CSE 421 Winter 2025
Lecture 5: Graph Search and Greedy

Nathan Brunelle

http://www.cs.uw.edu/421

Graph Traversal

Learn the basic structure of a graph

Walk from a fixed starting vertex s to find all vertices reachable from s

Three states of vertices

- **unvisited**
- **visited/discovered** (in R, i.e. reachable)
- **fully-explored** (in R and all neighbors have been visited)

BFS(s)

Global initialization: mark all vertices "unvisited"

BFS(s)

Mark s "visited"

Add s to Q

i = 0

Mark s as "layer i"

while Q not empty

u = next item removed from Q

i = "layer of u"

for each edge (u, v)

if (v is "unvisited")

add v to Q

mark v "visited"

mark v as "layer $i + 1$"

mark u "fully-explored"

Properties of BFS

BFS(s) visits x iff there is a path in G from s to x.

Edges followed to undiscovered vertices define a breadth first spanning tree of G

Layer i in this tree:

L_i = set of vertices u with shortest path in G from root s of length i.

Properties of BFS

Claim: For undirected graphs:

All edges join vertices on the same or adjacent layers of BFS tree

Proof: Suppose not...

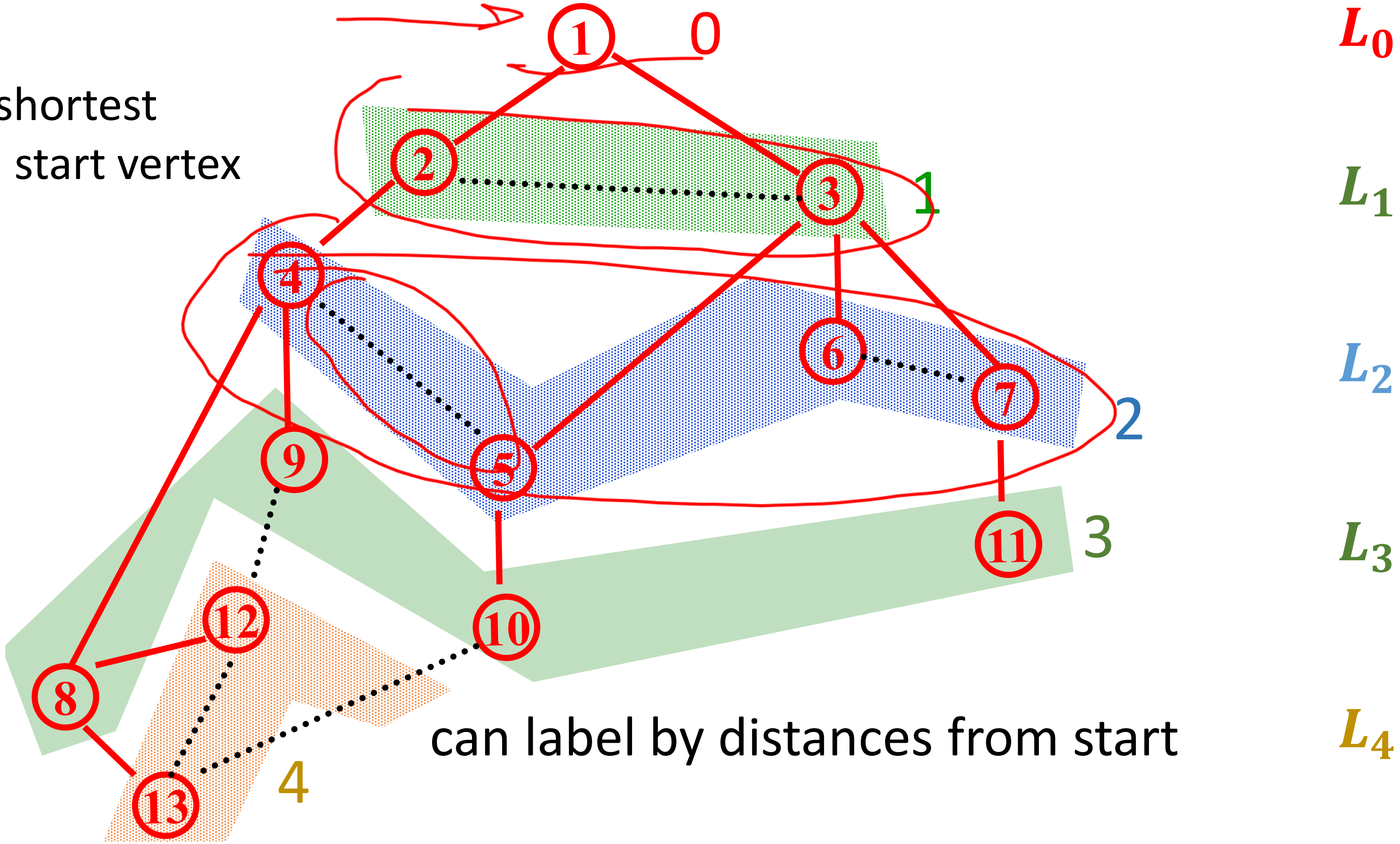
Then there would be vertices (x, y) s.t. $x \in L_i$ and $y \in L_j$ and $j > i + 1$.

Then, when vertices adjacent to x are considered in BFS,
y would be added with layer $i + 1$ and not layer j.

Contradiction.

BFS Application: Shortest Paths

Tree gives shortest paths from start vertex



can label by distances from start

Applications of Graph Traversal: Bipartiteness Testing

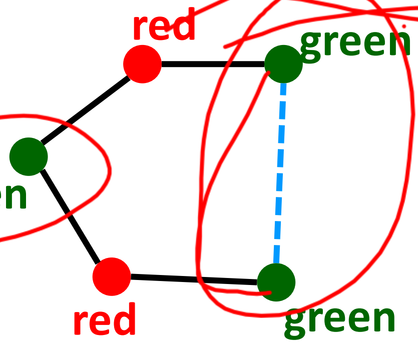
Definition: An undirected graph G is bipartite iff we can color its vertices **red** and **green** so each edge has different color endpoints

Input: Undirected graph G

Goal: If G is bipartite, output a coloring;
otherwise, output "NOT Bipartite".

Fact: Graph G contains an odd-length cycle ⇒ it is not bipartite

Just coloring the cycle part of G is impossible



On a cycle the two colors must alternate, so

- **green** every 2nd vertex
- **red** every 2nd vertex

Can't have either if length is not divisible by 2.

Applications of Graph Traversal: Bipartiteness Testing

WLOG ("without loss of generality"): Can assume that G is connected

- Otherwise run on each component

Simple idea: start coloring nodes starting at a given node s

- Color s **red**
- Color all neighbors of s **green**
- Color all their neighbors **red**, etc.
- If you ever hit a node that was already colored
  - the **same** color as you want to color it, ignore it
  - the **opposite** color, output "NOT Bipartite" and halt

BFS gives Bipartiteness

Run BFS assigning all vertices from layer $\boldsymbol{L_i}$ the color $\boldsymbol{i} \bmod \mathbf{2}$

- i.e., **red** if they are in an even layer, **green** if in an odd layer

- if no edge joining two vertices of the same color
  - then it is a good coloring
- otherwise
  - there is a bad edge; output "Not Bipartite"

Why is that "Not Bipartite" output correct?

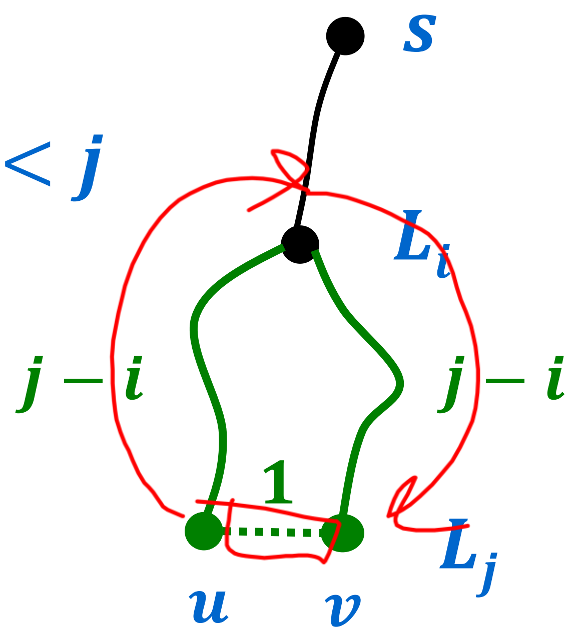
Why does BFS work for Bipartiteness?

Recall: All edges join vertices on the same or adjacent BFS layers

⇒ Any "bad" edge must join two vertices u and v in the same layer

Say the layer with u and v is L_j

u and v have common ancestor at some level L_i for $i < j$



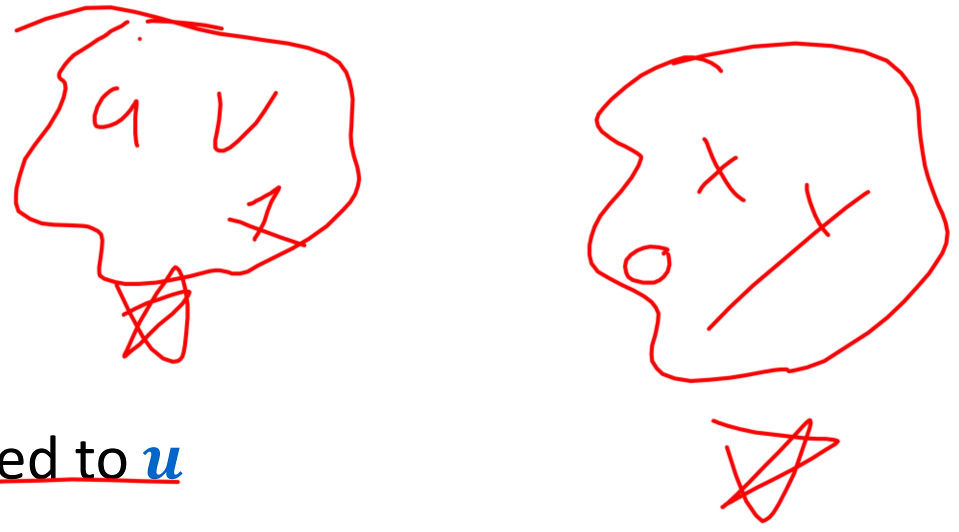
Odd cycle of length $2(j - i) + 1$
⇒ Not Bipartite

Undirected Graph Search Application: Connected Components

Want to answer questions of the form:

Given: vertices u and v in G

Is there a path from u to v?



Idea: create array A s.t

$A[u]$ = smallest numbered vertex connected to u

Answer is yes iff $A[u] = A[v]$

Q: Why is this better than an array $\mathbf{Path}[u, v]$?

Undirected Graph Search Application: Connected Components

Initial state: all $\boldsymbol{v}$ unvisited

for $\boldsymbol{s}$ from 1 to $\boldsymbol{n}$ do:

if state($\boldsymbol{s}$) $\neq$ fully-explored then

BFS($\boldsymbol{s}$): setting $\boldsymbol{A}[\boldsymbol{u}] = \boldsymbol{s}$ for each $\boldsymbol{u}$ found

(and marking $\boldsymbol{u}$ visited/fully-explored)

Total cost: $O(\boldsymbol{n} + \boldsymbol{m})$

- Each vertex is touched once in outer procedure and edges examined in different BFS runs are disjoint
- Works also with Depth First Search ...

DFS($\boldsymbol{u}$) – Recursive Procedure

Global Initialization: mark all vertices "unvisited"

DFS($\boldsymbol{u}$)

mark $\boldsymbol{u}$ “visited” and add $\boldsymbol{u}$ to $\boldsymbol{R}$

for each edge $(\boldsymbol{u}, \boldsymbol{v})$

if ($\boldsymbol{v}$ is “unvisited”)

DFS($\boldsymbol{v}$)

mark $\boldsymbol{u}$ “fully-explored”

Properties of DFS($\boldsymbol{s}$)

Like BFS($\boldsymbol{s}$):

- DFS($\boldsymbol{s}$) visits $\boldsymbol{x}$ iff there is a path in $\boldsymbol{G}$ from $\boldsymbol{s}$ to $\boldsymbol{x}$
- Edges into undiscovered vertices define depth-first spanning tree of $\boldsymbol{G}$

Unlike the BFS tree:

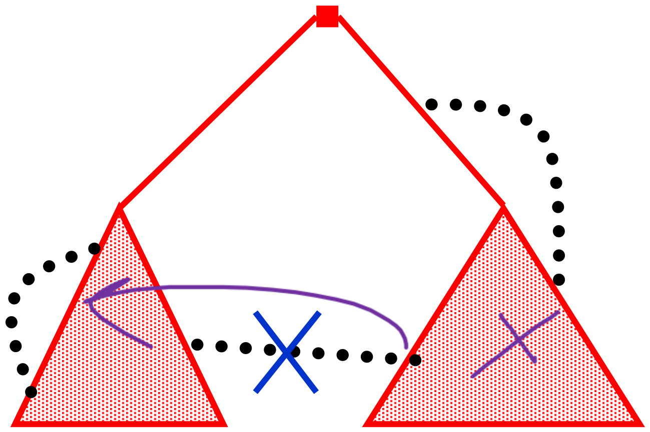
- the DFS spanning tree *isn't* minimum depth
- its levels *don't* reflect min distance from the root
- non-tree edges *never* join vertices on the same or adjacent levels

BUT…

Non-tree edges in DFS tree of **undirected** graphs

Claim: All non-tree edges join a vertex and one of its descendents/ancestors in the DFS tree

- In other words ... No "cross edges".



