Intent in Social Tagging Sytems

Markus Strohmaier

Univ. Ass. / Assistant Professor
Knowledge Management Institute
Graz University of Technology, Austria

e-mail: markus.strohmaier@tugraz.at
web: http://www.kmi.tugraz.at/staff/markus

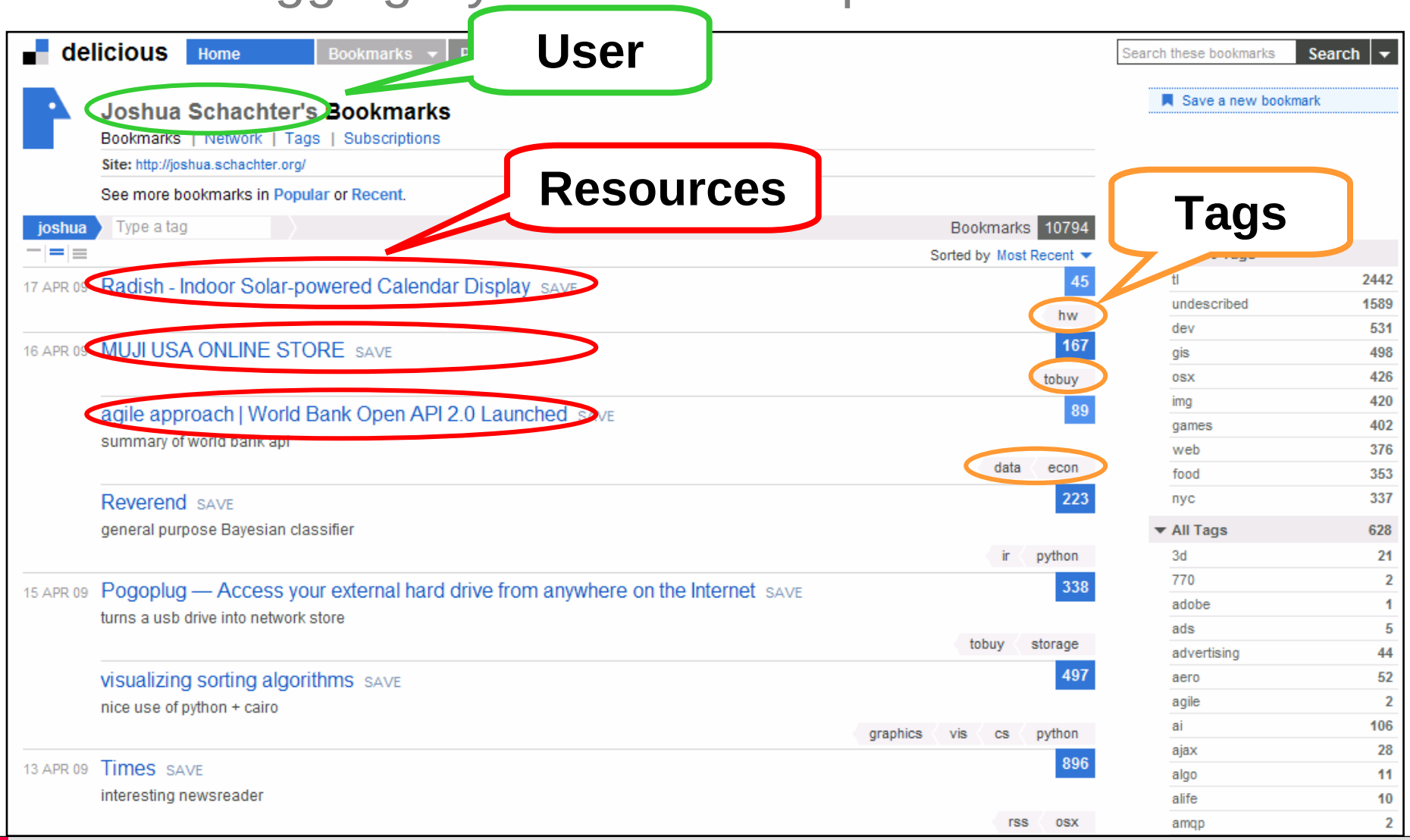
Vision

Opportunity: Use user generated data on the web to construct the world's most comprehensive common-sense knowledge base.

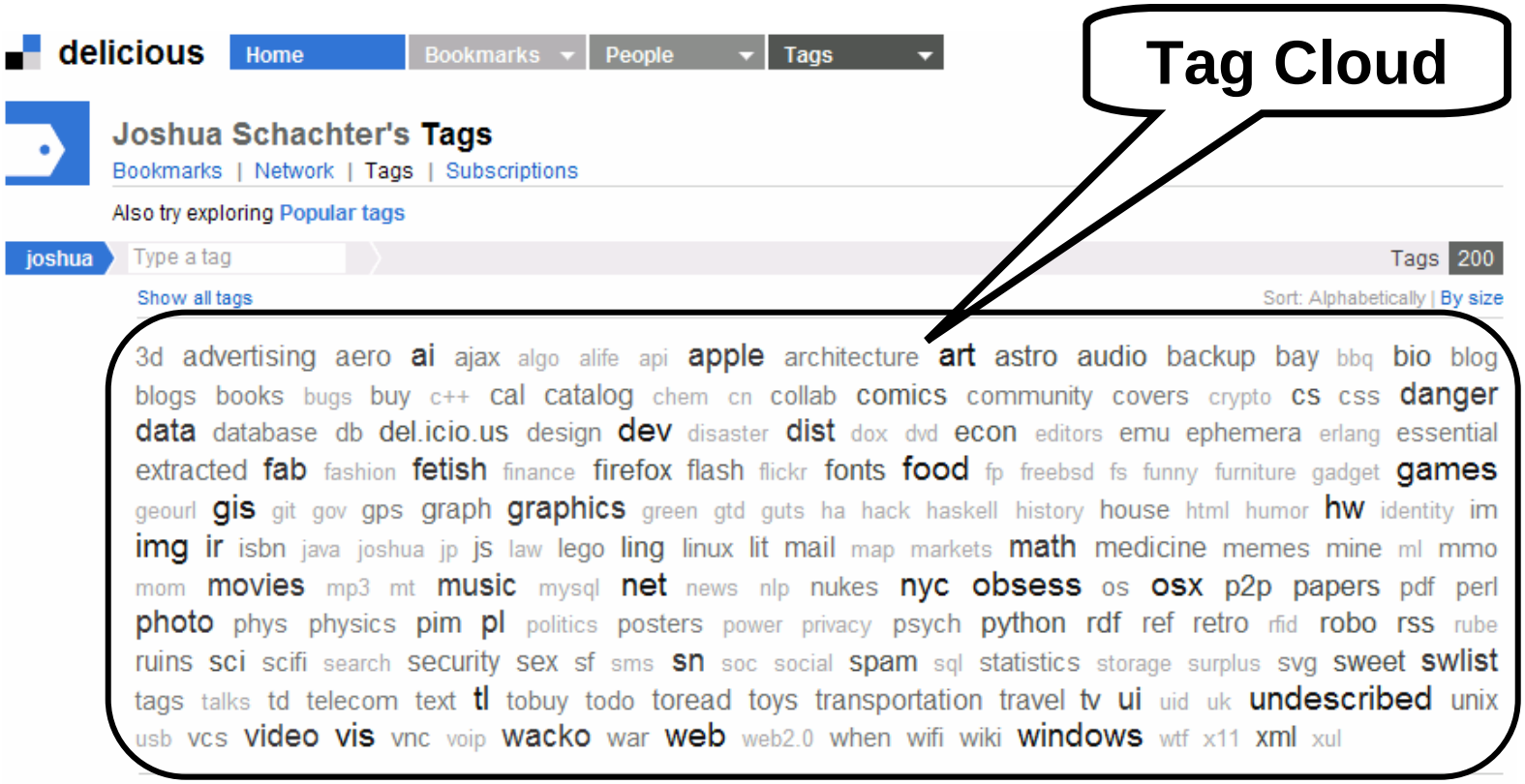
History:

- CYC (1984 -)
- Volunteer-based Knowledge Acquisition (2000 -)
  Openmind ConceptNet
- Knowledge Acquisition from the Web (2002 -)
- Human Computation (2004 -)
  Games with a Purpose

