

# cTag

## Semantic Contextualisation of Social Tags

Ignacio Fernández-Tobías, Iván Cantador, Alejandro Bellogín

ign.fernandez01@estudiante.uam.es, {ivan.cantador, alejandro.bellogin}@uam.es

Information Retrieval Group

Universidad Autónoma de Madrid, Spain

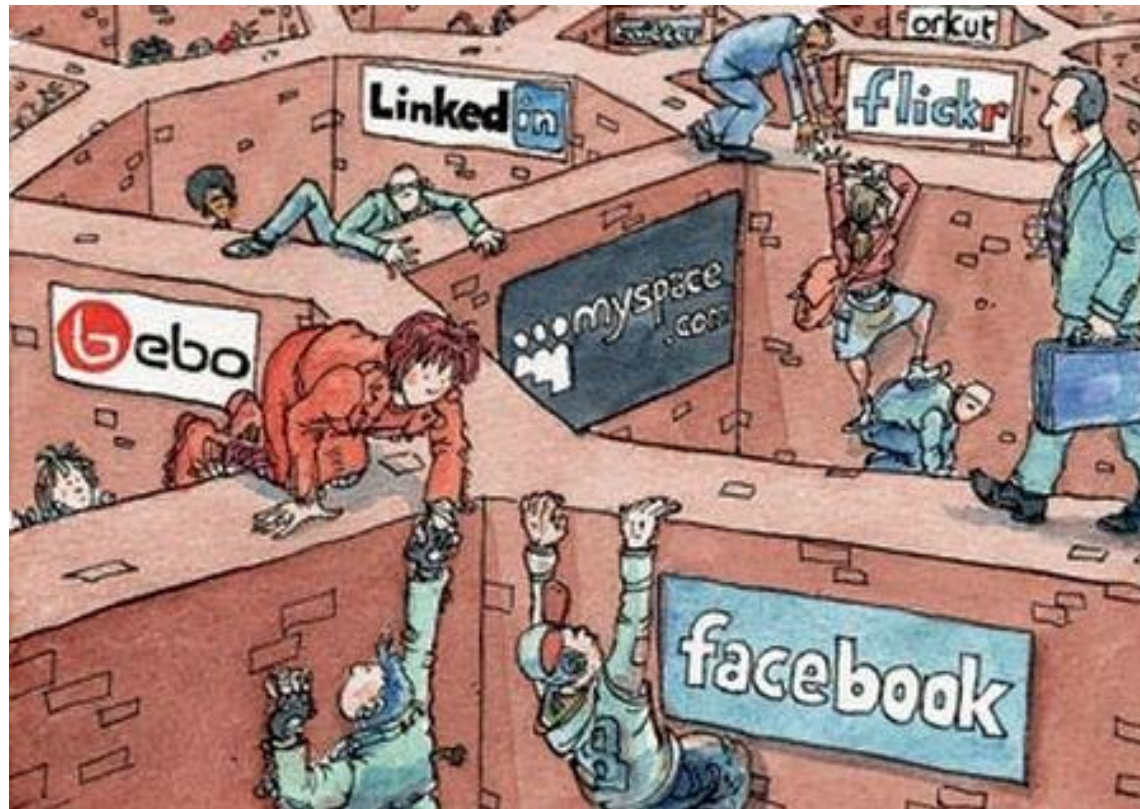
- Web 2.0 and social tagging
- Semantic contextualisation of social tags
- Future work
- cTag

- **Web 2.0 and social tagging**
- Semantic contextualisation of social tags
- Future work
- cTag

