# CPSC156: The Internet Co-Evolution of Technology and Society

Lecture 13: February 27, 2007

Review for First Exam

# In-class exam: Thurs., 3/1/07

- Test on first five weeks of CPSC156 (through 2/15/07)
- Lecture notes
- Homework assignments and solution sets
- Exams and HWs from earlier version of 156 and 155.
- Reading assignments

# Topics

- **Internet design**
- The web
- Information goods and information industries
- Basics of B2C and C2C e-commerce
- Copyright law
- Online music distribution
- Search

# Internet Design

- Lectures 1 and 2, first half of lecture 3
- First HW assignment
- Reading assignments from January 18, 2007 and January 23, 2007

# Internet Protocols Design Philosophy

- Ordered set of goals:
  1. multiplexed utilization of **existing networks**
  2. survivability in the face of failure
  3. support multiple types of communications service
  4. accommodate a variety of network types
  5. permit distributed management of resources
  6. cost effective
  7. low effort to attach a host
  8. account for resources
- Not all goals have been met

# Packets!

- Basic decision: use packets not circuits (Kleinrock)
- Packet (*a.k.a.* datagram)


  - self contained
  - handled independently of preceding or following packets
  - contains destination and source **internetwork** address
  - may contain processing hints (*e.g.*, QoS tag)
  - **no delivery guarantees**
    - net may drop, duplicate, or deliver out of order
    - reliability (where needed) done at higher levels

## Telephone Network

- Connection-based
- Admission control
- Intelligence is "in the network"
- Traffic carried by relatively few, "well-known" communications companies

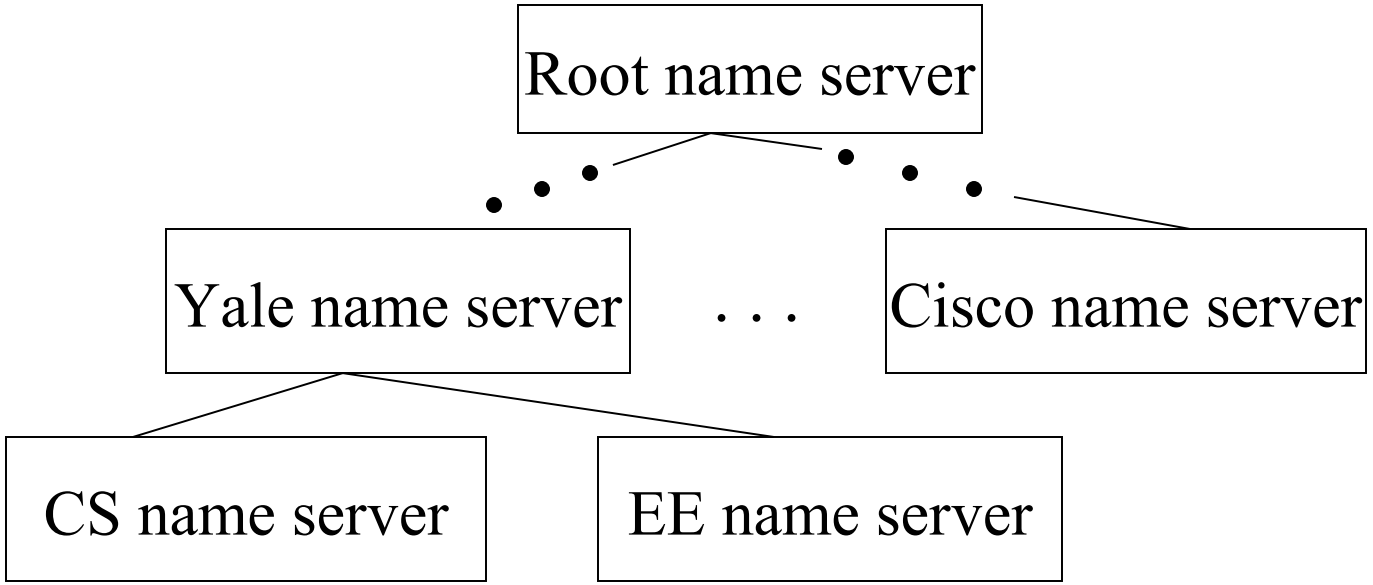
## Internet

- Packet-based
- Best effort
- Intelligence is "at the endpoints"
- Traffic carried by many routers, operated by a changing set of "unknown" parties