Social Tagging Systems - Example from Delicious

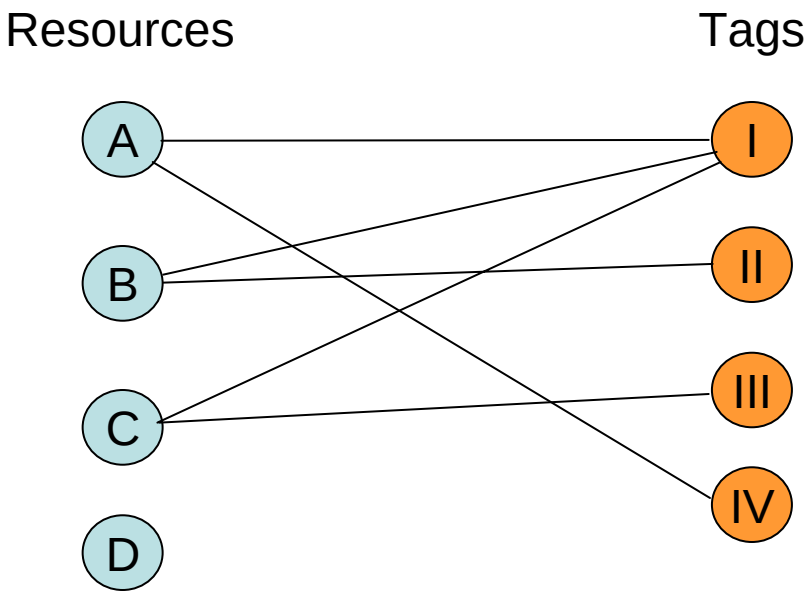


Social Tagging Systems - Example from Delicious

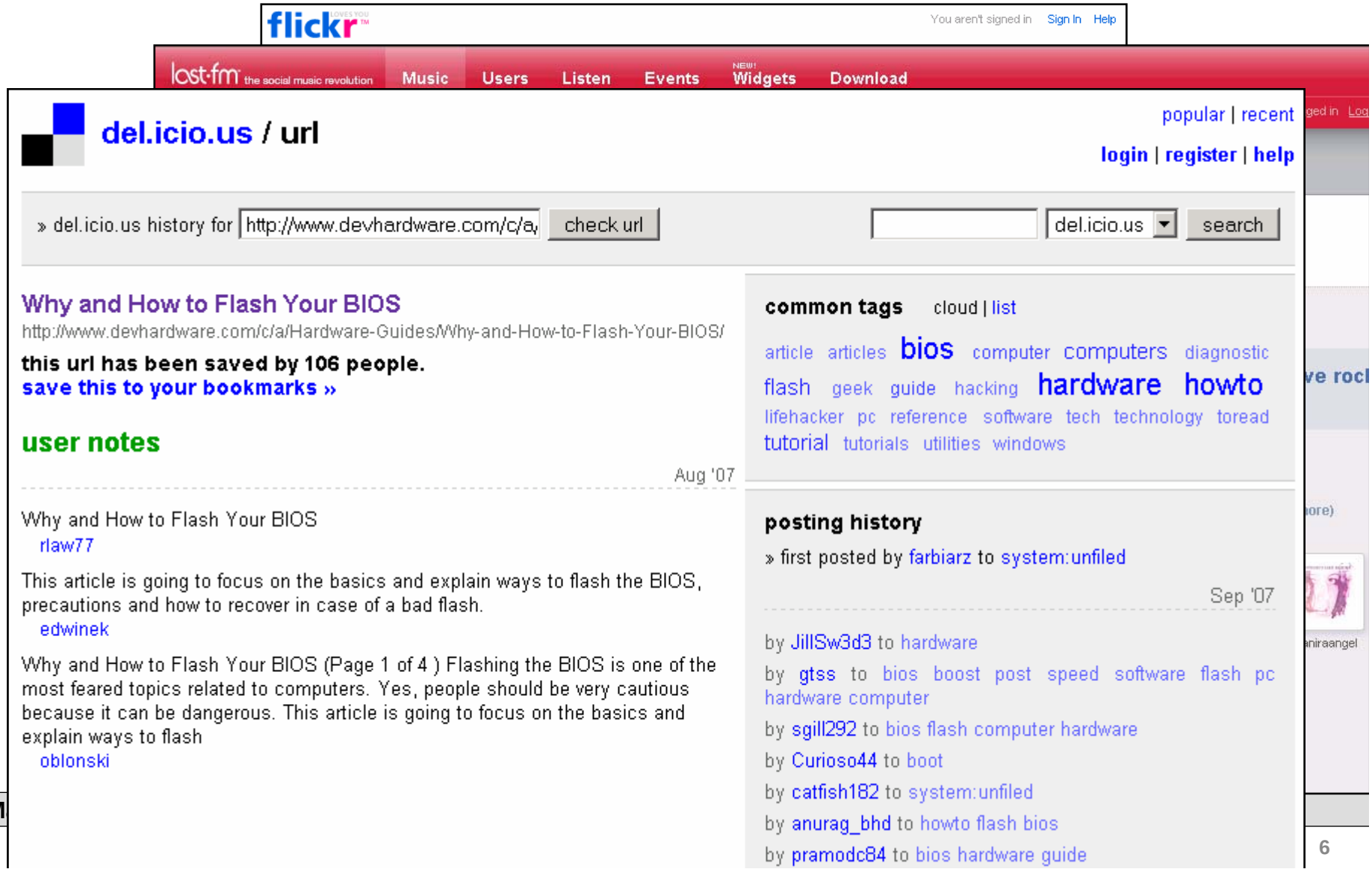


Two Mode Networks

- Two types of nodes
  e.g. Users and Tags, Tags and Resources

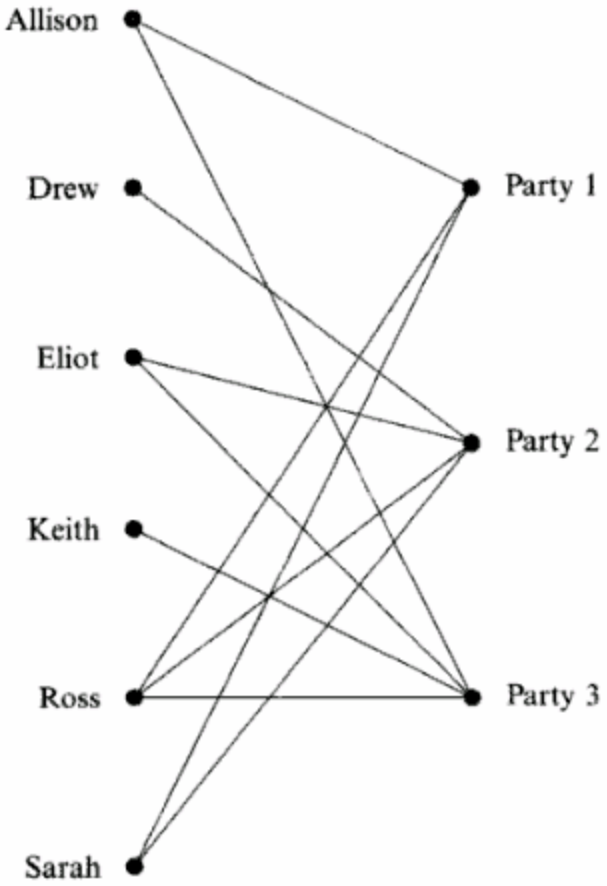


Reminder: Social Networks Examples



Representing Two-Mode Networks As Two Mode Sociomatrices

[Wasserman Faust 1994]



General form:

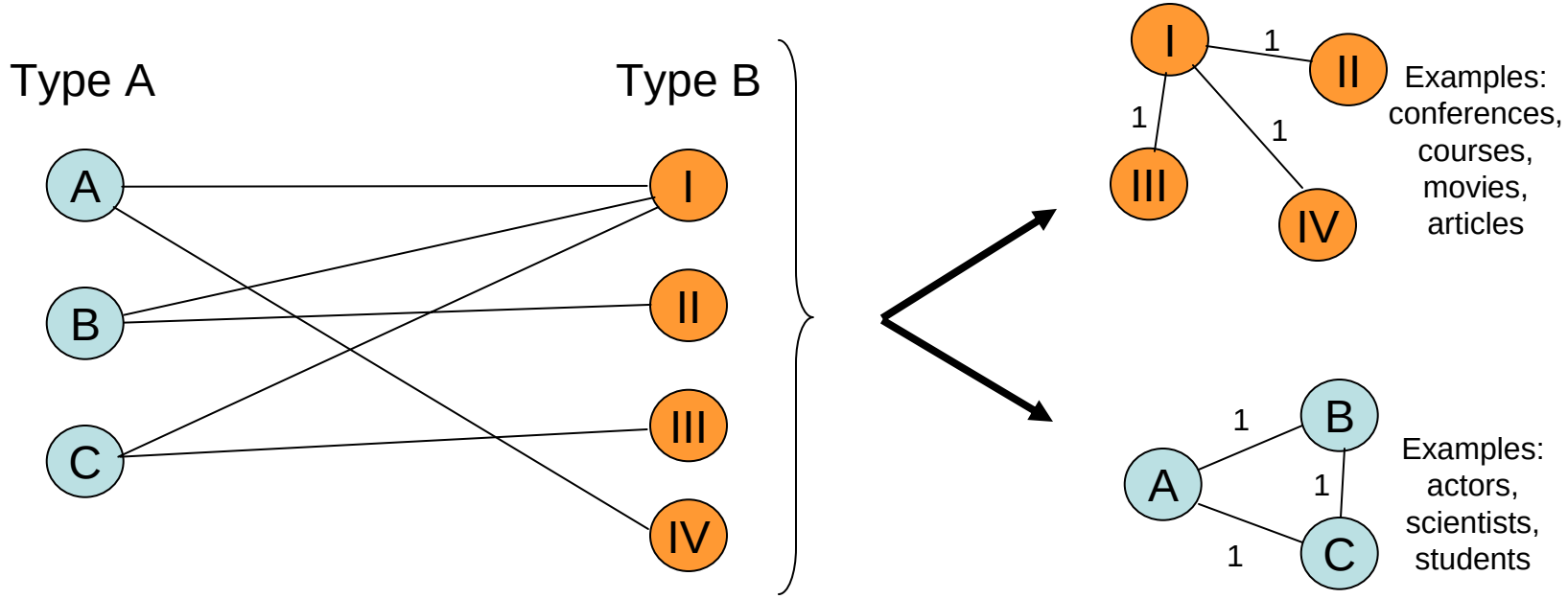
$$\begin{pmatrix} 0 & A \\ A' & 0 \end{pmatrix}$$

	Allison	Drew	Eliot	Keith	Ross	Sarah	Party 1	Party 2	Party 3
Allison	-	0	0	0	0	0	1	0	1
Drew	0	-	0	0	0	0	0	1	0
Eliot	0	0	-	0	0	0	0	1	1
Keith	0	0	0	-	0	0	0	0	1
Ross	0	0	0	0	-	0	1	1	1
Sarah	0	0	0	0	0	-	1	1	0
Party 1	1	0	0	0	1	1	-	0	0
Party 2	0	1	1	0	1	1	0	-	0
Party 3	1	0	1	1	1	0	0	0	-

Fig. 8.3. Sociomatrix for the bipartite graph of six children and three parties

Two Mode Networks and One Mode Networks

- **Folding** is the process of transforming two mode networks into one mode networks
  - Also referred to as: **T**, ⊥ projections [Latapy et al 2006]
- Each two mode network can be folded into 2 one mode networks



Two mode network **2 One mode networks**

Transforming Two Mode Networks into One Mode Networks

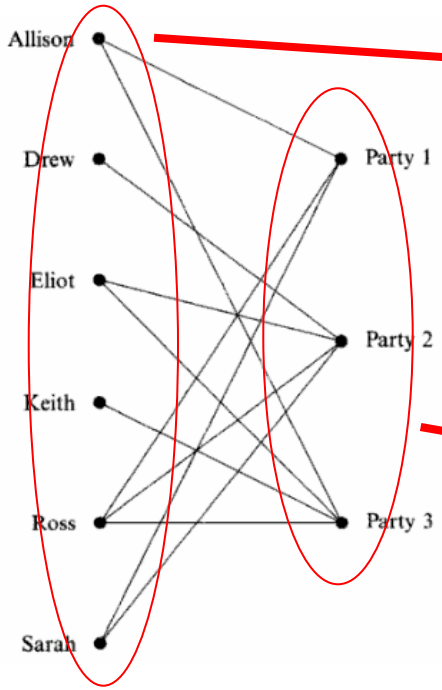
[Wasserman Faust 1994]

$M_P = M_{PC} * M_{PC}'$

C…Children

P…Party

- Two one mode (or co-affiliation) networks (folded from the children/party affiliation network)



	n_1	n_2	n_3	n_4	n_5	n_6
n_1	2	0	1	1	2	1
n_2	0	1	1	0	1	1
n_3	1	1	2	1	2	1
n_4	1	0	1	1	1	0
n_5	2	1	2	1	3	2
n_6	1	1	1	0	2	2