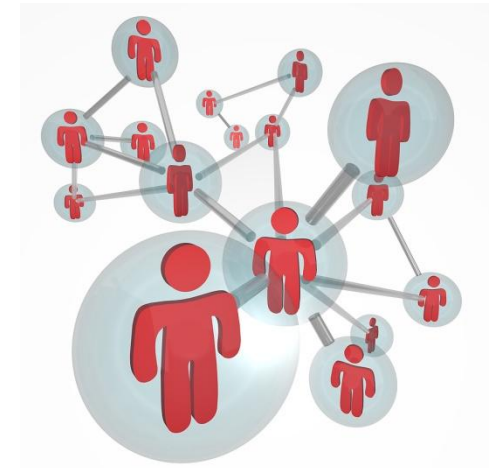
# Web 2.0 and social tagging

3

- Web 2.0 = The Social Web
- Creation of diverse formats of user generated content



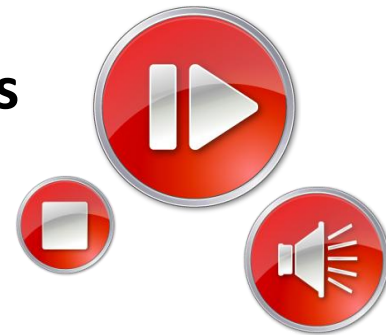
- **Web 2.0 = The Social Web**
  - Creation of diverse formats of **user generated content**
    - People:
      - **Communicate online with contacts**
- Social networks - Facebook, Twitter, LinkedIn, ...*



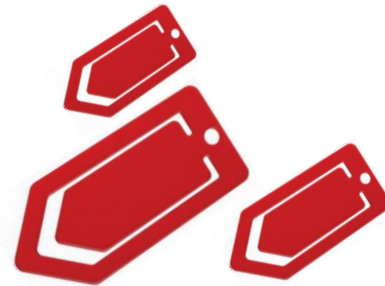
- **Web 2.0 = The Social Web**
- Creation of diverse formats of **user generated content**
  - People:
    - Communicate online with contacts
    - **Contribute to wiki-style repositories**  
*Wikipedia, AboutUs, wikiHow, WikiMapia, ...*



- **Web 2.0 = The Social Web**
- Creation of diverse formats of **user generated content**
  - People:
    - Communicate online with contacts
    - Contribute to wiki-style repositories
    - **Upload or create multimedia objects**  
*YouTube, Flickr, Last.fm, ...*



- **Web 2.0 = The Social Web**
- Creation of diverse formats of **user generated content**
  - People:
    - Communicate online with contacts
    - Contribute to wiki-style repositories
    - Upload or create multimedia objects
    - **Maintain personal bookmarks**  
*Delicious, Digg, CiteULike, ...*





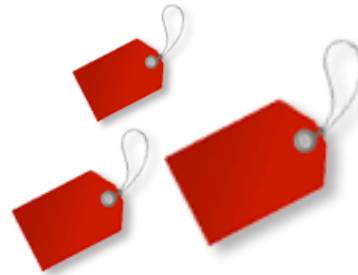
- **Web 2.0 = The Social Web**
- Creation of diverse formats of **user generated content**
  - People:
    - Communicate online with contacts
    - Contribute to wiki-style repositories
    - Upload or create multimedia objects
    - Maintain personal bookmarks
    - **Post comments, reviews and ratings**

*Amazon, Epinions, Blogger, Rotten Tomatoes, ...*

*Blogs and recommender systems*



- **Web 2.0 = The Social Web**
  - Creation of diverse formats of **user generated content**
    - People:
      - Communicate online with contacts
      - Contribute to wiki-style repositories
      - Upload or create multimedia objects
      - Maintain personal bookmarks
      - Post comments, reviews and ratings
      - **Annotate resources**
- Social tagging systems*



