

Social Web Search

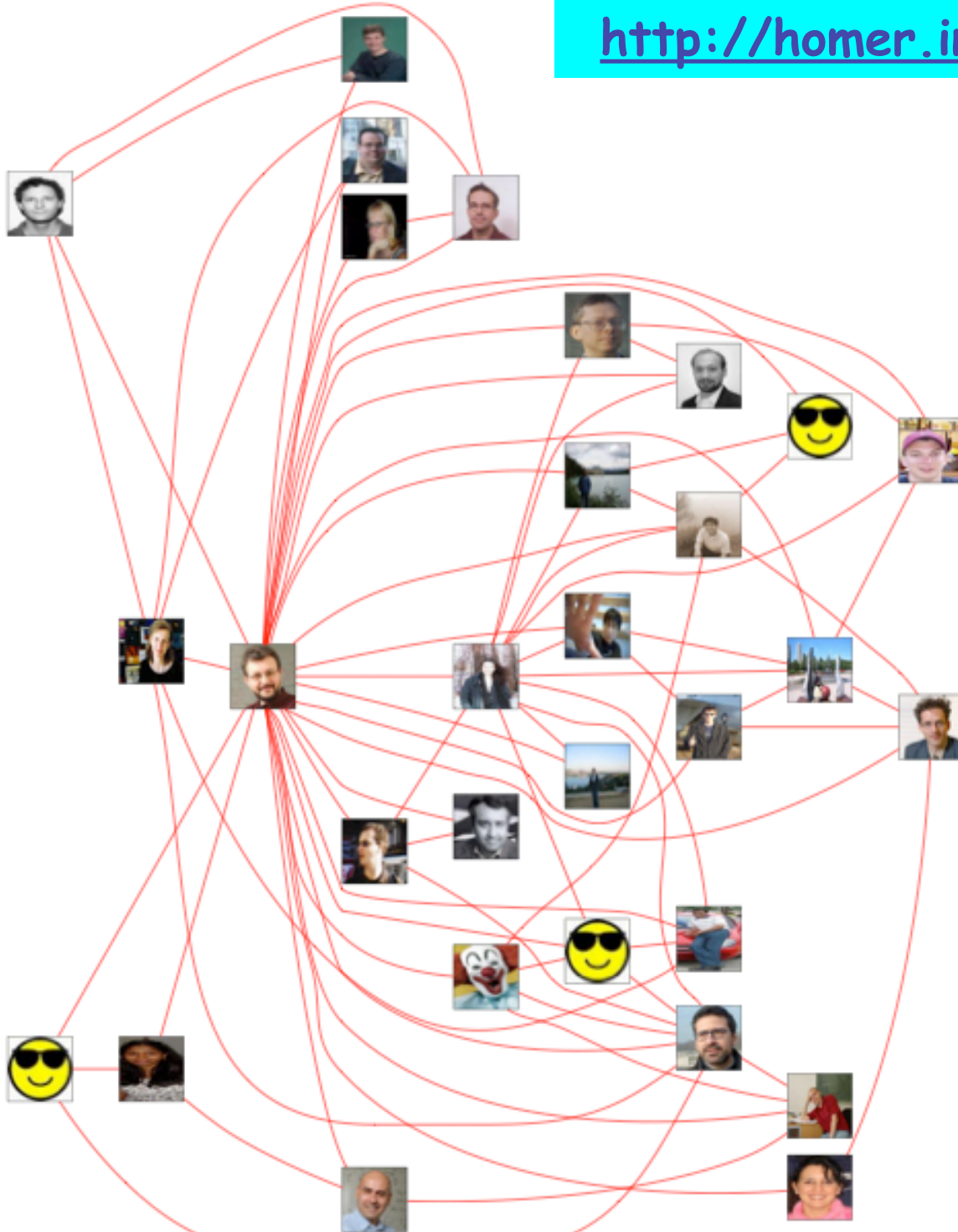
Filippo Menczer

Informatics and Computer Science

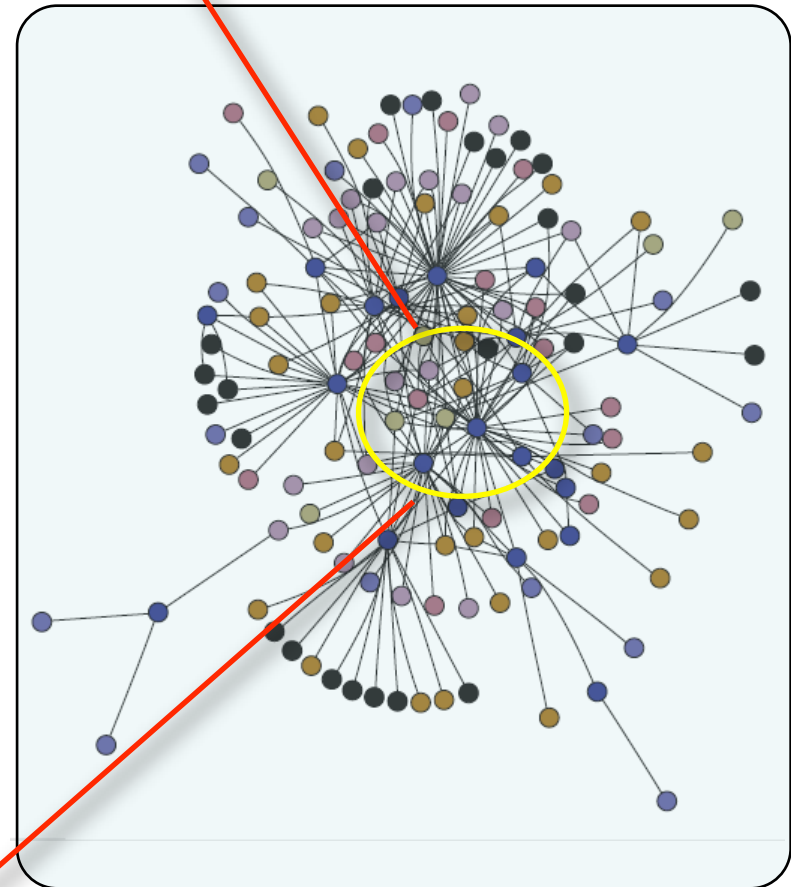
Leveraging online social behavior
for collaborative Web search



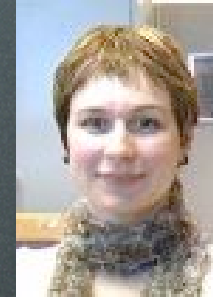
<http://homer.informatics.indiana.edu/~nan/>



