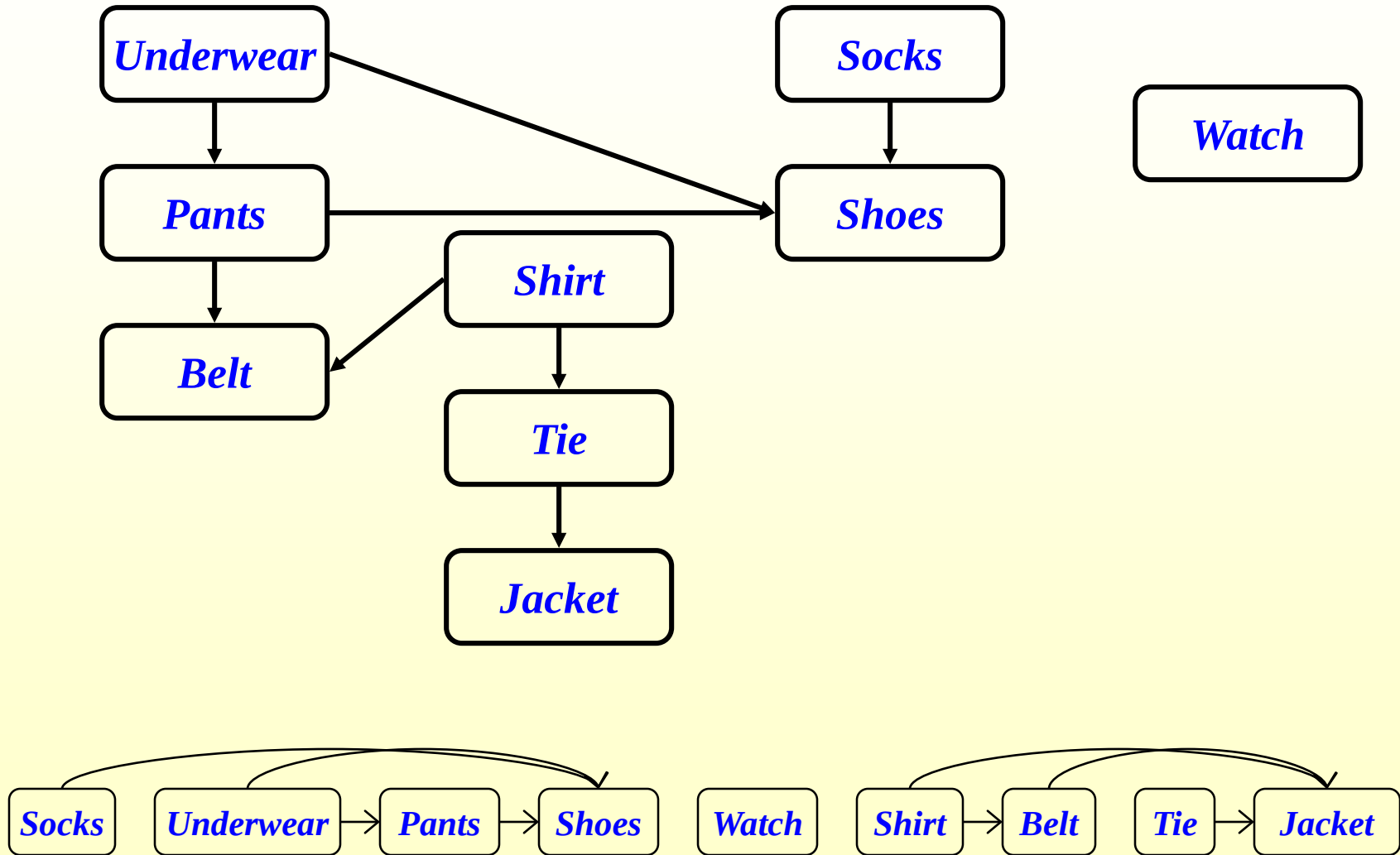
Topological Sort

- ◆ *Topological sort* of a DAG:
 - ◆ Linear ordering of all vertices in graph G such that vertex u comes before vertex v if edge $(u, v) \in G$
- ◆ Real-world example: getting dressed

Getting Dressed



Getting Dressed



Topological Sort Algorithm

Topological-Sort()