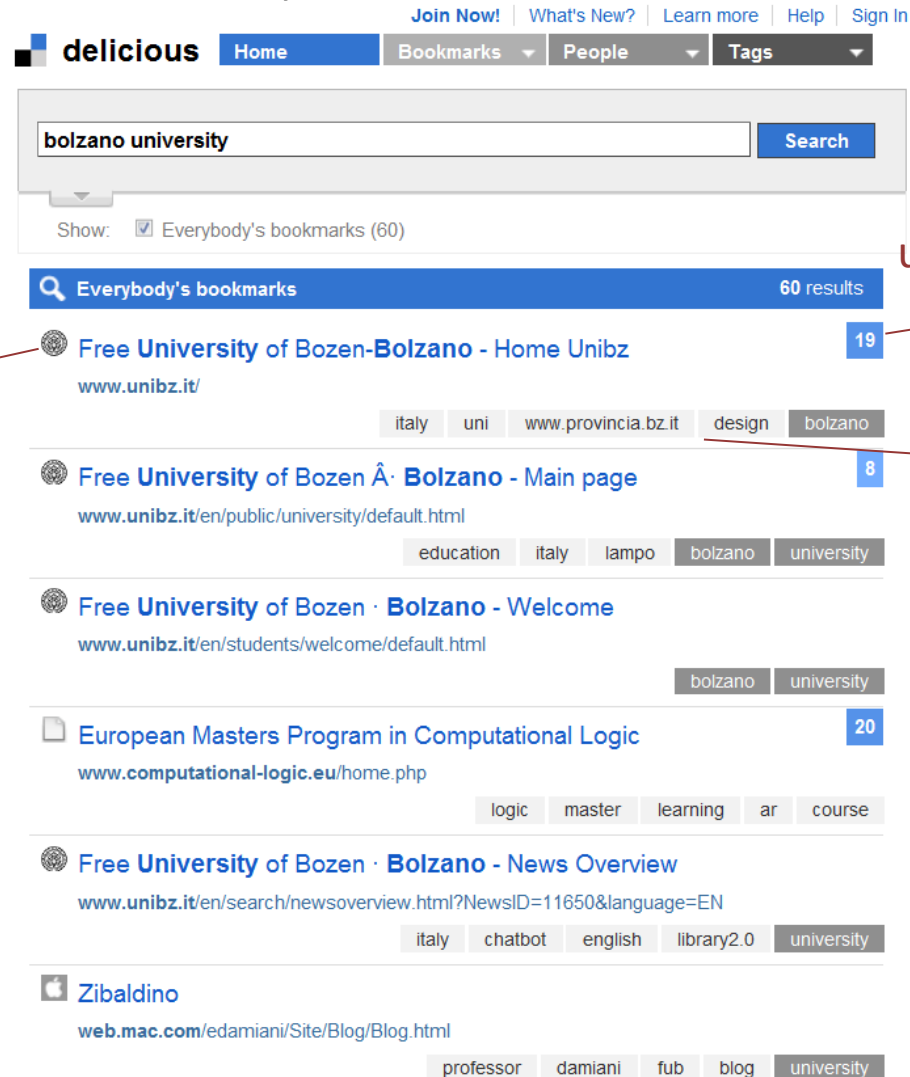
- In social tagging systems, **users** upload or create content (**items**), annotate it with freely chosen words (**tags**), and share these annotations with others
- The nature of tagged items is manifold:
  - *photos* (Flickr)
  - *music tracks* (Last.fm)
  - *video clips* (YouTube)
  - *movies* (MovieLens)
  - *web pages* (Delicious)
  - *scientific articles* (CiteULike, BibSonomy)
  - ...

The Flickr logo, featuring the word "flickr" in a blue and pink sans-serif font.The Last.fm logo, with "last.fm" in red and "the social music revolution" in smaller text below.The YouTube logo, with "You" in black and "Tube" in white inside a red rounded rectangle, with "Broadcast Yourself™" below.The MovieLens logo, with "movie lens" in red and "helping you find the right movies" in smaller text below.The Delicious logo, featuring a blue and black square icon, the word "delicious" in bold, and "social bookmarking" below.The CiteULike logo, with "citeulike" in blue and green, and a small icon of two books.The BibSonomy logo, featuring a network diagram icon and the text "BibSonomy" in blue.