No cross edges in DFS on undirected graphs

Claim: During DFS(x) every vertex marked "visited" is a descendant of x in the DFS tree T

Claim: For every x, y in the DFS tree T, if (x, y) is an edge *not* in T then one of x or y is an ancestor of the other in T

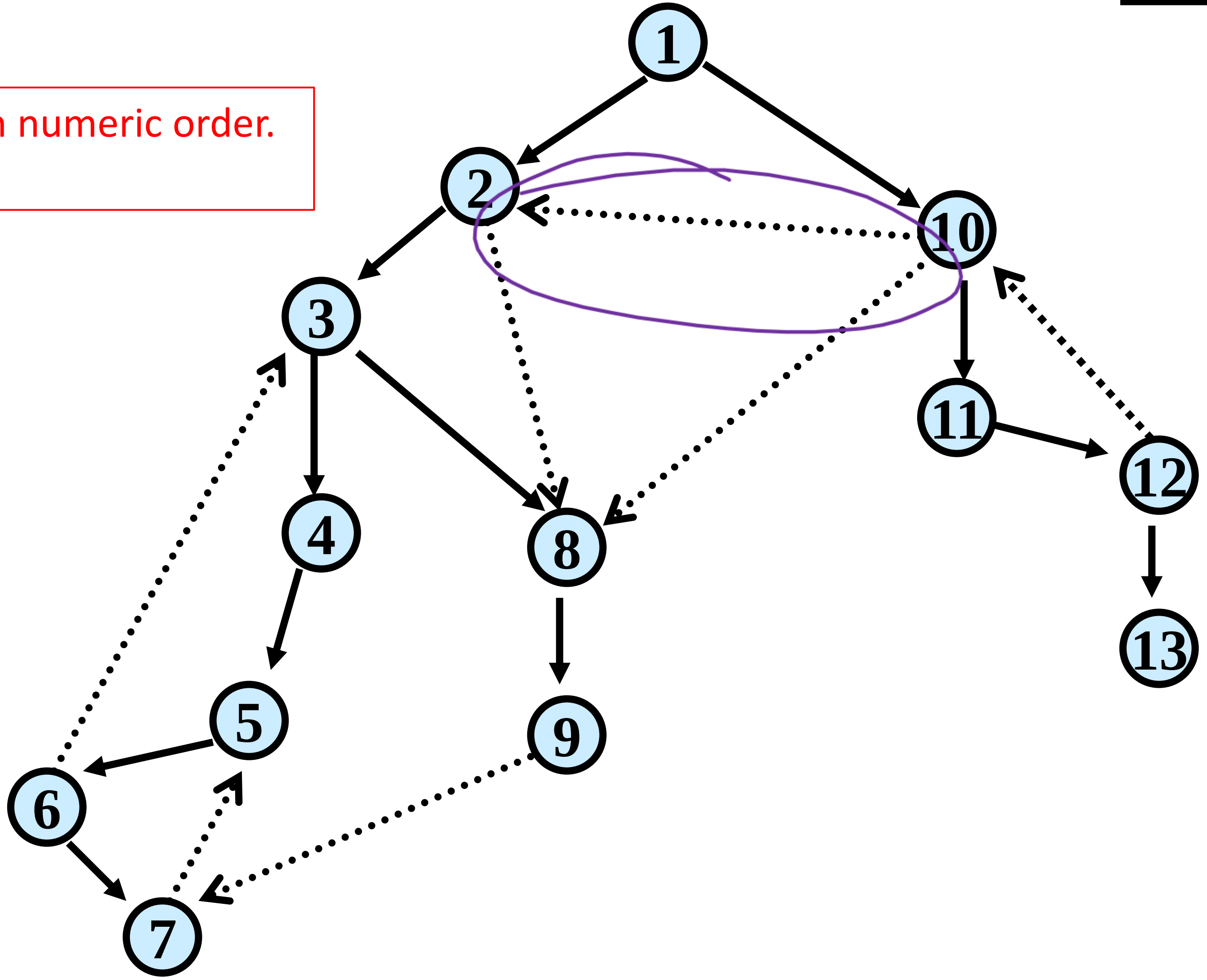
Proof:

- One of DFS(x) or DFS(y) is called first, suppose WLOG that DFS(x) was called before DFS(y)
- During DFS(x), the edge (x, y) is examined
- Since (x, y) is a *not* an edge of T, y was already visited when edge (x, y) was examined during DFS(x)
- Therefore y was visited during the call to DFS(x) so y is a descendant of x.

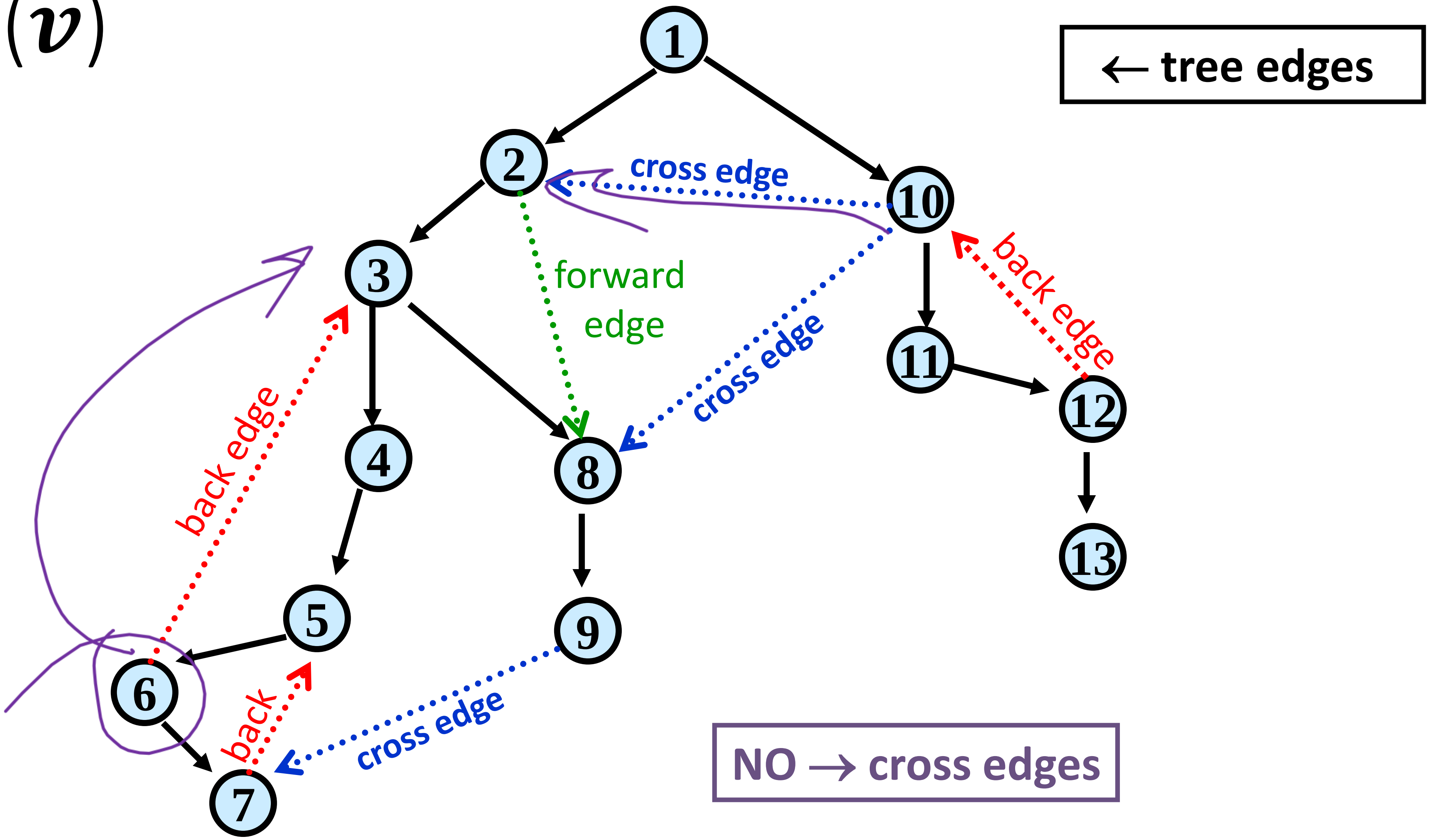
DFS($\boldsymbol{v}$) for a directed graph

← tree edges

Nodes were visited in numeric order.
How can we tell?



DFS(v)


1
2
10
3
4
8
11
12
13
5
9
6
7
cross edge
forward edge
cross edge
back edge
back edge
back
cross edge
← tree edges
NO → cross edges

Properties of Directed DFS

- Before DFS(s) returns, it visits all previously unvisited vertices reachable via directed paths from s

- Every cycle contains a back edge in the DFS tree

Directed Acyclic Graphs

A directed graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ is acyclic iff it has no directed cycles

Terminology: A directed acyclic graph is also called a DAG

After shrinking the strongly connected components of a directed graph to single vertices, the result is a DAG

Topological Sort

Given: a directed acyclic graph (DAG) $\boldsymbol{G = (V, E)}$

Output: numbering of the vertices of $\boldsymbol{G}$ with distinct numbers from $\mathbf{1}$ to $\boldsymbol{n}$ so that edges only go from lower numbered to higher numbered vertices

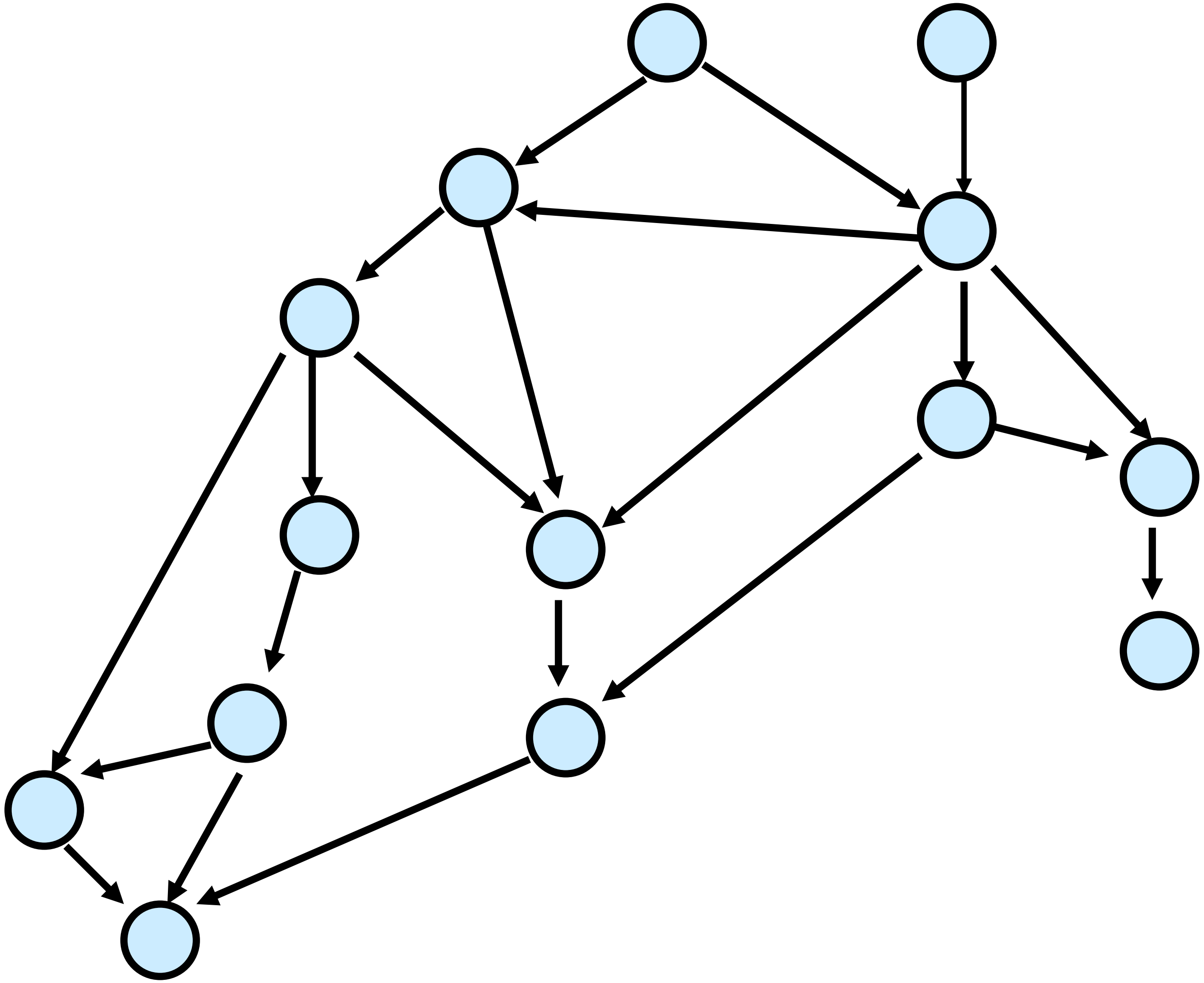
Applications:

- nodes represent tasks
- edges represent precedence between tasks
- topological sort gives a sequential schedule for solving them

Nice algorithmic paradigm for general directed graphs:

- Process strongly connected components one-by-one in the order given by topological sort of the DAG you get from shrinking them.

