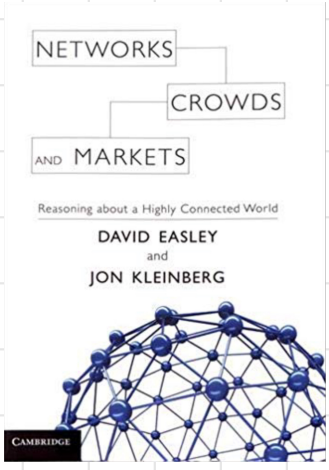
Lecture 11

Network Science

The Structure of the Web

Today's Topics

- the World Wide Web
- Information Nets, Hypertext, and Associative Memory
- the Web as a Directed Graph
- the Bow-tie structure of the Web
- the Emergence of Web 2.0



Chapter 13
The Structure of the Web

Information Network

units: are pieces of information

links: relationships between such units.

an example: www

citation networks

To reduce the web in terms of measures we studied so far

short paths, giant component, ...

- which is the informative power of the underlying graph structure?

the World Wide Web

created by Tim Berners Lee
1989 - 1991

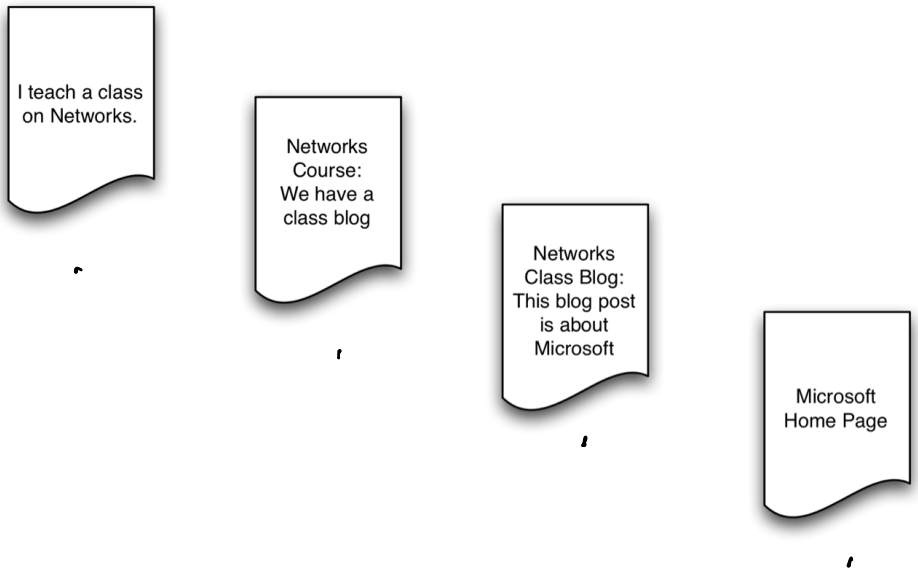
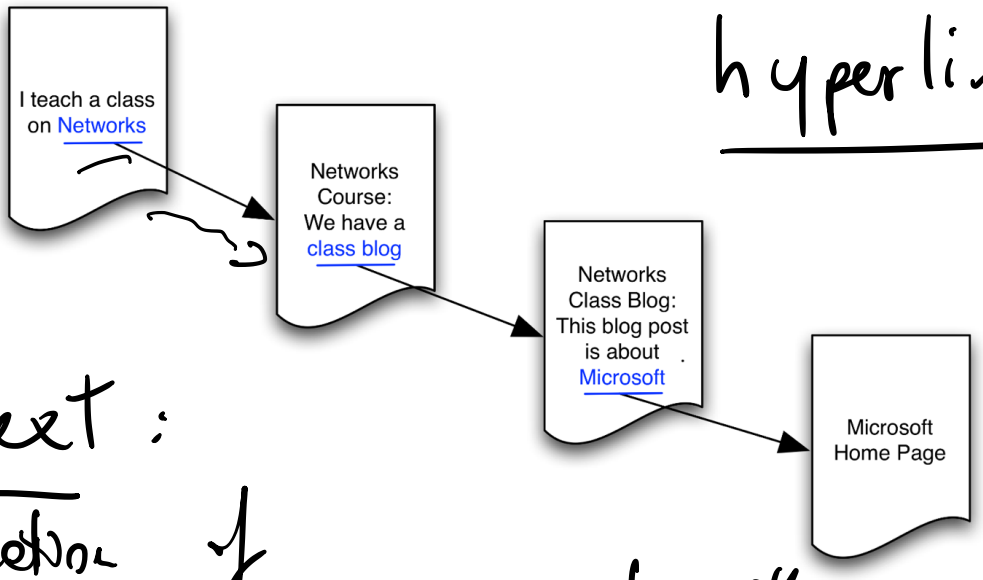


