Navigability in Social Tagging Systems

Markus Strohmaier

Knowledge Management Institute,
Graz University of Technology, Austria
e-mail: markus.strohmaier@tugraz.at
web: http://www.kmi.tugraz.at/staff/markus

In collaboration with:
Christian Körner, Denis Helic, Keith Andrews, Christoph Trattner

Classification Systems in Information and Library Sciences

Overview of the Dewey Decimal Classification

The ten main classes are:

000	Computers, information & general reference
100	Philosophy & psychology
200	Religion
300	Social science
400	Language
500	Science
600	Technology
700	Arts & recreat
800	Literature
900	History & geography

Goal: to arrange items so that books / articles on a given subject are found close to similar ones.

to support easy **navigation**.

The ACM Computing Classification System (1998)

- A. General Literature
  - A.0 GENERAL
    - *Biographies/autobiographies*
    - *ction, plays)*
    - Y
    - es, encyclopedias, glossaries)
  - B.1 CONTROL STRUCTURES AND MICROPROGRAMMING (D.3.2)
    - B.1.0 General
    - B.1.1 Control Design Styles
      - *Hardwired control* [**]
      - *Microprogrammed logic arrays* [**]
      - *Writable control store* [**]
    - B.1.2 Control Structure Performance Analysis and Design Aids

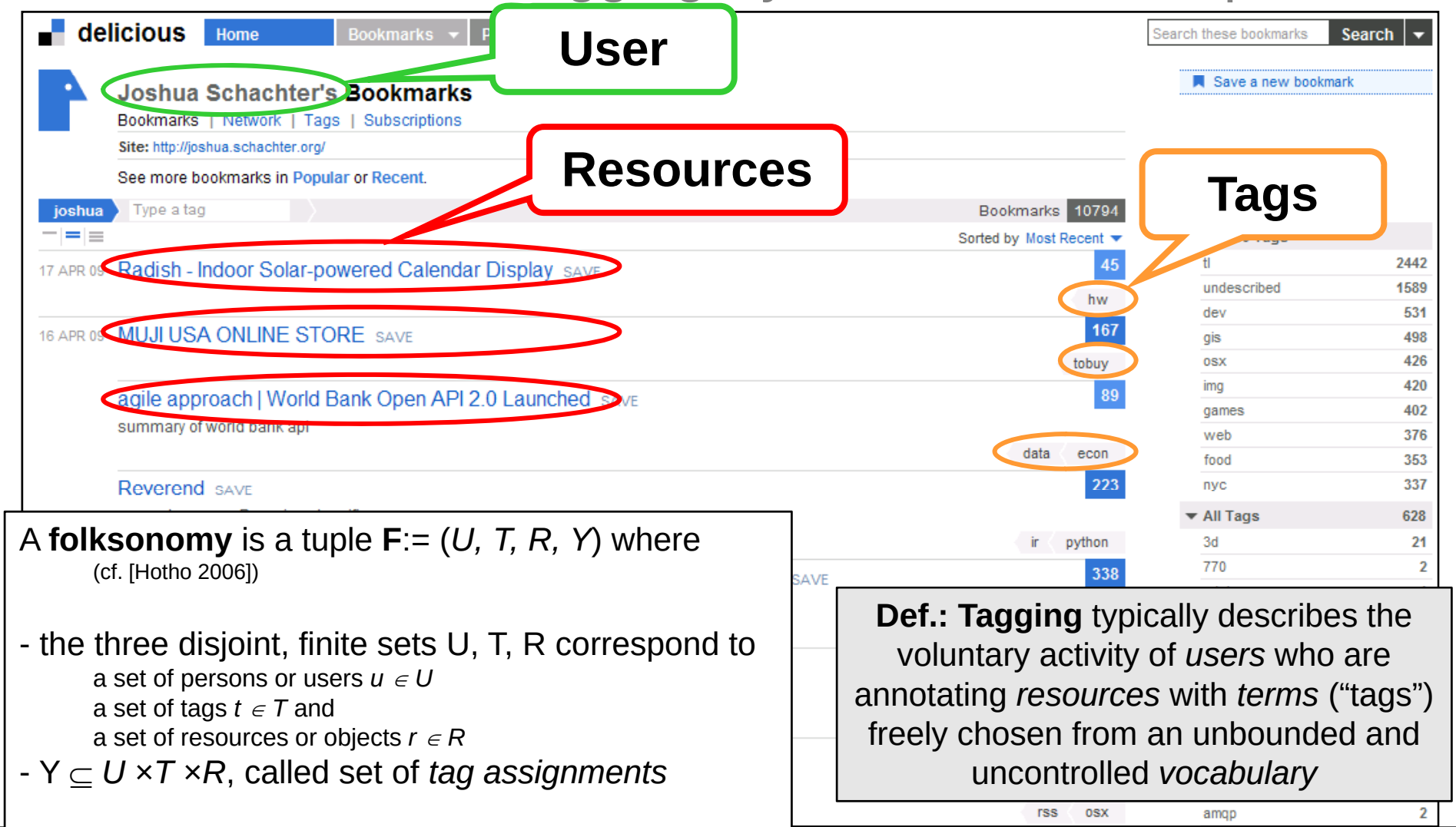
Example: Library of Congress

The **Library of Congress Classification** System:

- 21 basic classes, detailed subclasses
- In one year (2001-02), the LoC cataloged 310,235 bibliographic volumes at an avg. cost of $94.58 per volume
- each area is developed by an expert according to demands of cataloging

Issues: cost of classification, rigidity of schema, training of librarians, consistency of categories, number of categories, low adoption (vs. Dewey C.), …

What are Social Tagging Systems? An Example



A **folksonomy** is a tuple **F**:= (*U, T, R, Y*) where
(cf. [Hotho 2006])

- the three disjoint, finite sets U, T, R correspond to
  - a set of persons or users $u \in U$
  - a set of tags $t \in T$ and
  - a set of resources or objects $r \in R$
- $Y \subseteq U \times T \times R$, called set of *tag assignments*

A **personomy P_u** of user u is the restriction of F to u

Def.: Tagging typically describes the voluntary activity of *users* who are annotating *resources* with *terms* (“tags”) freely chosen from an unbounded and uncontrolled *vocabulary*

Tags: Why and What

Motivations for Tagging

- Future Retrieval
- Contribution and Sharing
- Attracting Attention (Flickr)
- Play and Competition (ESP Game)
- Self Presentation
- Opinion Expression
- Task Organization (“toread”)
- Social Signalling (“for:scott”)
- Money (Amazon Mechanical Turk)
- Technological Ease

Kinds of Tags

- Content-based
- Context-based
- Attribute Tags
- Ownership Tags
- Subjective Tags
- Organizational Tags
- Purpose Tags
- Factual Tags
- Personal Tags
- Self-referential tags
- Tag Bundles

Gupta et al. 2010

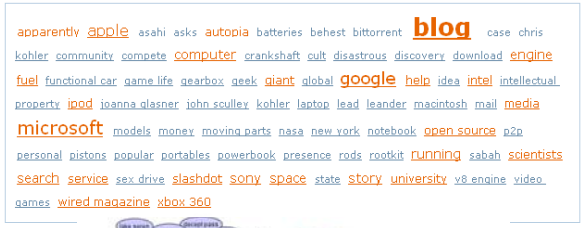
Visualization of Tags

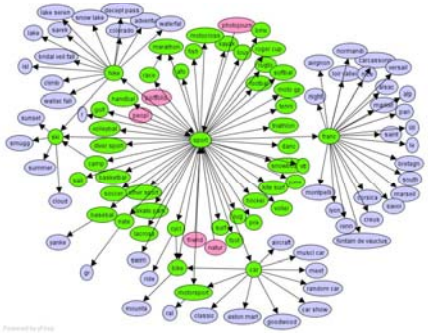
1. Tag clouds & tag selection strategies

2. Tag hierarchy generation

3. Tag Cloud Display Formats

4. Tag Evolution





(f) Tag-cloud dice (h) Tag-cloud slice

Gupta et al. 2010

Tagging Simulation

Models

- Basic Polya Urn Model
- Yule-Simon Model
- Yule-Simon Model with Long Term Memory
- Information Value Based Model
- Language Models

Taxonomy / Ontology Learning

Models

- Popularity Based

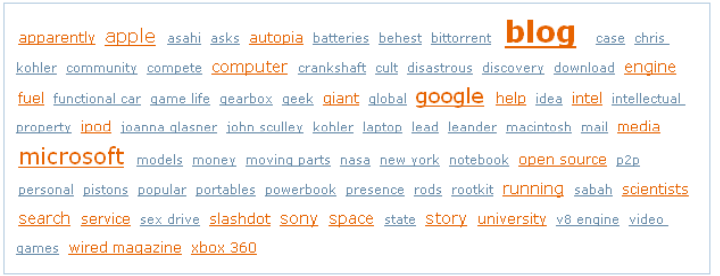
Gupta et al. 2010

Taxonomies vs. Folksonomies

A folksonomy is a **user-generated classification, emerging through bottom-up consensus** [1]

- Network of tags, users and resources (e.g. URLs)
- Users describe resources with tags
- Yields a “democratic” and emergent classification
- No explicitly defined relationship between terms, oka flat namespace

Tag clouds are a popular means for navigating folksonomies



[1] http://www.iskoi.org/doc/folksonomies.htm

Navigability

Informal Description:

If / how quick one can get from document A to document B in a hypertext system

(more precise definition follows later)

Designing for Navigability:

In traditional hypertext systems, this property used to be **within the control of system designers**

