IEEE/WIC/ACM Joint International Conference on Web Intelligence and Intelligent Agent Technology 2007

Workshop on Collective Intelligence on Semantic Web (CISW 2007)

Tag Meaning Disambiguation

through Analysis of Tripartite Structure of Folksonomies

Ching-man Au Yeung, Nicholas Gibbins, Nigel Shadbolt

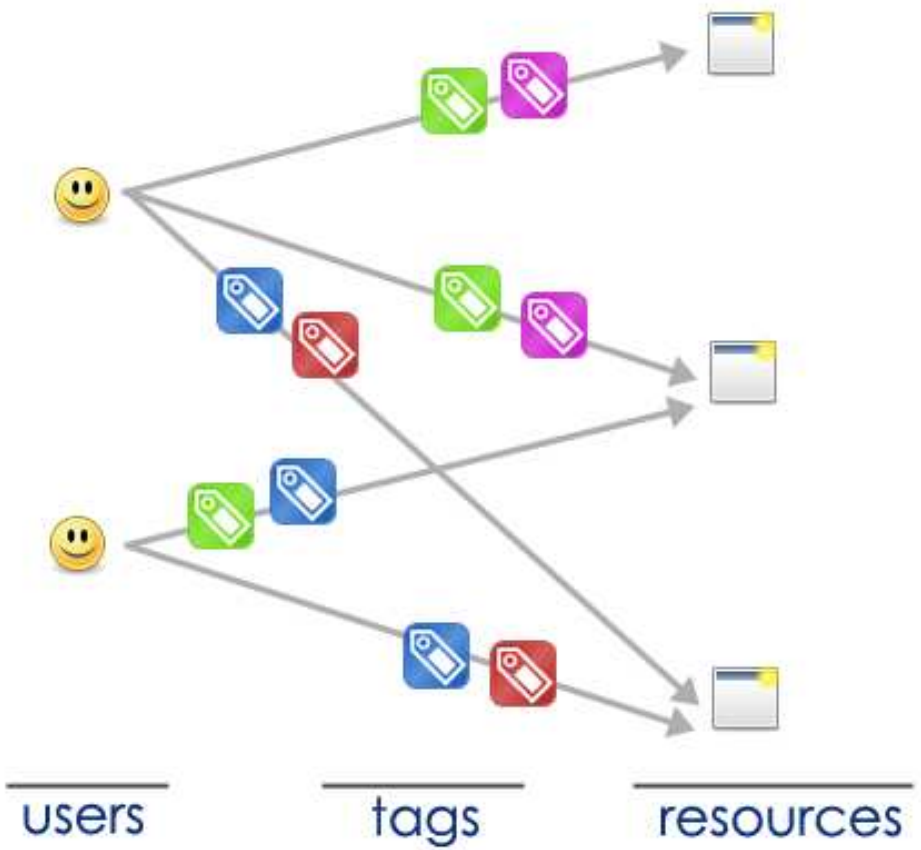
University of Southampton

ECS
Electronics & Computer Science
University of Southampton

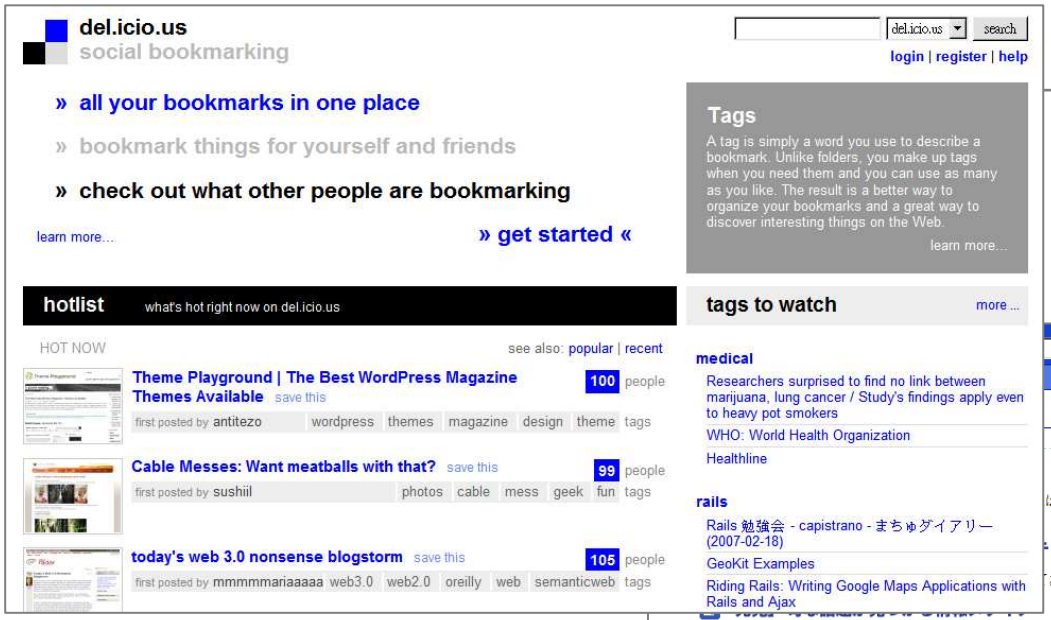
Overview

- Background
- Motivations
- Tripartite structure of folksonomies
- Tag meaning disambiguation
- Experiments
- Conclusions and future work

- ***Collaborative tagging systems*** and ***folksonomies***



- Examples of collaborative tagging systems



http://del.icio.us/

http://b.hatena.ne.jp/

Background

- Advantages [Adam 2004, Wu et al. 2006]
  - Freedom and flexibility
  - Quick adaptation to changes in vocabulary (e.g. ajax, youtube)
  - Convenience and serendipity
- Disadvantages [Adam 2004, Wu et al. 2006]
  - Ambiguity (e.g. apple, sf, opera)
  - Lack of format (e.g. how multiword tags are handled)
  - Existence of synonyms (e.g. semweb, semanticweb, semantic_web)
  - Lack of semantics