# Web 2.0 and social tagging

11

- An example: Delicious, <http://www.delicious.com>



users who tagged the item

tagged item (web page)

tags assigned to the item

What is the meaning of tag **sf**?

# Web 2.0 and social tagging

14

- What is the meaning of **sf**?



**sf** · sanfrancisco · california · usa ·  
city · bridge · goldengate · sea



**sf** · sciencefiction · fiction · starwars ·  
movie · tatooine · androids · c3po · r2d2

Different meanings!

# Web 2.0 and social tagging

15

- What is the meaning of **web**?



**web** · webbrowser · browser · browsing ·  
opera · explorer · firefox · chrome · safari



**web** · socialweb · web20 · socialmedia ·  
facebook · twitter · flickr · lastfm

Different semantic contexts!



- **Research questions**

- **RQ1:** How to accurately and efficiently identify the **actual meaning or semantic context** of a tag?
- **RQ2:** Is a semantic contextualization of tags useful for folksonomy-based user modelling and **personalisation/recommendation** strategies?



- Web 2.0 and social tagging
- **Semantic contextualisation of social tags**
- Future work
- cTag

# approach

The diagram illustrates a five-step process for semantic contextualization of folksonomy data:

- 1** **folksonomy** (represented by a cloud of people and documents) is processed into **tag-based profiles** (represented by a cylinder database icon).
- 2** **tag-based profiles** are used for **semantic distance computation** (represented by an oval).
- 3** **semantic distance computation** leads to **semantic clustering** (represented by an oval).
- 4** **semantic clustering** results in **semantic tag contexts** (represented by a cluster of circles containing tags).
- 5** **semantic tag contexts** are used for **semantic contextualization** (represented by an oval).

The **semantic contextualization** step is part of a larger process (indicated by a dashed box) that also involves **semantic distance computation** and **semantic clustering**. The final output is **contextualised tag-based profiles** (represented by a cylinder database icon with a person and document icon), which are then used for **personalisation / recommendation** (represented by a cylinder database icon with a gear icon).

- Computing semantic distances between tags
- Clustering tag graphs
- Building semantically contextualised profiles
- Building contextualised tag-based recommenders

- **Computing semantic distances between tags**
- Clustering tag graphs
- Building semantically contextualised profiles
- Building contextualised tag-based recommenders

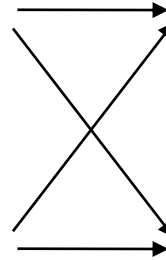
- **Aggregation methods** ( $[\text{user}, \text{item}, \text{tag}] \rightarrow [\text{item}, \text{tag}]$ )
  - Projection
  - Distributional
  - Macro-aggregation
  - Collaborative
- **Similarity measures** ( $[\text{item}, \text{tag}] \rightarrow [\text{tag}, \text{tag}]$ )
  - Matching
  - Overlapping
  - Jaccard
  - Dice
  - Cosine
  - Mutual information

*Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., Stumme, G. 2009. Evaluating Similarity Measures for Emergent Semantics of Social tagging. In Proceedings of the 18th International Conference on World Wide Web (WWW'09), 641-650.*

- Aggregation methods ([user, item, tag] → [item, tag])

Alice	conference	recommender	research
www.umap2011.org	1	1	
www.delicious.com		1	
ir.ii.uam.es		1	1

Bob	conference	recommender	research
www.umap2011.org	1	1	1
www.delicious.com		1	
ir.ii.uam.es			



Projection	conference	recommender	research
www.umap2011.org	1	1	1
www.delicious.com		1	
ir.ii.uam.es		1	1

Distributional	conference	recommender	research
www.umap2011.org	2	2	1
www.delicious.com		2	
ir.ii.uam.es		1	1

- Similarity measures ([item, tag] → [tag, tag])

Overlap similarity

$$\text{sim}(t_1, t_2) = \sum_{r \in \mathcal{R}} w_{t_1} \cdot w_{t_2} = |\mathcal{R}_{t_1} \cap \mathcal{R}_{t_2}|$$