NaN:
Networks
& agents
Network



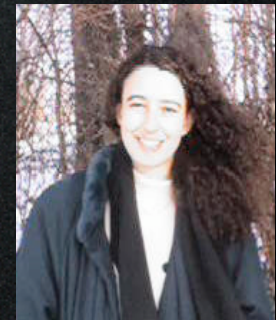
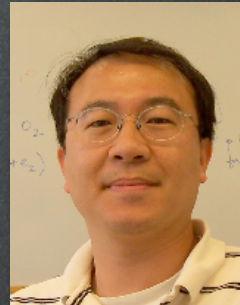
Outline



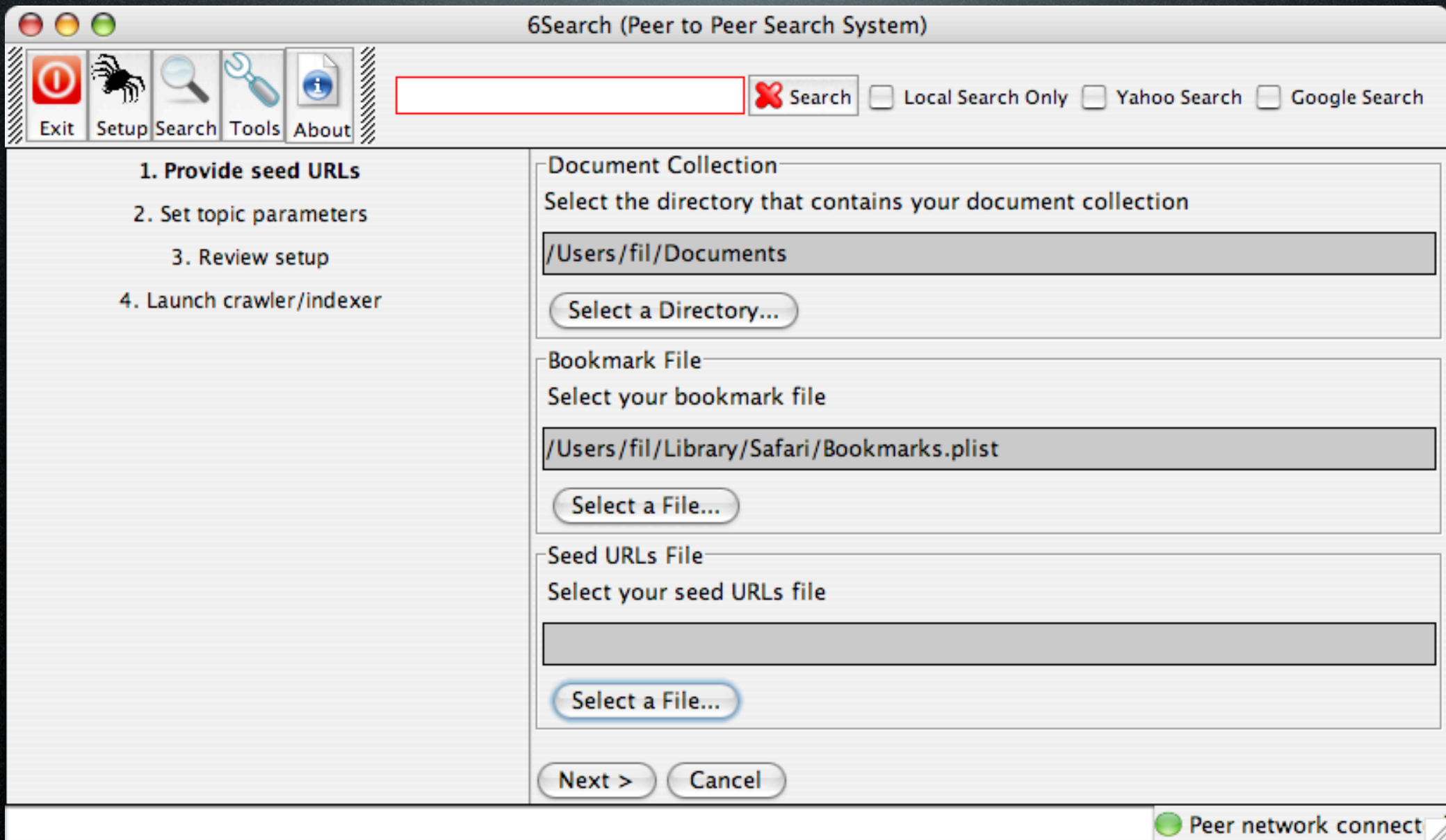
- Part 1: GiveALink.org (last semester)
- Part 2: 6S