Directed Acyclic Graph



In-degree 0 vertices

Claim: Every DAG has a vertex of in-degree 0

Proof: By contradiction

Suppose every vertex has some incoming edge

Consider following procedure:

while (true) do

$\boldsymbol{v}$ = some predecessor of $\boldsymbol{v}$

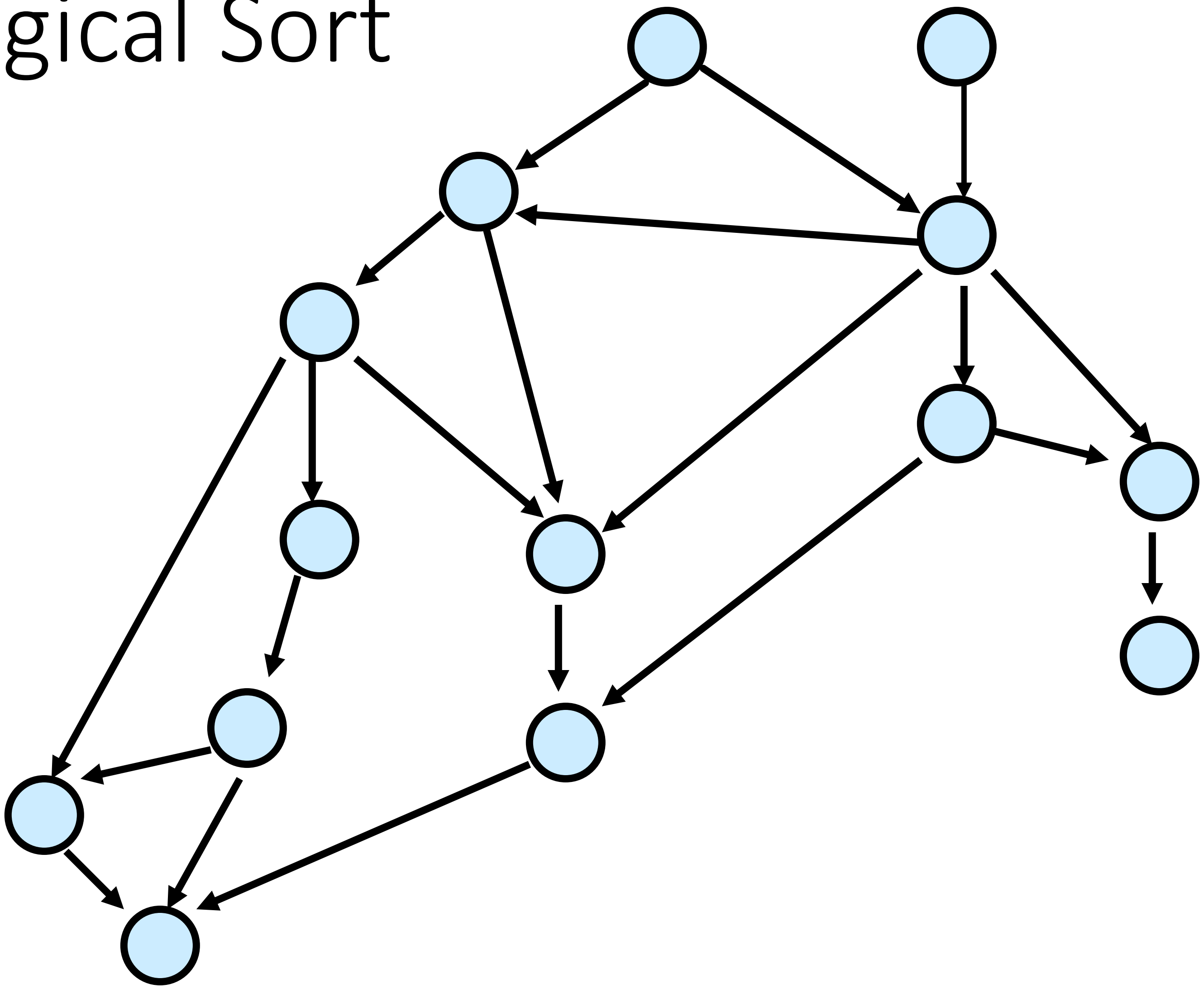
- After $\boldsymbol{n+1}$ steps where $\boldsymbol{n = |V|}$ there will be a repeated vertex
  - This yields a cycle, contradicting that it is a DAG.

Topological Sort

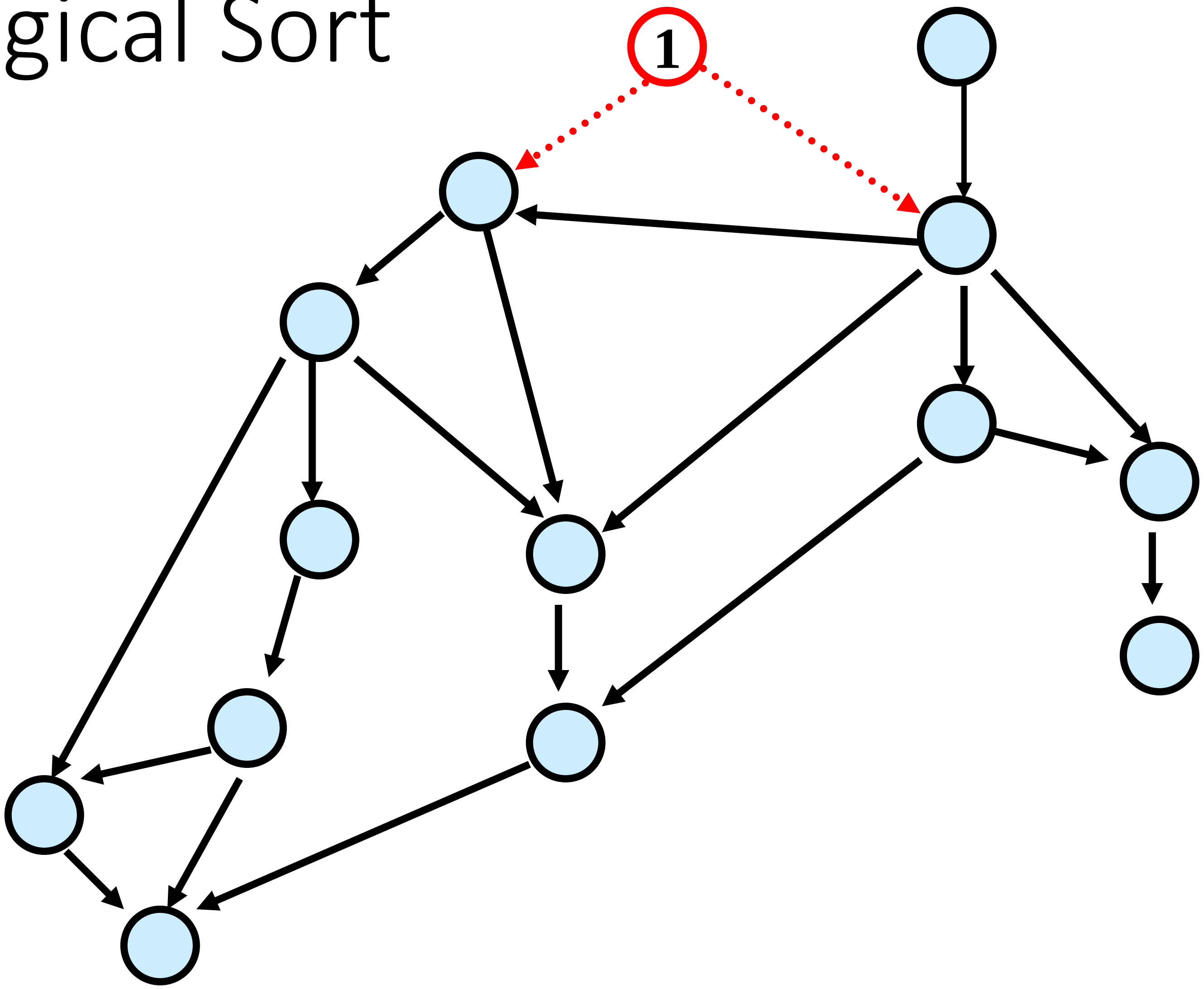
- Can do using DFS

- Alternative simpler idea:
  - Any vertex of in-degree 0 can be given number 1 to start
  - Remove it from the graph
  - Then give a vertex of in-degree 0 number 2
  - Etc.

