

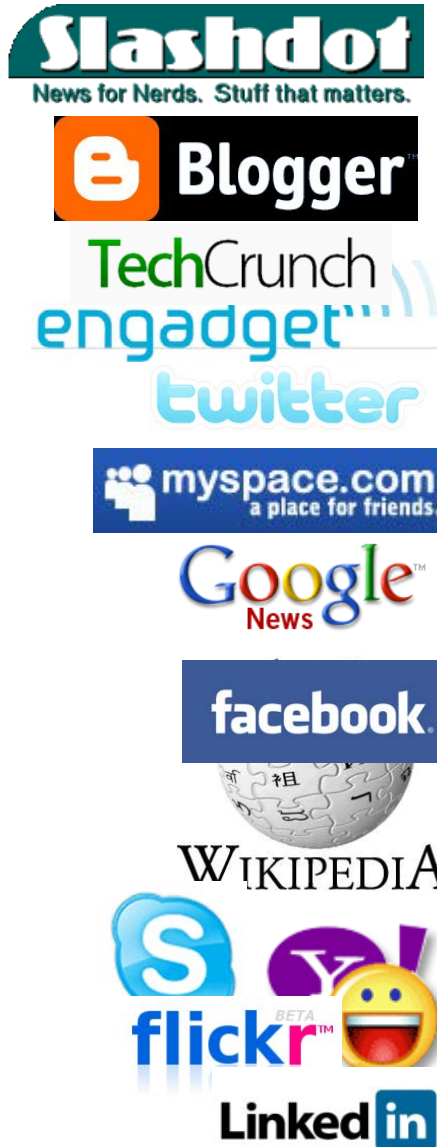


Social data management

Bogdan Cautis
TÉLÉCOM ParisTech

École Thematique BDA « Masses de données distribuées » 2012, Aussois

The social Web today



The Web is no longer a static library that we browse -> billions of users and connections, massive user-generated data, interactions, preferences, trends.

- Facebook generates more traffic than Google, has « 847.573.840 » users
- Wikipedia has 4 million pages with descriptions of entities, 10 million views per hour
- Flickr users have uploaded 6.5 billion photos
- YouTube has 3 billion views per day, 50h of videos uploaded each minute
- Twitter users generate 175 million tweets a day (Lady Gaga rapidly closing in on 20 million followers)

Exciting time for database researchers

Complex, dynamic, user-centric environments for publishing and disseminating information

- **collaborative** (the **Wikipedia**) and **social** applications (social networking, tagging, blogging, micro-blogging, video/photo sharing).

A new generation of tools is required to solve both classic data management tasks and emerging ones.

Social data management @ Telecom ParisTech: **efficient** access to **relevant** information in applications centered users, their relationships, their interactions and their data.

Roadmap

- Concepts, definitions, taxonomy
- Exemples of applications
- Models for social data
- Research challenges
- Predicting signed social links, inference from interactions
- Top-k search in social applications

Taxonomy of social applications

Content-oriented:

- Catalogues: Delicious, StumbleUpon, CiteULike, Last.fm
- Images / videos: Flickr, YouTube, DailyMotion, Instagram
- Wikis: Wikipedia, WikiTravel
- E-commerce: Ebay, Leboncoin, Groupon, LivingSocial
- Q&A: Yahoo! Answers
- News: Slashdot, Digg, Reddit
- Product reviews: Amazon, Ebay, Epinions

User-oriented:

- Social networking: Facebook, Google+, MySpace, Orkut, Hi5, LinkedIn
- Blogging / micro-blogging: Twitter, Blogspot, Wordpress

Among the most popular applications (according to Alexa traffic statistics)

	Worldwide	US	France
Facebook	2	2	2
Youtube	3	3	4
Wikipedia	6	6	6
Twitter	8	8	11
LinkedIn	12	10	15
Blogspot	10	11	12
EBay	21	7	13
Wordpress	18	23	25
Flickr	48	37	15

Not to mention Google, Yahoo, Amazon, MSN, Baidu, QQ, Taobao...

Concepts and definitions

Social Web :

« *The Social Web is currently used to describe how people **socialize or interact with each** other throughout the World Wide Web. »*

« *The Social Web may also be used to refer to the **description of web 2.0 technologies that are focused on social interaction and community** before anything else. »*

Source :Wikipedia

Concepts and definitions

Blog

« blog (a portmanteau of the term web log) is a personal journal published on the World Wide Web consisting of discrete entries ("posts") typically displayed in reverse chronological order.»

Wiki

«A wiki is a website whose users can add, modify, or delete its content via a web browser using a simplified markup language or a rich-text editor.»

Bookmarking (social bookmarking)

« *Social bookmarking is a method for Internet users to store, organize, search, and manage bookmarks of web pages on the Internet with the help of metadata. »*

Concepts and definitions

Tag (meta-data) :

« a non-hierarchical **keyword** or term assigned to a piece of information (such as an internet bookmark, digital image, or computer file) »

Folksonomy : fusion de Folk+Taxonomy

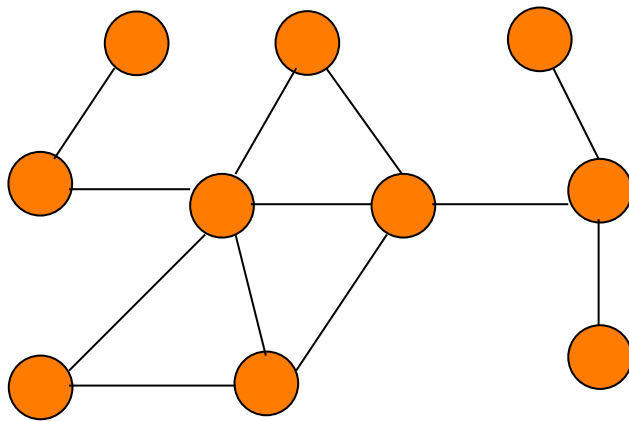
« Folksonomy (also known as collaborative tagging, social classification, social indexing, and social tagging) is the practice and method of collaboratively creating and managing tags to annotate and categorize content »

Source : wikipedia

Concepts and definitions

Social network

« A social network is a **social structure** made of nodes (which are generally individuals or organizations) that are tied by one or more specific types of **interdependency**, such as values, visions, ideas, financial exchange, friendship, kinship, dislike, conflict or trade. The resulting graph-based structures are often very complex. »



● Individual / organisation

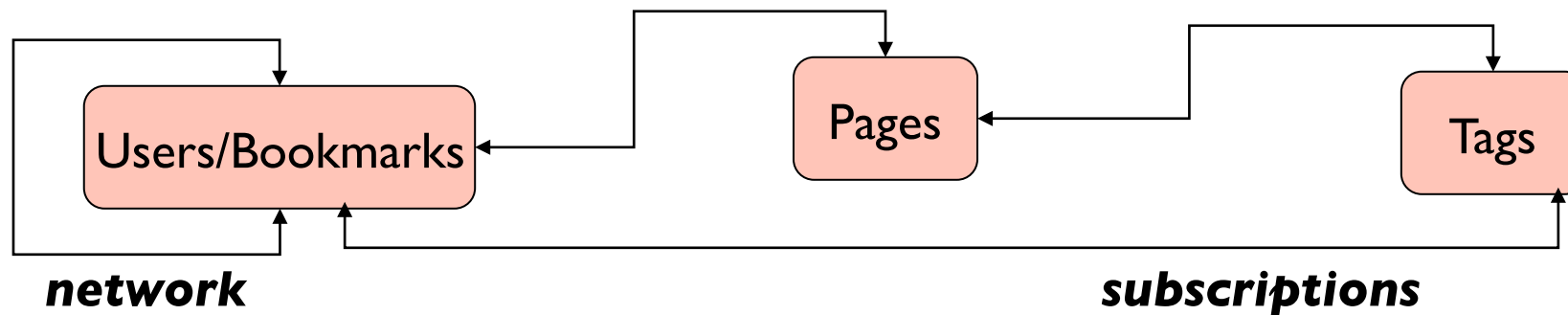
Source : wikipedia

Roadmap

- Concepts, definitions, taxonomy
- Exemples of applications
- Models for social data
- Research challenges
- Predicting signed social links, inference from interactions
- Top-k search in social applications

Delicious – tool for social bookmarking

Main entities



- Bookmarks (Web pages): may be public or private
- Keywords (given by the user)
 - Tag name and description
- Users
 - Login, e-mail, name, homepage
- Access: through social links or tag subscription
- Search: by tags (various scopes)

Delicious – tool for social bookmarking

The screenshot shows the Delicious interface for a user named 'gromuad'. The main content area displays a list of bookmarks, each with a date, title, URL, and a set of tags. Annotations are present:

- Tags:** A red arrow points to the 'research' tag of the first bookmark, with the word 'Tags' written above it.
- Occurrences:** A red arrow points to the number '465' (representing the number of occurrences) of the 'legifrance' bookmark, with the word 'Occurrences' written below it.
- Bookmarks: title, URL and notes:** A bracket on the left side of the page groups the first four bookmarks, with this text written next to it.
- Tags of the user:** A bracket on the right side of the page groups the 'Top 10 Tags' and 'All Tags' sections, with this text written next to it.

The bookmark list includes:

- 25 SEP 08 Centre de recherche en droit public www.crdp.umontreal.ca/fr/ Tags: research french canada droit Occurrences: 5
- Chaire L.R Wilson sur le droit des technologies de l'information et du commerce électronique www.chairelrwilson.ca/ Tags: privacy law canada droit Occurrences: 3
- Welcome to IEEE Xplore 2.0: Security and Privacy, IEEE Symposium on ieeexplore.ieee.org/xpl/conhome.jsp?punumber=1000646 Tags: research conference iee proceedings privacy Occurrences: 2
- 24 SEP 08 Ralf Bendrath bendrath.blogspot.com/ Tags: privacy blog theory law security politics Occurrences: 21
- ISECOM - Making Sense of Security www.isecom.org/osstmm/ Tags: sysadmin standards security reference policy Occurrences: 442
- 23 SEP 08 Journal Officiel - Accueil www.journal-officiel.gouv.fr/ Tags: politique officiel journal france legislation Occurrences: 46
- Legifrance - Le service public de l'accès au droit www.legifrance.gouv.fr/ Tags: service reference portail officiel loi france recherche Occurrences: 465
- Pricasso pricasso.com/ Tags: humor fun art pricasso Occurrences: 14
- Max-Planck-Institut für Informatik - Automation of Logic: Spass www.spass-prover.org/ Tags: logic spass prover software Occurrences: 1

The right sidebar shows a search bar and a 'Tags' section with 'Top 10 Tags' and 'All Tags' (741 total tags).

Delicious – tool for social bookmarking

Search example

*Search in tags,
titles and notes*

The screenshot shows the Delicious search interface. At the top, there is a navigation bar with 'delicious' and links for 'Home', 'Bookmarks', 'People', and 'Tags'. On the right, there are links for 'Join Now!', 'What's New?', 'Learn more', 'Help', and 'Sign In', along with a search box labeled 'Search Delicious' and a 'Search' button. Below the navigation bar, the search results are displayed under the heading 'Searching Everybody's bookmarks for:'. A search input field contains 'AFIA 2009' and a 'Search' button. Below the search bar, there is a link to 'Sign in to search your own bookmarks' and a button to 'Search all of Delicious for "AFIA 2009"'. The search results are shown in a blue bar labeled 'Everybody's bookmarks' with '2 results - show more detail'. The first result is 'Nouveau PortAI AFIA : Facture Inscription Plate-Forme AFIA 2009' with a 'SAVE' button and tags 'facture', 'ic2009', 'inscription', and 'afia'. The second result is 'Atelier Système d'Organisation des Connaissances IC 2009' with a 'SAVE' button and tags 'atelier', 'cms', 'ic2009', 'soc', and 'afia'. Below the results, there is a section for 'Everyone's Related Tags' showing 'afia', 'ic2009', 'atelier', 'cms', and 'soc'. At the bottom, there is a footer with links for 'delicious', 'about', 'blog', 'terms of service', 'privacy policy', 'copyright policy', 'forums', and 'support', and a 'What's new?' button.

