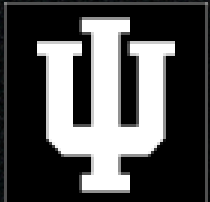


# Social Web Search

Filippo Menczer

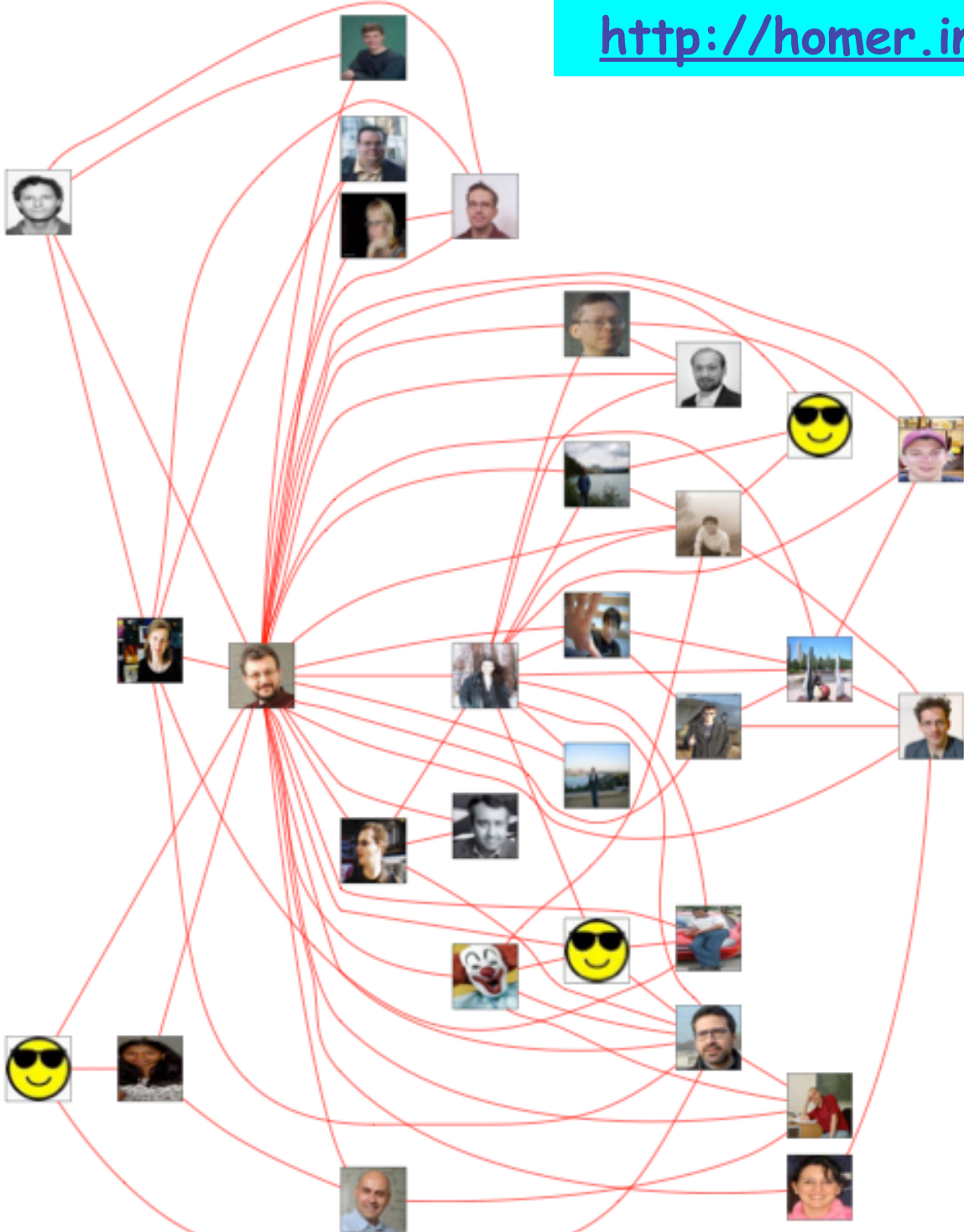
Informatics and Computer Science

Leveraging online social behavior  
for collaborative Web search

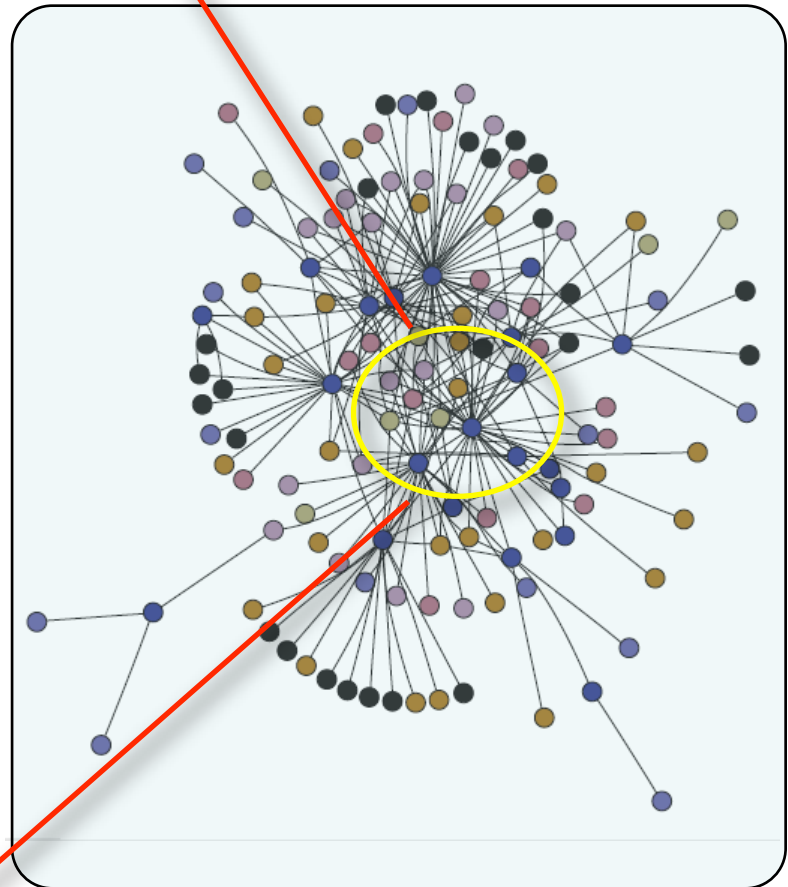




<http://homer.informatics.indiana.edu/~nan/>

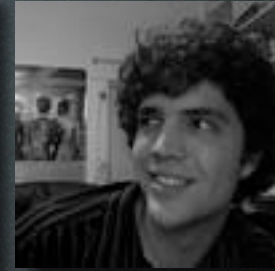
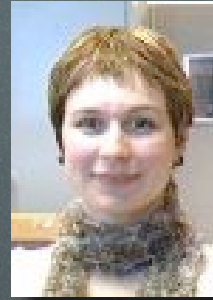


NaN:  
Networks  
& agents  
Network



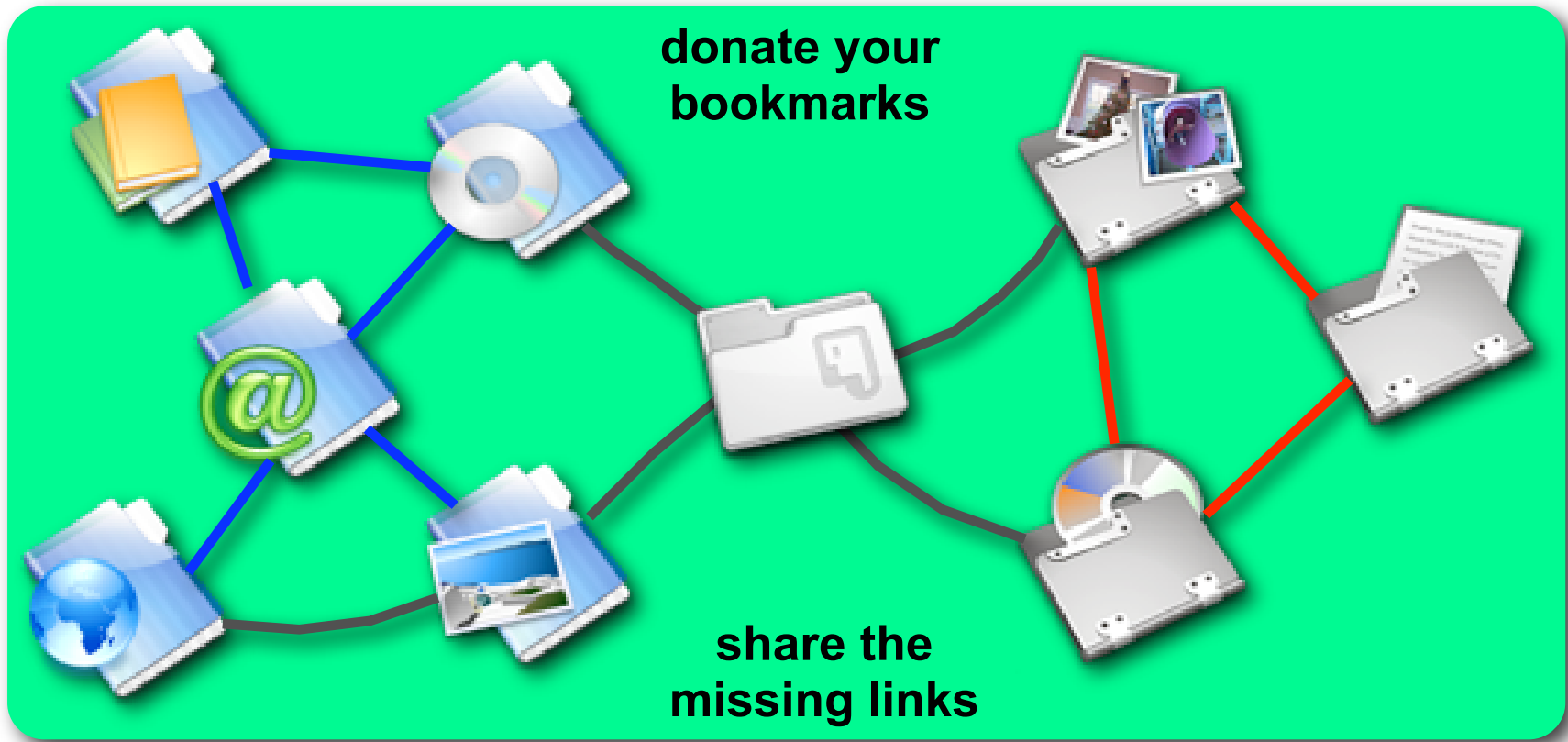


# Outline



- Part 1: GiveALink.org
- Part 2: 6S





**personalized recommendations**



**social search**

### Search GiveALink

Related Links

Related Links

### Personalized Search

Personalized Links

Personalized Links

### Surprise Me

Novel Links

Novel Links

### Popular Tags

amazon archives blog  
blogs book computer  
development download free  
information java langages linux  
live media mp3  
music news

## Today's Personalized Recommendations

[kuro5hin.org || technology and culture, from the trenches](#)

[The Register: Sci/Tech News for the World](#)

[Auditor.main - Hackingdefined](#)

[de.lirio.us: Store, share and tag your favourite links](#)

[Gizmodo, The Gadget Guide](#)

[anandgraves.com](#)

[ExtremeTech - Deep technology for enthusiasts and professionals](#)

[Questionable Content: New comics every Monday through Friday](#)

[Neil Gaiman - Home](#)

[eWEEK.com - Enterprise Technology News and Reviews](#)

[Click Here](#) for New Recommendations

### Top Ten Bookmarks <sup>?</sup>

1. [Zvon Index](#)
2. [XML Files - Extensible Markup Language Tutorials for Developers](#)
3. [rockcritics.com main page](#)
4. [XML.com: XML From the Inside Out -- XML development, XML resources, XML specifications](#)
5. [the Harmony Music List](#)
6. [:::Activa Podcasts:::](#)
7. [http://www.diigo.com/toolbar](#)
8. [TrouserPress.com](#)
9. [lucky on All Consuming](#)
10. [SHOUTcast | Free internet radio!](#)