Fig. 8.5. Actor co-membership matrix for the six children

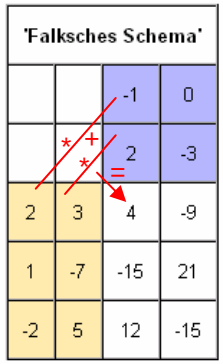
	m_1	m_2	m_3
m_1	3	2	2
m_2	2	4	2
m_3	2	2	4

Fig. 8.6. Event overlap matrix for the three parties

[Images taken from Wasserman Faust 1994]

Transforming Two Mode Networks into One Mode Networks

[Wasserman Faust 1994]



$$M_P = M_{PC} * M_{PC}'$$

C…Children

P…Party

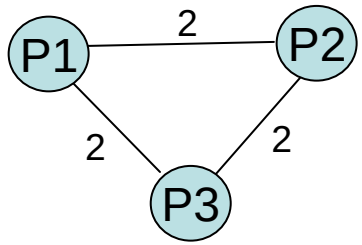
	Allison	Drew	Eliot	Keith	Ross	Sarah
Party 1	1	0	0	0	1	1
Party 2	0	1	1	0	1	1
Party 3	1	0	1	1	1	0

*

	Party 1	Party 2	Party 3
Allison	1	0	1
Drew	0	1	0
Eliot	0	1	1
Keith	0	0	1
Ross	1	1	1
Sarah	1	1	0

=

	Party 1	Party 2	Party 3
Party 1	3	2	2
Party 2	2	4	2
Party 3	2	2	4

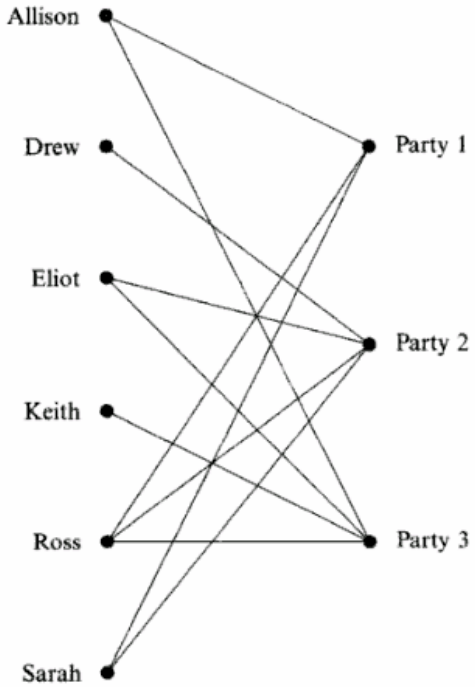


Output: Weighted regular graph

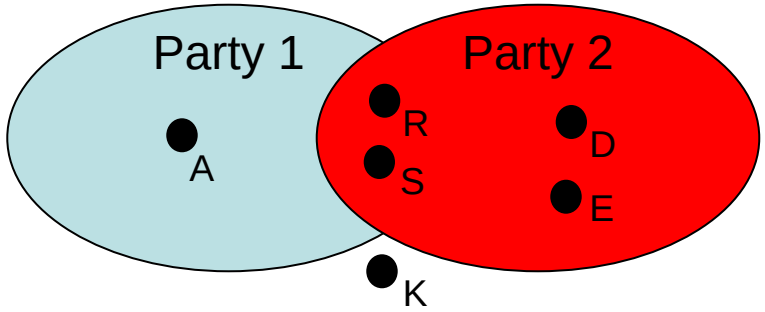
Transforming Two Mode Networks into One Mode Networks

[Wasserman Faust 1994]

Bi-partite representation (entire bipartite graph)



Set theoretic interpretation (P1, P2)

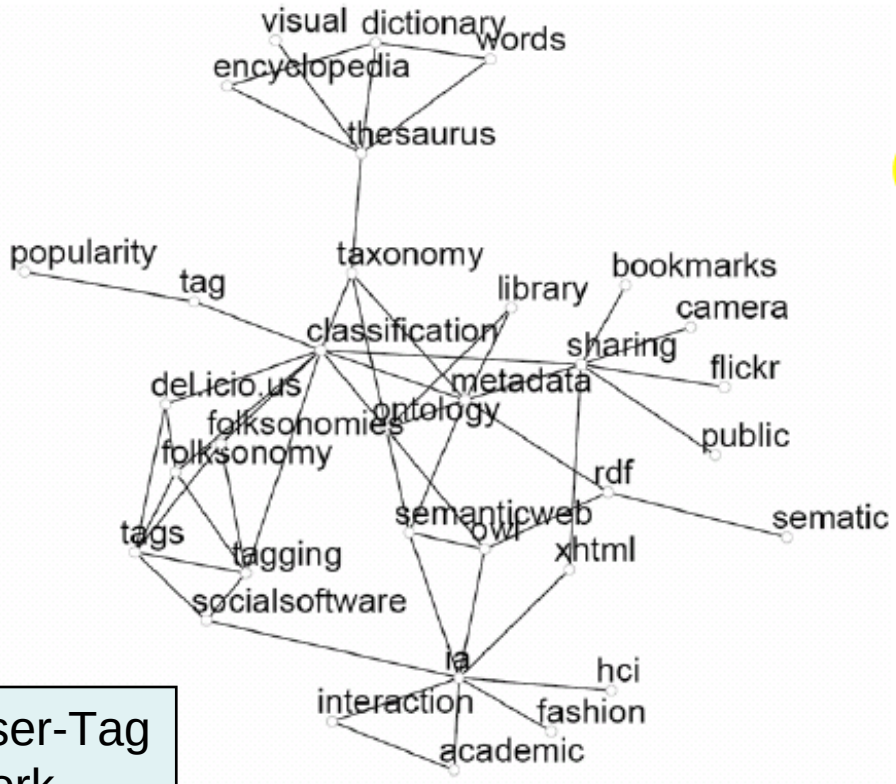


Vector interpretation (P1, P2)

	Party 1	Party 2
Allison	1	0
Drew	0	1
Eliot	0	1
Keith	0	0
Ross	1	1
Sarah	1	1

Broader / narrower term relations

P. Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics. International Semantic Web Conference, 522-536, Springer,2005



Folded User-Tag network

Fig. 2. Detail view of the del.icio.us tags associated through users: a 3-neighborhood of the term ontology. Note that the term *sematic* is correctly associated, despite the obvious typo.

Table 3. Broader/narrower term relations in the technology domain, based on sub-communities in del.icio.us

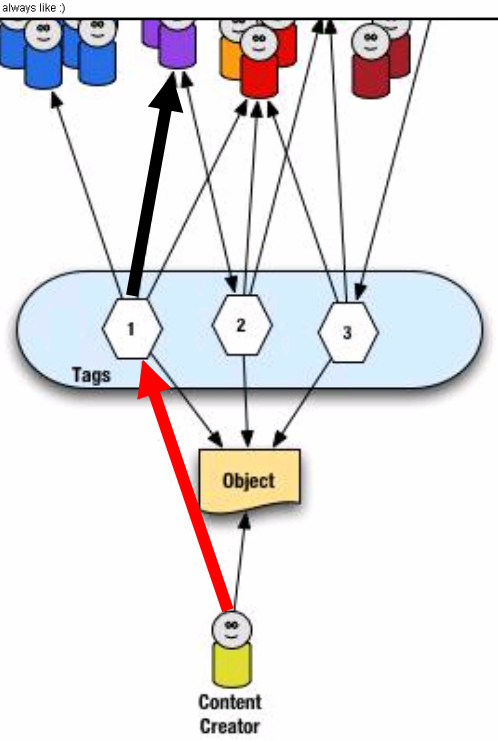
Broader	Narrower
rss	atom
cmyk	rgb
cell	umts, wcdma, ev-do
phone	cell
ajax	json
xml	xslt
rdf	owl
flickr	gmail, picasa
ruby	rails
mac	iphoto
java	j2ee
google	gds
search	a9, engine
linux	ubuntu, gnome
flash	actionscript
flickr	lickr, photoset
javascript	xmlhttprequest, dom, sarissa

Types of Folksonom

[Thomas Vander Wal http://www.personalinfocloud.com/20

Narrow folksonomies

- tagging objects that are **not easily searchable** or have no other means of using text to describe or find the object
- done by **one or a few people** providing tags that the person uses to get back to that information.
- The tags, unlike in the broad folksonomy, are **singular in nature**
- tags are **directly associated with the object**.
- Example: Flickr

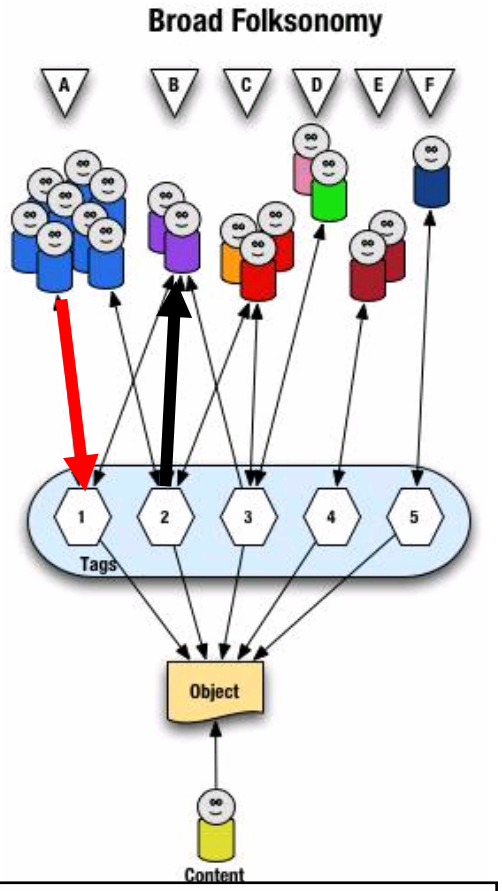


Types of Folksonomies

[Thomas Vander Wal http://www.personalinfocloud.com/2005/02/explaining_and_.html]

Broad folksonomies

- many people **tagging the same object** and
- every person can **tag the object with their own tags** in their own vocabulary
- Example: Social bookmarking
- The broad folksonomy provides a means to see trends in how a broad range of people are tagging one object.
- power law curves and long-tail are relevant phenomena