Delicious – tool for social bookmarking

Tag subscription

The screenshot shows the Delicious website interface. At the top, there's a navigation bar with 'delicious', 'Home', 'Bookmarks', 'People', and 'Tags'. The main content area displays 'Talel Abdessalem's Subscriptions' with a list of bookmarks. Each bookmark entry includes a date, title, URL, user name, and a count. For example, 'Sen. John McCain [R-AZ] - Voting Record - GovTrack.us' has a count of 9. A 'Add a subscription' dialog box is open on the right, showing a search for 'obama' and an option to 'From a specific user'. Below the dialog, a 'Subscriptions' list shows 'All Subscriptions' with a count of 4, and a list of tags: 'mccain', 'salsa', 'social', and 'tango'. A red arrow points to the 'obama' tag in the dialog box.

From one user or all public bookmarks

My subscriptions

Delicious – tool for social bookmarking

Related tags

The screenshot shows the Delicious website interface. At the top, there's a navigation bar with 'delicious' and menu items like 'Home', 'Bookmarks', 'People', and 'Tags'. A search bar is visible with the text 'Search these bookmarks' and a 'Search' button. Below the navigation, there's a section for 'Recent obama Bookmarks' with a 'Recent | Popular' filter. A tag input field shows 'obama' and a 'Type another tag' prompt. The main content area displays several search results for the 'obama' tag, each with a title, URL, description, and associated tags. For example, the first result is 'Obama introduces Biden as running mate - CNN.com' with tags like 'michellehoover', 'Technology', and 'obama'. To the right, there's a 'Tags' sidebar with a 'Related Tags' section showing 11 related tags: '+politics', '+mccain', '+election', '+election08', '+election2008', '+2008', '+economy', '+palin', '+race', '+usa', and '+2008election'. A red bracket on the right side of the page groups these tags under the heading 'Tags related to Obama'.

Tags related to Obama

Delicious – tool for social bookmarking

Network and fans

The screenshot displays the Delicious website interface for a user named JeanSaul. The top navigation bar includes 'Home', 'Bookmarks', 'People', and 'Tags'. The user is signed in as 'tal.icio.us'. The main content area shows a list of bookmarks with details such as dates, titles, descriptions, and tags. On the right side, there is a sidebar with a search bar and two sections: 'People' and 'Fans'. The 'People' section shows a list of users in the network, including 'cpicard', 'gbrocker', 'iurics', 'margoot', 'MLeitzelman', and 'speedykoala'. The 'Fans' section shows a list of users who are fans of the user, including 'ereteog'. Red arrows point from the text 'Connected users' to the 'People' section and from 'Fans' to the 'Fans' section.

delicious Home Bookmarks People Tags

What's New? Signed in as tal.icio.us | Inbox | Settings | Help | Sign Out

Search these bookmarks Search

Save a new bookmark Add to my Network

JeanSaul's Network Bookmarks | Network | Tags | Subscriptions

Also see more bookmarks in Popular or Recent.

JeanSaul Network Type a tag Bookmarks 2603

11 MAY 09 Litchi-PC SAVE 5
distributeur mini pc, barebone, accessoires pour mini pc, mini itx industriel, panel pc
gbrocker hardware

Cas Orange - iPhone SAVE 7
analyse de l'information sur l'iPhone
mleitzelman visualization graph RTGI

FORMATION AU WEB 2.0 APPLIQUE A L'INTELLIGENCE ECONOMIQUE - LE SUPPORT SAVE 2
le web 2 et l'IE
mleitzelman IE slideshare web2.0

10 MAY 09 Web 2.0 Diagrams : un album sur Flickr SAVE 35
diagramme de toutes les images de Dion Hinchcliffe sur le web 2
mleitzelman web2.0 diagram visualization

trendwatching.com's May 2009 Trend Briefing covering INNOVATION JUBILATION SAVE 39
liste de trends sur les idées de business
mleitzelman innovation business trends prospective

08 MAY 09 Linux Manua: 10 antidotes anti-Hadopi SAVE 175
speedykoala p2p tips antidote web privacy hadopi

07 MAY 09 coreboot SAVE 26
coreboot (formerly known as LinuxBIOS) is a Free Software project aimed at replacing the proprietary BIOS (firmware) you can find in most of today's computers. It performs just a little bit of hardware initialization and then executes a so-called payload.
gbrocker software linux opensource free programming hardware flash firmware system admin

Flashrom SAVE 74
flashrom is a utility for reading, writing, verifying and erasing flash ROM chips. It's often used to flash BIOS/EFI/coreboot/firmware images.
gbrocker software linux flash firmware free opensource debian admin system

06 MAY 09 FraonomiA ca SAVE

People Network 6
cpicard
gbrocker
iurics
margoot
MLeitzelman
speedykoala
New fan
Mutual fan

Fans 1
ereteog

Connected users

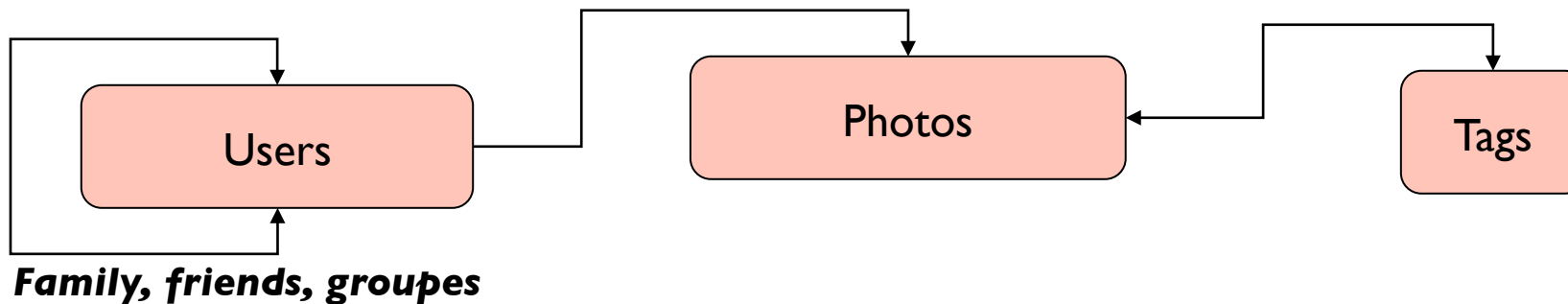
Fans

Flickr – social photo hosting and sharing

- Allows users to store photos, associate them keywords (tags), share photos with contacts or everyone
- Social features allow users to discover new users and photos, to communicate, to form communities

Flickr – social photo hosting and sharing

Main entities



- Photos
 - Title, description, tags, dates, geo-location
- Keywords
 - Tag name and description
- Users
 - alias, e-mail, name, homepage, profile, domains of interest

Flickr – social photo hosting and sharing

Searching for

- Photos
 - Description of photos or tags, or both
- Groups
 - Name and description or discussions
- Users
 - Name or profile

Flickr – social photo hosting and sharing

Geo tagging

The screenshot displays the Flickr map interface. At the top, navigation tabs include "Organiser par lot", "Albums", "Groupes", and "Carte". A search bar contains the text "eau 07500 Guierand-Granges". A map of the town of Guierand-Granges is shown, with a red location pin placed on "222 Avenue Georges Clemenceau". A pop-up window for this location shows a photo thumbnail and the text "bda2008 par talel.abdessalem". The map includes various street names, such as "av de la République" and "av Georges Clem", and road markers for "N532", "N7", and "A7/E74". The bottom of the interface shows a search bar with "RECHERCHER" and "Plus d'options", and a list of 3 elements with 0 selected.

Flickr – social photo hosting and sharing

Popular tags

Last 24h



Tags du moment

Au cours des dernières 24 heures

folsomstreetfair2008, nationalbookfestival, riverwood, wordcampdx, folsomstreetfair, day272, coolhandluke, barcampvancouver08, kubix, horseshit, festaalcel, hideseek, pp7, ayearofsundays, fg27092008, welshflickrmyru, wordonthestreet, 15p, maidenhead, skol

Au cours de la semaine passée

dilosep08, flyingfeet, crossmediaweek, ta08, startonthestreet, seacthebest, twestival, blogorlando, acl2008, gdd08mad, wind0mph, presidentialdebate, picnic2008, firstdayoffall, westergasterrein, attentionwhore, todaysart, atpny, holidaygifts, piromusical



Last week

Afficher : OK

Tags les plus populaires

Tag cloud



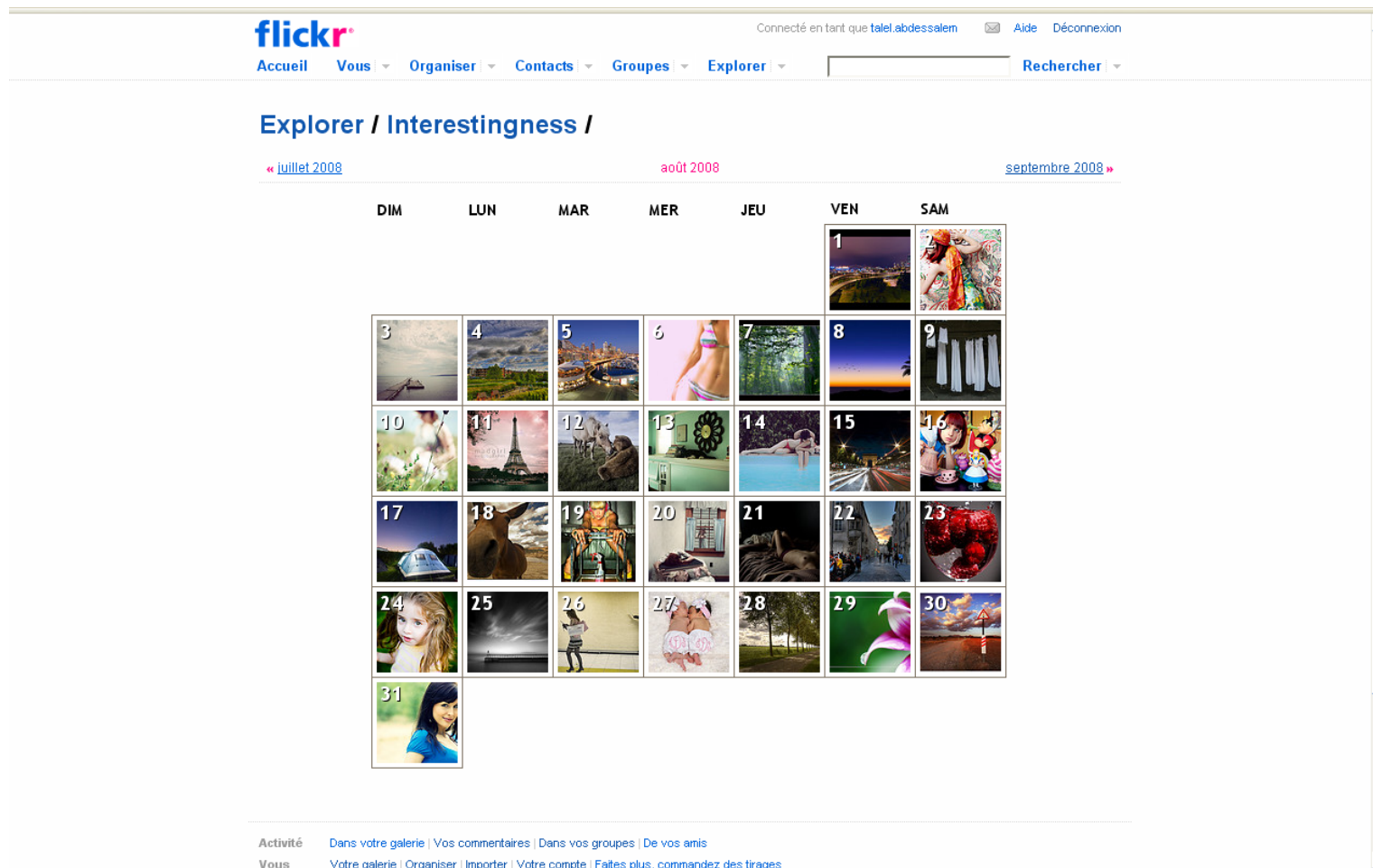
africa amsterdam animals architecture art august australia baby band barcelona beach
berlin bird birthday black blackandwhite blue boston bw california cameraphone camping
canada canon car cat chicago china christmas church city clouds color concert
cute dance day de dog england europe family festival film florida flower flowers
food football france friends fun garden geotagged germany girl girls graffiti green
halloween hawaii hiking holiday home house india ireland island italia italy japan july june
kids la lake landscape light live london macro may me mexico mountain mountains museum
music nature new newyork newyorkcity night nikon nyc ocean paris park
party people photo photography photos portrait red river rock rome san sanfrancisco
scotland sea seattle show sky snow spain spring street summer sun sunset
taiwan texas thailand tokyo toronto tour travel tree trees trip uk urban usa
vacation vancouver washington water wedding white winter yellow york zoo

Que sont les tags ?