Figure 13.1: A set of four Web pages.

A way to exchange documents via the Internet

- Web page: public
- browser: connects to the Internet



hyperlinks

hypertext: organization of information using a network metaphor

Figure 13.2: Information on the Web is organized using a network metaphor: The links among Web pages turn the Web into a directed graph.

Networks of citations

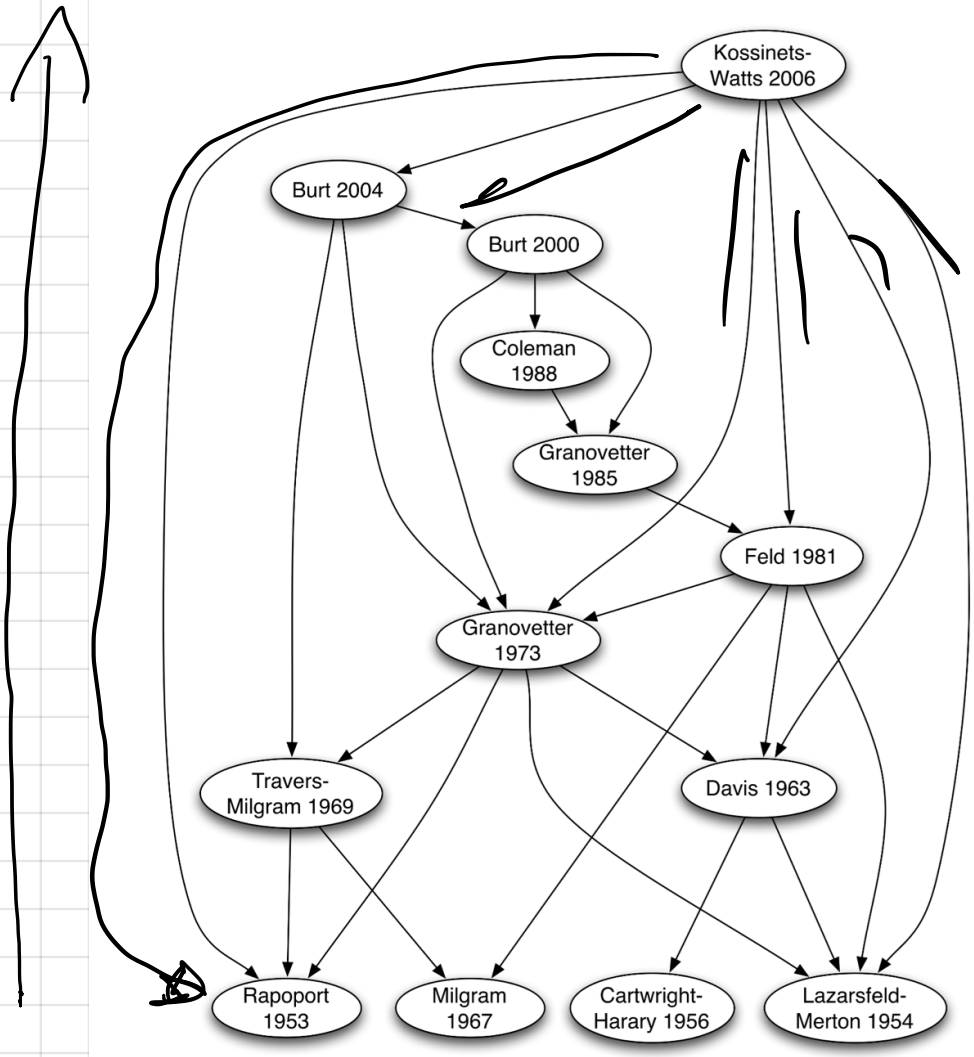


Figure 13.3: The network of citations among a set of research papers forms a directed graph that, like the Web, is a kind of information network. In contrast to the Web, however, the passage of time is much more evident in citation networks, since their links tend to point strictly backward in time.