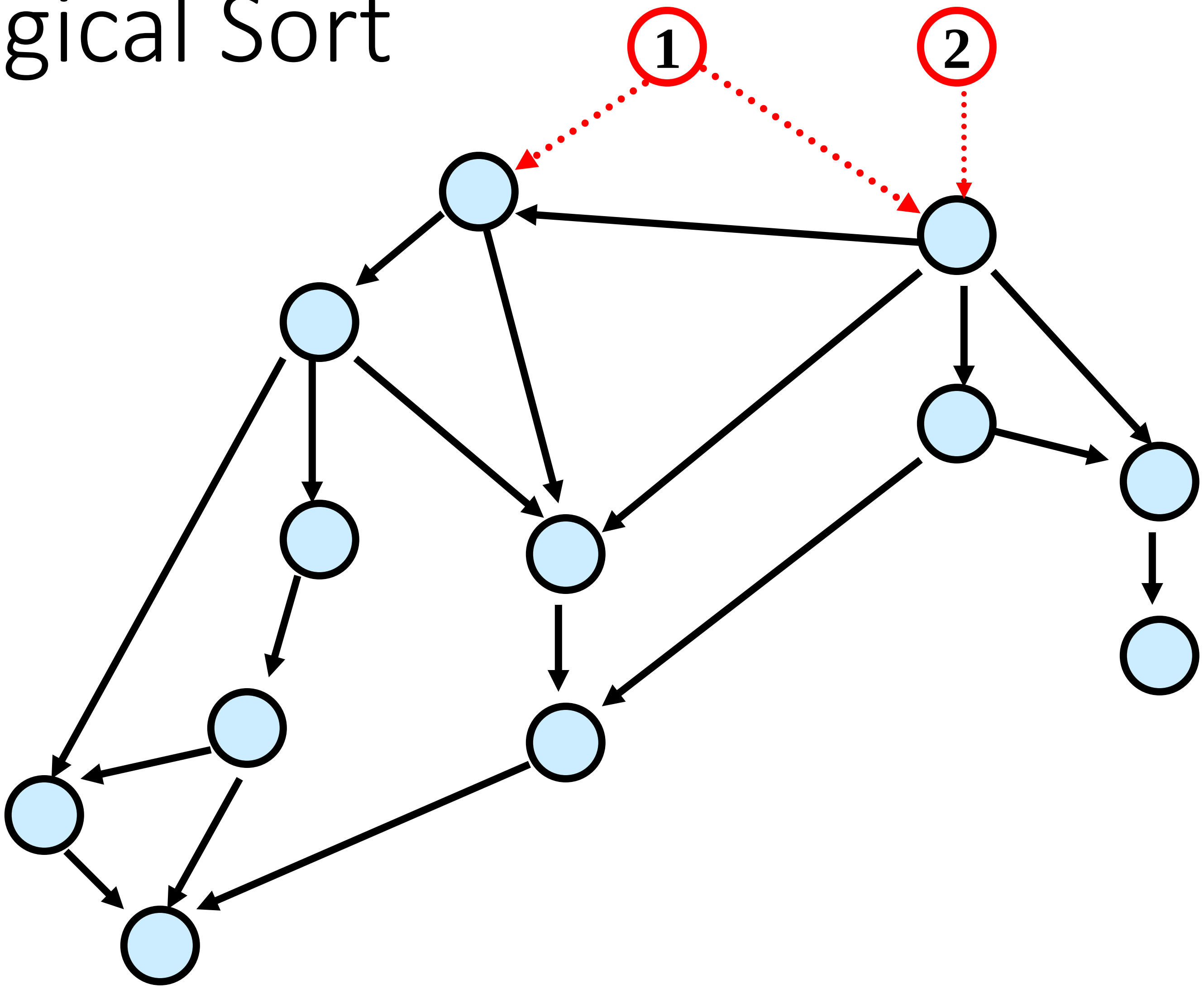
Topological Sort



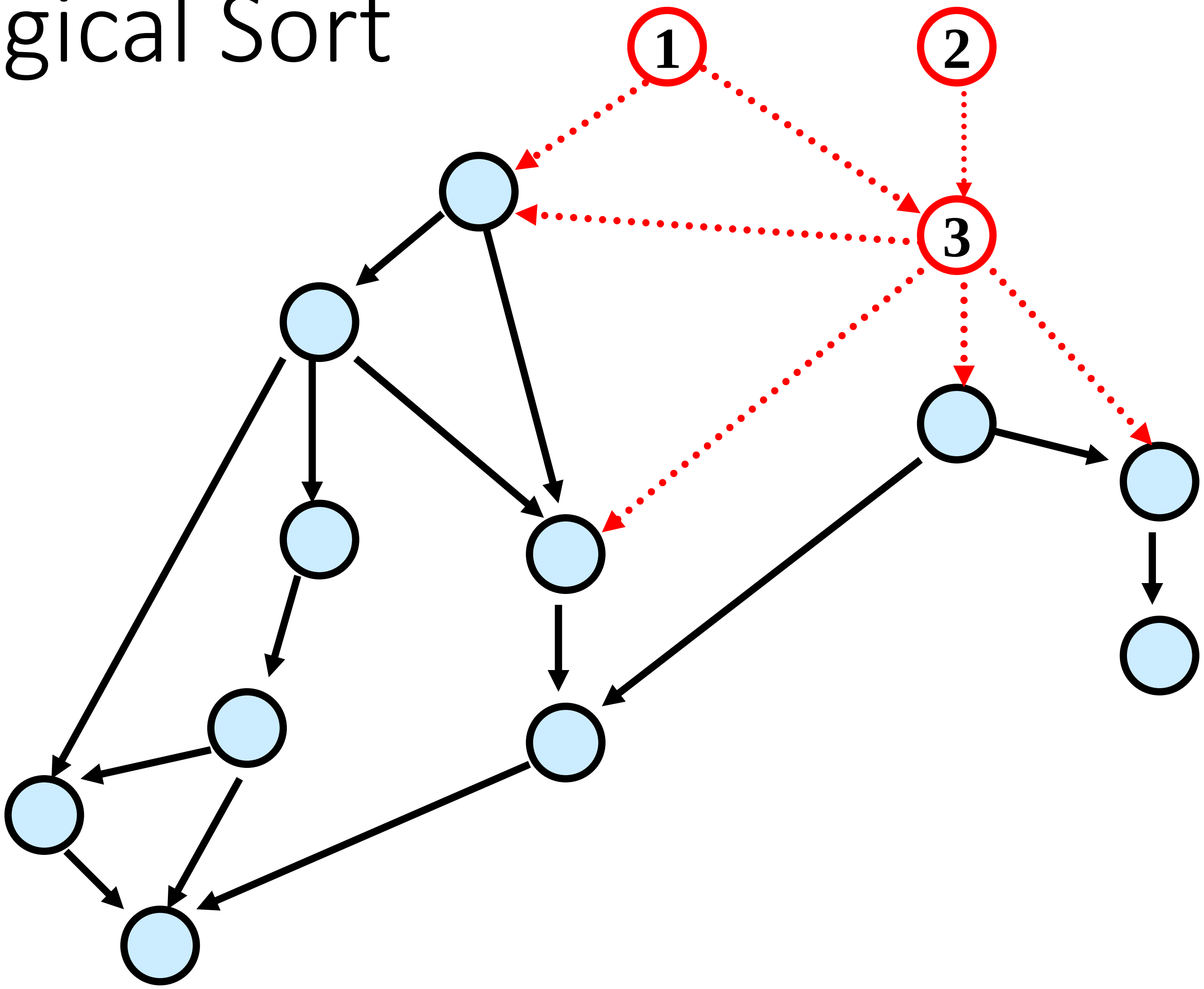
Topological Sort



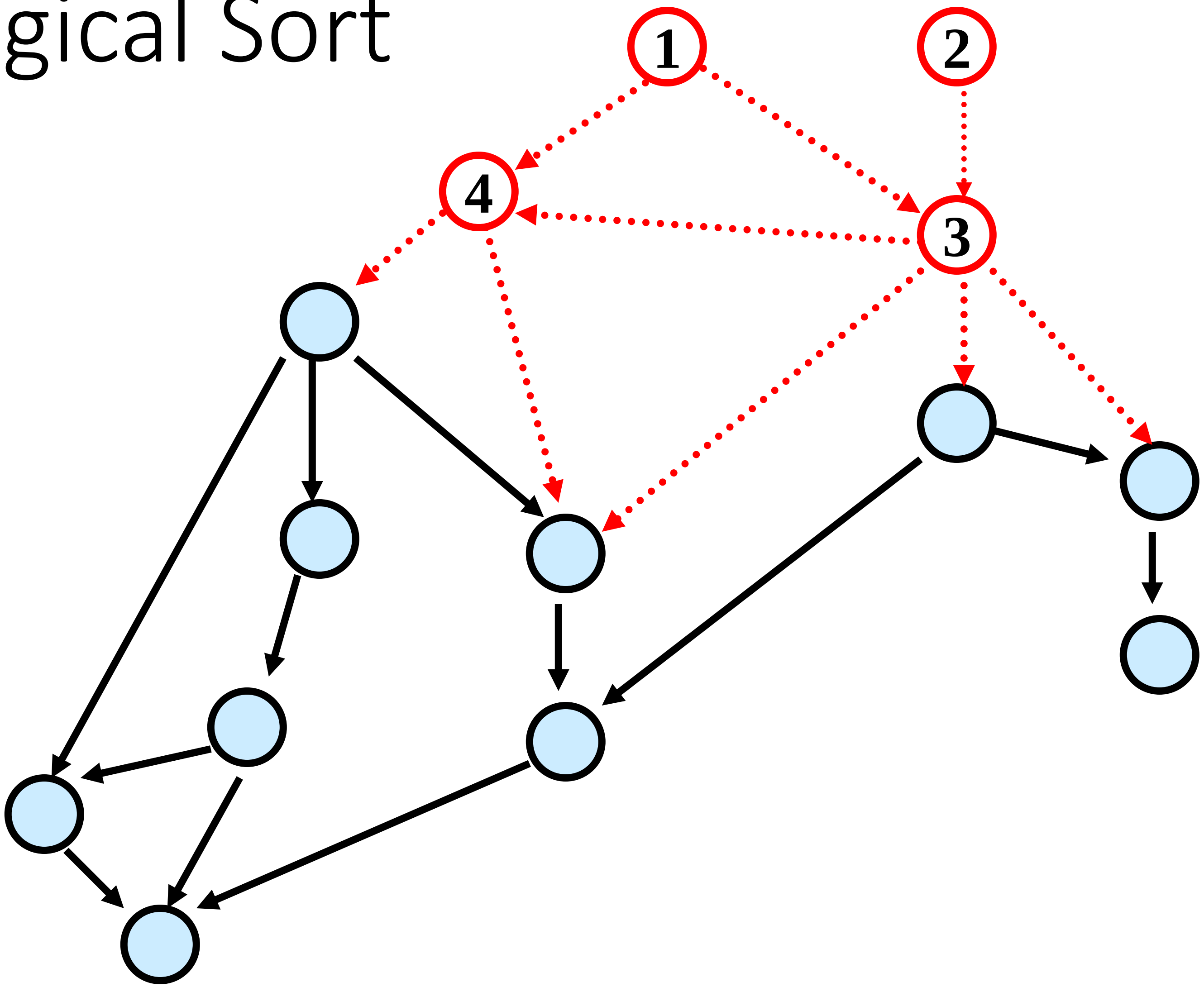
Topological Sort



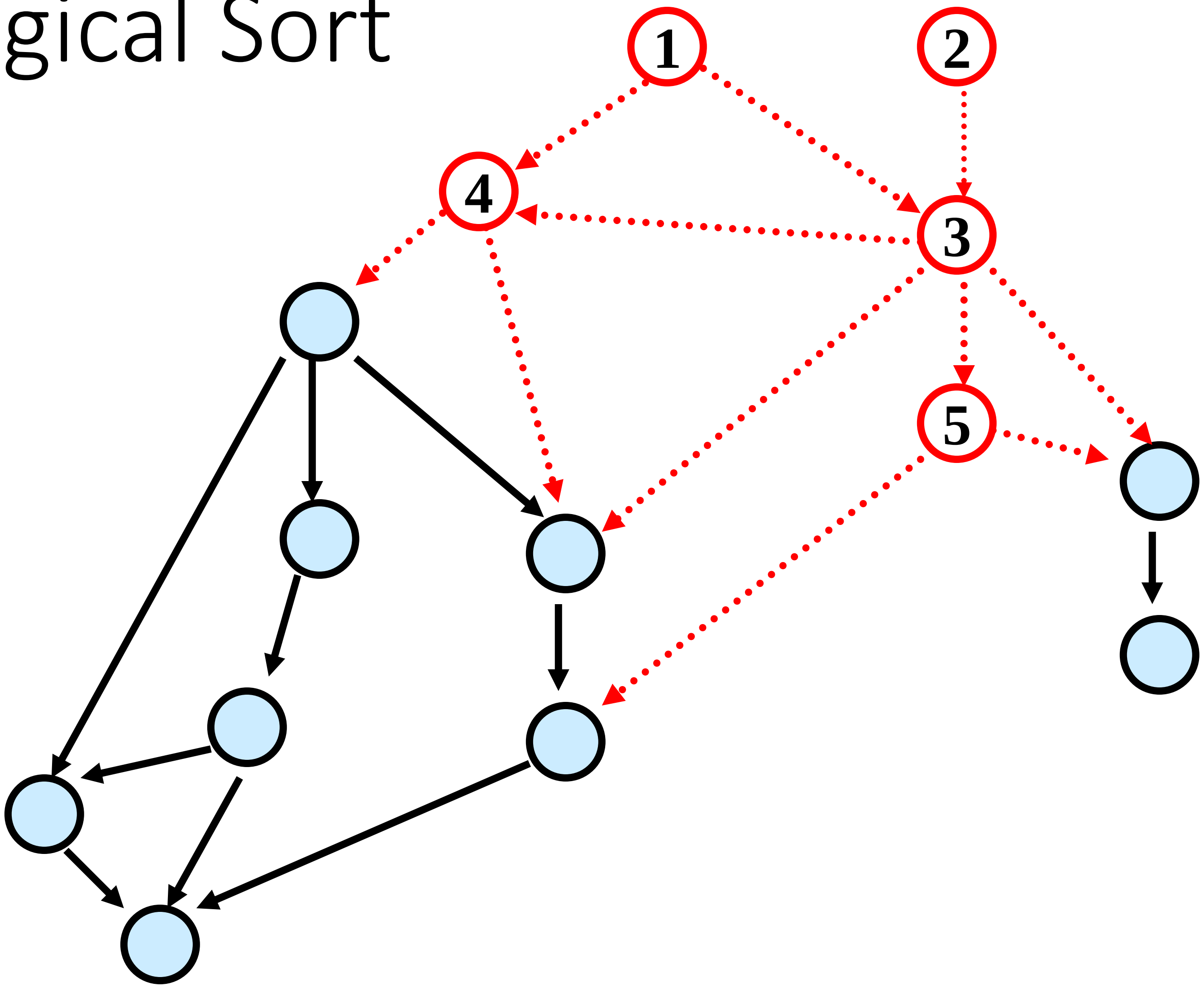
Topological Sort



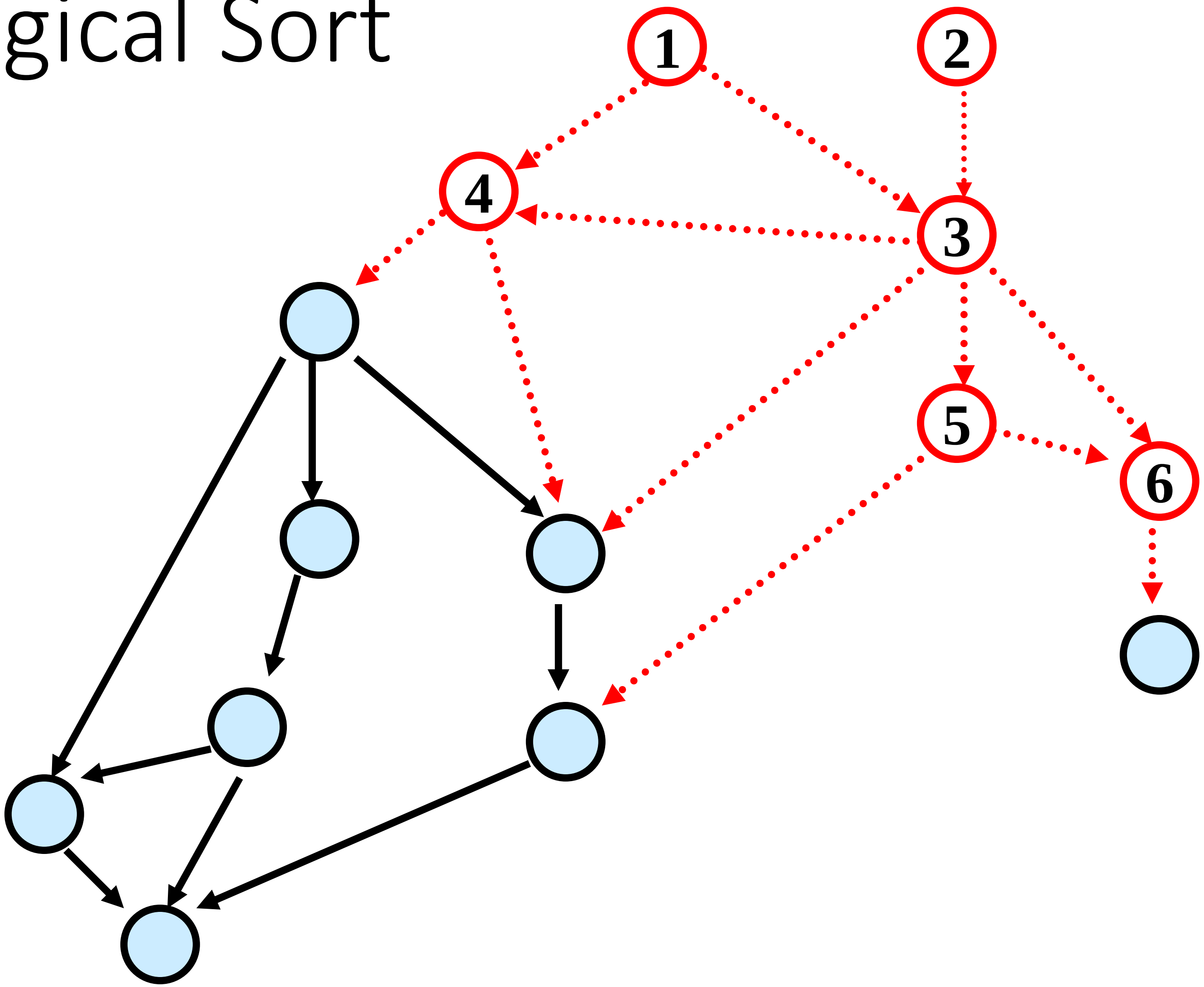
Topological Sort



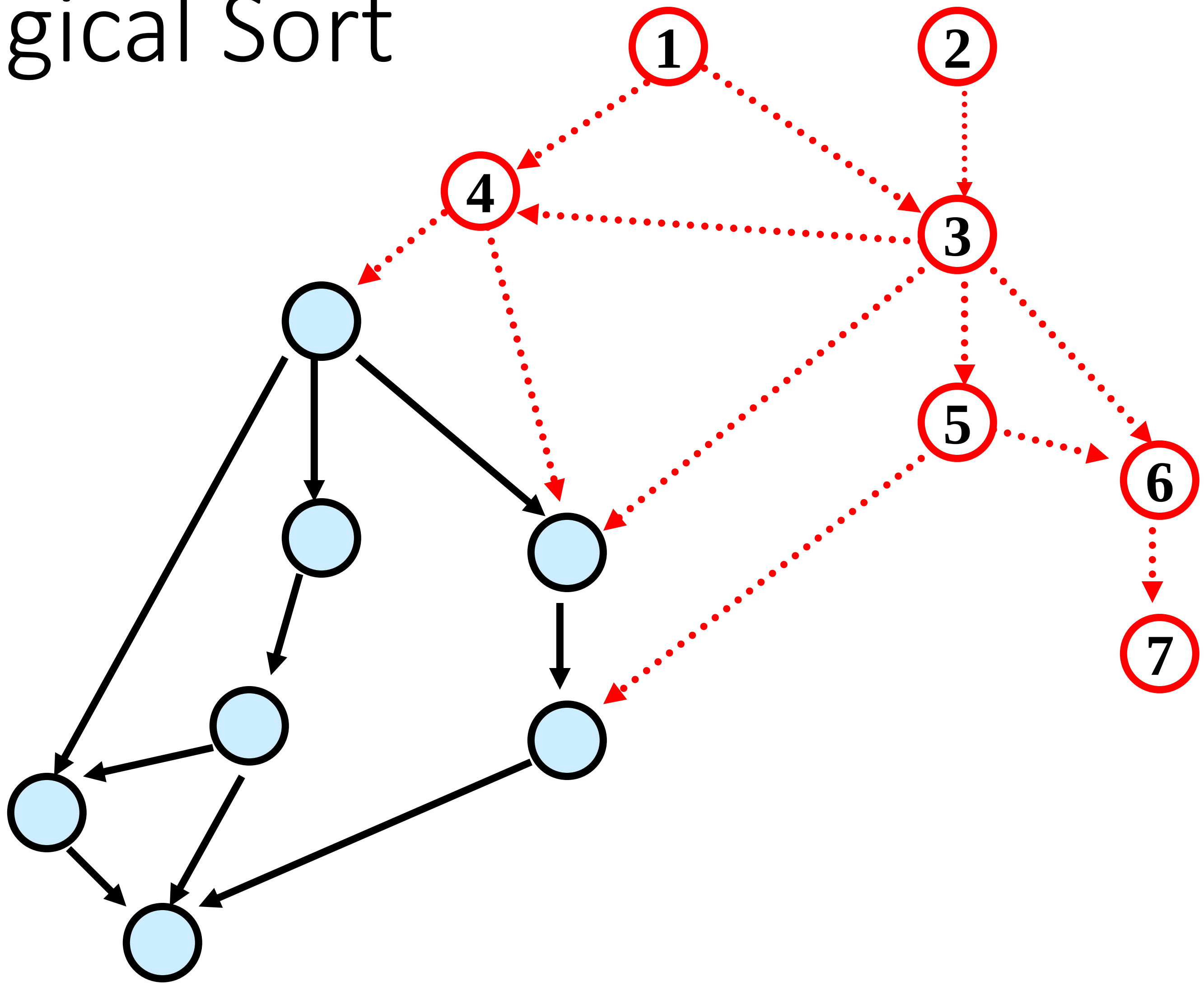
Topological Sort



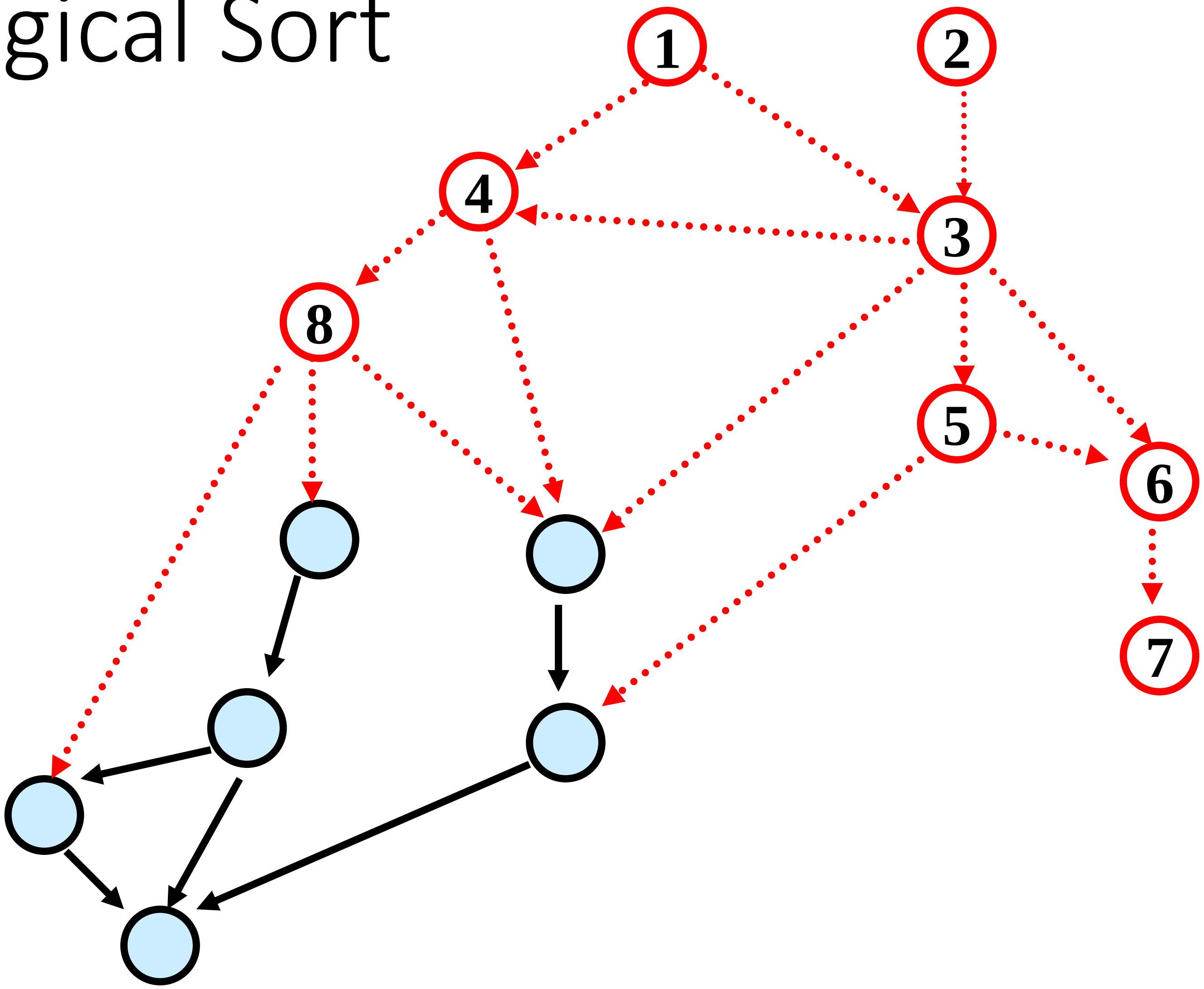
Topological Sort



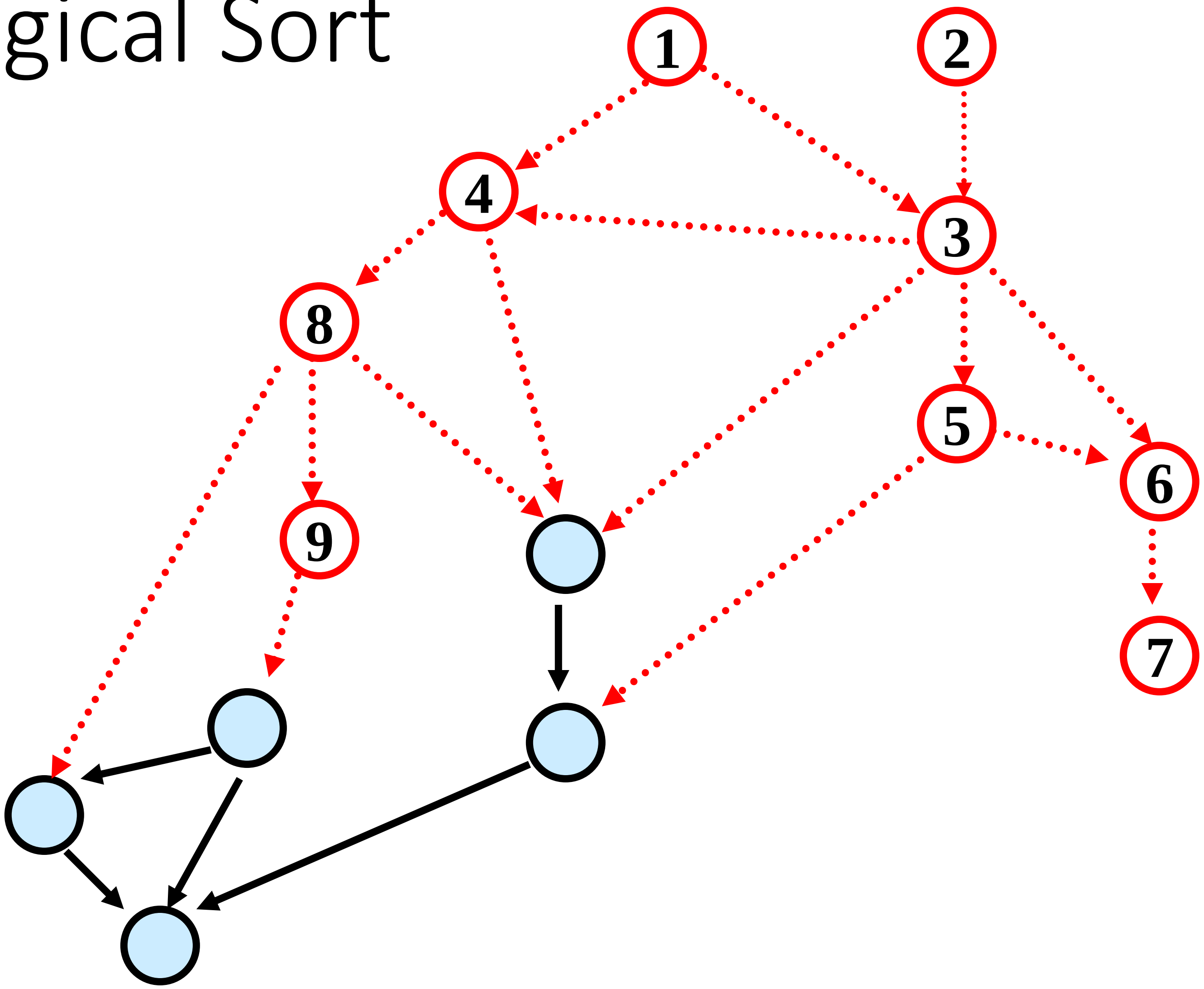
Topological Sort