# IP Addresses and Host Names

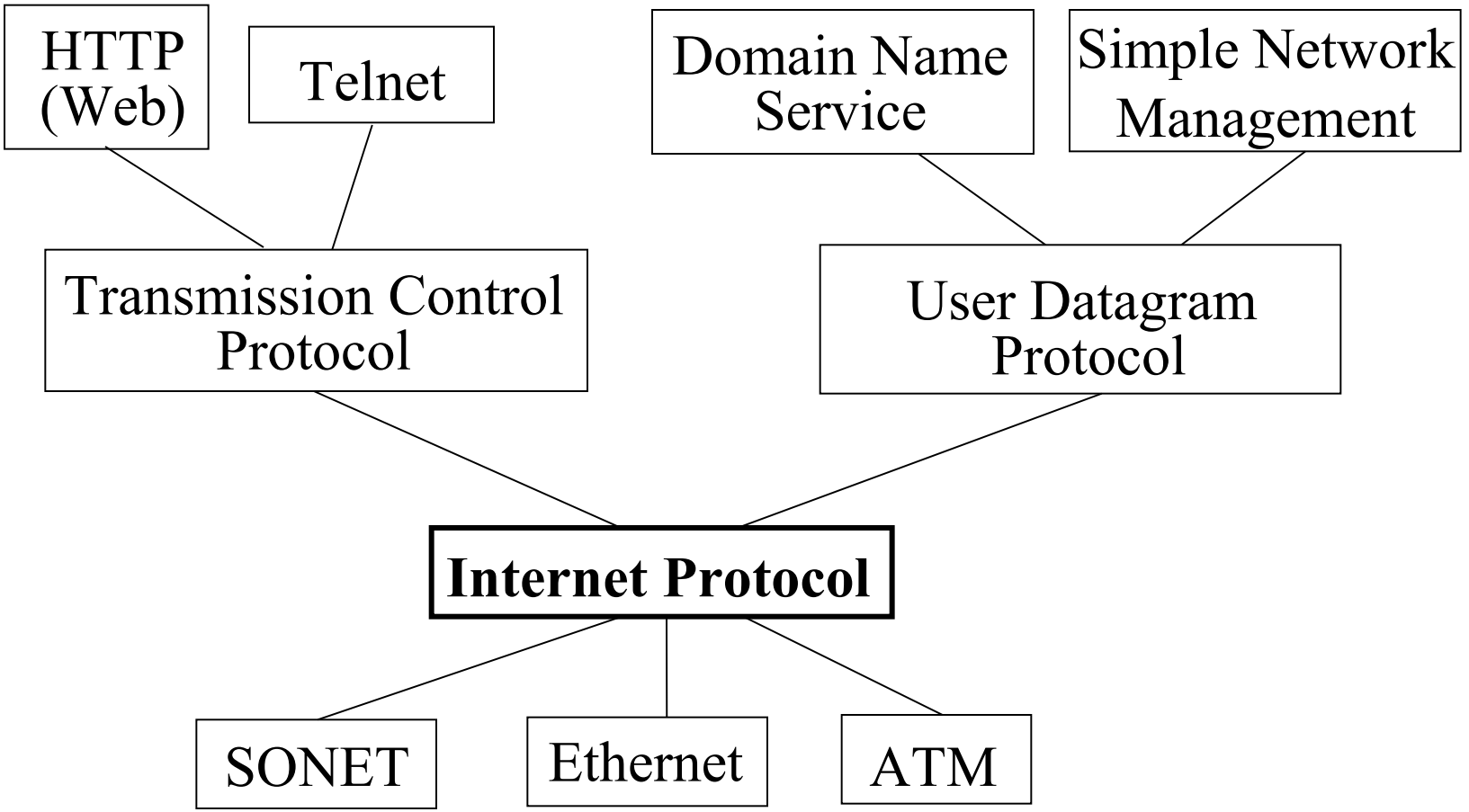
- Each machine is addressed by an integer, its **IP address**, written down in a "dot notation" for "ease" of reading, such as `128.36.229.231`
- IP addresses are the universal IDs that are used to name everything.
- For convenience, each host also has a human-friendly host name. For example, `128.36.229.231` was `concave.cs.yale.edu`.
- Question: How do you translate names into IP addresses?

