Graph Algorithms

Textbook reading

Chapter 3
Chapter 4
Midterm

Alternate Tue. June 21

Thu June 23

Midterm

Tue

New Midterm Date

Thu June 30
Overview

Design principle:
- Learn the structure of a graph by systematic exploration

Proof techniques:
- Proof by contradiction

Problems:
- Bipartiteness
- Connectivity
- Strong connectivity
- Topological sorting
A **graph** is an ordered pair $G = (V, E)$.

- $V$ is the set of **vertices** of $G$.
- $E$ is the set of **edges** of $G$.
- The elements of $E$ are pairs $(v, w)$ of vertices.
A **graph** is an ordered pair $G = (V, E)$.

- $V$ is the set of **vertices** of $G$.
- $E$ is the set of **edges** of $G$.
- The elements of $E$ are pairs $(v, w)$ of vertices.

For an edge $(v, w) \in E$, we call vertices $v$ and $w$ **adjacent**.

Edge $(v, w)$ is **incident** to $v$ and $w$.

Vertices $v$ and $w$ are the **endpoints** of edge $(v, w)$.

The **degree** of a vertex $v$ is the number of edges incident to $v$. 
A graph is **undirected** if its edges are unordered pairs \((v, w)\), that is, \((v, w) = (w, v)\).

A graph is **directed** if its edges are ordered pairs \((v, w)\), that is, \((v, w) \neq (w, v)\).

Edge \((v, w)\) is an **out-edge** of \(v\) and an **in-edge** of \(w\).

The **in-degree** and **out-degree** of a vertex \(v\) are the numbers of in-edges and out-edges incident to \(v\).
A path $P = (x = v_0, v_1, \ldots, v_k = y)$ from a vertex $x$ to a vertex $y$ is a sequence of vertices such that $(v_{i-1}, v_i)$ is an edge, for all $1 \leq i \leq k$.

A cycle is a path from a vertex $x$ back to itself.

A path or cycle is simple if it contains every vertex of $G$ at most once.
A graph is **connected** if there exists a path from $x$ to $y$, for any two vertices $x$ and $y$ of $G$. 

connected

not connected
n vertices
m edges

O(1) is there an edge (v,w)
O(n^2) space
ntm

100 vertices
2 edges for each vertex
100 + 200 = 300 input size AM: 100^2

sparse: relatively few edges
dense: O(n^2)
Adjacency-List Representation of Graphs

- Doubly-linked vertex list
- Doubly-linked edge list
- One doubly-linked adjacency list per vertex
- Pointers from adjacency list entries to vertices
- Cross-pointers between edges and adjacency list entries
Many problems are quite naturally expressed as graph problems:

**Example:** Stable matching is a special case of bipartite perfect matching.

A graph is **bipartite** if its vertices can be divided into sets $X$ and $Y$ so that every edge has one endpoint in $X$ and the other in $Y$.

A **matching** of a graph is a subset of edges so that no two edges in the set share an endpoint.

A matching is **perfect** if every vertex is the endpoint of an edge.
**Modelling Real-World Problems (2)**

*Example:* Airline scheduling

- There are $n$ lucrative flight segments to be serviced
- Flight segment = (source, destination, departure time, arrival time)

**Question:** Can we service all $n$ segments using the $k$ planes in our fleet?

**Rules:**
- Same plane can service two segments $(s_1, t_1, d_1, a_1)$ and $(s_2, t_2, d_2, a_2)$ if $t_1 = s_2$
- There is enough time for maintenance between arrival time $a_1$ and departure time $d_2$
- We can add other flight segments to get a plane that arrived at destination $t_1$ to service segment from destination $s_2$ by adding extra flight segments that get us from $t_1$ to $s_2$. The same rules about maintenance periods between arrivals and departures apply.
Graph-theoretic formulation:
Graph-theoretic formulation:
Graph-theoretic formulation:

Are there $k$ paths in this network whose union includes all solid edges?
Graph-theoretic formulation:

Are there $k$ paths in this network whose union includes all solid edges?
Graph-theoretic formulation:

Are there $k$ paths in this network whose union includes all solid edges?