{ // condition: the graph is a DAG

Run DFS

When a vertex is finished, output it

**Vertices are output in reverse
topological order**

}

◆ **Time: $O(V+E)$**

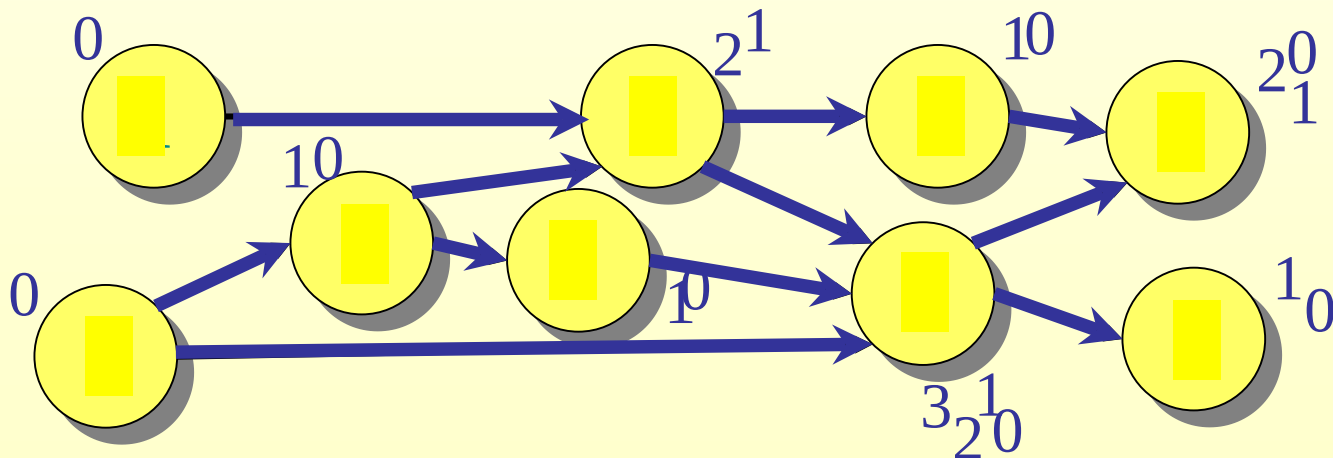
◆ **Correctness: Want to prove that**
$$(u,v) \in G \Rightarrow u \rightarrow f > v \rightarrow f$$

Correctness of Topological Sort

- ◆ Claim: $(u, v) \in G \Rightarrow u \rightarrow f > v \rightarrow f$
 - ◆ When (u, v) is explored, u is grey
 - ◆ $v = \text{grey} \Rightarrow (u, v)$ is back edge. Contradiction (*Why?*)
 - ◆ $v = \text{white} \Rightarrow v$ becomes descendent of $u \Rightarrow v \rightarrow f < u \rightarrow f$
(since must finish v before backtracking and finishing u)
 - ◆ $v = \text{black} \Rightarrow v$ already finished $\Rightarrow v \rightarrow f < u \rightarrow f$

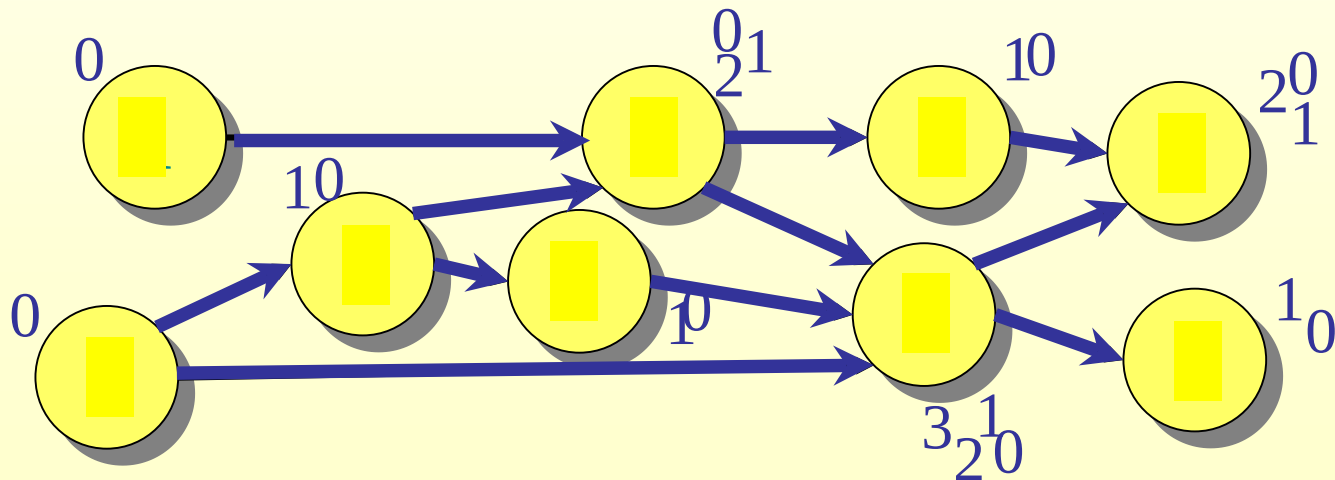
Another Algorithm

- Store vertices in a priority min-queue, with the **in-degree** of the vertex as the key
- While queue is not empty
 - Extract minimum vertex v , and give it next number
 - Decrease keys of all adjacent vertices by 1



Another Algorithm

- Store vertices in a priority min-queue, with the **in-degree** of the vertex as the key
- While queue is not empty
 - Extract minimum vertex v , and give it next number
 - Decrease keys of all adjacent vertices by 1



Topological Sort Runtime

Runtime:

- $O(|V|)$ to build heap + $O(|E|)$ DECREASE-KEY ops
⇒ $O(|V| + |E| \log |V|)$ with a binary heap
⇒ $O(|V|^2)$ with an array

Compare to DFS:

⇒ $O(|V| + |E|)$