Announcements (3/4/09)

- Project 3 code is due **tonight**.
- Reading for this lecture: Chapter 9.

CSE 326: Data Structures
Graphs

Steve Seitz
Winter 2009

Graphs

- A formalism for representing relationships between objects

Graph $G = (V, E)$

- Set of vertices:
  $V = \{v_1, v_2, \ldots, v_n\}$

- Set of edges:
  $E = \{e_1, e_2, \ldots, e_m\}$
  where each $e_i$ connects one vertex to another $(v_j, v_k)$

For directed edges, $(v_j, v_k)$ and $(v_k, v_j)$ are distinct. (More on this later...)

Graphs

Notation

$|V| = \text{number of vertices}$

$|E| = \text{number of edges}$

- $v$ is adjacent to $u$ if $(u, v) \in E$
  - *neighbor* of $u$ adjacent to
  - Order matters for directed edges
- It is possible to have an edge $(v, v)$, called a *loop*.
  - We will assume graphs without loops.
Examples of Graphs

- The web
  - Vertices are webpages
  - Each edge is a link from one page to another
- Call graph of a program
  - Vertices are subroutines
  - Edges are calls and returns
- Task graph for work flow
  - Vertices are tasks
  - Edge from $u$ to $v$, if $u$ must be completed before $u$ begins
- Social networks
  - Vertices are people
  - Edges connect friends

Directed Graphs

In directed graphs (a.k.a., digraphs), edges have a specific direction:

Thus, $(u, v) \in E$ does not imply $(v, u) \in E$.
I.e., $v$ adjacent to $u$ does not imply $u$ adjacent to $v$.

In-degree of a vertex: number of inbound edges.
Out-degree of a vertex: number of outbound edges.

Undirected Graphs

In undirected graphs, edges have no specific direction (edges are always two-way):

Thus, $(u, v) \in E$ does imply $(v, u) \in E$. Only one of these edges needs to be in the set; the other is implicit.

Degree of a vertex: number of edges containing that vertex. (Same as number of adjacent vertices.)

Weighted Graphs

Each edge has an associated weight or cost.

Clinton → Mukilteo 20
Kingston → Edmonds 30
Bainbridge → Seattle 35
Bremerton → Seattle 60
**Paths and Cycles**

- A path is a list of vertices \( \{w_1, w_2, ..., w_q\} \) such that \( (w_i, w_{i+1}) \in E \) for all \( 1 \leq i < q \).
- A cycle is a path that begins and ends at the same node.

\[
P = \{\text{Seattle, Salt Lake City, Chicago, Dallas, San Francisco, Seattle}\}
\]

**Path Length and Cost**

- Path length: the number of edges in the path
- Path cost: the sum of the costs of each edge

For path \( P \):
- \( \text{length}(P) = 5 \)
- \( \text{cost}(P) = 11.5 \)

Set \( \text{cost}(e) = 1 \)

How would you ensure that \( \text{length}(p) = \text{cost}(p) \) for all \( p \)?

---

**Simple Paths and Cycles**

A simple path repeats no vertices (except that the first can also be the last):
- \( P = \{\text{Seattle, Salt Lake City, San Francisco, Dallas}\} \)
- \( P = \{\text{Seattle, Salt Lake City, Dallas, San Francisco, Seattle}\} \)

A cycle is a path that starts and ends at the same node:
- \( P = \{\text{Seattle, Salt Lake City, Dallas, San Francisco, Seattle}\} \)
- \( P = \{\text{Seattle, Salt Lake City, Seattle, San Francisco, Seattle}\} \)

A simple cycle is a cycle that is also a simple path (in undirected graphs, no edge can be repeated).

For undirected "cycle" means "simple cycle".
Undirected Graph Connectivity

Undirected graphs are *connected* if there is a path between any two vertices:

![Connected graph](image1)

![Disconnected graph](image2)

A **complete undirected** graph has an edge between every pair of vertices:

![Complete undirected graph](image3)

(Complete = *fully connected.*)

Directed Graph Connectivity

Directed graphs are *strongly connected* if there is a path from any one vertex to any other.

![Directed graph](image4)

Directed graphs are *weakly connected* if there is a path between any two vertices, ignoring direction.

![Weakly connected directed graph](image5)

A **complete directed** graph has a directed edge between every pair of vertices. (Again, complete = *fully connected.*)

Trees as Graphs

A tree is a graph that is:
- **undirected**
- **acyclic**
- **connected**

Hey, that doesn’t look like a tree!

![Tree diagram](image6)

Rooted Trees

We are more accustomed to:
- Rooted trees (a tree node that is “special”)
- Directed edges from parents to children (parent closer to root).

![Rooted tree](image7)

A rooted tree (root indicated in red) drawn two ways

Rooted tree with directed edges from parents to children.

