

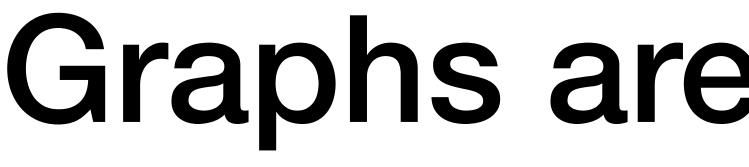
图及其遍历 Graphs and Graph Traversal

The slides are mainly adapted from the original ones shared by Chaodong Zheng and Kevin Wayne. Thanks for their supports!

钮鑫涛 Nanjing University 2024 Fall



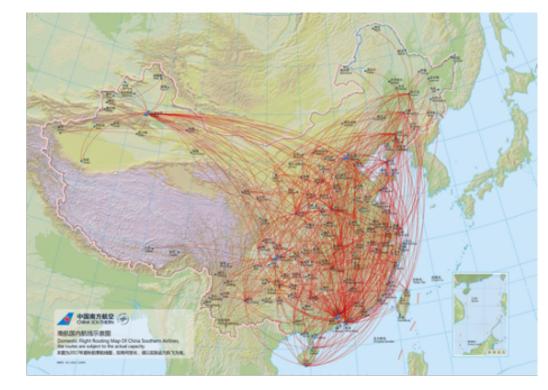


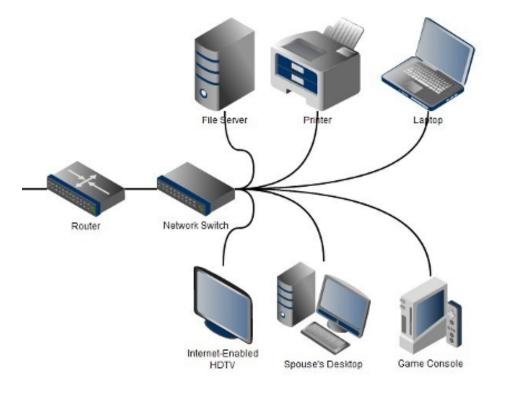


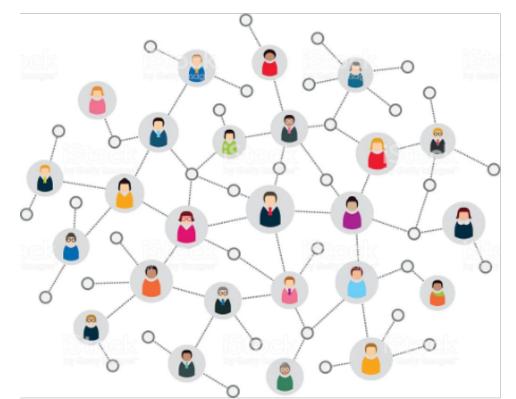
- Transportation Networks.
 - Nodes: Airports; Edges: Nonstop flights.
- Communication Networks.
 - Nodes: Computers; Edges: Physical links.
- Social Networks.
 - Nodes: People; Edges: Friendship.



Graphs are Everywhere!







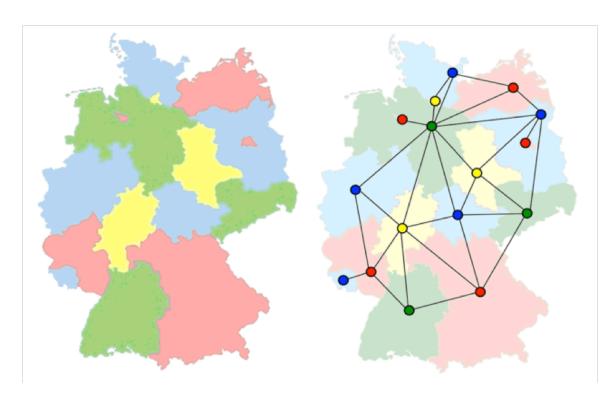




Followed by many graph problems

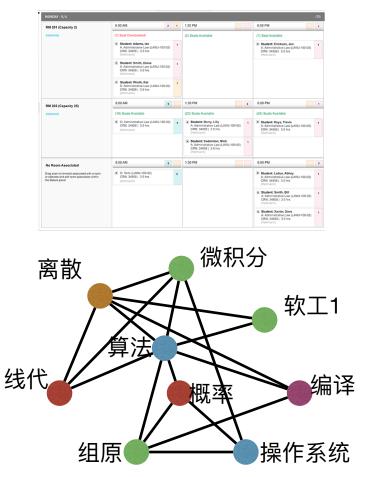
Coloring Maps

- Nodes: Countries; **Edges:** Neighboring countries.
- Question of Interest: Chromatic number?



Scheduling Exams

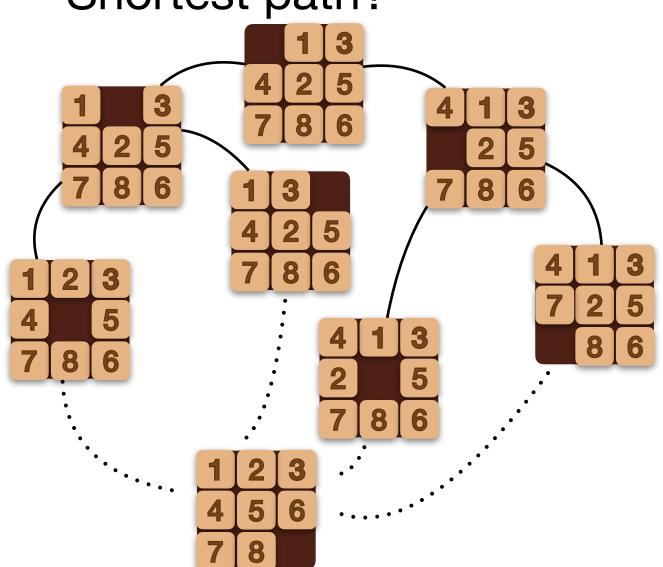
- Nodes: Exams; **Edges:** Conflicts.
- Question of Interest: Chromatic number?



both taken by at least one student

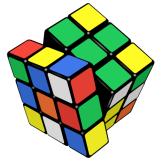
Solving Sliding Puzzle

- Nodes: States; Edges: Legit moves
- Question of Interest: Shortest path?

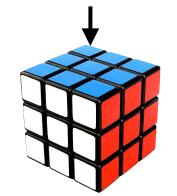


Solving Rubik's Cube

- Nodes: States; Edges: Legit moves
- Question of Interest: God's Number?



minimal number of turns?











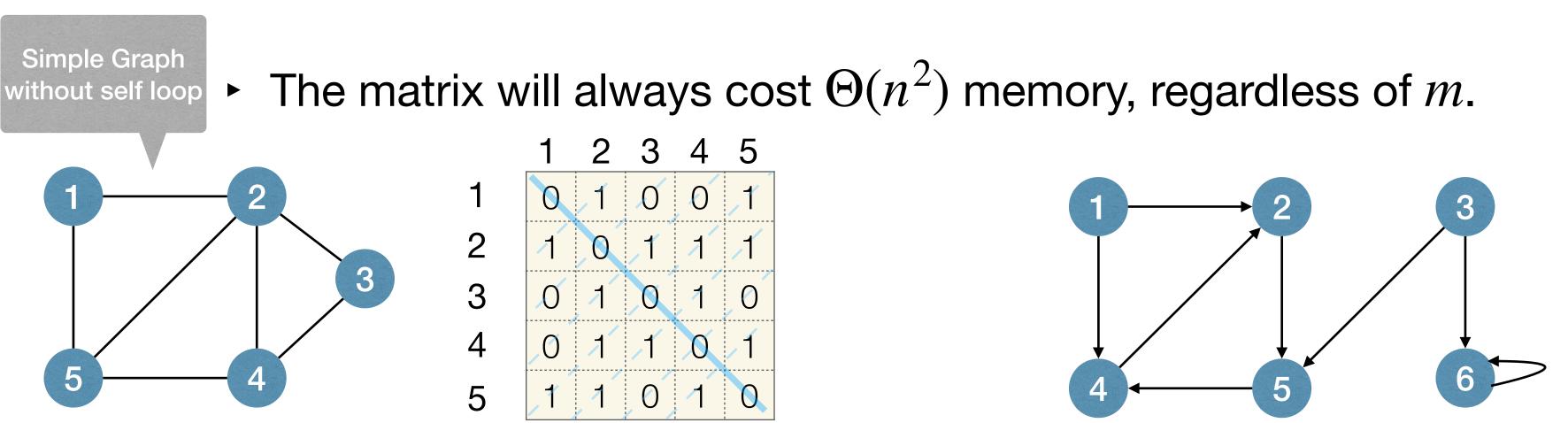


Representing graphs in computers

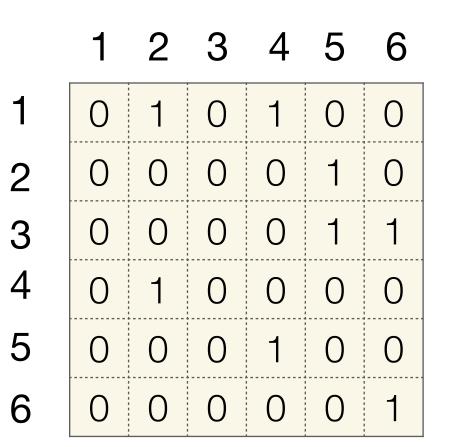
- Adjacency Matrix (邻接矩阵)
 - Consider a graph G = (V, E), where |V| = n and |E| = m
 - The Adjacency Matrix of G is an $n \times n$ matrix $A = (a_{ii})$ where

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

The matrix will be symmetry if G is undirected.



Quick Question: What does A^2 mean, if anything?