### Stats

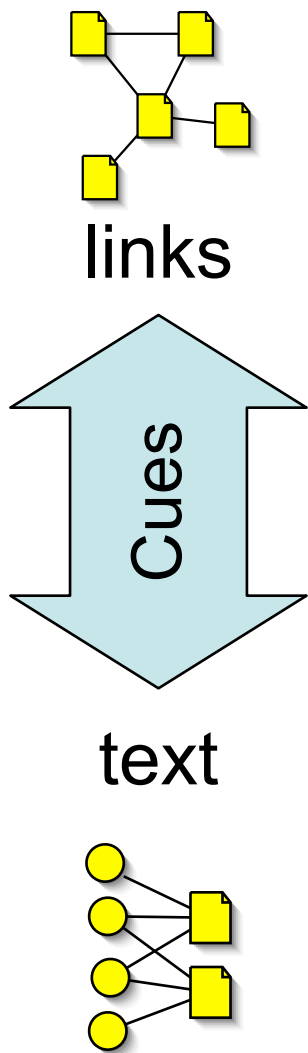
**1609285** links  
**4976512** relationships  
**2499** donations  
**376** registered users  
**Last Updated:** Fri Dec 15  
01:15:52 2006



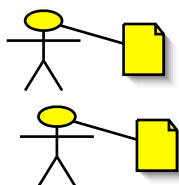


givealink.org

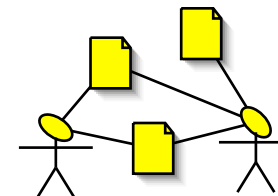
*I donated my bookmarks to science. Did you?*




author



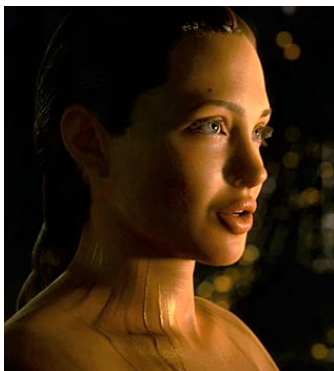
social





givealink.org

*I donated my bookmarks to science. Did you?*



news

research

fun



$\sigma_A$



+

$\sigma_B$



=

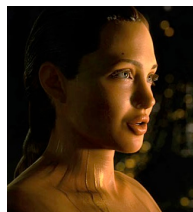
$\sigma$



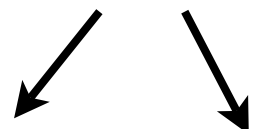


givealink.org

*I donated my bookmarks to science. Did you?*



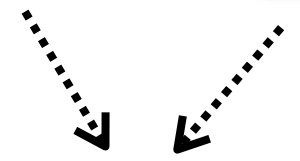
$$\sigma(x, y) = \frac{1}{N} \sum_{u=1}^N \frac{2 \log \left( \frac{|F_u[a(x,y)]|}{|R_u|} \right)}{\log \frac{|F_u(x)|}{|R_u|} + \log \frac{|F_u(y)|}{|R_u|}}$$



FA(x)

FA(y)

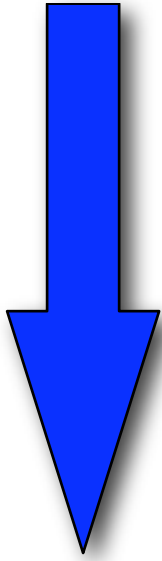
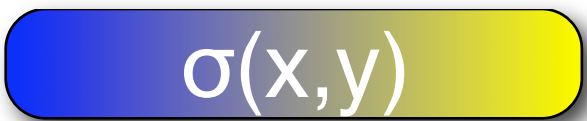
FB(a(x,y))



+



=

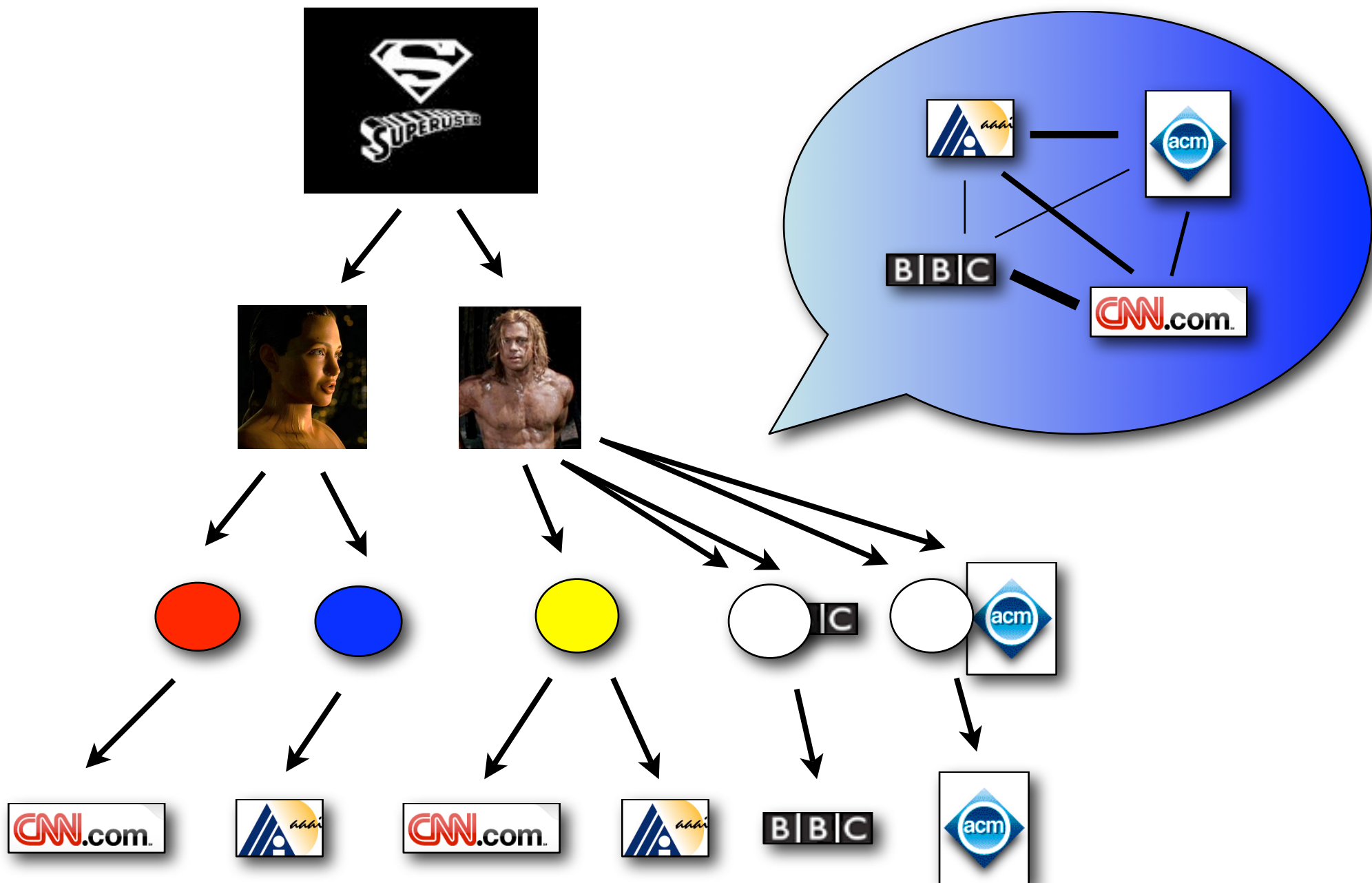






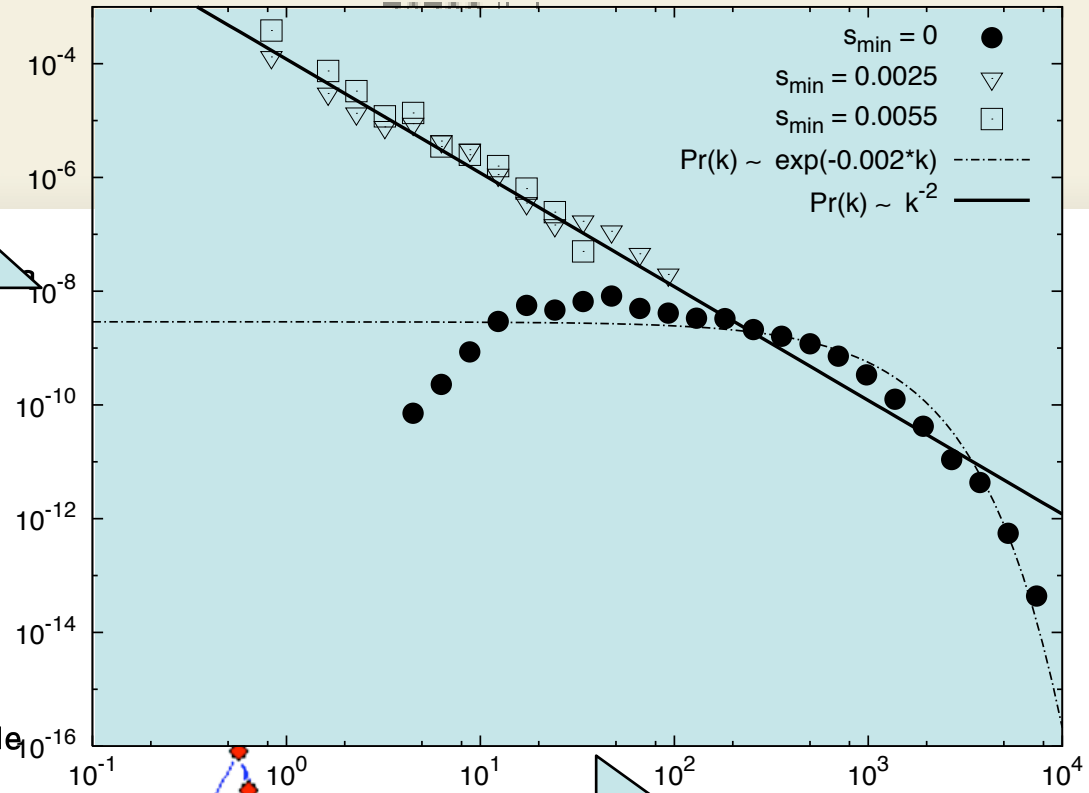
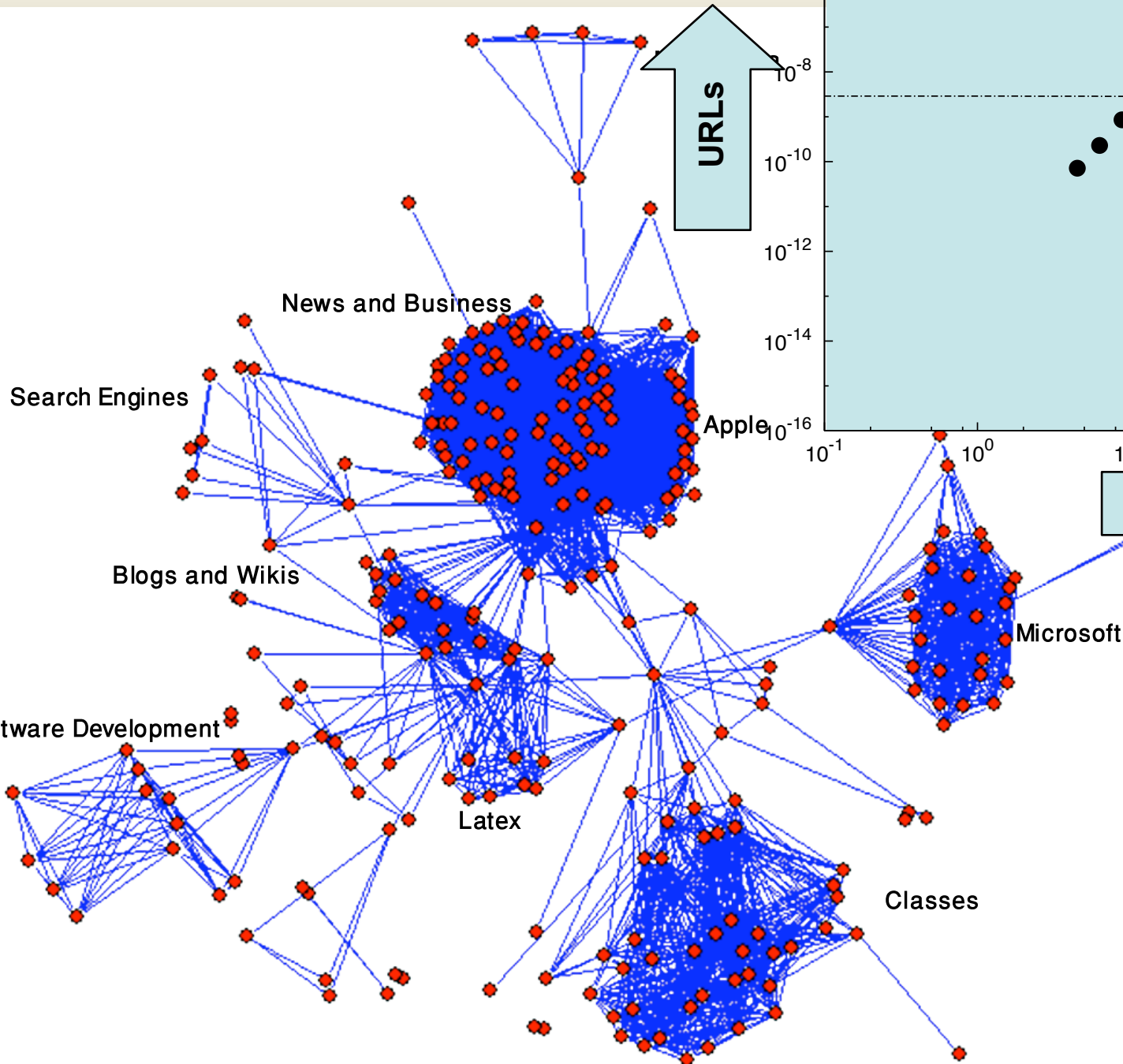
givealink.org