there is an "arrow of time" citation network: returns a sense of flow from present to past

Wikipedia Pages

T.B.L. was influenced by the pioneering work of Vannevar BUSH

We seen information with e not linear procen

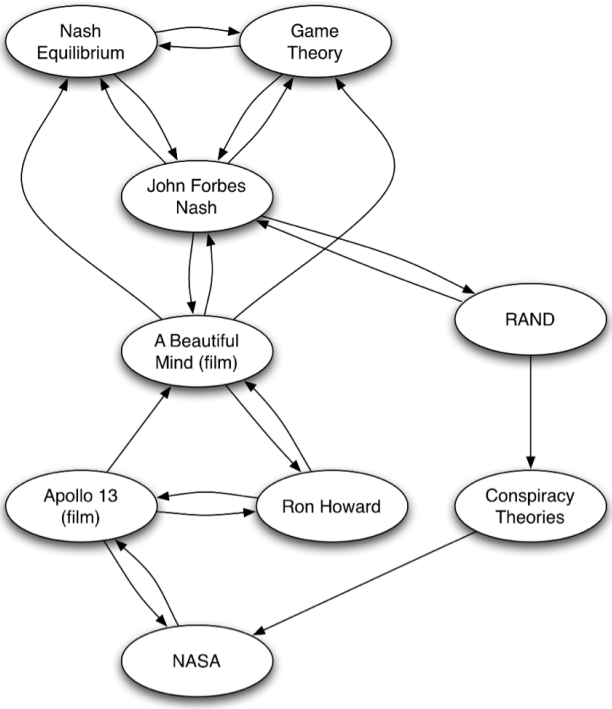


Figure 13.4: The cross-references among a set of articles in an encyclopedia form another kind of information network that can be represented as a directed graph. The figure shows the cross-references among a set of Wikipedia articles on topics in game theory, and their connections to related topics including popular culture and government agencies.

Bush : "memex"
tech. device that implents the idea of associative memory.

Gigantic BRAIN

"As We May Think" paper (1945)

the Web and its Evolution

Navigational function

↓

To transport the user from one page to another (not linearly)

transactional function

↓

a link may trigger a server side script that computes input and creates on "the fly" a web page.

---

we focus on the old fashioned web that was mainly "navigational"

The Web as a directed Graph

Why ?

a) We can better understand the hyperlinked relationships expressed by its links

b) break its structure into smaller units

c) identify important pages as a step in organizing results of the web

No symmetric connections We have a "from" and a "to"