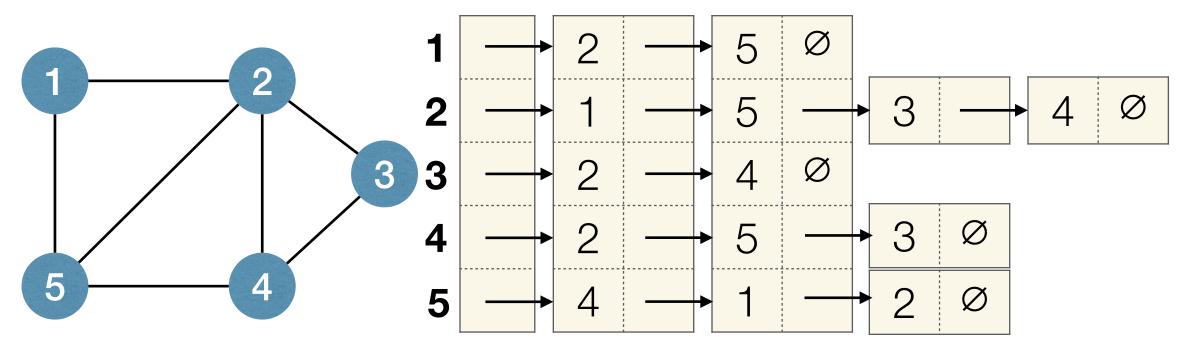


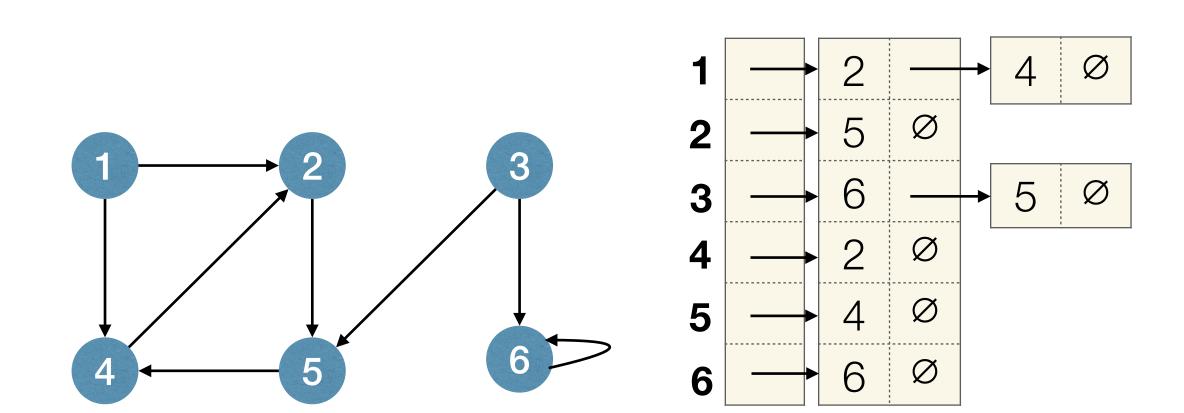




Representing graphs in computers

- Adjacency List (邻接表)
 - Consider a graph G = (V, E), where |V| = n and |E| = m
 - The Adjacency List of G is a collection of n lists:
 - One for each vertex $u \in V$
 - In the list for u, vertex v exists iff edge $(u, v) \in E$
 - Each edge appears twice if G is undirected.
 - The space cost is $\Theta(n+m)$









Adjacency Matrix

- **Fast Query**: Are *u* and *v* neighbors?
- **Slow Query**: Find me any neighbor of *u*.
- **Slow Query**: Enumerate all neighbors of *u*.

Queries: What types of queries are needed and/or frequent? Space usage: Is the graph dense or sparse?

Irade

)-offs

Adjacency Matrix and Adjacency List

Adjacency List

- **Fast Query**: Find me any neighbor of *u*.
- **Fast Query**: Enumerate all neighbors of *u*.
- **Slow Query**: Are *u* and *v* neighbors?

Important question to ask







Searching in a Graph (or, Graph Traversal)

- **Goal**: Start at source node s and find some node t.
- Or: Visit all nodes reachable from s.
- Two Basic Strategies:
 - Breadth-First Search (BFS)
 - Depth-First Search (DFS)
- Many applications, beside searching and traversal!

• Usually use adjacency list when discussing BFS/DFS. (At least in this course...)





Breadth-First Search

Surger and the second



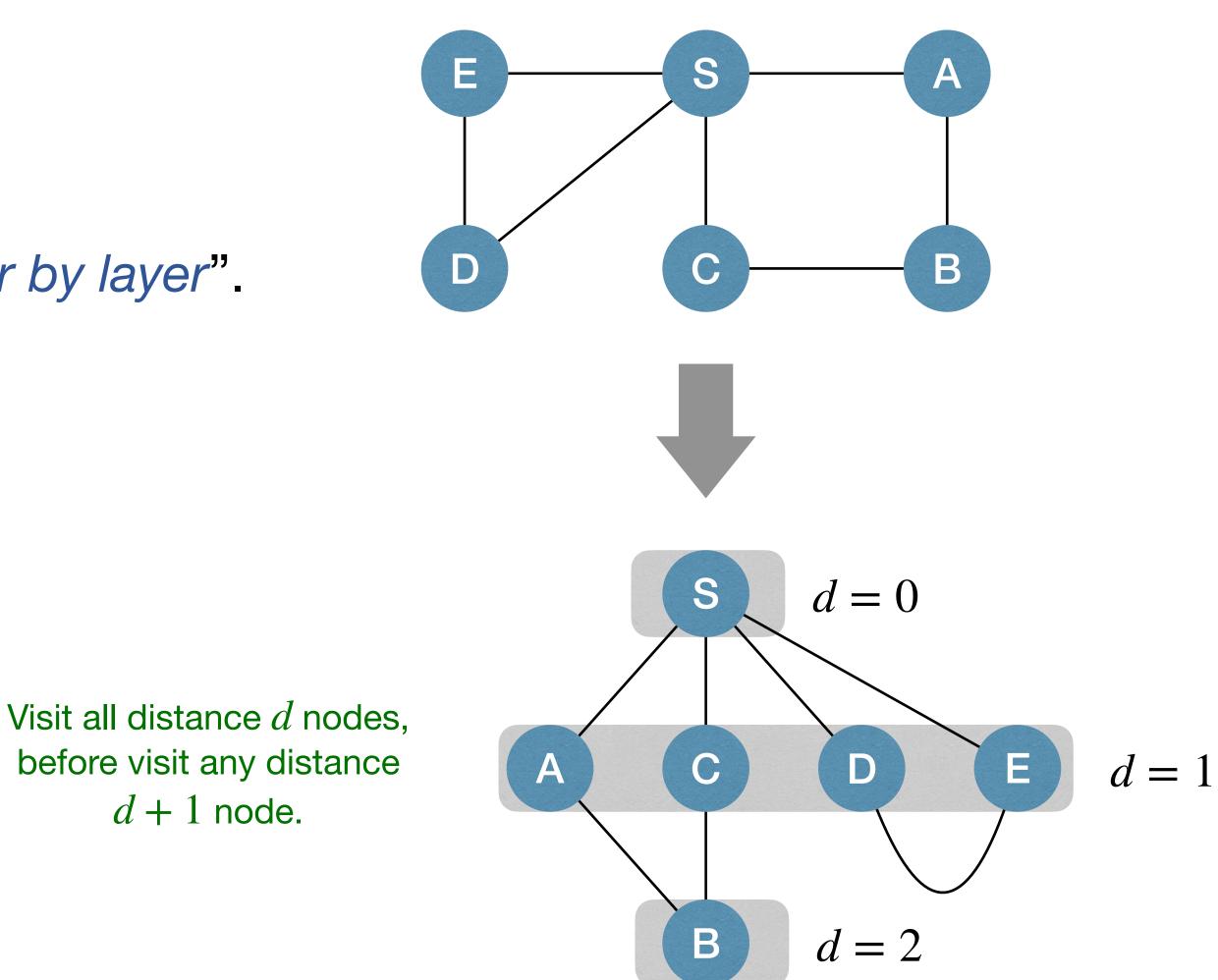


. . .

Breadth-First Search (BFS)

- Basic Idea of BFS:
 - Start at the source node s;
 - Visit other nodes (reachable from s) "layer by layer".
- More precise description:
 - Start at the source node s;
 - Visit nodes at *distance* 1 from s;
 - Visit nodes at *distance* 2 from s;

These nodes are neighbors of distance 1 nodes!





BFS Implementation

- How to implement BFS? (Hint: recall traversal-by-layer in trees) **<u>BFSSkeleton(G, s)</u>**:
 - Use a FIFO Queue!
- Nodes have 3 status: \bullet
 - Undiscovered ([WHITE]): Not in queue yet.
 - Discovered but not visited (GRAY): In queue but not processed.
 - Visited (BLACK): Ejected from queue and processed.

