The 6th International Semantic Web Conference (ISWC 2007)

International Workshop on Emergent Semantics and Ontology Evolution

Understanding the Semantics of Ambiguous Tags in Folksonomies

Ching-man Au Yeung, Nicholas Gibbins, Nigel Shadbolt

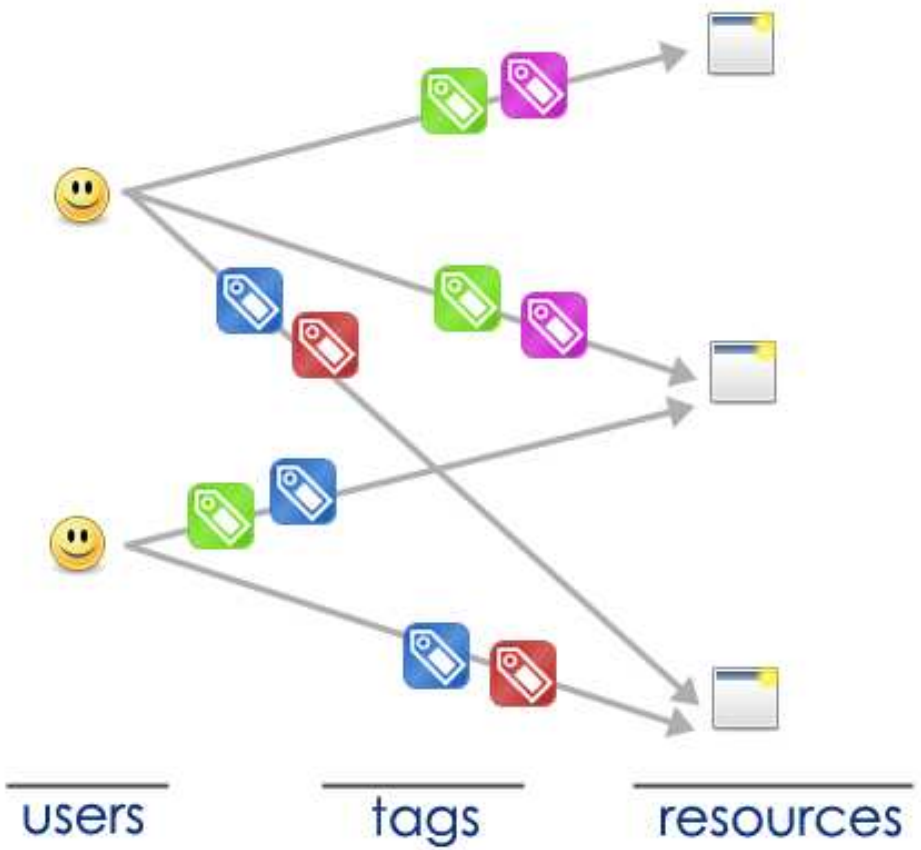
University of Southampton

Electronics & Computer Science
University of Southampton

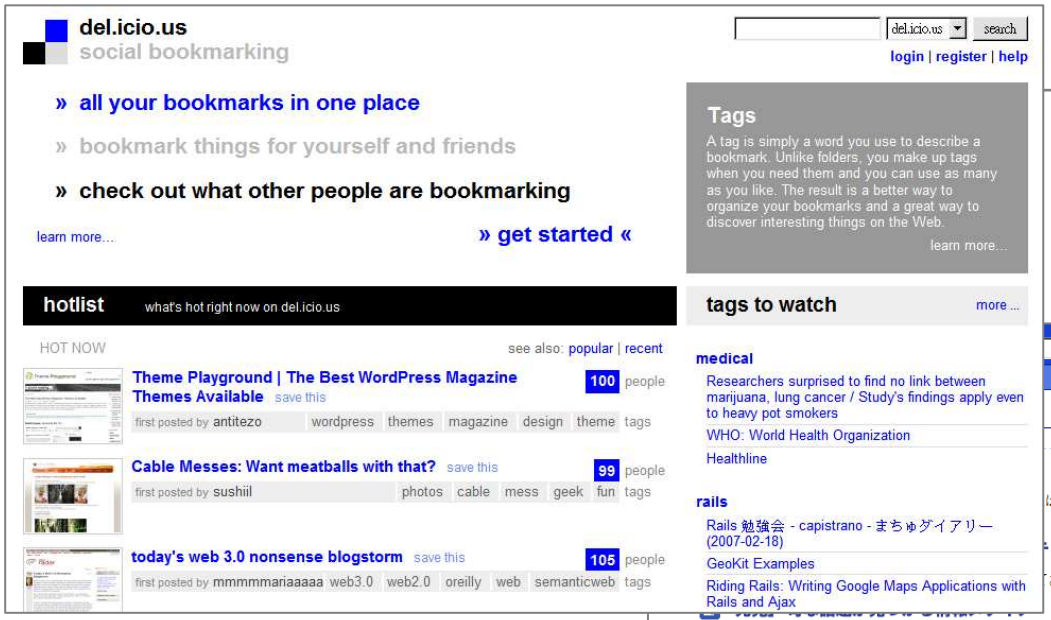
Overview

- Background (Collaborative tagging systems, folksonomies)
- Mutual contextualization in folksonomies
- Semantics of tags
- Discussions
- Conclusion and Future Work
- Recent Development

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



- Examples of collaborative tagging systems



http://del.icio.us/

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

- 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

Are folksonomies really so chaotic?

Mutual contextualization in folksonomies

- Folksonomies are actually associations between the three types of entity – users, tags and resources [Mika 2005]