Vous pouvez attribuer des tags à vos photos et vidéos, qui agiront comme des mots-clés ou des étiquettes de catégorie. Les tags vous

Flickr – social photo hosting and sharing

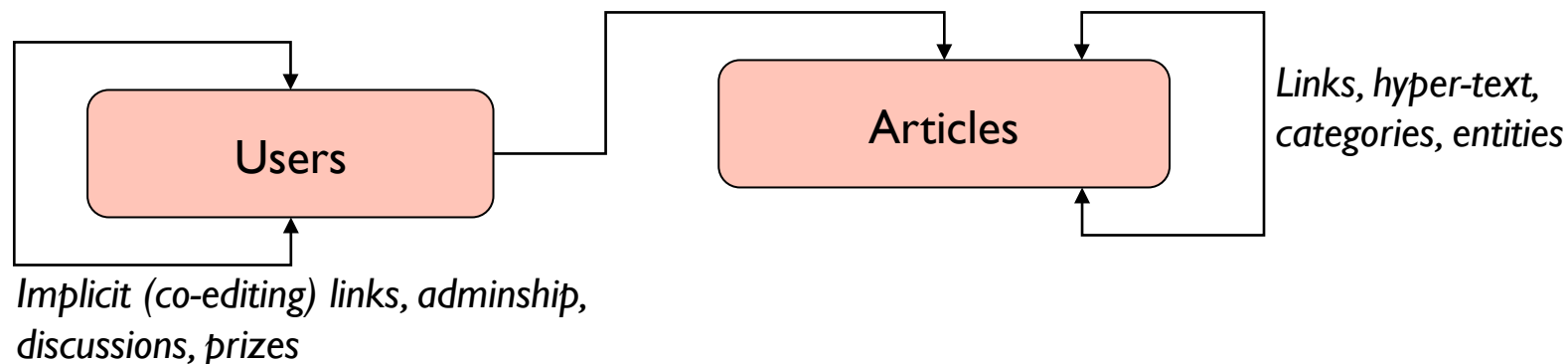
Interestingness : interesting photos -> provenance, comments, favorites lists, tags



Wikipedia – collaborative encyclopaedia

A network of articles, enriched with descriptions of entities.

Complete history of revisions is available.



Search: full-text search over articles, acces to revision history, access to list of contributors of an article.

Wikipedia – collaborative encyclopaedia



WIKIPÉDIA
L'encyclopédie libre

- Accueil
- Portails thématiques
- Index alphabétique
- Article au hasard
- Contactez Wikipédia
- Contribuer
- Premiers pas
- Aide
- Communauté
- Modifications récentes
- Faire un don
- Imprimer / exporter
- Boîte à outils
- Autres langues
- Català
- Cebuano
- Deutsch
- English
- Español
- Euskara
- Italiano
- Latina
- Lumbaart
- Bahasa Melayu
- Nederlands
- Occitan
- Polski
- Slovenčina
- Српски / Srpski
- Українська
- Tiếng Việt
- Volapük
- Winaray

Créer un compte ou se connecter

Article Discussion Lire Modifier Afficher l'historique Rechercher

Nos Conditions d'utilisation mises à jour entreront en vigueur le 25 mai 2012. En savoir plus.

Aussois

45° 13′ 38″ N 6° 44′ 32″ E (carte)



Cet article est une ébauche concernant une commune de la Savoie.
Vous pouvez partager vos connaissances en l’améliorant (comment ?); pour plus d’indications, visitez le projet Communes de France.
Consultez la liste des **tâches à accomplir** en page de discussion.

Aussois est une commune française, située dans le département de la Savoie et la région Rhône-Alpes.

Sommaire [masquer]
1 Géographie
1.1 Voies de communication et transports
2 Administration
3 Démographie
4 Toponymie
5 Histoire
6 Lieux et monuments
6.1 Barrière de l'Essillon
6.2 Sites touristiques
6.2.1 Sites naturels
6.2.2 Edifices
6.2.2.1 Structures touristiques
6.2.2.2 Sites industriels
7 Économie
7.1 Politique environnementale
7.2 Tourisme
7.3 Production agricole
8 Événements
9 Notes et références
10 Voir aussi
10.1 Article connexe
10.2 Liens externes

Géographie [modifier]

Station de sports d'hiver et d'été, située sur un plateau aux portes du Parc national de la Vanoise en Maurienne, dans le département de la Savoie.
Autrefois village traditionnel de montagne.
Latitude : 45° 13' N, Longitude : 6° 44' E, Altitude : 1489 m.

La station est gérée par une régie communale : 21 pistes balisées entre 1530 et 2675 m d'altitude, 55 km de pistes au pied de la Dent Parrachée (3 697 m). La station, exposée au sud, souffre parfois d'un manque d'enneigement, qui est néanmoins compensé par un bon équipement en neige de culture et la possibilité de skier à partir de la station intermédiaire, située à un peu plus de 2150 mètres d'altitude.

En 1970 est construite la route D215 à partir de Modane, menant ainsi directement au village et évitant ainsi l'ancienne route (D215 e, f & g) en comble des forts de l'Essillon, désormais fermée l'hiver.

En 2005, la station a réorganisé une partie importante de son domaine d'altitude par le remplacement d'un des plus longs téléskis de France (550 de dénivellée et plus de 2 km de long) par un télésiège débrayable 6 places. Le tracé étant différent, l'altitude maximale de la station a été réduite de quelque 20 mètres. Le sommet actuel de la station se trouve au sommet du télésiège 2 places de la Fournache à 2675 m d'altitude.

Des accords particuliers, conclus avec des stations alentours, permettent de skier à La Norma, Valfréjus, Val Cenise et Val Thorens grâce à la télécabine d'Ornillo.

La station profite d'une évolution économique très positive par la construction de plusieurs ensembles résidentiels de tourisme de standing, à proximité immédiate des pistes.

Voies de communication et transports [modifier]

La commune d'Aussois est reliée :

- vers l'ouest à Modane (à 7,5 km) par la route départementale D215. De là, en continuant vers l'ouest, on peut emprunter en direction de Chambéry l'autoroute à péage A43 (E70) ou la route D1006 (anciennement "RN6"), qui lui est parallèle. De Modane, on peut aussi se rendre en Italie (Bardonecche, à 21 km) par le tunnel du Fréjus.
- vers l'est par la D83 à Sardières (à 4,5 km) puis à Sollières (Aérodrome de Sollières-Sardières). De Sollières, on rejoint la route D1006 qui remonte la vallée de l'Arc, vallée qu'elle quitte à Lanslebourg pour atteindre la frontière italienne, via le col du Mont-Cenis (fermé en hiver).

Aussois



Centre ville d'Aussois

Administration	
Pays	France
Région	Rhône-Alpes
Département	Savoie
Arrondissement	Saint-Jean-de-Maurienne
Canton	Modane
Code commune	73023
Code postal	73500
Maire	Alain Marmézy
Mandat en cours	2008-2014
Site web	Aussois.com
Démographie	
Population	691 hab. (2008)
Densité	16 hab./km ²
Gentilé	Aussoyens / Aussoyennes
Géographie	
Coordonnées	45° 13′ 38″ Nord 6° 44′ 32″ Est
Altitudes	mini. 1120 m — maxi. 3600 m
Superficie	41,94 km ²

Localisation carte nationale [départements]



Wikipedia – collaborative encyclopaedia

Nos Conditions d'utilisation mises à jour entreront en vigueur le 25 mai 2012. En savoir plus.

Historique des versions de « Aussois »

Voir les opérations sur cette page

Naviguer dans l'historique

À partir de l'année (et précédentes) : À partir du mois (et précédents) : tous Filtrer les balises : Masqués seulement

Outils externes et statistiques

[Liste des auteurs](#) - [Rechercher l'auteur d'un passage de l'article](#) - [Modifications](#) - [Consultations](#) - [Nombre de contributeurs qui suivent cette page](#)

Autres discussions

[Suppression](#) - [Neutralité](#) - [Droit d'auteur](#) - [Article de qualité](#) - [Bon article](#) - [Lumière sur](#) - [À faire](#) - [Archives](#) - [Traduction](#)

Légende : (actu) = différence avec la version actuelle - (diff) = différence avec la version précédente - m = modification mineure

(dernière page | première page) Voir (50 plus récentes | 50 plus anciennes) (20 | 50 | 100 | 250 | 500).

Comparer les versions sélectionnées

- (actu | diff) 1 avril 2012 à 03:09 EmausBot (discuter | contributions) **m** . . (20 579 octets) (+15) . . (r2.7.2+) (robot Ajoute : sk:Aussois) (défaire)
- (actu | diff) 25 mars 2012 à 22:31 82.251.46.140 (discuter) . . (20 564 octets) (+19) . . (→Tourisme) (défaire)
- (actu | diff) 25 mars 2012 à 22:30 82.251.46.140 (discuter) . . (20 545 octets) (+42) . . (→Tourisme) (défaire)
- (actu | diff) 1 mars 2012 à 22:50 Gzen92 (discuter | contributions) **m** . . (20 503 octets) (-1) . . (défaire)
- (actu | diff) 22 février 2012 à 21:24 79.80.65.163 (discuter) . . (20 504 octets) (+762) . . (défaire)
- (actu | diff) 22 février 2012 à 20:59 86.197.208.179 (discuter) . . (19 742 octets) (+6) . . (→Histoire) (défaire)
- (actu | diff) 22 février 2012 à 20:58 86.197.208.179 (discuter) . . (19 736 octets) (+136) . . (→Histoire) (défaire)
- (actu | diff) 31 janvier 2012 à 13:38 Arrotta (discuter | contributions) . . (19 600 octets) (+4) . . (→Lieux et monuments) (défaire)
- (actu | diff) 26 janvier 2012 à 00:28 Wikialine (discuter | contributions) **m** . . (19 596 octets) (+94) . . (→Notes et références) (défaire)
- (actu | diff) 26 janvier 2012 à 00:27 Wikialine (discuter | contributions) **m** . . (19 512 octets) (-315) . . (→Démographie) (défaire)
- (actu | diff) 16 janvier 2012 à 14:21 HerculeBot (discuter | contributions) **m** . . (19 827 octets) (+18) . . (Résolution de redirection; changement de type cosmétique) (défaire)
- (actu | diff) 25 décembre 2011 à 13:08 Arrotta (discuter | contributions) . . (19 809 octets) (+1) . . (défaire)
- (actu | diff) 13 décembre 2011 à 09:49 147.100.145.242 (discuter) . . (19 808 octets) (+157) . . (ajout colloques bactériologie) (défaire)
- (actu | diff) 5 décembre 2011 à 20:41 EmausBot (discuter | contributions) **m** . . (19 851 octets) (+16) . . (r2.6.4) (robot Ajoute : lmo:Aussois) (défaire)
- (actu | diff) 27 novembre 2011 à 12:31 Lomita (discuter | contributions) **m** . . (19 635 octets) (-19) . . (WPCleaner (v1.09) Item de liste terminé par un saut de ligne (projet correction syntaxique)) (défaire)
- (actu | diff) 24 novembre 2011 à 13:56 Sergelucas (discuter | contributions) **m** . . (19 648 octets) (+82) . . (Corrections de forme) (défaire)
- (actu | diff) 24 novembre 2011 à 11:10 Sergelucas (discuter | contributions) . . (19 586 octets) (+493) . . (→Lieux et monuments) (défaire)
- (actu | diff) 21 novembre 2011 à 09:09 Luckas-bot (discuter | contributions) **m** . . (19 073 octets) (+15) . . (r2.7.1) (robot Ajoute : la:Aussois) (défaire)
- (actu | diff) 14 novembre 2011 à 20:56 EmausBot (discuter | contributions) **m** . . (19 058 octets) (+15) . . (r2.6.4) (robot Ajoute : oc:Aussois) (défaire)
- (actu | diff) 27 septembre 2011 à 12:56 147.100.145.242 (discuter) . . (19 043 octets) (+10) . . (ajout nom centre CNRS) (défaire)
- (actu | diff) 27 septembre 2011 à 12:51 147.100.145.242 (discuter) . . (19 033 octets) (+131) . . (ajout précision colloque) (défaire)
- (actu | diff) 14 septembre 2011 à 19:14 93.25.86.240 (discuter) . . (18 802 octets) (+20) . . (→Administration) (défaire)
- (actu | diff) 13 septembre 2011 à 01:11 ZetudBot (discuter | contributions) **m** . . (18 882 octets) (-1) . . (→Administration : - espace avant Ref., replaced: commune <ref> → commune<ref>) (défaire)
- (actu | diff) 27 juillet 2011 à 18:40 78.232.95.206 (discuter) . . (18 883 octets) (+43) . . (défaire)
- (actu | diff) 26 juillet 2011 à 14:35 Bernarddb (discuter | contributions) . . (18 840 octets) (+356) . . (→Démographie : écoquartier de la Fintan) (défaire)
- (actu | diff) 26 juillet 2011 à 14:15 Bernarddb (discuter | contributions) . . (18 484 octets) (+7) . . (→Démographie) (défaire)
- (actu | diff) 26 juillet 2011 à 14:14 Bernarddb (discuter | contributions) . . (18 477 octets) (+13) . . (→Démographie) (défaire)
- (actu | diff) 26 juillet 2011 à 14:13 Bernarddb (discuter | contributions) . . (18 464 octets) (0) . . (→Démographie) (défaire)
- (actu | diff) 26 juillet 2011 à 14:12 Bernarddb (discuter | contributions) . . (18 464 octets) (+6) . . (→Démographie) (défaire)
- (actu | diff) 26 juillet 2011 à 14:10 Bernarddb (discuter | contributions) . . (18 458 octets) (+210) . . (→Démographie) (défaire)
- (actu | diff) 26 juillet 2011 à 12:33 Bernarddb (discuter | contributions) . . (18 248 octets) (-13) . . (→Tourisme) (défaire)
- (actu | diff) 26 juillet 2011 à 12:16 Bernarddb (discuter | contributions) . . (18 261 octets) (+48) . . (→Tourisme : PTT) (défaire)
- (actu | diff) 26 juillet 2011 à 01:16 Bernarddb (discuter | contributions) . . (18 213 octets) (+393) . . (→Histoire) (défaire)
- (actu | diff) 18 juillet 2011 à 20:06 78.232.95.206 (discuter) . . (17 820 octets) (+62) . . (→Administration) (défaire)