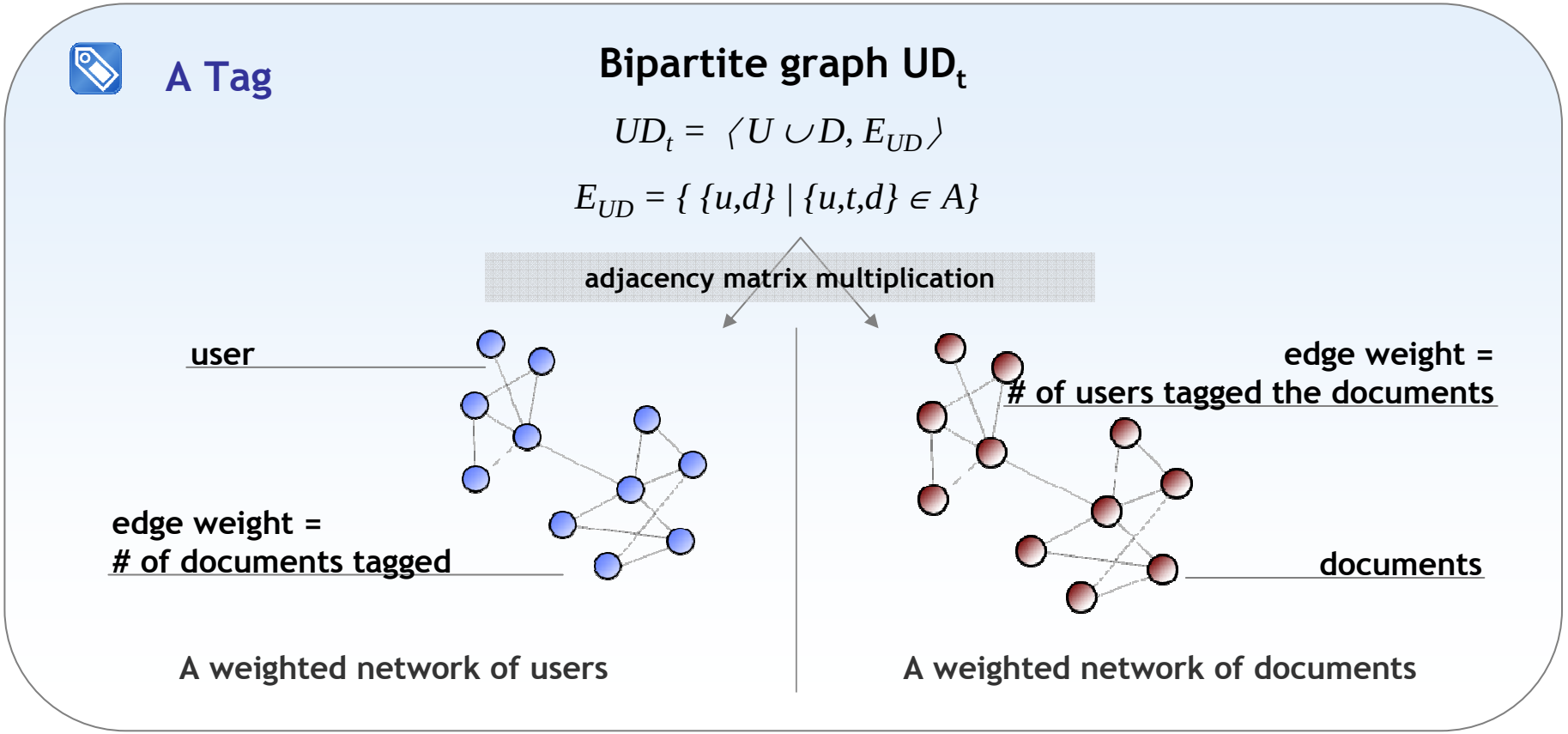
Motivations

- Many tags are ambiguous (possess multiple meanings)
- This affects the precision of retrieval and annotation of shared resources
- Current research works mainly focus on clustering of tags
- Few works deal with ambiguous tags, and in indirect ways only (e.g. [Wu et al. 2006])

Tripartite structure of folksonomies

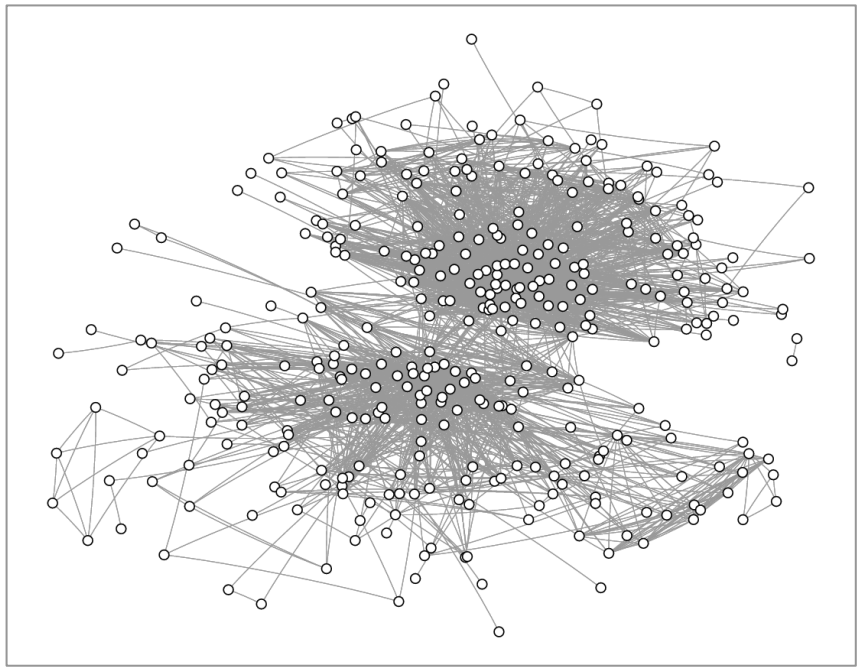
Folksonomy (A hypergraph)