Outline

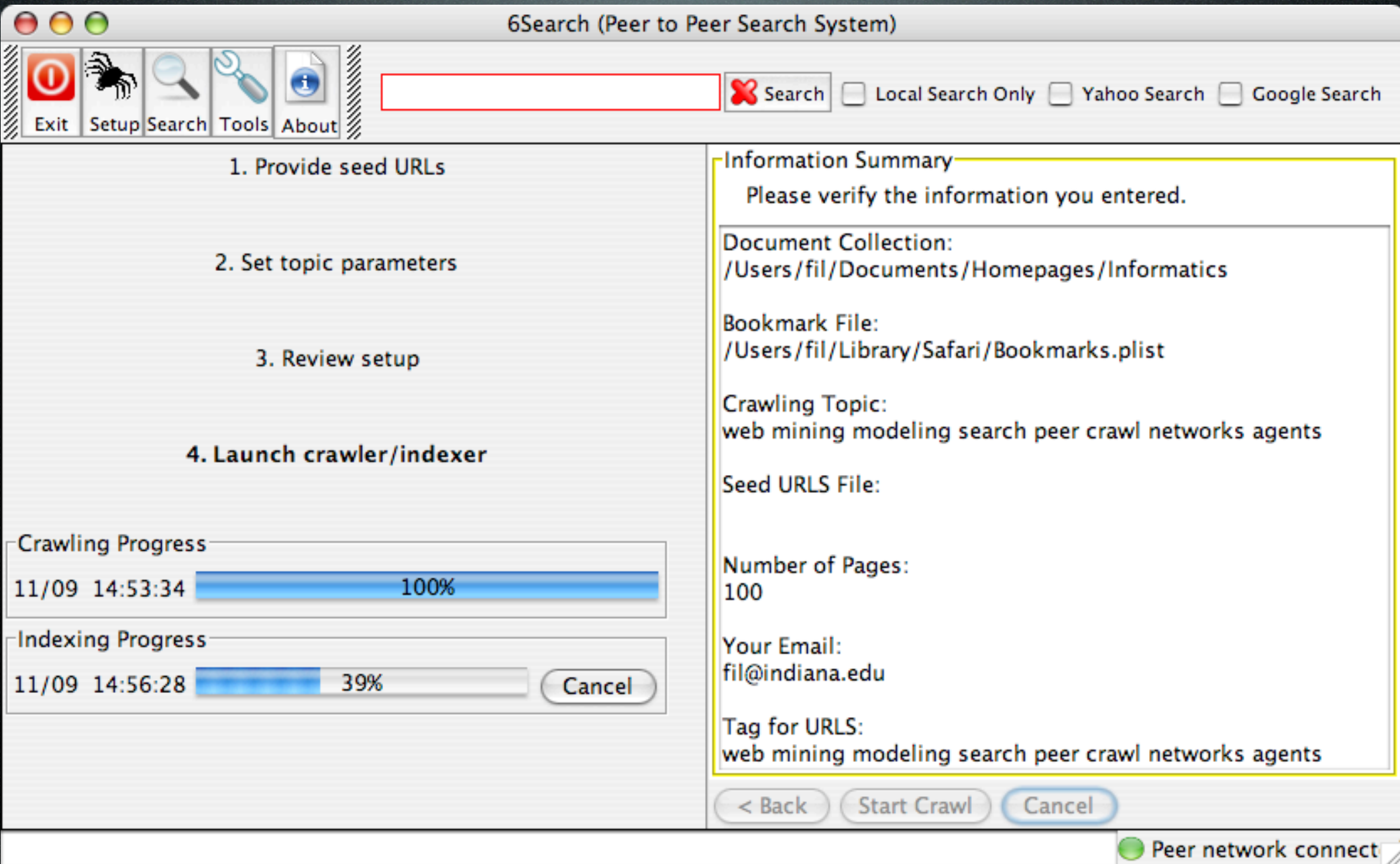
- Part 1: GiveALink.org
- Part 2: 6S



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



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



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

6Search (Peer to Peer Search System)

Exit Setup Search Tools About

☐ Local Search Only ☐ Yahoo Search ☐ Google Search

Indexing Date	Document Title	Document Url	Document Tag
2006/11/09 ...	ClickZ Internet Marketing Solutions for Marketers	http://www.clickz.com/	web mining modeling...
2006/11/09 ...		http://www.di.fc.ul.pt/sobre/bibtex.php?entr...	web mining modeling...
2006/11/09 ...		http://wwwconf.ecs.soton.ac.uk/perl/oai2?ve...	web mining modeling...
2006/11/09 ...	Geeking with Greg	http://glinden.blogspot.com/	web mining modeling...
2006/11/09 ...	Crawling the Web Speakers	http://www.roundtable.com/speakersbureau...	web mining modeling...
2006/11/09 ...	Things an Observatory should have (institutiona...	http://www.sdsc.edu/NEON/jan2000/broadc...	web mining modeling...
2006/11/09 ...	Luminous Landscape Forum (Powered by Invisio...	http://www.luminous-landscape.com/forum/	web mining modeling...
2006/11/09 ...	VDS Technologies - Map Data	http://www.vdstech.com/map_data.htm	web mining modeling...
2006/11/09 ...	Mark Pincus Blog	http://markpincus.typepad.com/	web mining modeling...
2006/11/09 ...	W3C I18N Tutorial: CSS3 and International Text	http://www.w3.org/International/tutorials/css...	web mining modeling...
2006/11/09 ...	OnoTech	http://onotech.blogspot.com/	web mining modeling...
2006/11/09 ...	Search	http://www.roundtablegroup.com/about/sear...	web mining modeling...
2006/11/09 ...	First Tuesdays at the Urban Institute 03/2002	http://www.urban.org/Presentations/FT/FT3-...	web mining modeling...
2006/11/09 ...	Clusty Cloud Creator	http://cloud.clusty.com/	web mining modeling...
2006/11/09 ...	ピッツバーグ便り	http://mawaha.exblog.jp/	web mining modeling...
2006/11/09 ...	Powerset: See what people are saying right now...	http://www.technorati.com/tag/Powerset/	web mining modeling...
2006/11/09 ...	AI Resources and Links	http://labs.rightnow.com/resources.php	web mining modeling...
2006/11/09 ...	The Next Net: The Search for the Next Big Leap...	http://business2.blogs.com/business2blog/2...	web mining modeling...
2006/11/09 ...	slides.dvi	http://www.comp.hkbu.edu.hk/WI05/speake...	web mining modeling...
2006/11/09 ...	Amazon.com: The Search: How Google and Its ...	http://www.amazon.com/Search-Rewrote-Bu...	web mining modeling...
2006/11/09 ...	Industry Solutions from RightNow Technologies ...	http://www.rightnow.com/products/industry-...	web mining modeling...
2006/11/09 ...	ODP - Open Directory Project	http://www.dmoz.org/	web directory

Advanced Tool

Delete row Undelete All Delete All

Peer network connect

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

The screenshot shows a web browser window titled "6S Search (Peer to Peer Search System)". The interface includes a top navigation bar with icons for Exit, Setup, Search, Tools, and About. A search input field contains the text "collaborative web search", followed by a green "Search" button. To the right of the search field are checkboxes for "Local Search Only" (unchecked), "Yahoo Search" (checked), and "Google Search" (checked). Below the search bar are three tabs: "6S search" (selected), "Google search", and "Yahoo search". The main content area displays a list of search results, each with a title, a brief description, a URL, and a "Similar pages (Powered by GiveALink)" link. The results are as follows:

- Information Retrieval & Extraction**
... the future." Also hear: **Web Search** as a Force ... to use this site **Search** ...
<http://www.aaai.org/AITopics/html/info.html> [Similar pages \(Powered by GiveALink\)](#)
Contributors: Unknown#10,wls_iceman#5
- BUBL LINK: Information retrieval**
... of Internet Resources Home | **Search** | Subject Menus | Countries | Types | BUBL ...
<http://bubl.ac.uk/link/i/informationretrieval.htm> [Similar pages \(Powered by GiveALink\)](#)
Contributors: Unknown#10,wls_iceman#5
- Autonomy**
... Technical Benefits Conceptual Retrieval Enterprise **Search** Entry-Level **Search** - Ultraseek Legacy **Search** Categorization Channels
Clustering Taxonomy Generation ...
<http://www.verity.com/content/home/> [Similar pages \(Powered by GiveALink\)](#)
Contributors: wls_dell2#8
- MacËOSËX**
... Mac OS X Advanced **Search** Restrict to Mac ...
<http://developer.apple.com/macosx/> [Similar pages \(Powered by GiveALink\)](#)
Contributors: wls_XP2#3
- Amazon.com: Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management: Book**
... item Make a **search** suggestion | Manage your **search** suggestions Tag this product (What ... Cart | Your Lists | Help | Advanced ...
<http://www.amazon.com/Data-Mining-Techniques-Relationship-Management/dp/0471470643> [Similar pages \(Powered by GiveALink\)](#)
Contributors: wls_angel#6
- UMBC Agent Web -- news and information on software agent technology**
... retrieval and knowledge management **Search web** for similar pages ... UMBC Agent **Web** -- news and information ...
http://agents.umbc.edu/Topics/Related_Topics/Information_retrieval_and_knowledge_management/ [Similar pages \(Powered by GiveALink\)](#)
Contributors: Unknown#10

At the bottom of the window, there is a status bar with the text "Done" on the left and a green icon followed by "Peer network connect" on the right.