Wikipedia – collaborative encyclopaedia

Utilisateur:Gzen92

- Projets de Maintenance :
- Illustration des communes
 - Orthographe à vérifier
 - Pages orphelines
 - Date naissance/décès
 - Articles à fusionner
 - Suppression d'articles
 - Restauration lien rouge

- Liens utiles :
- [Le Bistro](#)
 - [Annonces](#)
 - [Requête aux administrateurs](#)
 - [Statistiques France](#)
 - [Nombre d'éditeurs](#)
 - [Fautes d'orthographe](#)
 - [Upload sur Commons](#)

scripts	bandeaux
~~~~	#redirect[[xxx]]
[[pagetexte]]	{{...}} (vide / incomplet)
[http://xxx texte]	{{ébauche}}
<ref> ... </ref>	{{Article principal NOM}}
<references/>	{{Article connexe NOM}}
<gallery>	{{suppression}}
Fichier:nom alt= texte	{{orthographe}}
</gallery>	{{à wikifier date=xxx}}
=== ... xxx === ...	[[en:article]] (autre langue)
<span> xxx </span>	{{langue alttexte}} (autre langue)
*xxx*	{{homonymie}}
" xxx "	{{voir homonymes nom}}
... ..	{{référence souhaitée texte date}}
*** ...	{{tit1 ...}}... (tr tableau avec autre nom)
### ...	[[Catégorie:...]]
{{s XV}}	[[Catégorie:Naissance/Décès en ...]]
{{s mini XV}}	[[Catégorie:Date de naissance/décès inconnue (...e siècle)]]
{{refnec ...}}	[[Catégorie:Naissance/Décès au ...e siècle av. J.-C.]]

Base de travail [modifier](#)

**Haut-Rhin**  
J'habite dans le Haut-Rhin. ▾

 Je contribue à Wikipédia depuis **6 ans, 10 mois et 19 jours**

 **Monuments historiques**  
Je participe au projet sur les monuments historiques. Mais je n'en suis pas un pour autant ! ▾

 **Maintenance**  
Je participe à la maintenance de Wikipédia. ▾

**90 000+** Cet utilisateur a effectué plus de **90 000** contributions sur Wikipédia.

 <?php echo("Bonjour !"); ?> ▾  
Je parle PHP.

 **SQL**  
J'aime coder en SQL. ▾

 *print("Hello World!");* ▾  
Je parle couramment le C.

 **Mathématiques**  
J'aime les mathématiques. ▾

 **Non au SMS**  
Je ne supporte pas le langage SMS. ▾

 **Facebook**  
J'ai horreur de Facebook. ▾

 Je n'ai pas tué John Fitzgerald Kennedy le 22 novembre 1963... Ou alors je ne m'en souviens pas... ▾

## Contributions de l'utilisateur

Pour 79.80.65.163 ([discuter](#) | [journal des blocages](#) | [imports](#) | [journaux](#) | [journal des filtrages](#)). Les contributions supprimées ne sont pas affichées.

Rechercher les contributions

Ne montrer que les contributions des nouveaux utilisateurs

Adresse IP ou nom d'utilisateur :

Espace de noms :   Inverser la sélection  Espace de noms associé

Filtrer les balises :

Masqués seulement  Ne montrer que les articles dont je suis le dernier contributeur

À partir de l'année (et précédentes) :  À partir du mois (et précédents) :

2012, Aussois.

# Roadmap

- Concepts, definitions, taxonomy
- Exemples of applications
- Models for social data
- Research challenges
- Predicting signed social links, inference from interactions
- Top-k search in social applications

# Social network models

- As an undirected graph of users: adapted when relationships are symmetrical (Facebook, LinkedIn)
  - Also for user similarity (e.g., tagging in Delicious)
- As an directed graph of users: adapted in asymmetrical settings like Twitter
  - Also the model of Wikipedia's article graph
- As bipartite, n-partite graph: photos-tags-users in Flickr, products-reviews-users in Epinions.
- Edges may be labeled: e.g., Taobao marketplace trading, messages, contacts.

# Explicit vs. implicit networks

Traditional social network analysis distinguishes between pairs of people that are linked or not.

But interactions in social media are much richer -> may reveal new kinds of relationships (implicit), strength, trust/distrust, similarity/antagonism, friends/foes

- Implicit network of editors of the Wikipedia
- Similarity in tagging in Delicious

Some applications have explicit signed (+ or -) links: trust/distrust in Epinions, friends/foes in Slashdot.

# Some characteristics of social graphs

- Small-world phenomenon: distance between any two users is small (logarithmic in the size of the graph)
- Sparse graphs: much fewer edges than in complete graph (Facebook users has  $\sim 100$  friends in average)
- High-transitivity: if A is connected to B and B is connected to C, then A is more likely to be connected to C.
- Degree distribution: often power-low.
- Captured often by preferential-attachment models for random graph generation

# Roadmap

- Concepts, definitions, taxonomy
- Exemples of applications
- Models for social data
- Research challenges
- Predicting signed social links, inference from interactions
- Top-k search in social applications



# Challenges (1)

- How to collect massive amounts of social media data, deal with incomplete and missing data
- How to extract structured data from unstructured user-generated content
- How to anonymize and publish social data
- How to find the flow of interesting pieces of information that are disseminated among users
- How to find relevant content, how to identify authoritative or influential users
- How to quantify the influence of users
- How to recommend and predict individual social links
- How to model/predict global structure
- How to go beyond link/ no link – type (sign), weight
- How to identify in real-time emerging topics of discussion

# Challenges (2)

- How to model the patterns by which information evolves over time
- How to predict the popularity of a piece of information
- How to identify implicit networks, for data diffusion, influence
- How to filter out the spam/offensive content
- How to rank social media based on relevance or importance
- How can social data improve search, question answering, entity disambiguation, etc.