Paths and Strong Connectivity

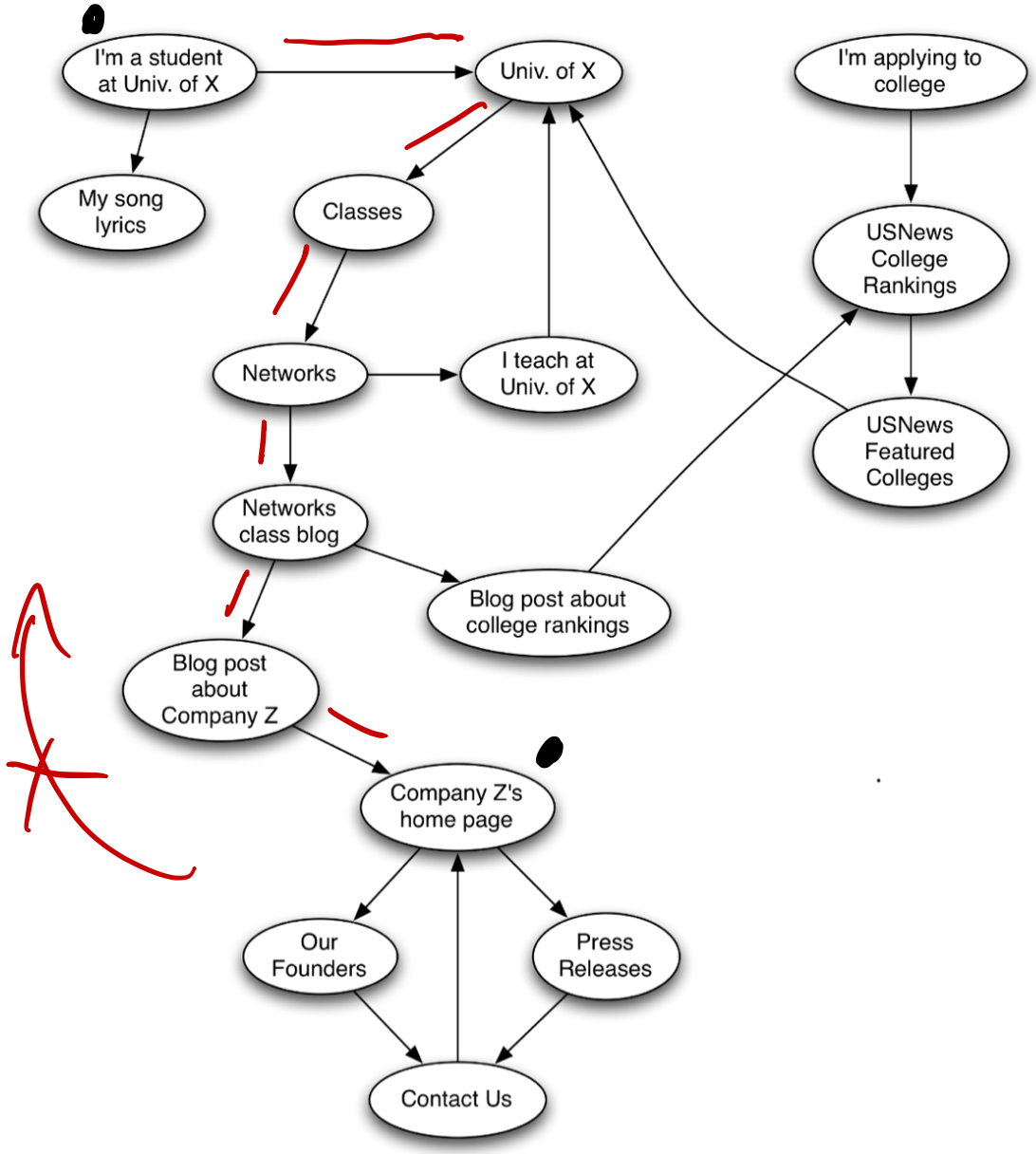


Figure 13.5: A directed graph formed by the links among a small set of Web pages.

We have connections

We have components

Strongly Connected Components (SCC)

Broder et al (1999)
they used AltaVista
A "map" of the Web
dividing the graph into few
large pieces

First step ⇒ they found the
Giant SCC
only one containing a
significant fraction
of all the nodes

they started from "starting
pages": reach large
components Web sites

("Largest Connected Component")