We can "store" a shortest path, instead of only "computing" the length of the path —> by additionally recording the node's parent info.

for each *u* in *V u.dist* := *INF*, *u.discovered* := *False* s.dist := 0, s.discovered := True Q.enque(s)while !Q.empty() u := Q.dequeue()for each edge (u, v) in E if !v.discovered v.dist := u.dist + 1v.discovered := True Q.enque(v)



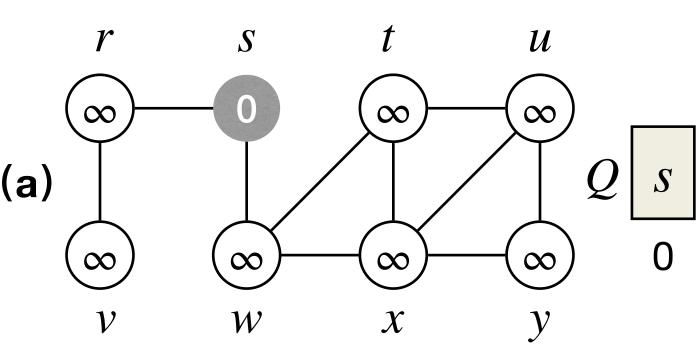


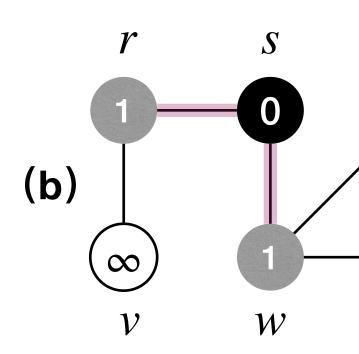


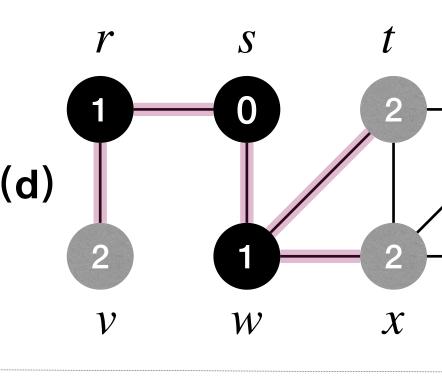


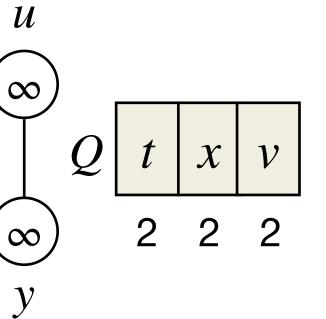
BFS Implementation <u>BFS(G, s):</u> for each *u* in *V* u.c := WHITE, u.d := INF, u.p := NILs.c := GRAY, s.d := 0, s.p := NILQ.enque(s)while !Q.empty() u := Q.dequeue()u.c := BLACKfor each edge (u, v) in Eif v.c = WHITEv.c := GRAYv.d := u.d + 1v.p := uQ.enque(v)

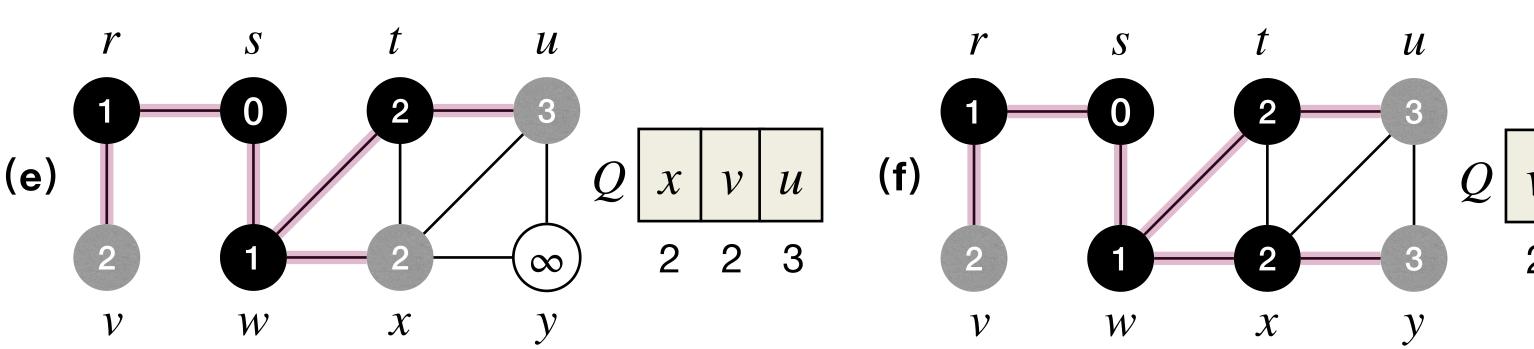


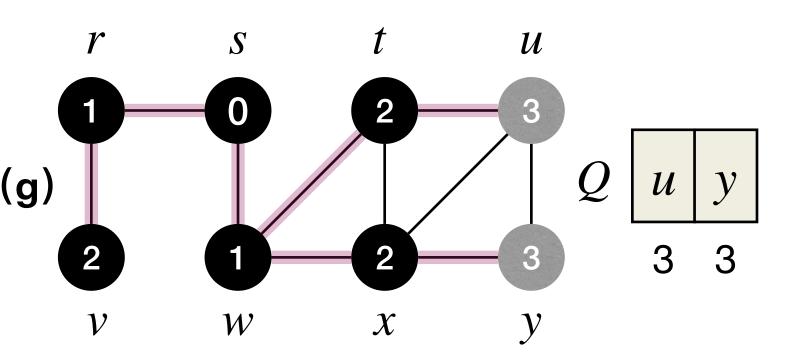


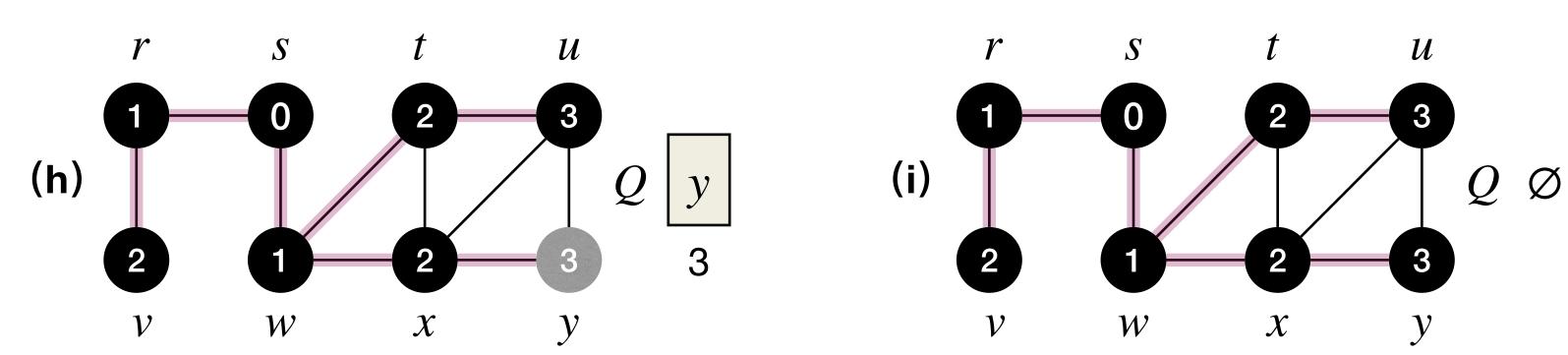




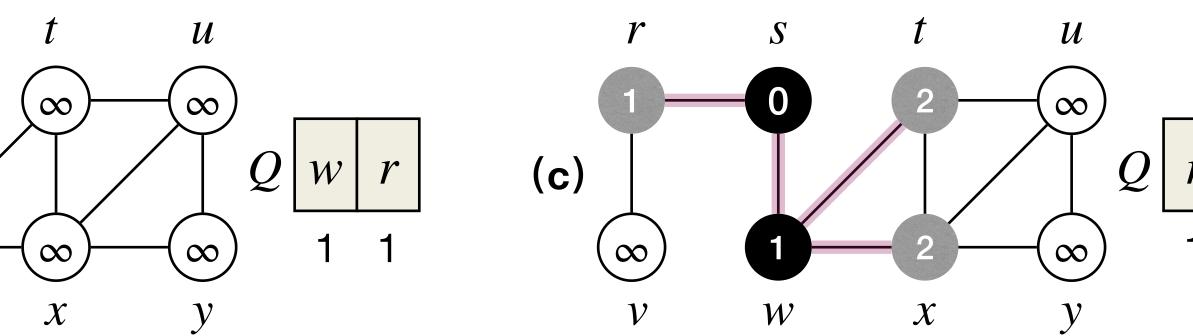


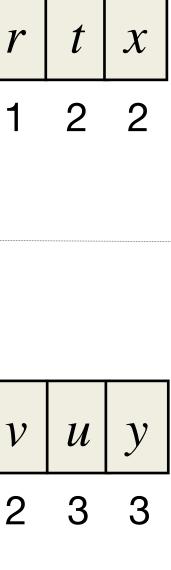






Sample Execution









Performance of BFS

- Runtime of BFS? (Assuming G is connected.)
- "While" loop $\Theta(n)$ times.
 - Each node in Q at most once.
- "For" loop $\Theta(m)$ times.
 - Each edge visited at most once or twice.
- Runtime of BFS is $\Theta(n+m)$.

What if we use adjacency matrix instead of adjacency list?

<u>BFS(G, s):</u>

for each *u* in *V* u.c := WHITE, u.d := INF, u.p := NILs.c := GRAY, s.d := 0, s.p := NILQ.enque(s)while !*Q.empty*() u := Q.dequeue()u.c := BLACKfor each edge (u, v) in Eif v.c = WHITEv.c := GRAYv.d := u.d + 1v.p := uQ.enque(v)







Theorem BFS visits a node iff it is reachable from s.

Proof:

- moves along edges.
- [*if*] If a node is reachable from s, then BFS visits it.
 - Claim: For all $k \ge 0$, all nodes within k hops of s are visited.
 - [Basis]: Clearly s is visited.
 - [Hypothesis]: All nodes within k-1 hops of s are visited.

Correctness and Properties of BFS

• [only if] If a node is not reachable from s, then BFS does not visit it, since BFS only





Correctness and Properties of BFS

Theorem BFS visits a node iff it is reachable from *s*.

- v's neighbor on (one of) v's shortest path back to s
 - By induction hypothesis, *u* gets visited.

 - Either way, v eventually gets visited.

- [Inductive Step]: Consider a node v that is k hops away from s. Let u be

When BFS visits u, node v is already **GRAY** or **BLACK**, or will be put in Q.

Will this really happen?!







Theorem BFS correctly computes *u.dist*, for every node *u* that is reachable from s

- [*Proof Idea*] Use induction to show:
 - For all $d \ge 0$, there is a moment at which:
 - (a) every node u with $dist(s, u) \leq d$ correctly computes u.dist;
 - (b) every other node v has $v.dist = \infty$;
 - (c) Q contains exactly the nodes d hops away from s.