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

The search results of 6S system

http://localhost:1999/result/ 6S web mining

Mozilla.org Plug-in FAQ CSN givealink.org Oncourse OneStart CenSEARCHip MySpiders Wind

[KDnuggets: Data Mining, Web Mining, and Knowledge Discovery Guide](#)
Newsletter on the data mining and knowledge industries, offering information on data mining, know
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
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/
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/
Contributors: Unknown#7,wls_iceman#1
[Similar pages \(Power by GiveALink\)](#)

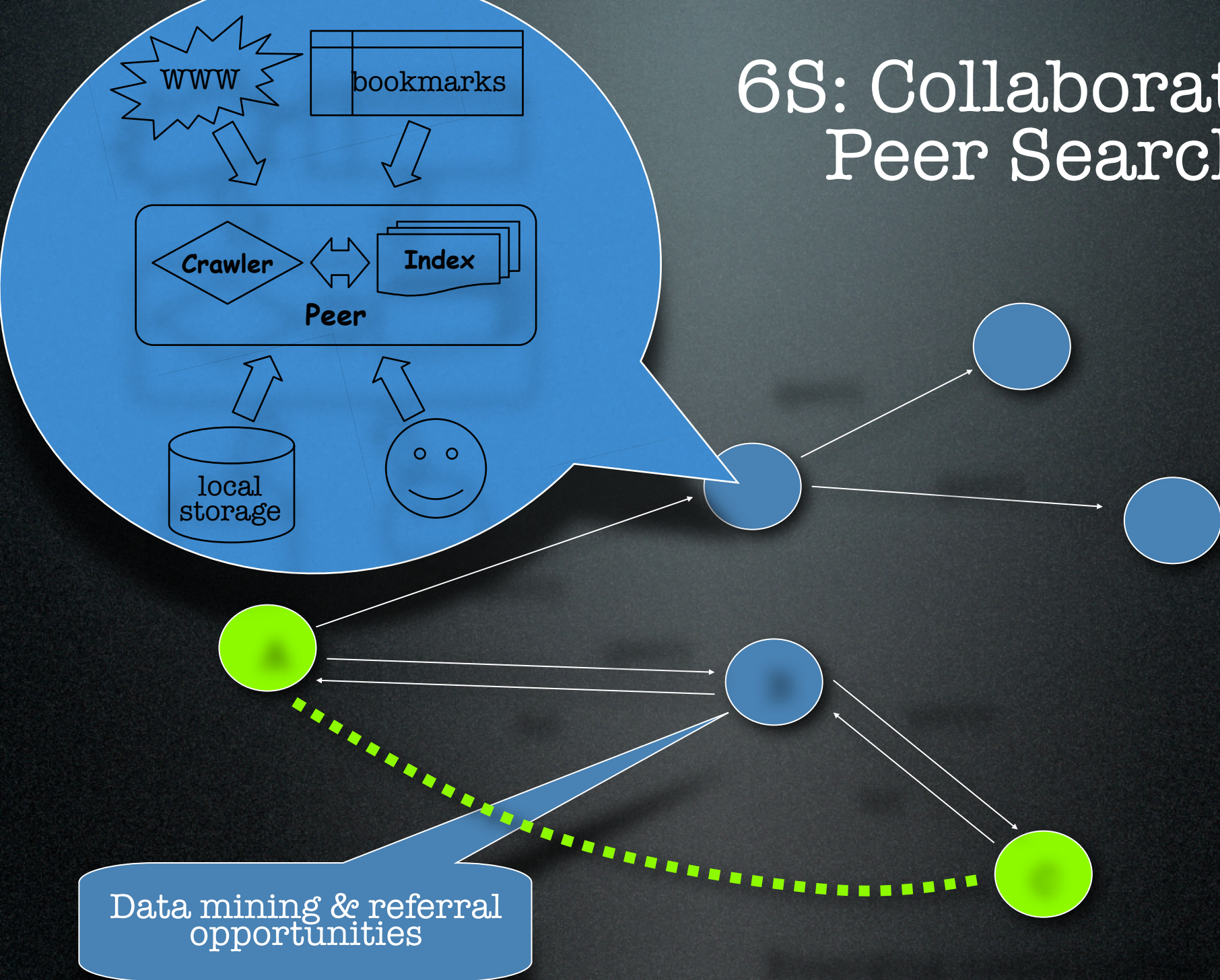
[BUBL LINK: Information retrieval](#)
... and Forward Knowledge Approach **Data Mine: Data Mining** and Knowledge Discovery ... the Gaps KD Mine: **Data** ...
<http://bubl.ac.uk/link/i/informationretrieval.htm>
Contributors: Unknown#7,wls_iceman#1
[Similar pages \(Power by GiveALink\)](#)

[Library and Information Science > Information Retrieval in the Yahoo! Directory](#)