Social Navigation of Tagging Systems

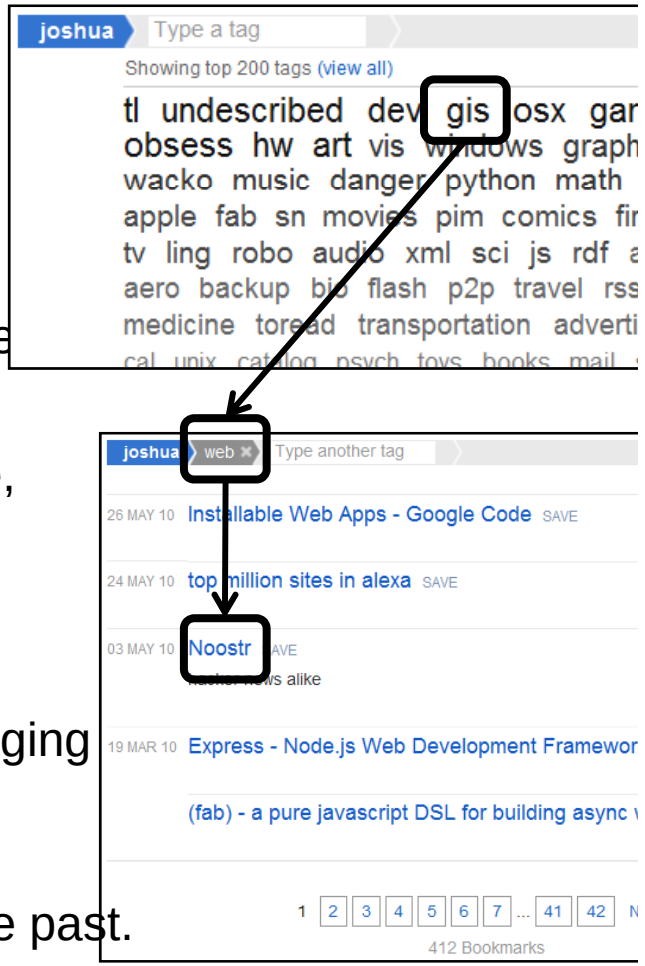
Definition: Social navigation refers to systems in which a user's navigation is guided by the behavior of others [4].
In such systems, the link structure is not created by a single person, but it is **the result of aggregating information from a group of users**.

In this sense, navigability of tagging systems can be understood as a result of social computation processes, which lie **mostly beyond the control of system designers**.

Tag Clouds are Supposed to be Efficient Tools for Navigating Tagging Systems

Navigating tagging systems via tag clouds:

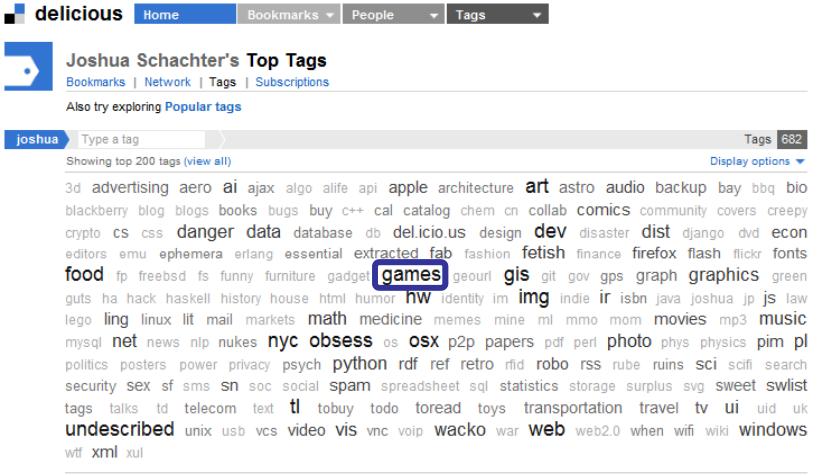
1) The system presents a tag cloud to the user.
2) The user selects a tag from the tag cloud.
3) The system presents a list of resources tagged with the
4) The user selects a resource from the list of resources.
5) The system transfers the user to the selected resource, and the process potentially starts anew.