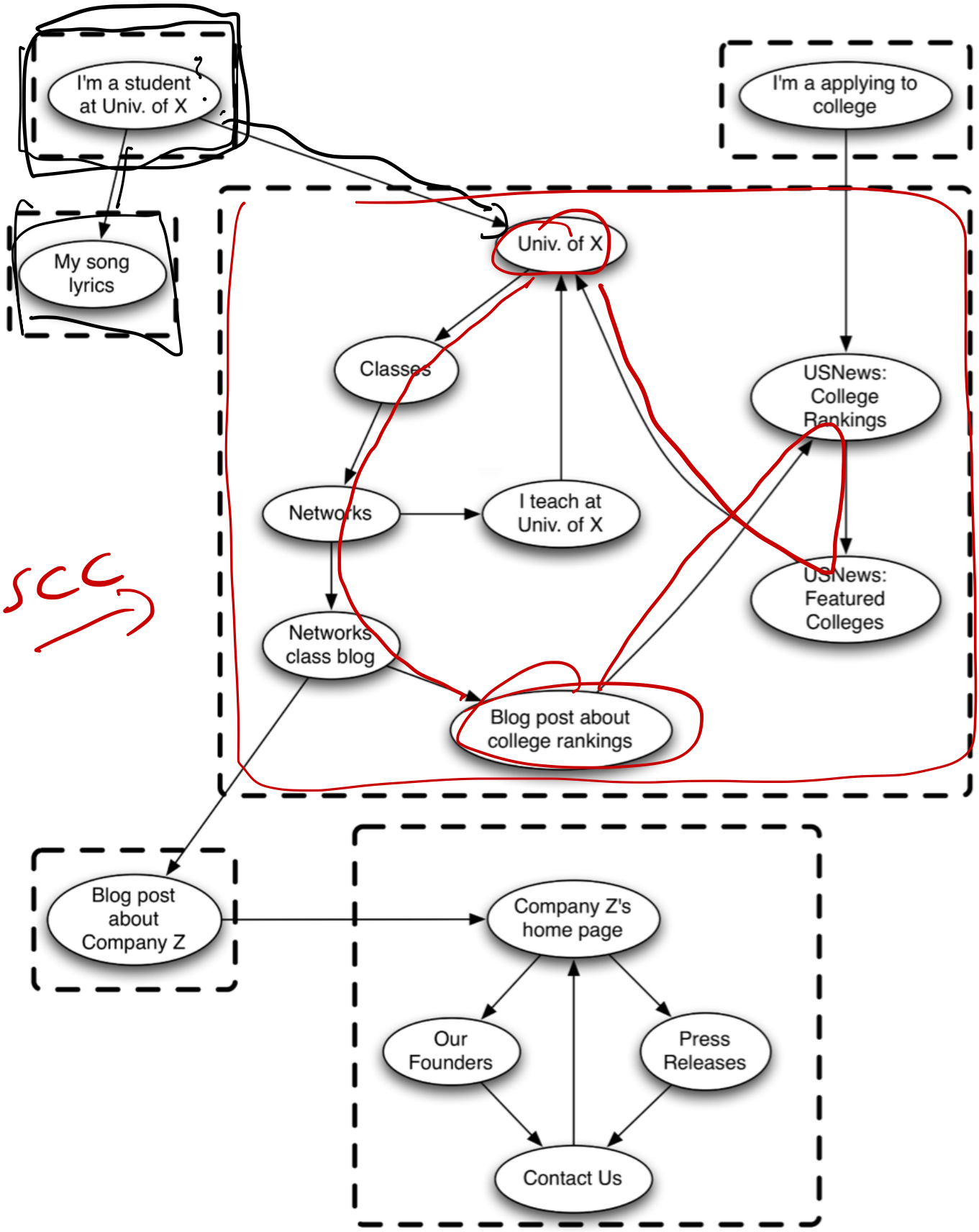


Figure 13.6: A directed graph with its strongly connected components identified.

Web: a GIANT SCC

reachability and paths

the largest subgraph s.t.
I can find paths leading
from a node to another
and back : SCC