# Hierarchy of Name Servers




- Clients send queries to name servers.
- Name servers reply with answers or forward requests to other name servers.
- Most name servers perform "lookup caching."

# Basic Architectural Principle: Layering



# Getting from A to B: Summary

- Need IP addresses for:
  - Self (to use as source address)
  - DNS Server (to map names to addresses)
  - Default router to reach other hosts (e.g., gateway)
- Use DNS to get destination address
- Pass message through TCP/IP handler
- Send it off! **Routers** will do the work:
  - Physically connecting different networks
  - Deciding where to next send packets

# Topics

- Internet design
- **The web**
- Information goods and information industries
- Basics of B2C and C2C e-commerce
- Copyright law
- Online music distribution
- Search

# The Web

- Second half of lecture 3 (Jan. 23, 2007), all of lecture 6 (Feb. 1, 2007), beginning of Nov. 6, 2001 lecture of CPSC155
- Second HW assignment
- Reading assignment from Feb. 1, 2007

# HTTP (Hypertext Transfer Protocol)

- Standard protocol for web transfer
- "Request-response" interaction between clients and servers
- Request methods: GET, HEAD, PUT, POST, DELETE,…
- Response: Status line + additional info (*e.g.*, a web page)

Example of a request line:

```
<form action="http://lab.zoo.cs.yale.edu/cs156/cgi-
  bin/sendform.cgi" method="post">
```

# HTML (Hypertext Markup Language)

- Language in which web pages are written
- Contains formatting commands
- Tells browser what to display and how to display

```
<TITLE> Welcome to Yale </TITLE>
```

- The title of this page is "Welcome to Yale"

```
<B> Great News! </B>
```

- Set "**Great News!**" in boldface

```
<A HREF="http://www.cs.yale.edu/index.html">
Yale Computer Science Department </A>
```

- A link pointing to the web page http://www.cs.yale.edu/index.html with the text "Yale Computer Science Department" displayed.

# What does "http://www.cs.yale.edu/index.html" mean?

| Protocol | Host, Domain Name | Local File |
|---|---|---|
| http | www.cs.yale.edu | index.html |

# Topics

- Internet design
- The web
- **Information goods and information industries**
- **Basics of B2C and C2C e-commerce**
- Copyright law
- Online music distribution
- Search

# Information Goods and Information Industries, B2C, C2C

- Lectures 4 and 5
- Reading assignment from Jan. 30, 2007

# E-Commerce, cont.

- Information is anything that can be digitized, *i.e.*, encoded as bits. Examples include books, magazines, movies, music, web pages, software, and databases.
- Information industries are those that produce information goods and/or deliver information services.
- Networked industries are those that rely on customers' interaction. Networks can be real (as in the telecomm industry) or virtual (as in the PC-software industry).

# Existing Business Models for Information Products

- **Fee models**: Subscription purchase, Single-transaction purchase, Single-transaction license, Serial-transaction license, Site license, Payment per electronic use
- **Advertising models**: Combined subscription and advertising income, Advertising income only
- **"Free" distribution models**: Free distribution (no hidden motives), Free samples (*e.g.*, coming attractions), Free first version, Free information when you buy something else (complementary products, bundling)

# Less Traditional Business Models for Information Products

- **Extreme customization**: Make the product so personal that few people other than the purchaser would want it.
- Provide a **large product in small pieces**, making it easy to browse but difficult to get in its entirety.
- **Give away digital content** because it **complements** (and increases demand for) the traditional product.
- Give away the product, **sell the service contract.**
- Allow free distribution of the product but **request payment** (Shareware).
- Position the product for **low-priced, mass market distribution.**

# Network Effects

- A product or service exhibits network effects if its value to any single user is strongly positively correlated with the total number of users. Communication products and services are prime examples.
- Network-effected products and services exhibit long lead times followed by explosive growth. Example: Fax invented in 1843, offered by AT&T in 1925, and widely adopted in 1980s.
- "Network-effected" ≠ "mass-market"

* Network effects cut both ways!