- How to identify groups/communities, topics and sentiment
- How do social networks shape purchasing decisions
- ...

Need to scale to large datasets, real-time / online constraints.

# Search & recommendation in social media

- Users consume and create information -> use global importance measures (a la PageRank) to rank users and the data they produce
- Use social links and social data to improve Web search (e.g., use Delicious)
- Recommend users and data based on the social network.
- Use the social links to improve the search quality: social-aware search - > results « biased » to the seeker's social network.

# Social link prediction

Facebook's « people you may know » :  
friend-of-friend -> many common friends

Many possible criteria:

- distance in the graph
- common neighbours
- Jaccard coefficient
- Adamic/Adar coefficient
- preferential attachment
- profiles and user attributes

# Link inference

Often links are more than just ties: strength, may indicate sentiment or opinion

Users do have rich interactions, do express positive and negative attitudes/opinions through actions: rating a product, editing one's text in a Wikipedia article, pressing the like button, commenting, reviewing.

Opinions about

- products: imdb, amazon.
- people: epinions
- items created by others: Yahoo Answers, StackOverflow

# Understanding information flow

Social media is disseminated through social interactions -> understand information dynamics and consumption

Real-time spread of information, ideas, influence, opinion, decision -> can be modelled by **cascade graphs**

Motivation examples:

- select trend setters for viral marketing
- detect big stories before they become one, which websites to follow to get important stories
- predicting information attention and popularity
- which blogs to read to be up-to-date

# Identifying information flow

First challenge: identify and track units of information corresponding to pieces of information (events, articles, entities, etc): the « contagion »

Bloggers write posts and refer (link) to other posts -> cascading hyperlinks

Twitter: users generate streams of tweets, users subscribe to follow streams of others

- Trace the spread of hashtag
- Trace the spread of a particular URL
- Re-tweets

# Tracking information flow

Second challenge:

- how do messages spread,
- how to predict it,
- how to identify networks over which the messages spread.

Tracking information through implicit networks -> we do not see who « infected » whom

Infer a diffusion network: find the optimal network that best explains the observed infection times.



# Roadmap

- Concepts, definitions, taxonomy
- Exemples of applications
- Models for social data
- Research challenges
- Predicting signed social links, inference from interactions
- Top-k search in social applications

# A Wikipedia study [WWW2011, DBSocial2011]

Inference of a **signed** social network  
from **interactions** in Wikipedia.

# Context

A plethora of Web-based applications for collaborative publishing and sharing data (Wikipedia, Flickr, YouTube).

Often these have an underlying social network (user relationships)

- can be exploited at the application level: recommendation, search, access control, etc.
- even more valuable when relationships are **signed** (i.e., indicative of attitude)

Epinions ([www.epinions.com](http://www.epinions.com)) - trust / distrust tags

Slashdot ([www.slashdot.com](http://www.slashdot.com)) - friend / foe tags

# Epinions

### Web of Trust

**Suzi trusts:**

- [quasar](#)
- [2buzy](#)
- [jdhauer](#)
- [jo.com](#)
- [tombarnes](#)

[View all 67 members whom Suzi trusts](#)

**Suzi is trusted by:**

- [disartain](#)
- [mrbean200200](#)
- [jdomoff](#)
- [gaelkm](#)
- [pmontey](#)

[View all 147 members who trust Suzi](#)

### Suzi's Profile



**About Suzi**  
(POPULAR AUTHOR) - Top 200

Epinions.com ID: **Suzi**  
 Location: **California**  
 Member Since: **Feb 20 '00**

**Activity Summary**  
 Reviews Written: **190**  
 Member Visits: **12,566**  
 Total Visits: **401,808**

Suzi is an active trustor and enjoys scuba

### Web of Trust

 [Trust Suzi](#)  
 [Block Suzi](#)  
[Whom should I trust?](#)

Product / Topic	Product Rating	Review Rating
<a href="#">Royal Caribbean Monarch of the Seas in Cruises</a>	★★★★★	Very Helpful
<a href="#">Dawn Princess in Cruises</a>	★★★★☆	Very Helpful
<a href="#">Royal Caribbean Majesty of the Seas in Cruises</a>	★★★★★	Very Helpful
<a href="#">Liberty of the Seas in Cruises</a>	★★★★★	Very Helpful

# Motivation

In collaborative, social applications: often interactions are indicative of trust or affinity, distrust or antagonism.

Motivating a new line of research:

- very few social applications have explicit signed links
- but these might be deduced from previous interactions between users in the network

# Signed social networks

Trust and distrust: a subjective measure of the relation between users in a social network

- positive or negative (**trust** or **distrust**)
  - trust (also proximity, similarity, affinity)
  - distrust (also dissimilarity, antagonism)
- **explicit** (declared by the users of an online system) or **implicit** (inferred from interactions between users)

# Our thesis

User interactions in online social applications can provide good indicators of implicit relationships.

We applied this idea on Wikipedia, building a signed network of its **editors**

- a **local model** of relationships: from a link generator to a link recipient
- can be interpreted as “trust to improve the Wikipedia”

First work to consider the inference of a signed network based on interactions in social media.