$$F = \langle U, T, D, A \rangle; A \subseteq U \times T \times D$$

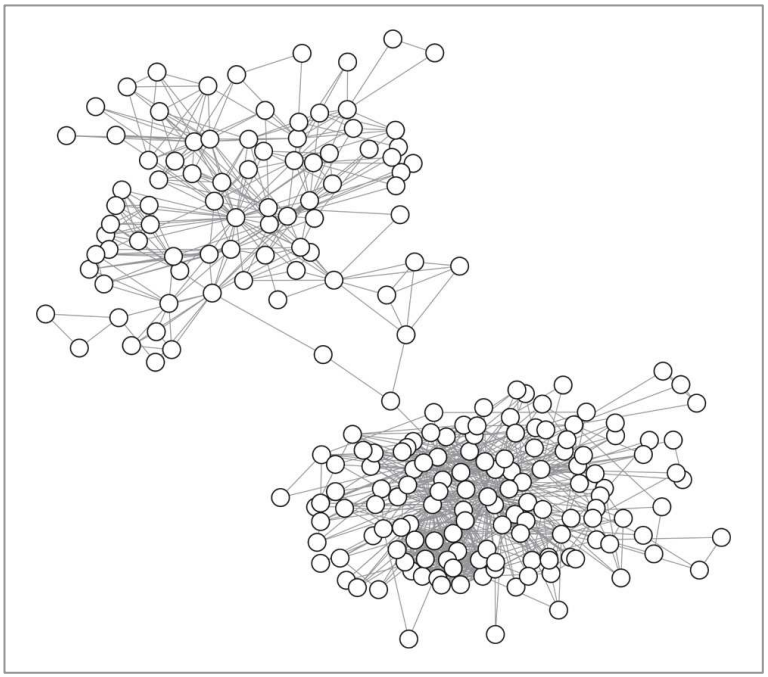


A case study

- *sf* in *del.ici.ous* [Au Yeung et al. 2007]



