Graphs

\( G = (V, E) \): 

- \( V \) = set of nodes (a.k.a., vertices) 
- \( E \) = set of edges 

\( G \) is usually assumed to be directed, so that an edge is a pair of nodes \((v, w)\) (graphically, \(v \rightarrow w\)) 

If \((v, w) \in E\), let’s call \(w\) a successor of \(v\), and \(v\) a predecessor of \(w\) 

Successor(\(v\)) := set of all successors of \(v\) 

An undirected graph can be thought of as a directed graph, where \((v, w) \in E \Rightarrow (w, v) \in E\)
Graph representations

- **Sparse**: an array of *adjacency lists*
  - an array $A$ indexed by $V$, where $A[v]$ is a linked list containing all successors of $v$
  - size: $O(|V| + |E|)$
  - this will be the “default”

- **Dense**: an *adjacency matrix*
  - a 2D boolean array $A$ indexed by $V \times V$, where $A[v, w] = true$ iff $(v, w) \in E$
  - size: $O(|V|^2)$
  - May be preferable when $|E| \approx |V|^2$
Example:

\[ V = \{1, 2, 3, 4, 5, 6\} \]
\[ E = \{(1, 2), (1, 4), (2, 3), (2, 4), (2, 5), (3, 4), (5, 3), (6, 3), (6, 5)\} \]
Weighted graphs

Edges can have weights, representing costs or distances.

Weighted Digraph

Adjacency Lists

Adjacency matrix
Directed graphs: cycles

A **cycle** in a graph is a path (with at least one edge) from a node back to itself:

\[ V_0 \rightarrow V_1 \rightarrow V_2 \cdots \rightarrow V_{k-1} \rightarrow V_0 \]

Example:

A **DAG (Directed Acyclic Graph)** is a directed graph with no cycles
Topological Sort

Suppose $G = (V, E)$ is a DAG

A topological sort of $G$ is an ordering of the vertices $(v_1, v_2, \ldots, v_n)$ such that $(v_i, v_j) \in E \Rightarrow i < j$

“all arrows go from left to right”

Example:
Topological Sort

Applications:

• “Generic application”:
  ○ Nodes represents “tasks”
  ○ Edges represent “scheduling constraints”
    · $v \rightarrow w$ means “task $v$ must be performed before task $w$”
  ○ A topological sort is a schedule for performing all tasks that respects all of the constraints

• Concrete applications:
  ○ Software build systems (e.g., Unix make command)
  ○ Software packaging systems (e.g., apt-get from Linux)
  ○ Spreadsheets (ordering of formula cell evaluation)
Why only DAGs?

There was only one catch and that was Catch-22, which specified that a concern for one’s safety in the face of dangers that were real and immediate was the process of a rational mind. Orr was crazy and could be grounded. All he had to do was ask; and as soon as he did, he would no longer be crazy and would have to fly more missions. Orr would be crazy to fly more missions and sane if he didn’t, but if he was sane he had to fly them. If he flew them he was crazy and didn’t have to; but if he didn’t want to he was sane and had to. Yossarian was moved very deeply by the absolute simplicity of this clause of Catch-22 and let out a respectful whistle.

“That’s some catch, that Catch-22,” he observed.

“It’s the best there is,” Doc Daneeka agreed.

– Joseph Heller, Catch-22
Theorem: A directed graph $G$ is a DAG $\iff$ it has a topological sort

Proof: ( $\iff$ ) — easy direction

Want to show: if $G$ has a top. sort, then $G$ is a DAG

We prove the contrapositive: if $G$ is not a DAG, then $G$ cannot have a top. sort

Suppose $G$ has a cycle: $v_0 \to v_1 \to v_2 \cdots \to v_{k-1} \to v_0$

There is a path from $v_0$ to $v_1$, so $v_0$ must come before $v_1$ in any top. sort

There is also a path from $v_1$ to $v_0$, so $v_1$ must come before $v_0$ in any top. sort

$\therefore G$ cannot have a top. sort
Theorem: A directed graph $G$ is a DAG $\iff$ it has a topological sort

Proof: ( $\iff$ ) — harder direction

Want to show: if $G$ is a DAG, then $G$ has a top. sort

Proof by induction on $n = \#$ nodes

Base case: $n = 1 \checkmark$

Let $n > 0$, and assume every DAG with $n - 1$ nodes has a top. sort (induction hypothesis)

We show that every DAG with $n$ nodes has a top. sort
Topological Sort

(Proof cont’d)

Let $G$ be a DAG with $n$ nodes

Want to show: $G$ has a top. sort

There must be a node $v_1$ that has in-degree 0, i.e., with no incoming edges (why?)