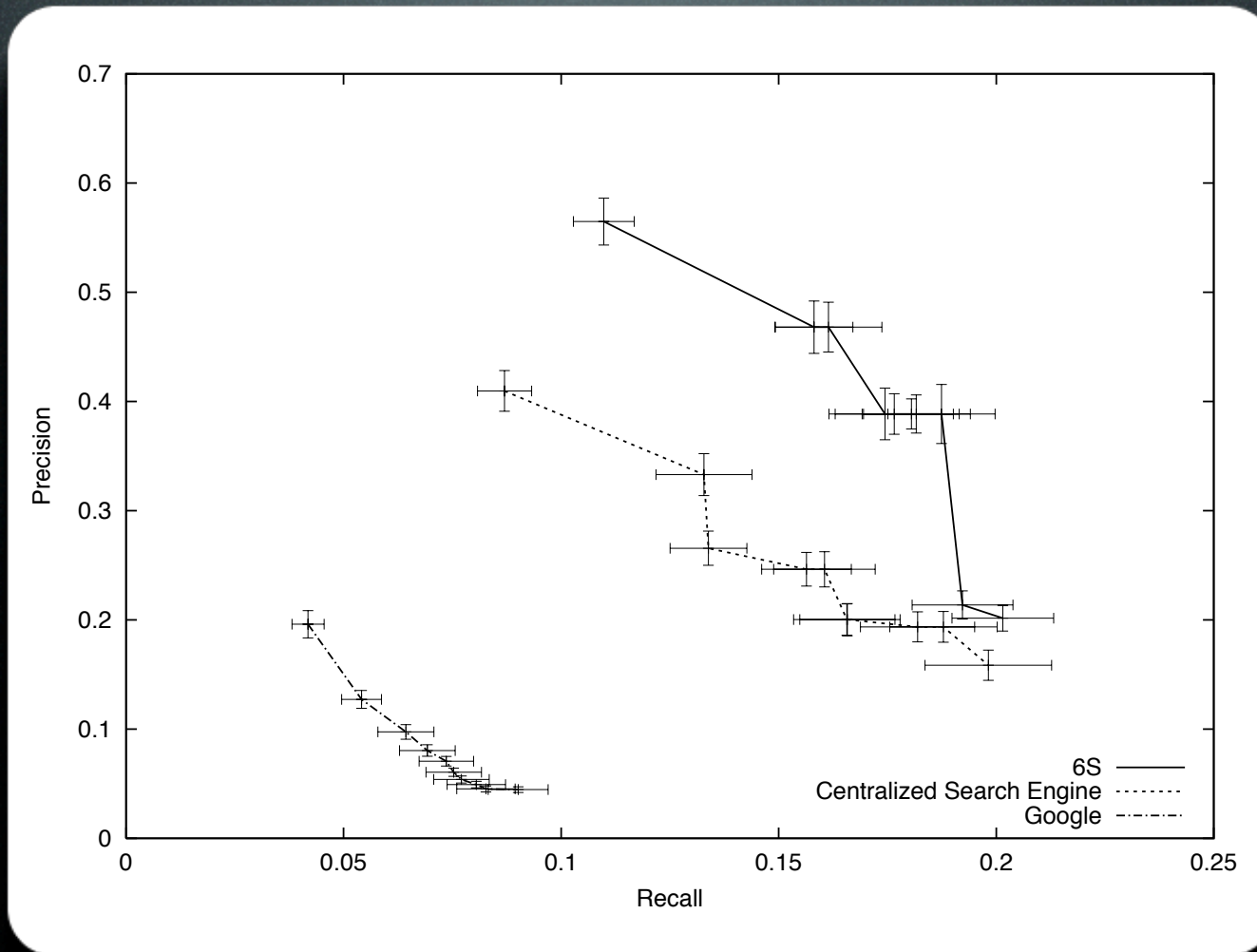
Done

Google
Yahoo
6S 6Search (local)
6S 6Search (peers)
Amazon.com
Answers.com
Creative Commons
eBay
Manage Search Engines...