Characteristics of this one?
Directed Acyclic Graphs (DAGs)

DAGs are directed graphs with no (directed) cycles.

Aside: If program call-graph is a DAG, then all procedure calls can be inlined

\[
\{\text{rooted trees}\} \subset \{\text{DAGs}\} \subset \{\text{graphs}\}
\]

|E| and |V|

How many edges |E| in a graph with |V| vertices?

\[
0 \leq |E| \leq \binom{|V|}{2} + |V| = O(|V|^2)
\]

What if the graph is directed?

\[0 \leq |E| \leq |V|^2\]

What if it is undirected and connected?

\[|V| - 1 \leq |E| \leq \frac{1}{2} |V| (|V| - 1) = O(|V|^2)\]

Can the following bounds be simplified?

- Arbitrary graph: \(O(|E| + |V|)\)
- Arbitrary graph: \(O(|E| + |V|^2)\)
- Undirected, connected: \(O(|E| \log |V| + |V| \log |V|)\)

\[O(|E| \log |V|)\]

Some (semi-standard) terminology:

- A graph is **sparse** if it has \(O(|V|)\) edges (upper bound).
- A graph is **dense** if it has \(O(|V|^2)\) edges.

What’s the data structure?

- Common query: which edges are adjacent to a vertex

\[
O(|V|^2)
\]

List of edges \(O(|V|^3)\)

- Extract \(v \rightarrow u\) \(\Rightarrow \ \ O(1)\)

Representation 2: Adjacency List

A list (array) of length \(|V|\) in which each entry stores a list (linked list) of all adjacent vertices

\[
A \rightarrow B
\]

\[
B \rightarrow A
\]

\[
C \rightarrow B, D
\]

\[
D
\]

Runtimes:

- Iterate over vertices? \(O(|V|)\)
- Iterate over edges? \(O(|E|)\)
- Iterate edges adj. to vertex? \(O(1)\)
- Existence of edge? \(O(1)\)

Space requirements: \(|E| + |V|\)

Best for what kinds of graphs?
**Representation 1: Adjacency Matrix**

A $|V| \times |V|$ matrix $M$ in which an element $M[u, v]$ is true if and only if there is an edge from $u$ to $v$.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Runtimes:
- Iterate over vertices? $O(|V|)$
- Iterate over edges? $O(|V|^2)$
- Iterate edges adj. to vertex? $O(|V|)$
- Existence of edge? $O(1)$

Space requirements? $|V|^2$  
Best for what kinds of graphs? Dense

**Representing Undirected Graphs**

What do these representations look like for an undirected graph?

**Adjacency matrix:**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

**Adjacency list:**

- A: B
- B: A, C
- C: A, B, D
- D: C

**Some Applications:**

**Moving Around Washington**

What’s the *shortest* way to get from Seattle to Pullman?

Edge labels:  

**Moving Around Washington**

What’s the *fastest* way to get from Seattle to Pullman?

Edge labels:  

*distance*
Application: Topological Sort

Given a graph, \( G = (V, E) \), output all the vertices in \( V \) sorted so that no vertex is output before any other vertex with an edge to it.

**DAG**

What kind of input graph is allowed? **No**

Is the output unique? **No**

Topological Sort: Take One

1. Label each vertex with its in-degree (# of inbound edges)
2. **While** there are vertices remaining:
   a. Choose a vertex \( v \) of in-degree zero; output \( v \)
   b. Reduce the in-degree of all vertices adjacent to \( v \)
   c. Remove \( v \) from the list of vertices

**Runtime:**
**Topological Sort: Take Two**

1. Label each vertex with its in-degree
2. Initialize a queue $Q$ to contain all in-degree zero vertices
3. While $Q$ not empty
   a. $v = Q$.dequeue; output $v$
   b. Reduce the in-degree of all vertices adjacent to $v$
   c. If new in-degree of any such vertex $u$ is zero $Q$.enqueue($u$)

**Runtime:** $O(|V| + |E|)$

```cpp
void Graph::topsort(){
    Vertex v, w;

    labelEachVertexWithItsInDegree(); $O(|V|)$

    for (int counter=0; counter < NUM_VERTICES; counter++){
        v = findNewVertexOfDegreeZero(); $O(|V|)$
        v.topologicalNum = counter;
        for each w adjacent to v $O(|E|)$
            w.indegree--;
    }

    Total $O(|V|^2 + |E|)$
}
```

```cpp
void Graph::topsort(){
    Queue q(NUM_VERTICES);
    int counter = 0;
    Vertex v, w;
    labelEachVertexWithItsIn-degree();
    q.makeEmpty();
    for each vertex v
        if (v.indegree == 0)
            q.enqueue(v);

    while (!q.isEmpty()){
        v = q.dequeue();
        v.topologicalNum = ++counter;
        for each w adjacent to v
            if (--w.indegree == 0)
                q.enqueue(w);}
```