This reduces to a flow problem. See textbook, Chapter 7.
Example: Ordering tasks under constraints

Building a shack

1. Buy boards
2. Buy nails
3. Buy hammer
4. Buy hinges
5. Add roof
6. Erect sides of shack
7. Assemble door
8. Insert door in door frame

buy boards → buy nails → add roof

CSci 3110 • Graph Algorithms • 11/41
Example: Ordering tasks under constraints

Building a shack

1 Buy boards
2 Buy nails
3 Buy hammer
4 Buy hinges
5 Add roof
6 Erect sides of shack
7 Assemble door
8 Insert door in door frame

Topological sorting:
Number the vertices so that, for every edge \((v, w)\), \(v < w\).
Modelling Real-World Problems (4)

More examples:
- Communication networks
- Transportation networks
- Data structures
- ...
A graph with $n$ vertices and $n-1$ edges forms a tree.
**Graph Exploration (1)**

**EXPLORE-GRAPH** ($G$)

1. Mark every vertex and every edge of $G$ as unexplored
2. **for** every vertex $v$ of $G$
3. **do if** $v$ is unexplored
4. **then** Explore-from-Vertex($G, v$)

**EXPLORE-FROM-VERTEX** ($G, v$)

1. Mark $v$ as explored
2. $S \leftarrow$ Adj($v$)
3. **while** $S$ is not empty
4. **do** Remove an edge $(x, y)$ from $S$
5. **if** $(x, y)$ is not explored
6. **then if** $y$ is not explored
7. **then** Mark $(x, y)$ as a tree edge
8. Mark $y$ as explored
9. $S \leftarrow S \cup$ Adj($y$)
10. **else** Mark $(x, y)$ as a non-tree edge
**Graph Exploration Variants (1): Depth-First Search**

**DFS-FROM-VERTEX**$(G, v)$

1. Mark $v$ as explored
2. $\triangleright S$ is a stack
3. <for every edge $(v, w)$ incident to $v$
   4. do Push$(S, (v, w))$ $O(m)$
5. **while** $S$ is not empty $O(m)$
   6. do $(x, y) \leftarrow$ Pop$(S)$
   7. if $(x, y)$ is not explored
      8. **then if** $y$ is not explored
         9. then Mark $(x, y)$ as a tree edge
         10. Mark $y$ as explored
      **for** every edge $(y, z)$ incident to $y$
         12. do Push$(S, (y, z))$ $O(m)$
   13. **else** Mark $(x, y)$ as a non-tree edge

$O(n+m)$

$n$ vertices

$m$ edges
Graph Exploration Variants (1): Depth-First Search

DFS-FROM-VERTEX($G, v$)
1 Mark $v$ as explored
2 ▷ $S$ is a stack
3 for every edge $(v, w)$ incident to $v$
4 do Push($S, (v, w)$)
5 while $S$ is not empty
6 do ($x, y$) ← Pop($S$)
7 if ($x, y$) is not explored
8 then if $y$ is not explored
9 then Mark $(x, y)$ as a tree edge
10 Mark $y$ as explored
11 for every edge $(y, z)$ incident to $y$
12 do Push($S, (y, z)$)
13 else Mark $(x, y)$ as a non-tree edge

Lemma: Depth-first search takes $O(n + m)$ time.
Graph Exploration Variants (2): Breadth-First Search

**BFS-FROM-VERTEX** \((G, v)\)

1. Mark \(v\) as explored
2. \(\triangleright S\) is a queue
3. for every edge \((v, w)\) incident to \(v\)
4. \hspace{1em} do Enqueue \((S, (v, w))\)
5. \hspace{1em} while \(S\) is not empty
6. \hspace{2em} do \((x, y) \leftarrow\) Dequeue \((S)\)
7. \hspace{2em} if \((x, y)\) is not explored
8. \hspace{3em} then if \(y\) is not explored
9. \hspace{4em} then Mark \((x, y)\) as a tree edge
10. \hspace{3em} Mark \(y\) as explored
11. \hspace{2em} for every edge \((y, z)\) incident to \(y\)
12. \hspace{3em} do Enqueue \((S, (y, z))\)
13. \hspace{2em} else Mark \((x, y)\) as a non-tree edge
Graph Exploration Variants (2): Breadth-First Search