Correctness and Properties of BFS





Correctness and Properties of BFS

Theorem BFS correctly computes *u.dist*, for every node *u* that is reachable from s

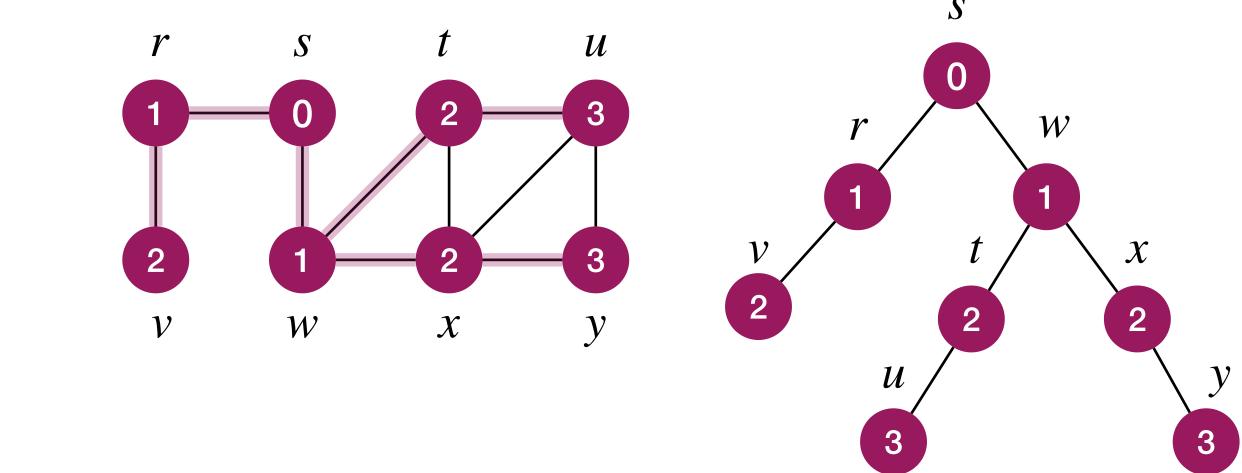
s to u is a shortest path from s to u. p followed by the edge $(u \cdot p, u)$

• $G_p = (V_p, E_p)$ is a **breadth-first tree**, which can print on a shortest path from any node v to the source node s. Here:

•
$$V_p = \{u \in V : u \cdot p \neq \mathsf{NIL}\} \cup \{s\}$$

•
$$E_p = \left\{ (u \cdot p, u) : u \in V_p - \{s\} \right\}$$

Corollary For any $u \neq s$ that is reachable from s, one of the shortest path from





One last note on BFS

- What if the graph is not connected?
 - Easy, do a BFS for each connected component! for each u in V

Runtime of this procedure?

<u>BFS(G):</u>

for each u in V u.c := WHITE, u.d := INF, u.p := NILif u.c = WHITEu.c := GRAY, u.d := 0, u.p := NILQ.enque(u)while !*Q.empty*() v := Q.dequeue()v.c := BLACKfor each edge (v, w) in E if w.c = WHITEw.c := GRAYw.d := v.d + 1w.p := wQ.enque(w)





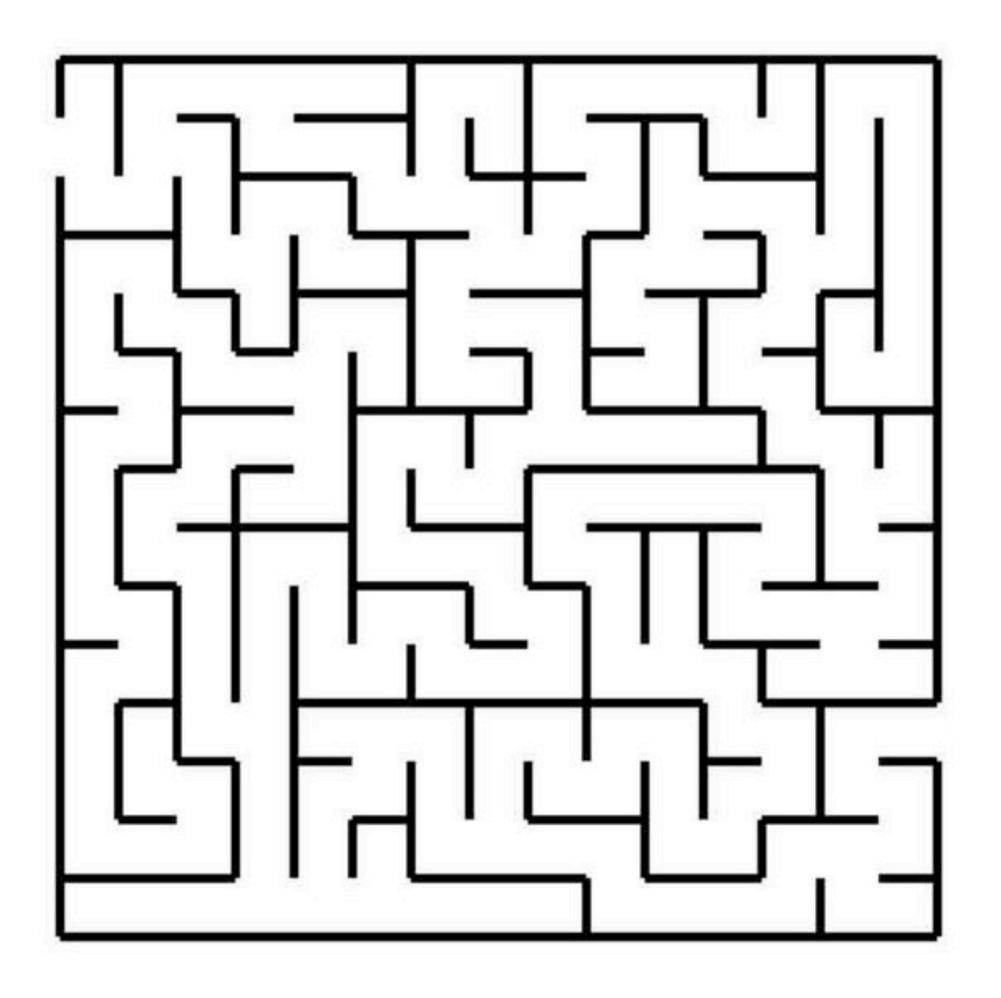


Depth-First Search



Depth-First Search (DFS)

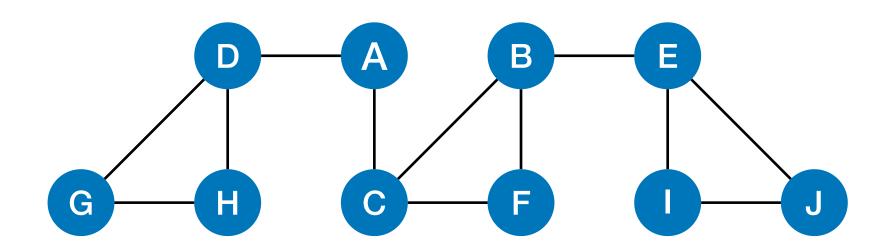
- Much like exploring a maze:
 - Use a ball of string and a piece of chalk.
 - Follow path (unwind string and mark at intersections), until stuck (reach dead-end or already-visited place).
 - Backtrack (rewind string), until find unexplored neighbor (intersection with unexplored direction).
 - Repeat above two steps.





Depth-First Search (DFS)

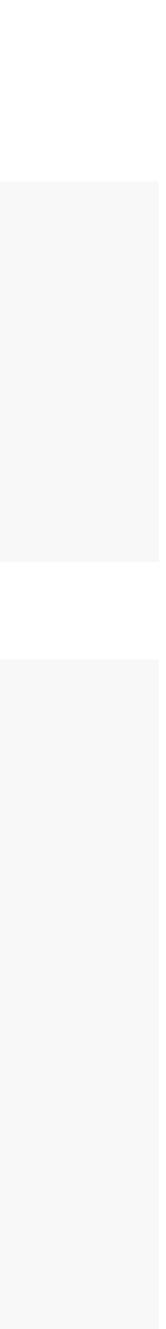
- How to do this for a graph, in computer?
 - Chalk: boolean variables.
 - String: a stack.



DFSSkeleton(G, s):

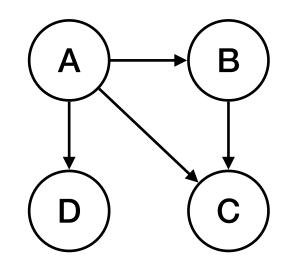
s.visited := *True* for each edge (s, v) in Eif !v.visited DFSSkelecton(G, v)

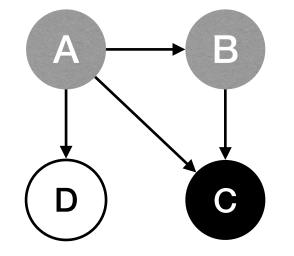
DFSIterSkeleton(G, s): Stack Q Q.push(s)while !Q.empty() u := Q.pop()if !u.visited *u.visited* := *True* for each edge (u, v) in EQ.push(v)