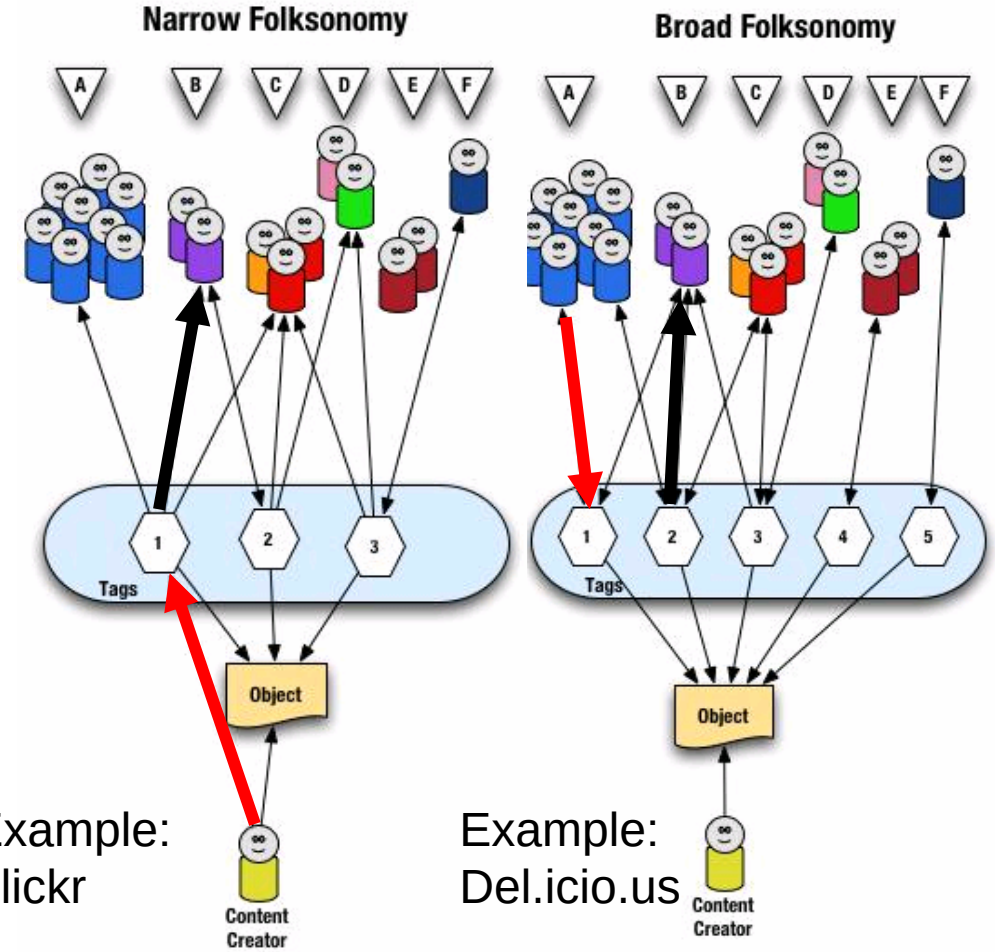


Types of Folksonomies

[Thomas Vander Wal http://www.personalinfocloud.com/2005/02/explaining_and_.html]

Differences

- Number of people tagging a single object
- Narrow folksonomies are more sparse
- Purpose
- Narrow ones allow for enhanced metadata for an object



Example: Flickr

Example: Del.icio.us

Tagging

- Metadata at large, finally!
  - User generated data at large scale
- Not standardized, because no meta-meta information
  - Does „BernersLee“ refer to *DC creator* or *DC subject* [Dublin Core]?
- useful, because intrinsically motivated
  - Useful to somebody: users tag for a reason

Q: What are the motivations and intentions of users when **tagging resources**?

Agenda

Structure of this presentation:

1. Relating Content (of Resources) and Intent (of Users) via Tagging

2. Detecting User Motivation in Tagging Systems

A Simple Model of Folksonomies

But:

Variability in the set of Users U

- at least four user roles including 1) resource author, 2) resource collector 3) indexer or tagger and 4) searcher [Voss 2007].

Variability in the set of Tags T

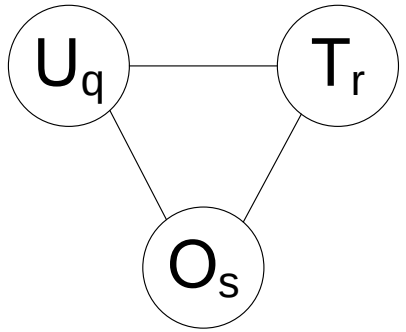
- For example, types of tags include: 1) Identifying what a resource is about 2) Identifying what it is 3) Identifying who owns it 4) Refining categories 5) Identifying qualities or characteristics 6) Self reference 7) Task organizing [Golder und Hubermann 2005]

Variability in the set of Objects O

- Different „*Objects of sociality*": movies (youtube), URLs (delicious), photos (flickr), music (last.fm), etc..

$$(F \subseteq U \times T \times O)$$

$$F \subseteq U_q \times T_r \times O_s$$



Extended Model of Folksonomies

q...types of users
r...types of tags
s...types of objects

Motivating Example: Content vs. Intent

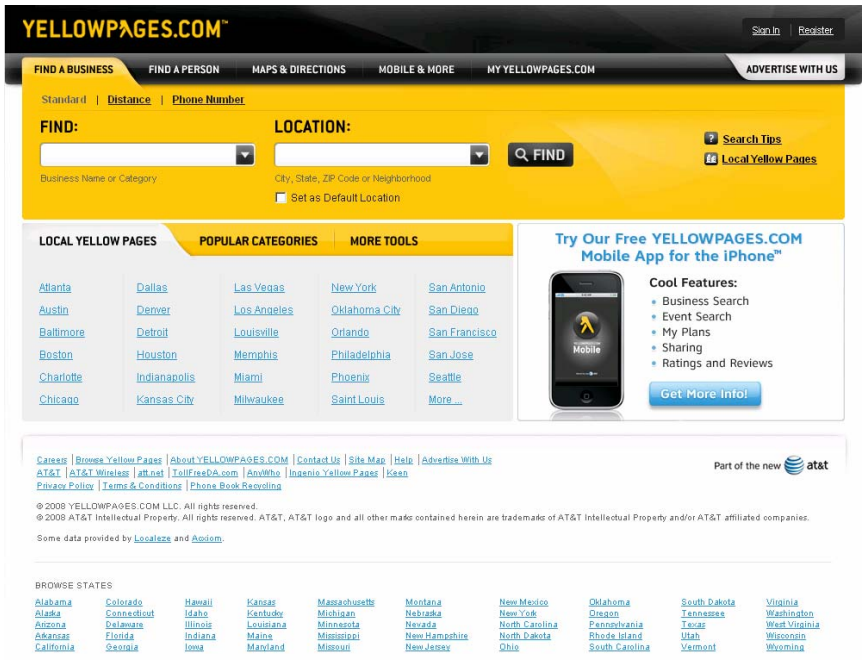
Intent

(What goals it aims at / helps to achieve)

- find a physician
- organize a high-school reunion
- contact an old friend
- organize a marketing campaign
- find others who share the same family name
- find my way to an address
- …

Content

(What it is)



Websites, Blogs, Images, Web Services, …

What factors influence the type of tags being used?

Terminological and contextual mismatch: While **search queries** tend to express user *intent*, **tags** tend to express aspects of *content*

(94% According to one of today's talks)

CIKM'08 Papers …

on Search *In*tent

- **Understanding the Relationship between Searchers' Queries and Information Goals**, D. Downey, D. Liebling, S. Dumais
- **Matching Task Profiles and User Needs in Personalized Web Search**, J. Luxenburger, S. Elbassuoni, G. Weikum
- **Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs**, R. Jones, K. Klinkner
- **Keynote** B. Croft „Long Queries / Intent statements"

on Tagging *Con*tent

- **Can All Tags Be Used for Search?**, K. Bischoff, C. Firan, W. Nejdl, R. Paiu
- **Social Tags: Meanings and Suggestions**, F. Suchanek, M. Vojnovic, D. Gunawardena
- **Tag-Based Filtering for Personalized Bookmark Recommendations**, P. K. Vatturi, W. Geyer, C. Dugan, M. Muller, B. Brownholtz [Poster]
- + related work in WWW, Hypertext, etc (see paper)

Observation: terms used to craft search queries are usually **different** from the terms that are used to tag resources in social media [Heyman 2008]

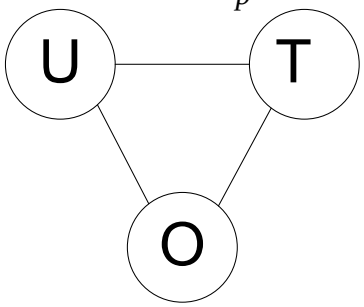
Why ?

Exploratory Research Questions

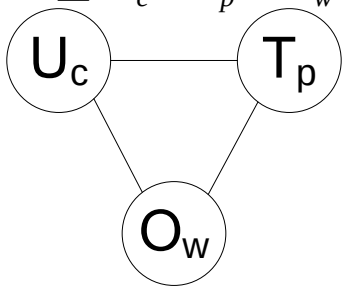
1. **Feasibility**: Would users assign meaningful purpose tags?
2. **Accuracy***:* Do purpose tags accurately reflect plausible purposes of resources?
3. **Utility**: Can purpose tagging improve search in social software?
4. **Coverage**: Can purpose tags expand the vocabulary of existing tags?
5. **Meaning***:* Are purpose tag graphs meaningful?

