We’ll talk about...

- A Different kind of Network
- The World Wide Web
  - Hypertext
- Information Networks, Hypertext, and Associative Memory
  - Vannevar Bush and the Memex
  - The Web and its Evolution
- The Web as a Directed Graph
  - Paths and Strong Connectivity
  - Strongly Connected Components
- The Bow-Tie Structure of the Web
  - A Giant Strongly Connected Component
  - The Bow-Tie Structure
- The Emergence of Web 2.0

The World Wide Web

- A computer application
- Developed by Tim Berners-Lee between 1989-1991

A Different Kind of Network

<table>
<thead>
<tr>
<th>Feature</th>
<th>Social</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>People</td>
<td>Pieces of information</td>
</tr>
<tr>
<td>Edges</td>
<td>Friendship/relationship</td>
<td>Links</td>
</tr>
<tr>
<td>Directed</td>
<td>Usually not</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Hypertext

- Key feature:
  - Allow the published document to hold virtual links to other published documents
  - This transforms the set of documents into a directed graph structure, in which every node is a Web page and the directed edges are the links
  - The idea to organize Web pages as a network was non-obvious [alphabetically, library]

The World Wide Web

Features:
1. Lets users make documents available to anyone on the internet
2. Gives users access to such Web pages using a browser to browse through the public spaces of the internet [example]
The World Wide Web

Hypertext

- Key feature:
  - Use of this structure is allowing anyone publishing a Web page to highlight a relationship with any other existing page.
  - Hypertext was explored and refined in the middle of the last century
  - The main idea is to link logical parts of a text directly
  - The web is an Hypertext system at a scale no one could have anticipated

Information Networks, Hypertext, and Associative Memory

- Intellectual Precursors of Hypertext
  - citation among scholarly books and articles
  - cross-references within a printed encyclopedia

Vannevar Bush and the Memex

- Published an article in 1945 in the Atlantic Monthly, entitled “As We May Think”
- Methods for storing information in a book, a library, or a computer memory are highly linear — consist of a collection of items sorted in some sequential order
- Our conscious experience of thinking exhibits what might be called an associative memory
The Web and its Evolution

- In its first decade the Web was mostly used for static pages with mainly navigational links.
- Computers that held these pages were mainly used to serve them to the user.
- As the Web and the computational power developed, links now trigger complex programs on the hosting computer.
- “Buy Now” button may navigate the browser to the receipt’s page but it also triggers a transaction that bills the buyer’s credit card and the delivery of the product.
- These are transactional links.

Vannevar Bush and the Memex

- Memex – hypothetical machine that mimics the way our mind works and includes all of the human knowledge.
- Its scope and the usages he described are remarkably similar to that of today’s Web.
- This is not a coincidental fact as he had a prominent role in US governments’ scientific funding.
- The creators of hypertext and later Tim Berners-Lee explicitly invoked Bush’s ideas.

The Web and its Evolution

- Although many of the links in the Web are both transactional and navigational it is useful for Web analysis purposes to classify them as one or the other.
- While a lot of content on the Web now has a primarily transactional nature, it still remains largely linked by a navigational “backbone” — it is reachable via relatively stable Web pages connected to each other by more traditional navigational links.
- Distinguishing them is essential to Web search engines, when they build their indexes of the available content on the Web.

The World Wide Web

- Although many of the links in the Web are both transactional and navigational it is useful for Web analysis purposes to classify them as one or the other.
- While a lot of content on the Web now has a primarily transactional nature, it still remains largely linked by a navigational “backbone” — it is reachable via relatively stable Web pages connected to each other by more traditional navigational links.
- Distinguishing them is essential to Web search engines, when they build their indexes of the available content on the Web.