*I donated my bookmarks to science. Did you?*





givealink.org







givealink.org

*I donated my bookmarks to science. Did you?*

Username:

[Register?](#)

Password:

[About:](#)

[Project](#)

[Help](#)

[FAQ](#)

[Privacy](#)

[Links:](#)

[Share](#)

[Manage](#)

[Download](#)

[RSS](#)

### Search GiveALink

### Surprise Me

### Related URLs sorted by Related URL List

1-10 of 1000 results for <http://www.cnn.com>

[BBC NEWS | News Front Page](#)

[Los Angeles, California, national and world news, jobs, real estate, cars - Los Angeles Times](#)

[The New York Times - Breaking News, World News & Multimedia](#)

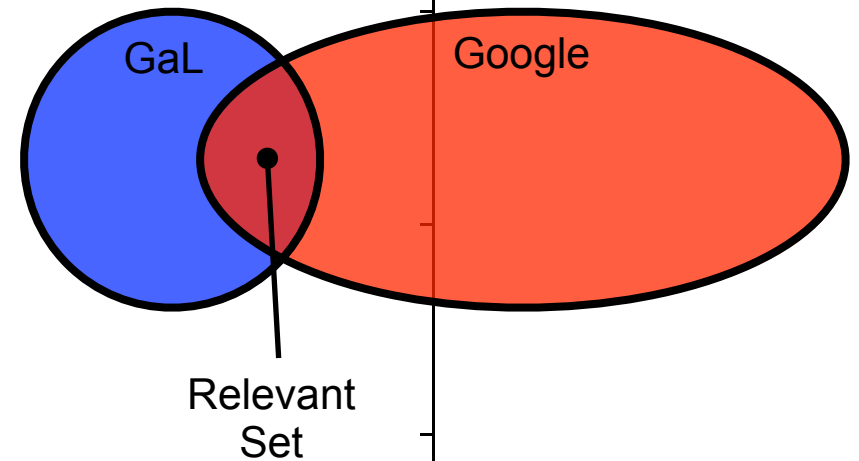
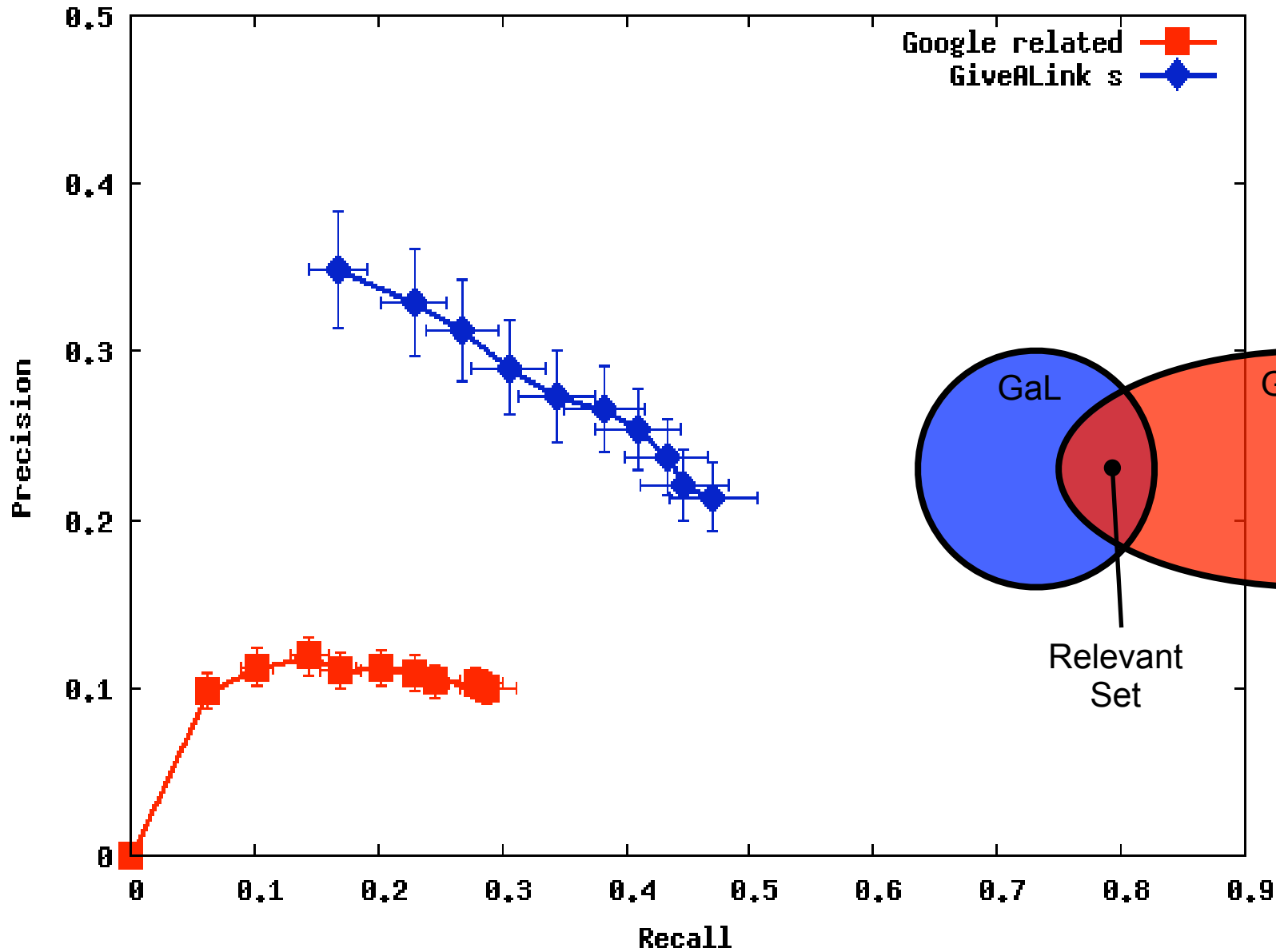
[CNET News.com -- Technology news and business reports](#)

[CNET News.com -- Technology news and business reports](#)

[The New York Times - Breaking News, World News & Multimedia](#)



# User Study



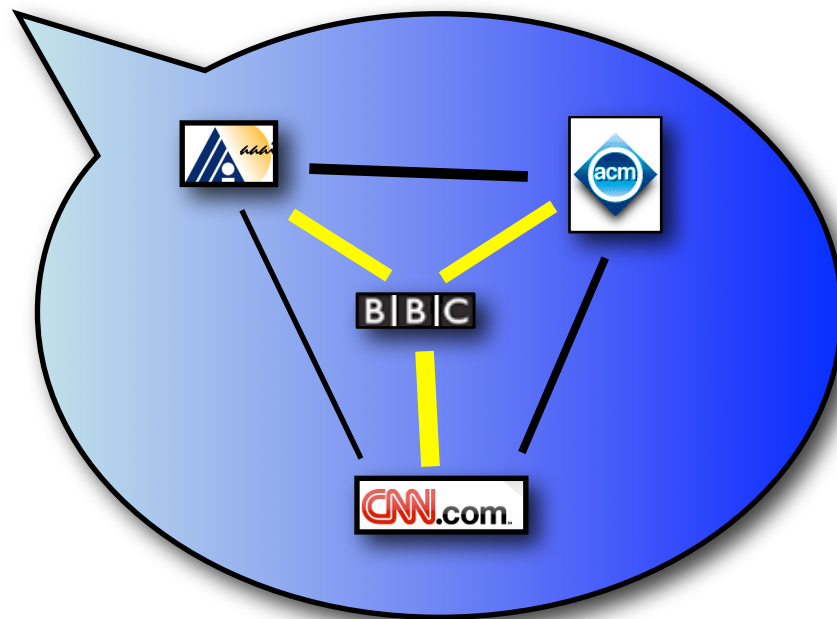




*I donated my bookmarks to science. Did you?*

$$c(x) = \frac{1}{|U|} \sum_{y \in U} \left[ 1 + \min_{x \rightsquigarrow y} \sum_{(u,v) \in x \rightsquigarrow y} \left( \frac{1}{\sigma(u,v)} - 1 \right) \right]^{-1}$$

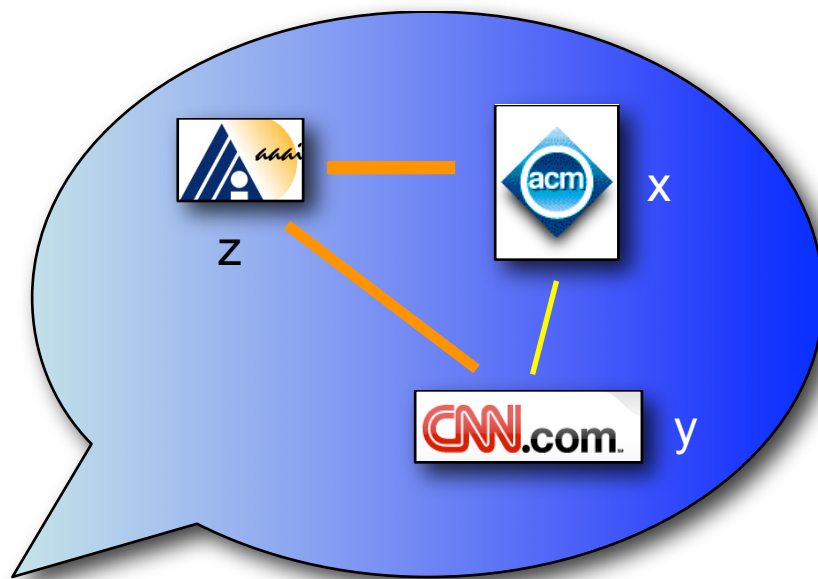
# Centrality & Prestige



$$p_{t+1}(x) = (1 - \alpha) + \alpha \cdot \sum_{y \in U} \frac{\sigma(x, y) \cdot p_t(y)}{\sum_{z \in U} \sigma(y, z)}$$



# Ranking & Recommendation by Novelty



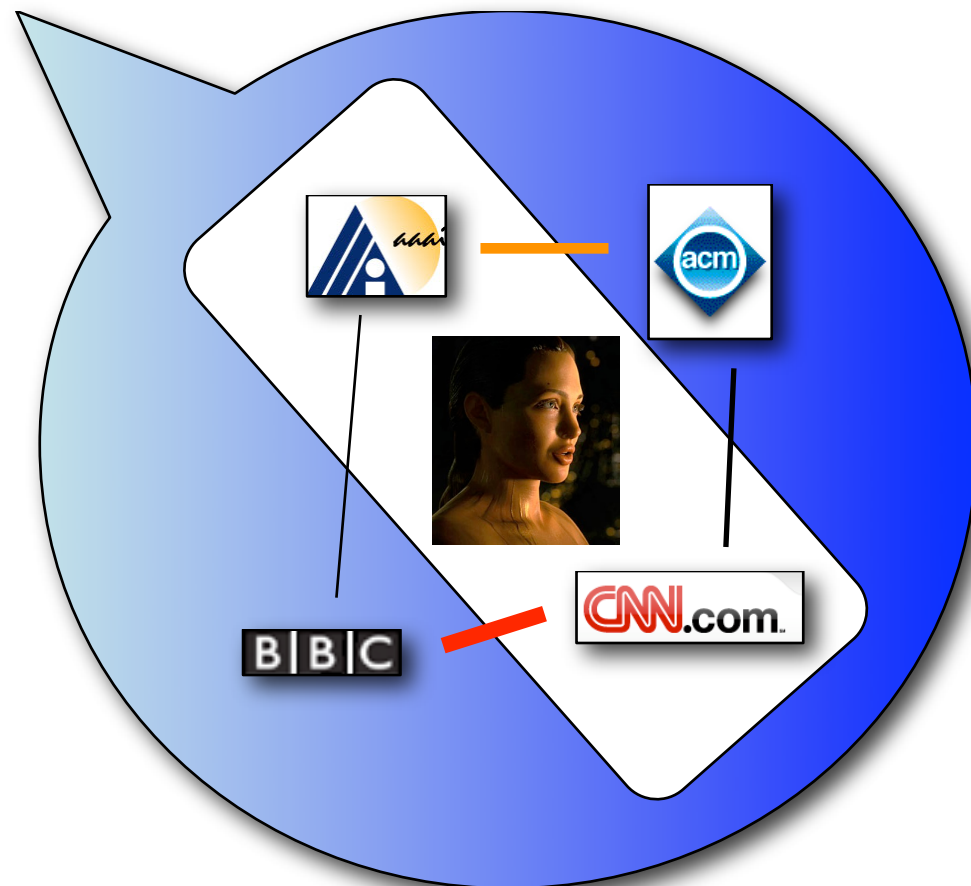
$$\nu(x, y) = \frac{\left[ 1 + \min_{z \in U} \left( \frac{1}{\sigma(x, z)} + \frac{1}{\sigma(z, y)} - 2 \right) \right]^{-1}}{\sigma(x, y)}$$





$$\eta(x, A) = \max_{y \in A} \left[ \sigma(x, y) \cdot \log \left( \frac{N}{N(y)} \right) \right]$$