Delete $v_1$ (along with all of its outgoing edges) from $G$, giving us a DAG $G'$ with $n - 1$ nodes

By induction hypothesis, $G'$ has a top. sort $(v_2, \ldots, v_n)$

$(v_1, v_2, \ldots, v_n)$ is a top. sort for $G$ (why?)

QED
Topological Sort

The proof suggests an algorithm:

\[ L \leftarrow () \quad // \text{Initialize empty list} \]
while \( G \) contains a node \( v \) of in-degree 0 do
  delete \( v \) (and its outgoing edges) from \( G \)
  append \( v \) to \( L \)

if \( G \) is empty
  then output \( L \)
else output “graph contains a cycle”
Topological Sort

Example:

There's only one vertex with in-degree 0 (vertex 1), so we start with that one (and think of it as removed from the graph):
Topological Sort

Example:

1

Topological sort

● An example:
● As we “remove” 1, now 2 and 3 have in-degree 0, so the topological sort could continue with either one of these — we’ll choose 2:
Topological Sort

Example:

1, 2
Topological Sort

Example:

1, 2, 4
Topological Sort

Example:

1, 2, 4, 8
Topological Sort

Example:

1, 2, 4, 8, 5
Topological Sort

Example: ...

1, 2, 4, 8, 5, 9, 11, 3, 6, 10, 7, 12, 13
Topological Sort

Kahn’s Algorithm: a linear time — $O(|V| + |E|)$ — implementation

We assume:

- $V = \{0, \ldots, n - 1\}$
- Adjacency list representation

We need:

- An array $deg[0..n)$ (initially all zero)
  // Keeps track of in-degrees
- A set $S$ of nodes (initially empty)
  // tracks all nodes of in-degree 0
- A list $L$ (initially empty)
  // The topological sort
Topological Sort

// Initialize in-degrees
for \( \nu \) in \([0..n]\) do
    for each \( \omega \in \text{Successor}(\nu) \) do
        increment \( \text{deg}[\omega] \)

// Initialize \( S \) to all nodes of in-degree 0
for \( \nu \) in \([0..n]\) do
    if \( \text{deg}[\nu] = 0 \) then add \( \nu \) to \( S \)

// Main loop
while \( S \) is not empty do
    remove some \( \nu \) from \( S \)
    append \( \nu \) to \( L \)
    for each \( \omega \in \text{Successor}(\nu) \) do
        decrement \( \text{deg}[\omega] \)
        if \( \text{deg}[\omega] = 0 \) then add \( \omega \) to \( S \)

output \( L \)  // Graph is a DAG \iff \( L \) has length \( n \)
Topological Sort

Running time analysis:

- “Initialize in-degrees” and “Main loop”: each vertex $v$ and each edge $v \rightarrow w$ is visited once
- \( \therefore \) total running time is \( O(|V| + |E|) \)
Example: gathering coins

Given a DAG $G = (V, E)$

On each node $v$ there are $N[v]$ coins

Goal: find the max number of coins that can be gathered on any one path through $G$
Solution:
Assume $V = [0..n)$ and let $TopSort[0..n)$ be an array that lists vertices in a topological order.

We will compute $P[
u] = \text{maximum number of coins you can gather on any path starting at } \nu$

for $i$ in reverse $[0..n)$

$\quad u \leftarrow \text{TopSort}[i]$

$\quad P[u] \leftarrow N[u] + \max \left[ \{ P[\nu] : \nu \in \text{Successor}(u) \} \cup \{0\} \right]$

A more concrete implementation of “$d[u] \leftarrow N[u] + \max \cdots$”

$max \leftarrow 0$

for each $\nu \in \text{Successor}(u) do$

$\quad$ if $P[\nu] > max$ then $max \leftarrow P[\nu]$

$\quad P[u] \leftarrow N[u] + max$