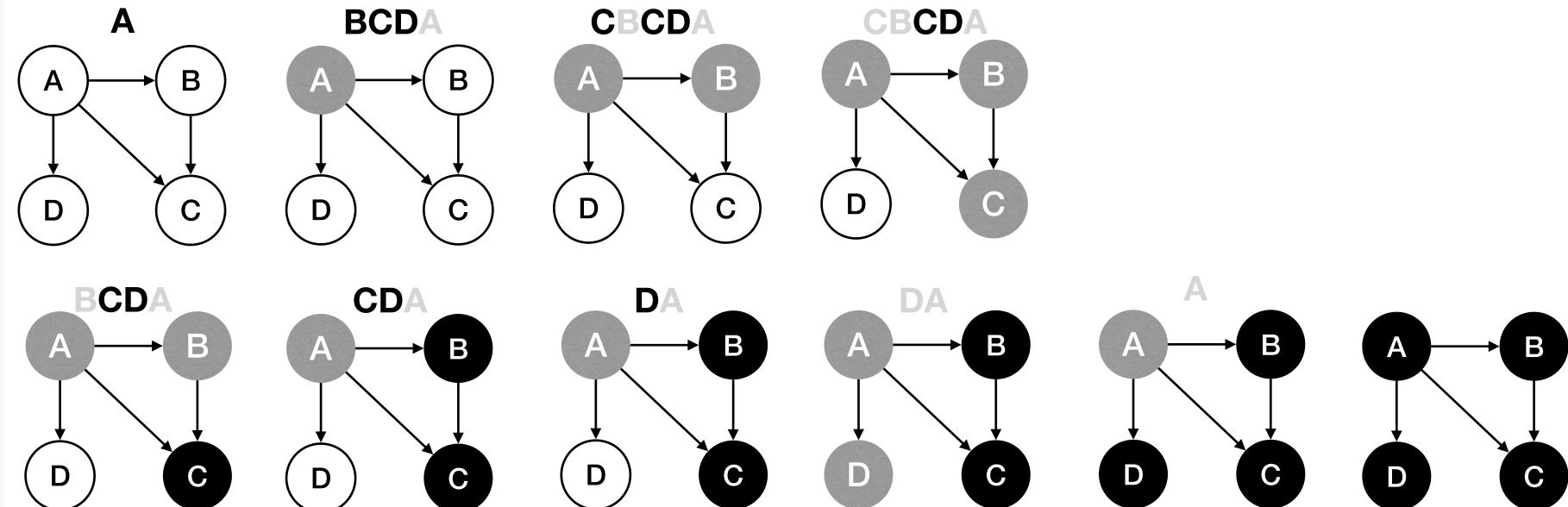
DFSSkeleton(G, s): *s.visited* := *True* for each edge (s, v) in Eif !v.visited DFSSkelecton(G, v)

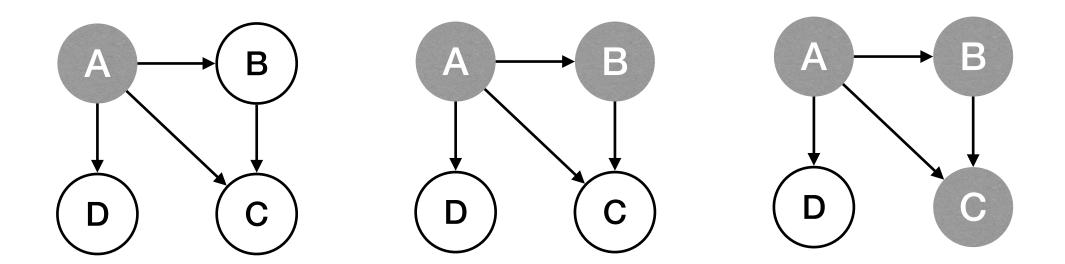


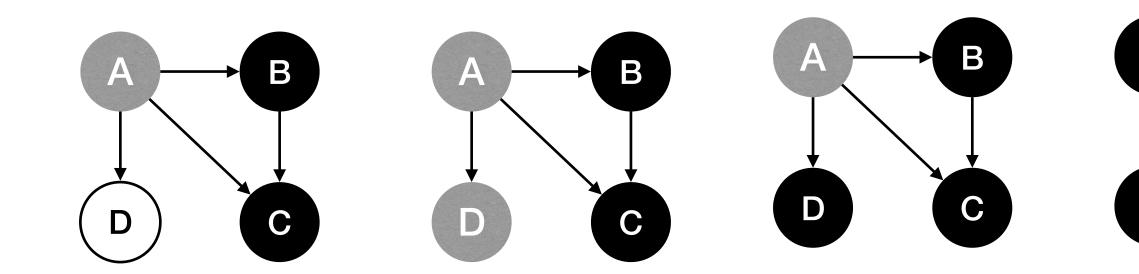


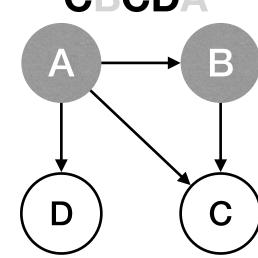
DFSIterSkeleton(G, s):

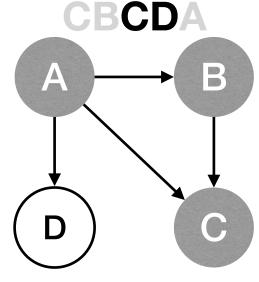
Stack Q Q.push(s)while !Q.empty() u := Q.pop()if !u.visited *u.visited* := *True* for each edge (u, v) in EQ.push(v)

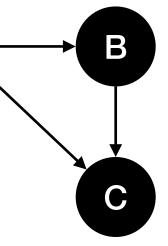














Depth-First Search (DFS)

- What if the graph is not (strongly) connected?
 - Do DFS from multiple sources.

DFSAll(G): for each node *u* in *V* v.visited := False for each node *u* in *V* if !u.visited DFSSkelecton(G, u)DFSSkeleton(G, s): *s.visited* := *True* for each edge (s, v) in E if !v.visited DFSSkelecton(G, v)

DFSAll(G): for each node *u* in *V v.visited* := *False* for each node *u* in *V* if !u.visited DFSIterSkelecton(G, u) DFSIterSkeleton(G, s): Stack Q Q.push(s)while !Q.empty() u := Q.pop()if !u.visited *u.visited* := *True* for each edge (u, v) in EQ.push(v)

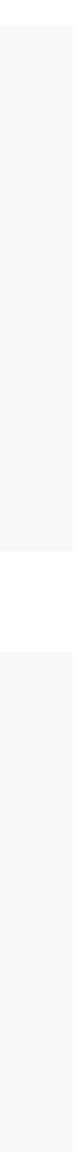


Depth-First Search (DFS) DFSAll(G):

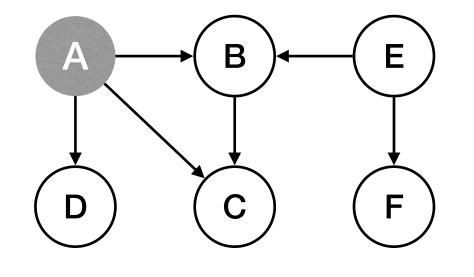
- Each node *u* have 3 status during DFS:
 - Undiscovered [(WHITE)]: before calling DFSSkeleton (G, u)
 - Discovered [GRAY]: during execution of DFSSkeleton (G, u)
 - Finished [BLACK]: DFSSkeleton (G, u) returned
- DFS(G,u) builds a tree among nodes reachable from u:
 - Root of this tree is u.
 - For each non-root, its parent is the node that makes it turn GRAY.
- DFS on entire graph builds a forest.

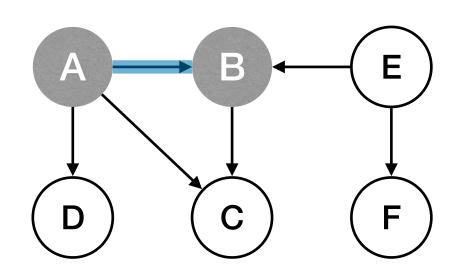
for each node *u* in *V* u.color := WHITE*u.parent* := *NIL* for each node *u* in *V* if u.color = WHITEDFS(G, u)

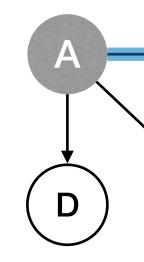
<u>DFS(G, s):</u> s.color := GRAYfor each edge (s, v) in Eif v.color = WHITEv.parent := sDFS(G, v)s.color := BLACK