# Lock-in and Switching Costs

- Information industries often involve systems of interoperating components and durable complementary assets. Prime examples are Intel processors, Windows PC Platform, and numerous PC application programs.
- Often leads to technology lock-in and high switching costs
- Modular architectures and open standards are mitigating forces.
- "Network effects" ≠ "Strong lock-in"
- "High market share" ≠ "High switching costs"

# Netscape Used Many "Information Business Models"

(esp. those that involve making money by "giving away" an information product)

Complementary products (esp. server code)

- Bundling
  - Communicator includes browser, email tool, collaboration tool, calendar and scheduling tool, *etc.* One "learning curve," integration, compatibility, *etc.*
- Usage monitoring
  - Data mining, strategic alliances
  - "Installed base" ≠ "Active installed base"

# Pluses and Minuses of Network Effects

+ Initial "Metcalf's Law"- based boom
+ Initial boom <u>accelerated</u> by bundling, complementary products, etc.
- Network effects ≠ strong lock in

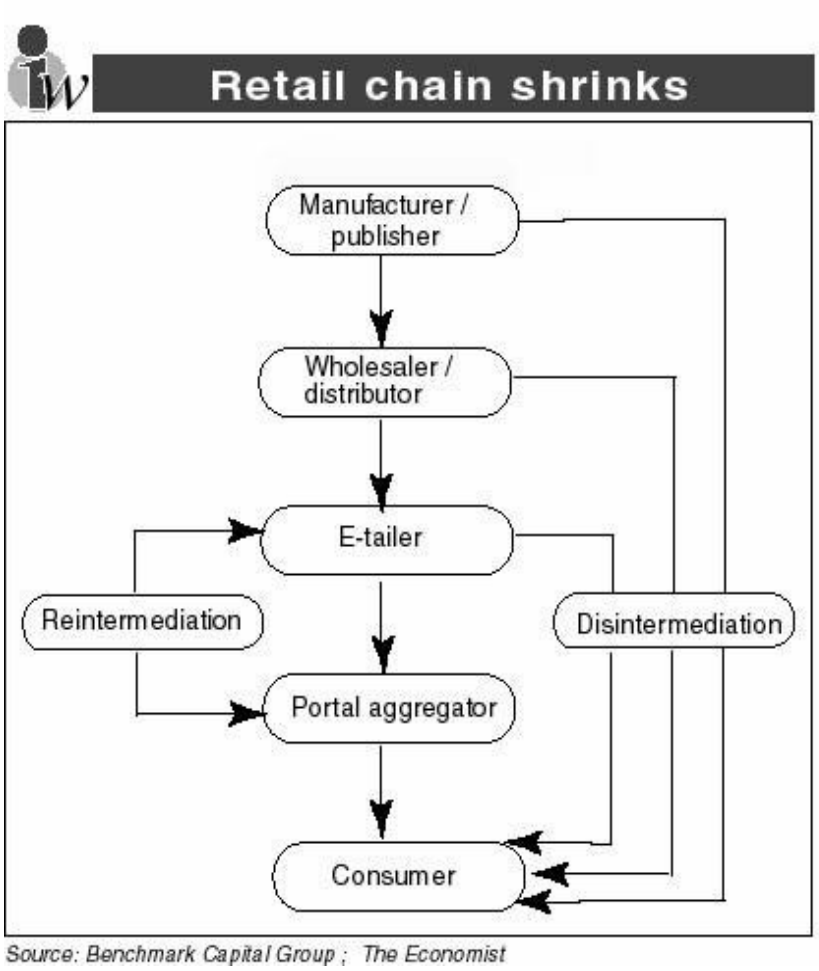
  high market share ≠ high switching costs
- Network effects are strong for "browser" but weak for any particular browser.