Jaccard similarity

$$\text{sim}(t_1, t_2) = \frac{|\mathcal{R}_{t_1} \cap \mathcal{R}_{t_2}|}{\min(|\mathcal{R}_{t_1}|, |\mathcal{R}_{t_2}|)}$$

Dice similarity

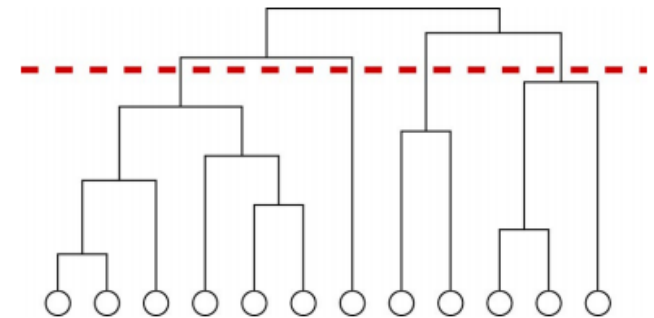
$$\text{sim}(t_1, t_2) = \frac{|\mathcal{R}_{t_1} \cap \mathcal{R}_{t_2}|}{|\mathcal{R}_{t_1} \cup \mathcal{R}_{t_2}|}$$

Cosine similarity

$$\text{sim}(t_1, t_2) = \frac{\mathcal{R}_{t_1}}{\sqrt{|\mathcal{R}_{t_1}|}} \cdot \frac{\mathcal{R}_{t_2}}{\sqrt{|\mathcal{R}_{t_2}|}} = \frac{|\mathcal{R}_{t_1} \cap \mathcal{R}_{t_2}|}{\sqrt{|\mathcal{R}_{t_1}| \cdot |\mathcal{R}_{t_2}|}}$$

- Computing semantic distances between tags
- **Clustering tag graphs**
- Building semantically contextualised profiles
- Building contextualised tag-based recommenders

- Compute the semantic distance between each pair of tags
- Create the whole tag **graph**  $G$  based on the above distances
- For each tag  $t$ , extract a **subgraph**  $G_t$  from  $G$  with the tags “most” similar to  $t$
- **Clustering**  $G_t$  by using the algorithm proposed by Newman and Girvan, 2004
  - Fast computation
  - Automatic clustering stop criterion



Newman, M. E. J., and Girvan, M. 2004. Finding and Evaluating Community Structure in Networks. *Physical Review, E* 69, 026113