6S: Collaborative Peer Search

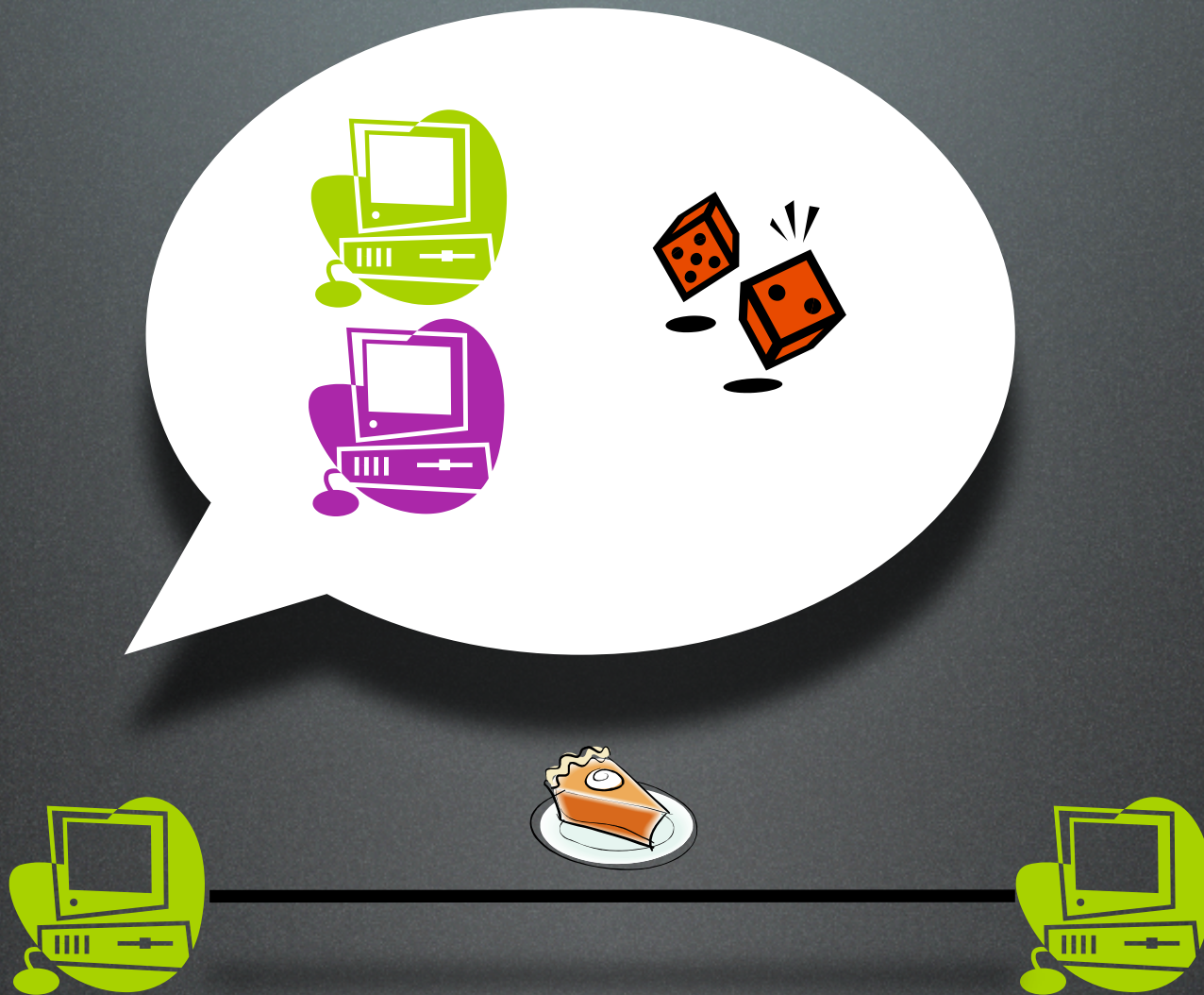


WWW2004
WWW2005
WTAS2005
P2PIR2006

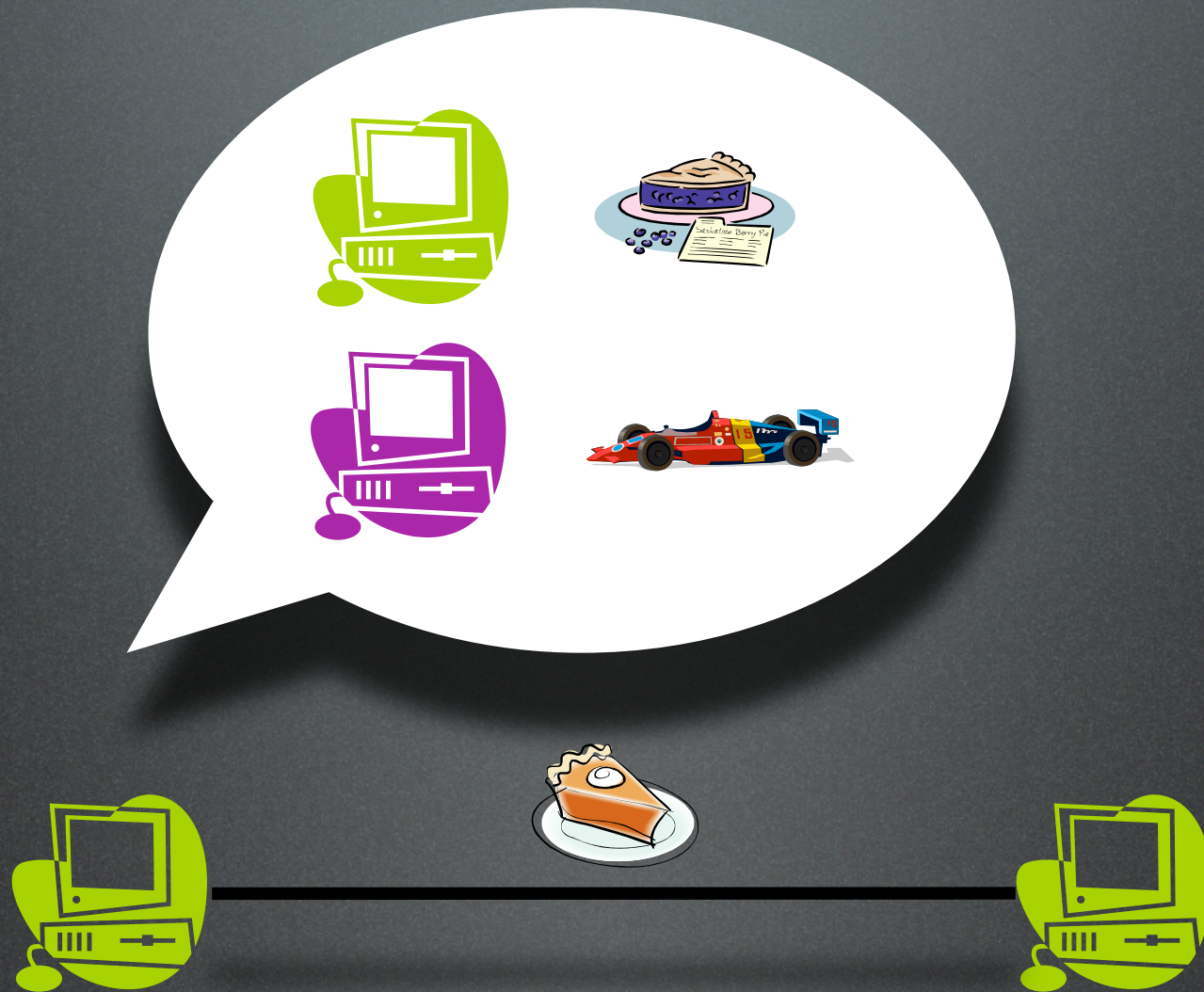


Distributed vs Centralized

Algorithm 1: Random Known



Algorithm 2: Greedy



Algorithm 3: Reinforcement



Learning Peers

- A focused profile W^f of other peers for terms in queries
- An expanded profile W^e of other peers for terms not in queries, but that co-occur with query terms and with higher frequency

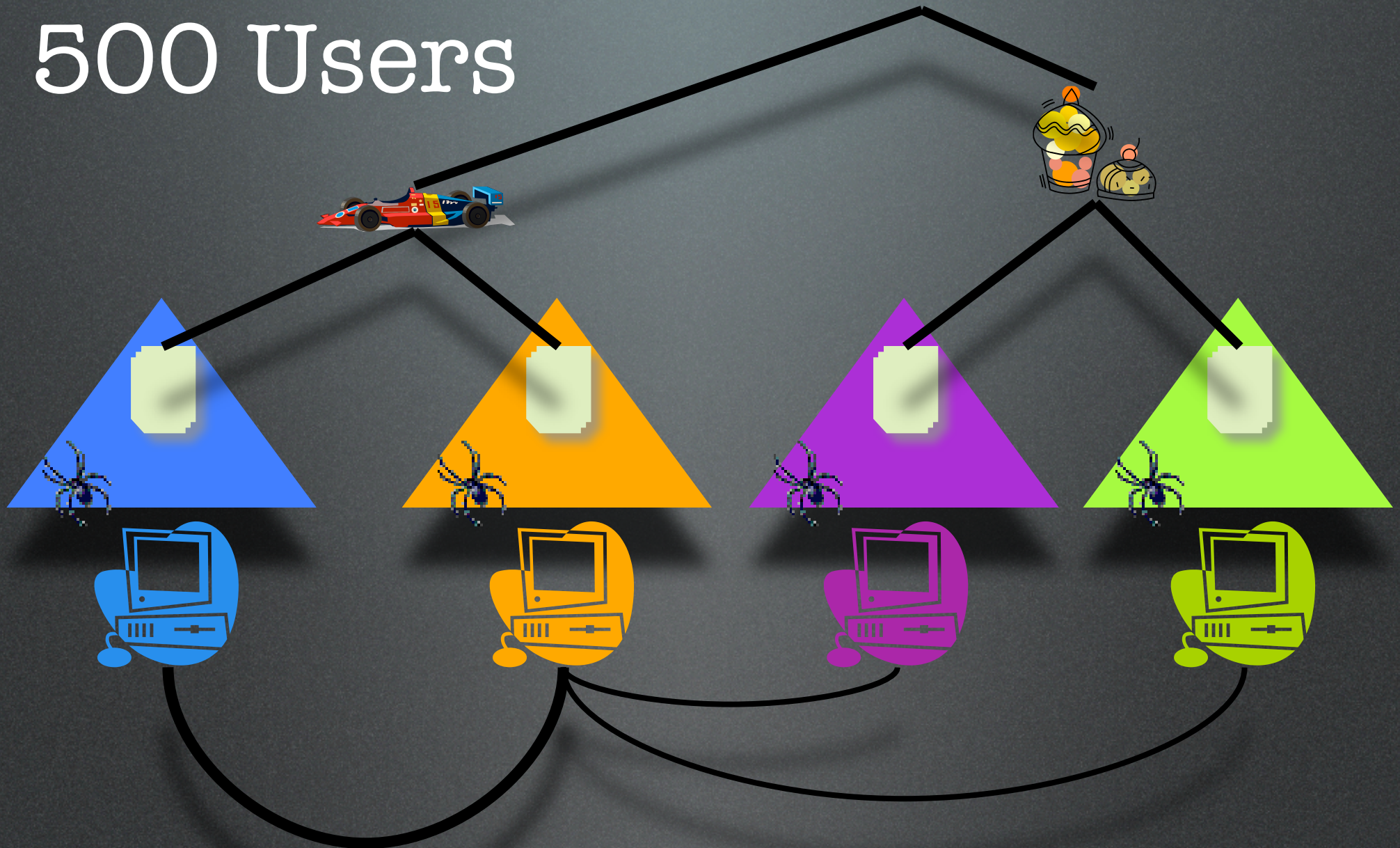
$$w_{i,p}(t+1) = (1 - \gamma) \cdot w_{i,p}(t) + \gamma \cdot \left(\frac{S_p + 1}{S_l + 1} - 1 \right)$$