# Terminology

- B2C Commerce: Interactions relating to the purchase and sale of goods and services between a business and consumer—retail transactions.
- "Novelty" is that retail transaction is done on the Internet, rather than in a "brick and mortar" store location.
  - All the customer needs is a browser!
- Technical evolution of B2C from "brick and mortar" model not new.

# First-Generation B2C

- Main Attraction: Lower Retail Prices
- "B2C Pure Plays" could eliminate intermediaries, storefront costs, some distribution costs, *etc.*
- Archetype: www.amazon.com



# "Multi-Channel" Retail (B2C w/ B&M)

- Exploit multiple marketing and distribution channels simultaneously
  - B&M ("bricks and mortar") stores: Customers browse on the web before going to the store.
  - Catalog sales, telephone, tv advertising,…
- Since 2002, multi-channel retailers (*i.e.,* B&Ms or traditional catalog companies that also sell online) have accounted for most of B2C e-commerce. Originally, they focused mostly on high-margin sales, *e.g.*, computers, travel, and automotive.
- Multi-channel retailers are more profitable, on average, than web-based and store-based retailers.



(source: Boston Consulting Group)

# eBay Business Model

- Sellers pay small fee (<$2) per listed item.
- eBay takes a cut (~2.5%) of each sale.

  Sellers are willing to pay this fee, because it's a very small price to pay compared to the global exposure they get.
- Although the percentage earned on any given item is small, this is profitable for eBay precisely because the market is global: Millions of new items are added to the site everyday.

# Business Model (continued)

- Buyers and sellers handle exchange and payment (but, eBay offers support for PayPal exchanges).
- eBay has no inventory, no transportation, *no costs at all except website operation*.

**Conventional wisdom:** Service is *technically commoditizable*, but strong network effects favor eBay.

# Terminology

- A product or service is technically commoditizable if it is built using standard parts or protocols (*i.e.*, "commodities"), and its functionality can easily be reproduced by competitors. Examples:
  - eBay auctions
  - Netscape browser
- A product or service that requires significant proprietary or specialized knowledge to produce, deliver, or maintain is not technically commoditizable. Examples:
  - MS Windows
  - Mac OS

# Technical Foundations of Internet C2C Commerce

- Market Design (*e.g.*, Auction Types)
- Payment Systems (can't always use credit cards)
- E-Market Operations
  - Website Design Issues (*e.g.*, UI)
  - ★System Reliability and Availability

# Topics

- Internet design
- The web
- Information goods and information industries
- Basics of B2C and C2C e-commerce
- **Copyright law**
- Online music distribution
- Search

# Copyright Law

- Lecture 7 and DMCA material from lecture 9
- Reading assignments from Feb. 6, 2007

# Basis of US Copyright Law

U.S. Constitution:

[Article I, Section 8]

"The Congress shall have Power…
[Clause 8] To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries…"

Note: The founding fathers did not feel the need to empower Congress to create physical property rights.

# Examples of Exclusive Rights

- to reproduce the copyrighted work
- to prepare derivative works
- to distribute copies through sales, rental, lease, or lending
- to perform the copyrighted work publicly (applies, *e.g.*, to plays)
- to display the copyrighted work publicly (applies, *e.g.*, to sculpture)
- digital audio transmission

[These are paraphrases.]

# Exception: "4-factors" test for "Fair Use"

- The purpose and character of the use, including whether such use is of a commercial nature or is for non-profit educational purposes
- The nature of the copyright work
- The amount and substantiality of the portion used in relation to the copyright work as a whole
- The effect of the use upon the potential market for or value of the copyrighted work

# Exception: First-Sale Rule

- When a copyright owner sells a copy of a work, he relinquishes control over that copy but not over the work.
- The work cannot be reproduced by the purchaser, but the copy can be loaned, resold, or given to someone else.
- "Promotes progress" by enabling, *e.g.*
  - libraries
  - used book stores

# General Structure of Copyright Law

- Copyright owners' rights stated explicitly.
- General public has no explicitly stated rights, just exceptions to owners' rights.
- Fair use is a *defense* against a charge of infringement.

This structure works fairly well for traditional media, particularly books.