- Associations between these entities are not randomly made
- There is always a reason why a particular user uses a particular tag to describe a particular Web resources
- Semantics embedded in folksonomies → ***mutual contextualization*** between the entities

Mutual contextualization in folksonomies

Folksonomy (A hypergraph)

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

A User

Bipartite graph TD_u

$$TD_u = \langle T \cup D, E_{TD} \rangle$$

$$E_{TD} = \{ \{t,d\} \mid \{u,t,d\} \in A \}$$

adj matrix multiplication

tag network | document network

A Tag

Bipartite graph UD_t

$$UD_t = \langle U \cup D, E_{UD} \rangle$$

$$E_{UD} = \{ \{u,d\} \mid \{u,t,d\} \in A \}$$

adj matrix multiplication

user network | document network

A Document

Bipartite graph UT_d

$$UT_d = \langle U \cup T, E_{UT} \rangle$$

$$E_{UT} = \{ \{u,t\} \mid \{u,t,d\} \in A \}$$

adj matrix multiplication

user network | tag network

Mutual contextualization in folksonomies

A Tag

Bipartite graph UD_t

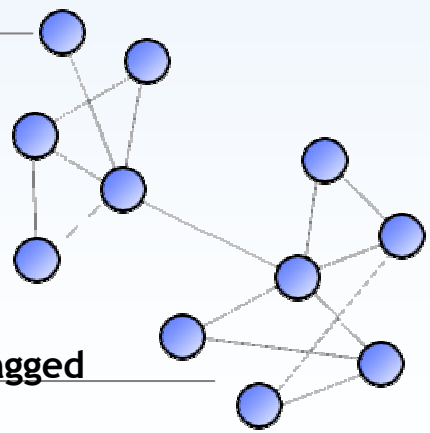
$$UD_t = \langle U \cup D, E_{UD} \rangle$$

