

Machine Learning Techniques in Social Computing (社会计算)

Irwin King

king@cse.cuhk.edu.hk

<http://www.cse.cuhk.edu.hk/~king>

Department of Computer Science & Engineering
The Chinese University of Hong Kong



Alexa as of Nov. 2008	USA	CHINA	Global
1	Google	Baidu	Yahoo
2	Yahoo	QQ	Google
3	Myspace	Sina	YouTube
4	YouTube	Google.cn	Windows Live
5	Facebook	Taobao	Facebook
6	Windows Live	163	MSN
7	MSN	Yahoo	Myspace
8	Wikipedia	Google	Wikipedia
9	EBay	Sohu	Blogger
10	AOL	Youku	Yahoo.jp



What's On the Menu?

- Web 2.0 and Social X
- Social Computing (社会计算)
- Some Interesting Problems
 - Collaborative Filtering (協同過濾)
 - Query Suggestion (查询建议)



What's On the Menu?

- Web 2.0 and Social X
- Social Computing (社会计算)
- Some Interesting Problems
 - Collaborative Filtering (協同過濾)
 - Query Suggestion (查询建议)



Web 2.0

- Web as a medium vs. **Web as a platform**
- Read-Only Web vs. **Read-and-Write Web**
- Static vs. **Dynamic**
- Restrictive vs. **Freedom & Empowerment**
- Technology-centric vs. **User-centric**
- Limited vs. **Rich User Experience**
- Individualistic vs. **Group/Collective Behavior**
- Consumer vs. **Producer**
- Transactional vs. **Relational**
- Top-down vs. **Bottom-up**
- People-to-Machine vs. **People-to-People**
- Search & browse vs. **Publish & Subscribe**
- Closed application vs. **Service-oriented Services**
- Functionality vs. **Utility**
- Data vs. **Value**



Social Networking



The screenshot shows the Facebook profile of Irwin King. The profile includes a profile picture, a cover photo, and a bio. The bio states: "What are you doing right now? Networks: CUHK Faculty, Sex: Male, Hometown: Taipei, Taiwan, Religious Views: Christian". The Mini-Feed shows several updates, including a new address, profile edits, and new friendships. The Information section lists contact info, education, and work history. The Friends section shows 3 friends: Chi Chung Chan, Tom Hung, and David Shepherd.

facebook Profile edit Friends Networks Inbox home account privacy logout

Search

Applications edit

- Photos
- Video
- Groups
- Events
- Marketplace

more

Irwin King
What are you doing right now?

Networks: CUHK Faculty
Sex: Male
Hometown: Taipei, Taiwan
Religious Views: Christian

Mini-Feed
Displaying 7 stories. See All

Today

- Irwin has a new address: Department of Computer Science and Engineering, The Chinese University of Hong Kong. 8:14pm
- Irwin edited Education Info, Hometown, Work Info, Website and Religious Views in his profile, and changed his profile picture. 8:14pm
- Irwin and Chi Chung Chan are now friends. 7:54pm
- Irwin and Tom Hung are now friends. 7:54pm
- Irwin and David Shepherd are now friends. 7:54pm
- Irwin joined the CUHK network. 7:54pm
- Welcome to Facebook! We're glad you're here. 7:54pm

Information

Contact Info
Email: king@cse.cuhk.edu.hk
Current Address: Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, Hong Kong
Website: http://www.cse.cuhk.edu.hk/~king

Education and Work

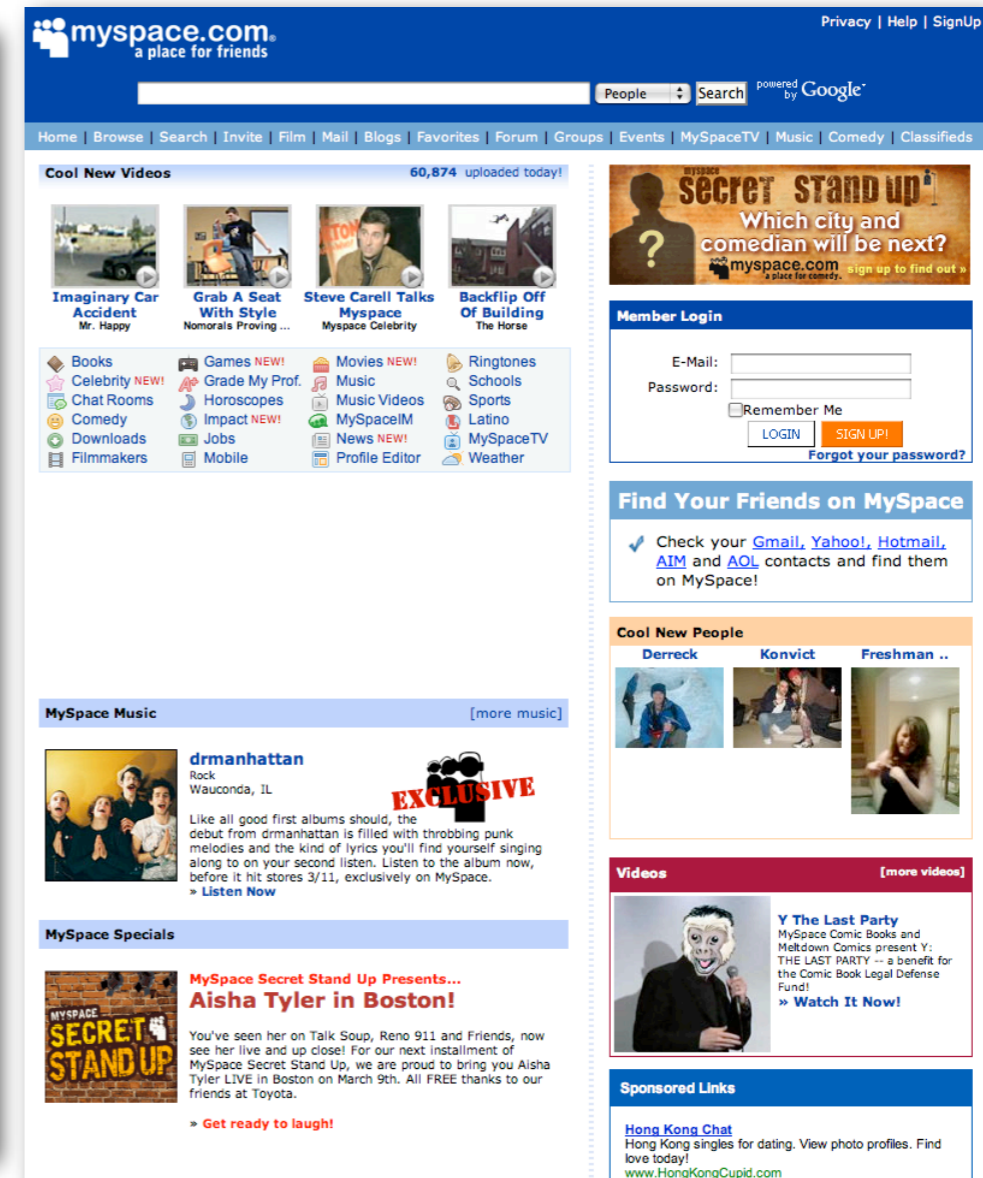
Education Info
Grad Schools: University Of Southern California '88 M.Sc., Computer Science
University Of Southern California '03 Ph.D., Computer Science

Friends
3 friends. See All

Chi Chung Chan Tom Hung David Shepherd

Search for friends
Use the Friend Finder

Friends in Other Networks
Networks with the most friends
Hong Kong (2)
Show All Networks | View All Friends



The screenshot shows the MySpace profile of drmanhattan. The profile includes a profile picture, a cover photo, and a bio. The bio states: "Like all good first albums should, the debut from drmanhattan is filled with throbbing punk melodies and the kind of lyrics you'll find yourself singing along to on your second listen. Listen to the album now, before it hits stores 3/11, exclusively on MySpace. Listen Now". The profile also features a "Cool New Videos" section, a "MySpace Music" section, and a "MySpace Specials" section. The "MySpace Specials" section features a video titled "MySpace Secret Stand Up Presents... Aisha Tyler in Boston!".

myspace.com a place for friends Privacy | Help | SignUp

Search powered by Google

Home Browse Search Invite Film Mail Blogs Favorites Forum Groups Events MySpaceTV Music Comedy Classifieds

Cool New Videos 60,874 uploaded today!

- Imaginary Car Accident Mr. Happy
- Grab A Seat With Style Nomorals Proving ...
- Steve Carell Talks Myspace Myspace Celebrity
- Backflip Off Of Building The Horse

Member Login

E-Mail:
Password:
 Remember Me
LOGIN SIGN UP!
Forgot your password?

Find Your Friends on MySpace

Check your [Gmail](#), [Yahoo!](#), [Hotmail](#), [AIM](#) and [AOL](#) contacts and find them on MySpace!

Cool New People

Derreck Konvict Freshman ..

MySpace Music [more music]

drmanhattan
Rock
Wauconda, IL

EXCLUSIVE

Like all good first albums should, the debut from drmanhattan is filled with throbbing punk melodies and the kind of lyrics you'll find yourself singing along to on your second listen. Listen to the album now, before it hits stores 3/11, exclusively on MySpace.
Listen Now

MySpace Specials

MySpace Secret Stand Up Presents... Aisha Tyler in Boston!

You've seen her on Talk Soup, Reno 911 and Friends, now see her live and up close! For our next installment of MySpace Secret Stand Up, we are proud to bring you Aisha Tyler LIVE in Boston on March 9th. All FREE thanks to our friends at Toyota.
Get ready to laugh!

Videos [more videos]

Y The Last Party
MySpace Comic Books and Meltdown Comics present Y: THE LAST PARTY -- a benefit for the Comic Book Legal Defense Fund!
Watch It Now!