Second step : other components

IN : from every node in
IN there is a path
leading to SCC

OUT : nodes in the SCC
are connected to nodes
in OUT, no way back

the bow-tie structure

tendrils: nodes reachable from IN reach nodes in the OUT

WCC

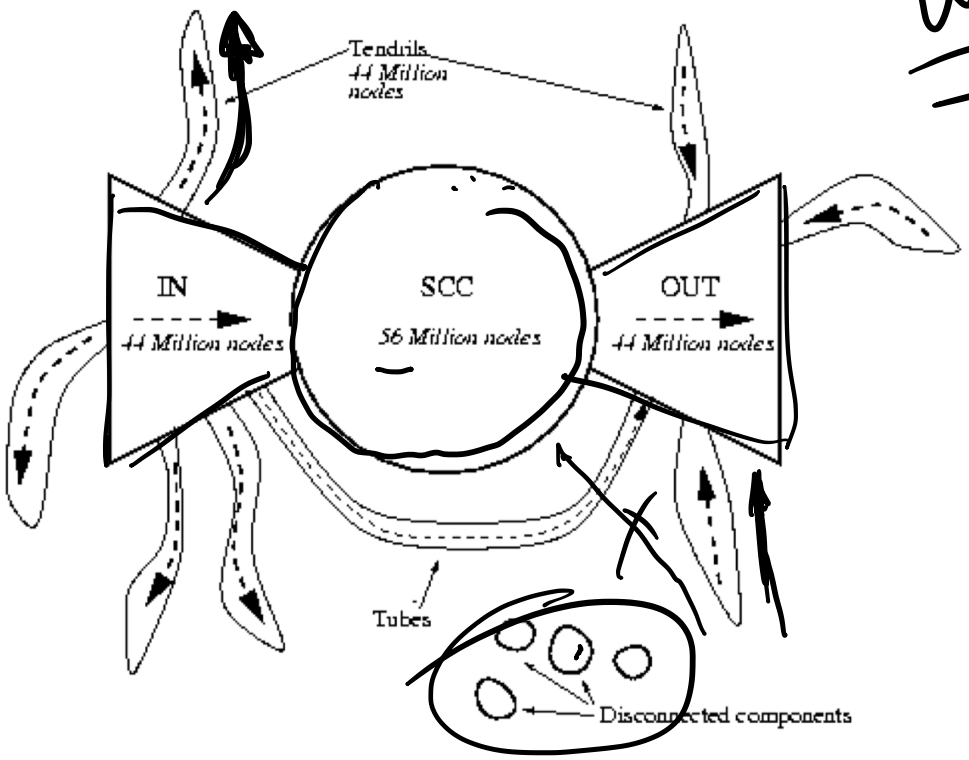
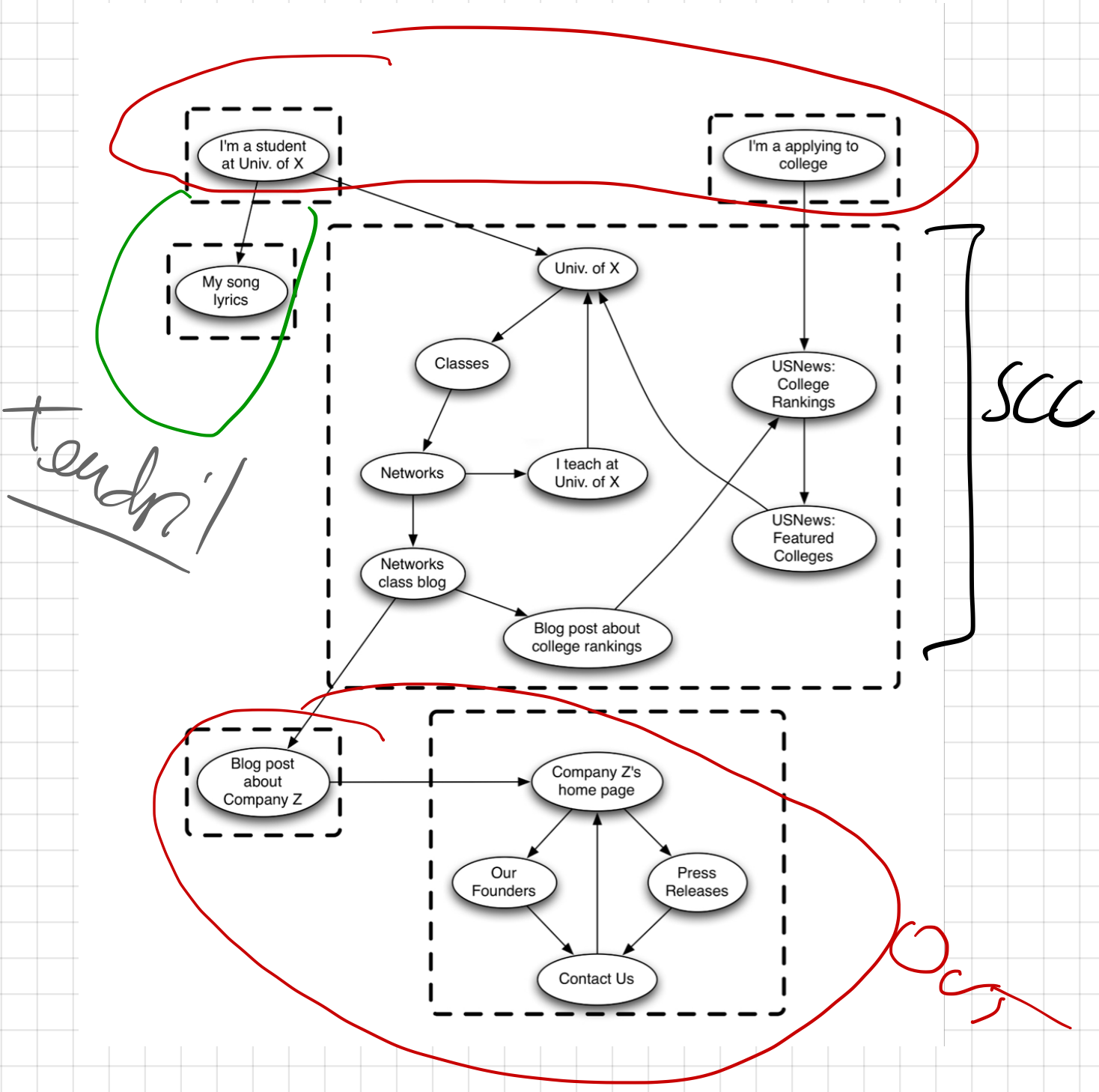