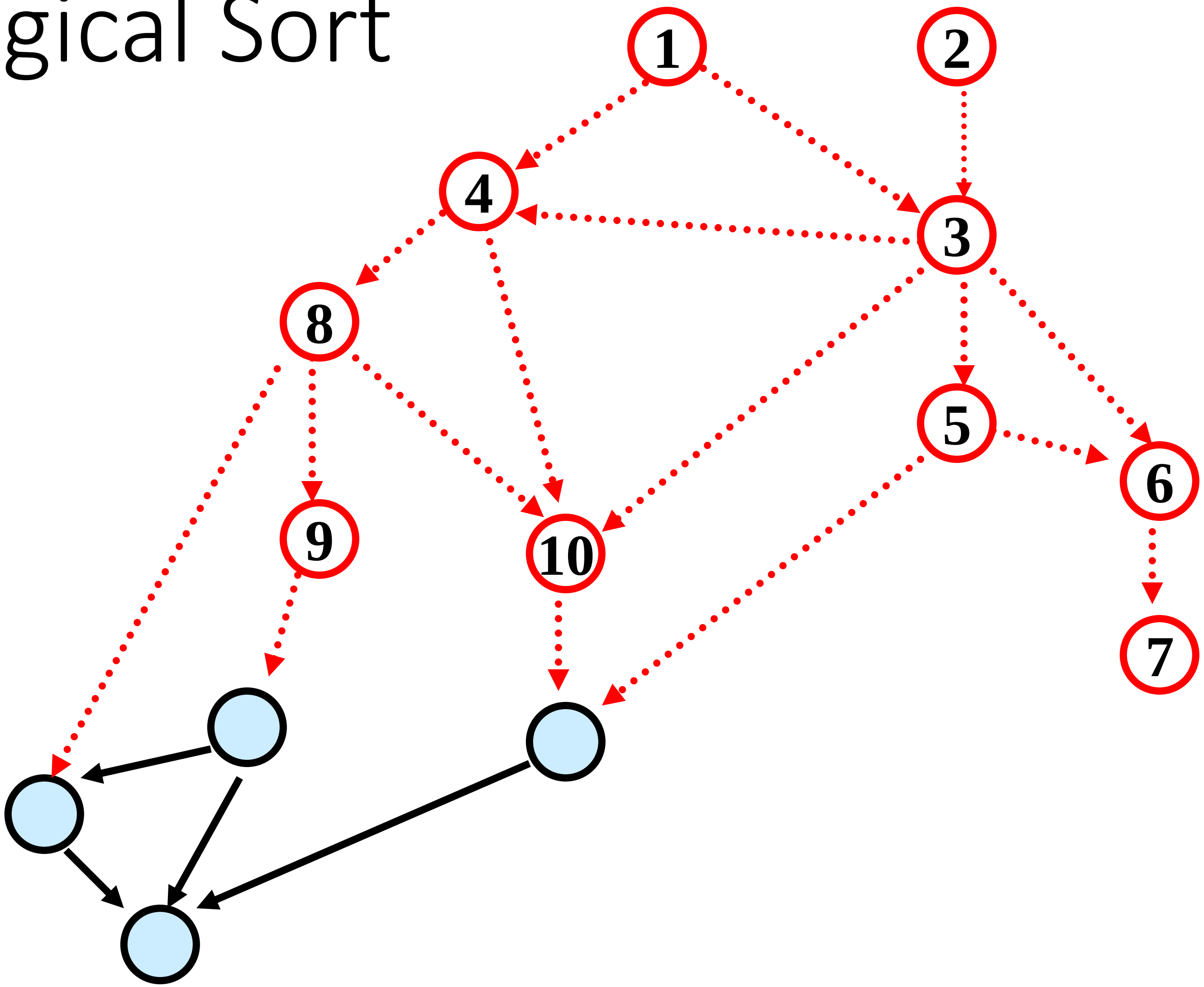
Topological Sort



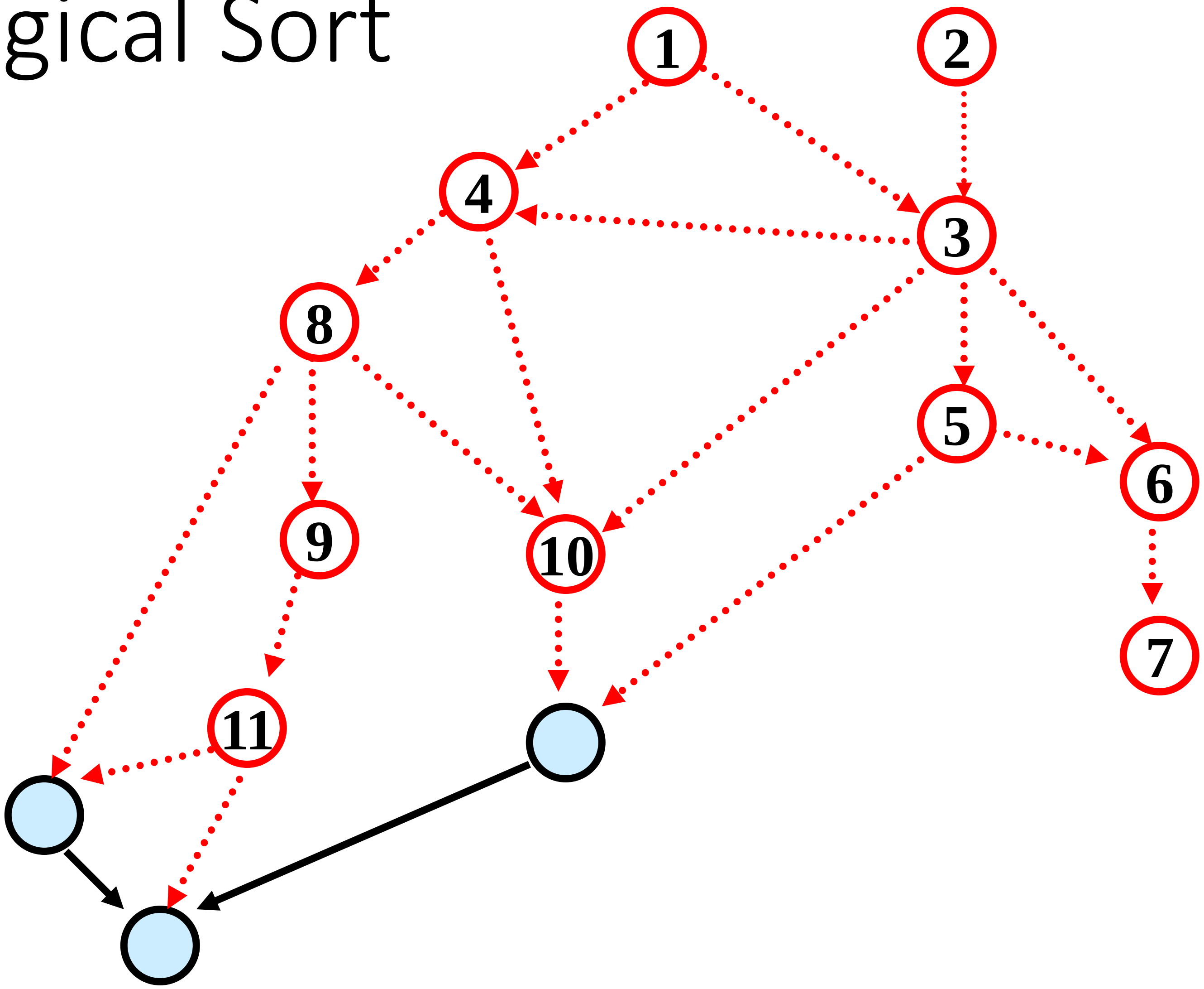
Topological Sort



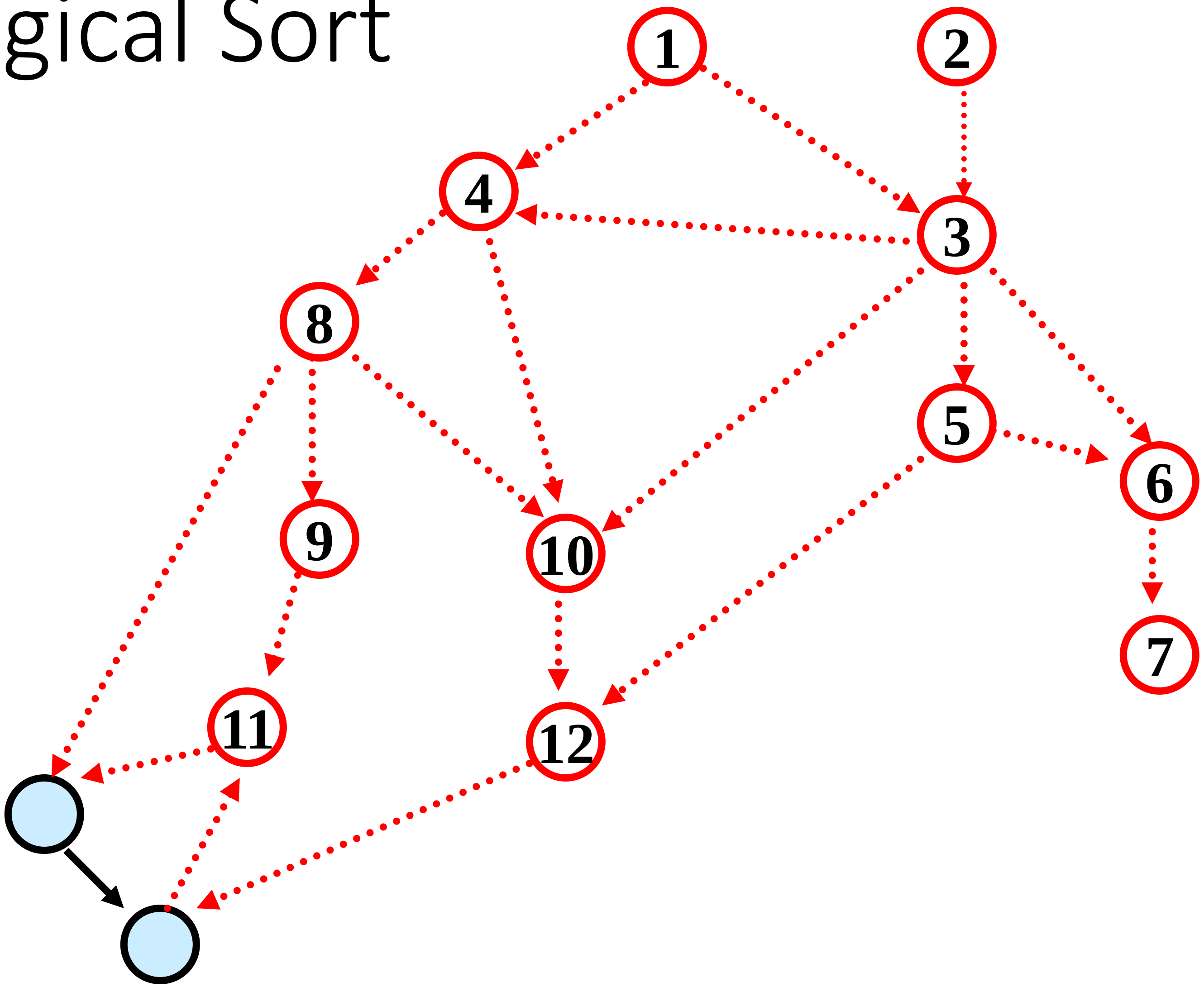
Topological Sort



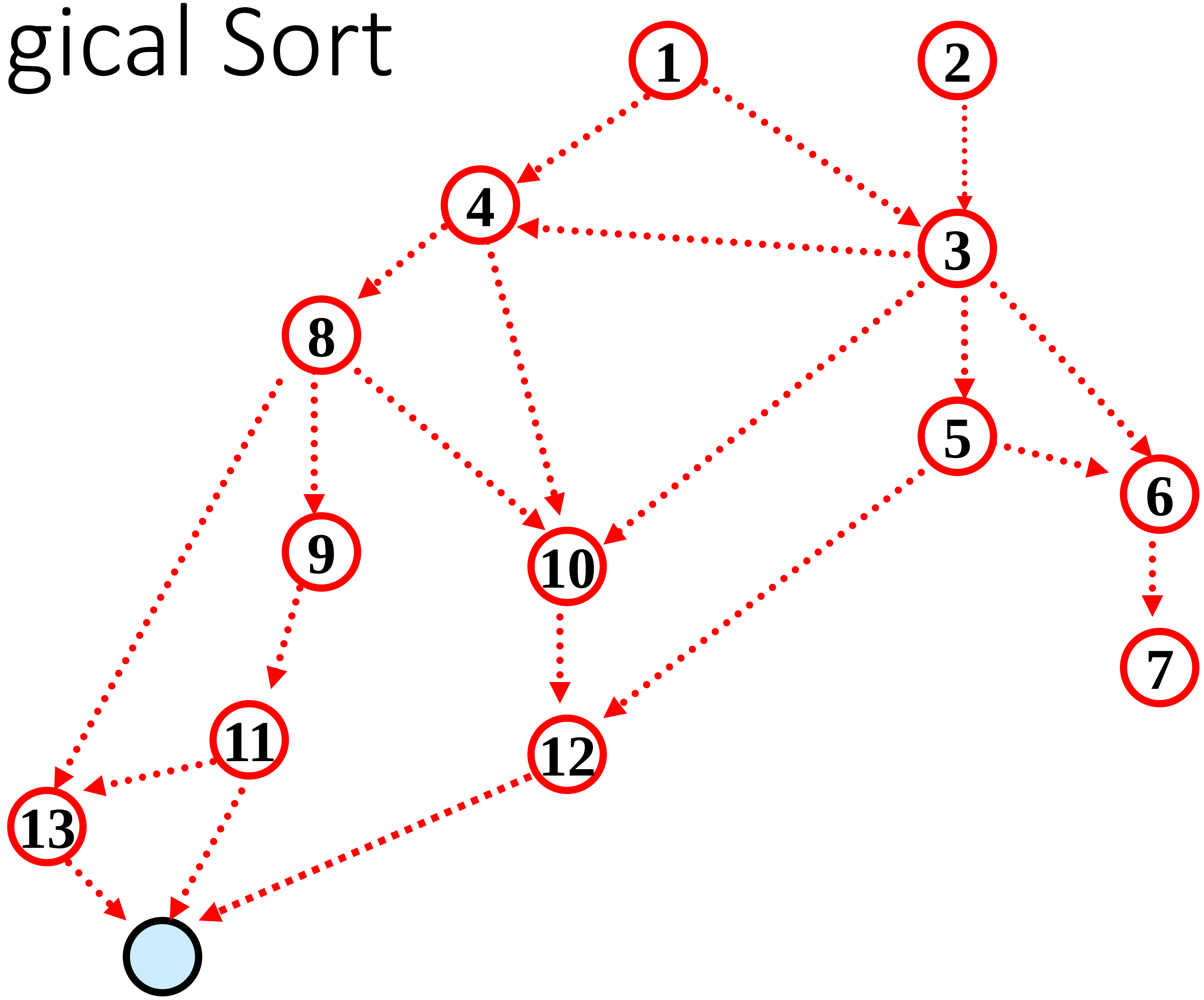
Topological Sort



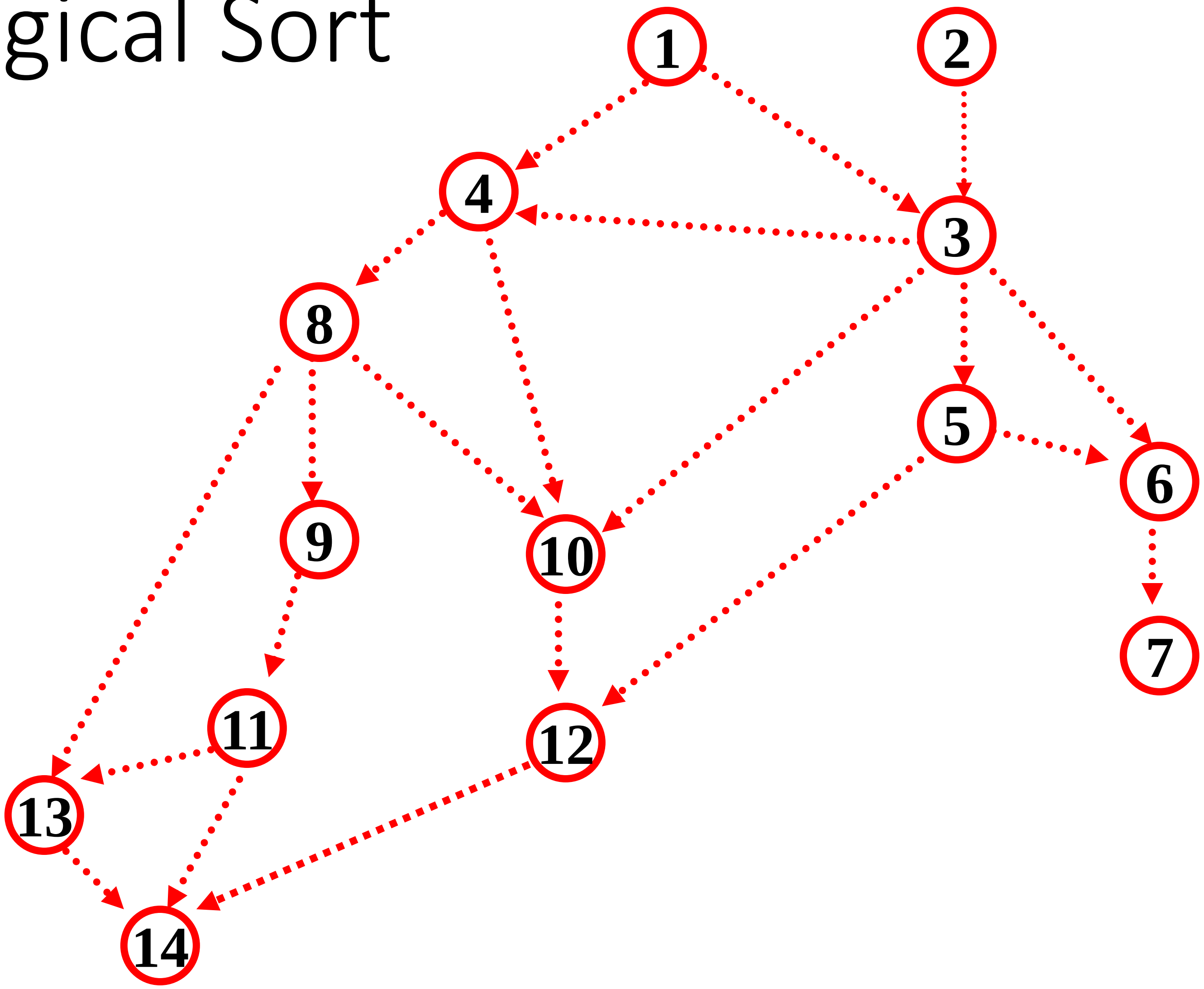
Topological Sort



Topological Sort



Topological Sort



Implementing Topological Sort

- Go through all edges, computing array with in-degree for each vertex
  $\boldsymbol{O(m+n)}$

- Maintain a list of vertices of in-degree $\mathbf{0}$