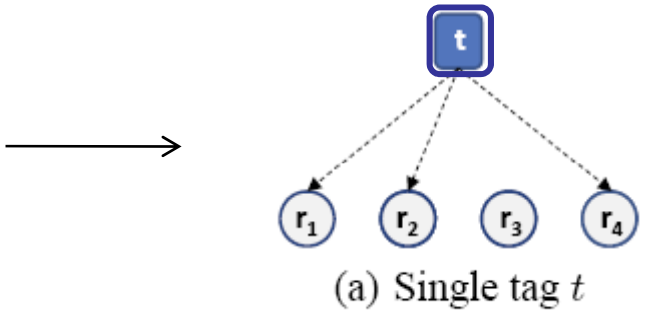


The **Navigability Assumption**:

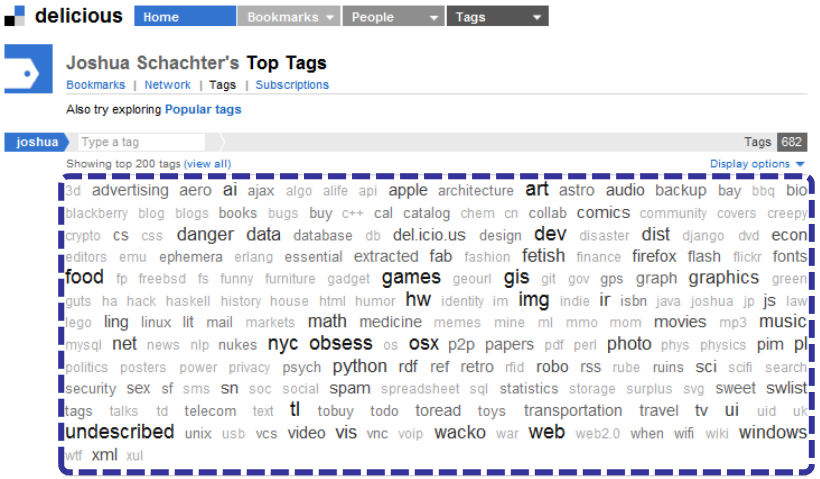
- An implicit assumption among designers of social tagging systems that tag clouds are specifically useful to support navigation.
- This has hardly been tested or critically reflected in the past.

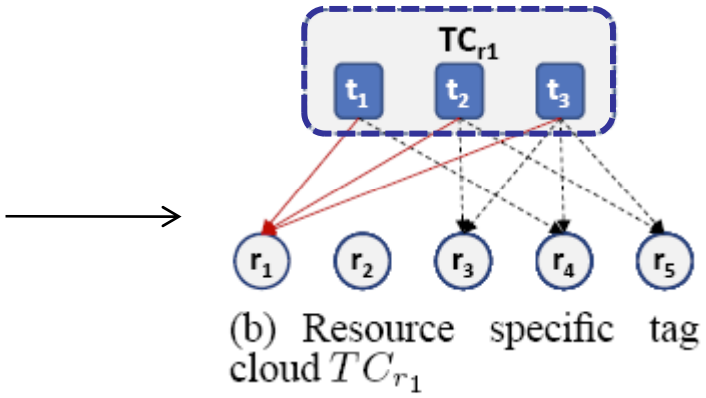
Bi-partite Nature of Tag Clouds





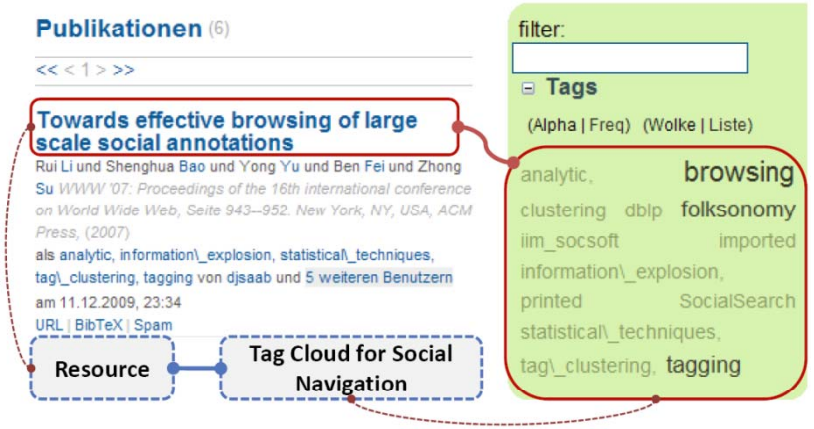
(a) Single tag t

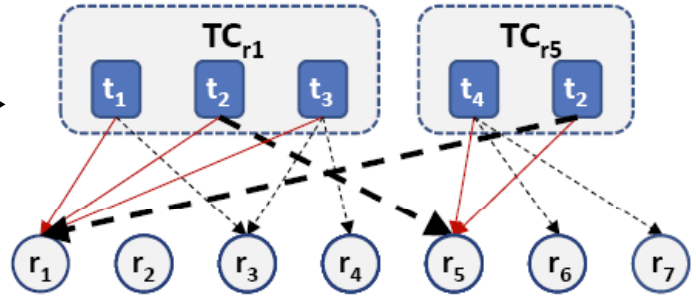




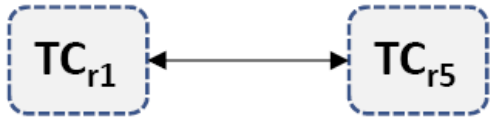
(b) Resource specific tag cloud TC_{r_1}