Figure 13.7: A schematic picture of the bow-structure of the Web (image from [80]). Although the numbers are now outdated, the structure has persisted.

Disconnected components

Weakly Connected Component:

IN + OUT + SCC + tendrils


I'm a student at Univ. of X
I'm a applying to college
My song lyrics
Univ. of X
Classes
USNews: College Rankings
Networks
I teach at Univ. of X
USNews: Featured Colleges
Networks class blog
Blog post about college rankings
Blog post about Company Z
Company Z's home page
Our Founders
Press Releases
Contact Us

The Emergence of Web 2.0

- from 2000 on
- "invented" by a group of technologists led by Tim O'Reilly

Major forces

1. Web authoring styles ⇒ more user generated content
2. personal online data moved from the personal computer to the Web (email, calendars, photos, videos, ...)
3. linking styles ⇒ not only between pages, but also between people

Examples: Wikipedia, facebook, Twitter, Gmail, ...

New ideas that can be studied with networks

- software that gets better the more people use them
- the wisdom of the crowds
- the long tail

Take Home Message

- Directed Graph Analysis helped to better understand the Web (and influenced the Web itself as we know it today)
- Networks as an abstraction to find Components (SCC) and to drive their evolution
  - social networks (online)
  - reputation systems
  - Recommendation Engines

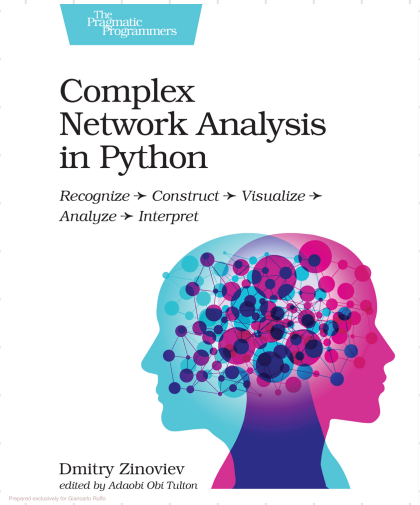
All of them are governed by laws that can be investigated by means of networks

Directed Networks

- Asymmetric Relationships
- Explore Directed Nets with Python
- Topological Sort

Notebooks: 16_ Directed Networks

17_ Adjectives



Chapter 17