- Remove any vertex in list and number it

- When a vertex is removed, decrease in-degree of each neighbor by $\mathbf{1}$ and add them to the list if their degree drops to $\mathbf{0}$

Total cost: $\boldsymbol{O(m+n)}$

Strongly Connected Components of Directed Graphs

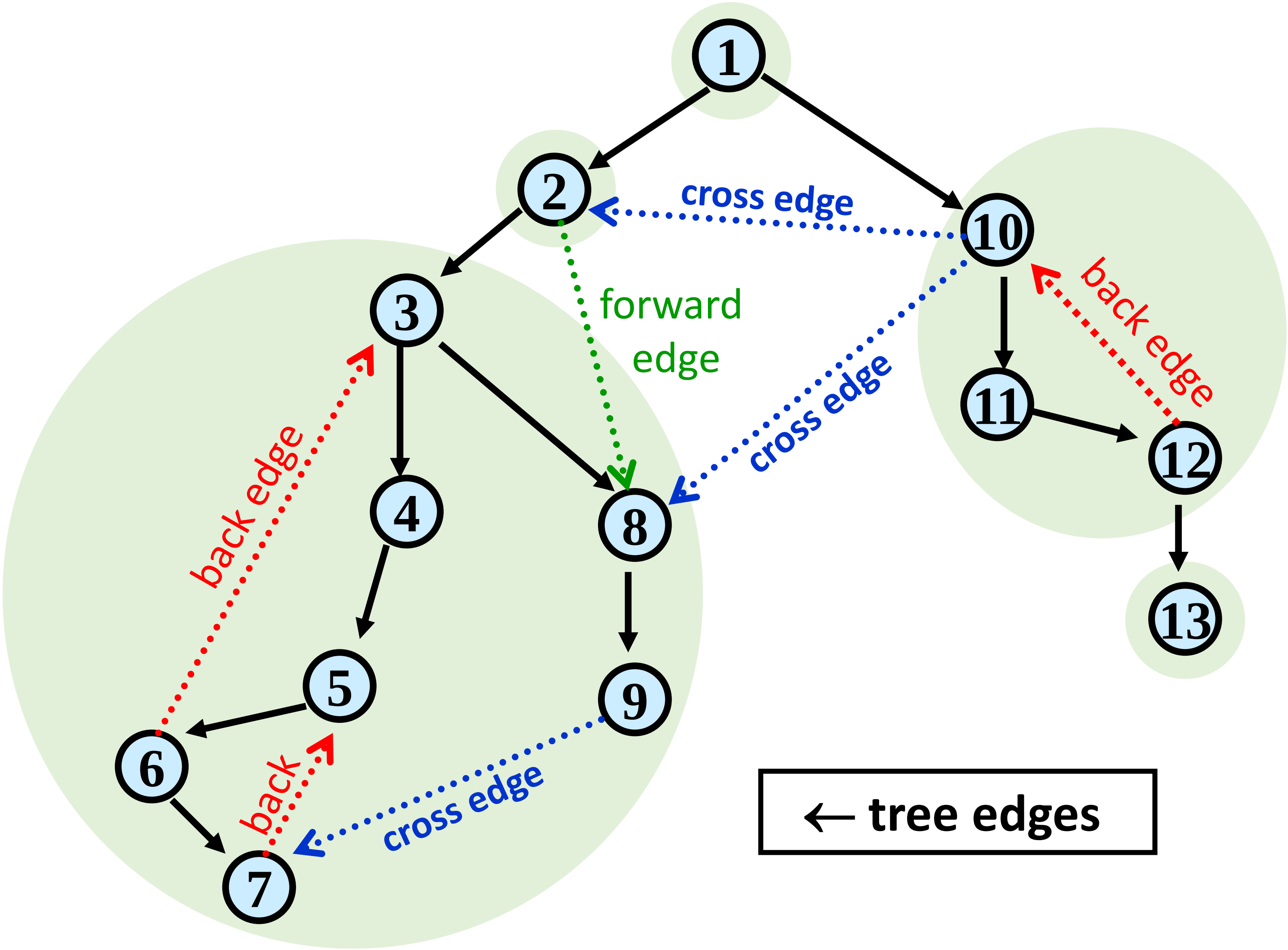
Defn: Vertices u and v are strongly connected iff they are on a directed cycle (there are paths from u to v and from v to u).

Defn: Can partition vertices of any directed graph into strongly connected components:

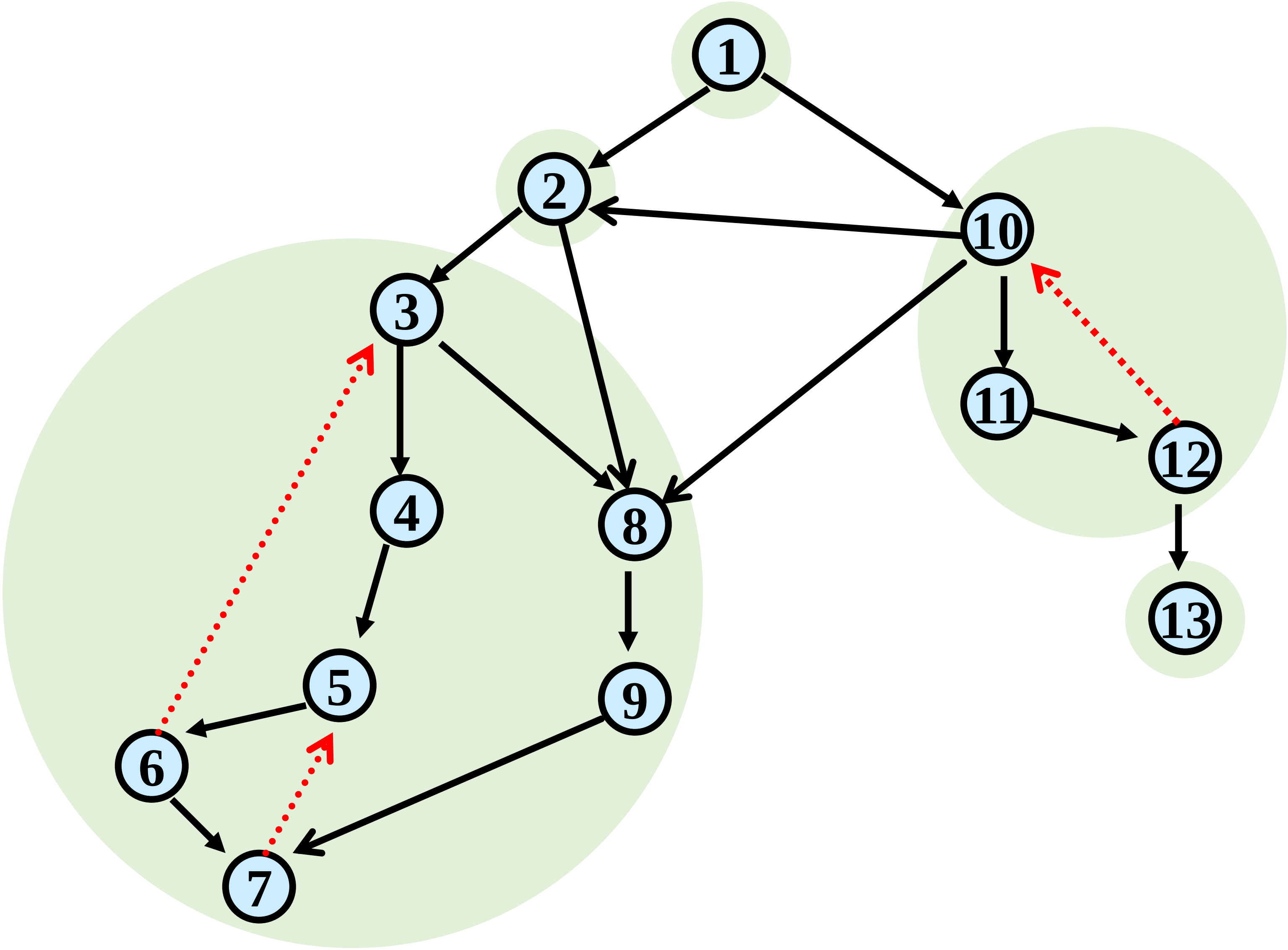
1. all pairs of vertices in the same component are strongly connected
2. can't merge components and keep property 1

- Strongly connected components can be stored efficiently just like connected components
- Can be found in $O(n+m)$ time using a DFS then a BFS
  - Do a depth-first sort, keeping track of the order nodes are marked "fully-explored"
  - Going in order from least recent to most recent, run connected components