An Intentional Social Bookmarking Prototype

$$F \subseteq U \times T_p \times O$$

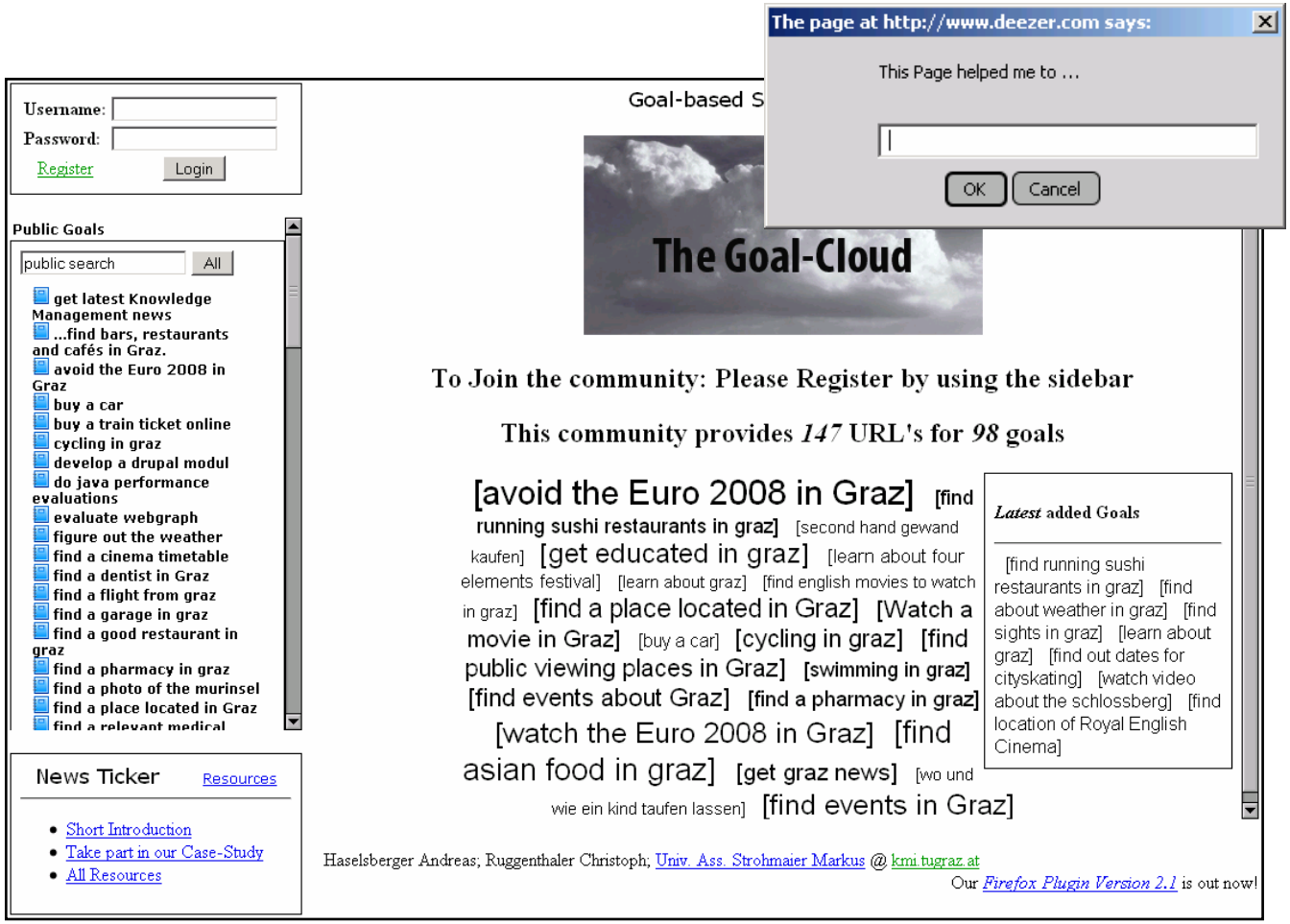


$$F \subseteq U_c \times T_p \times O_w$$



Intentional Social Bookmarking

c...consumer
p...purpose
w...websites



with students Andreas Haselsberger and Christoph Ruggenthaler

Purpose Tagging

Data Collection

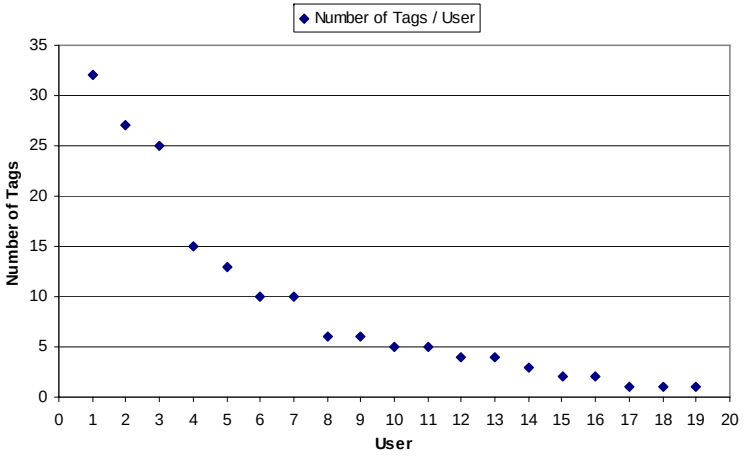
- **Duration**: 2 weeks
- **Population**: Computer graduate students and employees of a research organization
- **Task**: Bookmark resources related to „Graz“

	URIs	Tags	Users
Dataset A (Purpose Tagging)	123	75	19
Dataset B (del.icio.us)	36	849	2801

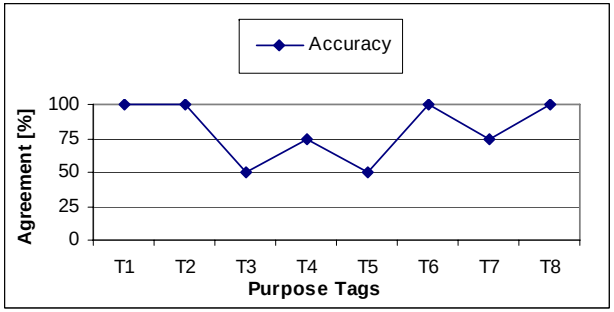
Table 1 Characteristics of Dataset A and B

Purpose Tagging

1. Would Users Assign Purpose Tags?



2. Do Purpose Tags Accurately Reflect Plausible Purposes of Resources?



Purpose Tagging

3. Can Purpose Tagging Improve Search in Social Software?

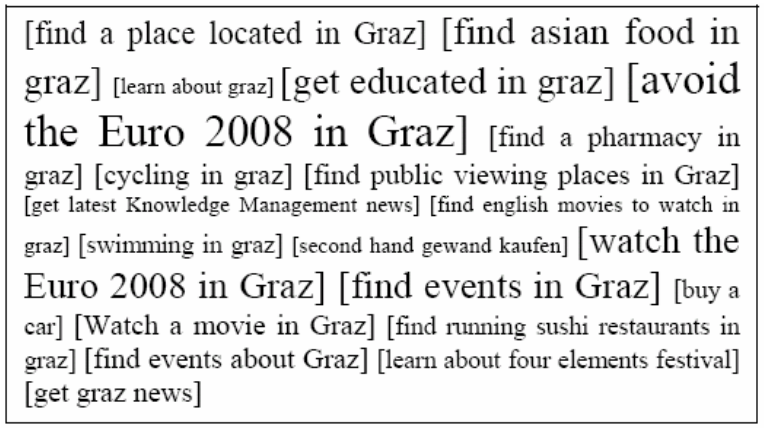
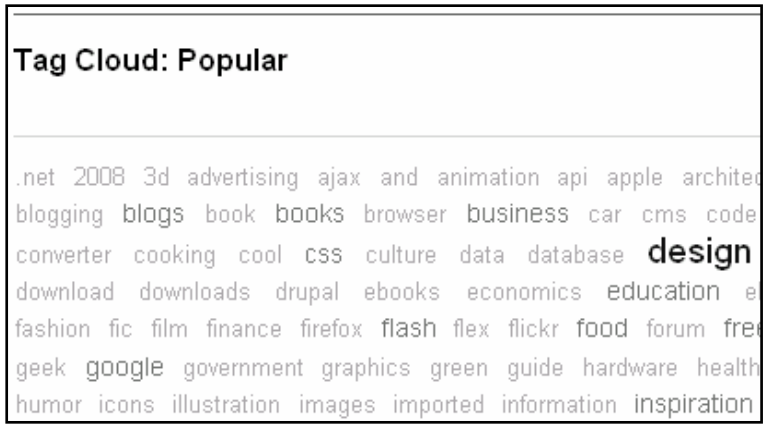


Figure 5 Purpose Tag Cloud Generated from Dataset A



delicious

Purpose Tagging

3. Can Purpose Tagging Improve Search in Social Software?

Four users / four search tasks each:
- "find an overview of restaurants in Graz"
- "get a weather forecast for Graz"
- "find information about local events in Graz"
- "find information about movie showtimes in Graz"

Observations (Audio/Screen casts):
- Purpose tags used to narrow search / disambiguate
- Users „felt guided"
- Purpose tags „felt natural" to accomplish search goals
- easier to assess relevance
- One user felt a particular purpose tag was misleading
- Overspecified queries in delicious search

Figure 7 Comparison of Search Results for Query "Graz", left: Intentional Social Bookmarking, right: del.icio.us

Purpose Tagging

4. Can Purpose Tags Expand the Vocabulary of Existing Tags?

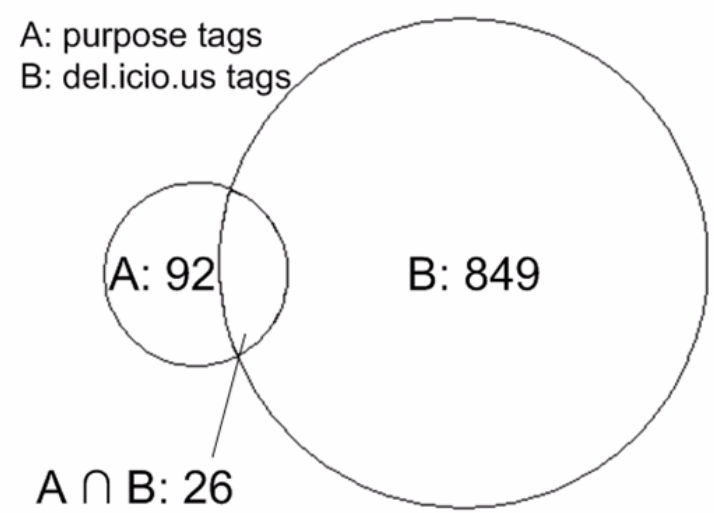


Figure 9 Vocabulary Overlap between Purpose Tags and Delicious Tags

new terms	overlap terms
$A \setminus B$	$A \cap B$
pharmacy, to, via, location, buy, for, street, timetable, euro, where, asian, of, medical, more, results, facts, ticket, located, on, place, educated, transport., get, how, watch, offers, learn, practitioner, map., climb, specific, about, insight, use, from, running, swimming, figure, dates, detailed, tourists., relevant, restaurants, find, graz., airport., cityskating, b, news, a, at, into, the, learnpartner, in, know, dentist, flight, new, schedule, overview, what, arrival, public, first, out	car, 2008, knowledge, technology, cycling, events, visit, online, em2008, friends, transport, sushi, weather, train, conference, attention, metadata, information, university, movie, graz, tu, cinema, program, food, management

Table 2 $A \setminus B$ and $A \cap B$

~72% of the vocabulary of purpose tags was novel
(created by 19 purpose tagging users vs. 2801 users of delicious)

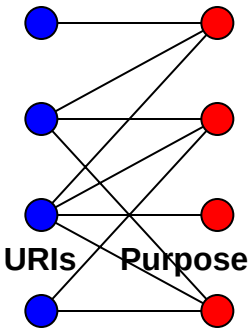
Example: „find a girlfriend“ for german version of facebook.com