BFS-FROM-VERTEX\((G, v)\)

1. Mark \(v\) as explored
2. ▷ \(S\) is a queue
3. for every edge \((v, w)\) incident to \(v\)
   4. do Enqueue\((S, (v, w))\)
5. while \(S\) is not empty
   6. do \((x, y) \leftarrow\) Dequeue\((S)\)
   7. if \((x, y)\) is not explored
       8. then if \(y\) is not explored
           9. then Mark \((x, y)\) as a tree edge
       10. Mark \(y\) as explored
   11. for every edge \((y, z)\) incident to \(y\)
       12. do Enqueue\((S, (y, z))\)
   13. else Mark \((x, y)\) as a non-tree edge

Lemma: Breadth-first search takes \(O(n + m)\) time.
**Graph Exploration Variants (3): Dijkstra’s Algorithm**

**DIJKSTRA-FROM-VERTEX**(*G, v*)

1. Mark *v* as explored
2. \(^\triangleright\) *S* is a priority queue
3. **for** every edge \((v, w)\) incident to *v*
4. **do** Insert\((S, (v, w), w(v, w))\)
5. **while** *S* is not empty
6. **do** \((x, y) \leftarrow\) Delete-Min\((S)\)
7. **if** \((x, y)\) is not explored
8. **then if** *y* is not explored
9. **then** Mark \((x, y)\) as a tree edge
10. Mark *y* as explored
11. Let dist\((v, y)\) be the priority of edge \((x, y)\)
12. **for** every edge \((y, z)\) incident to *y*
13. **do** Insert\((S', (y, z), \text{dist}(s, y) + w(y, z))\)
14. **else** Mark \((x, y)\) as a non-tree edge

CSci 3110 • Graph Algorithms • 17/41
**Prim’s Algorithm**

**Prim-from-Vertex** \((G, v)\)

1. Mark \(v\) as explored
2. \(\triangleright S\ is a priority queue\)
3. **for** every edge \((v, w)\) incident to \(v\)
4. **do** Insert \((S, (v, w), w(v, w))\)
5. **while** \(S\) is not empty
6. **do** \((x, y) \leftarrow \text{Delete-Min}(S)\)
7. **if** \((x, y)\) is not explored
8. **then if** \(y\) is not explored
9. **then** Mark \((x, y)\) as a tree edge
10. Mark \(y\) as explored
11. **for** every edge \((y, z)\) incident to \(y\)
12. **do** Insert \((S, (y, z), w(y, z))\)
13. **else** Mark \((x, y)\) as a non-tree edge
Depth-first search, breadth-first search, Dijkstra’s algorithm, and Prim’s algorithm are variants of the same graph exploration procedure:

Maintain a set of explored vertices. Grow this set by exploring edges incident to these vertices.

What differs is the order in which edges are explored:

- **DFS**: Choose the edge whose source vertex has been discovered most recently.
- **BFS**: Choose the edge whose source vertex has been discovered first.
- **Dijkstra**: Choose the edge whose target is unexplored and has the minimum tentative distance among all unexplored vertices.
- **Prim**: Choose the edge whose target is unexplored and which has minimum weight.
Lemma: Let $s$ be a vertex of $G$, let $T$ be a BFS-tree rooted at $s$, and let $(u, v)$ be an edge of $G$. Then $|\text{dist}_T(s, u) - \text{dist}_T(s, v)| \leq 1$.

In other words, $u$ and $v$ are on the same level or on adjacent levels of $G$. 
Suppose two vertices $uv$ are more than one level apart but share an edge $(uv)$

- assume an edge $(wuv)$ was explored instead.
- but $u$’s edges would have been examined before $w$’s because $u$ is at a higher level of the tree.
- at this step in the algorithm, $v$ is unexplored so $(uv)$ must be a tree edge.
- contradiction
Problem: Given a graph $G$, decide whether it is bipartite.
Problem: Given a graph $G$, decide whether it is bipartite.
Problem: Given a graph $G$, decide whether it is bipartite.
**Problem:** Given a graph $G$, decide whether it is bipartite.