# Personalization





givealink.org

*I donated my bookmarks to science. Did you?*

**givealink.org**

Bookmark Donations for Science

<http://www.givealink.org> <http://www.backflip.com/login.ihtml>

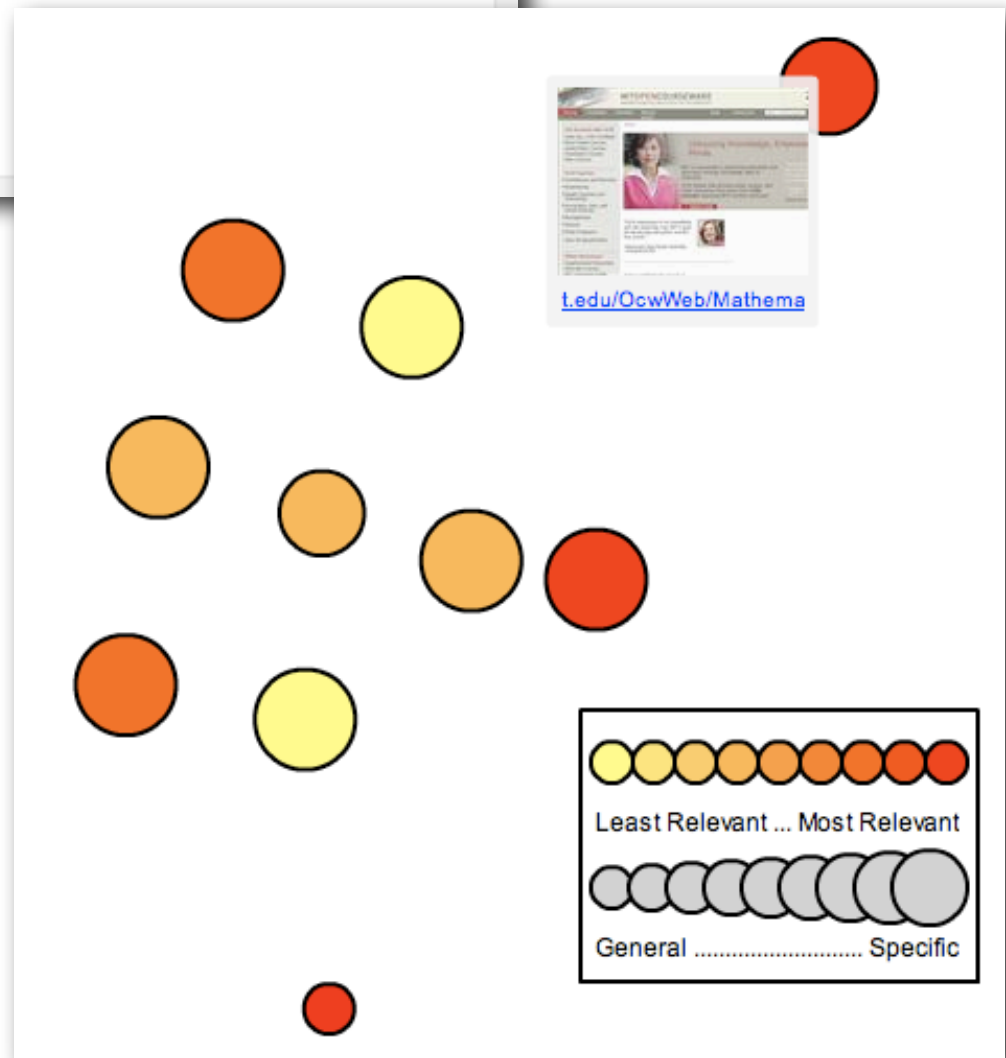
similarity:0.00177462

RSS

1.0

REST (XML) Web service

# Navigation





*I donated my bookmarks to science. Did you?*

Links: Share Manage Download RSS

### 3 Ways to Share

#### 1. Donate Bookmarks

[Donate as registered user](#) to get personalized recommendations and search results based on your bookmarks

[Donate anonymously](#) to help us without giving us your email address; we will not be able to personalize your searches, nor give you personalized recommendations

#### 2. Import Bookmarks from del.icio.us

username:

password:

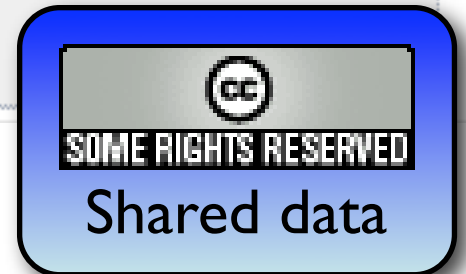
#### 3. Bookmarklet

Install the bookmarklet by dragging the link to the toolbar of your browser; then click when visiting a page you would like to donate:

[Give This Link](#)

Install the pop-up bookmarklet by dragging the link to the toolbar of your browser; then click when visiting a page you would like to donate:

[Give This Link Pop Up](#)







collaborative filtering,  
social semantic similarity,  
unlinked pages,  
multimedia content, trust

scalability,  
density



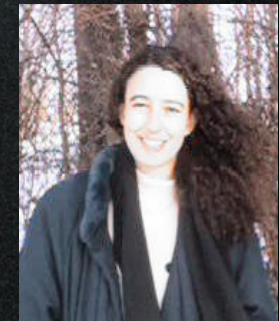
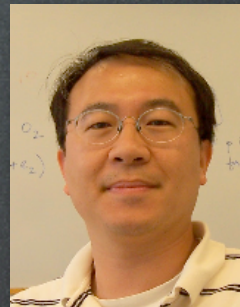
spam

Questions?



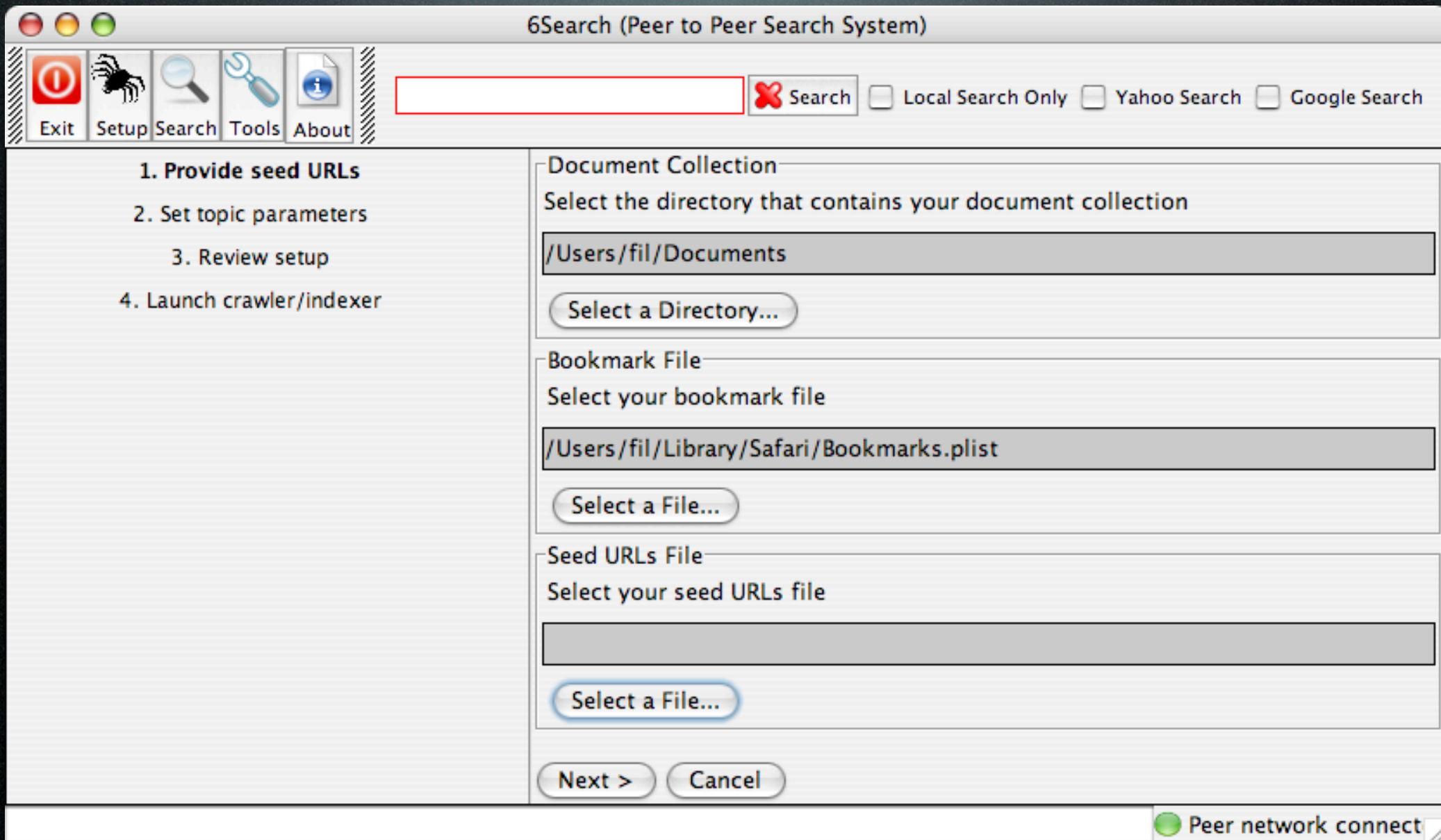
# Outline

- Part 1: GiveALink.org
- Part 2: 6S





<http://homer.informatics.indiana.edu/6S/>





<http://homer.informatics.indiana.edu/6S/>

6Search (Peer to Peer Search System)

Exit Setup Search Tools About

Local Search Only  Yahoo Search  Google Search

1. Provide seed URLs
2. Set topic parameters
3. Review setup
4. Launch crawler/indexer

Crawling Progress  
11/09 14:53:34

Indexing Progress  
11/09 14:56:28

Information Summary

Please verify the information you entered.