Sponsored Links

[Hong Kong Chat](#)
Hong Kong singles for dating. View photo profiles. Find love today!
[www.HongKongCupid.com](#)

Machine Learning Techniques for Social Computing, Irwin King, MLA'08, Nanjing, China, November 8-9, 2008



Social Search

- Social Search Engine
- Leveraging your social networks for searching

eurekasterswicki login | sign up

build new swicki swicki directory about swickis about eurekaster

Search and vote for your faves

swicki search

a custom search portal around the topic of your choice powered by your community

Build a swicki!

A swicki is a custom social search portal on the topic of your choice. With every search, vote and click, your swicki generates more relevant results and turns into a valuable asset for you and your community. [Take a tour](#) to find out more about how swickis work.

- Choose from text, multimedia or video content
- Customize the swicki widget look and feel
- Share your swicki widget with your community

[Build a swicki](#)

Eurekaster news

Now out of beta!

- Come join the network for swicki builders
- Swicki Users Go Green
- CEO Speaking at SES New York

Get swicki illustrated
For the latest news and trends in social search, subscribe now.

Build a swicki!

Even fresher swickis with RSS and Autodetect. [Learn More.](#)

Browse the directory

Try searching one of over 100,000 swickis already created, or grab one to add to your site or blog.

Recently created

- askforkids
- e-learning et didactique ...
- denver news
- home repairs any gal can ...
- creative ideas for green ...
- easy woodworking projects...

[More >](#)

Top swickis

- techcrunch
- born2ikea
- popular science
- readwriteweb
- lockergnome
- neopets
- larkswicki

[More >](#)

DIY: home improvement swicki showcase

- Home Repairs Any Gal Can Do
- Make Yourself a Man Pad
- Making Room for Baby
- Creative Ideas for Green Home Improvement

Computers

- dot net search engl...
- php resource search
- rails on ruby
- software factories
- web 2.0 workgroup

[More >](#)

Business

- adblogging
- alternative search ...
- bubblegeneration - ...
- contextual advertis...
- digging into search
- freelance tipster
- green building reso...

[More >](#)

Home

- about color for hom...
- gardening and plant...
- home improvement se...
- homemade baby food ...
- homemaking
- salmon

[More >](#)

Regional

- amazon river
- atlanta business se...
- atlanta home and ga...
- berkeley public lib...
- pittsburgh news
- pittsburgh wedding ...
- ski tahoe

[More >](#)

delver:: liad agmon edit

My Profile | My Network

Your friends are the best source of information!
Look for information, media and people within your network

(Go)

Noa Rabiner
Noa Rabiner is connected to you directly

- This is me!
- I know this person
- Add as Connection
- Send Message



Social Bookmarking

The screenshot shows the del.icio.us homepage. At the top left is the logo and the text "del.icio.us social bookmarking". To the right is a search bar with "del.icio.us" and a "search" button, and links for "login", "register", and "help". Below the logo are three main navigation links: "» all your bookmarks in one place", "» bookmark things for yourself and friends", and "» check out what other people are bookmarking". A "learn more..." link is also present. In the center, there is a "» get started «" link. On the right side, there is a "Tags" section with a definition: "A tag is simply a word you use to describe a bookmark. Unlike folders, you make up tags when you need them and you can use as many as you like. The result is a better way to organize your bookmarks and a great way to discover interesting things on the Web." Below this is a "tags to watch" section with a "more ..." link. The main content area is titled "hotlist" and "what's hot right now on del.icio.us". It features a "HOT NOW" section with a "see also: popular | recent" link. The list includes: "Video: Twitter in Plain English | Common Craft - Explanations In Plain English" (130 people, posted by jtyerse, tags: twitter, video, howto, commoncraft, web2.0), "Home | NotchUp Beta" (212 people, posted by sokrates_af, tags: jobs, interview, career, search, job), "PrimeTimeRewind - The TV Cube" (145 people, posted by david.rothman, tags: tv, video, streaming, television, media), "The Simple Dollar » Planning a Kitchen Garden" (133 people, posted by lantzilla, tags: gardening, food, garden, cooking, vegetables), and "Office Live Workspace vs Google Docs: Feature-by-Feature Comparison - ReadWriteWeb" (135 people, posted by gariig, tags: microsoft, google, office, google_docs, live). On the right side of the hotlist, there are several category sections: "illustration" (karenklassenillustration, Dave Devries's Monster Engine, current work), "family" (Cozi, Comeeko - Creating comic strips from your photos, Let's Have More Teen Pregnancy), "living" (Home - tiny living, Eartheasy homepage, The Simple Dollar » Nourishment on a Desperate Income), "cool" (Browse Goods, rssWheel, Laptop Stand By LapDawg - A Revolutionary, Ergonomic Laptop Holder), and "itunes".



Social Media

The screenshot shows the YouTube homepage with the following sections:

- Header:** YouTube logo, "Broadcast Yourself™", navigation tabs (Home, Videos, Channels, Community), search bar, and an "Upload" button.
- Videos being watched right now...:** A row of five video thumbnails with durations: 02:13, 03:29, 01:58, 07:01, and 03:53.
- Promoted Videos:** Four video thumbnails with titles: "Think Again Awards", "Think Again Awards", "第14屆十大電視廣告頒獎典禮 - 搬出...", and "紅館觀眾向肥姐獻花".
- Featured Videos:** A list of four featured videos:
 - David Sedaris delivers a pizza:** From weaknights, 11,313 views, 5 stars, 01:01 duration. Description: "Author and NPR personality David Sedaris delivers piping hot dinner right to your door in 30 bleak anecdotes or less-- or your iro (more)".
 - Erbert and Gerbert's Candle Cannon:** From candlecannon, 109,029 views, 5 stars, 02:34 duration. Description: "See the world's largest and most powerful air vortex cannon in action - www.candlecannon.com Erbert & Gerbert's has been makin (more)".
 - Girl's Night Out:** From dianidevine, 169,435 views, 5 stars, 03:49 duration. Description: "Newly single Mary Olson (Etta Devine) goes to the grocery store to find true love. Written and directed by Gabriel Diani. Shot a (more)".
 - Lionel Neykov - Freeze My Senses:** From LionelNeykov, 150,758 views, 5 stars, 03:35 duration. Description: "Hey! If you like this song, you can download the mp3 from itunes. Just type my name in the search box, and you'll find me. You ca (more)".
- What's New:** A yellow box containing:
 - YouTube Mobile:** "New! Watch ALL YouTube videos on your mobile device".
 - Warp!** "Visually fly through YouTube videos in the Fullscreen player".
 - RSS Feeds:** "Click on the 'RSS this page' link to get fresh videos delivered".
 - SXSW on YouTube:** "For the next week and a half, the SXSW festival is taking over Austin, Texas, to celebrate music, film and all things interactive. Read more in our Blog".

The screenshot shows the Flickr homepage with the following elements:

- Header:** Flickr logo, "Sign In", and "Create Your Account" button.
- Main Content:** A large photo of a small plant growing in a crack in the pavement, credited to "By Jam343". Below it, text says "3,602 photos uploaded in the last minute · 558,832 photos tagged with urban · 2.2 million photos geotagged this month · Take the tour".
- Navigation:** A search bar and a "SEARCH" button.
- Footer:** Four icons representing different features: "Share & stay in touch", "Upload & organize", "Make stuff!", and "Explore...".

The screenshot shows the Second Life website with the following elements:

- Header:** "SECOND LIFE" logo, "Your World. Your Imagination.", and navigation links (WHAT IS SECOND LIFE?, SHOWCASE, COMMUNITY, BLOG, SUPPORT).
- Main Content:** A large image of two avatars flying in a virtual sky. Text on the right says "Get Started! Membership is FREE! Second Life is an online, 3D virtual world imagined and created entirely by its Residents." Below this, it says "Discover a whole new world of friends, fashion, music, videos and fun! Explore the best of Second Life >>".
- Footer:** A section titled "Your Organization in Second Life!" with a "VISIT SECOND LIFE GRID" button and a small image of a virtual grid.

Social News/Mash Up

The image shows a screenshot of a social news/mashup website. On the left, there is a list of news items with their digg counts and titles. On the right, there is a detailed view of the artist Björk, including her profile information, a list of genres, and several embedded content widgets.

News List:

- 104 diggs: **Microsoft Demos "ADD TO DIGG" Feature in IE8**
microsoft.com — Using the new "Activities" XML feature in IE8, users can right-click on any page. Check out the screenshot on Microsoft's site! Other new IE8 features: Atom feeds in normal we...
16 Comments | Share | Bury | renesisx made popular 6 min ago
- 161 diggs: **It was only a matter of time, The SIMS 3 Official**
strategyinformer.com — Maxis have already showcased The Sims 3 in a non-disclosure agreement... the blockade is said to end on March 19. Wasting no time to rally the Sim communities around...
36 Comments | Share | Bury | JamieSI made popular 14 min ago
- 151 diggs: **Universe submerged in a sea of chilled neutrinos**
space.newscientist.com — 'Cosmic' neutrinos produced in the big bang fill space, according to five... NASA's WMAP satellite - they can't be detected on Earth [More...](#) (Space)
21 Comments | Share | Bury | MaggieMcKee made popular 26 min ago
- 180 diggs: **Unique locks on microchips could reduce hardware piracy**
ns.umich.edu — The technique is called EPIC, short for Ending Piracy of Integrated Circuits. [Mor...](#)
45 Comments | Share | Bury | cosmicdebris made popular 45 min ago
- 519 diggs: **Warren Buffett Passes Gates To Become World's Richest Man**
...
42 Comments | Share | Bury | ... made popular 42 min ago

Björk Profile:

Björk
Björk Guðmundsdóttir ([ˈpjo̥r̥k ˈkʏðmʏnstou̯htɪr] (help·)) (born November 21, 1965 in Reykjavík, Iceland) is an Icelandic singer/songwriter and com... (more)
Share this artist | Blog this artist

Genres: Pop, Trip-Hop, Rock, Vocal Jazz, Ambient, Electronica, Dance, Alternative, Experimental

Widgets:


- Videos on YouTube:** All is full of love (4:09), bjork-hunter (3:38), Bjork - Human Behaviour (4:17)
- Lyrics from Yahoo! Music:** 5 Years, Alarm Call, All Is Full of Love, All Neon Like, An Echo, A Stain, Army of Me, Aurora, Bachelorette, Big Time Sensuality, Cetacea, Cocoon, Come to Me, Crying, Desired Constellation
- Flickr Photos:** Selected Photos, More on Flickr
- Artist on Last.fm:** The Sugarcubes, Goldfrapp
- Music on Hype Machine:** Play All



Social Knowledge Sharing

WIKIPEDIA

<p>English <i>The Free Encyclopedia</i> 2 268 000+ articles</p>	<p>Deutsch <i>Die freie Enzyklopädie</i> 718 000+ Artikel</p>
<p>Français <i>L'encyclopédie libre</i> 631 000+ articles</p>	<p>Polski <i>Wolna encyklopedia</i> 477 000+ haseł</p>
<p>日本語 フリー百科事典 474 000+ 記事</p>	<p>Italiano <i>L'enciclopedia libera</i> 421 000+ voci</p>
<p>Nederlands <i>De vrije encyclopedie</i> 414 000+ artikelen</p>	<p>Português <i>A enciclopédia livre</i> 364 000+ artigos</p>
<p>Español <i>La enciclopedia libre</i> 339 000+ artículos</p>	<p>Svenska <i>Den fria encyklopedin</i> 277 000+ artiklar</p>



search · suche · rechercher · szukaj · 検索 · ricerca · zoeken · busca
 buscar · søk · поиск · 搜索 · søk · haku · suk · cerca · căutare · ara

English ▼ ➤

KNOL™
BETA

Welcome to Knol

Share what you know

Write and post a knol (nōl) — a unit of knowledge.



Create
easy to write and manage





Search
searchable through popular search engines

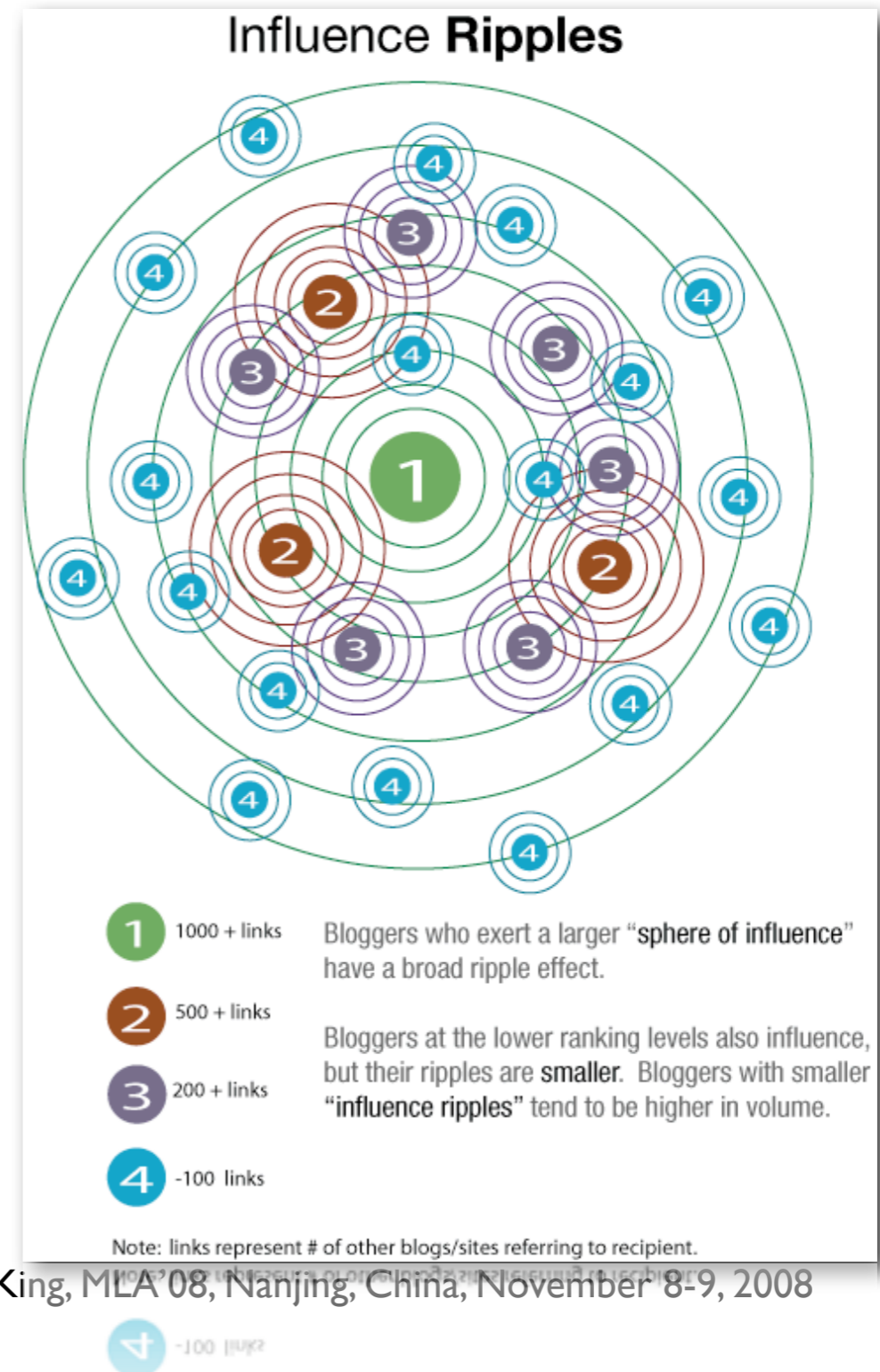


Control
each knol is owned by you, the author



Social Marketing

- Viral marketing
- Who are the **brokers**?
- Who can exert the **most influence** on buying/selling?
- How **much** should one advertise?



Social/Human Computation

Security Check: Enter **both** words below, separated by a space. [What's This?](#)
Can't read this? [Try another.](#)
[Try an audio captcha](#)

discharge *carolina*

Text in the box:

I have read and agree to the [Terms of Use and Privacy Policy](#)

[Sign Up](#)

[Problems signing up? Check out our help pages](#)

Security Check: Enter **both** words below, separated by a space. [What's This?](#)
Can't read this? [Try another.](#)
[Try an audio captcha](#)

discharge *tesbiten*

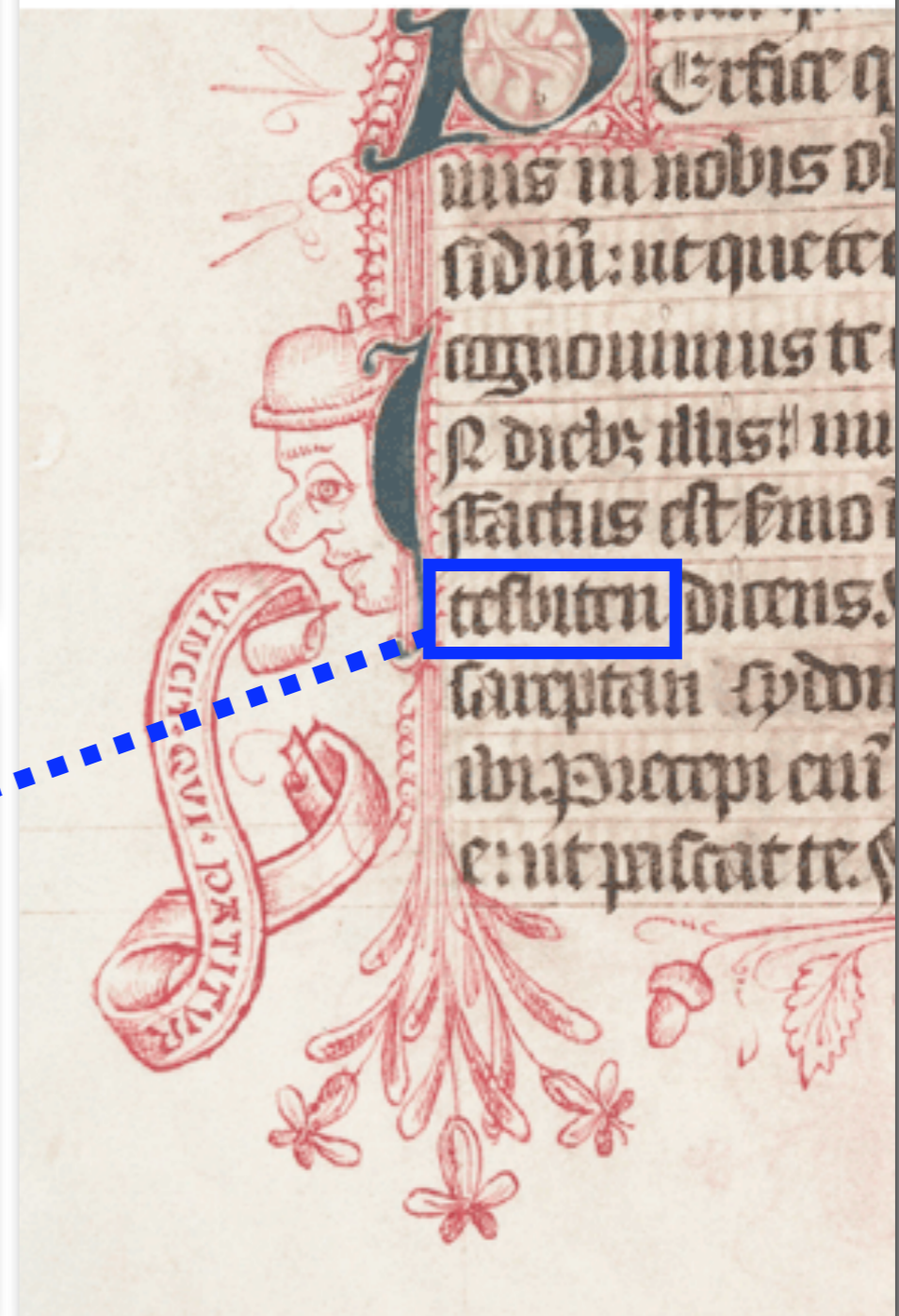
Text in the box:

I have read and agree to the [Terms of Use and Privacy Policy](#)

[Sign Up](#)

[Problems signing up? Check out our help pages](#)

MS. Don. b. 6, fol. 48v (detail) © Bodleian Library, University of Oxford



Web 2.0 Revolution

革命 vs. 演变

The Three C's

Connectivity (人等之间的关系)

Collaboration (协作)

Communities (社会团体)



What's On the Menu?