**Lemma:** A graph is bipartite if and only if it does not contain an odd cycle.
Proof Sketch

If there is no odd cycle then we can take two arbitrary sets $X, Y$ and modify them to be bipartite.

If there is an odd cycle then there must be two vertices $x_1, x_2 \in X$ that share an edge.

→

even cycle

odd cycle
Lemma: A graph $G$ is bipartite if and only if there are no two adjacent vertices that are on the same level in a BFS-tree of $G$. 
Lemma: A graph $G$ is bipartite if and only if there are no two adjacent vertices that are on the same level in a BFS-tree of $G$. 
**Lemma:** A graph $G$ is bipartite if and only if there are no two adjacent vertices that are on the same level in a BFS-tree of $G$.  

If:

Only if:
**Test-Bipartiteness** \((G)\)

1. Run BFS on \(G\) to label all vertices with their distance \(d(v)\) from some vertex \(s\).

2. For every edge \((v, w)\) of \(G\),
   
   - do if \(d(v) = d(w)\),
   
   - then Report that \(G\) is not bipartite and exit.

3. \(X \leftarrow \{v \in G \mid d(v)\text{ is odd}\}\)

4. \(Y \leftarrow \{v \in G \mid d(v)\text{ is even}\}\)
**TEST-BIPARTITENESS**($G$)

1. Run BFS on $G$ to label all vertices with their distance $d(v)$ from some vertex $s$
2. for every edge $(v, w)$ of $G$
3. do if $d(v) = d(w)$
4. then Report that $G$ is not bipartite and exit
5. $X \leftarrow \{v \in G \mid d(v) \text{ is odd}\}$
6. $Y \leftarrow \{v \in G \mid d(v) \text{ is even}\}$

Lemma: *Given a graph $G$, one can decide in $O(n + m)$ time whether $G$ is bipartite.*
A Recursive Depth-First Search Procedure

\textbf{DFS}(G)
\begin{enumerate}
\item Mark every vertex and every edge of \( G \) as unexplored
\item \textbf{for} every vertex \( v \) of \( G \)
\item \textbf{do if} \( v \) is unexplored
\item \textbf{then} \text{DFS-from-Vertex}(G, v)
\end{enumerate}

\textbf{DFS-from-Vertex}(G, v)
\begin{enumerate}
\item Mark \( v \) as explored
\item \textbf{for} every out-edge \((v, w)\) of \( v \)
\item \textbf{do if} \( w \) is unexplored
\item \textbf{then} Mark \((v, w)\) as a tree edge
\item \text{DFS-from-Vertex}(G, w)
\item \textbf{else} Mark \((v, w)\) as a non-tree edge
\end{enumerate}
Lemma: For every non-tree edge \((u, v)\) in an undirected graph \(G\) w.r.t. a DFS-tree \(T\) of \(G\), either \(u\) is an ancestor of \(v\) or vice versa; that is, there are no cross edges.
Lemma: Every call to DFS-from-Vertex in Line 4 of Procedure DFS completely explores a connected component of $G$. 

We do not explore more. 

We do not explore less.
Computing Connected Components

CONNECTED-COMPONENTS($G$)
1 $c \leftarrow 0$
2 Mark every vertex and every edge of $G$ as unexplored
3 for every vertex $v$ of $G$
4   do if $v$ is unexplored
5     then $c \leftarrow c + 1$
6     Label-Component-from-Vertex($G, v, c$)

LABEL-COMPONENT-FROM-VERTEX($G, v, c$)
1 Mark $v$ as explored
2 component($v$) $\leftarrow c$
3 for every out-edge ($v, w$) of $v$
4   do if $w$ is unexplored
5     then Mark ($v, w$) as a tree edge
6     Label-Component-from-Vertex($G, w, c$)
7   else Mark ($v, w$) as a non-tree edge
Computing Connected Components

**CONNECTED-COMPONENTS**(G)
1. \( c \leftarrow 0 \)
2. Mark every vertex and every edge of \( G \) as unexplored
3. **for** every vertex \( v \) of \( G \)
4. **do if** \( v \) is unexplored
5. \( c \leftarrow c + 1 \)
6. **then** \( \) Label-Component-from-Vertex\((G, v, c)\)