Tag Cloud Networks



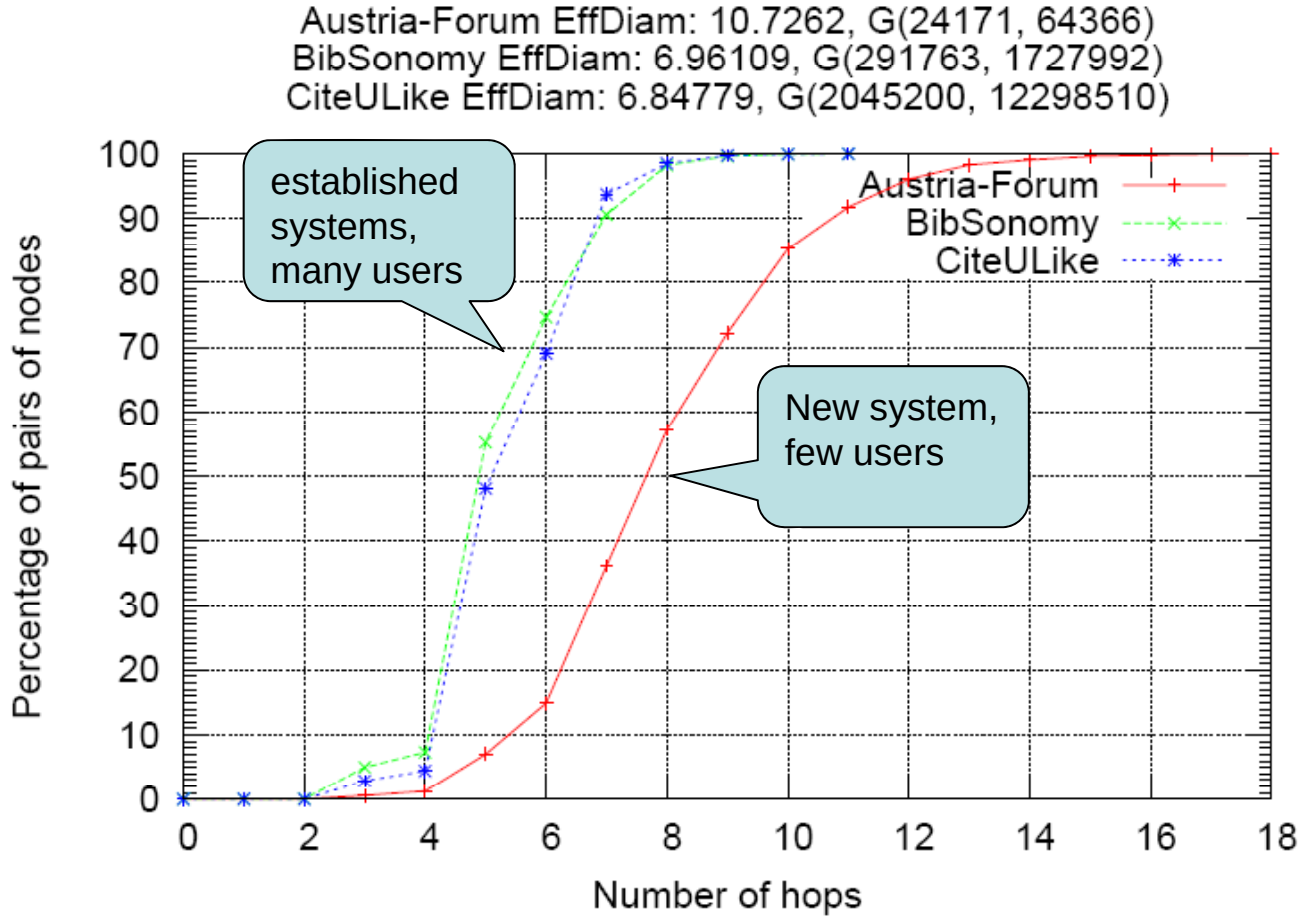


(a) Resource specific tag cloud network with tag clouds TC_{r_1} and TC_{r_5}



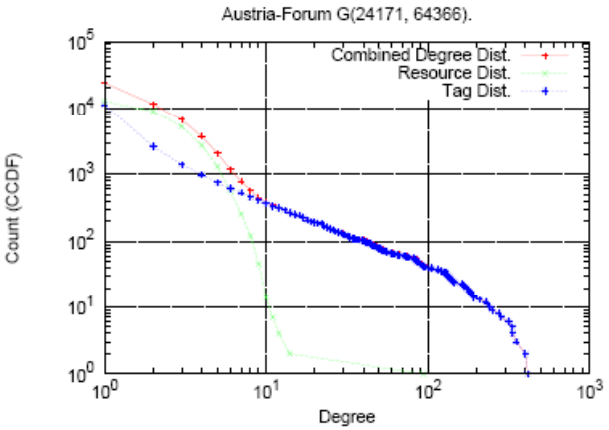
(b) Simplified resource specific tag cloud network with tag clouds TC_{r_1} and TC_{r_5}