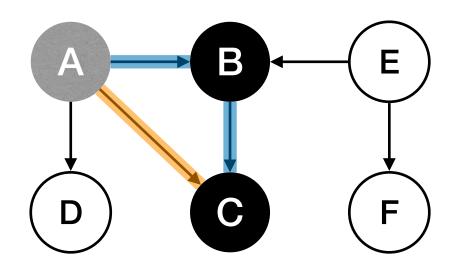


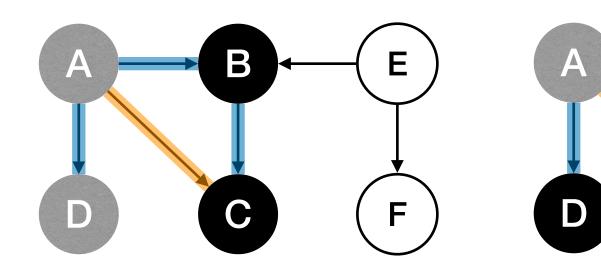


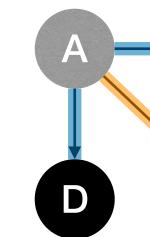


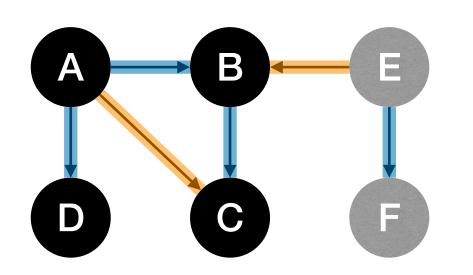


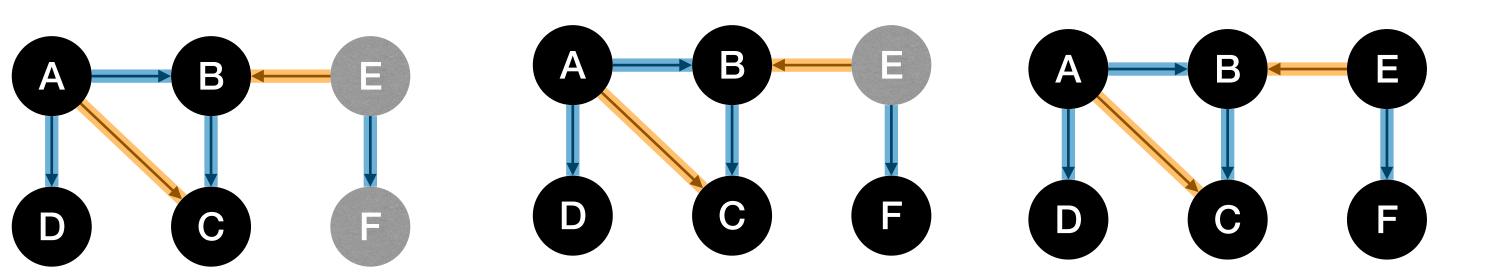


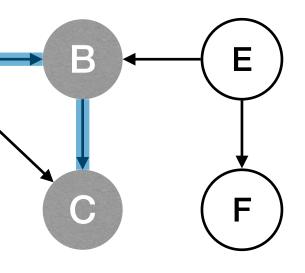


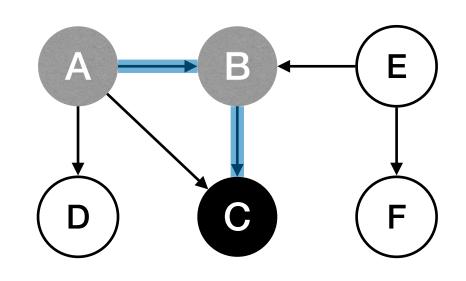


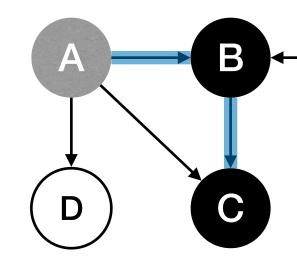


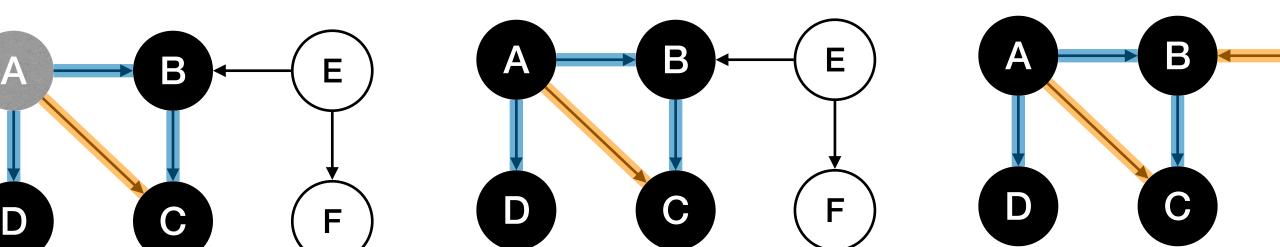


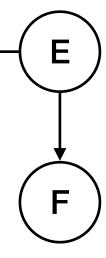


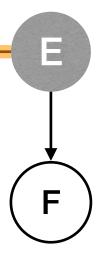














Depth-First Search (DFS)

- DFS provides (at least) two chances to process each node: 4 Sample application: Track active intervals of nodes
 - **Pre-Visit:** WHITE \rightarrow GRAY
 - ▶ **Post-Visit:** GRAY → BLACK

DFSAll(G):

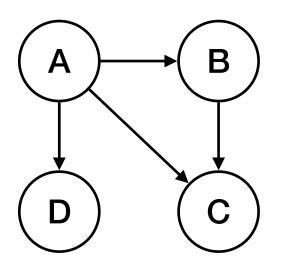
PreProcess(G)for each node u in V u.color := WHITE u.parent := NIL for each node u in V if !u.visited DFS(G, u) DFS(G, s):
PreVisit(s)
s.color := GRAY
for each edge (s, v) in E
 if v.color = WHITE
 v.parent := s
 DFS(G, v)
s.color := BLACK
PostVisit(s)

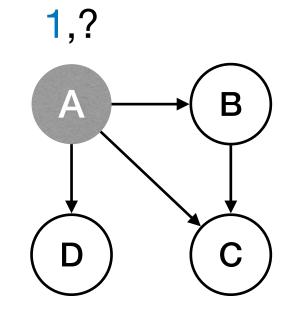
- Clock ticks whenever some node's color changes.
- **Discovery time**: when the node turns to GRAY.
- Finish time: when the node turns to BLACK.

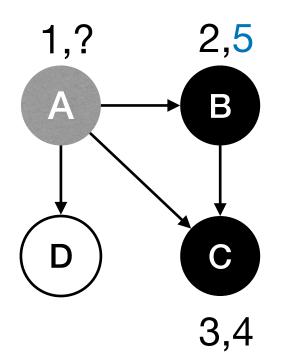
PreProcess(G):
time := 0PreVisit(s):
time := time + 1
s.d := timeNote: here it
indicates the
discovery timePostProcess(G):
time := time + 1
s.f := time

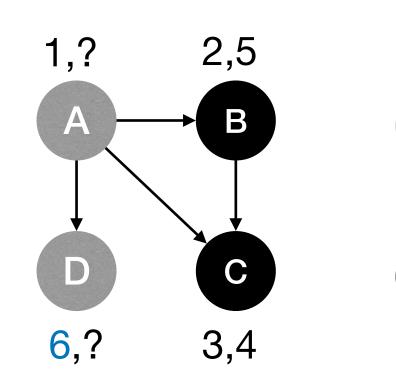












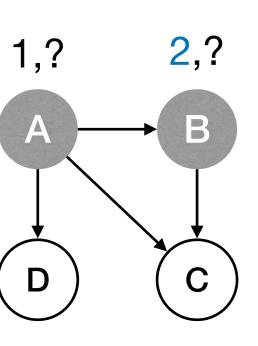
DFSAll(G):

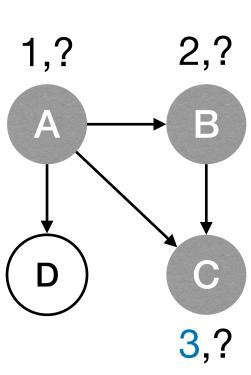
PreProcess(*G*)

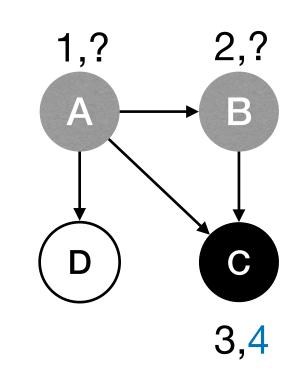
for each node *u* in *V* u.color := WHITE*u.parent* := *NIL* for each node *u* in *V* if !u.visited DFS(G, u)

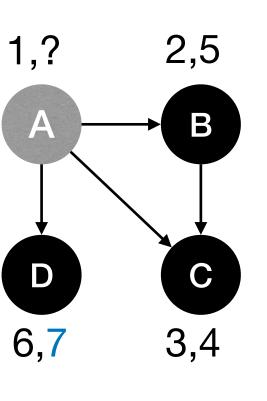
DFS(G, s): *PreVisit(s)* s.color := GRAYfor each edge (s, v) in Eif *v.color* = WHITE v.parent := sDFS(G, v)s.color := BLACK

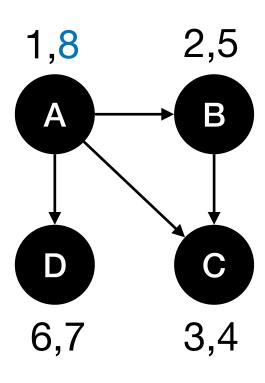
PostVisit(s)











PreProcess(G): time := 0

PreVisit(s): time := time + 1s.d := time

PostProcess(G): time := time + 1 $s_f := time$



Runtime of DFS

- Time spent on each node: O(1)
 - ► DFS(G,u) is called once for each node u.
- Time spent on each edge: O(1)
 - Each edge is examined O(1) times.

DFSAll(G):	DFS(G, s)
PreProcess(G)	PreVisit(s)
for each node <i>u</i> in <i>V</i>	s.color := G.
u.color := WHITE	for each edg
u.parent := NIL	if v.col
for each node <i>u</i> in <i>V</i>	v.p
if !u.visited	Di
DFS(G, u)	s.color := BI
	PostVisit(s)

Total runtime: O(n + m)

<u>S):</u>

GRAY dge (s, v) in Eolor = WHITE.parent := sDFS(G, v)BLACK

PreProcess(G):

time := 0

PreVisit(s): time := time + 1s.d := time

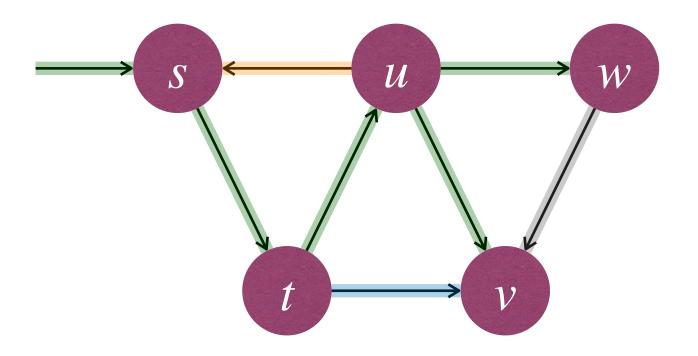
PostProcess(G): time := time + 1 $s_{f} := time$



Classification of edges

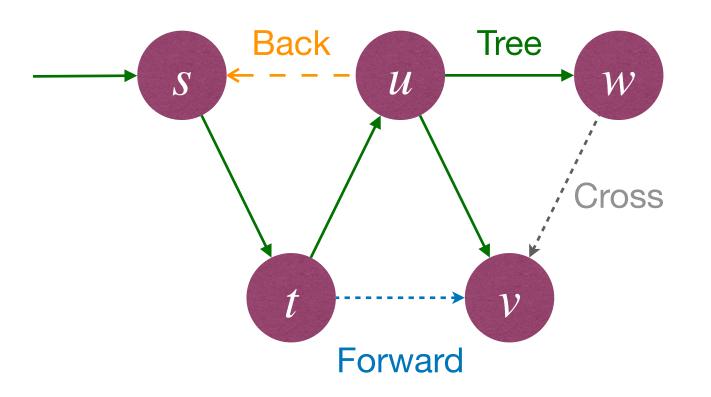
- DFS process classify edges of input graph into four types.
 - Tree Edges: Edges in the DFS forest.
 - Back Edges: Edges (u, v) connecting u to an ancestor v in a DFS tree.