Document Collection:  
/Users/fil/Documents/Homepages/Informatics

Bookmark File:  
/Users/fil/Library/Safari/Bookmarks.plist

Crawling Topic:  
web mining modeling search peer crawl networks agents

Seed URLs File:

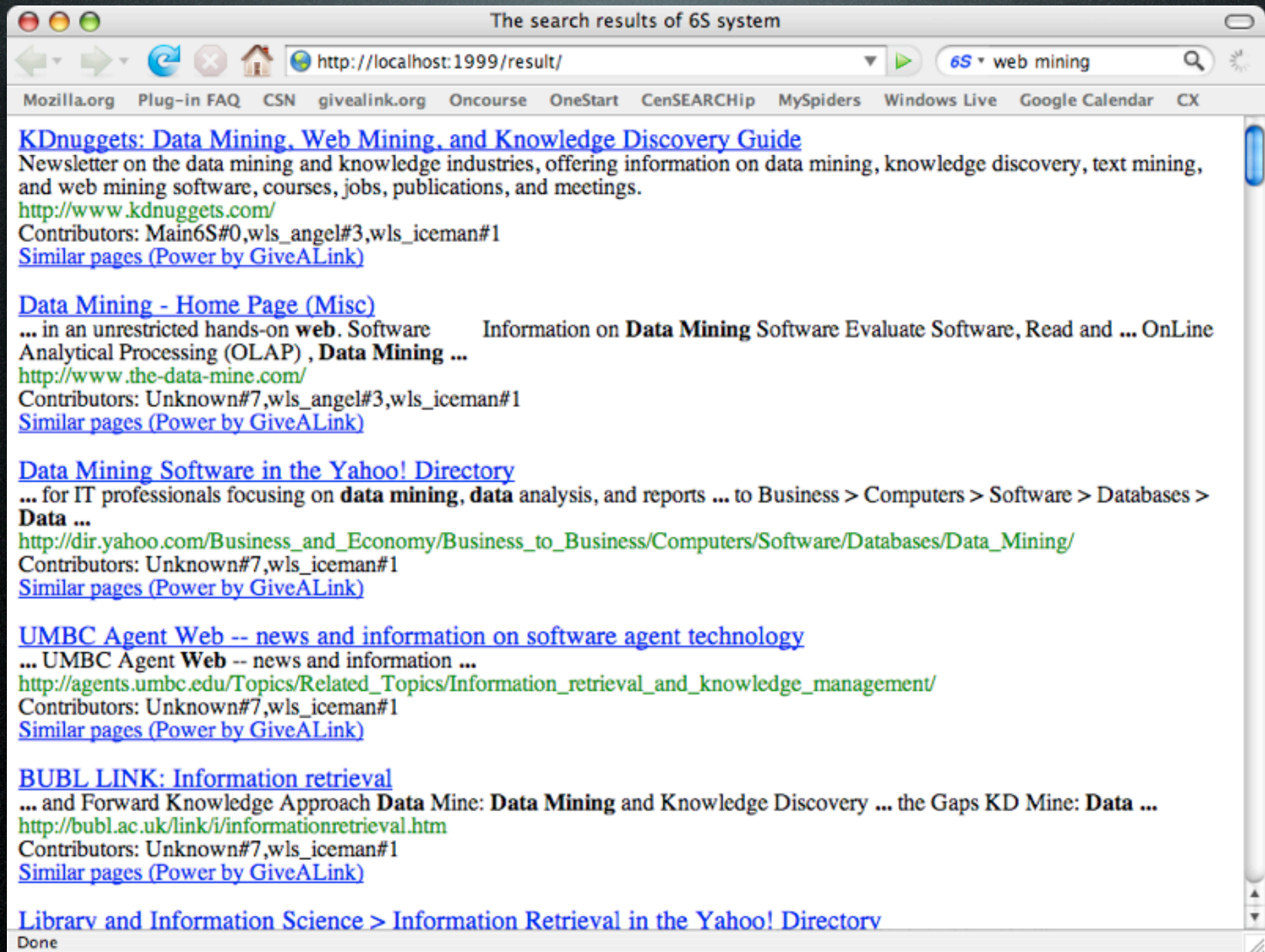
Number of Pages:  
100

Your Email:  
fil@indiana.edu

Tag for URLs:  
web mining modeling search peer crawl networks agents

Peer network connect

<http://homer.informatics.indiana.edu/6S/>



The screenshot shows a web browser window titled "The search results of 6S system". The address bar displays "http://localhost:1999/result/" and the search query is "6S web mining". The browser's toolbar includes navigation buttons and a search icon. Below the address bar, there is a list of search engines: Mozilla.org, Plug-in FAQ, CSN, givealink.org, Oncourse, OneStart, CenSEARCHip, MySpiders, Windows Live, Google Calendar, and CX.

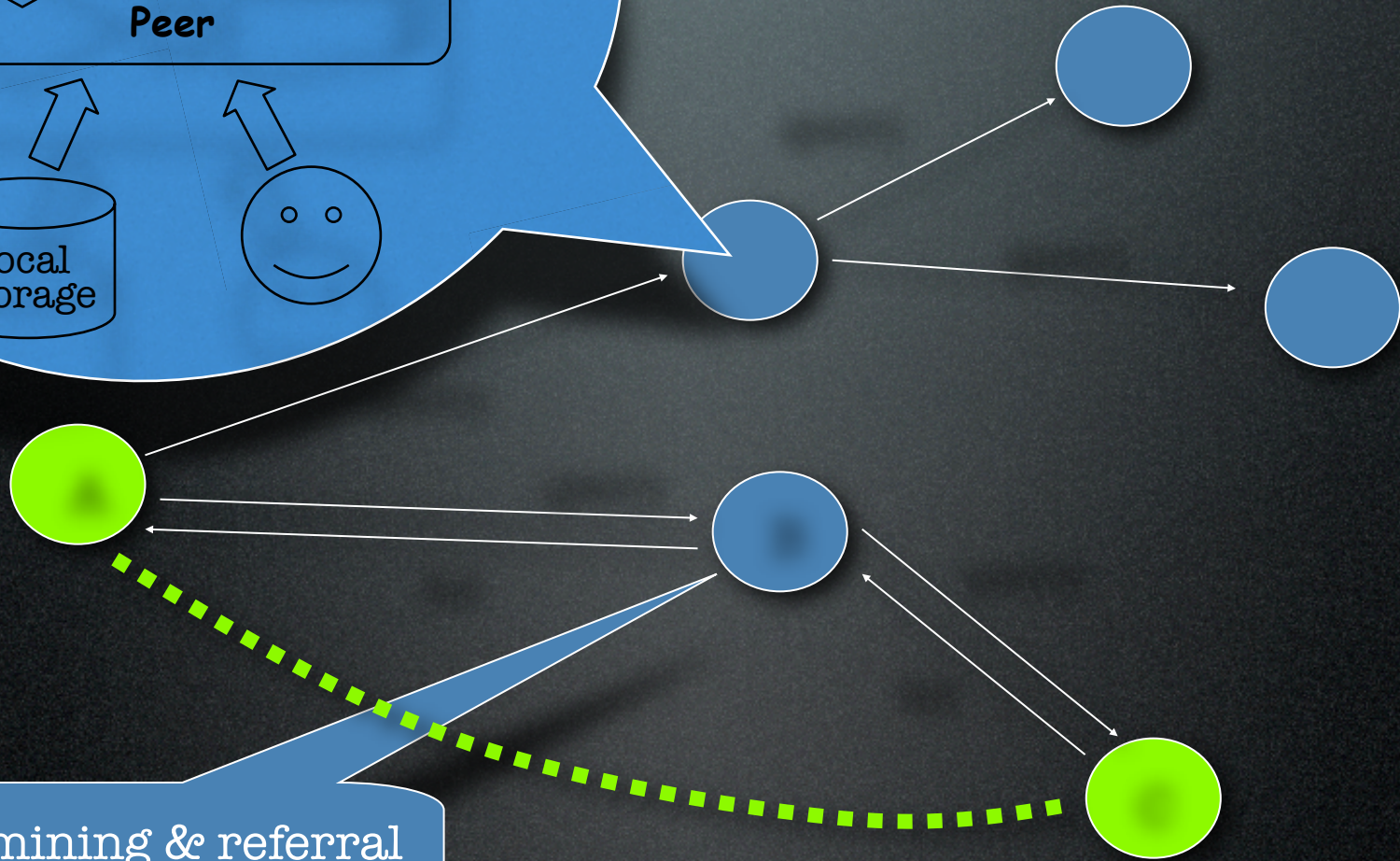
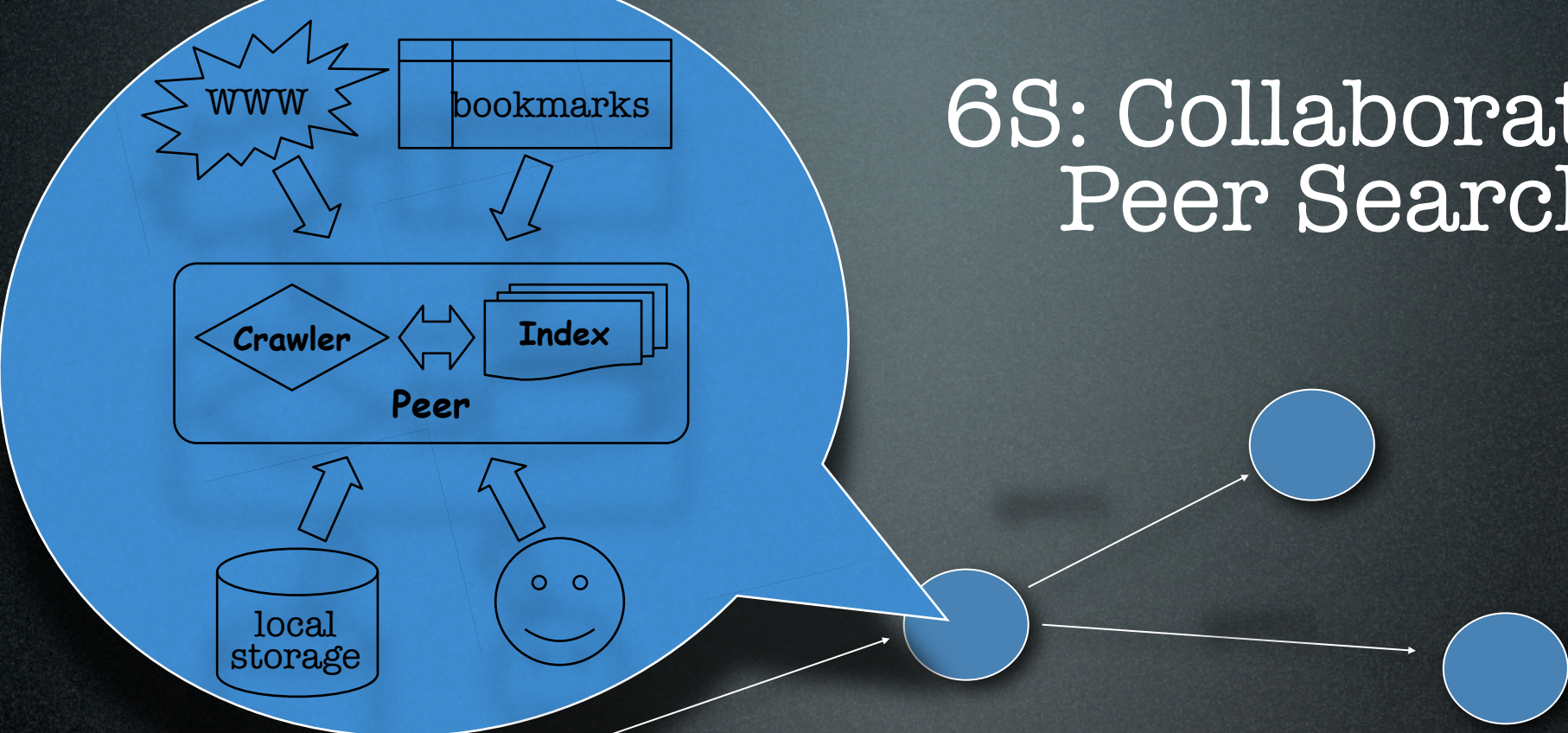
The search results are listed as follows:

- [KDnuggets: Data Mining, Web Mining, and Knowledge Discovery Guide](#)**  
Newsletter on the data mining and knowledge industries, offering information on data mining, knowledge discovery, text mining, and web mining software, courses, jobs, publications, and meetings.  
<http://www.kdnuggets.com/>  
Contributors: Main6S#0,wls\_angel#3,wls\_iceman#1  
[Similar pages \(Power by GiveALink\)](#)
- [Data Mining - Home Page \(Misc\)](#)**  
... in an unrestricted hands-on web. Software Information on **Data Mining** Software Evaluate Software, Read and ... OnLine Analytical Processing (OLAP) , **Data Mining** ...  
<http://www.the-data-mine.com/>  
Contributors: Unknown#7,wls\_angel#3,wls\_iceman#1  
[Similar pages \(Power by GiveALink\)](#)
- [Data Mining Software in the Yahoo! Directory](#)**  
... for IT professionals focusing on **data mining**, **data** analysis, and reports ... to Business > Computers > Software > Databases > **Data** ...  
[http://dir.yahoo.com/Business\\_and\\_Economy/Business\\_to\\_Business/Computers/Software/Databases/Data\\_Mining/](http://dir.yahoo.com/Business_and_Economy/Business_to_Business/Computers/Software/Databases/Data_Mining/)  
Contributors: Unknown#7,wls\_iceman#1  
[Similar pages \(Power by GiveALink\)](#)
- [UMBC Agent Web -- news and information on software agent technology](#)**  
... UMBC Agent Web -- news and information ...  
[http://agents.umbc.edu/Topics/Related\\_Topics/Information\\_retrieval\\_and\\_knowledge\\_management/](http://agents.umbc.edu/Topics/Related_Topics/Information_retrieval_and_knowledge_management/)  
Contributors: Unknown#7,wls\_iceman#1  
[Similar pages \(Power by GiveALink\)](#)
- [BUBL LINK: Information retrieval](#)**  
... and Forward Knowledge Approach **Data Mine: Data Mining** and Knowledge Discovery ... the Gaps KD Mine: **Data** ...  
<http://bubl.ac.uk/link/i/informationretrieval.htm>  
Contributors: Unknown#7,wls\_iceman#1  
[Similar pages \(Power by GiveALink\)](#)
- [Library and Information Science > Information Retrieval in the Yahoo! Directory](#)**