Network of Documents



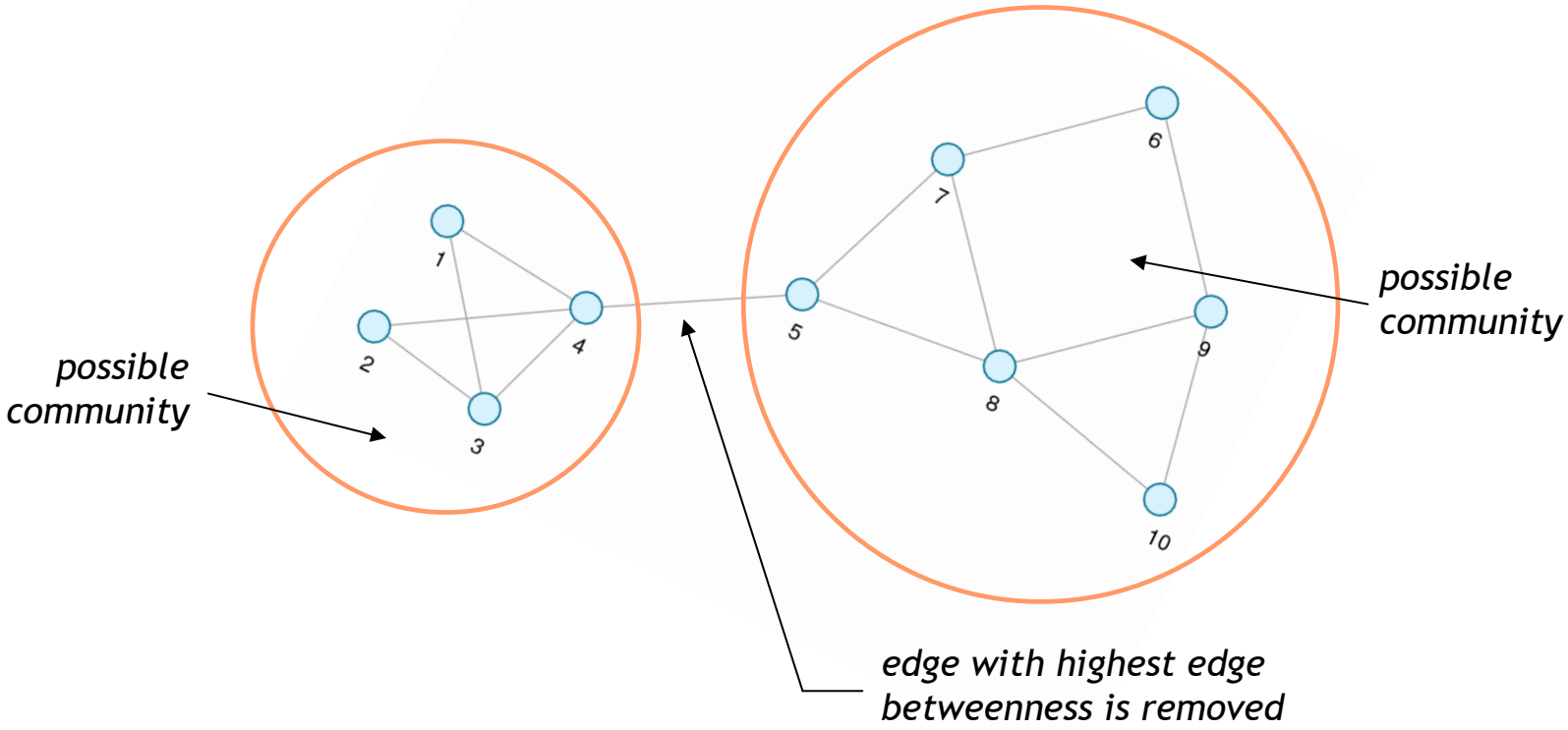
Network of Users

Tag Meaning Disambiguation

- Basic ideas
  - Different clusters of nodes in the network correspond to different meanings of the tag
  - Different meanings of ambiguous tags can be obtained by partitioning the network into communities of nodes
  - The meanings can be understood by examining the most frequently used tags within a cluster

- Algorithms for discovering communities in a network