- Web 2.0 and Social X
- Social Computing (社会计算)
- Some Interesting Problems
 - Collaborative Filtering (协同过滤)
 - Query Suggestion (查询建议)



Social Relations

同伴;支持者;共谋人;市场;社区;团体;组织

presence (存在)

crew

族,队;组;班;群;窝

identity (身份)

teams

populations

binary

social role (社会角色)

squad

organizations

cardinal

reputation (信誉)

cohorts

markets

integer

expertise (专长)
trust (信任)

communities

partners

real

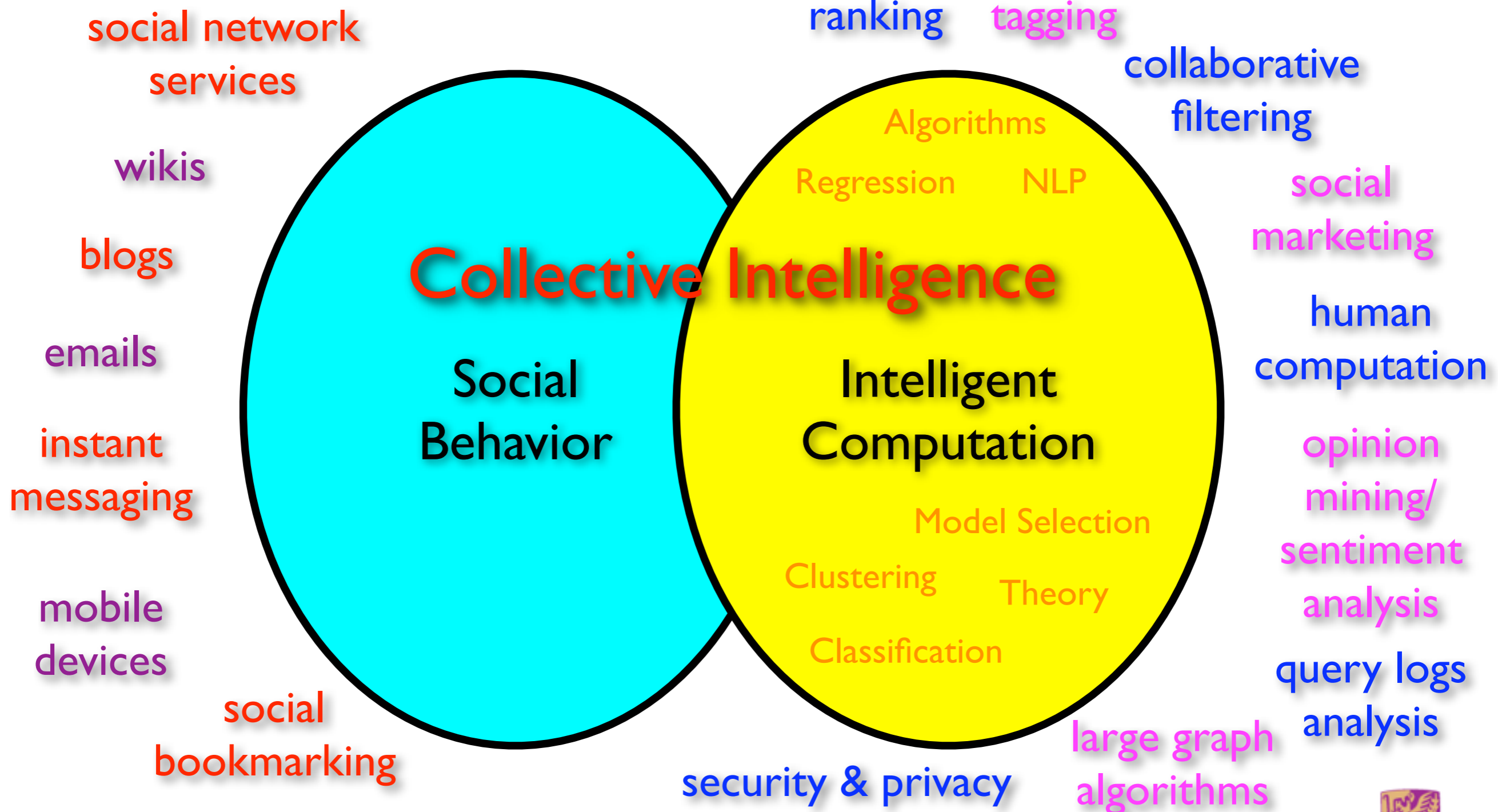
ownership (所有权)
accountability (问责制)

groups

knowledge (知识)



Social Computing



Social Computing (SC)

- Social computing is a general term for an area of computer science that is concerned with the intersection of *social behavior* (社会行为) and *computational systems*. Wikipedia
- *A social structure in which technology puts power in communities, not institutions.* Forrester
- *Forms of web services where the value is created by the collective contributions (集体贡献) of a user population.*



Issues

- **Theory** and models (理论与模型)
- **Mining** (数据挖掘) of existing information, e.g., spatial (relations) and temporal (time) domains
- Dealing with **partial** and **incomplete** (局部和不完整) information, e.g., collaborative filtering, ranking, tagging, etc.
- **Scalability** and algorithmic (可扩展性和算法) issues
- **Security** and **privacy** (安全和隐私) issues
- **Monetization** (货币) of social interactions



Machine Learning in SC

- Classification, clustering, regression, etc.
- New insights on the data
 - Social relations are often **hidden** (latent)
 - Change data from (x, y) to $(x, c_1(x), c_2(x), \dots, y)$
- $c(x)$ = context in **tags, relations, ratings**, etc.
- data type = *binary, integer, real, cardinal*, etc.



What's On the Menu?

- Web 2.0 and Social X
- Social Computing (社会计算)
- Some Interesting Problems
 - Collaborative Filtering (协同过滤)
 - Query Suggestion (查询建议)



What's On the Menu?

- Web 2.0 and Social X
- Social Computing (社会计算)
- Some Interesting Problems
 - Collaborative Filtering (協同過濾)
 - Query Suggestion (查询建议)



A Better Mousetrap?

cuil

how to be

- How to Be a Player
- How to be Happy
- How to Be Popular
- How to Be a DJ
- How to Be Cool
- How to be Rich

617,892,992 web pages

out Cuil | Your Privacy

© 2008 Cuil, Inc

Yahoo! Mail

Web Images Video Local Shopping more

dudley man

YAHOO!

dudley manlove

dudley manlove quartet

dudley mansion

dudley management

dudley mandy

- Muppets**
www.Target.com - Find muppets Online. Shop and Save at Target Today.
- Muppet Show Ringtones**
www.hiptunez.com - Download Muppet Show ringtones to your phone today.

1. **Muppets.com**
Official site of Jim Henson's Muppets. Includes games, music, downloads, news, and information about all the characters.
muppets.go.com - 4k - Cached



Live Search | MSN | Windows Live | Hotmail

United States | Options | cashback | Sign in

Live Search

Web 1-10 of 132,000,000 results · [Advanced](#)
See also: [Images](#), [Video](#), [News](#), [Maps](#), [More](#)

Apple Items - www.ebay.com Sponsored sites
Buy **Apple** Items. You may get 25% off with PayPal if eligible.

macintosh - Search.Live.com/cashback
Earn cashback on millions of products from sites you trust!

Official Apple® Store - store.apple.com
Buy the new MacBook, Air, and Pro direct from **Apple**. Free shipping.

Apple
Apple designs and creates iPod and iTunes, Mac laptop and desktop computers, the OS X operating system, and the revolutionary iPhone.
www.apple.com · [Cached page](#)

[iPod+iTunes](#)
[The Apple Store](#)
[Downloads](#)
[Support](#)
[Show more results from www.apple.com](#)

Related searches

- [Apple iPod](#)
- [iTunes](#)
- [Apple Computers](#)
- [Apple Vacations](#)
- [Apple Store](#)
- [Apple Trailers](#)
- [Best Buy](#)

Sponsored sites

You're a PC Too
Post a Picture. Tell Us How You Use Your PC. See Yourself in the Ads.
ImAPC.LifeWithoutWalls.com

[See your message here...](#)

ews, fan forum, and Muppet Show episode guide.

al.com liberal a

liberal arts	13,100,000 results
liberal arts colleges	818,000 results
liberal arts college	2,450,000 results
liberal arts degree	513,000 results
liberal arts education	1,710,000 results
liberal arts schools	1,320,000 results
liberal arts major	6,130,000 results
liberal arts college rankings	341,000 results
liberal art	2,950,000 results
liberal arts college ranking	457,000 results

[close](#)



Challenges

- Queries contain **ambiguous** and **new** terms
- **apple**: “apple computer” or “apple pie”?
- **NDCG**:?
- Users tend to submit **short queries** consisting of only one or two words
- almost **20%** one-word queries
- almost **30%** two-word queries

- Users may have **little or even no knowledge** about the topic they are searching for!