 - descendant relation, or connecting nodes in different DFS trees.)

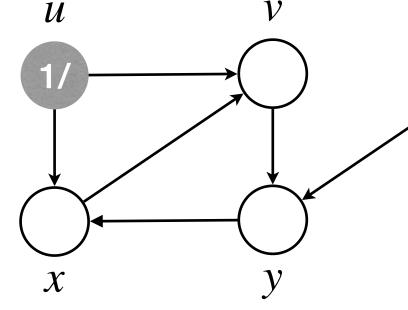


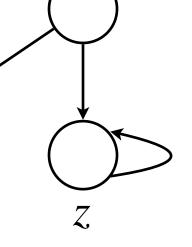
• Forward Edges: Non-tree edges (u, v) connecting u to a descendant v in a DFS tree.

Cross Edges: Other edges. (Connecting nodes in same DFS tree with no ancestor-

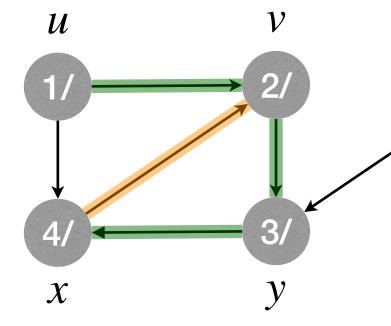


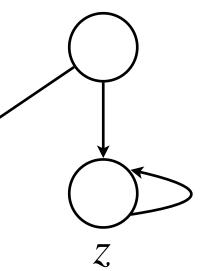




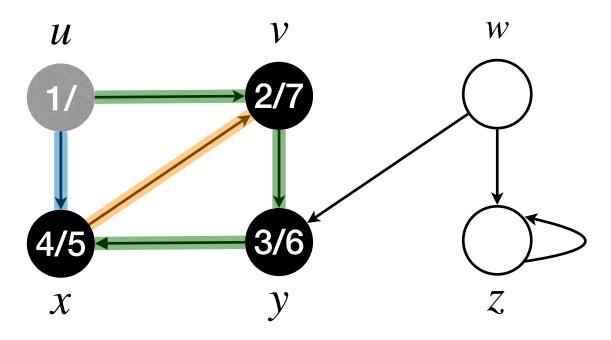


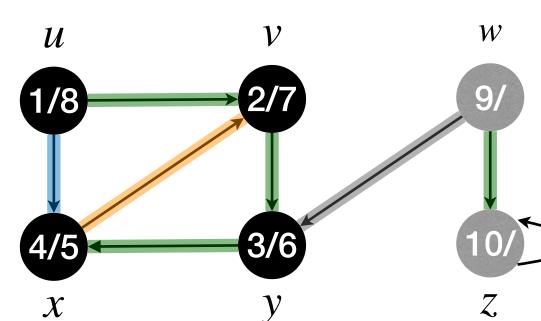
W

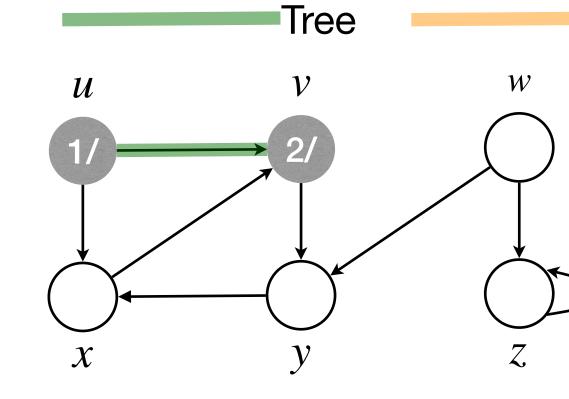


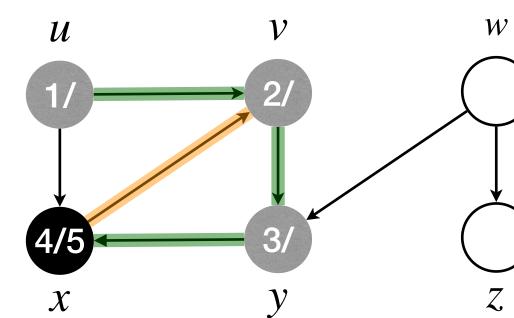


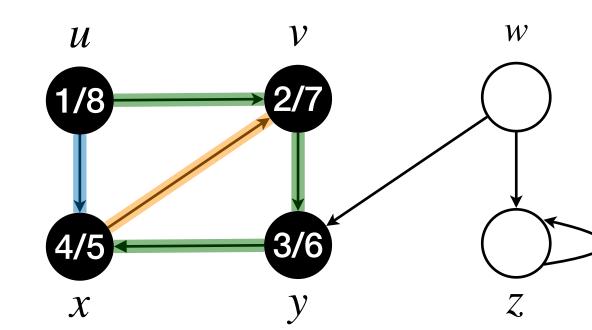
 ${\mathcal W}$

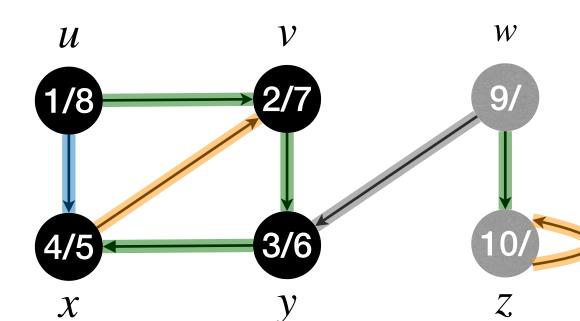


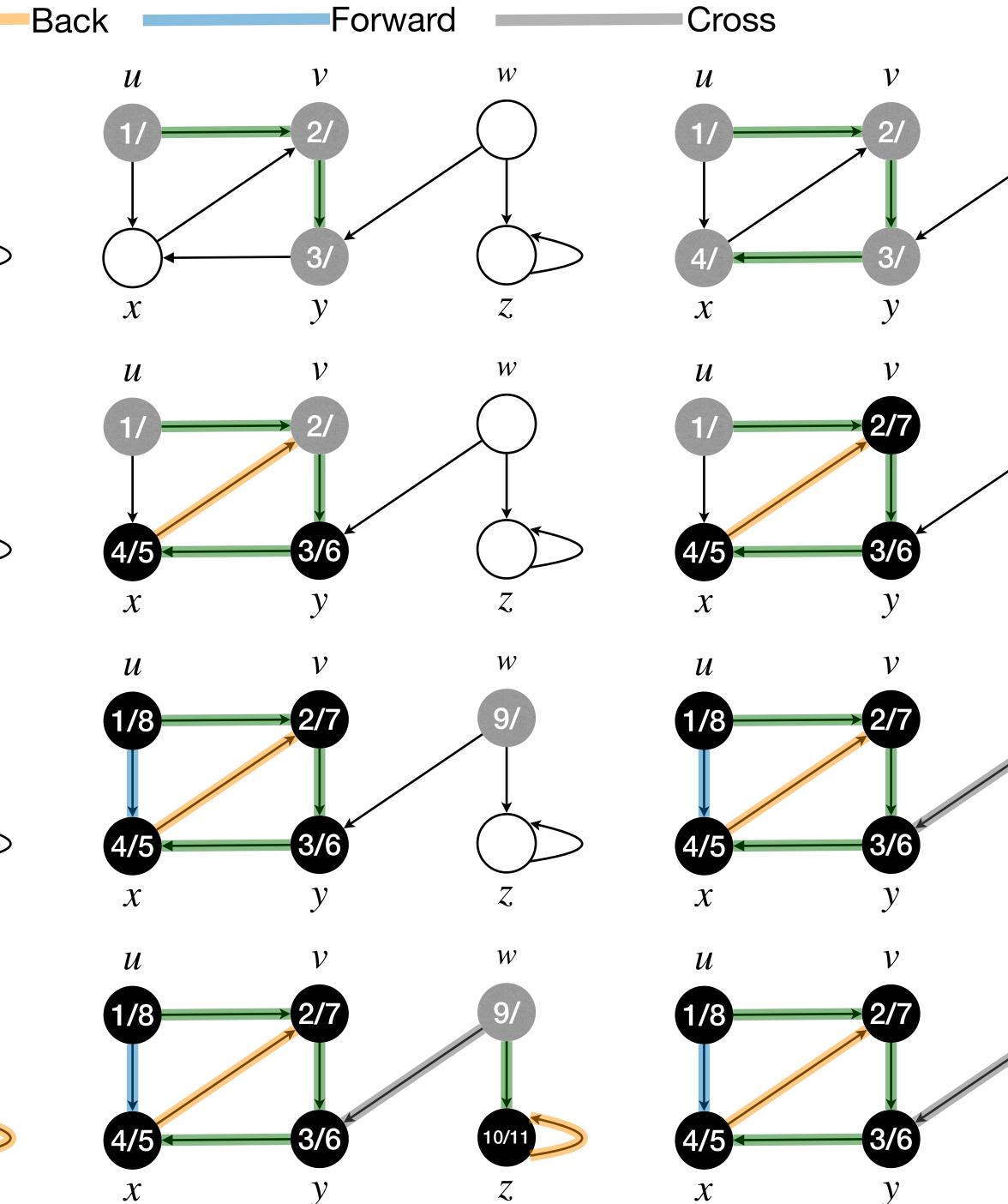


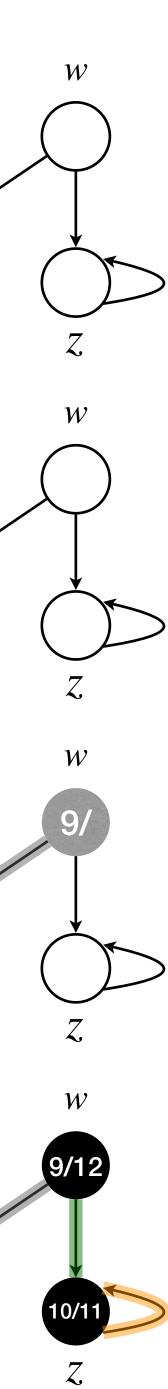










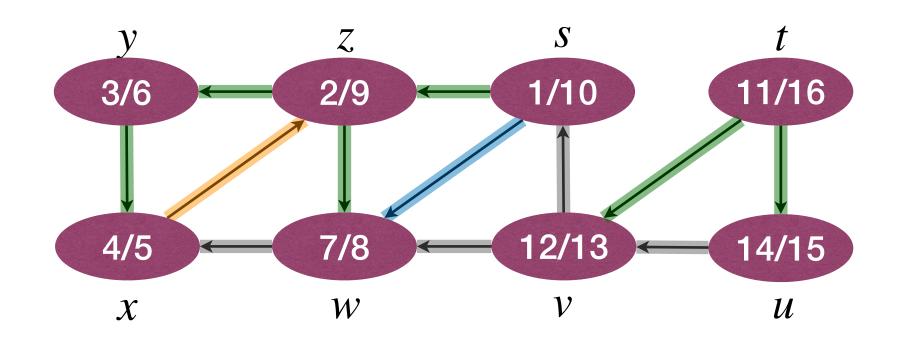




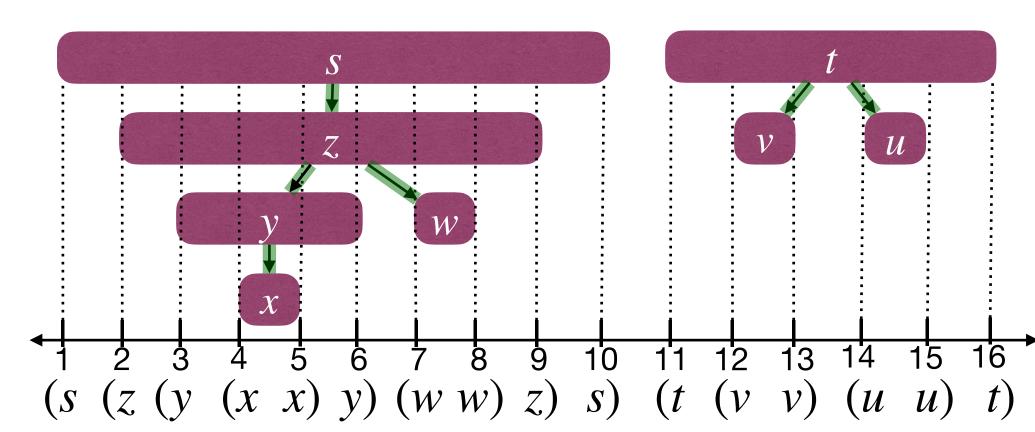
Properties of DFS: Parenthesis Theorem

one is entirely contained within another.

- For any two nodes *u* and *v*, <u>exactly</u> one of following holds:
 - (a) [u.d, u.f] and [v.d, v.f] are disjoint, and u, v have no ancestor-descendant relation in the DFS forest;
 - (b) $[u.d, u.f] \subset [v.d, v.f]$, and u is a descendant of v in a DFS tree;
 - (c) $[v.d, v.f] \subset [u.d, u.f]$, and u is an ancestor of v in a DFS tree.



Theorem: Active intervals of two nodes are either: (a) entirely disjoint; or (b)





Properties of DFS: Parenthesis Theorem

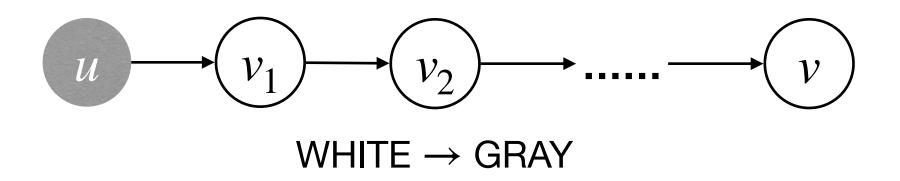
- **Proof**: Consider two nodes u and v. W.I.o.g., assume u.d < v.d.
- If v.d < u.f, then v is discovered (WHITE \rightarrow GRAY) while u is being processed (GRAY); and DFS will finish v first, before returning to u.
 - In this case, $[v.d, v.f] \subset [u.d, u.f]$, and u is an ancestor of v.
- If v.d > u.f, then obviously u.d < u.f < v.d < v.f; and DFS has finished exploring u (BLACK), before v is discovered (WHITE \rightarrow GRAY).
 - In this case, [u.d, u.f] and [v.d, v.f] are disjoint, and u, v have no ancestordescendant relation.



Properties of DFS: White-path Theorem

Theorem In the DFS forest, v is a descendant of u iff when u is discovered, there is a path in the graph from u to v containing only WHITE nodes.

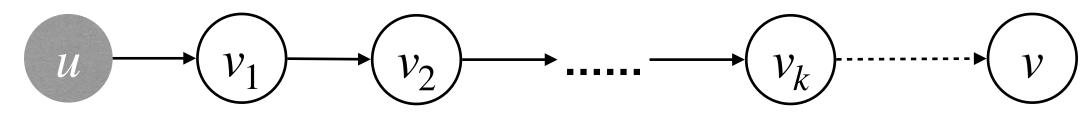
- Proof of \implies
 - Claim: If v is a proper descendant of u, then v is WHITE when u is discovered.
 - Since if v is a is a proper descendant of u, then u.d < v.d.
 - For any node along the path from u to v in the DFS forest, above claim holds.
 - Therefore, $[\Longrightarrow]$ direction of the theorem holds.