Paths and Strong Connectivity

<table>
<thead>
<tr>
<th>Path</th>
<th>Undirected</th>
<th>Directed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two nodes are linked by a path if we can follow a sequence of edges from one to the other.</td>
<td>Sequence of nodes, beginning with A and ending with B, with the property that each consecutive pair of nodes in the sequence is connected by an edge pointing in the forward direction.</td>
<td></td>
</tr>
<tr>
<td>A graph is connected if every pair of nodes is linked by a path.</td>
<td>Strongly connected if there is a path from every node to every other node.</td>
<td></td>
</tr>
</tbody>
</table>

The Web as a Directed Graph

- Like social and economic networks, visualizing the information networks as a graph helps us better understand the structure of the network.
- In our analysis edges are navigational links.
- Unlike the previous networks, information networks imply a directed graph.
Strongly Connected Components

<table>
<thead>
<tr>
<th>Undirected</th>
<th>Directed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connected component</td>
<td>If two nodes belong to the same component, then they can reach each other by paths; and if two nodes belong to different components then they can't.</td>
</tr>
<tr>
<td>Component connectivity</td>
<td>Connected components have no edges between them; SCC's might have edges between them.</td>
</tr>
</tbody>
</table>

The Bow-Tie Structure of the Web

- In 1999 Andrei Broder and his colleagues tried to create a map of the Web.
- They used data from the largest commercial Web search engine at the time, AltaVista.
- Although the numbers are now outdated, the structure has persisted.

A Giant Strongly Connected Component

- In the center of the Web there is a giant SCC.
- The fact is not hard to believe if you consider the main search engines and other “starting points” have links to directory type sites which link to all of the major educational institutes, large companies and government agencies.
- Further, many pages within those sites contain links back to the search engines and starting points.
- This amounts to a one giant SCC.

The Bow-Tie Structure of the Web
**The Bow-Tie Structure**

- **IN**: nodes that can reach the giant SCC but cannot be reached from it
- **OUT**: nodes that can be reached from the giant SCC but cannot reach it
- **Tendrils**: (a) the nodes reachable from IN that cannot reach the giant SCC, and (b) the nodes that can reach OUT but cannot be reached from the giant SCC. Tubes are nodes that satisfy both (a) and (b)
- **Disconnected**: the nodes that would not have a path to the giant SCC even if we completely ignored the directions of the edges

**The Bow-Tie Structure**

- The bow tie structure resembles a river in which pages go from the IN (upstream) into the core that is the giant SCC an then maybe OUT (downstream)
- The details of the structure change continuously but the overall structure remains the same

**The Emergence of Web 2.0**

- Tim O’Reilly and others began speaking in 2004 and 2005 about the emergence of Web 2.0
- An attitude, not a technology
- No concrete definition of the term
- But, a major next step in the evolution of the Web triggered by the above forces and not arising from a single organization’s decisions

**The Emergence of Web 2.0**

1. The growth of Web authoring styles that enabled many people to collectively create and maintain shared content
2. The movement of people’s personal on-line data (including e-mail, calendars, photos, and videos) from their own computers to services offered and hosted by large companies
3. The growth of linking styles that emphasize on-line connections between people, not just between documents
We talked about...

- A Different kind of Network
- The World Wide Web
- Hypertext
- Information Networks, Hypertext, and Associative Memory
  - Vannevar Bush and the Memex
  - The Web and its Evolution
- The Web as a Directed Graph
  - Paths and Strong Connectivity
  - Strongly Connected Components
- The Bow-Tie Structure of the Web
  - A Giant Strongly Connected Component
  - The Bow-Tie Structure
- The Emergence of Web 2.0

The Emergence of Web 2.0

- Some sites that feature one or more of the 3 principals between 2004-2006 are:
  - Wikipedia
  - Gmail
  - Myspace
  - Facebook
  - Flickr
  - YouTube
  - Twitter

Exercises

1. Classify SCC’s in the graph according to the bow tie classification
2. What edge can we remove or add to increase the size of the:
   1. Giant SCC?
   2. IN set?
   3. OUT set?

Questions?
This is Hypertext

The End
THANKS – AND HAPPY NEW YEAR