Graph Primitives

Structure of the Web
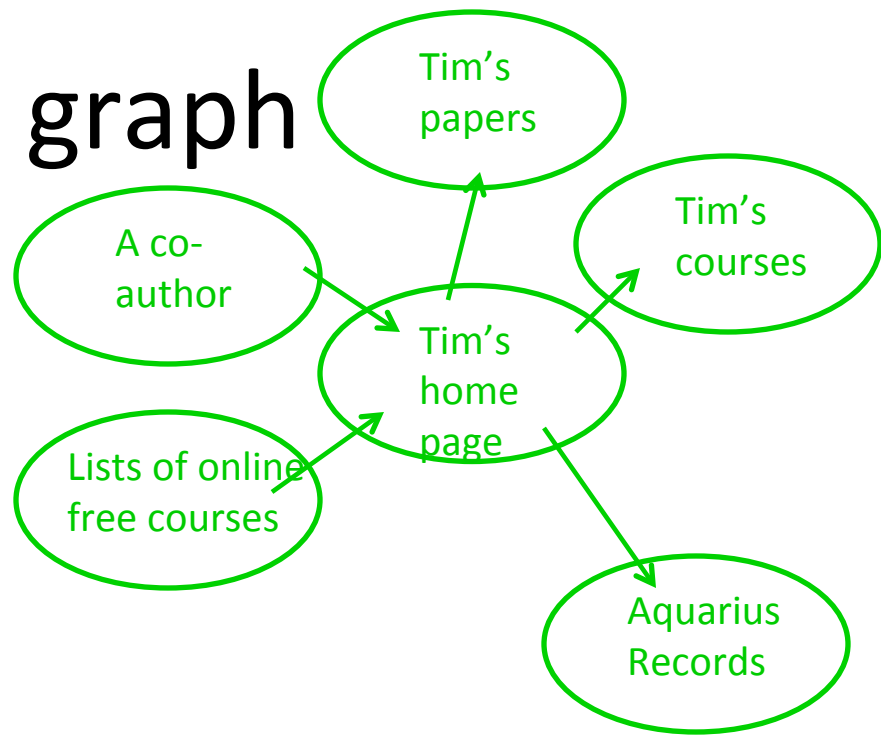
Design and Analysis of Algorithms I

The Web graph

-Vertices = Web pages
- (directed) edges = hyperlinks

Question : what does the web graph

look like ?

(assume you've already "crawled" it)

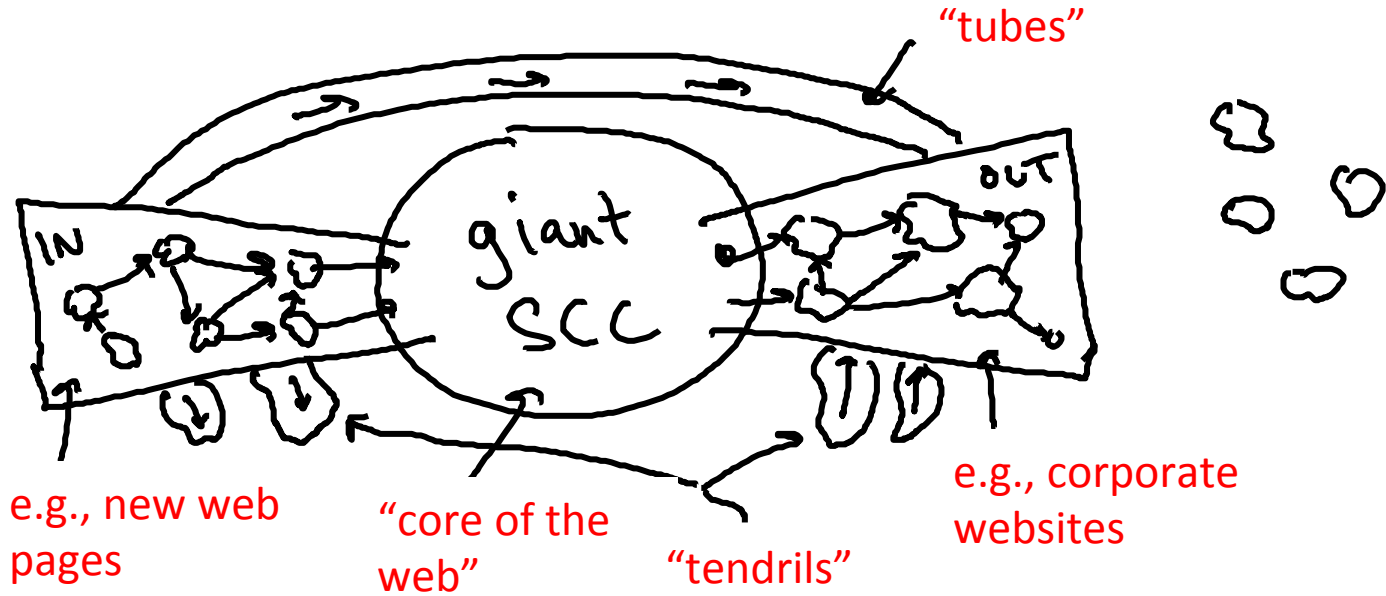Size : ~ 200 million nodes, ~ 1 billion edges

Reference : [Broder et al WWW 2000]
computed the SCCs of the Web graph.

Tim's papers

A co-author

Tim's courses

Tim's home page

Lists of online free courses

Aquarius Records

ETC.
ETC.

(pre map-reduce/hadoop)

Tim Roughgarden

The Bow Tie

"tubes"

IN

OUT

giant SCC

e.g., new web pages

"core of the web"

"tendrils"

e.g., corporate websites

Main Findings

1. All 4 parts (giant, IN, OUT, tubes + tendrils) have roughly the same size

2. Within CORE, very well connected (has the "small world" property) [Milgram]

3. Outside, surprisingly poorly connected

Tim Roughgarden

Modern Web Research

1. Temporal aspects – how is the web graph evolving over time ?

2. Informational aspects – how does new information propagate throughout the Web (or blogosphere, or Twitter, etc.)

3. Finer-grained structure – how to define and compute "communities" in information and social networks ?
   Recommended Reading : Easley + Kleinberg, "Networks, Crowds, & Markets"

Tim Roughgarden