  - ***Modularity optimization by removing edges based on edge betweenness*** [Newman & Girvan 2004]
  - ***Modularity***: a measure of the "goodness" of a partition of a network
  - ***Edge betweenness***: a measure of how likely an edge is a bridge between two communities

Tag Meaning Disambiguation



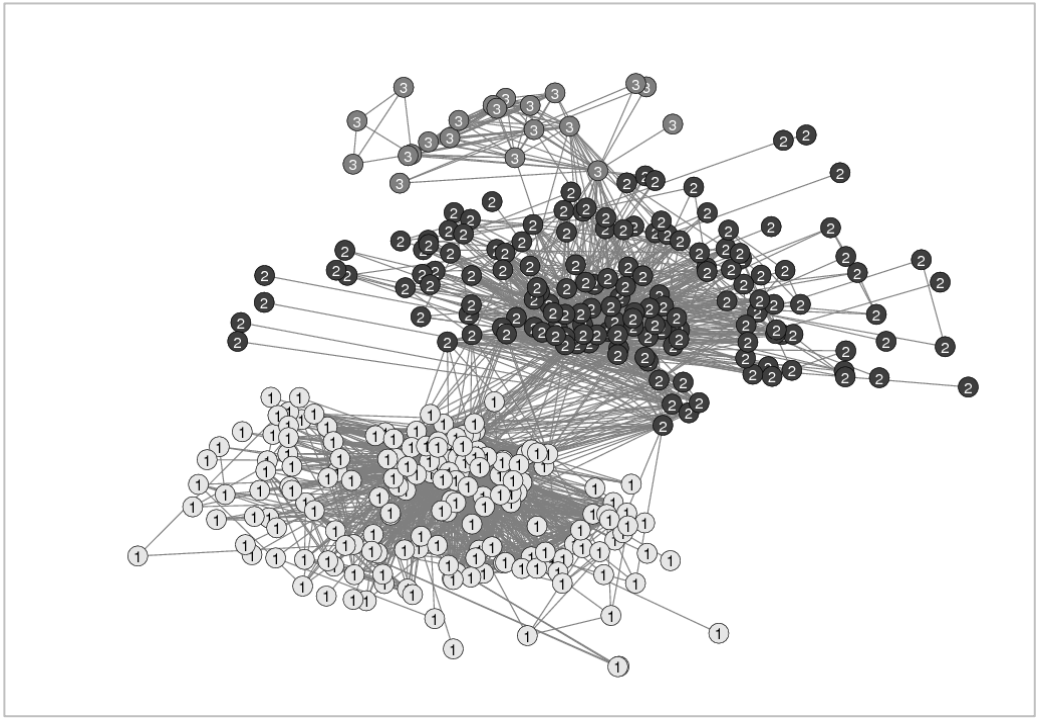
edge betweenness of edge ***e*** = number of shortest path running through ***e***