# Structure is Challenged by Digital Works

- Digital documents are fundamentally different:
  - Copies are perfect.
  - Copies can be made at zero cost.
  - Copying is not necessarily a good proxy for infringement.
- TPSs are imperfect:
  - A perfect TPS could moot fair use: no infringement, no charge, no defense.
  - But no TPS can be perfect in today's computers. General purpose PCs are programmable, and hence TPSs are circumventable (at least by experts).

# Digital Millennium Copyright Act (1998)

- Illegal, except under narrowly defined special circumstances, **to circumvent effective technological protection measures**
- Illegal to distribute **circumvention tools**
- Gives content owners a property right in TPS as well as the content that the TPS protects. In SAT terms, circumvention is to infringement as breaking and entering is to burglary.

# Techies' Objection to DMCA

- What is an "**effective** technological protection measure?"
  - If a skilled hacker can break it, is it "**effective**"?
  - If an average computer-literate person can break it, *but few do*, is it "**effective**?"
- Weakens incentives for content owners to pay for good IP-management technology.
- Shifts costs from content owners to society at large, by shifting responsibility from TPSs to courts and police.
- **Exceptions for R&D are vague**.

# Topics

- Internet design
- The web
- Information goods and information industries
- Basics of B2C and C2C e-commerce
- Copyright law
- **Online music distribution**
- Search

# Online Music Distribution

- Lecture 9
- Reading assignment from Feb. 8, 2007

# Origin of the "Internet Problem" for Music Distributors

- Music is sold unencrypted in digital form on CDs.
- Music CDs are readable by PCs.
- Digital content read off music CDs is easily convertible to the compact MP3 format.
- MP3 files are easy to distribute using standard Internet protocols.

# Three Major "Enforcers" Support a Content-Distribution Business

- Copyright law
- Technical Protection System (TPS)

* Business Model

# Dual Doomsday Scenarios

Rights Holders and Distributors:

TPSs don't suffice. Digital copying, modification, and distribution are uncontrollable. We need more legal and social sanctions.

Fair-Use Advocates and (Some) Consumers: TPSs work too well. Some rights holders now have *more* control than they do in the analog world. *Normal* use can often be monitored and controlled in the digital world.

# Discussion Point

After many years of online music distribution, may failed business models, and the success of iTunes, we have Jobs's suggestion to do away with DRM.

Key component of the argument: People who are currently paying for music ***are not doing it because DRM has forced them to***.

# Topics

- Internet design
- The web
- Information goods and information industries
- Basics of B2C and C2C e-commerce
- Copyright law
- Online music distribution
- **Search**

# Search

- Lecture 10
- Third HW assignment
- Reading assignment from Feb. 15, 2007

# Two Aspects of WWW Searching

- Analyze *contents* of pages
  - Text (*e.g.*, search terms)
  - Structure (*e.g.*, HTML tags)
- Analyze *structure* of WWW digraph
  - Links to page P indicate *interest* in the contents of of P.
  - *Importance* depends on *who* is interested.
  - Requires *global* analysis of digraph.

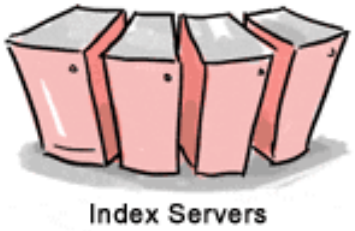
# Technical Highlights

- PageRank Technology: Linear-algebraic, objective calculations of the "importance" of a webpage.
  - Link from Page A to Page B is a "vote" for B.
  - Importance of A is factored into the vote.
  - Page owners cannot pay to have their PageRanks modified. (Note the difference between buying a "sponsored link" and getting a higher PageRank.)
  - Google employees can modify a PageRank in exceptional circumstances (*e.g.*, security threats).

# Life of a Query


**1.** The user enters a query on a web form sent to the Google web server.


**2.** The web server sends the query to the Index Server cluster, which matches the query to documents.




**3.** The match is sent to the Doc Server cluster, which retrieves the documents to generate abstracts and cached copies.



**4.** The list, with abstracts, is displayed by the web server to the user, sorted (using a secret formula involving PageRank).