# Wikipedia interactions

Extracted interactions between contributors:

- **operations on text**: amount of text inserted, deleted and replaced
- **reverts and restores**: we keep the count of each of these for each contributor pair
- **votes**: votes in the administrator elections
- **barnstars**: prizes given on the user pages of the authors



# Revisions

- (cur | prev) ● 18:10, 15 May 2011 TpbBradbury (talk | contribs) (116,358 bytes) (*external link more appropriate for battle of waterloo article, remove circular link*)
- (cur | prev) ● 17:59, 15 May 2011 TpbBradbury (talk | contribs) m (116,667 bytes) (*correct to article title*)
- (cur | prev) ● 04:03, 5 May 2011 Farslayer (talk | contribs) (116,671 bytes) (*Removed ad for book again.*)
- (cur | prev) ● 03:30, 5 May 2011 Tinu1 (talk | contribs) (117,108 bytes)
- (cur | prev) ● 01:25, 5 May 2011 ProperlyRaised (talk | contribs) (116,495 bytes) (*Let's not speculate*)
- (cur | prev) ● 01:22, 5 May 2011 ProperlyRaised (talk | contribs) m (116,618 bytes) (*Removed a commercial - as in, a plug for a book.*)
- (cur | prev) ● 23:25, 4 May 2011 Tinu1 (talk | contribs) (117,247 bytes)
- (cur | prev) ● 19:45, 4 May 2011 Benea (talk | contribs) (116,616 bytes) (*rv, overly promotional examination of a single book, not suitable for the lead*)
- (cur | prev) ● 19:43, 4 May 2011 Tinu1 (talk | contribs) (117,215 bytes)
- (cur | prev) ● 14:58, 3 May 2011 RibotBOT (talk | contribs) m (116,616 bytes) (*r2.6.5*) (*robot Modifying: sa:नेपोलियन बोनापार्ट*)
- (cur | prev) ● 03:34, 2 May 2011 Koavf (talk | contribs) m (116,604 bytes) (*new key for Category:Napoleon I: " " using HotCat*)
- (cur | prev) ● 16:06, 26 April 2011 CrimeCentral (talk | contribs) (116,613 bytes)
- (cur | prev) ● 13:04, 26 April 2011 TpbBradbury (talk | contribs) (116,612 bytes) (*→Reforms: error in last edit deleted the word 'education'*)
- (cur | prev) ● 10:07, 26 April 2011 Ohconfucius (talk | contribs) (116,602 bytes) (*align date formats by script; per WP:MOS, WP:MOSTEXT and WP:Linking*)
- (cur | prev) ● 15:39, 24 April 2011 PragmaticStatistic (talk | contribs) m (116,658 bytes)
- (cur | prev) ● 13:07, 16 April 2011 Picaballo (talk | contribs) (116,378 bytes) (*+ His personal Imperial Flag*)
- (cur | prev) ● 08:19, 13 April 2011 Deposuit (talk | contribs) (116,271 bytes) (*→Titles: not yet Duc de Plaisance*)
- (cur | prev) ● 22:44, 11 April 2011 Tktru (talk | contribs) (116,246 bytes) (*→Titles, styles, honours and arms*)
- (cur | prev) ● 19:37, 11 April 2011 AndreasJS (talk | contribs) m (116,244 bytes) (*→Marriages and children: fix wikilink*)
- (cur | prev) ● 16:18, 9 April 2011 TpbBradbury (talk | contribs) (116,244 bytes) (*→Hundred Days: face in towards text wp:mos*)
- (cur | prev) ● 20:39, 8 April 2011 Mathiasrex (talk | contribs) m (116,247 bytes) (*→Hundred Days*)
- (cur | prev) ● 20:37, 8 April 2011 Mathiasrex (talk | contribs) (116,243 bytes) (*→Hundred Days: image*)
- (cur | prev) ● 11:20, 6 April 2011 TpbBradbury (talk | contribs) m (115,990 bytes) (*formatting*)
- (cur | prev) ● 10:28, 6 April 2011 TpbBradbury (talk | contribs) (115,994 bytes) (*disambiguate civil law. images need to go after main/see also links due to wp:accessibility. remove duplicate links*)

# Adminship votes

## Wikipedia:Requests for adminship/Diannaa

From Wikipedia, the free encyclopedia

< Wikipedia:Requests for adminship

The following discussion is preserved as an archive of a **successful** request for adminship. *Please do not modify it.*

### Contents [hide]

- 1 Diannaa
  - 1.1 Nomination
  - 1.2 Questions for the candidate
  - 1.3 General comments
  - 1.4 Discussion
    - 1.4.1 Support
    - 1.4.2 Oppose
    - 1.4.3 Neutral

### Diannaa [edit]

Final (90/10/9); ended 06:33, 28 October 2010 (UTC) - …[日本蕨?](#) · [投稿](#) · [Talk to Nihonjoe](#) · [Join WikiProject Japan!](#) 06:33, 28 October 2010 (UTC)

### Nomination [edit]

[Diannaa](#) ([talk](#) · [contribs](#)) – This is my second attempt at nominating someone else for adminship. Ladies and Gentlemen, I would like to give Diannaa the administrative functions. She's been around for just over a year, and since then has made over 12,000 edits, with nearly 60% of them into the article space. Now, the first time I came across this user was when she was doing lots of rollback, giving warnings to vandals and even warns rollbackers about any questionable revert they make. (See [this](#) [[↗](#)] one which I did wrong recently!)

Apart from rollbacking she's also done a lot of article promoting. She promoted four lists to featured status and one article to featured; all of which you can see on [her userpage](#).

With her experience sufficient enough to know many areas of Wikipedia and the fact that she has a clean block log during her year I would hope that the community agrees with my decision that Diannaa would become one of Wikipedia's newest administrators. [Minimac](#) ([talk](#)) 10:00, 20 October 2010 (UTC)

*Candidate, please indicate acceptance of the nomination here:* I have thought it over and have decided to accept the nomination. I would like to point out that the featured articles etc shown on my user page were not written by me; this is material I helped other people promote through copy editing efforts. --[Diannaa](#) ([Talk](#)) 03:43, 21 October 2010 (UTC)

### Questions for the candidate [edit]

Dear candidate, thank you for offering to serve Wikipedia as an administrator. Please answer these questions to provide guidance for participants:

1. What administrative work do you intend to take part in?