(most likely to be a bridge between two communities in the network)

- Proposed method

1. Collect tagging data of the tag to be disambiguated (including ***documents*** with the tag, ***users*** and other ***tags*** involved)
2. Construct a ***document network*** out of the data
3. Apply the ***community-discovering algorithm*** to the network
4. For each community discovered, extract the ***10 most frequently used tags*** among those documents
5. The sets of tags should give different meanings of the tag being examined

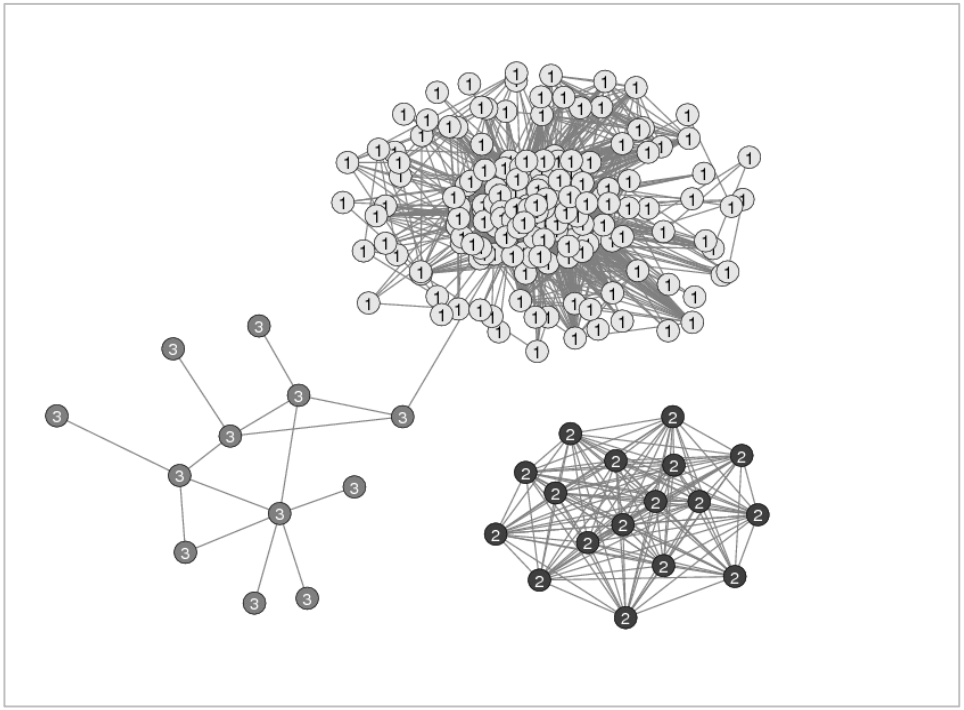
Experiments



Cluster	Tags
1	sf, scifi, fiction, books, sci-fi, writing, literature, science, sciencefiction, fantasy
2	sf, sanfrancisco, bayarea, san, francisco, california, travel, events, art, san_francisco
3	sf, sanfrancisco, design, bayarea, blog, food, todo, california, shopping, san