Done



# 6S: Collaborative Peer Search

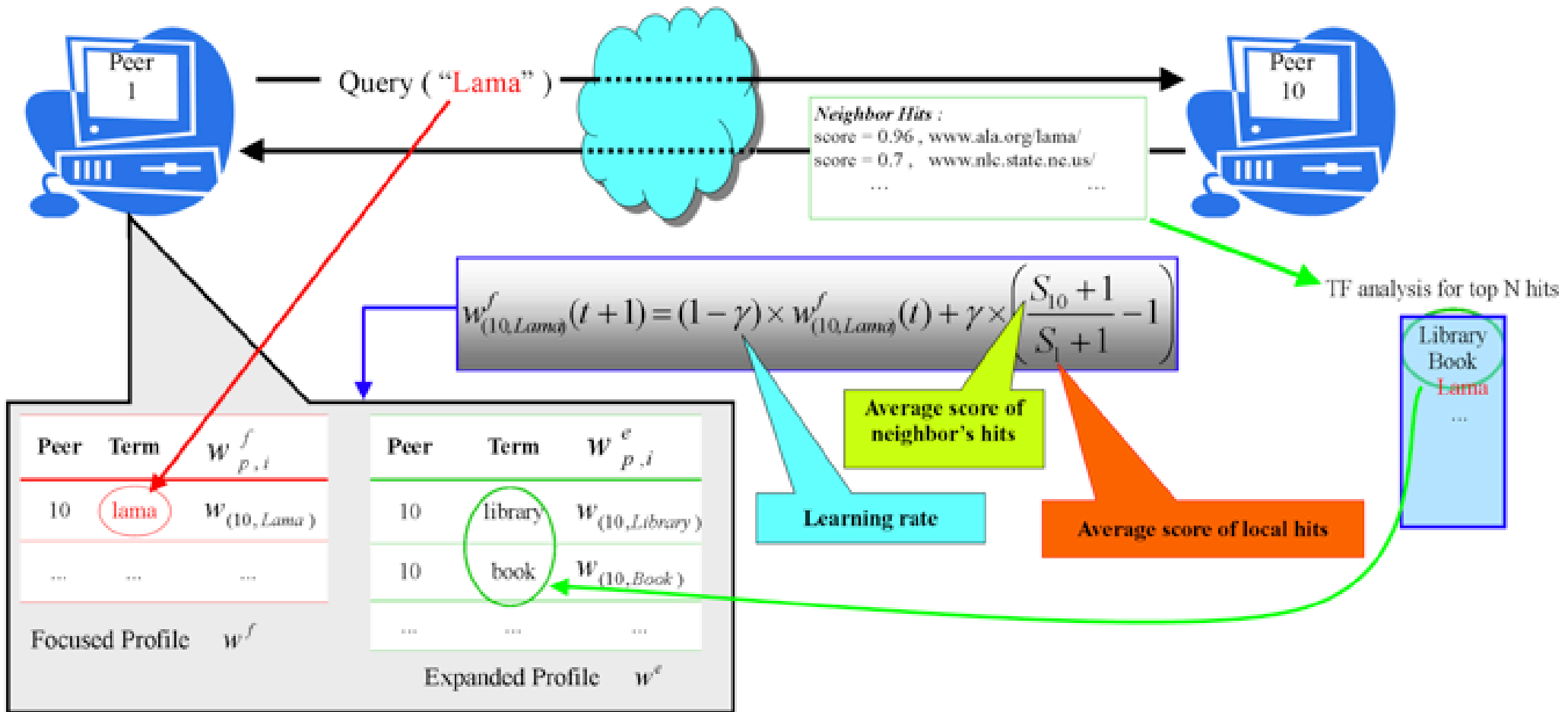


Data mining & referral opportunities

WWW2004  
WWW2005  
WTAS2005  
P2PIR2006



# Algorithm 3: Reinforcement



# Query Routing



Query ( "Lama library book" )

$$\sigma(p, Query) = \sum_{i \in Query} [\alpha \cdot w_{p,i}^f + (1 - \alpha) \cdot w_{p,i}^e]$$

N top ranked among known peers are selected as neighbors and sent Query.

*Forwarding the query*



Peer	Term	$w_{p,i}^f$
10	lama	.98
...	...	...

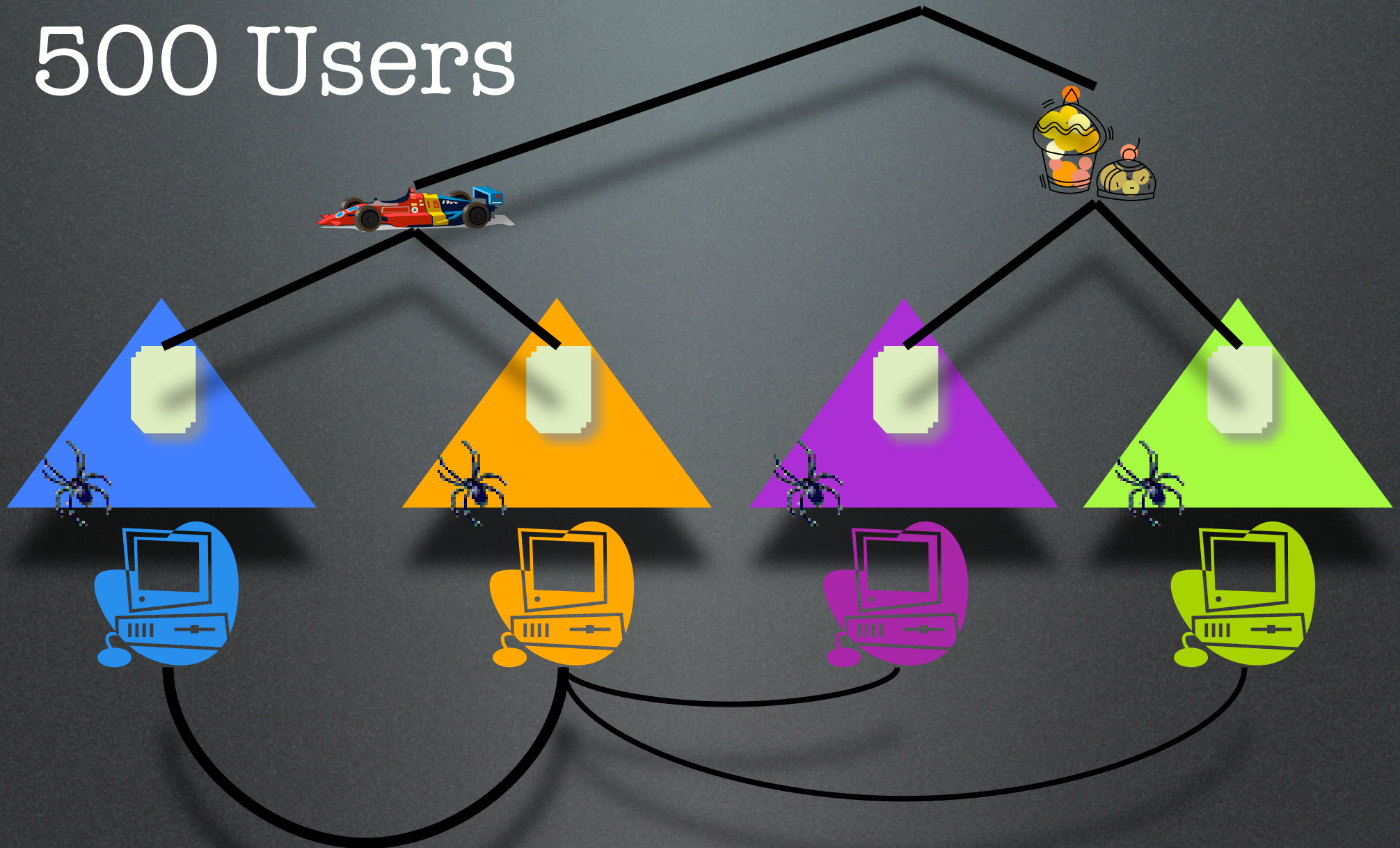
Focused Profile  $w^f$

Peer	Term	$w_{p,i}^e$
10	library	.87
10	book	.67
...	...	...

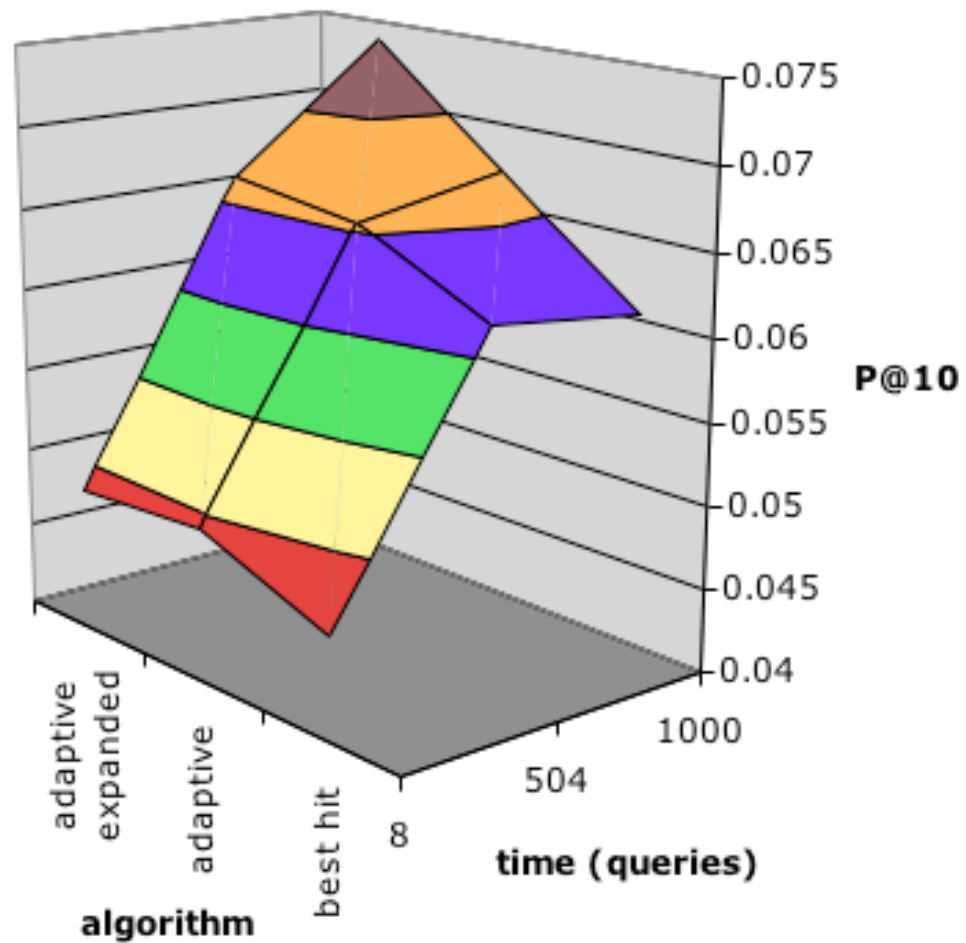
Expanded Profile  $w^e$

# Simulating 500 Users

ODP (dmoz.org)