**A:** I would like to help out with the articles that are tagged as copyright violations. There is often a large backlog of work there and I think my research skills and attention to

# Barnstars

Good job

[\[edit\]](#)



## The Working Man's Barnstar

For your work with the last [Award Center Collaboration of the Fortnight](#), I hereby award you this barnstar. Thank you for your improvements to the article on [United States-Australia Relations](#). --Sharkface^T/C 18:20, 27 April 2008 (UTC)

Barnstarred

[\[edit\]](#)



## The Random Acts of Kindness Barnstar

For curtailing terpidiot vandalism on my userpage, I, Matt, award you, Enigma, this RAoK Barnstar. Rock on. [Non Curat Lex \(talk\)](#) 20:25, 6 May 2008 (UTC)

Congrats

[\[edit\]](#)



## The Original Barnstar

I hereby award you this barnstar for your edits and also your highly intellectual philosophy. Thank You! [Buddha24 \(talk\)](#) 05:54, 13 May 2008 (UTC)

# Aggregated interactions

+	-	-	+	-	+	-	+
insert	replace	delete	restores	reverts	support	oppose	barnstar
<hr/>			<hr/>		<hr/>		<hr/>
operations on article text			operations on article revisions		adminship elections (RFAs)		user pages

- 563 articles extracted from the Politics domain (910,209 revisions by 197,798 contributors)
- number of aggregated interaction vectors was 17,262,082
- sign of the link was decided via a simple voting heuristic
- final network (WikiSigned): 138,592 nodes, 740,397 edges (87.9% positive)

A much broader study (complete English Wikipedia) under way.

# Validating WikiSigned

Evaluation with respect to social theories: balance and status

Accuracy in predicting new links:

- link sign prediction over our network
  - cut out one A -> B link, predict it from the remaining ones
  - features based on the types of link triads involving A and B

Comparing WikiSigned with explicit networks:

- cross learning-prediction of other signed networks

Application-level validation:

- predict the quality and importance of articles using WikiSigned

# Link prediction accuracy

The predictive power using **triads** has been previously tested in explicit signed networks.

We learned a predictor for link signs over WikiSigned:

- feature vector consisting of the number of triads of each type
- logistic regression model, balanced dataset, 10-fold cross validation
- each triad type is given a coefficient by the trained model

**Good prediction accuracy:** 0.852 with an AUC of 0.924.

# Predicting the explicit networks

Same prediction model: do cross learning-predicting with the other networks.

	Epinions	Slashdot	Elections	WikiSigned
Epinions	0.926	0.905	0.787	0.765
Slashdot	0.929	0.806	0.792	0.716
Elections	0.922	0.895	0.814	0.775
WikiSigned	0.882	0.839	0.755	0.852

Predictive accuracy of training on the row data and testing on the column data.

Similar properties both at local and global level to the existing (explicit) signed networks.

# Classifying articles

- 5 article qualities (Featured Articles, Great Articles, A-class articles, B-class articles, C-class articles)
- 4 article importance levels (Top, High, Medium, Low)

Regression learning on a feature vector

- number of **contributors**
- **outgoing** links, **incoming** links, **inside** links,
- the **proportion** of incoming positive and negative links of contributors

Knowledge about the link structure of articles improves the classification of articles.

Features	Importance	Quality
contributors	0.691	0.518
contribs. + links	0.743	0.835
contribs. + soc. links	0.749	0.895
contribs. + soc. links + rep.	0.756	0.935



# Roadmap

- Concepts, definitions, taxonomy
- Exemples of applications
- Models for social data
- Research challenges
- Predicting signed social links, inference from interactions
- Top-k search in social applications

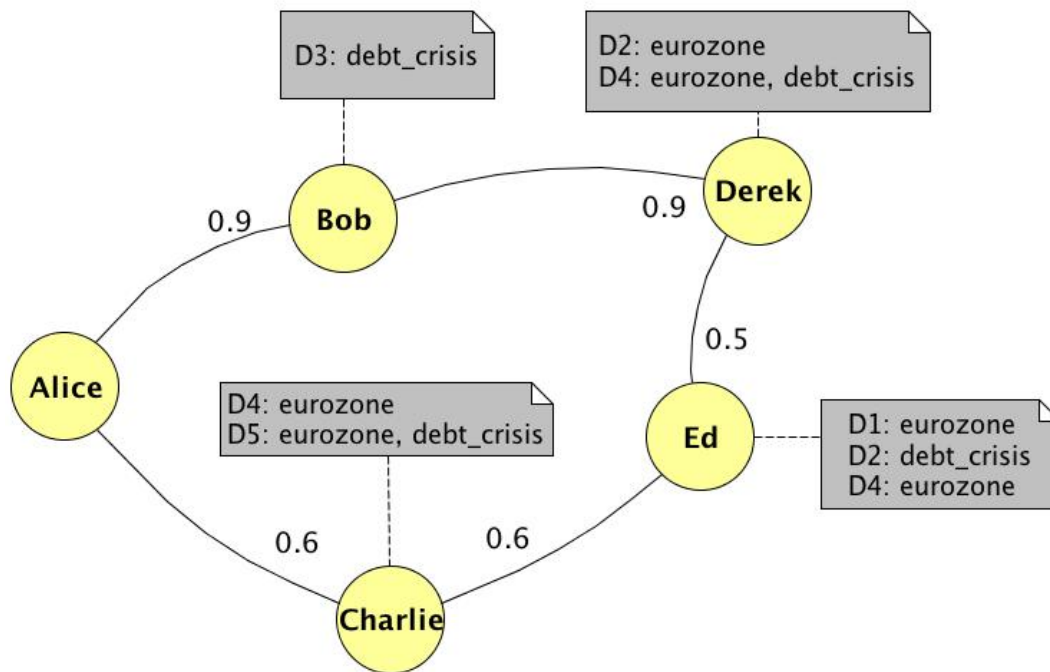
# A study on social tagging applications

[SIGMOD2012 & under review]

**Efficient** top-k search in **real**  
applications, allowing full **personalization**.

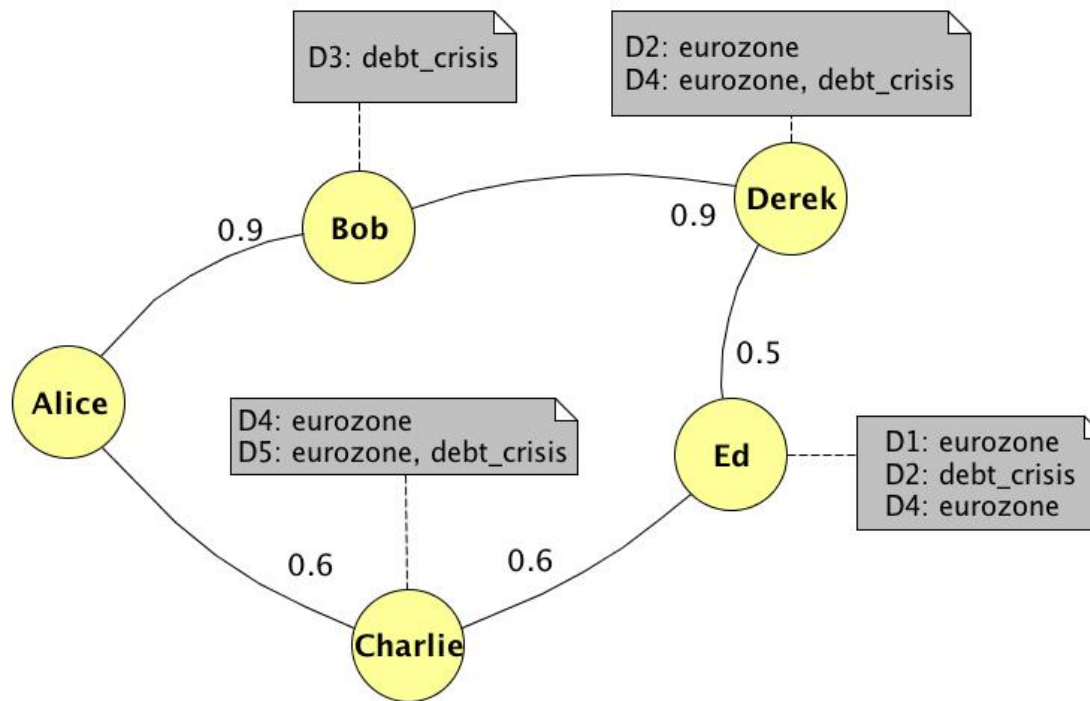
“Keep, share, discover the best of the Web”  
(Del.icio.us motto)

# Example



Alice asks for top 2 documents  
 «eurozone debt_crisis»

# Example



Alice asks for top 2 documents  
 «eurozone debt_crisis»

D1

eurozone

D2

eurozone

debt_crisis

D3

debt_crisis

D4

eurozone

debt_crisis

eurozone

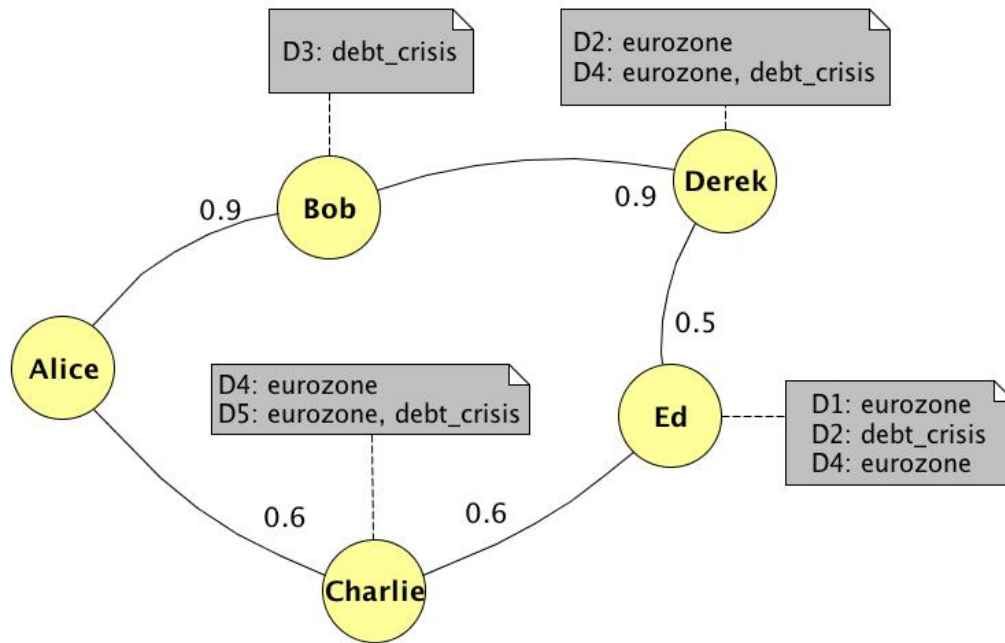
eurozone

D5

eurozone

debt_crisis

# Inverted index (classic Web search)

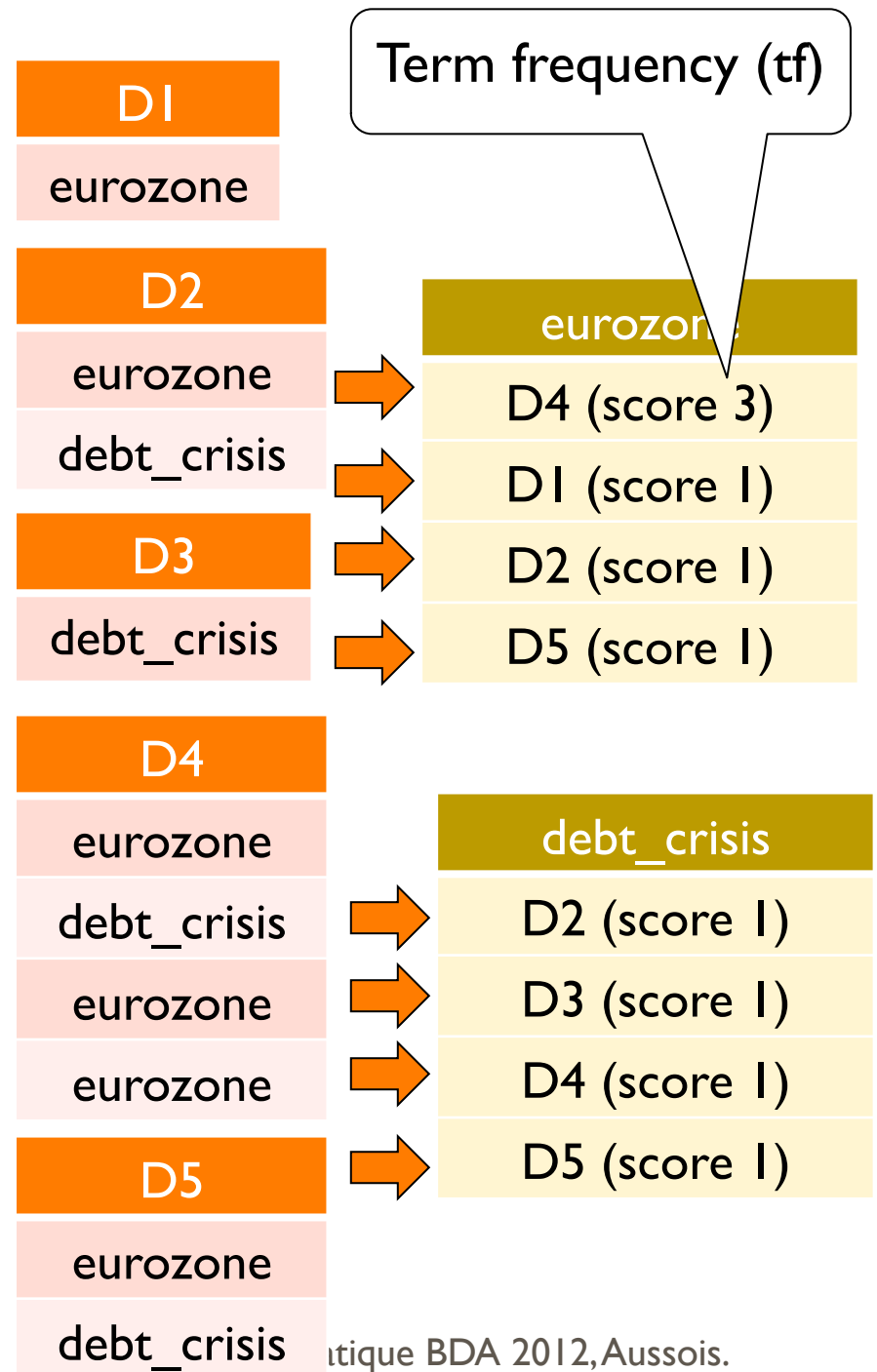


Alice asks for top 2 documents  
«eurozone debt_crisis»

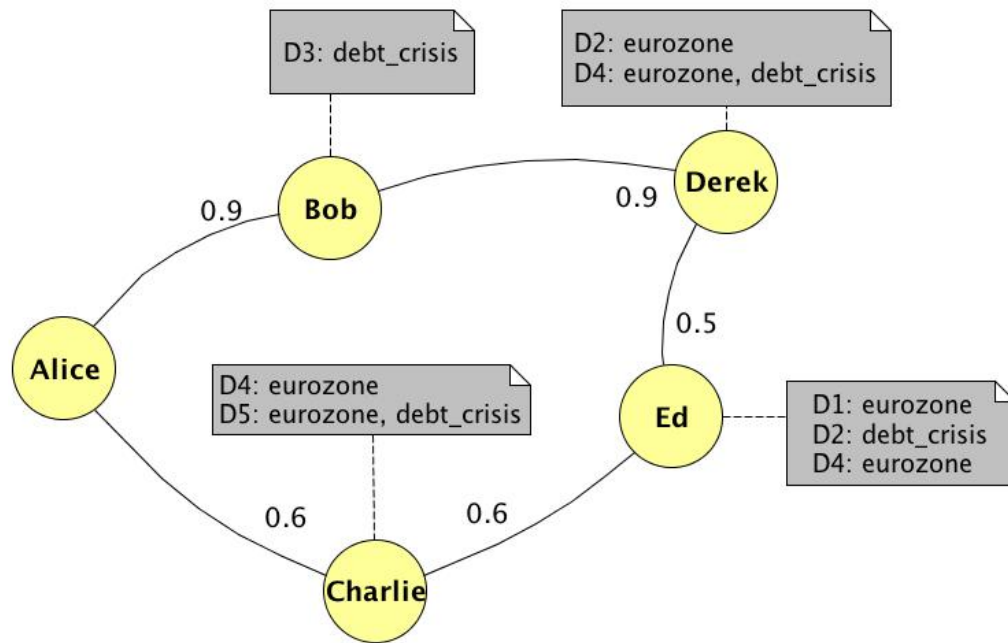
top-2 answer

**D4** : score 4

**D2** : score 2



# A social-aware search interpretation?



Alice asks for top 2 documents  
«eurozone debt_crisis»

D1	author	proximity
eurozone	Ed	?

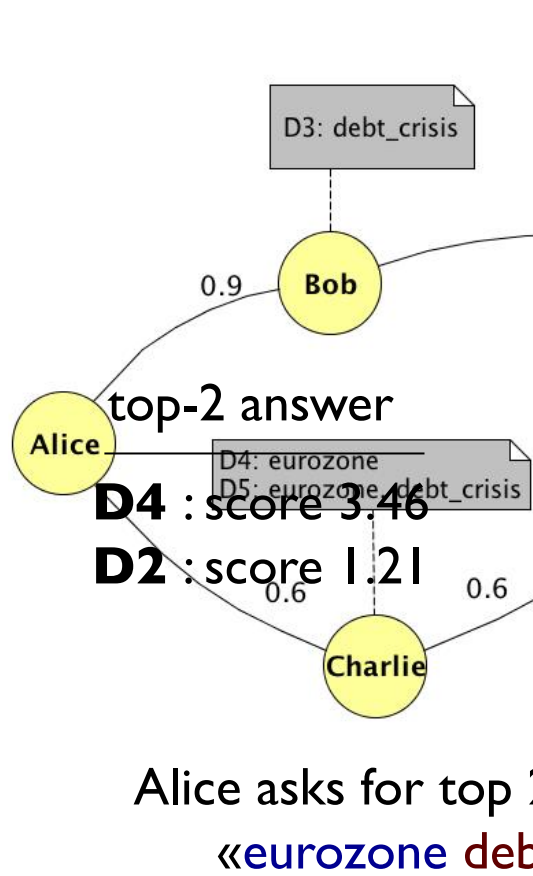
D2	author	proximity
eurozone	Derek	?
debt_crisis	Ed	?

D3	author	proximity
debt_crisis	Bob	0.9

D4	author	proximity
eurozone	Derek	?
debt_crisis	Derek	?
eurozone	Charlie	0.6
eurozone	Ed	?

D5	author	proximity
eurozone	Charlie	0.6
debt_crisis	Charlie	0.6

# A social-aware search interpretation?



eurozone
D4 (score 2.65)
D2 (score 0.81)
D5 (score 0.6)
D1 (score 0.4)

debt_crisis
D3 (score 0.9)
D4 (score 0.81)
D5 (score 0.6)
D2 (score 0.4)

D1	author	proximity
eurozone	Ed	$0.9 \times 0.9 \times 0.5$

D2	author	proximity
eurozone	Derek	$0.9 \times 0.9$
debt_crisis	Ed	$0.9 \times 0.9 \times 0.5$

D3	author	proximity
debt_crisis	Bob	0.9

D4	author	proximity
eurozone	Derek	$0.9 \times 0.9$
debt_crisis	Derek	$0.9 \times 0.9$
eurozone	Charlie	0.6
eurozone	Ed	$0.9 \times 0.9 \times 0.6$

D5	author	proximity
eurozone	Charlie	0.6