**LABEL-COMPONENT-FROM-VERTEX**(G, v, c)
1. Mark \( v \) as explored
2. component\((v) \leftarrow c \)
3. **for** every out-edge \((v, w)\) of \( v \)
4. **do if** \( w \) is unexplored
5. \( \) **then** Mark \((v, w)\) as a tree edge
6. **then** \( \) Label-Component-from-Vertex\((G, w, c)\)
7. **else** Mark \((v, w)\) as a non-tree edge

**Lemma:** The connected components of a graph with \( n \) vertices and \( m \) edges can be computed in \( \mathcal{O}(n + m) \) time.
Lemma: When depth-first search backtracks from a vertex $v$, all out-neighbours of $v$ are explored.
Topological Sorting (2)

**Top-Sort(G)**
1. \( c \leftarrow n \)
2. Mark every vertex and every edge of \( G \) as unexplored
3. for every vertex \( v \) of \( G \)
4. do if \( v \) is unexplored
5. then \( c \leftarrow \text{Label-Vertex}(G, v, c) \)

**Label-Vertex(G, v, c)**
1. Mark \( v \) as explored
2. for every out-edge \((v, w)\) of \( v \)
3. do if \( w \) is unexplored
4. then \( c \leftarrow \text{Label-Vertex}(G, w, c) \)
5. number\((v)\) \( \leftarrow c \)
6. \( c \leftarrow c - 1 \)
7. return \( c \)
\[ a \rightarrow b \rightarrow c \rightarrow d \rightarrow e \rightarrow f \]
\[ g \rightarrow b \rightarrow c \rightarrow e \]
Topological Sorting (2)

\[ \text{TOP-SORT}(G) \]
\[
1 \quad c \leftarrow n \quad O(1)
2 \quad \text{Mark every vertex and every edge of } G \text{ as unexplored}
3 \quad \textbf{for} \text{ every vertex } v \text{ of } G
4 \quad \textbf{do if} \text{ } v \text{ is unexplored}
5 \quad \textbf{then} \quad c \leftarrow \text{Label-Vertex}(G, v, c)
\]

\[ \text{LABEL-VERTEX}(G, v, c) \]
\[
1 \quad \text{Mark } v \text{ as explored}
2 \quad \textbf{for} \text{ every out-edge } (v, w) \text{ of } v
3 \quad \textbf{do if} \text{ } w \text{ is unexplored}
4 \quad \textbf{then} \quad c \leftarrow \text{Label-Vertex}(G, w, c)
5 \quad \text{number}(v) \leftarrow c \quad O(1)
6 \quad c \leftarrow c - 1 \quad \delta(1)
7 \quad \text{return } c
\]

Lemma: A directed acyclic graph can be topologically sorted in \(O(n + m)\) time.
Proof by Contradiction

Suppose we have two vertices $U$, $V$ so that there is an edge $(u,v)$ and number $(u) > number (v)$.

However, because $number (u) > number (v)$, we finished $u$ first, which is a contradiction because we would have followed the $(uv)$ edge and finished $v$ first.
**Detecting Cycles (1)**

**Lemma:** If graph $G$ contains a directed cycle $C$, the vertices of $C$ are discovered after the first vertex $v$ in $C$ is discovered and before the recursive call $\text{DFS-from-Vertex}(G, v)$ returns.
Lemma: If graph $G$ contains a directed cycle $C$, the vertices of $C$ are discovered after the first vertex $v$ in $C$ is discovered and before the recursive call DFS-from-Vertex$(G, v)$ returns.