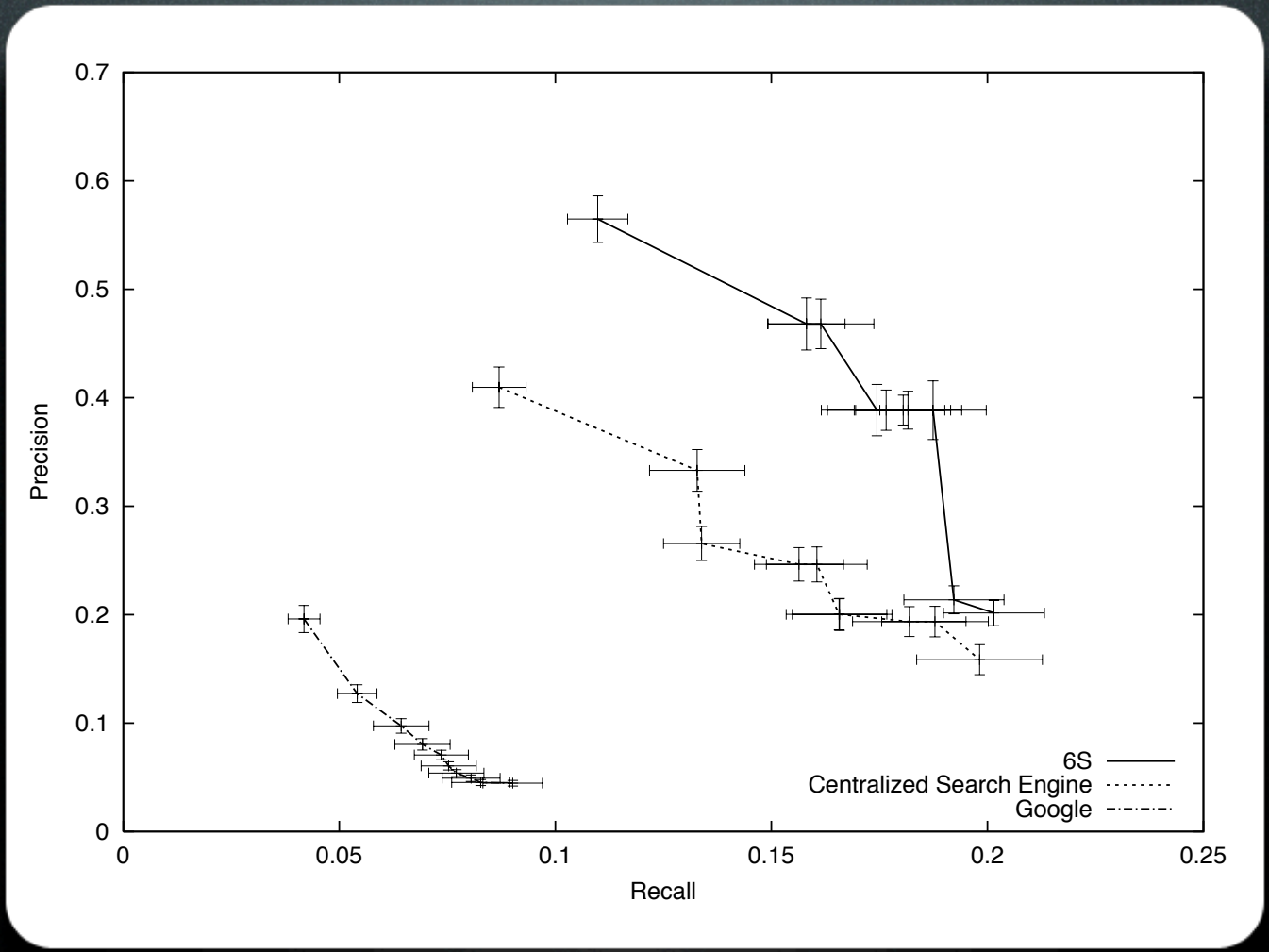






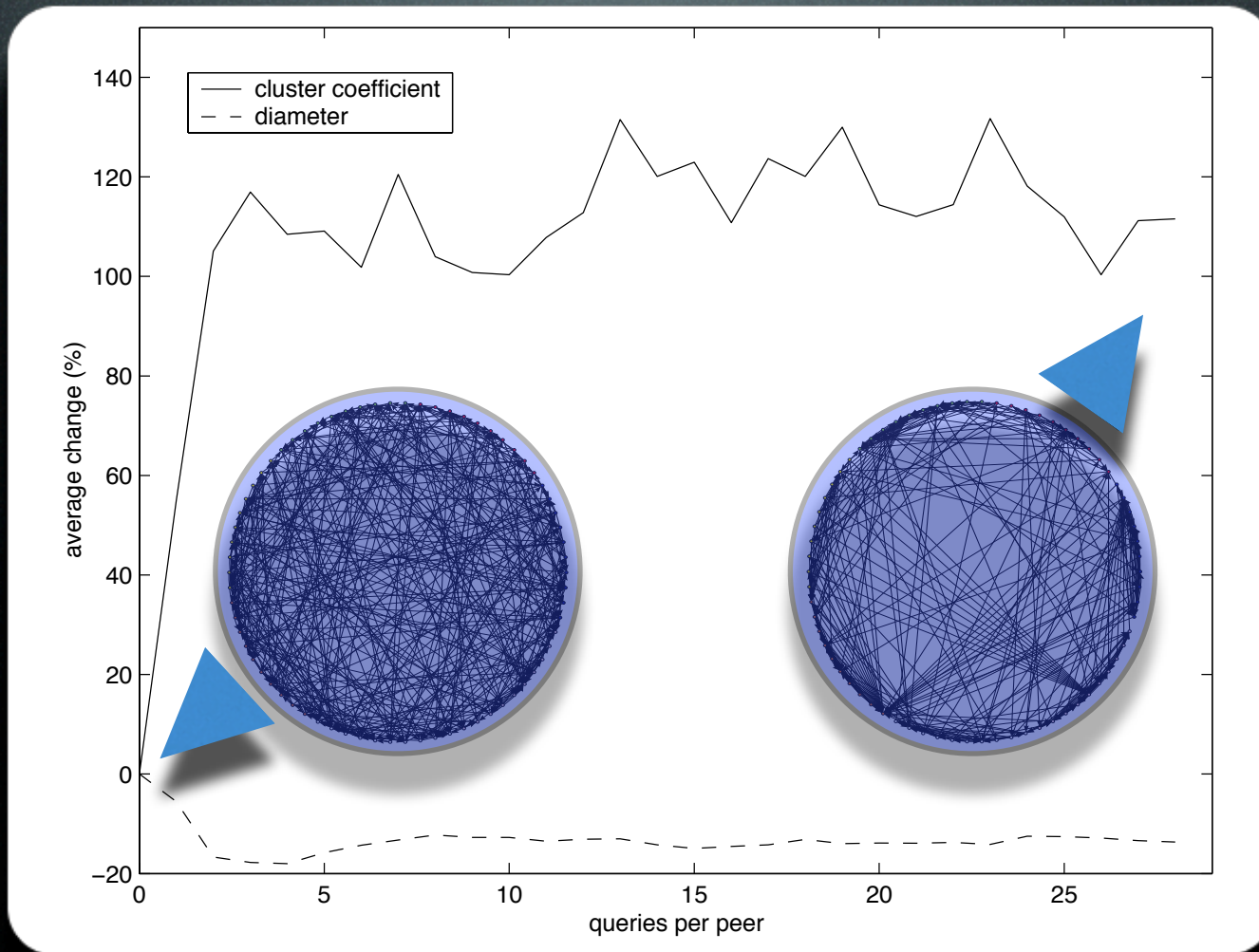
P@10





# Distributed vs Centralized

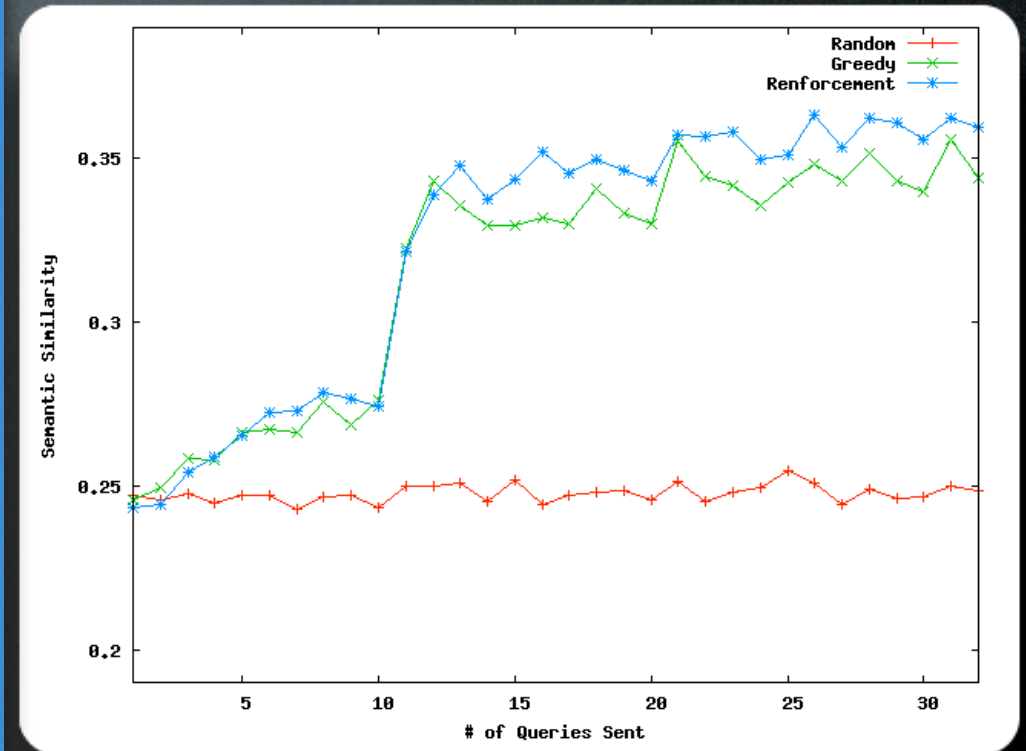




# Small-world



# Semantic Similarity

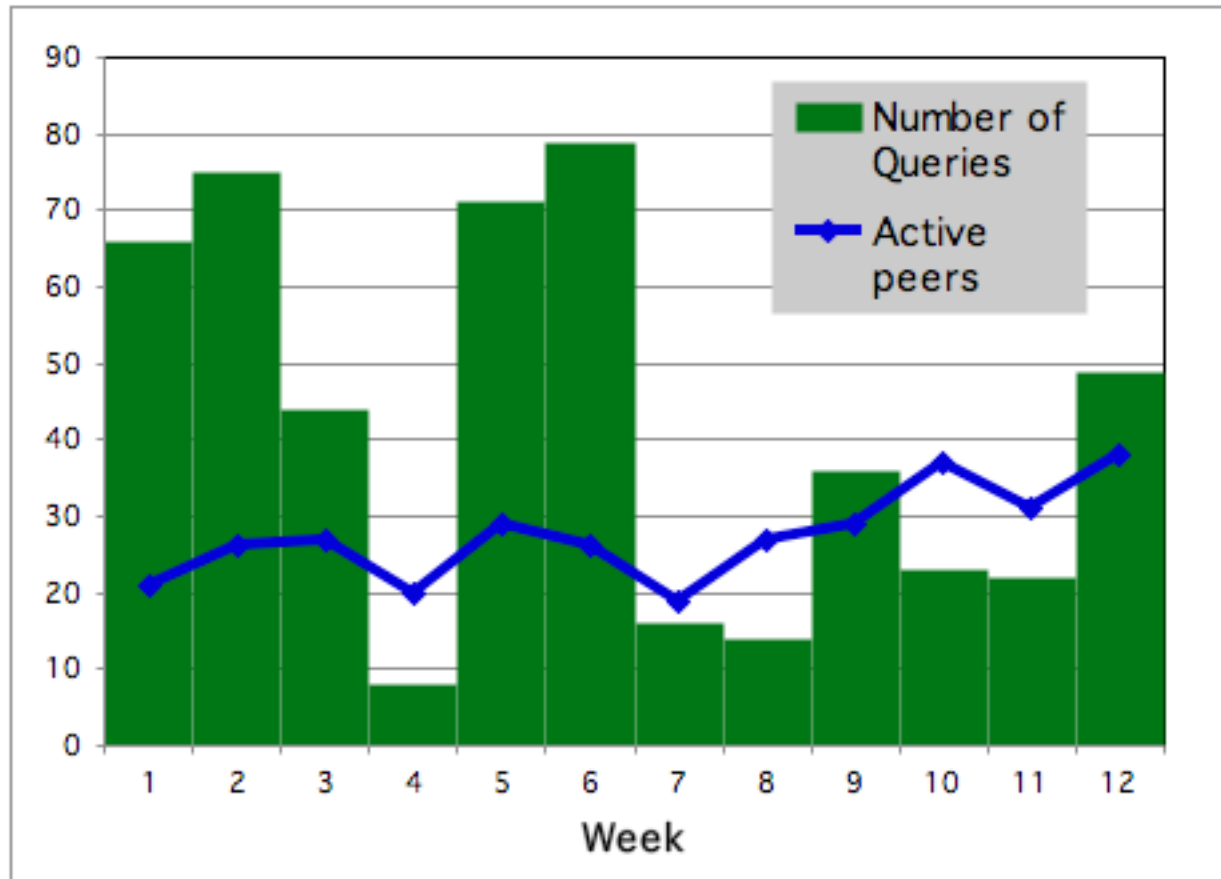




# Ongoing Work

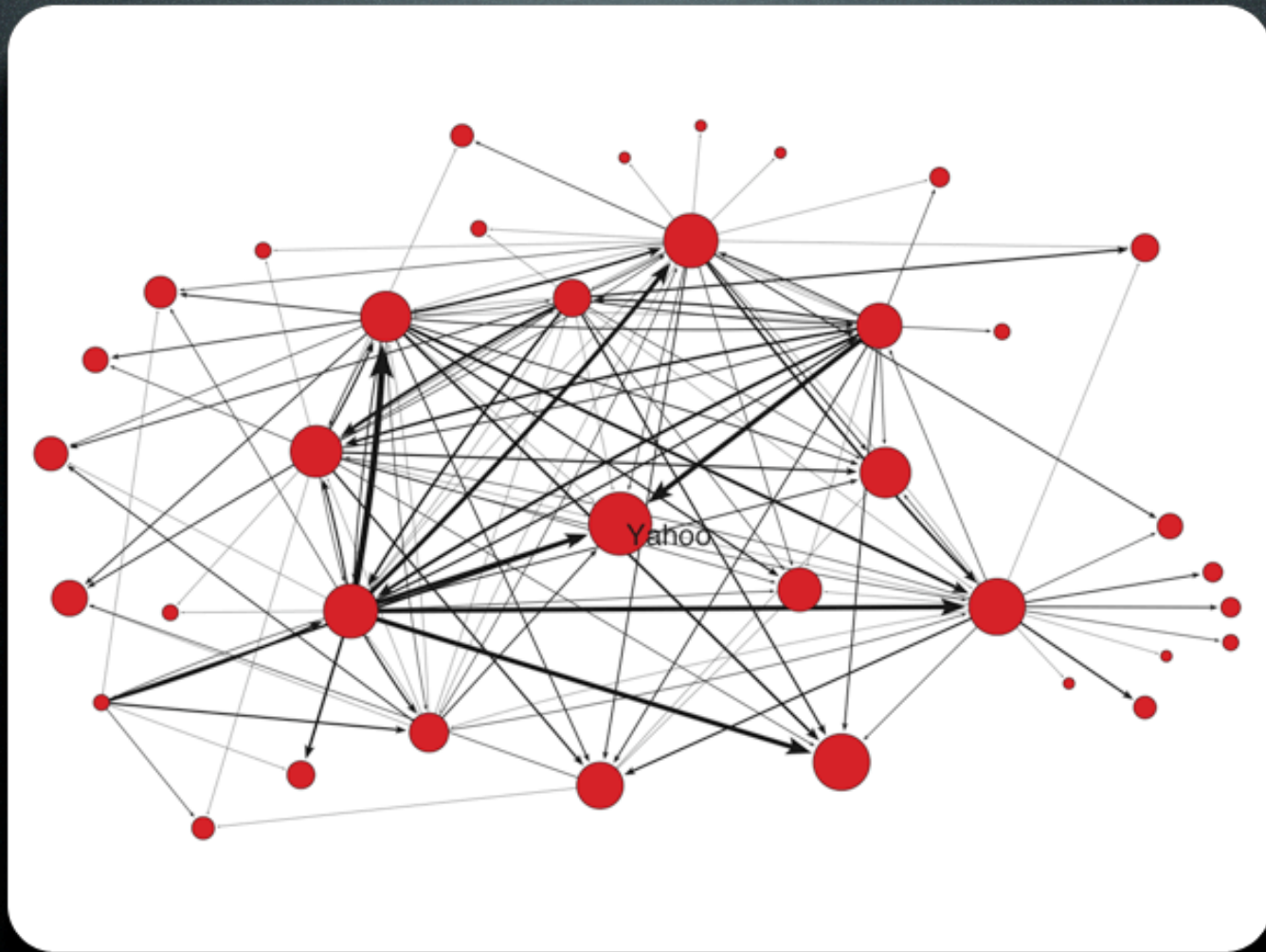
- Improve coverage/diversity in query routing algorithm
- Spam protection: trust/reputation subsystem
- User study with 6S application





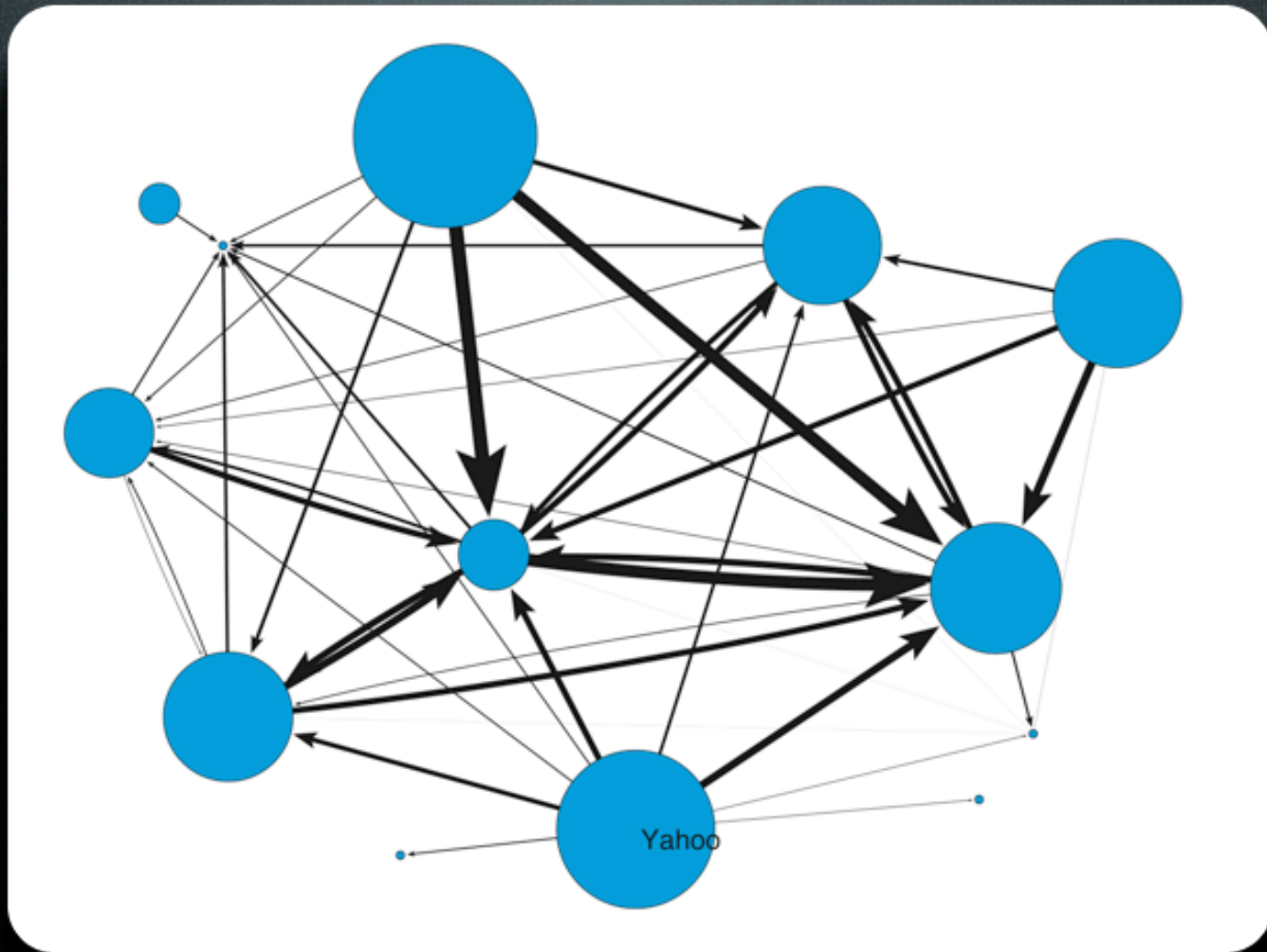
User study





Query network





Result network



# Questions?

<http://homer.informatics.indiana.edu/6S/>

6S peer search system

http://sulu.informatics.indiana.edu/6S/network/index.html

## 6S: P2P Web index collecting and sharing System

Home Members Download Publications Getting Started Network Viewer Forum

Mouse over a node to see peer information. Mouse over an edge to see queries. Click on the background for zooming to fit. Scroll or hold right mouse button for zooming. Drag the background for panning. The graph is generated by [Prefuse](#).