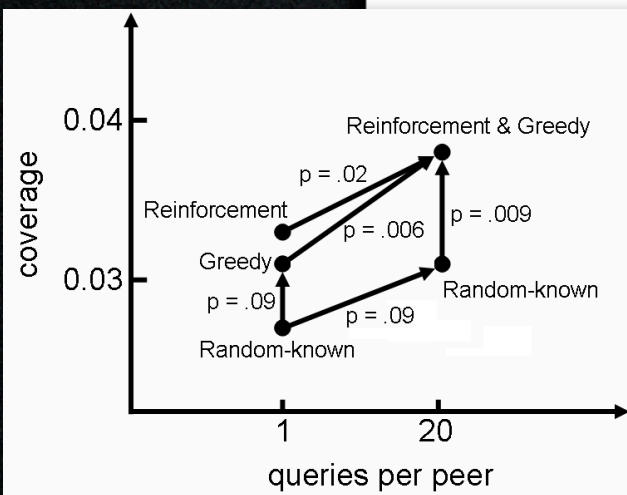
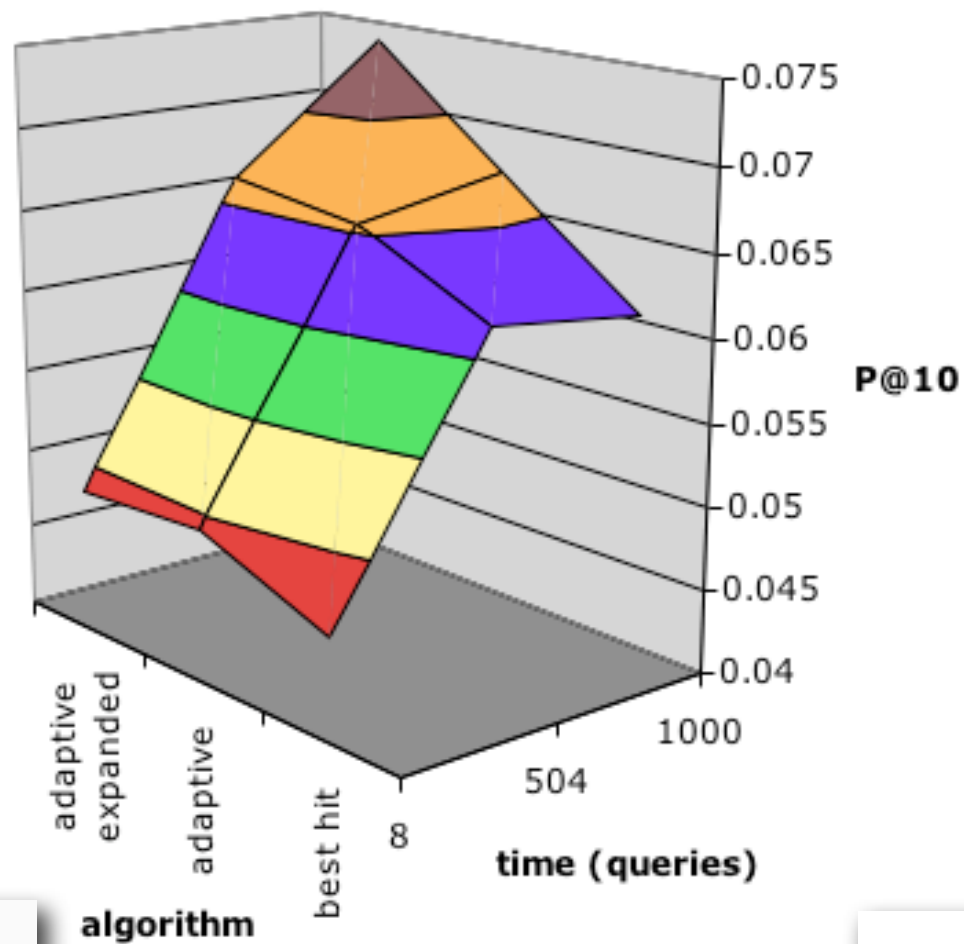
- Query routing:

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

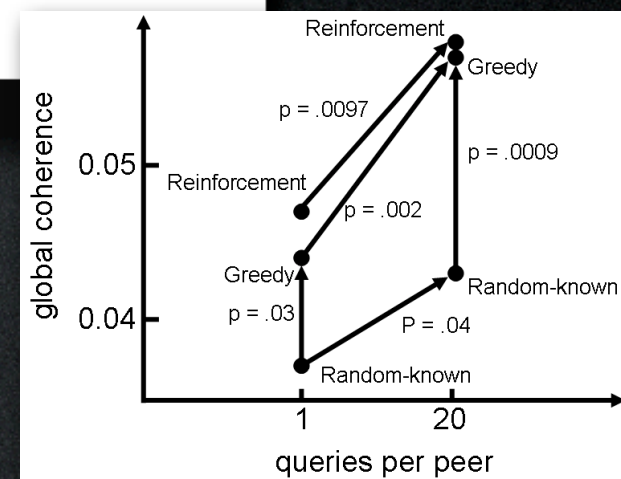
Simulating 500 Users

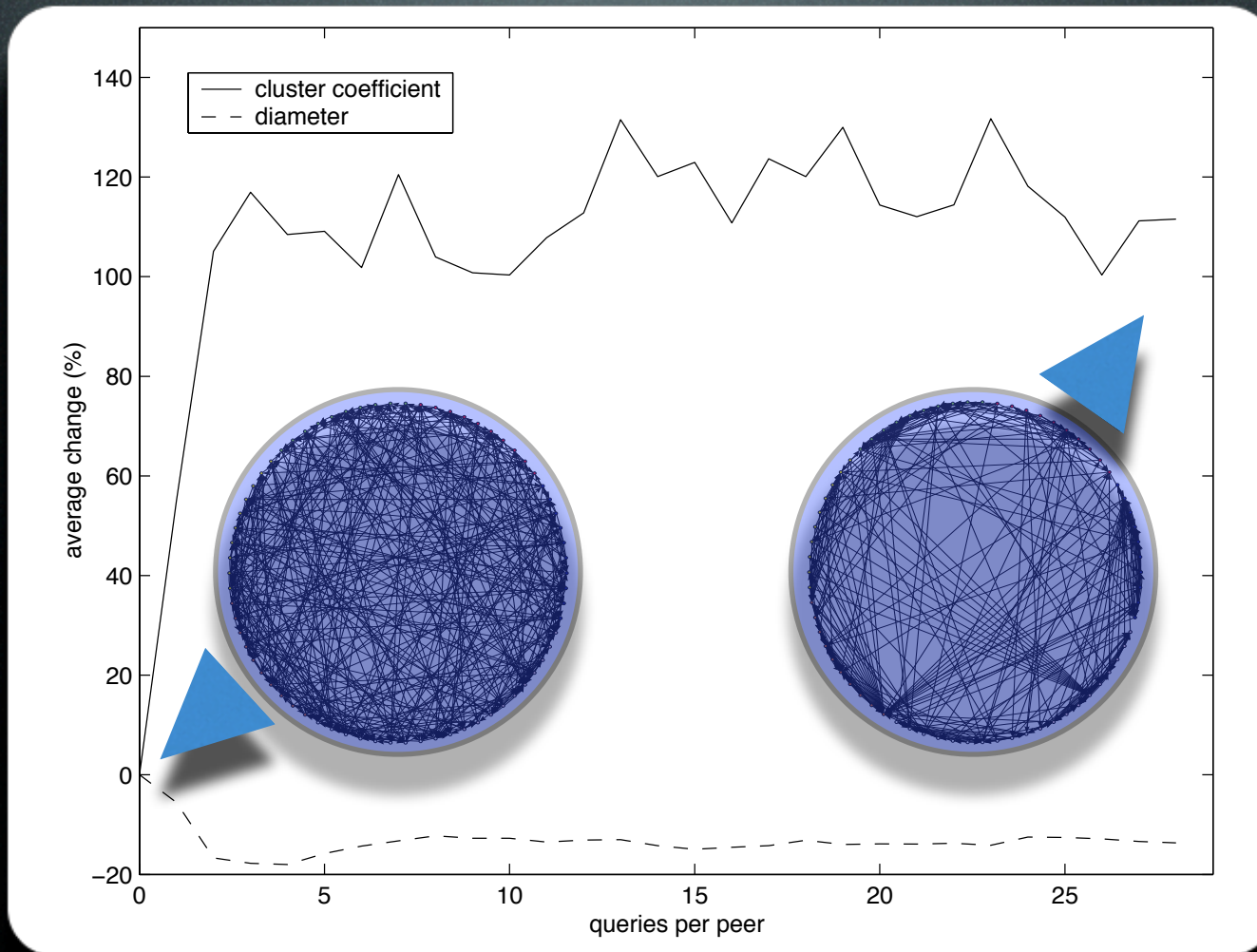
ODP (dmoz.org)





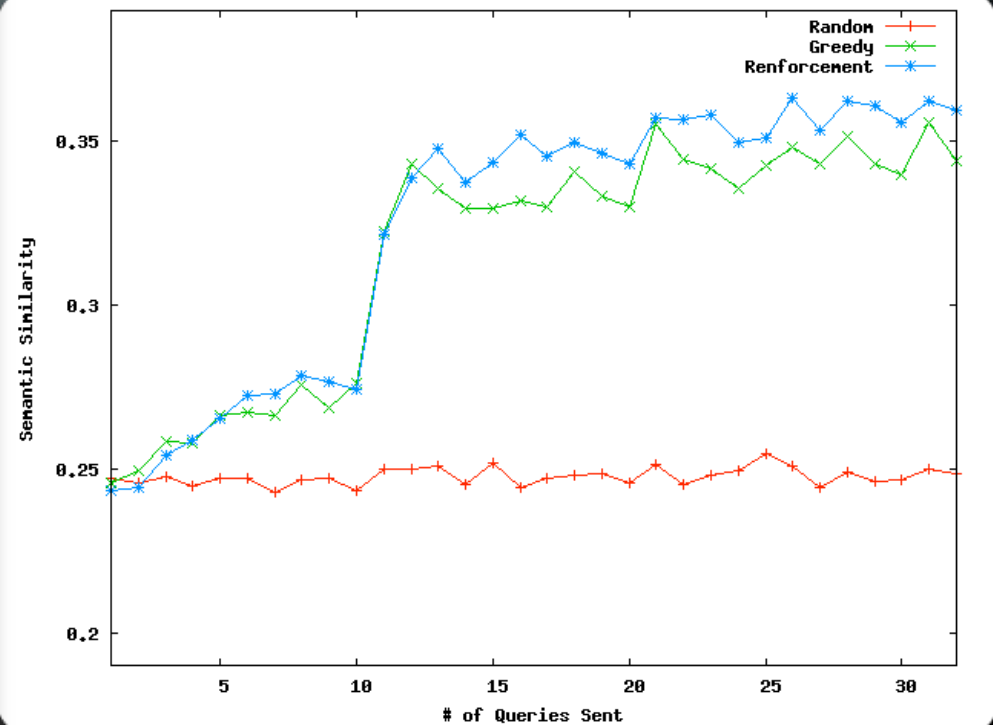
P@10





Small-world

Semantic Similarity



Ongoing Work

- TREC based evaluation framework (Klampanos et al. 2005)
- User study with 6Search application
- Improve query routing algorithm
 - Coverage
- Spam protection

Questions?

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