$$E_{UD} = \{ \{u,d\} \mid \{u,t,d\} \in A \}$$

adjacency matrix multiplication

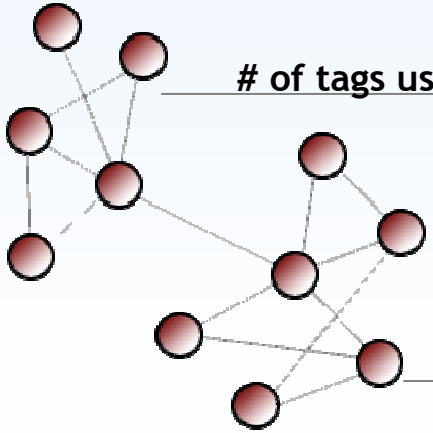
user

edge weight =
# of documents tagged



A weighted network of users

edge weight =
# of tags used on documents



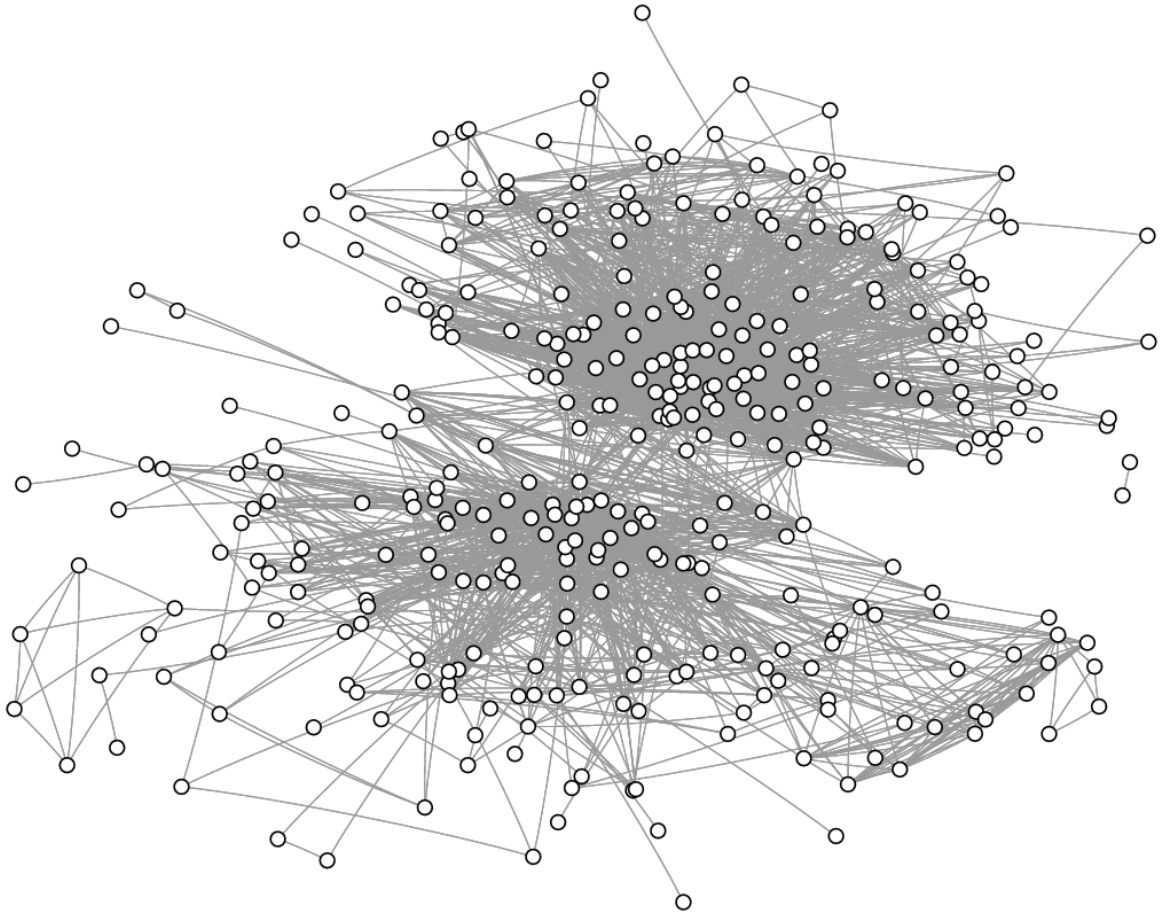
documents

A weighted network of documents

Understanding a single tag

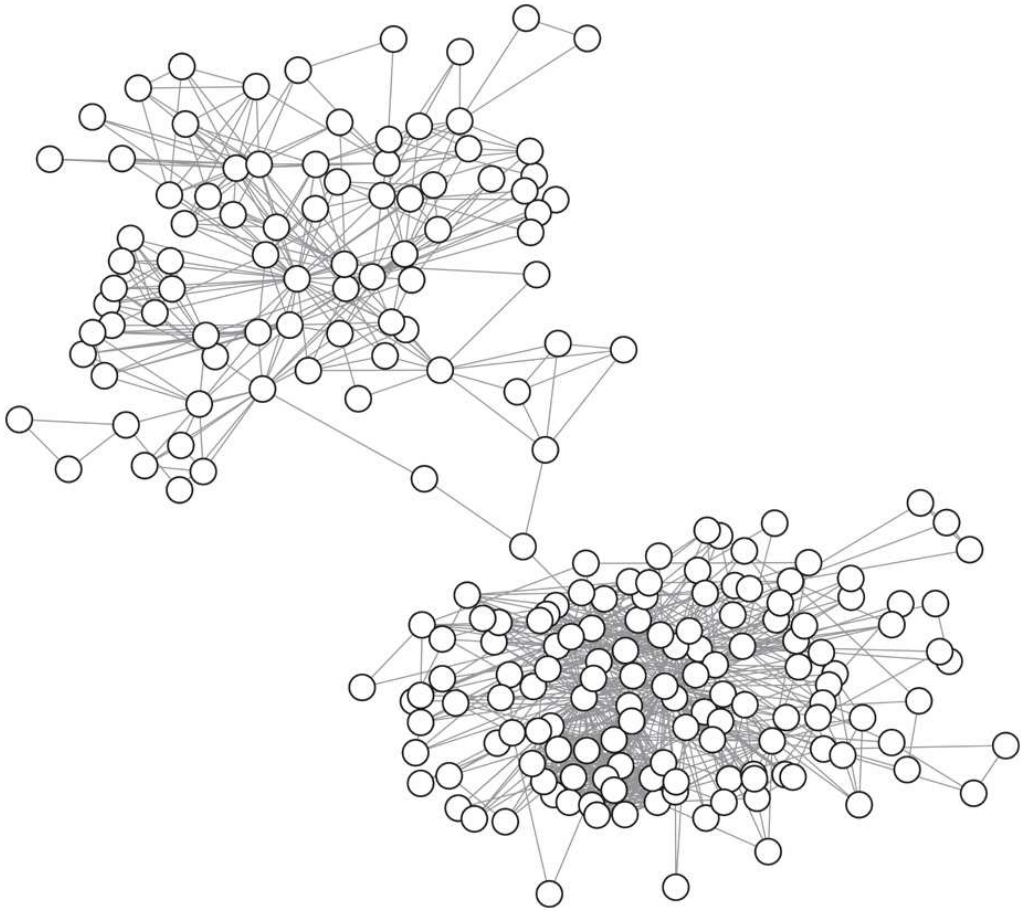
- A case study: *sf* in *del.ici.ous*
  - ***sf*** is a popular tag in delicious (427 URLs, 19979 users, 5852 triples)
  - ***sf*** is ambiguous (***Science fiction*** or ***San Francisco***?)
  - Are users using the same tag to refer to two different concepts? (Can the users/documents be divided into two groups?)
  - What would be the characteristics of the networks constructed around such ambiguous tag?