Disambiguation of the tag **"sf"**

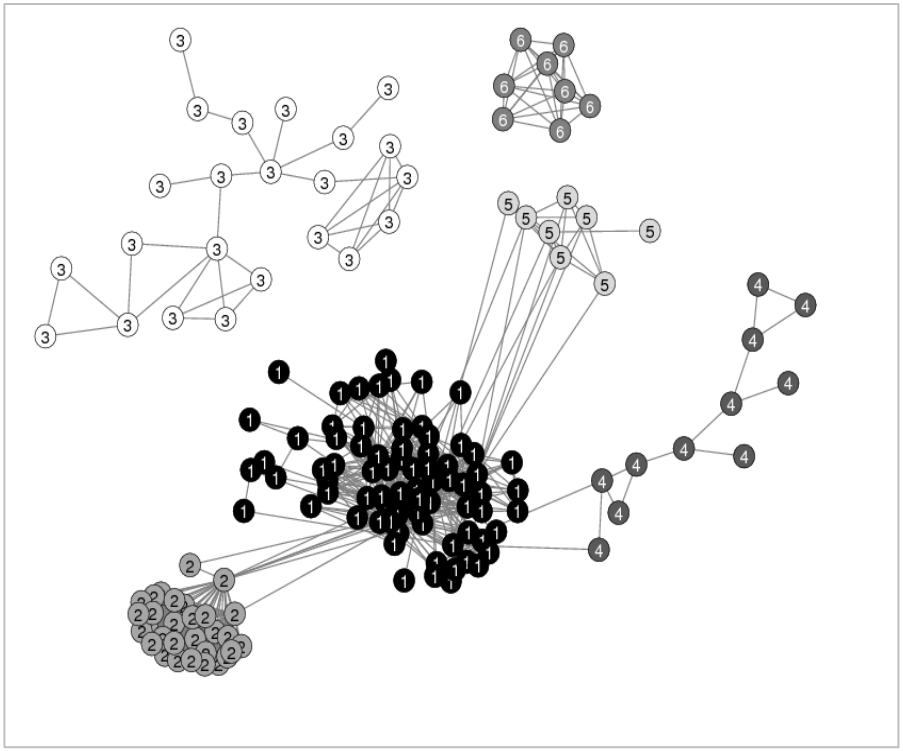
Experiments



Cluster	Tags
1	opera, browser, web, software, javascript, browsers, tips, tools, internet, firefox
2	opera, shopping, imported, shop, design, store, home, inspiration, work, personal
3	opera, music, musique, classical, art, culture, musica, musica, classic, travel