Navigability of Social Tagging Systems

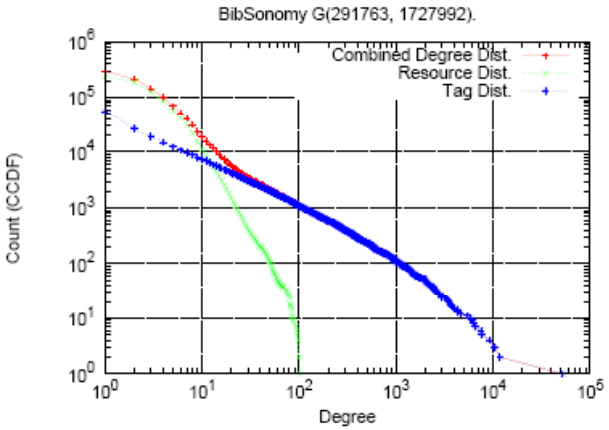


The usefulness of tag clouds for navigation is sensitive to the phase of adoption of the social tagging system

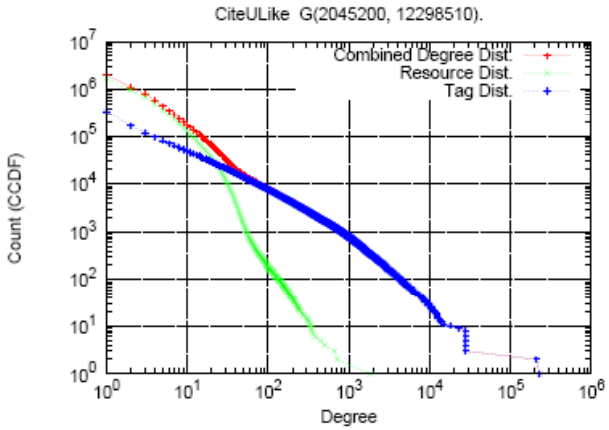
Navigability of Social Tagging Systems



(a) Austria-Forum



(b) BibSonomy



(c) CiteULike

Tagging networks are navigable power-law networks. For power law networks, efficient sub-linear decentralised navigation algorithms exist.

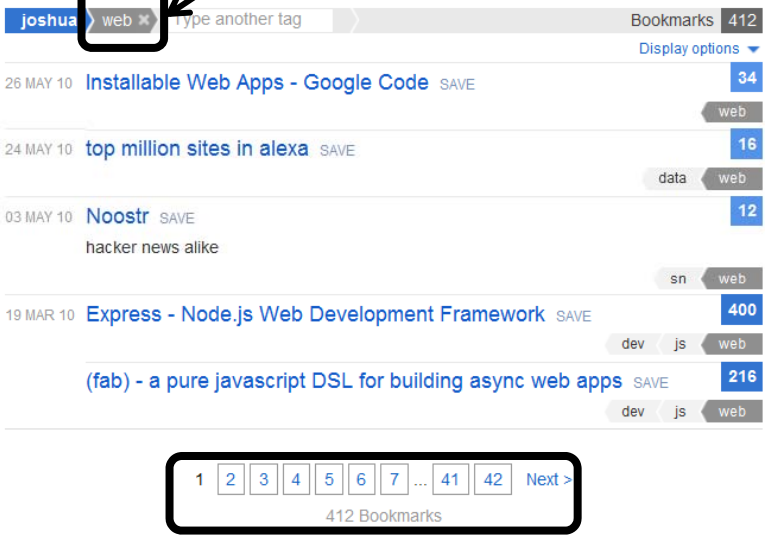
User Interface constraints

Tag Cloud Size
topN resources

(topN most common algorithm)

Pagination of resources / tag
k resources shown / page

(reverse chronological ordering most common)



Measuring Navigability

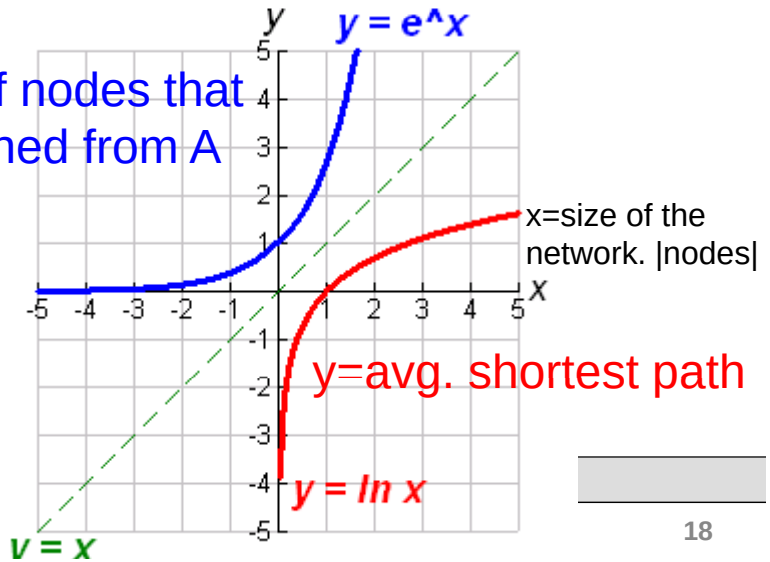
A network is **navigable** iff:

- there exists a giant component
- It has a low effective diameter

A network is **efficiently navigable** iff:

$p_{avg} <= \log |x|$

- The average shortest path p_{avg} is less than or equal the log size of the network |x|.
- For such networks, efficient decentralised search algorithms exist [Adamic].



How UI constraints effect Navigability

Tag Cloud Size

(a) Austria-Forum

UIR	GC	ED	UIA	NADT
none	0.77	10.73	none	sub-lin.
$n = 5$	0.75	10.99	TopN	sub-lin.
$n = 10$	0.76	11.3	TopN	sub-lin.
$n = 20$	0.76	11.97	TopN	sub-lin.
$n = 30$	0.76	11.05	TopN	sub-lin.
$k = 5$	0.36	12.04	Chron.	unnav.
$k = 10$	0.47	11.16	Chron.	unnav.
$k = 20$	0.56	10.31	Chron.	unnav.
$k = 30$	0.6	10.68	Chron.	unnav.

(b) BibSonomy

UIR	GC	ED	UIA	NADT
none	0.98	6.96	none	sub-lin.
$n = 5$	0.94	6.8	TopN	sub-lin.
$n = 10$	0.97	6.87	TopN	sub-lin.
$n = 20$	0.98	6.84	TopN	sub-lin.
$n = 30$	0.98	6.91	TopN	sub-lin.
$k = 5$	0.31	6.82	Chron.	unnav.
$k = 10$	0.4	6.62	Chron.	unnav.
$k = 20$	0.5	6.61	Chron.	unnav.
$k = 30$	0.54	6.65	Chron.	unnav.

(c) CiteULike

UIR	GC	ED	UIA	NADT
none	0.98	6.85	none	sub-lin.
$n = 5$	0.93	6.97	TopN	sub-lin.
$n = 10$	0.95	7.07	TopN	sub-lin.
$n = 20$	0.97	7.17	TopN	sub-lin.
$n = 30$	0.97	6.98	TopN	sub-lin.
$k = 5$	0.27	6.89	Chron.	unnav.
$k = 10$	0.36	6.95	Chron.	unnav.
$k = 20$	0.44	6.91	Chron.	unnav.
$k = 30$	0.48	7.05	Chron.	unnav.

UIR = UI Restriction, GC = Giant Component, ED = Effective Diameter, UIA = UI Algorithm, NADT = Navigation Algorithm Delivery Time
Chron. = Chronological algorithm, sub-lin. = sub-linear, unnav. = unnavigable network

Pagination

TABLE I
NAVIGATIONAL PROPERTIES OF THE AUSTRIA-FORUM, BIBSONOMY, AND CITEULIKE TAGGING SYSTEMS.

Limiting the tag cloud size n to practically feasible sizes (e.g. 5, 10, or more) does not influence navigability.

BUT: Limiting the out-degree of high frequency tags k (e.g. through pagination with resources sorted in reverse-chronological order) leaves the network vulnerable to fragmentation. This **destroys navigability** of prevalent approaches to tag clouds.

Findings

1. For certain specific, but popular, tag cloud scenarios, the so-called Navigability Assumption does not hold.
2. While we could confirm that tag-resource networks have efficient navigational properties in theory, we found that popular user interface decisions significantly impair navigability.

These results make a theoretical and an empirical argument against existing approaches to tag cloud construction. **New approaches are needed.**

Recovering Navigability in Social Tagging Systems

Instead of reverse-chronological ordering of resources, we apply a **random ordering.**

(a) Austria-Forum

UIR	GC	ED	UIA	NADT
k=5	0.86	11.7	Random	linear
k=10	0.86	11.02	Random	linear
k=20	0.85	10	Random	linear
k=30	0.84	10.42	Random	linear

(b) BibSonomy

UIR	GC	ED	UIA	NADT
k=5	0.99	8.75	Random	linear
k=10	0.99	6.97	Random	linear
k=20	0.99	6.75	Random	linear
k=30	0.99	6.46	Random	linear

(c) CiteULike

UIR	GC	ED	UIA	NADT
k=5	0.99	7.98	Random	linear
k=10	0.99	7.88	Random	linear
k=20	0.99	7.13	Random	linear
k=30	0.99	6.86	Random	linear

UIR = UI Restriction, GC = Giant Component, ED = Effective Diameter, UIA = UI Algorithm, NADT = Navigation Algorithm Delivery Time

TABLE II
NAVIGATIONAL PROPERTIES OF THE AUSTRIA-FORUM, BIBSONOMY, AND CITEULIKE TAGGING SYSTEMS WITH A RANDOM PAGINATION ALGORITHM.

