CS 103: Lecture 11 Information Networks and the Web

Dan Sheldon

October 27, 2015

Announcements

- HW 4 due Thursday
  - Fill out poll about office hours (see Piazza)
- Midterm next Tuesday
  - HWs returned, solutions posted by end of week
  - Outline of topics on Thursday

Plan for today

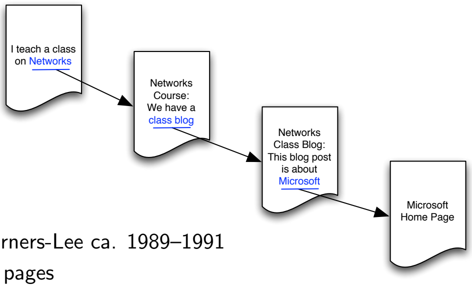
- History of the web
- Web structure
  - Directed graphs
  - Strongly-connected components
  - Bow-tie structure
- Web search

Information Networks

Information network

- nodes = pieces information
- links = connections between related pieces of information

World Wide Web


I teach a class on Networks
Networks Course: We have a class blog
Networks Class Blog: This blog post is about Microsoft
Microsoft Home Page

Tim Berners-Lee ca. 1989–1991

- Web pages
- Browser
- *Hypertext*

Seems obvious now, but Internet existed for ˜20 years *without* it.

Hypertext

Principle for organizing information. Vannevar Bush 1945 “As We May Think”

```
Our ineptitude in getting at the record is largely
    caused by the artificiality of systems of
    indexing. When data of any sort are placed in
    storage, they are filed alphabetically or
    numerically, and information is found (when it
    is) by tracing it down from subclass to
    subclass. It can be in only one place, unless
    duplicates are used; one has to have rules as
    to which path will locate it, and the rules are
    cumbersome. Having found one item, moreover,
    one has to emerge from the system and re-enter
    on a new path.
...
The human mind does not work that way. It operates
    by association
```

Hypertext

Vannevar Bush goes on to describe **memex** machine

- Mechanized information viewer
- “Associative trails”

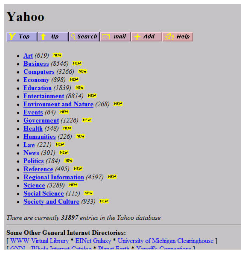
Idea refined throughout 20th century

- Substitute/enhancement for completely linear text
- First real implementation in WWW!

Hypertext inspired and non-obvious organizational principle

Organizing the Web

- Early efforts at *navigating* the web still based on directories and text-based search



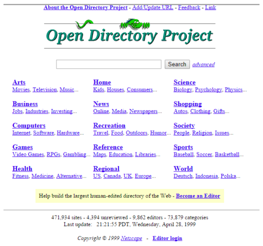
Yahoo

Top | Up | Search | mail | Add | Help

- Art (619) NEW
- Business (8546) NEW
- Computers (3266) NEW
- Economy (898) NEW
- Education (1839) NEW
- Entertainment (8814) NEW
- Environment and Nature (268) NEW
- Events (64) NEW
- Government (1226) NEW
- Health (548) NEW
- Humanities (226) NEW
- Law (221) NEW
- News (301) NEW
- Politics (184) NEW
- Reference (495) NEW
- Regional Information (4597) NEW
- Science (3289) NEW
- Social Science (115) NEW
- Society and Culture (933) NEW

*There are currently **31897** entries in the Yahoo database*

Some Other General Internet Directories:
[WWW Virtual Library * EINet Galaxy * University of Michigan Clearinghouse]
[GNN - Whole Internet Catalog * Planet Earth * Yanoff's Connections]



About the Open Directory Project - Add/Update URL - Feedback - Link

Open Directory Project

Search advanced

Arts Movies, Television, Music...
Home Kids, Houses, Consumers...
Science Biology, Psychology, Physics...
Business Jobs, Industries, Investing...
News Online, Media, Newspapers...
Shopping Autos, Clothing, Gifts...
Computers Internet, Software, Hardware...
Recreation Travel, Food, Outdoors, Humor...
Society People, Religion, Issues...
Games Video Games, RPGs, Gambling...
Reference Maps, Education, Libraries...
Sports Baseball, Soccer, Basketball...
Health Fitness, Medicine, Alternative...
Regional US, Canada, UK, Europe...
World Deutsch, Indonesia, Polska...

Help build the largest human-edited directory of the Web - **Become an Editor**

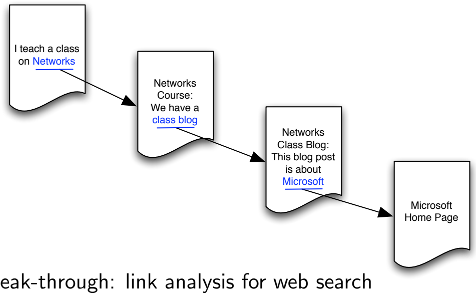
471,934 sites - 4,394 unreviewed - 9,862 editors - 73,879 categories
Last update: 21:21:55 PDT, Wednesday, April 28, 1999

Copyright © 1999 Netscape - **Editor login**

Yahoo!
1995 / DMOZ 1999

Why didn’t these work? What made the web really work?

Web as Directed Graph


I teach a class on Networks
Networks Course: We have a class blog
Networks Class Blog: This blog post is about Microsoft
Microsoft Home Page

Real break-through: link analysis for web search

- Google
- Use links to determine importance of pages

Paths and Strong Connectivity

Examples on board

What are appropriate notions of connectivity for a directed graph?

Directed path: sequence of nodes in which each consecutive pair is connected by an edge *in the forward direction*

Strongly connected component (SCC): a set of nodes that

- contains a directed path between each pair in the set
- is not part of a larger set with this property

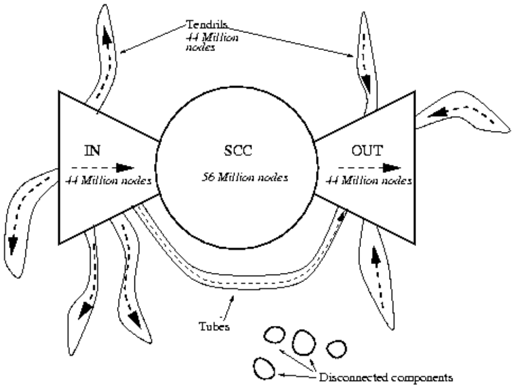
Thought Experiment

What do the strongly connected components of the web look like?

- How big is the SCC containing the MHC home page?
- Is it the biggest?
- How big are the other ones?
- How do they connect to the MHC SCC?

Discuss with partner, then as a class.

Bow-Tie Structure of the Web



Broder et al. 1999