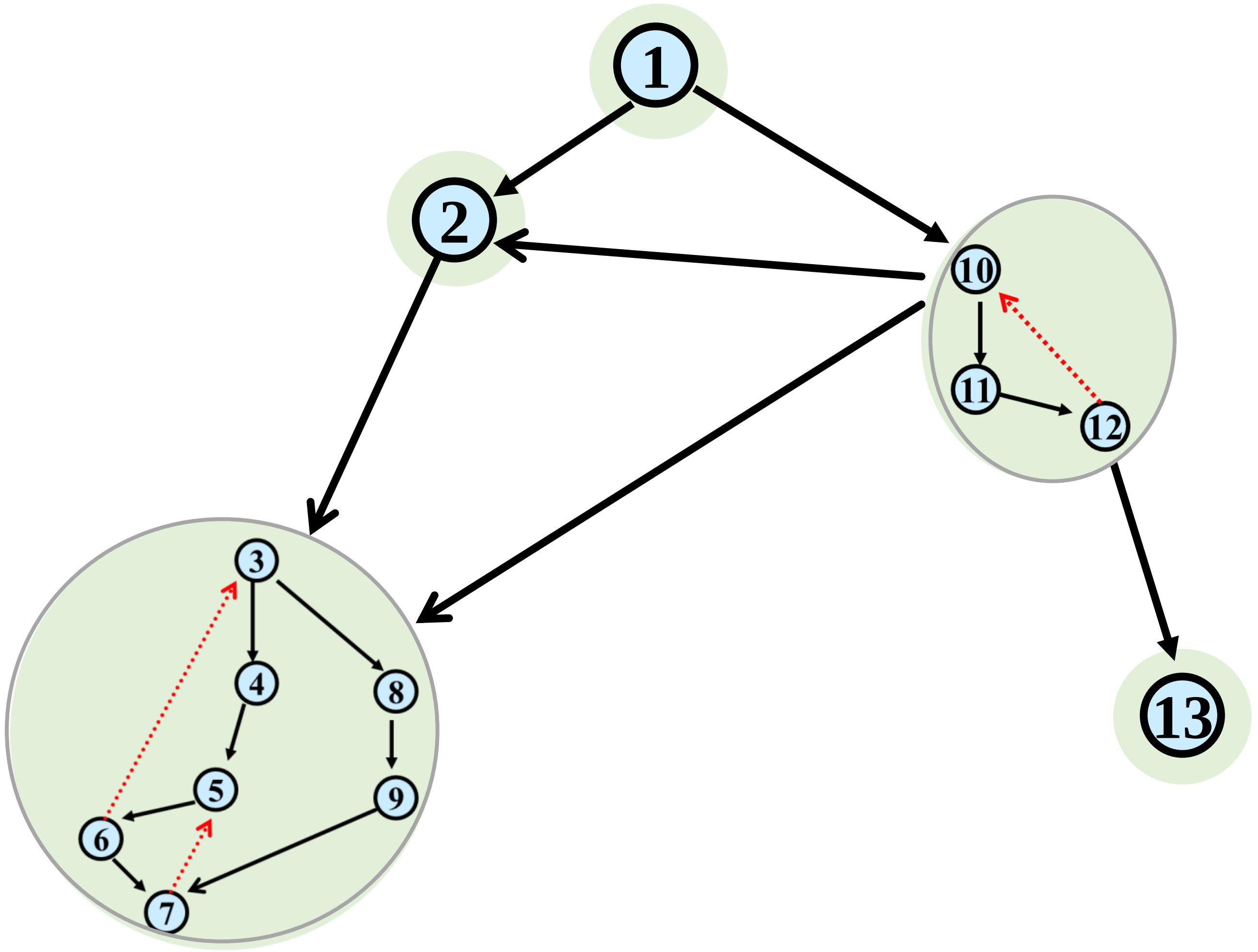
Strongly Connected Components



Strongly Connected Components



Strongly Connected Components



Strongly-Connected Components Usage

Common algorithmic paradigm for general directed graphs:

- Process strongly connected components one-by-one in the order given by topological sort of the DAG you get from shrinking them.

Greedy Algorithms

Hard to define exactly but can give general properties

- Solution is built in small steps
- Decisions on how to build the solution are made to maximize some criterion without looking to the future
  - Want the 'best' current partial solution as if the current step were the last step

May be more than one greedy algorithm using different criteria to solve a given problem

- Not obvious which criteria will actually work

Greedy Algorithms

- Greedy algorithms
  - Easy to describe
  - Fast running times
  - Work only on certain classes of problems
    - Hard part is showing that they are correct

- Focus on methods for proving that greedy algorithms do work

Interval Scheduling

Interval Scheduling:

- Single resource
- Reservation requests of form:

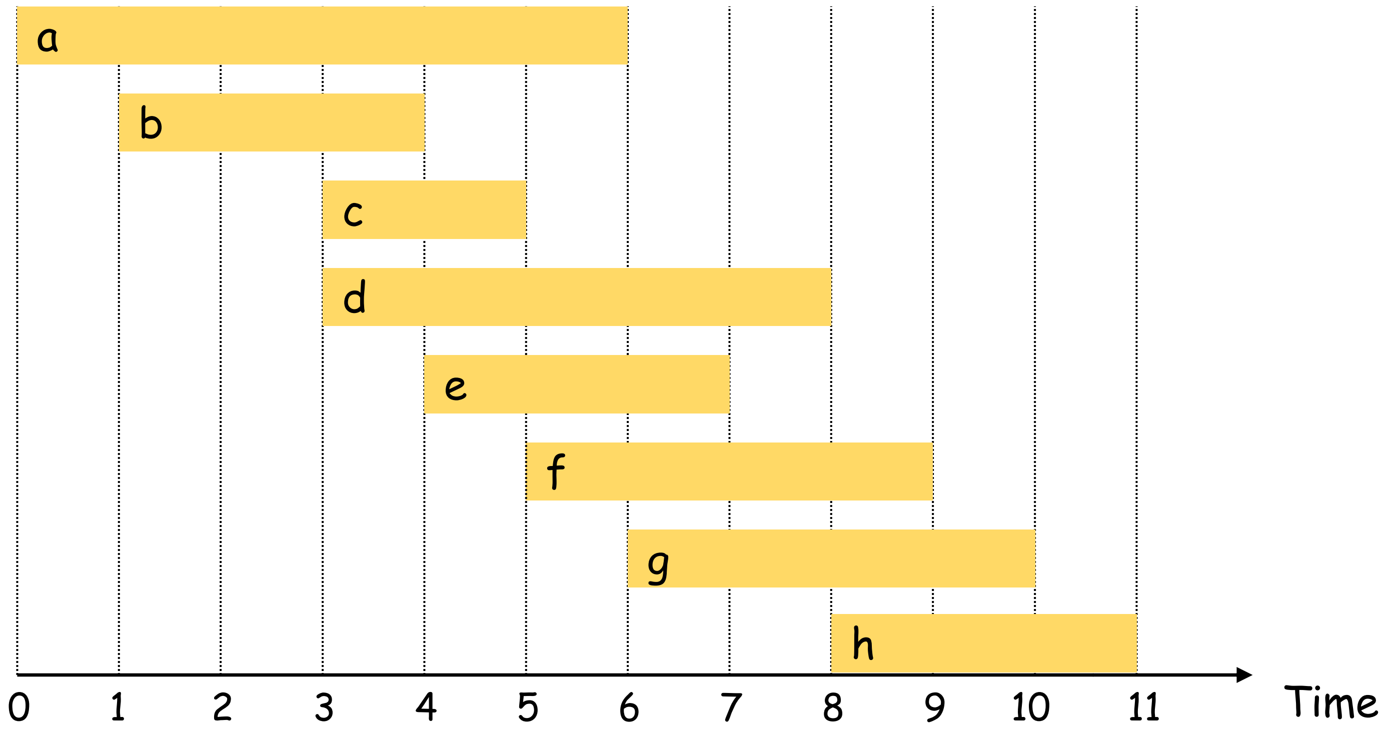
  "Can I reserve it from start time s to finish time f?"

  $$s < f$$

Interval Scheduling

Interval scheduling:

- Job $\boldsymbol{j}$ starts at $\boldsymbol{s_j}$ and finishes at $\boldsymbol{f_j > s_j}$.
- Two jobs $\boldsymbol{i}$ and $\boldsymbol{j}$ are compatible if they don't overlap: $\boldsymbol{f_i \leq s_j}$ or $\boldsymbol{f_j \leq s_i}$
- **Goal:** find maximum size subset of mutually compatible jobs.



Greedy Algorithms for Interval Scheduling

- What criterion should we try?

Greedy Algorithms for Interval Scheduling

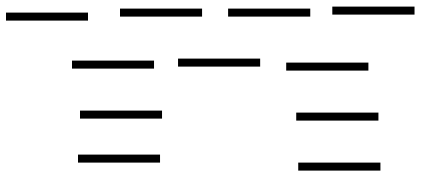
- What criterion should we try?
  - Earliest start time $\boldsymbol{s_i}$
  - Shortest request time $\boldsymbol{f_i - s_i}$
  - Fewest conflicts

Greedy Algorithms for Interval Scheduling

- What criterion should we try?
  - Earliest start time s_i
    - Doesn't work

  - Shortest request time $f_i - s_i$
    - Doesn't work

  - Fewest conflicts
    - Doesn't work



  - Earliest finish time f_i
    - Works!

Greedy (by finish time) Algorithm for Interval Scheduling

$\boldsymbol{R}$ = set of all requests

$\boldsymbol{A} = \varnothing$

while $\boldsymbol{R} \neq \varnothing$ do:

Choose request $\boldsymbol{i} \in \boldsymbol{R}$ with smallest finish time $\boldsymbol{f_i}$

Add request $\boldsymbol{i}$ to $\boldsymbol{A}$

Delete all requests in $\boldsymbol{R}$ not compatible with request $\boldsymbol{i}$

return $\boldsymbol{A}$

Greedy Analysis Strategies

Greedy algorithm stays ahead: Show that after each step of the greedy algorithm, its solution is at least as good as any other algorithm's

For interval scheduling: Show that after the greedy algorithm selects each interval, any alternative schedule's selection would have also been non-conflicting.

Conclusion: Each choice from the alternative selections can be swapped with a greedy choice, making greedy no worse off.

Interval Scheduling: Analysis

Claim: $\boldsymbol{A}$ is a compatible set of requests and
requests are added to $\boldsymbol{A}$ in order of finish time

- When we add a request to $\boldsymbol{A}$ we delete all incompatible ones from $\boldsymbol{R}$

Name the finish times of requests in $\boldsymbol{A}$ as $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_t$ in order.

Claim: Let $\boldsymbol{O} \subseteq \boldsymbol{R}$ be a set of compatible requests whose finish times in order are $\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_s$. Then for every integer $\boldsymbol{k} \geq 1$ we have:

a) if $\boldsymbol{O}$ contains a $\boldsymbol{k}^{\text{th}}$ request then $\boldsymbol{A}$ does too, and

b) $\mathbf{a}_k \leq \mathbf{o}_k$ "$\boldsymbol{A}$ is ahead of $\boldsymbol{O}$"

Note that a) alone implies that $\boldsymbol{t} \geq \boldsymbol{s}$ which means that $\boldsymbol{A}$ is optimal but we also need b) "stays ahead" to keep the induction going.

Inductive Proof of Claim

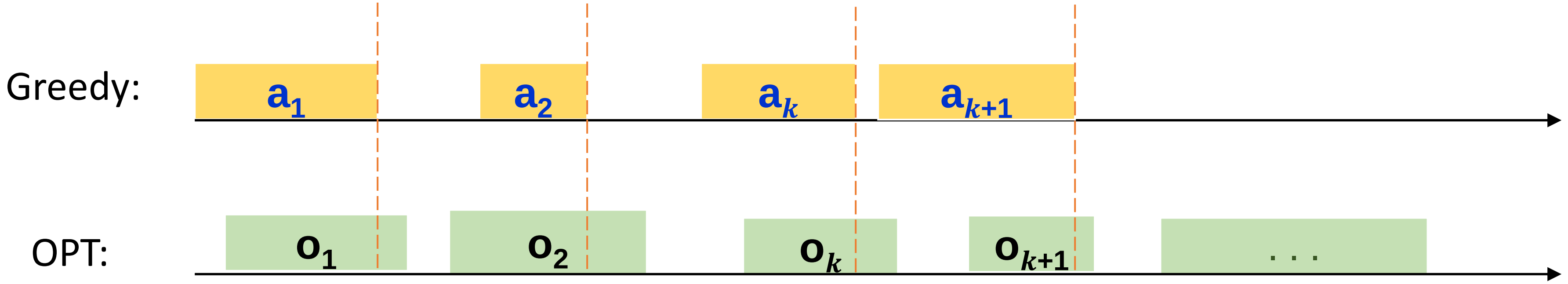
Base Case $k = 1$: A includes the request with smallest finish time, so if O is not empty then $\mathbf{a}_1 \le \mathbf{o}_1$

Inductive Step: Suppose that $\mathbf{a}_k \le \mathbf{o}_k$ and there is a k+1st request in O.

Then k+1st request in O is compatible with $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ since $\mathbf{a}_k \le \mathbf{o}_k$ and $\mathbf{o}_k \le$ start time of k+1st request in O whose finish time is $\mathbf{o}_{k+1}$

⇒ There is a k+1st request in A whose finish time is named $\mathbf{a}_{k+1}$.

Also, since A would have considered both requests and chosen the one with the earlier finish time, $\mathbf{a}_{k+1} \le \mathbf{o}_{k+1}$.



Interval Scheduling: Greedy Algorithm Implementation

```
Sort jobs by finish times so that 0 ≤ f1 ≤ f2 ≤ ... ≤ fn.

A = φ
last = 0
for j = 1 to n {
    if (last ≤ sj)
      A = A ∪ {j}
      last = fj
}
return A
```

$O(n \log n)$

$O(n)$