Understanding a single tag



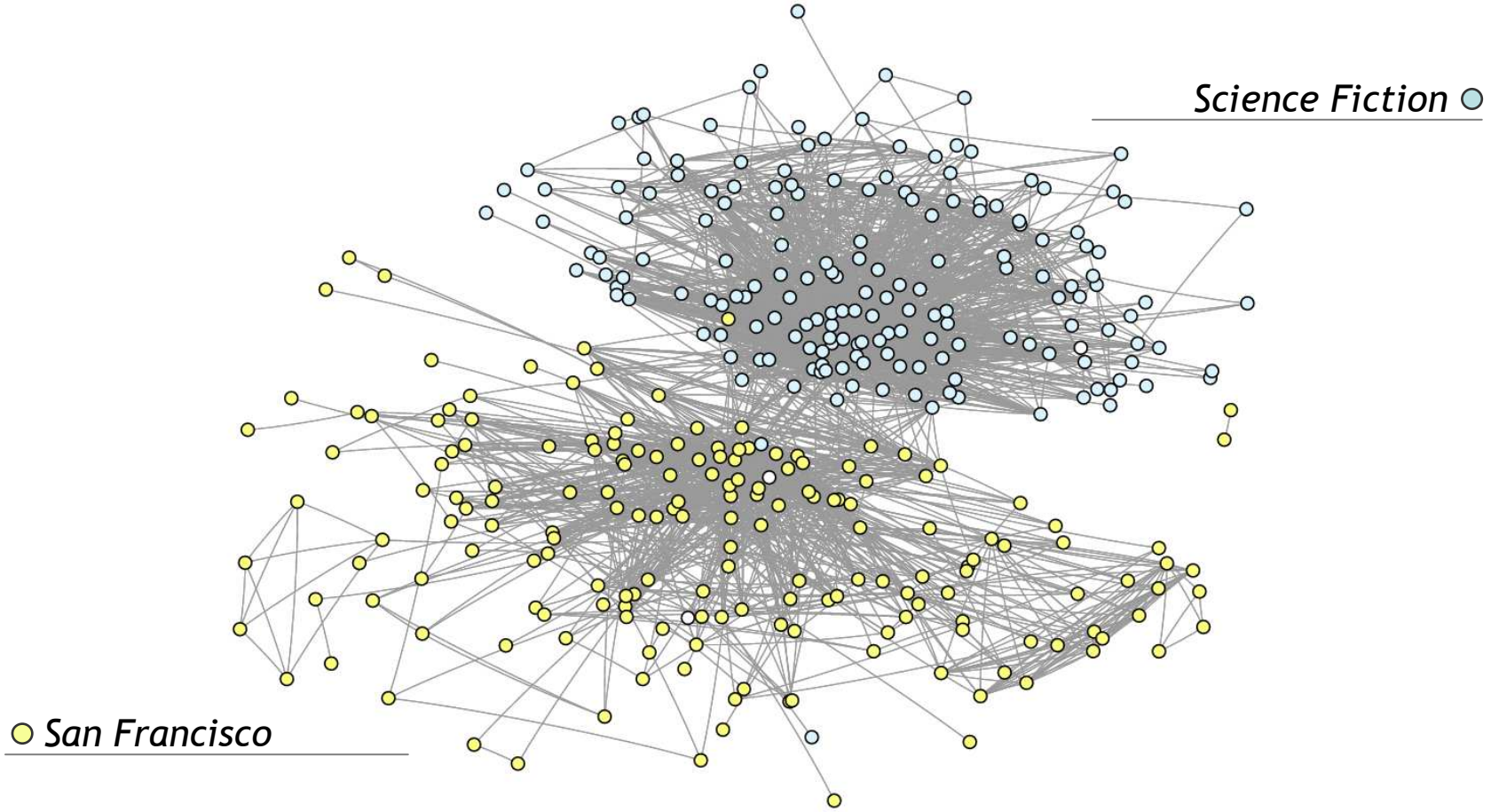
Network of Documents

Understanding a single tag



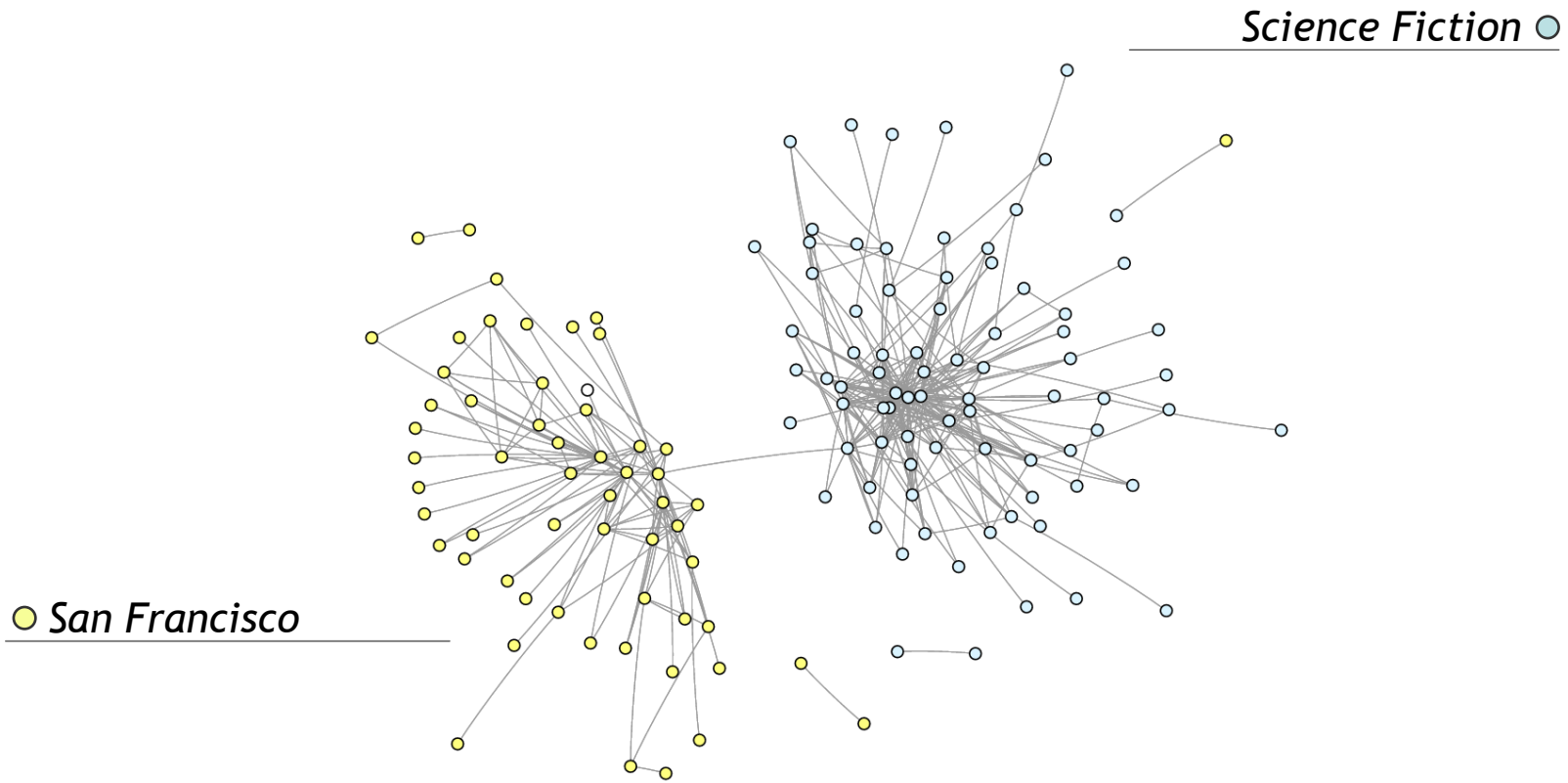
Network of Users

Understanding a single tag



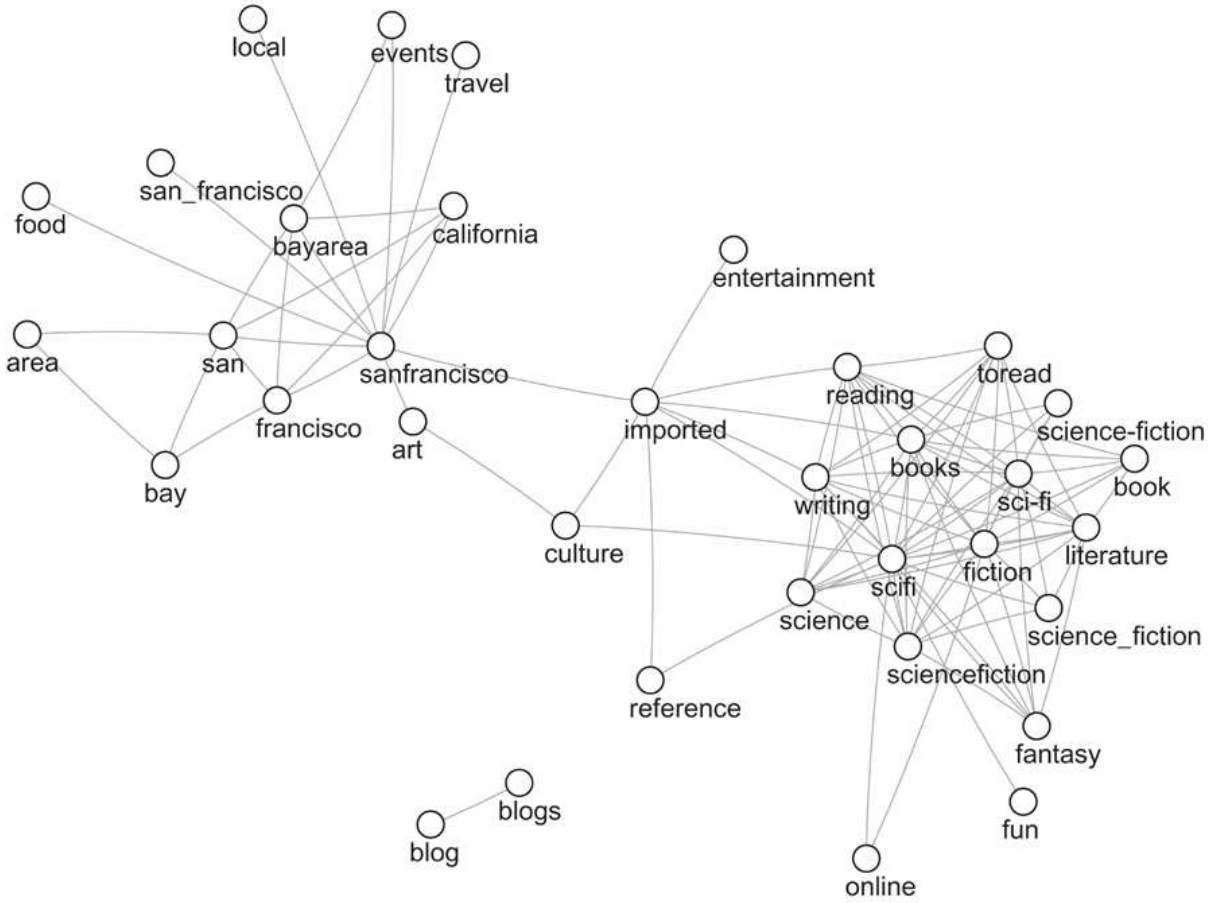
Network of Documents (Classified)

Understanding a single tag



Network of Documents (Removing edges with w < 2)

Understanding a single tag



Network of Tags (35 most frequently used)

Discussion

- Users' behaviour: majority of users tend to use the tag to refer to one concept only
- Possibility of automatic tag disambiguation by examining the network topology
- Possibility of identifying sub-topics (e.g. restaurant-related or arts-related under "San Francisco")
- Classification of documents which are not tagged with enough tags