- For procedure Top-Sort, this means that an out-edge of the last vertex $w$ on the cycle leads to a discovered, but unnumbered vertex, namely $v$.
- The vertices in $C$ are the vertices on the call stack between $v$ and $w$. 
Detecting Cycles (2)

Detect-Cycle-from-Vertex(G, v)
1 Mark v as visited
2 for every out-edge (v, w) of v
3 do if w is unvisited
4 then u ← Detect-Cycle-from-Vertex(G, w)
5 if u is not nil
6 then Output v
7 if u = v
8 then exit
9 else return u
10 else if w is not finished
11 then Output w
12 return w
13 Mark v as finished
14 return nil
Detecting Cycles (2)

**Detect-Cycle-from-Vertex**(\(G, v\))

1. Mark \(v\) as visited
2. for every out-edge \((v, w)\) of \(v\)
3. do if \(w\) is unvisited
4. then \(u \leftarrow \text{Detect-Cycle-from-Vertex}(G, w)\)
5. if \(u\) is not nil
6. then Output \(v\)
7. if \(u = v\)
8. then exit
9. else return \(u\)
10. else if \(w\) is not finished
11. then Output \(v\)
12. return \(w\)
13. Mark \(v\) as finished
14. return nil

**Lemma:** One can test in \(O(n + m)\) time whether a given directed graph \(G\) contains a directed cycle and, if so, output such a cycle.
A directed graph $G$ is *strongly connected* if, for any two vertices $v$ and $w$ in $G$, there exists a path from $v$ to $w$.

The *strongly connected components* of a directed graph $G$ are the maximal strongly connected subgraphs of $G$.

**Alternative formulation:** For any two vertices $v$ and $w$ in a strongly connected component, there exists a directed cycle that contains $v$ and $w$. 
A directed graph $G$ is strongly connected if, for any two vertices $v$ and $w$ in $G$, there exists a path from $v$ to $w$.

The strongly connected components of a directed graph $G$ are the maximal strongly connected subgraphs of $G$.

**Alternative formulation:** For any two vertices $v$ and $w$ in a strongly connected component, there exists a directed cycle that contains $v$ and $w$. 
A Strong Connectivity Algorithm (1)

Maintain partition of vertices into three sets:

- Finished + live components are strongly connected components of the graph defined by explored edges.
- Finished components are strongly connected components of $G$.
- Live components form a “path” and can merge into larger components as more edges are discovered.
Updating the Partition (1)

- Explore edges out of last live component.
- Three cases, depending on location of target of the edge:
  - Target is finished.
  - Target is unexplored.
  - Target is live.
- When last live component has no unexplored out-edges, mark it as finished and continue processing the previous live component.
Case 1: Edge to finished vertex
Case 1: Edge to finished vertex

Do nothing.
Case 2: Edge to unexplored vertex
Case 2: Edge to unexplored vertex
Case 3: Edge to live vertex
Case 3: Edge to live vertex

- Finished
- Live
- Unexplored
Representation of Live Components

Two stacks:

- **Vertex stack $S$:** Contains vertices in order of discovery, numbered in order of discovery
- **Component stack $C$:** Contains one entry per component, the number of the first vertex in this component
A Strong Connectivity Algorithm (2)

**Label-Components-from-Vertex**\((G, v, c, S, C)\)

1. \(c \leftarrow c + 1\)
2. Label \(v\) as live
3. \(\text{number}(v) \leftarrow c\)
4. Push\((S, v)\)
5. Push\((C, c)\)
6. for every out-edge \((v, w)\) of \(v\)
   7. if \(w\) is unexplored
   8. then **Label-Components-from-Vertex**\((G, w, c, S, C')\)
   9. else if \(w\) is live
      10. then repeat \(c' \leftarrow \text{Pop}(C)\),
          until \(c' \leq \text{number}(w)\)
      11. Push\((C, c')\)
      12. \(c' \leftarrow \text{Pop}(C)\)
   13. if \(\text{number}(v) = c'\)
      14. then repeat \(w \leftarrow \text{Pop}(S)\),
          Mark \(w\) as finished,
          number\((w) \leftarrow c'\)
      15. until \(w = v\)
   16. else Push\((C, c')\).
A Strong Connectivity Algorithm (3)

Lemma: The strongly connected components of a directed graph $G$ can be computed in $\mathcal{O}(n + m)$ time.
Summary

**Graphs are fundamental in computer science:**

- Many problems are quite natural to express as graph problems
  - Matching problems
  - Scheduling problems
  - . . .
- Data structures are graphs whose nodes store useful information

**Graph exploration lets us learn the structure of a graph:**

- Connectivity properties
- Distances between vertices
- Planarity
- . . .