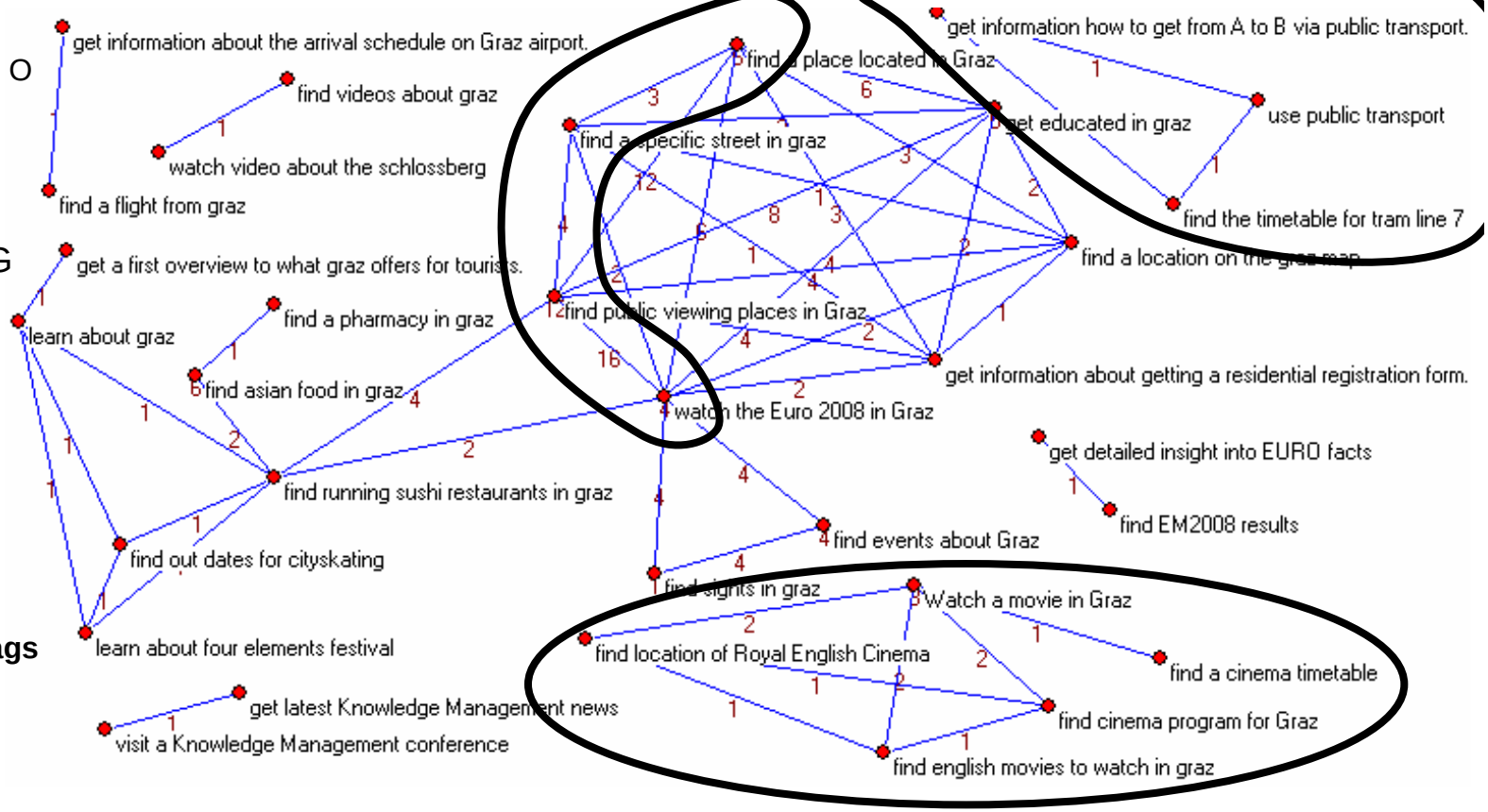
Purpose Tagging

5. Are Purpose Tag Graphs Meaningful?

Transforming the tripartite graph U, T, O into bipartite graphs UO, OT and UT.