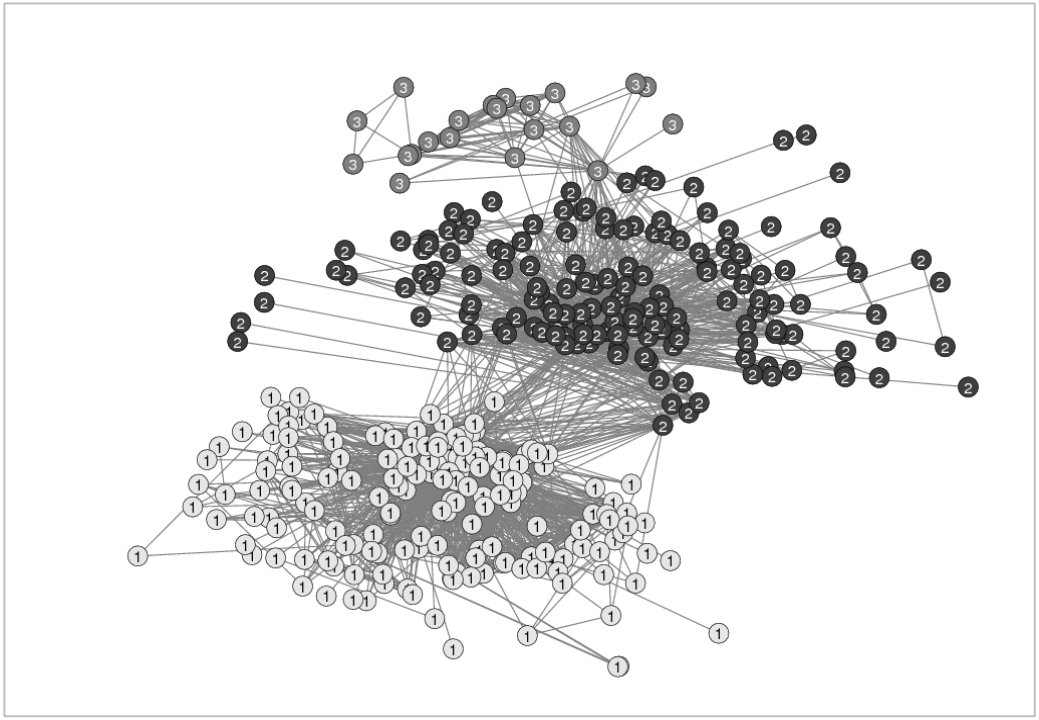
Conclusions and future work

- Conclusions
  - The semantics of a tag can be understood by studying the associations between users and documents
  - Automatic tag disambiguation is possible by exploring the topology of networks of users and documents around a tag
- Future Work
  - Develop automatic algorithms for tag disambiguation
  - Look for an appropriate representation for tag meanings
  - Apply similar techniques on a user or a document
    (e.g. to understand a user's interest/expertise; to study the social network and annotations of a document)

Recent development

- Applying community-discovery algorithms on the networks (e.g. modularity optimization [Newman & Girvan 2004])
- Attempt to break down the networks into communities (clusters of documents with similar contents/tags)
- Extract the most frequently used tags from each cluster
- Automatic tag meaning disambiguation
- A few case studies (Published in WI-IAT'07 [Au Yeung et al. 2007])

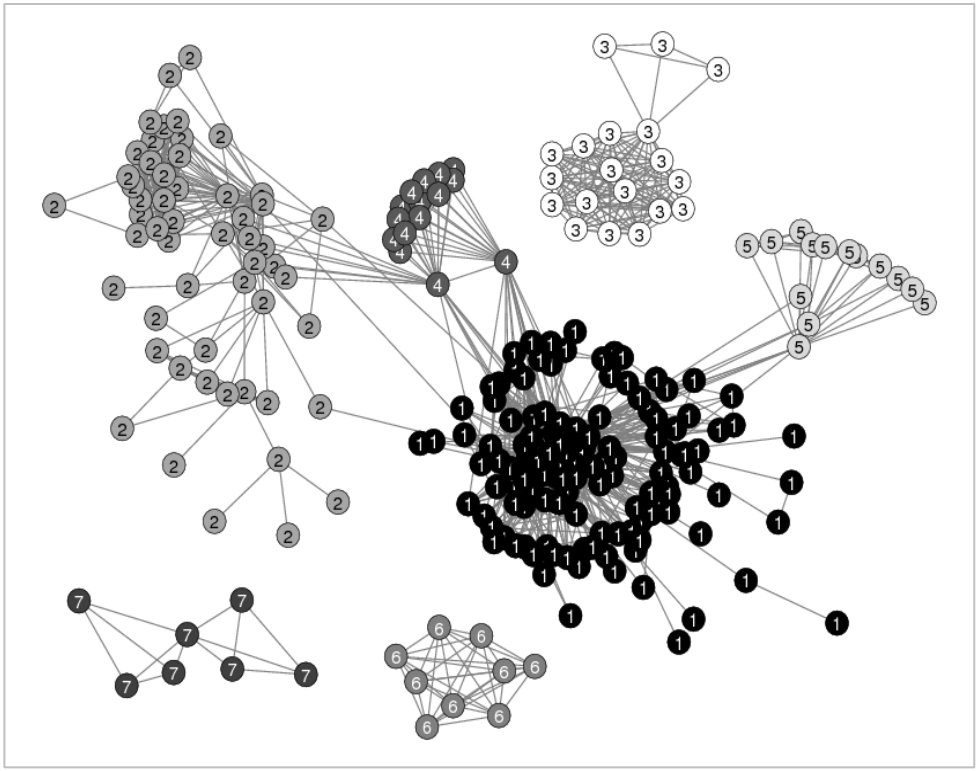
Recent development



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

Automatic disambiguation of the tag **"sf"**

Recent development



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

Automatic disambiguation of the tag **"tube"**

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. Tag meaning disambiguation through analysis of tripartite structure of folksonomies. In *Proceedings of 2007 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligence Agent Technology - Workshops*, Silicon Valley, California, USA, 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. 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.