Properties of DFS: White-path Theorem

- Proof of $[\Leftarrow]$:
 - of \mathcal{U} .
 - So we have $[v_k.d, v_k.f] \subset [u.d, u.f]$.
 - But v is discovered after u is discovered, and must before v_k is finished.
 - So we have $u.d < v.d < v_k.f \le u.f$.
 - Then it must be $[v.d, v.f] \subset [u.d, u.f]$, implying v is a descendant of u.



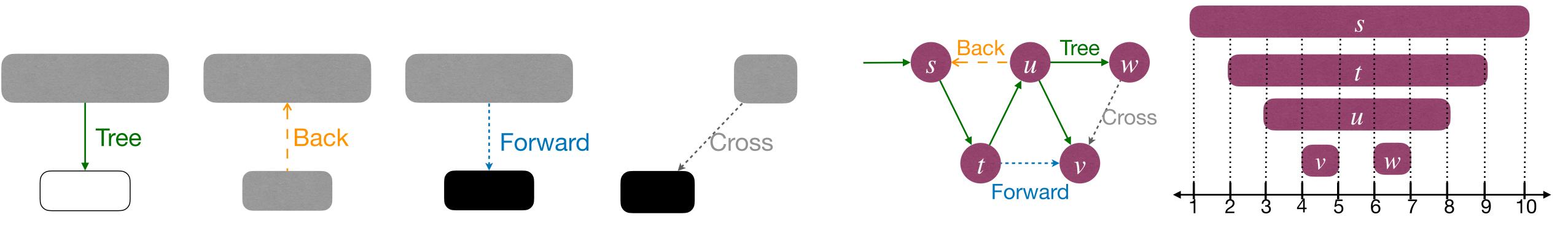
W.I.o.g., assume v is the first node along the path that does not become a descendant

Depth-first search until all the edges of v_k is explored!



Properties of DFS: Classification of edges

- Determine (u, v) type by color of v during DFS execution.
 - Tree Edges: Edges in the DFS forest.
 - Back Edges: Edges (u, v) connecting u to an ancestor v in a DFS tree.
 - Forward Edges: Non-tree edges (u, v) connecting u to a descendant v in a DFS tree.
 - Cross Edges: Other edges. (Connecting nodes in same DFS tree with no ancestor-descendant relation, or connecting nodes in different DFS trees.)



Node *v* is WHITE

Node *v* is BLACK

Node *v* is BLACK









Properties of DFS: Types of edges in undirected graphs

• Will all four types of edges appear in DFS of undirected graphs?

Theorem In DFS of an *undirected* graph G, every edge of G is either a tree edge or a back edge.

- **Proof**:
 - Consider an arbitrary edge (u, v). W.I.o.g., assume u.d < v.d.
 - Edge (u, v) must be explored while u is GRAY.
 - Consider the first time the edge (u, v) is explored.

WHY?





Properties of DFS: Types of edges in undirected graphs

- Proof (continued):
 - If the direction is $u \to v$. Then, v must be WHITE by then, for otherwise the edge would have been explored from direction $v \rightarrow u$ earlier.
 - In such case, the edge (u, v) becomes a tree edge.
 - If the direction is $v \rightarrow u$. Then, the edge is "GRAY \rightarrow GRAY".
 - In such case, the edge (u, v) becomes a **back edge**.





DFSIterSkeleton(G, s):

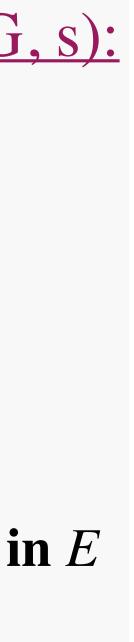
Stack Q Q.push(s)while !Q.empty() u := Q.pop()if !u.visited *u.visited* := *True* for each edge (u, v) in E Q.push(v)

BFSSkeletonAlt(G, s): FIFOQueue Q Q.enque(s)while !*Q.empty(*) u := Q.dequeue()if !u.visited *u.visited* := *True* for each edge (u, v) in E Q.enque(v)

Other queuing disciplines lead to more interesting algorithms!

DFS, BFS, and others...

GraphExploreSkeleton(G, s): GenericQueue Q Q.add(s)while !*Q.empty*() u := Q.remove()if !u.visited *u.visited* := *True* for each edge (u, v) in E Q.add(v)





Further reading

• [CLRS] Ch.22 (22.1-22.3)