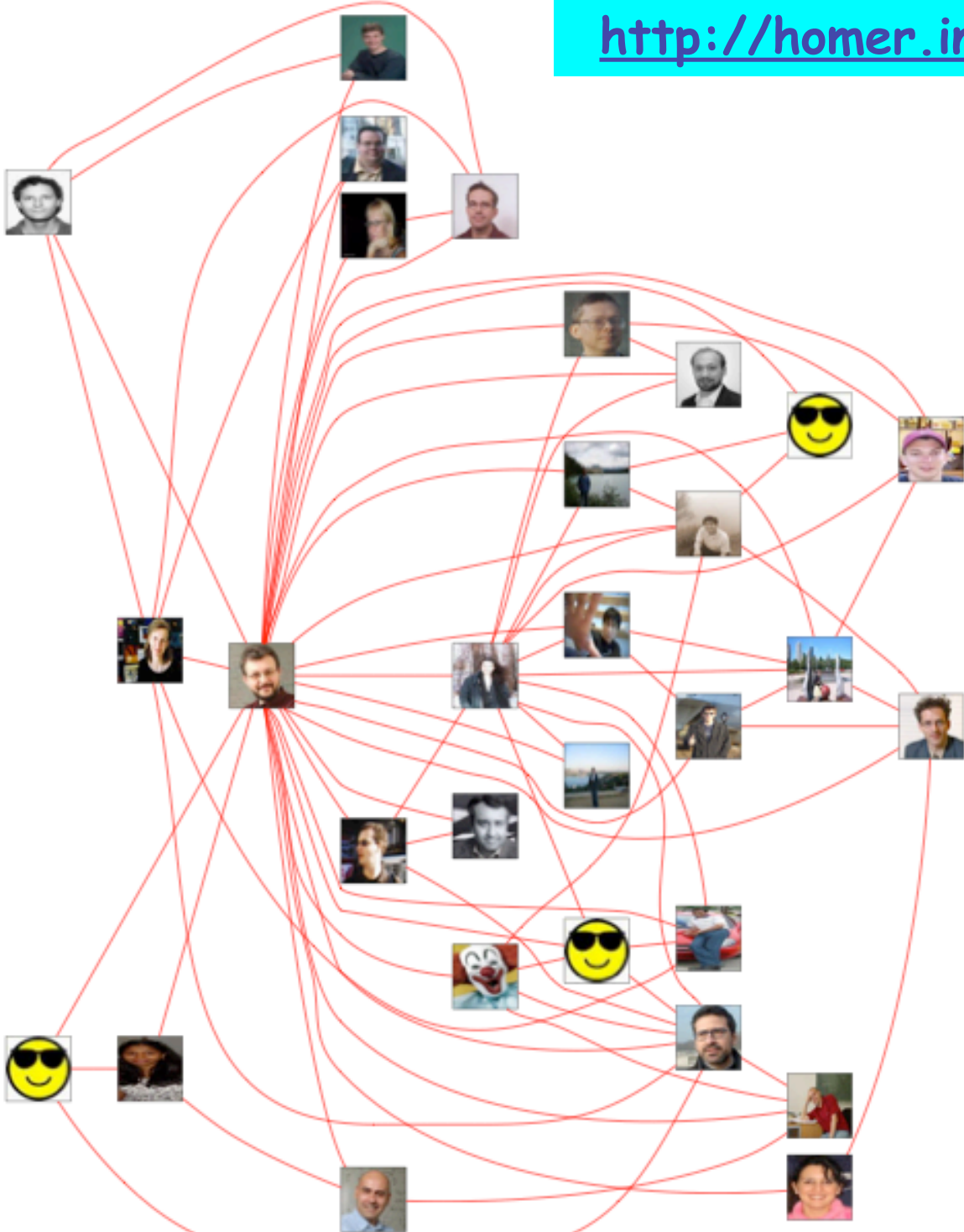
Peer Name: fil@isi  
Queries sent: 14  
Last active: 3 hours ago  
Collection: 1553 pages

Peer Name: testing  
Peer Name: bebe  
Peer Name: debug  
Peer Name: eigmac  
Peer Name: angel  
Peer Name: Peer#53  
Peer Name: Peer#32  
Peer Name: Peer#42  
Peer Name: Peer#40  
Peer Name: Peer#50  
Peer Name: Peer#54  
Peer Name: Peer#55  
Peer Name: Peer#59  
Peer Name: Peer#16  
Peer Name: wccwin

Indiana University Computer Science Department  
Comments: {rakavipa, lewu} at cs dot indiana dot edu.  
© Indiana University Page Last Updated: September 20, 2007

This material is based upon work supported by the National Science Foundation under award N. IIS-0133124 and IIS-0348940. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

<http://homer.informatics.indiana.edu/~nan/>



THANK  
YOU