Given G(OT)
Calculate $T^*=G^TG$

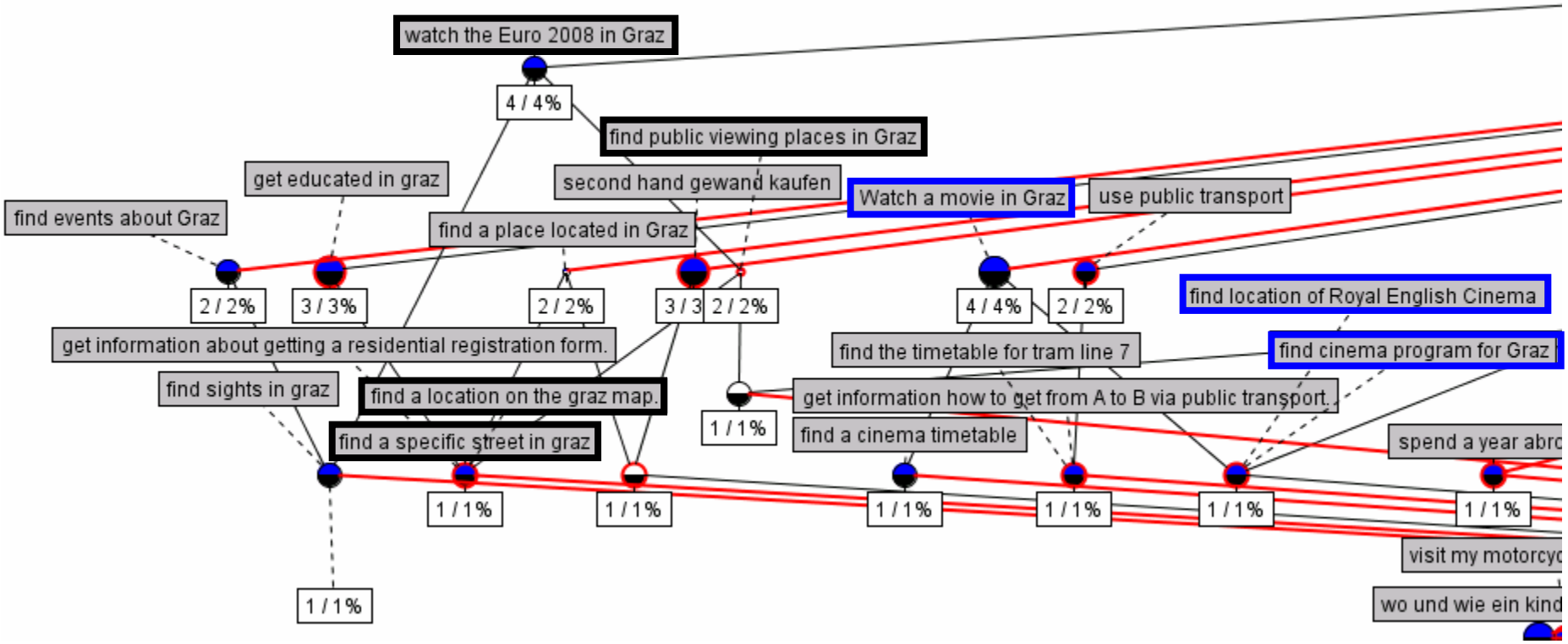




Purpose Tagging

5. Are Purpose Tag Graphs Meaningful?

Partially Ordered Sets over a Bi-Partite Graph

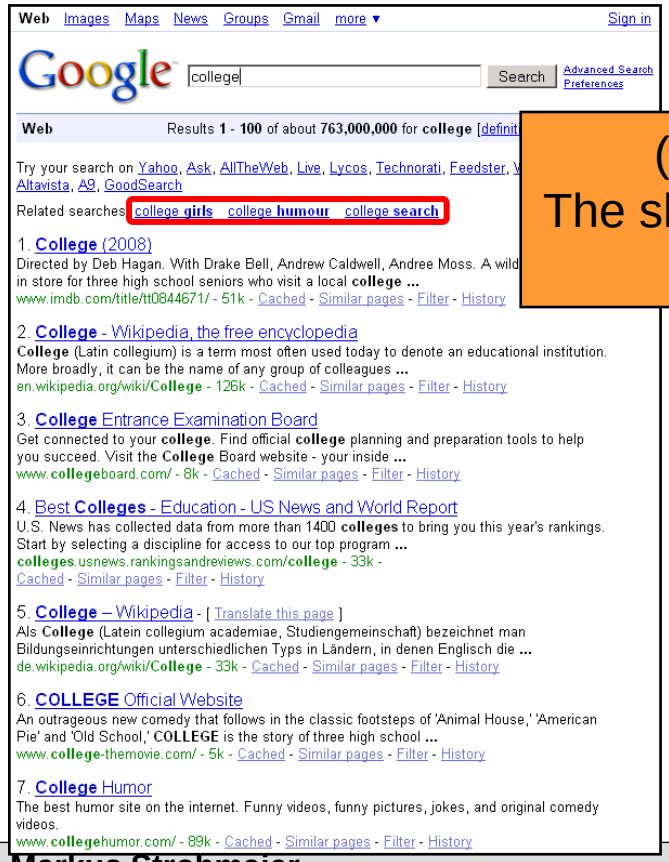


Based on Formal Concept Analysis [Wille 2005]
visualized with ConExp

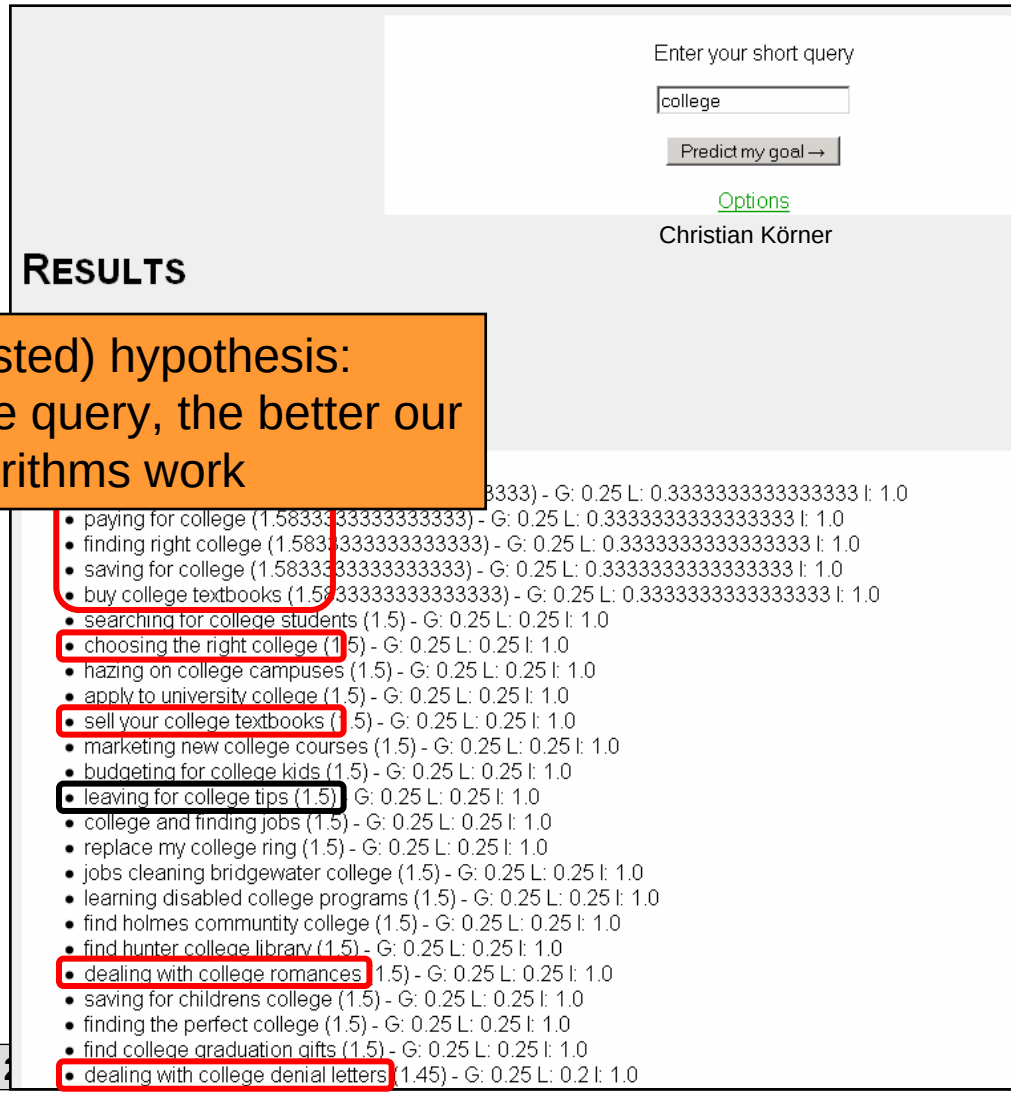
Applications

Relating Content Tags and Intent Tags

***AOL Search Query Log* based Intent Prediction**



(yet untested) hypothesis:
The shorter the query, the better our algorithms work



Conclusions

- More types of tags than currently studied
- Task-aware relevance
- Search result justification, Search intent estimation

Outlook:

- Large scale controlled experiment (~ 4.000 active users)
  - Modifying the tagging process in a social bookmarking system for scientists (BibSonomy)
  - User acceptance
  - Comparison of traditional tags vs. purpose tags
- Delicious study
  - Existence and nature of purpose tags in an existing bookmarking system

Detecting User Motivation of Tagging

Why do tagging systems work?

This was topic of a panel at CHI 2006, following conclusions were drawn:

Tagging has a benefit for the user

- Similar to bookmarking, integrated apps
- Benefit of accessibility from everywhere in the internet

Tagging allows social interaction

- Connecting a user to a community trough tags
- People can subscribe your stream

Benefits of Tagging

Tags are useful for retrieval

- Synonyms and typos vanish in the mass of tags
- Communities can retrieve “their” stuff (e.g. by special tag)

Tagging Systems have a low participation barrier

- Apps are easy to use, intuitive, responsive
- Free text is used to do the tagging
- Requires no previous considerations & training

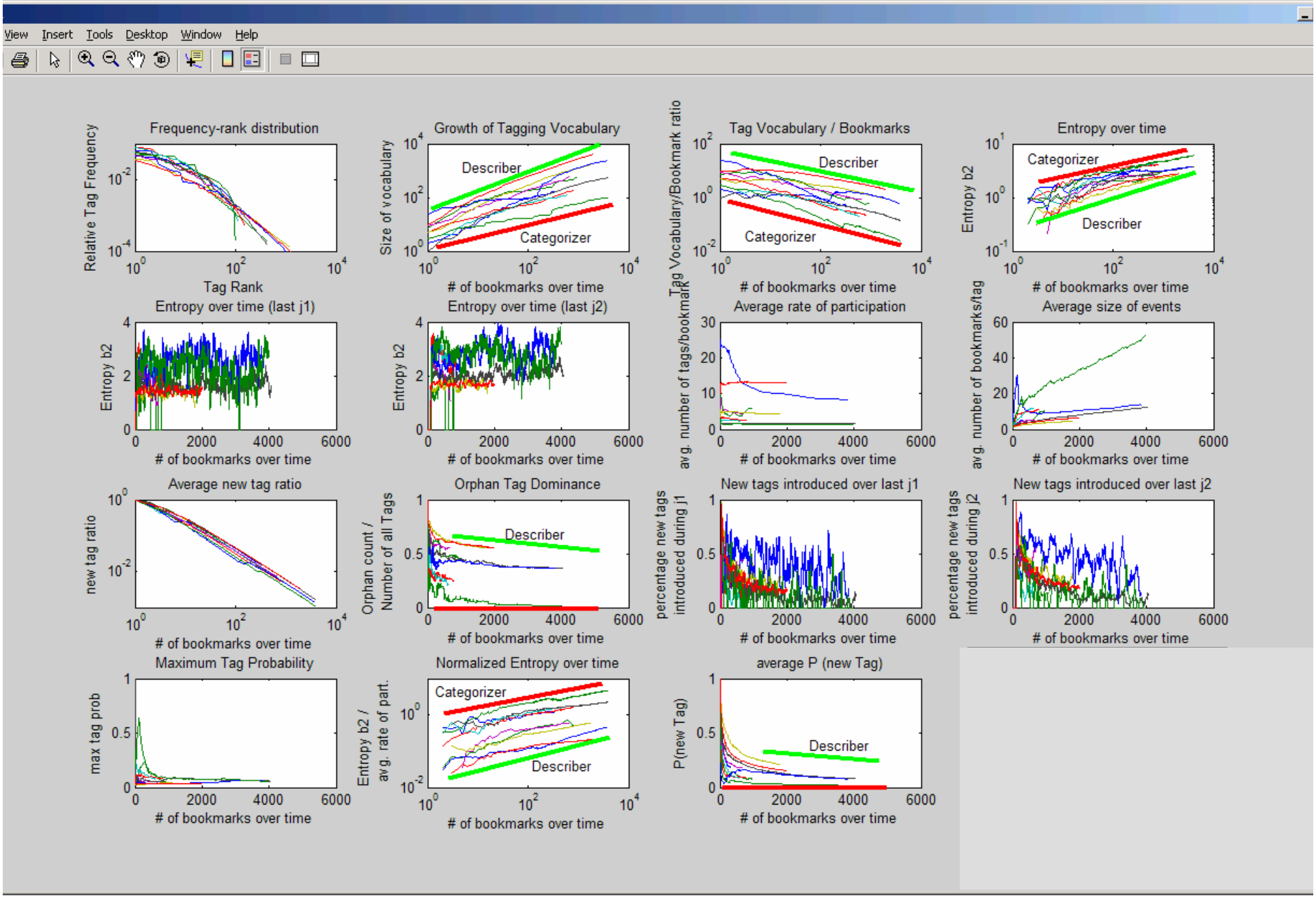
Categorization vs. Description

- **Categorization**:
  - Users who are motivated by Categorization engage in tagging because they want to construct and maintain a navigational aid to the resources (URLs, photos, etc) being tagged.
  - Resources are assigned to tags whenever they share some common characteristic important to the mental model of the user (e.g. ‘family photos’, ‘trip to Vienna’ or ‘favorite list of URLs’).
- **Description**:
  - Users who are motivated by Description engage in tagging because they want to accurately and precisely describe the resources being tagged.
  - Because the tags assigned are very close to the content of the resources, they can act as suitable facilitators for description and *searching*.

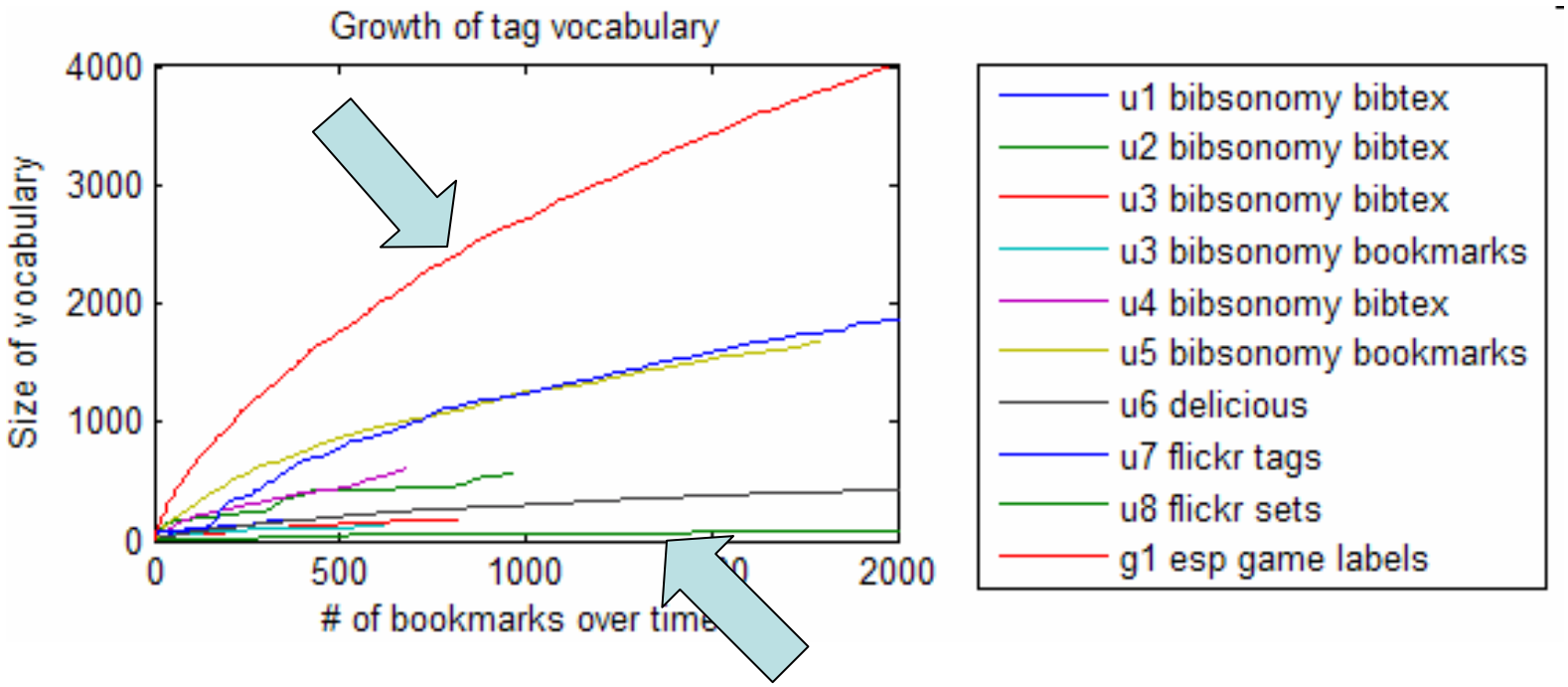
Detecting User Motivation

Potential Metrics:

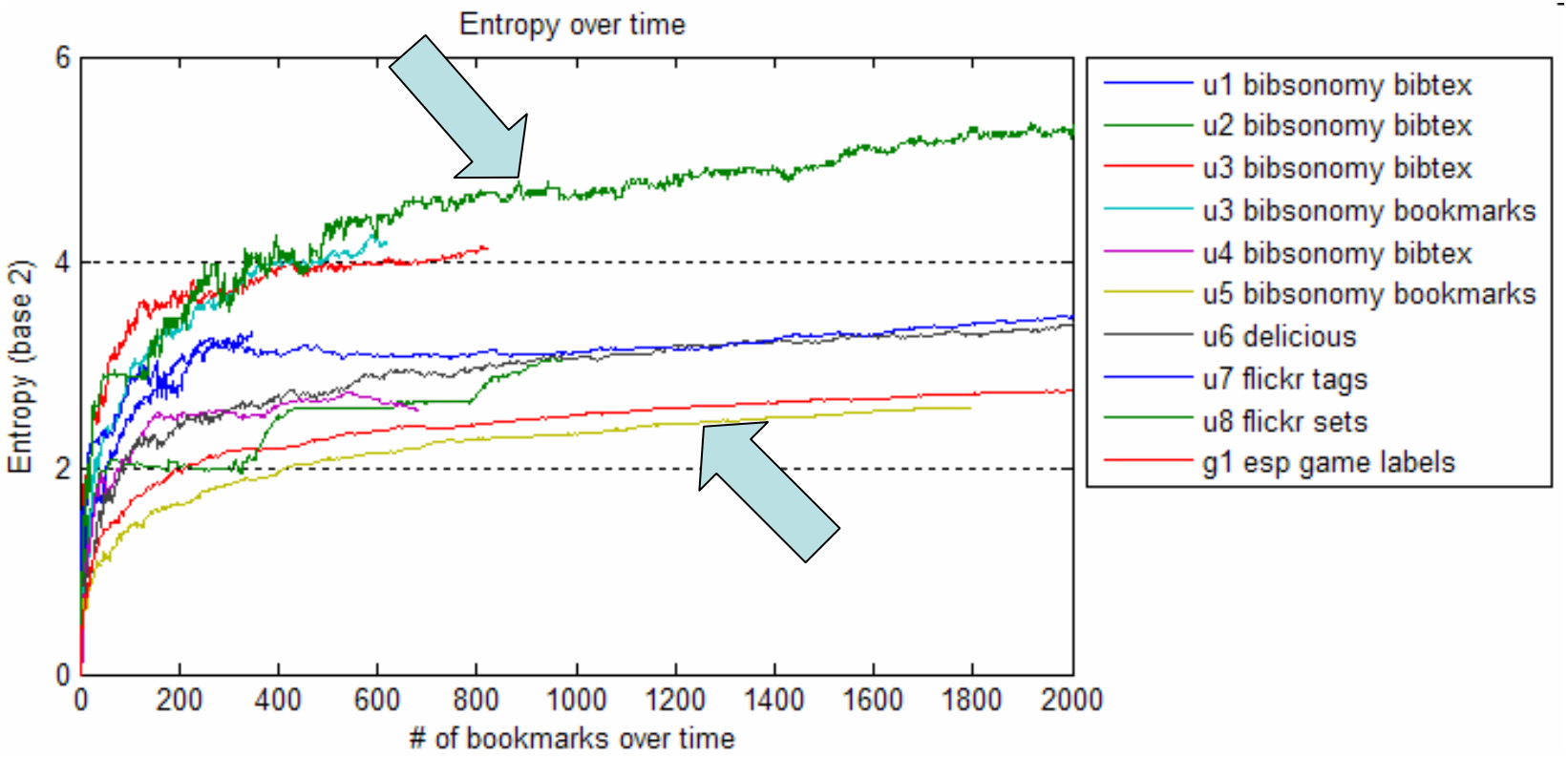
- Tag Vocabulary size
- Tag Entropy
- Percentage of Tag Orphans
- Tag Overlap


View Insert Tools Desktop Window Help
Frequency-rank distribution
Relative Tag Frequency
Tag Rank
Growth of Tagging Vocabulary
Size of vocabulary
Describer
Categorizer
# of bookmarks over time
Tag Vocabulary / Bookmarks
Tag Vocabulary/Bookmark ratio
Describer
Categorizer
# of bookmarks over time
Entropy over time
Entropy b2
Categorizer
Describer
# of bookmarks over time
Entropy over time (last j1)
Entropy b2
# of bookmarks over time
Entropy over time (last j2)
Entropy b2
# of bookmarks over time
Average rate of participation
avg. number of tags/bookmark
# of bookmarks over time
Average size of events
avg. number of bookmarks/tag
# of bookmarks over time
Average new tag ratio
new tag ratio
# of bookmarks over time
Orphan Tag Dominance
Orphan count / Number of all Tags
Describer
# of bookmarks over time
New tags introduced over last j1
percentage new tags introduced during j1
# of bookmarks over time
New tags introduced over last j2
percentage new tags introduced during j2
# of bookmarks over time
Maximum Tag Probability
max tag prob
# of bookmarks over time
Normalized Entropy over time
Entropy b2 / avg. rate of part.
Categorizer
Describer
# of bookmarks over time
average P (new Tag)
P(new Tag)
Describer
# of bookmarks over time

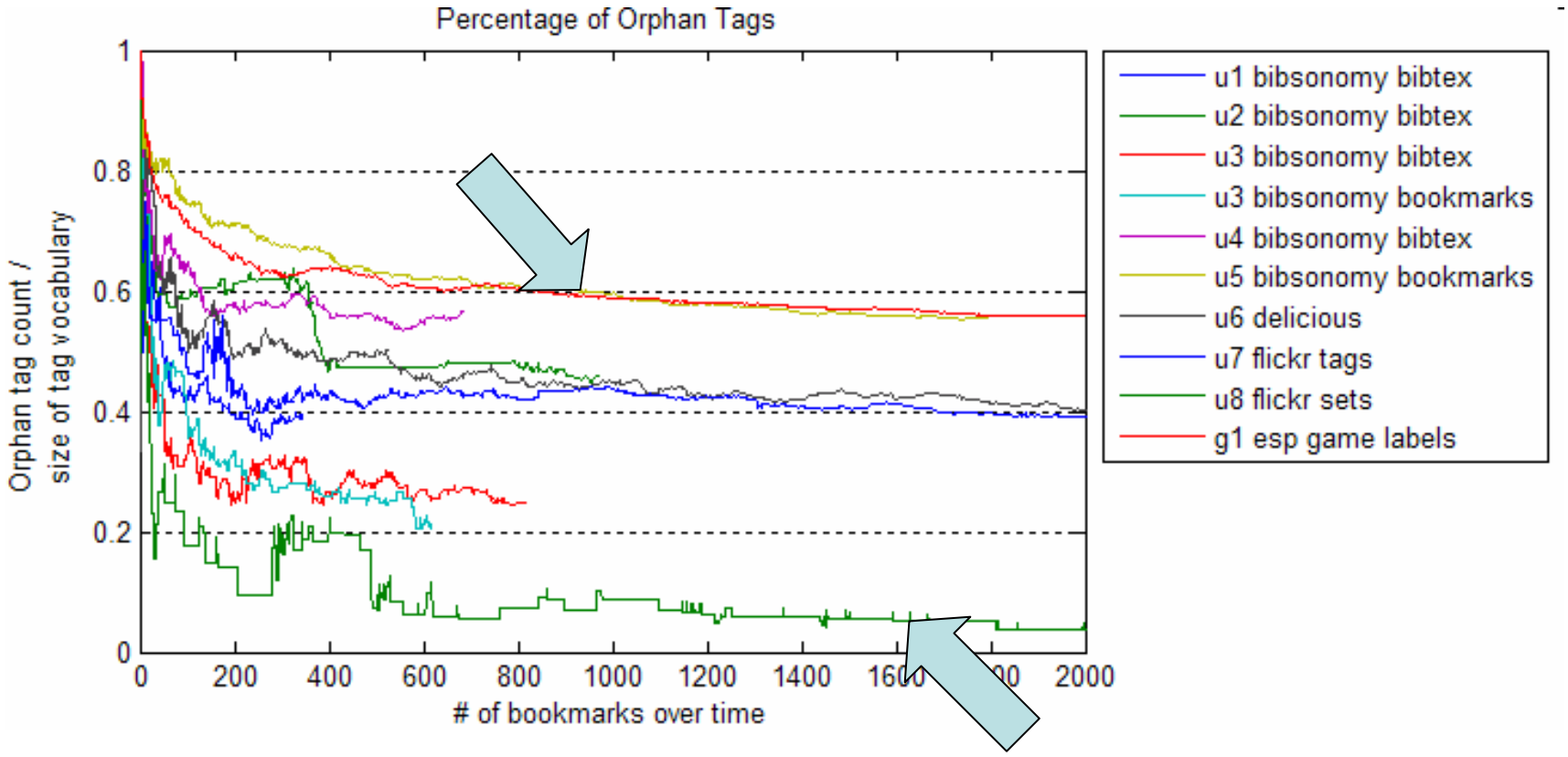
Preliminary Results: Vocabulary Size



Preliminary Results: Vocabulary Size



Preliminary Results: Tag Entropy



Categorization vs. Description

- **Implications and Relevance**:
  - *Tag Recommender Systems*:
    - Assuming that a user is a “Categorizer”, he will more likely reject tags that are recommended from a larger user population because he is primarily interested in constructing and maintaing “her” individual tag vocabulary.
  - *Search*:
    - Tags produced by “Describers” are more likely to be helpful for search and retrieval because they focus on the content of resources, where tags produced by “Categorizers” focus on their mental model. Tags by categorizers thus are more *subjective*, whereas tags by describers are more *objective*.
  - *Knowledge Acquisition*:
    - A tagging system primarily populated by categorizers is likely to give rise to a completely different set of possible folksonomies than tagging systems primarily populated by describers.

Thank you!

Any questions?