debt_crisis	Charlie	0.6

# Problem overview

- Users form a **weighted social network** (may reflect proximity, similarity, friendship, trust, etc)
- User **tag items** (e.g., documents, URLs, photos, etc) from a public pool of items

Examples: Del.icio.us, StumbleUpon, Digg, Reddit, IBM's Lotus Connections (a form of crowdsourcing); similar facilities in many other social applications

**Users search for items having certain tags (top-k)**

Why is this different from search on the Web?

- item relevance depends on the **proximity of taggers to the seeker** (users are both producers and consumers of information)
- this is **search with a context**



# Setting

**Online setting:** social network, tagging data, the seekers' search ingredients (parameters, model) can change at any moment

Few works in the literature: either assume certain model restrictions or cannot deal with the scale and dynamic context of real applications.

Main challenges: efficiency and applicability in real-world scenarios.

Inputs:

- the per-tag inverted lists
- a *Tagged* relation:  $\text{Tagged}(u,i,t)$  triples
- the weighted social network

All this data must reside on disk. I/O costs!

# Score model (per-tag)

Replace **term frequency (tf)** with a frequency measure depending on the seeker:

$$\text{freq}(\text{item}|\text{seeker},\text{tag}) = \alpha * \text{tf}(\text{tag},\text{item}) + (1-\alpha) * \text{sf}(\text{item}|\text{seeker},\text{tag})$$

- parameter  $\alpha$  in  $[0, 1]$  interval
- **social frequency**  $\text{sf}(\text{item}|\text{seeker},\text{tag})$  defined as:

$$\text{sf}(\text{item}|\text{seeker},\text{tag}) = \sum_{\text{user s.t. Tagged}(\text{user},\text{item},\text{tag})} \text{proximity}(\text{seeker},\text{user})$$

Two extreme cases:

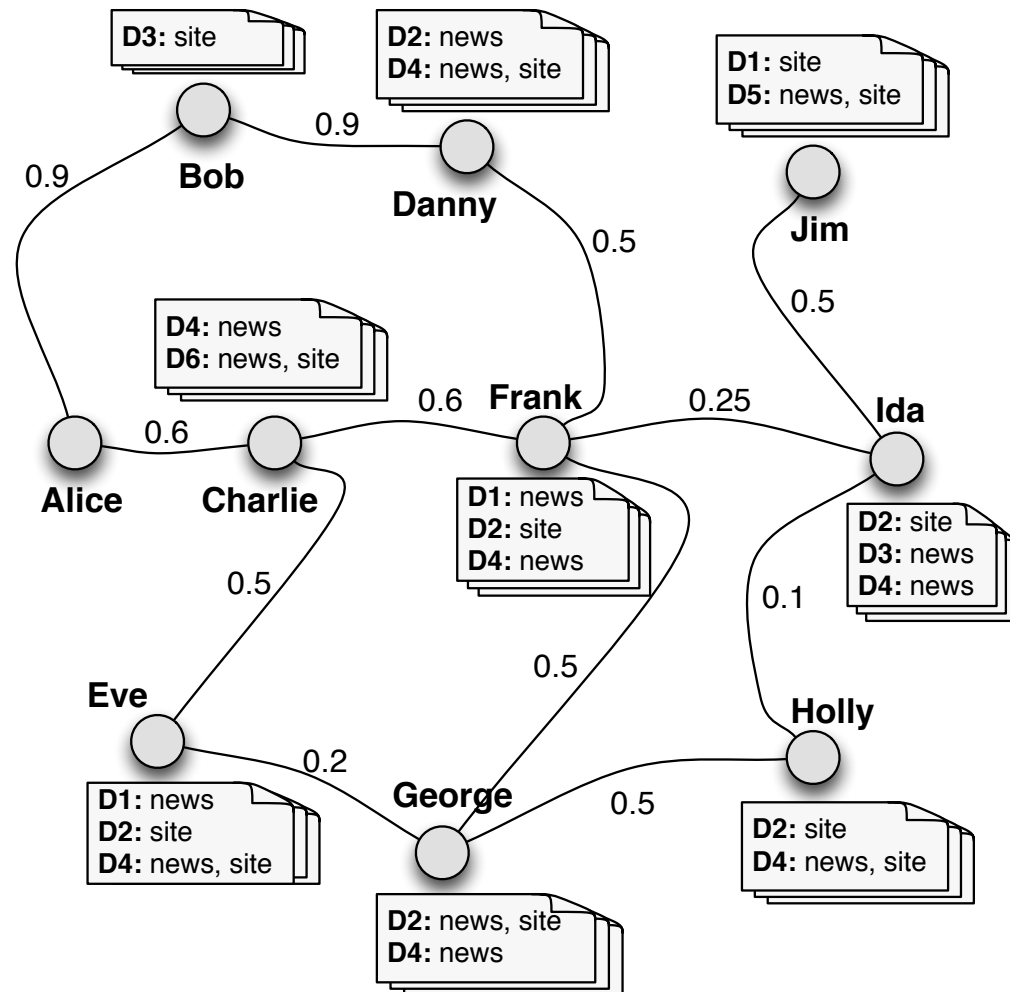
- $\alpha=1$  -> Web search
- $\alpha=0$  -> exclusively social search

# Extending the model to the entire network

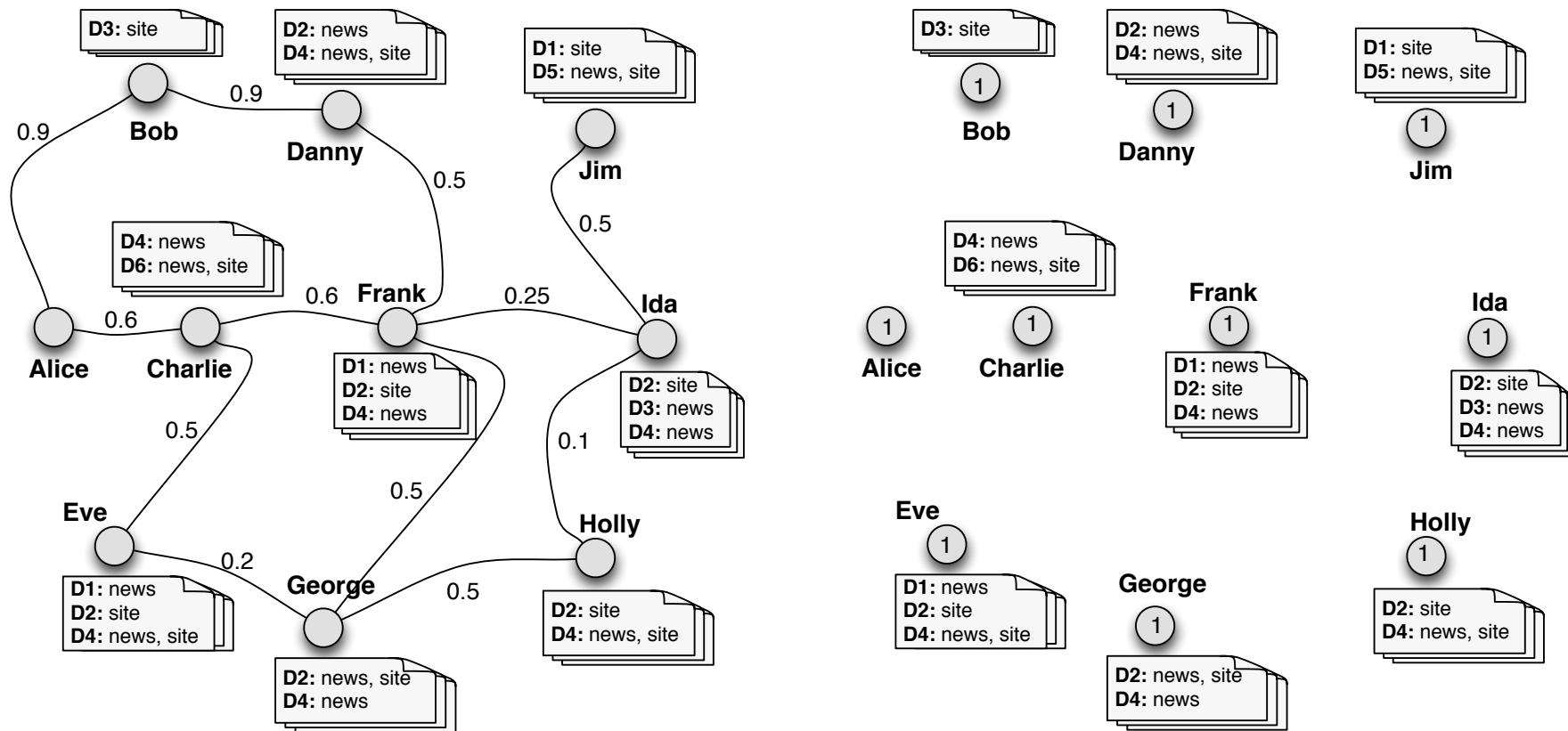
Proximity also w.r.t users indirectly connected to the seeker.

- multiplication over a path
- max over all connecting paths

{B:0.9, D:0.8, C:0.6, F:0.4, E:0.3, G:0.2, H:0.1, I:0.1, J:0.05}



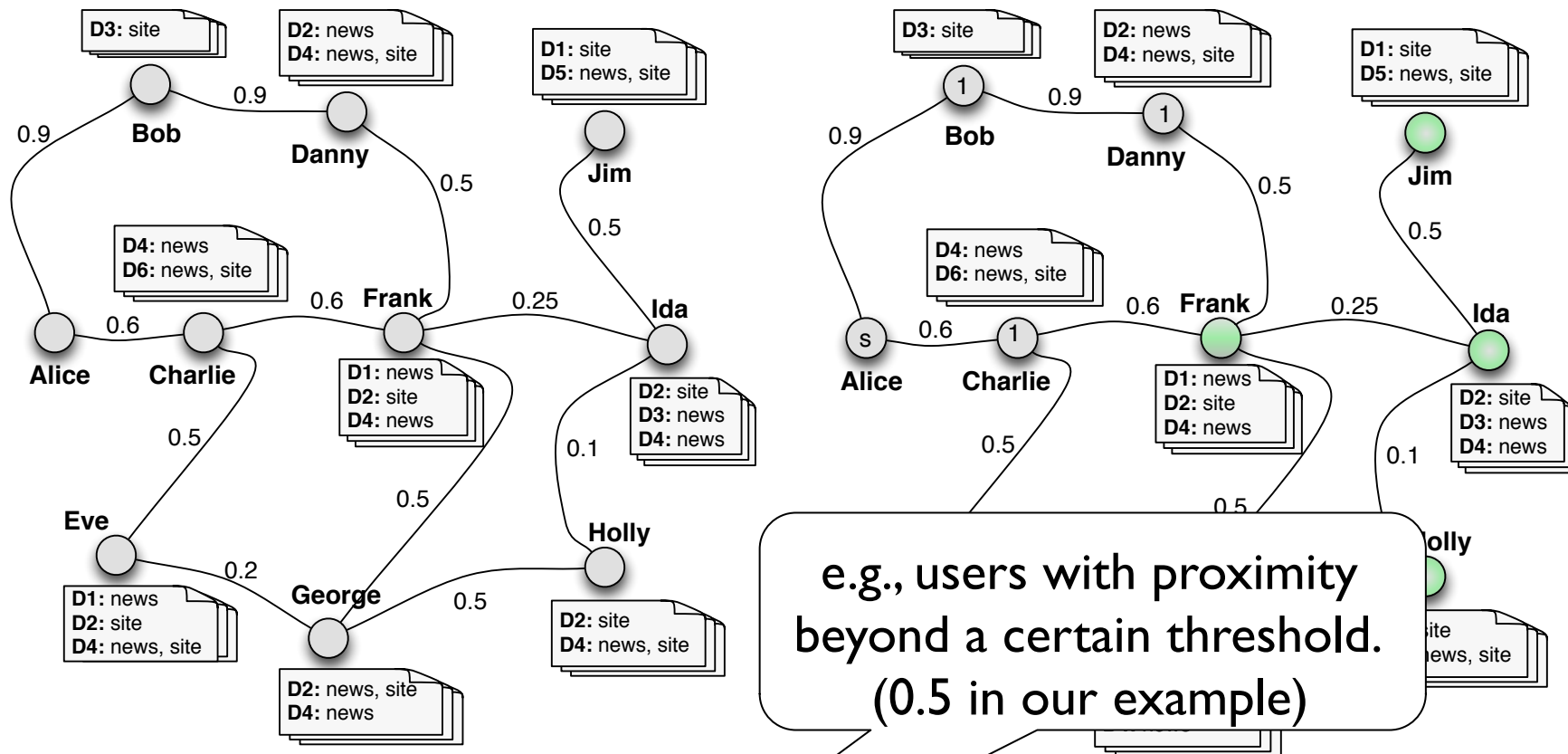
# Main related work: Threshold Algorithm (TA)



Web search setting: the Threshold Algorithm (TA) and NRA [Fagin01]

- pre-computed per-term inverted lists
- in a social context
  - either all users would have the same relevance (valued 1),
  - or we keep a list for each term and seeker

# Main related work: [Amer-Yahia et al., VLDB08]



## Global Upper-Bound algorithm

- integrates social proximity in top-k search in folksonomies, but with some restrictions: only the seeker's "direct" neighbors are considered relevant.

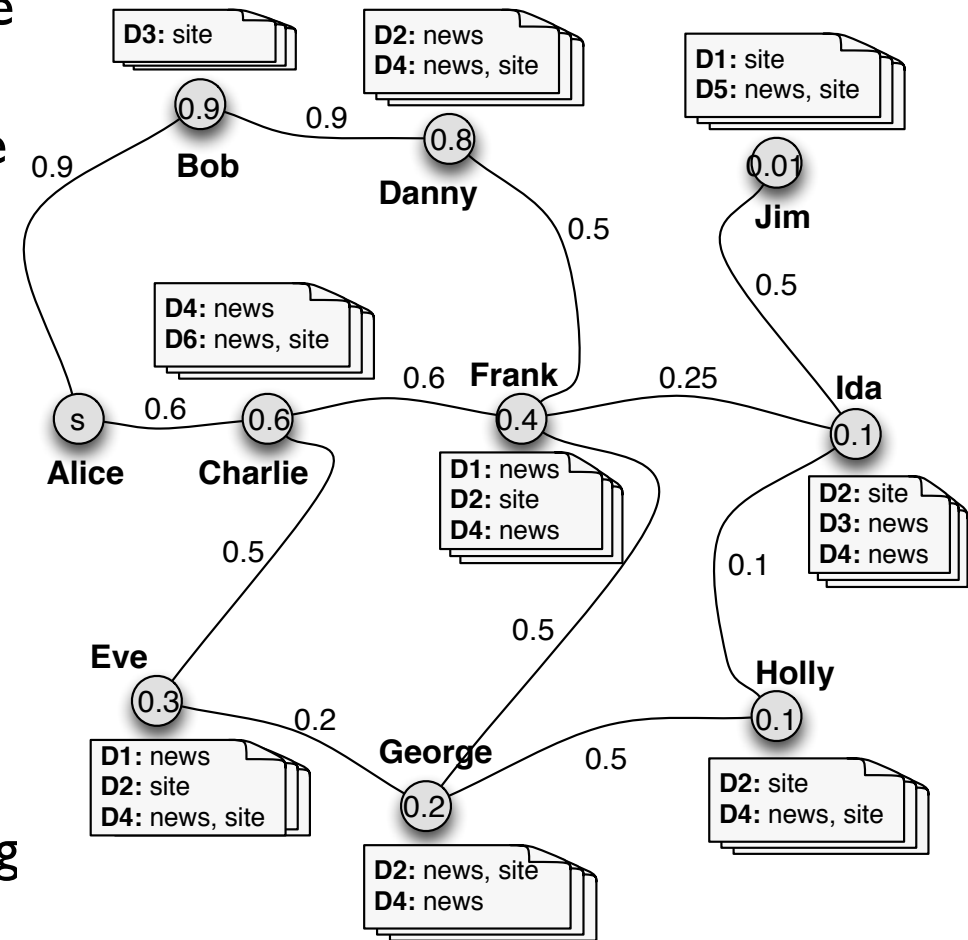
# Main related work: [Shenkel et al., SIGIR08]

**ContextMerge** algorithm: considers the general setting -> all users (even indirectly connected to the seeker) are relevant:

- all possible pairs of users have a **pre-computed** social score value

## Major drawbacks:

- high disk space cost: order of 700TB for Del.icio.us, much bigger for Facebook
- **limited applicability:**
  - social scores can evolve (e.g., tag similarity),
  - lists need to be kept up to date,
  - users should be able to chose key model parameters.



# Main contributions

We show that the visit of the network in decreasing order of proximity (w.r.t the seeker) can be done **on the fly** and as needed

- for a wide family of proximity functions (monotone ones)

Key advantages:

- changes and online model choices become a non-issue
- a typical network can fit in main-memory
- spare the potentially huge disk volumes required previously

A novel algorithm (TOPKS) that is:

- **instance optimal** for the exclusively social case,
- more efficient in general

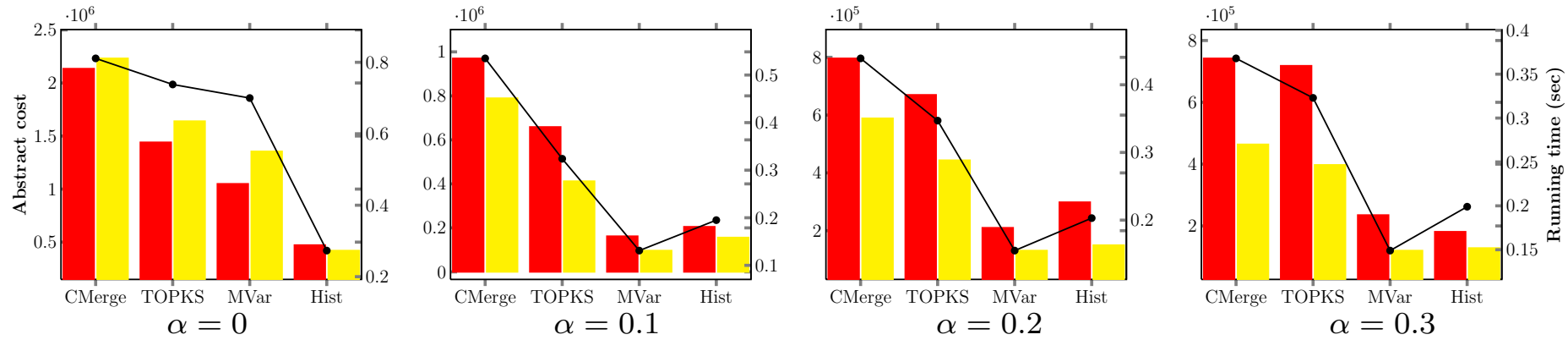
Two efficient **approximation algorithms**, based on high-level description of the social network (statistics on proximity values)

# Glimpse on experiments

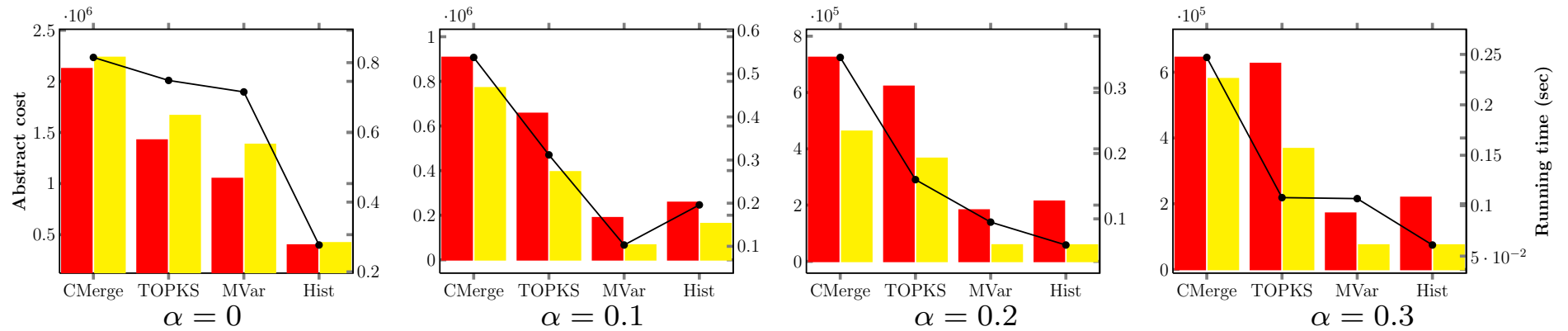
- Delicious dataset consisting of 80,000 users, 595,811 items, 198,080 tags
- 3 similarity networks (via Dice coefficient):
  - tag similarity network: 40,319 nodes, 8,335,544 links
  - item similarity network: 49,038 nodes, 3,329,540 links
  - item-tag similarity network: 40,353 nodes, 1,849,898 links.
- top-10 and top-20 answers, two ranking functions: tf-idf and BM15
- low alpha values: 0, 0.1, 0.2, 0.3
- two cost measures:
  - abstract cost: in terms of visited users and sequential access to data.
  - running time



# Tag-similarity network



BM15



tf-idf

# Conclusions

# The Web is now social

The social Web is growing at an unprecedented rate

- which major Web applications are not social ?
- Facebook's user base, Twitter's throughput, Wikipedia's richness

Graph models for social networks (directed or undirected, n-partite, hypergraph, weighted or unweighted)

Research playground for various communities: DB, IR, knowledge management, machine learning and data mining, human-computer interactions, game theory, networking, sociology, economics, ...

**Foundations of social data management**

# Perspectives

- Search: improve Web search, new kinds of search (real-time, social) for communities or data
- Mine behavior, opinions and personal information of hundreds of millions of individuals
- Go beyond explicit links: understand opinion, sentiment, behavior, influence
- Rank information and users by trustworthiness, importance
- Design social applications governed by the individuals
- Privacy and security issues

Thank you.