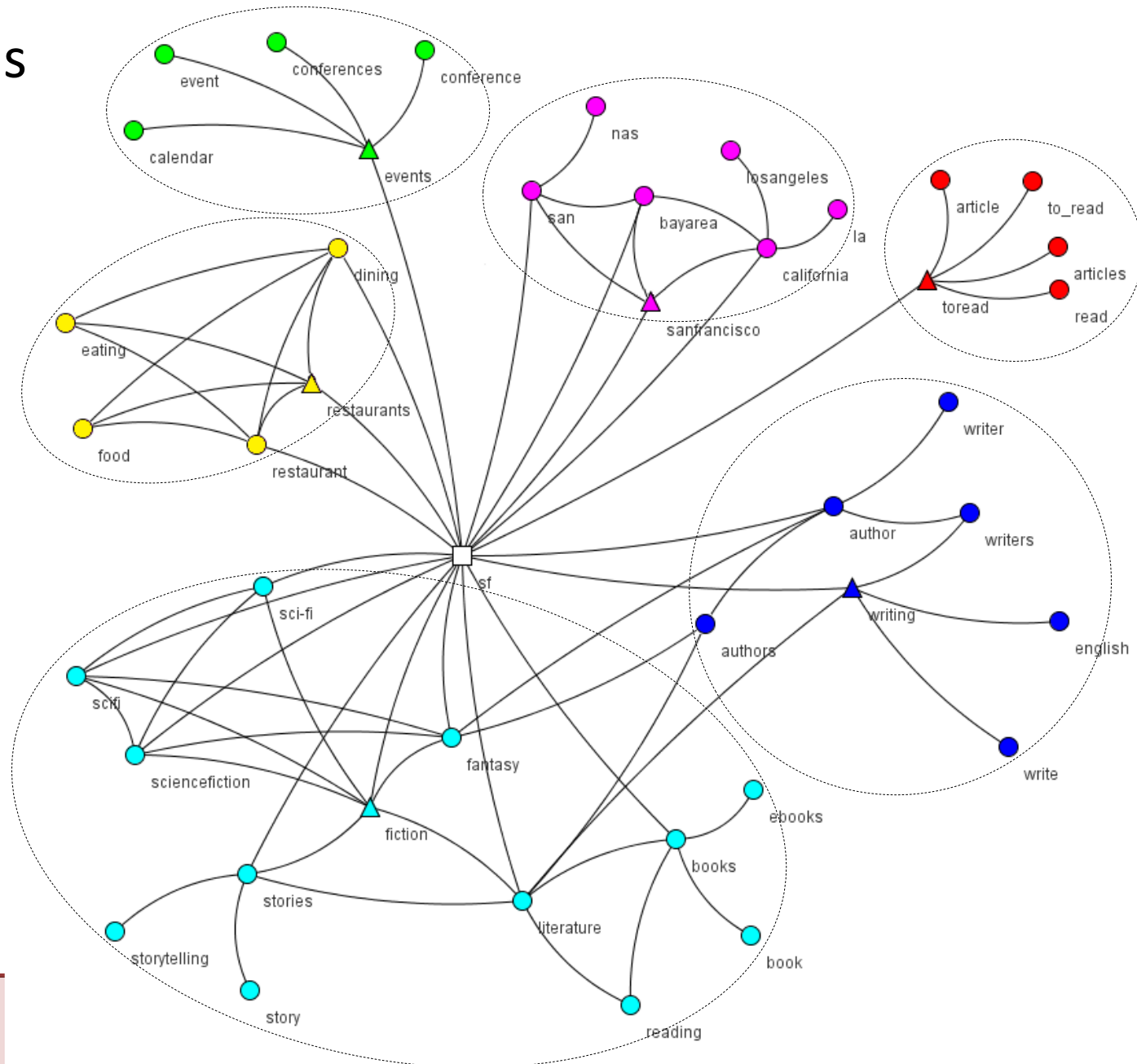
Newman, M. E. J. 2006. Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Physical Review, E* 74, 036104.