What is Clickthrough Data

- Query logs recorded by search engines

$$\langle u, q, l, r, t \rangle$$

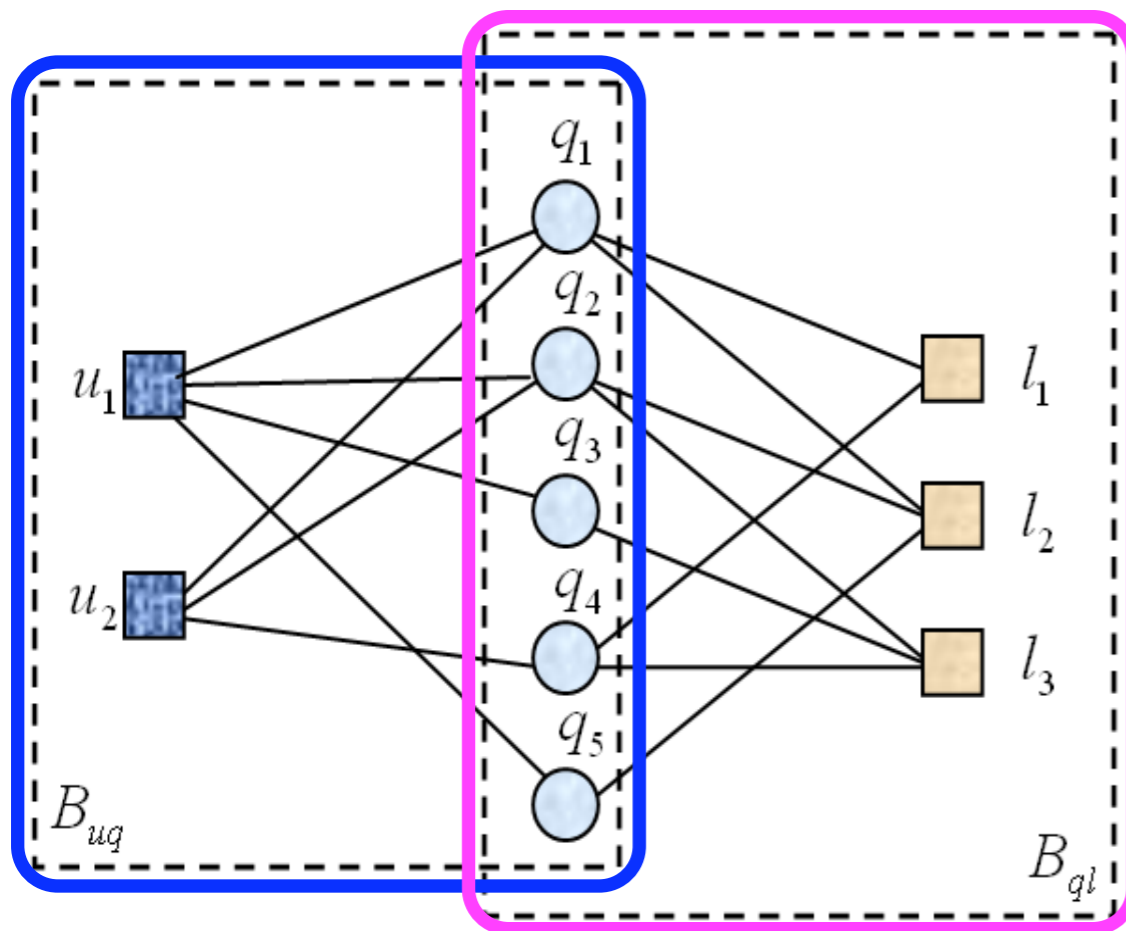
Table 1: Samples of search engine clickthrough data

ID	Query	URL	Rank	Time
358	facebook	http://www.facebook.com	1	2008-01-01 07:17:12
358	facebook	http://en.wikipedia.org/wiki/Facebook	3	2008-01-01 07:19:18
3968	apple iphone	http://www.apple.com/iphone/	1	2008-01-01 07:20:36
...

- Users' **relevance feedback** to indicate desired/preferred/target results



Joint Bipartite Graph



$$B_{uq} = (V_{uq}, E_{uq})$$

$$V_{uq} = U \cup Q$$

$$U = \{u_1, u_2, \dots, u_m\}$$

$$Q = \{q_1, q_2, \dots, q_n\}$$

$E_{uq} = \{(u_i, q_j) \mid \text{there is an edge from } u_i \text{ to } q_j\}$
is the set of all edges.

The edge (u_i, q_j) exists in this bipartite graph if and only if a user u_i issued a query q_j .

$$B_{ql} = (V_{ql}, E_{ql})$$

$$V_{ql} = Q \cup L$$

$$Q = \{q_1, q_2, \dots, q_n\}$$

$$L = \{l_1, l_2, \dots, l_p\}$$

$E_{ql} = \{(q_i, l_j) \mid \text{there is an edge from } q_i \text{ to } l_j\}$
is the set of all edges.

The edge (q_j, l_k) exists if and only if a user u_i clicked a URL l_k after issuing an query q_j .

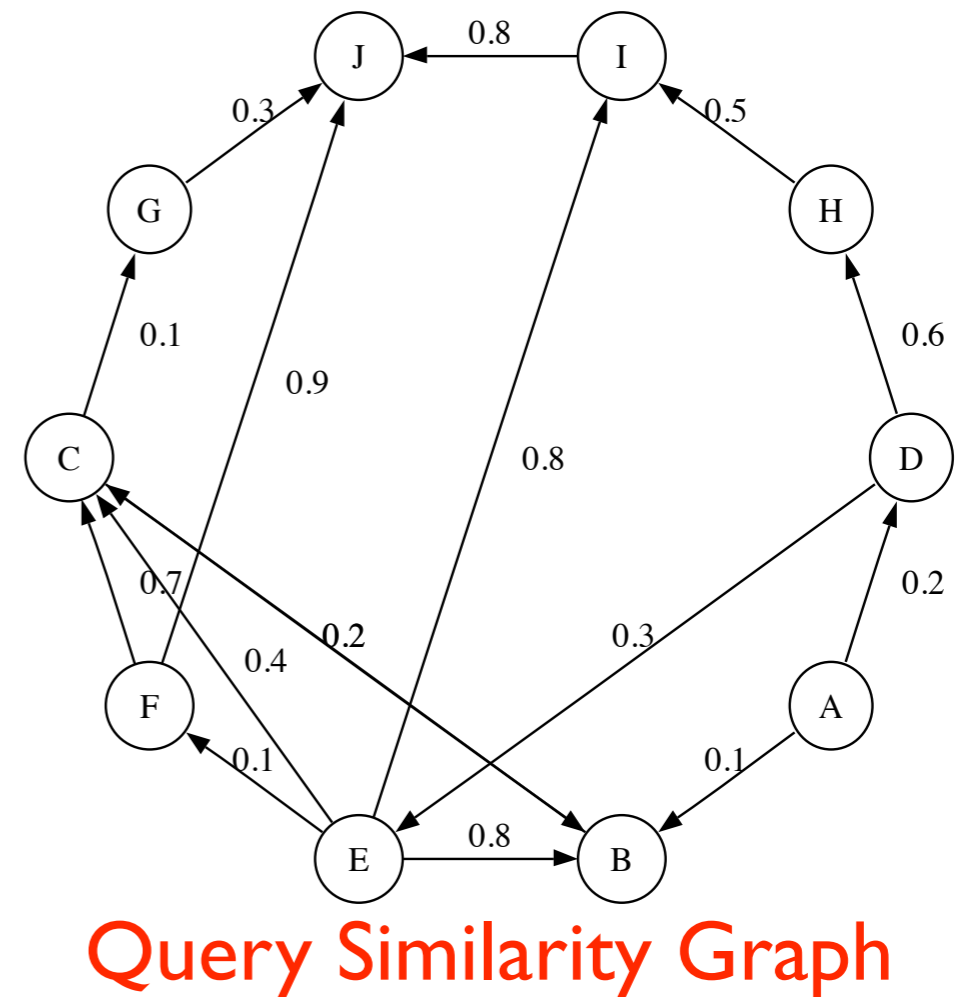
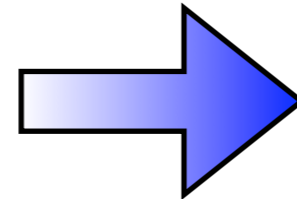
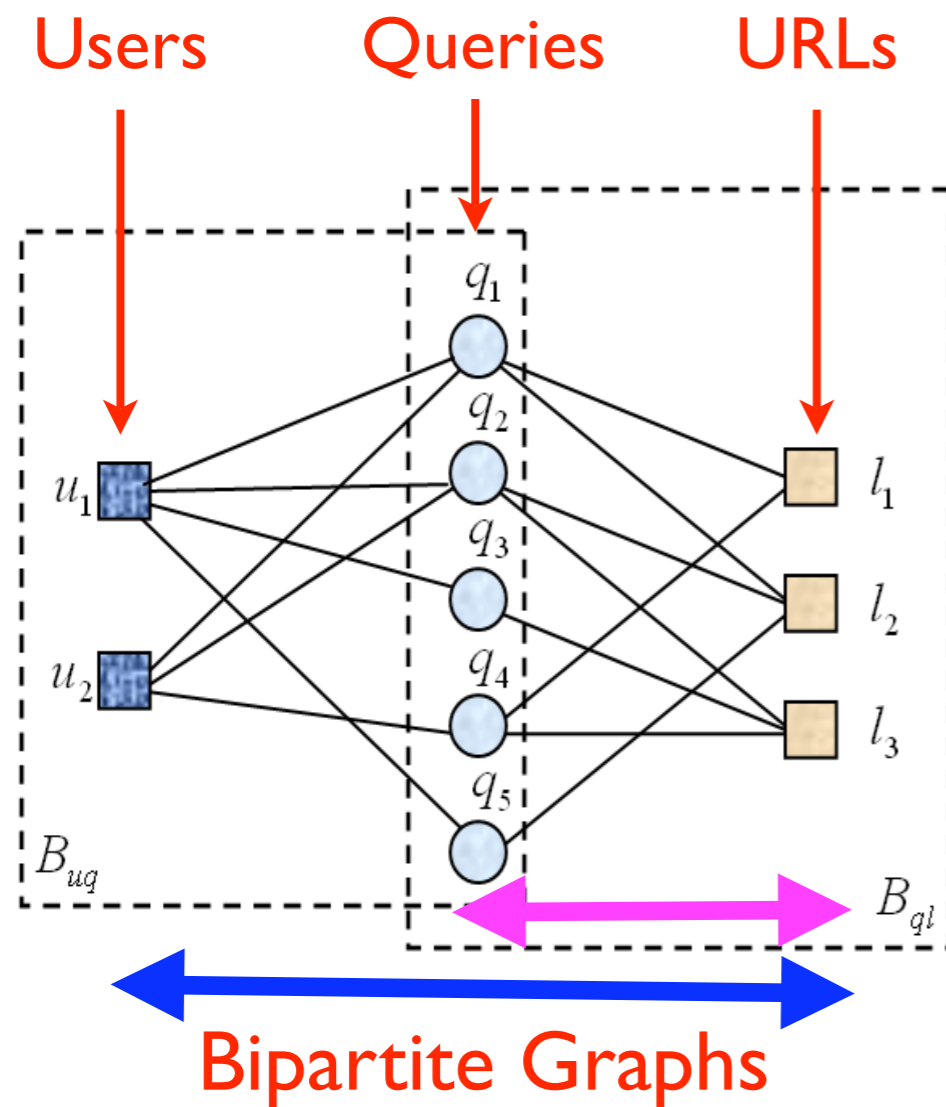