Disambiguation of the tag **"opera"**

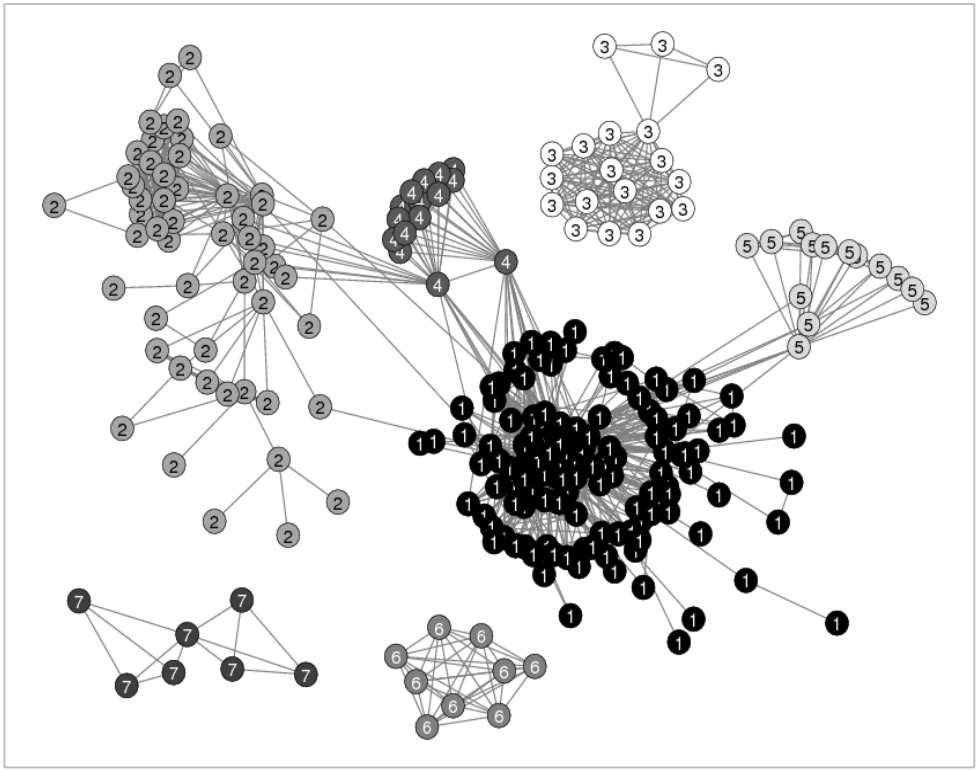
Experiments



Cluster	Tags
1	cambridge, university, uk, england, science, cam, local, cambridgeuniversity, research, community
2	cambridge, bcc_school, activism, education, community, contact, bcc, politics, critical_economy, blog
3	cambridge, boston, restaurants, food, massachusetts, imported, local, restaurant, venues, clubs
4	cambridge, english, cpe, cae, boston, online, fce, exam, inglés, esl
5	cambridge, mappingurbanism, visualisation, design, social, information, maps, mapping, infovis, toread
6	cambridge, letting, uk, photography, search, property, flats, cambsproperty, financial, fundraising

Disambiguation of the tag **"cambridge"**

Experiments



Cluster	Tags
1	tube, london, underground, travel, transport, maps, uk, map, subway, reference
2	tube, diy, audio, electronics, amp, amplifier, amps, tubes, guitar, music
3	tube, video, web, internet, tv, online, web2.0, media, videos, imported
4	tube, video, youtube, videos, funny, cool, interesting, sport, fun, humor
5	tube, video, videos, online, web2.0, youtube, free, media, movie, fun
6	tube, youtube, video, videos, cool, feel.good, fun, funny, flash, music
7	tube, radio, electronics, tubes, antique, amplifier, data, audio, info, incarnate

Disambiguation of the tag **"tube"**

Discussions

- Different meanings of a tag can be obtained from the result
- However, some problems exist:
  1. The meaning of a tag in some clusters is unclear (as in the results of *opera*, *cambridge*)
  2. Some clusters correspond to the same meaning of a tag (as in the result of *tube*)
  3. Some clusters correspond to the same meaning of a tag, but referring to such meaning in possibly different contexts (as in the result of *sf*)

Conclusions and future work

- Conclusions
  - The method is effective in clustering documents of different topic and discover the different meanings of a tag
  - Some post-processing may be required to clean up redundant or unclear clusters
- Future work
  - To improve the efficiency of the method by employing faster algorithms (e.g. [Clauset et al. 2004])
  - Investigate how the results can be refined to produce more useful disambiguation
  - Perform Larger scale of evaluation of the proposed method

References

1. Mathes Adam. Folksonomies – cooperative classification and communication through shared metadata. *http://www.adammathes.com/academic/computer-mediated-communication/folksonomies/html*, 2004.

2. C.M. Au Yeung, N. Gibbins and N. Shadbolt. Understanding the Semantics of Ambiguous Tags in Folksonomies. In *Proceedings of the International Workshop on Emergent Semantics and Ontology Evolutions*, ISWC 2007, Busan, South Korea, 2007.

3. Peter Mika. Ontologies are us: A unified model of social networks and semantics. In *Proceedings of International Semantic Web Conference*, pages 522-536, 2005.

4. M. E. J. Newman and M. Girvan. Finding and evaluating community structures in networks. *Physical Review E*, 69:026113, 2004.

5. Aaron Clauset, M. E. J. Newman and Christopher Moore. Finding community structure in very large networks. *Physical review E*, 70:066111, 2004.

6. Xian Wu, Lei Zhang, and Yong Yu. Exploring social annotations for the semantic web. In *WWW'06: Proceedings of the 15th international conference on World Wide Web*, pages 417-426, New York, NY, USA, 2006. ACM Press.