# Semantic contextualisation of social tags

25

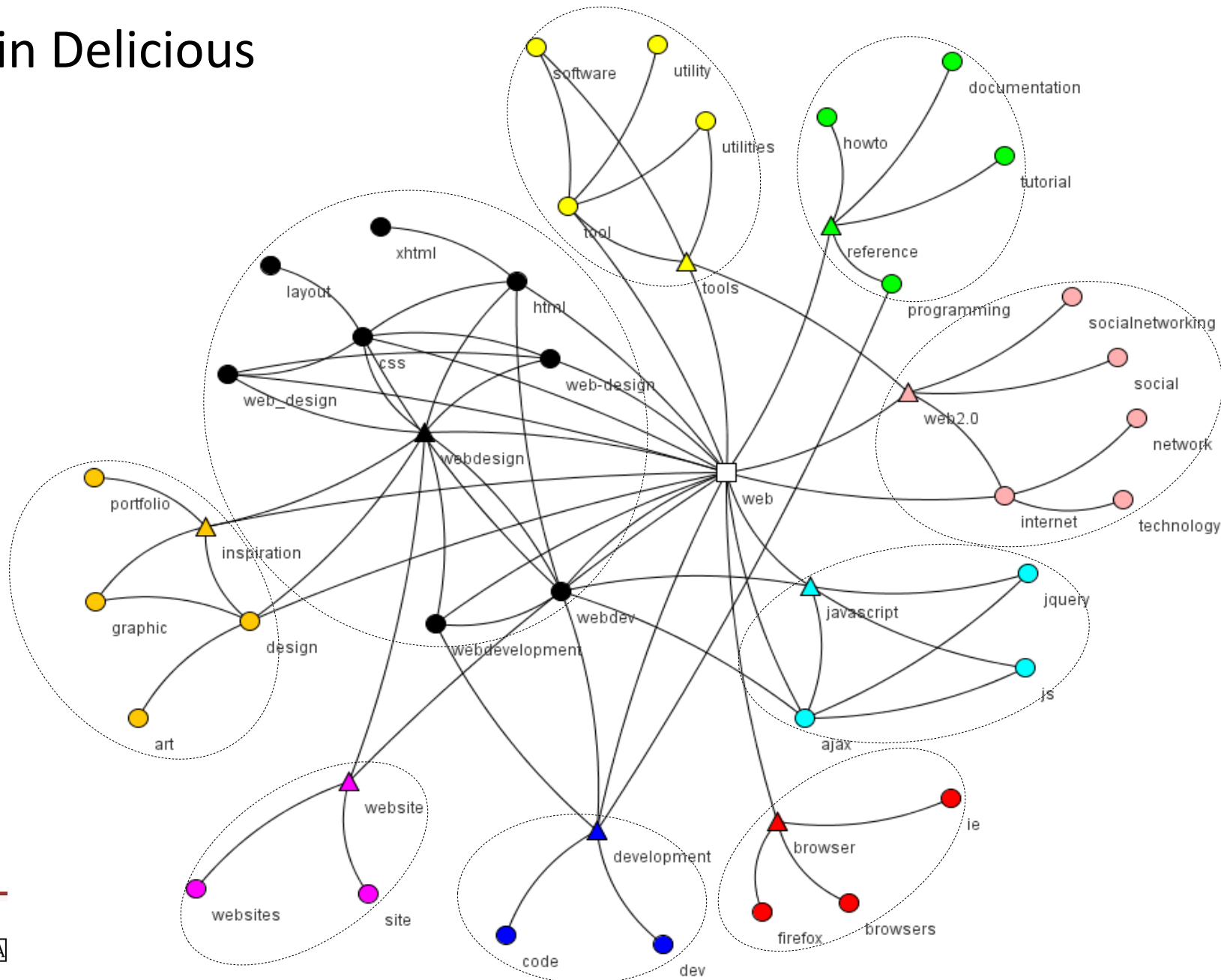
- **sf** in Delicious



# Semantic contextualisation of social tags

26

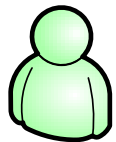
- **web** in Delicious



- Computing semantic distances between tags
- Clustering tag graphs
- **Building semantically contextualised tag-based profiles**
- Building contextualised tag-based recommenders

- Profile models

$$\forall (u, t, i) \in \mathcal{A}, \quad c_{t,u} = c(t, u) = \arg \max_{c_t} \cos(\mathbf{c}_t, \mathbf{u}) \Rightarrow t^u = t \cup c_{t,u}$$
$$c_{t,i} = c(t, i) = \arg \max_{c_t} \cos(\mathbf{c}_t, \mathbf{i}) \Rightarrow t^i = t \cup c_{t,i}$$



Basic

tag	weight
java	0.29
j2ee	0.17
web	0.42
tag	0.12



Contextualised

tag	tag context	weight
java	programming	0.29
j2ee	java	0.17
web	socialweb	0.42
tag	socialweb	0.12

Tag: java

Contextualised tag: java|programming

- Results - **user study**

- Delicious dataset: 2K users, 146.4K items, 54.6K tags
- Contextualisation of 2.9K popular tags (5.3% of the tags; 80.6% of the tag assignments)
- 30 subjects, 1,080 manual assessments (78 tags in different contexts)
- Subject agreement: kappa coefficient = 0.636
- Accuracy of the top context choices
  - Wrong context assignments: 8.2%
  - Correct context assignments for the top choice: 63.8%
  - Correct context assignments within the two top choices: 81.1%
  - Correct context assignments within the three top choice: 88.4%
  - ...