Key Points

- Two-level latent semantic analysis

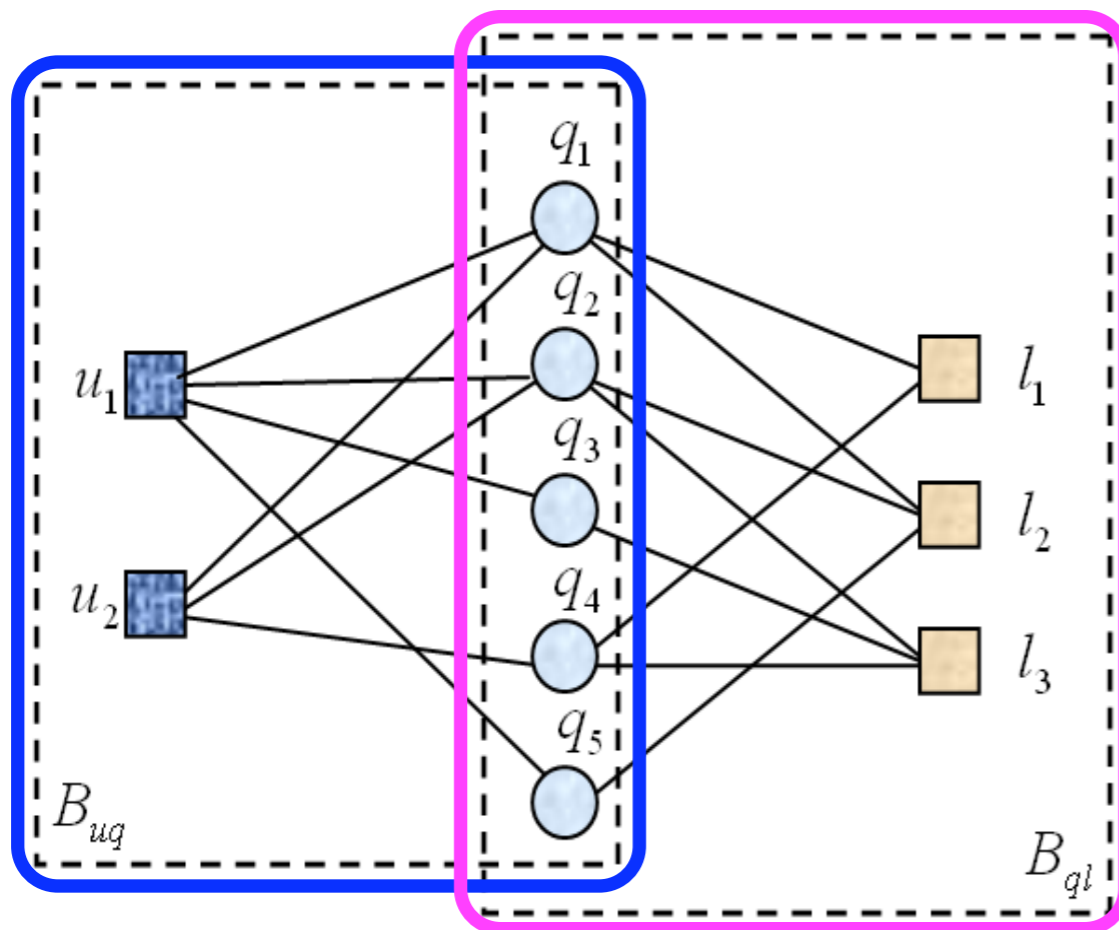
- Level 1
- Consider the use of a joint **user-query** and **query-URL bipartite graphs** for query suggestion
- Level 2
- Use **matrix factorization** for learning query features in constructing the Query Similarity Graph
 - Use **heat diffusion** for similarity propagation for query suggestions





- Queries are issued by the users, and which URLs to click are also decided by the users
- Two distinct users are similar if they issued **similar queries**
- Two queries are similar if they are issued by **similar users**



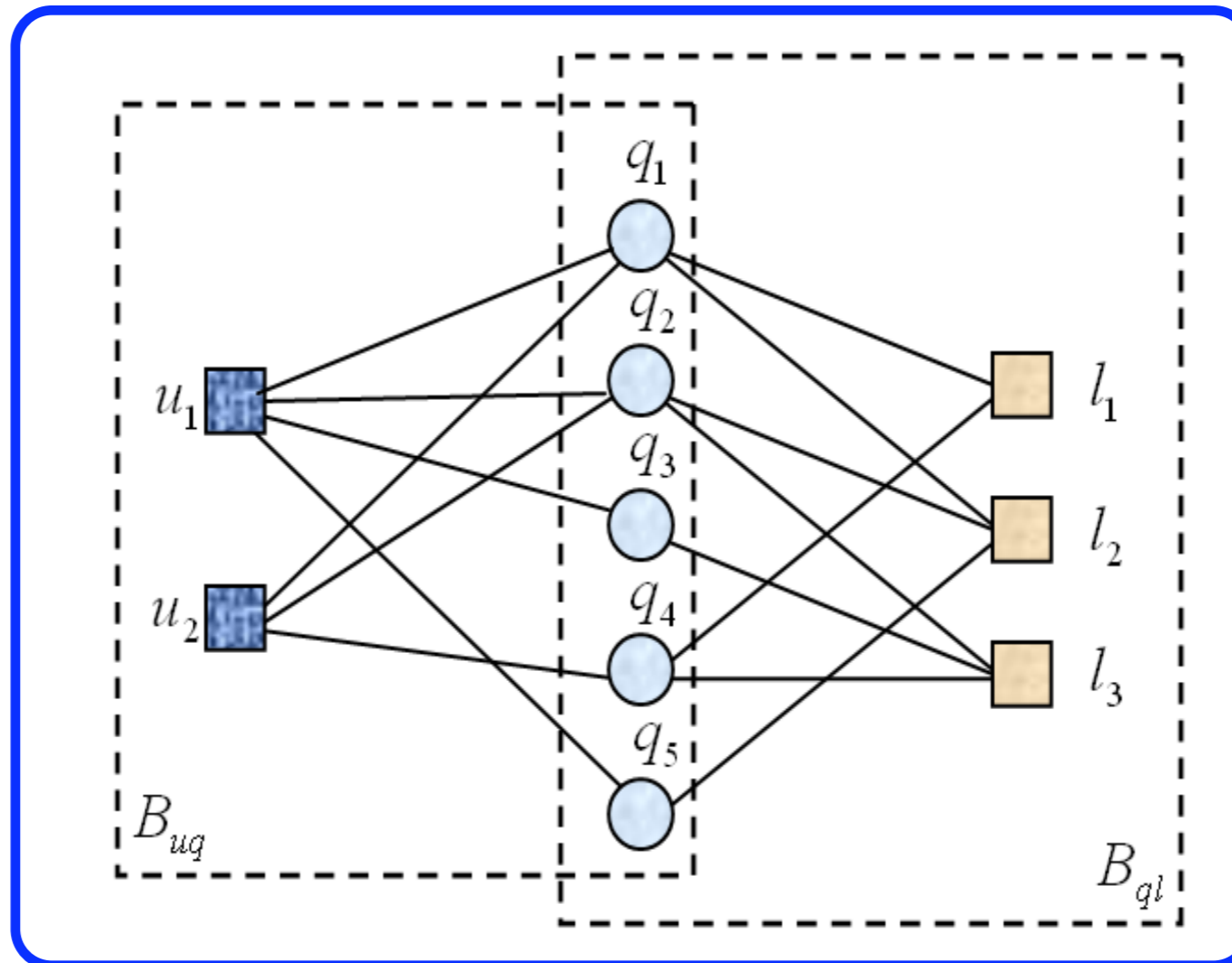


- r_{ij}^* Normalized weight, how many times u_i issued q_j
- $s_{j k}^*$ Normalized weight, how many times q_j is linked to l_k
- U_i L -dimensional vector of user u_i
- Q_j L -dimensional vector of query q_j
- L_k L -dimensional vector of URL l_k

$$\mathcal{H}(R, U, Q) = \min_{U, Q} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (r_{ij}^* - g(U_i^T Q_j))^2 + \frac{\alpha_u}{2} \|U\|_F^2 + \frac{\alpha_q}{2} \|Q\|_F^2$$

$$\mathcal{H}(S, Q, L) = \min_{Q, L} \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^p I_{jk}^S (s_{jk}^* - g(Q_j^T L_k))^2 + \frac{\alpha_q}{2} \|Q\|_F^2 + \frac{\alpha_l}{2} \|L\|_F^2$$





$$\mathcal{H}(S, R, U, Q, L) =$$

$$\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^p I_{jk}^S (s_{jk}^* - g(Q_j^T L_k))^2 + \frac{\alpha_r}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (r_{ij}^* - g(U_i^T Q_j))^2$$

$$+ \frac{\alpha_u}{2} \|U\|_F^2 + \frac{\alpha_q}{2} \|Q\|_F^2 + \frac{\alpha_l}{2} \|L\|_F^2,$$