Trade-off between Semantic and Navigational Properties

Semantic Penalty:

- Random ordering of links is counter-intuitive for navigation

Navigational Penalty:

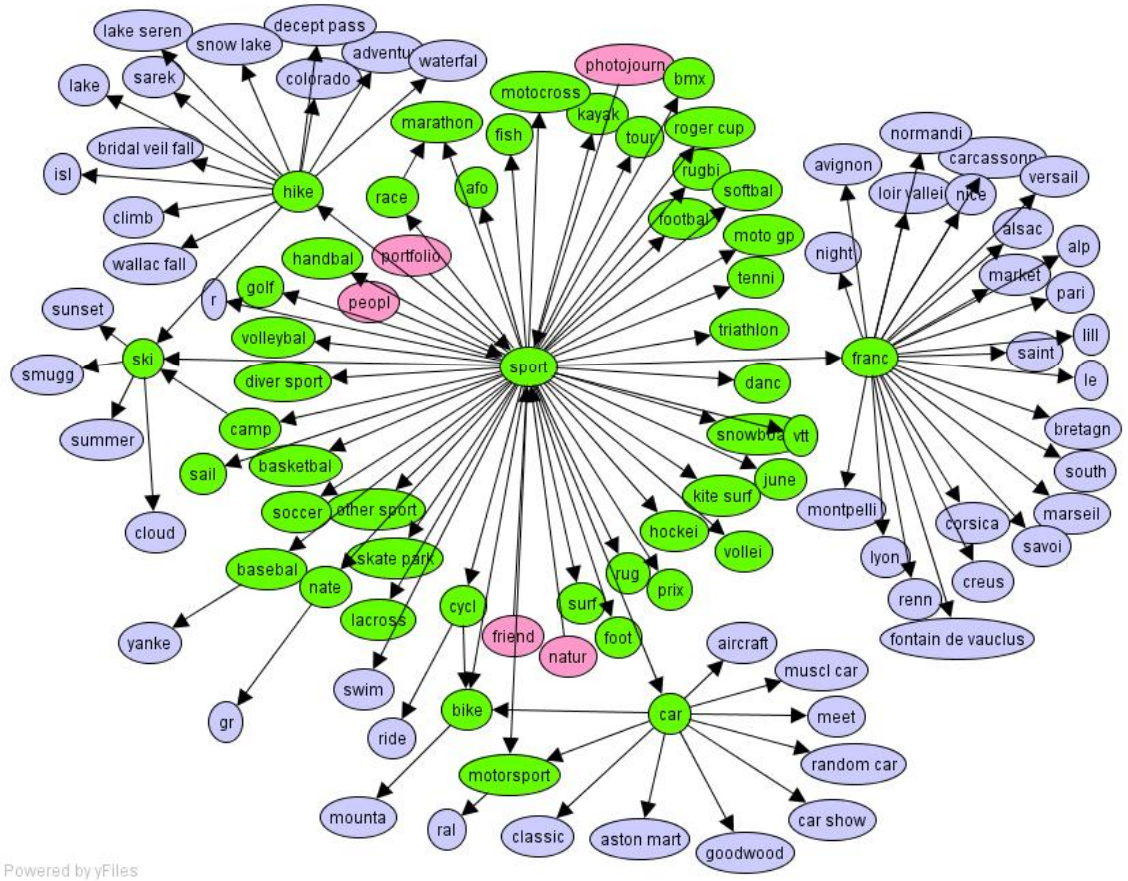
- Semantic ordering of links might impair overall navigability

Hypotheses:

- There is a semantic penality imposed by maximizing for navigability

- There is a navigational penality imposed by maximizing for semantic relatedness

Outlook: Evaluating Folksonomies



Lerman et al 2010

Conclusions

1. Certain properties of social media (such as navigability) are emergent properties, that are beyond the direct influence of system designers
2. User interface constraints can effect (but do not determine) these emergent properties
3. By studying the relationship between user interface and network phenomena, system engineers can influence (some aspects of) emergent system properties

Related Work on Tagging Systems

- Tagging motivation influences tagging behaviour
- Categorizer & Describer
- Recommendation
- Emergent Semantics

Next Lecture:

- **Wednesday!**

9.6.2010 10:15 - 11:45 HS i11 "SIEMENS AG Österreich"

On Tagging Motivation, held by Christian Körner

- **NO LECTURE next week!**

End of Presentation

Thank you!

Markus Strohmaier

markus.strohmaier@tugraz.at
Graz University of Technology, Austria