- Computing semantic distances between tags
- Clustering tag graphs
- Building semantically contextualised profiles
- Building contextualised tag-based recommenders

- (Some) Recommendation algorithms

- TF cosine-based similarity

$$g(u_m, i_n) = \cos_{tf}(u_m, i_n) = \frac{\sum_l tf_{u_m}(t_l) \cdot tf_{i_n}(t_l)}{\sqrt{\sum_l (tf_{u_m}(t_l))^2} \cdot \sqrt{\sum_l (tf_{i_n}(t_l))^2}}$$

- TF-IDF cosine-based similarity

$$g(u_m, i_n) = \cos_{tf-idf}(u_m, i_n) = \frac{\sum_l tf_{u_m}(t_l) \cdot iuf(t_l) \cdot tf_{i_n}(t_l) \cdot iif(t_l)}{\sqrt{\sum_l (tf_{u_m}(t_l) \cdot iuf(t_l))^2} \cdot \sqrt{\sum_l (tf_{i_n}(t_l) \cdot iif(t_l))^2}}$$

- BM25 cosine-based similarity

$$g(u_m, i_n) = \cos_{bm25}(u_m, i_n) = \frac{\sum_l (bm25_{u_m}(t_l) \cdot bm25_{i_n}(t_l))}{\sqrt{\sum_l (bm25_{u_m}(t_l))^2} \cdot \sqrt{\sum_l (bm25_{i_n}(t_l))^2}}$$

Cantador, I., Bellogín, A., Vallet, D. 2010. Content-based Recommendation in Social Tagging Systems. In Proceedings of the 4th ACM Conference on Recommender Systems (RecSys'10), 237-240.

- Results on a Delicious dataset - **offline evaluation**
  - Delicious dataset: 2K users, 146.4K items, 54.6K tags
  - Contextualisation of 2.9K popular tags (5.3% of the tags; 80.6% of the tag assignments)
  - Tags for training - test: 80% - 20%
  - 5 fold cross validation

	P@5	P@10	P@20	MAP	R@5	R@10	R@20	NDCG
tf	0.073	0.056	0.041	0.023	0.024	0.036	0.054	0.061
tfidf	0.135	0.103	0.074	0.044	0.044	0.067	0.096	0.113
bm25	0.149	0.109	0.077	0.048	0.048	0.071	0.100	0.121
tf*	0.093	0.069	0.049	0.029	0.030	0.045	0.064	0.077
tfidf*	0.162	0.117	0.083	0.052	0.053	0.076	0.107	0.131
bm25*	<b>0.171</b>	<b>0.123</b>	<b>0.085</b>	<b>0.069</b>	<b>0.055</b>	<b>0.080</b>	<b>0.109</b>	<b>0.136</b>
tf*	27.20%	23.18%	18.54%	23.77%	28.40%	23.98%	19.25%	24.81%
tfidf*	19.68%	14.49%	12.15%	18.07%	19.37%	14.18%	11.62%	18.07%
bm25*	15.25%	13.09%	9.85%	16.97%	15.09%	12.57%	9.13%	12.64%

Improvements in contextualised methods (those marked with \*)



- Web 2.0 and social tagging
- Semantic contextualisation of social tags
- **Future work**
- cTag

- Exploring alternative contextualisation models
- Evaluating other clustering techniques
- Using larger datasets from different folksonomies

- Web 2.0 and social tagging
- Semantic contextualisation of social tags
- Future work
- **cTag**

<http://ir.ii.uam.es/reshet/results.html>