- A local minimum can be found by performing **gradient descent** in U_i , Q_j and L_k



Gradient Descent Equations

$$\frac{\partial \mathcal{H}}{\partial U_i} = \alpha_r \sum_{j=1}^n I_{ij}^R g'(U_i^T Q_j) (g(U_i^T Q_j) - r_{ij}^*) Q_j + \alpha_u U_i,$$

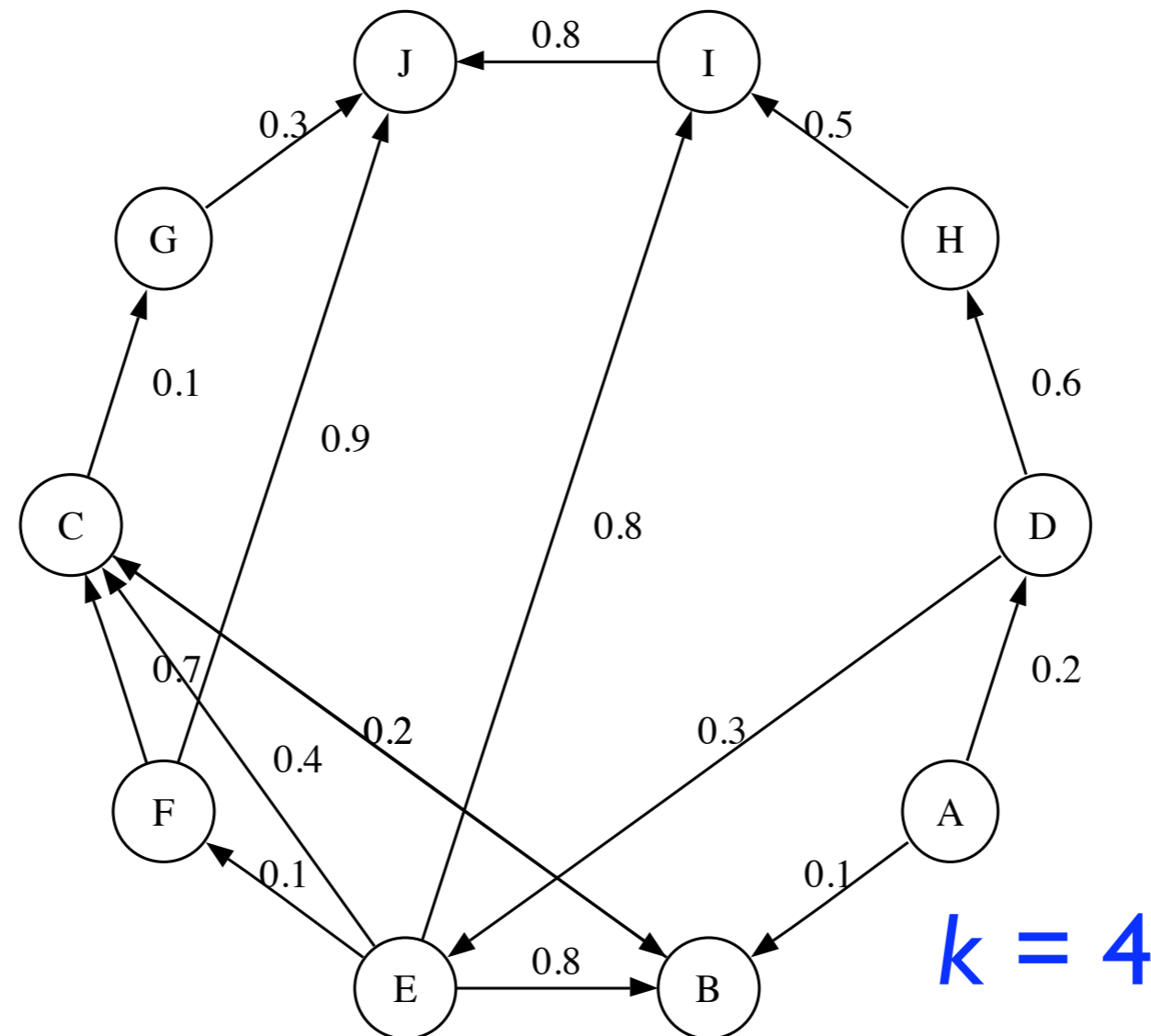
$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial Q_j} &= \sum_{k=1}^p I_{jk}^S g'(Q_j^T L_k) (g(Q_j^T L_k) - s_{jk}^*) L_k \\ &+ \alpha_r \sum_{i=1}^m I_{ij}^R g'(U_i^T Q_j) (g(U_i^T Q_j) - r_{ij}^*) U_i + \alpha_q Q_j, \end{aligned}$$

$$\frac{\partial \mathcal{H}}{\partial L_k} = \sum_{j=1}^n I_{jk}^S g'(Q_j^T L_k) (g(Q_j^T L_k) - s_{jk}^*) Q_j + \alpha_l L_k,$$

Only the **Q matrix**, the queries' latent features, is being used to generate the **query similarity graph!**



Query Similarity Graph



- Similarities are calculated using queries' latent features
- Only the **top-k** similar neighbors (terms) are kept



Similarity Propagation

- Based on the **Heat Diffusion Model**
- In the query graph, given the **heat sources** and the **initial heat values**, start the heat diffusion process and perform **P steps**
- Return the **Top- N** queries in terms of highest heat values for query suggestions



Heat Diffusion Model

- Heat diffusion is a **physical phenomena**
- Heat flows from **high** temperature to **low** temperature in a **medium**
- **Heat kernel** is used to describe the amount of heat that one point receives from another point
- The way that heat diffuse varies when the **underlying geometry**

$$\rho C_P \frac{\partial T}{\partial t} = Q + \nabla \cdot (k \nabla T)$$

ρ Density

C_P Heat capacity and constant pressure

$\frac{\partial T}{\partial t}$ Change in temperature over time

Q Heat added

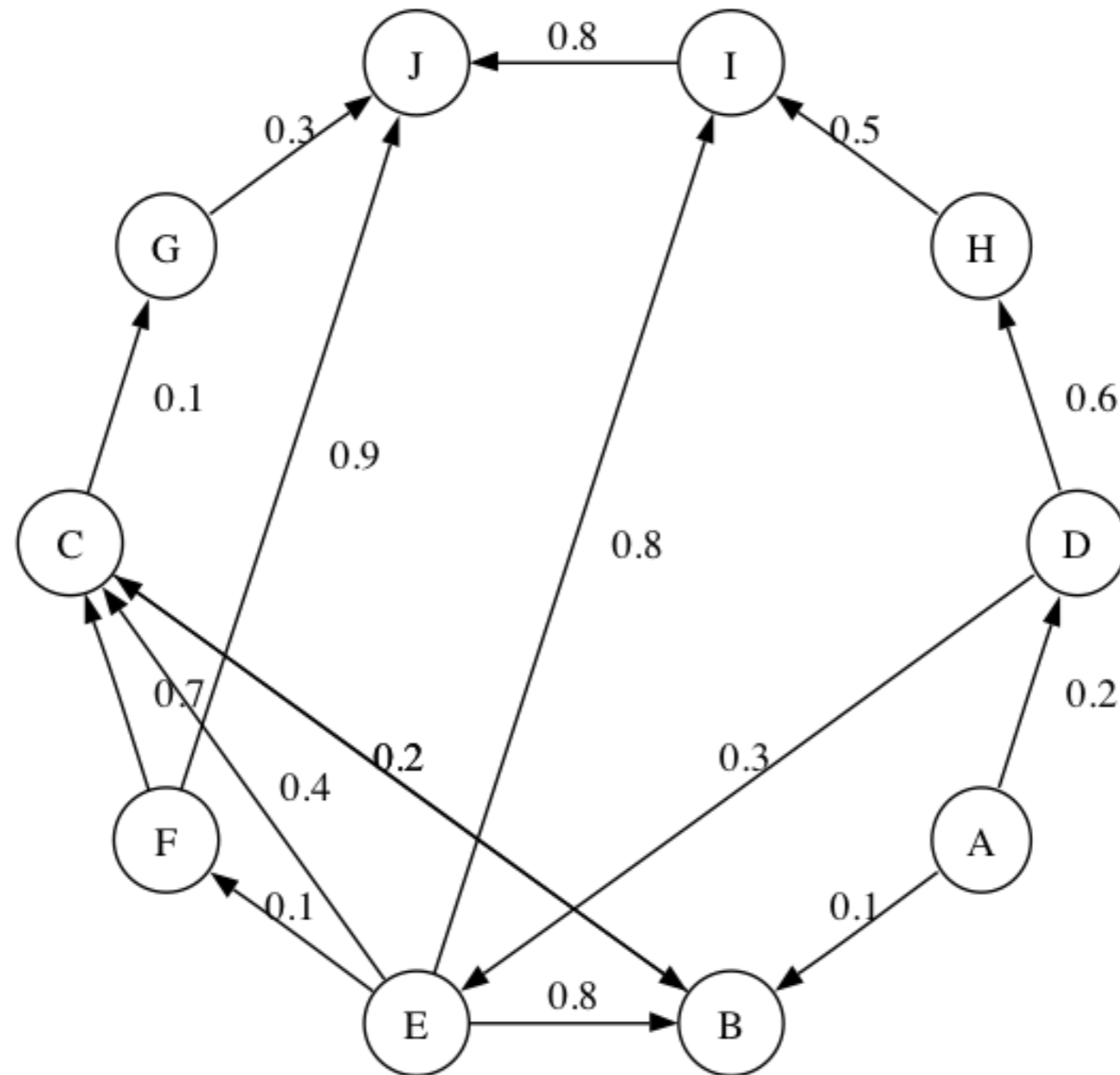
k Thermal conductivity

∇T Temperature gradient

$\nabla \cdot \mathbf{v}$ Divergence



Heat Diffusion Process



Similarity Propagation Model

$$\frac{f_i(t + \Delta t) - f_i(t)}{\Delta t} = \alpha \left(-\frac{\tau_i}{d_i} f_i(t) \sum_{k:(q_i, q_k) \in E} w_{ik} + \sum_{j:(q_j, q_i) \in E} \frac{w_{ji}}{d_j} f_j(t) \right) \quad (1)$$

$$\mathbf{f}(1) = e^{\alpha \mathbf{H}} \mathbf{f}(0) \quad (2)$$

$$H_{ij} = \begin{cases} w_{ji}/d_j, & (q_j, q_i) \in E, \\ -(\tau_i/d_i) \sum_{k:(i,k) \in E} w_{ik}, & i = j, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

$$\mathbf{f}(1) = e^{\alpha \mathbf{R}} \mathbf{f}(0), \quad \mathbf{R} = \gamma \mathbf{H} + (1 - \gamma) \mathbf{g} \mathbf{1}^T \quad (4)$$

α Thermal conductivity

d_i Heat value of node i at time t

$f_i(t)$ Heat value of node i at time t

w_{ik} Weight between node i and node k

$\mathbf{f}(0)$ Vector of the initial heat distribution

$\mathbf{f}(1)$ Vector of the heat distribution at time 1

τ_i Equal to 1 if node i has outlinks, else equal to 0

γ Random jump parameter, and set to 0.85

\mathbf{g} Uniform stochastic distribution vector



Discrete Approximation

- Compute $e^{\alpha \mathbf{R}}$ is time consuming
- We use the **discrete approximation** to substitute

$$\mathbf{f}(1) = \left(\mathbf{I} + \frac{\alpha}{P} \mathbf{R} \right)^P \mathbf{f}(0)$$

- For every heat source, only diffuse heat to its neighbors within **P steps**
- In our experiments, $P = 3$ already generates fairly good results



Query Suggestion Procedure

- For a given query q
 1. Select a set of n queries, each of which contains at least one word in common with q , as **heat sources**

2. Calculate the initial heat values by

$$f_{\hat{q}_i}(0) = \frac{|\mathcal{W}(q) \cap \mathcal{W}(\hat{q}_i)|}{|\mathcal{W}(q) \cup \mathcal{W}(\hat{q}_i)|}$$

$q = \text{"Sony"}$
 $\text{"Sony"} = 1$

$\text{"Sony Electronics"} = 1/2$

$\text{"Sony Vaio Laptop"} = 1/3$

3. Use $f(1) = e^{\alpha \mathbf{R}} f(0)$ to diffuse the heat in graph

4. Obtain the **Top-N** queries from $f(1)$



Physical Meaning of α

- If set α to a large value
 - The results depend more on the query graph, and **more semantically** related to original queries, e.g., **travel => lowest air fare**
- If set α to a small value
 - The results depend more on the initial heat distributions, and **more literally** similar to original queries, e.g., **travel => travel insurance**



Experimental Dataset

Data Source	Clickthrough data from AOL search	After Pre-Processing
Collection Period	March 2006 to May 2006 (3 months)	
Lines of Logs	19,442,629	
Unique user IDs	657,426	192,371
Unique queries	4,802,520	224,165
Unique URLs	1,606,326	343,302
Unique words		69,937



Query Suggestions

Table 2: Examples of LSQS Query Suggestion Results ($k = 50$)

Testing Queries	Suggestions				
	$\alpha = 10$			$\alpha = 1000$	
	Top 1	Top 2	Top 3	Top 4	Top 5
michael jordan	michael jordan shoes	michael jordan bio	pictures of michael jordan	nba playoff	nba standings
travel	travel insurance	abc travel	travel companions	hotel tickets	lowest air fare
java	sun java	java script	java search	sun microsystems inc	virtual machine
global services	ibm global services	global technical services	staffing services	temporary agency	manpower professional
walt disney land	world of disney	disney world orlando	disney world theme park	disneyland grand hotel	disneyland in california
intel	intel vs amd	amd vs intel	pentium d	pentium	centrino
job hunt	jobs in maryland	monster job	jobs in mississippi	work from home online	monster board
photography	photography classes	portrait photography	wedding photography	adobe elements	canon lens
internet explorer	ms internet explorer	internet explorer repair	internet explorer upgrade	microsoft com	security update
fitness	fitness magazine	lifestyles family fitness	fitness connection	womens health magazine	family fitness
m schumacher	schumacher	red bull racing	formula one racing	ferrari cars	formula one
solar system	solar system project	solar system facts	solar system planets	planet jupiter	mars facts
sunglasses	replica sunglasses	cheap sunglasses	discount sunglasses	safilo	marhon
search engine	audio search engine	best search engine	search engine optimization	song lyrics search	search by google
disease	grovers disease	liver disease	morgellons disease	colic in babies	oklahoma vital records
pizzahut	pizza hut menu	pizza coupons	pizza hut coupons	papa johns pizza coupon	papa johns
health care	health care proxy	universal health care	free health care	great west healthcare	uhc
flower delivery	global flower delivery	online florist	flowers online	send flowers	virtual flower
wedding	wedding guide	wedding reception ideas	wedding decoration	unity candle	centerpiece ideas
astronomy	astronomy magazine	astronomy pic of the day	star charts	space pictures	comet



Comparisons

Table 3: Comparisons between LSQS and SimRank

	Top 1	Top 2	Top 3	Top 4	Top 5
jaguar					
LSQS	jaguar cat	jaguar commercial	jaguar parts	jaguarundi	leopard
SimRank	american black bear	bottlenose dolphin	leopard	margay	jaguarundi
apple					
LSQS	apple computers	apple ipod	apple diet	apple vacations	apple bottom
SimRank	ipod troubleshooting	apple quicktime	apple ipods	apple computers	apple software

Table 4: Accuracy Comparisons

Accuracy	LSQS	SimRank
By Experts	0.8413	0.7101
By ODP	0.6823	0.5789

ODP, Open Directory Project, see <http://dmoz.org>



Summary

- Propose an offline novel **joint matrix factorization** method using **user-query** and **query-URL bipartite graphs** for learning query features
- Propose an online diffusion-based **similarity propagation** and **ranking method** for query suggestion



Conclusion

- Social Computing is a **paradigm shift** (範式轉變)!
- Novel views on the **spatial** and **temporal** relationship among **social entities**!
- Great **opportunities** in a new research direction!



On-Going Research

Machine Learning

- Direct Zero-norm Optimization for Feature Selection (ICDM'08)
- Semi-supervised Learning from General Unlabeled Data (ICDM'08)
- Learning with Consistency between Inductive Functions and Kernels (NIPS'08)
- An Extended Level Method for Efficient Multiple Kernel Learning (NIPS'08)
- Semi-supervised Text Categorization by Active Search (CIKM'08)
- Transductive Support Vector Machine (NIPS'07)
- Global and local learning (ICML'04, JMLR'04)

Web Intelligence

- Effective Latent Space Graph-based Re-ranking Model with Global Consistency (WSDM'09)
- Formal Models for Expert Finding on DBLP Bibliography Data (ICDM'08)

- Learning Latent Semantic Relations from Query Logs for Query Suggestion (CIKM'08)
- RATE: a Review of Reviewers in a Manuscript Review Process (WI'08)
- MatchSim: link-based web page similarity measurements (WI'07)
- Diffusion rank: Ranking web pages based on heat diffusion equations (SIGIR'07)
- Web text classification (WWW'07)

Collaborative Filtering

- Recommender system: accurate recommendation based on sparse matrix (SIGIR'07)
- SoRec: Social Recommendation Using Probabilistic Matrix Factorization (CIKM'08)

Human Computation

- An Analytical Study of Puzzle Selection Strategies for the ESP Game (WI'08)
- An Analytical Approach to Optimizing The Utility of ESP Games (WI'08)



Acknowledgments

- Prof. Michael R. Lyu
- Prof. Jimmy Lee
- Dr. Kaizhu Huang
- Dr. Haixuan Yang
- Thomas Chan (M.Phil)
- Hongbo Deng (Ph.D.)
- Zhenjiang Lin (Ph.D.)
- Hao Ma (Ph.D.)
- Haiqin Yang (Ph.D.)
- Xin Xin (Ph.D.)
- Zenglin Xu (Ph.D.)
- Chao Zhou (Ph.D.)



Q & A

<http://www.cse.cuhk.edu.hk/~king>

