The 2008 IEEE/WIC/ACM International Conference on Web Intelligence (WI'08)
International Workshop on Web Personalization, Reputation and
Recommender Systems (WPRRS 2008)

Discovering and Modelling

Multiple Interests of Users in Collaborative Tagging Systems

by Ching-man Au Yeung, Nicholas Gibbins, Nigel Shadbolt

Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science
University of Southampton

9 Dec 2008

User Interests and User Profiles

- User profiling is important in enabling recommendation systems
- (1) Collecting user data
  (2) Representing user interests
- Collaborative tagging systems provide rich information about user interests
- How can we make use of this information?

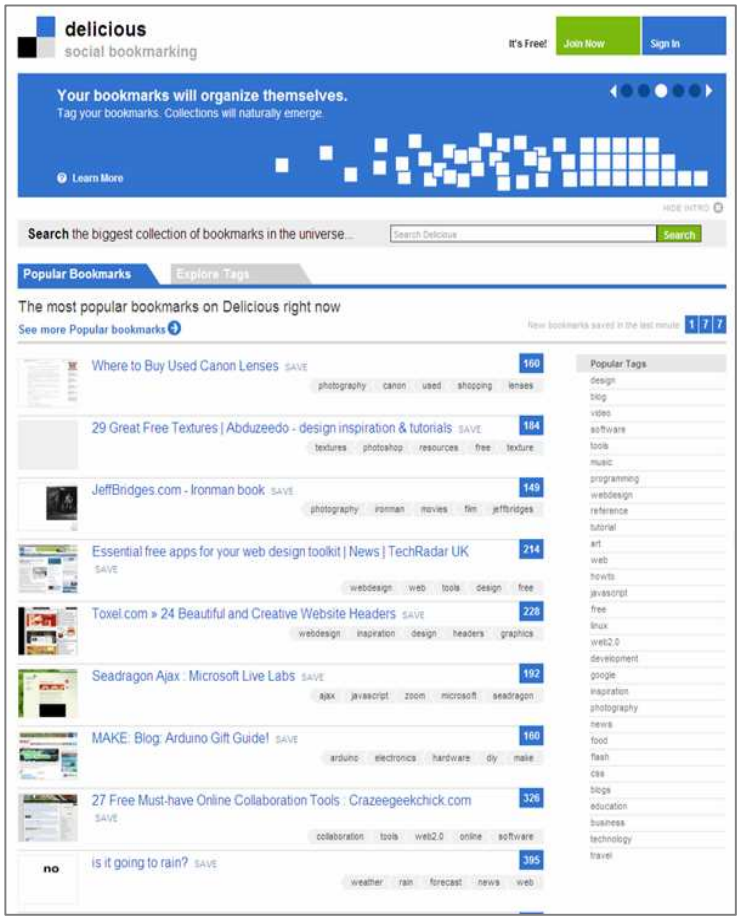
Folksonomies and Personomies

Folksonomies: products of collaborative tagging systems

$F = (U, T, D, A)$
U: User; T: Tag; D: Document;
$A = U \times T \times D$: Annotation

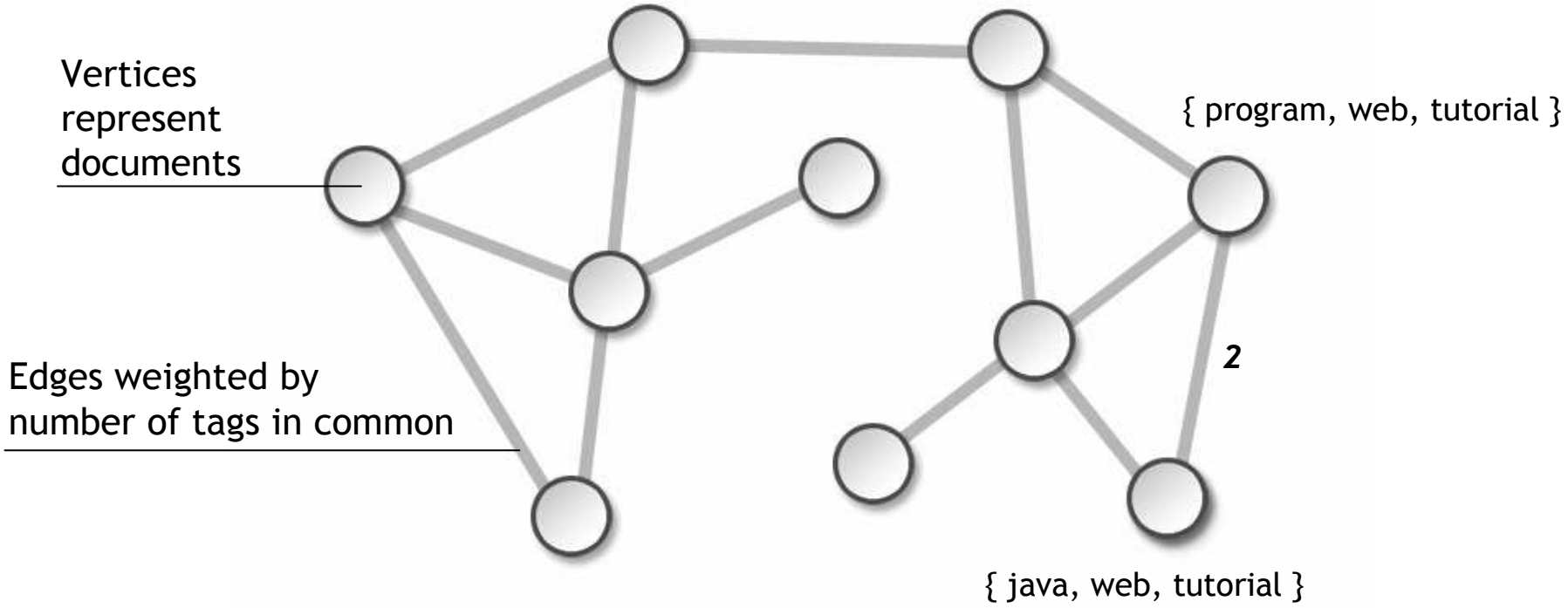
Personomies: data obtained by restricting F to a single user

$P_u = (T_u, D_u, A_u)$
$A_u = \{ (t,d) \mid (u,t,d) \in A \}$



Folksonomies and Personomies

- Personomies can be represented as a network



Analysis of Personomies

Objectives

- To understand the characteristics of personomies
- To study the diversity of user interests

Data

- 9,815 users from Delicious
  423 users from Bibsonomy
  8531 users from LibraryThing
- All associated tags and documents collected

Measuring Diversity of User Interests

Tag Utilisation

Measures the proportion of bookmarks a tag is assigned to. If tags are used on only a small portion of bookmarks, user interest should be more diversity.

Average Tag Co-occurrence Ratio

Measures the extent to which tags co-occur with each other. If tags are more frequently used with each other, they are more related. This also reflects the diversity of user interests.

Analysis of Personomies

Tag Utilisation

$$TagUtil(u) = \frac{1}{|T_u|} \sum_{t \in T_u} \frac{|D_{u,t}|}{|D_u|}$$

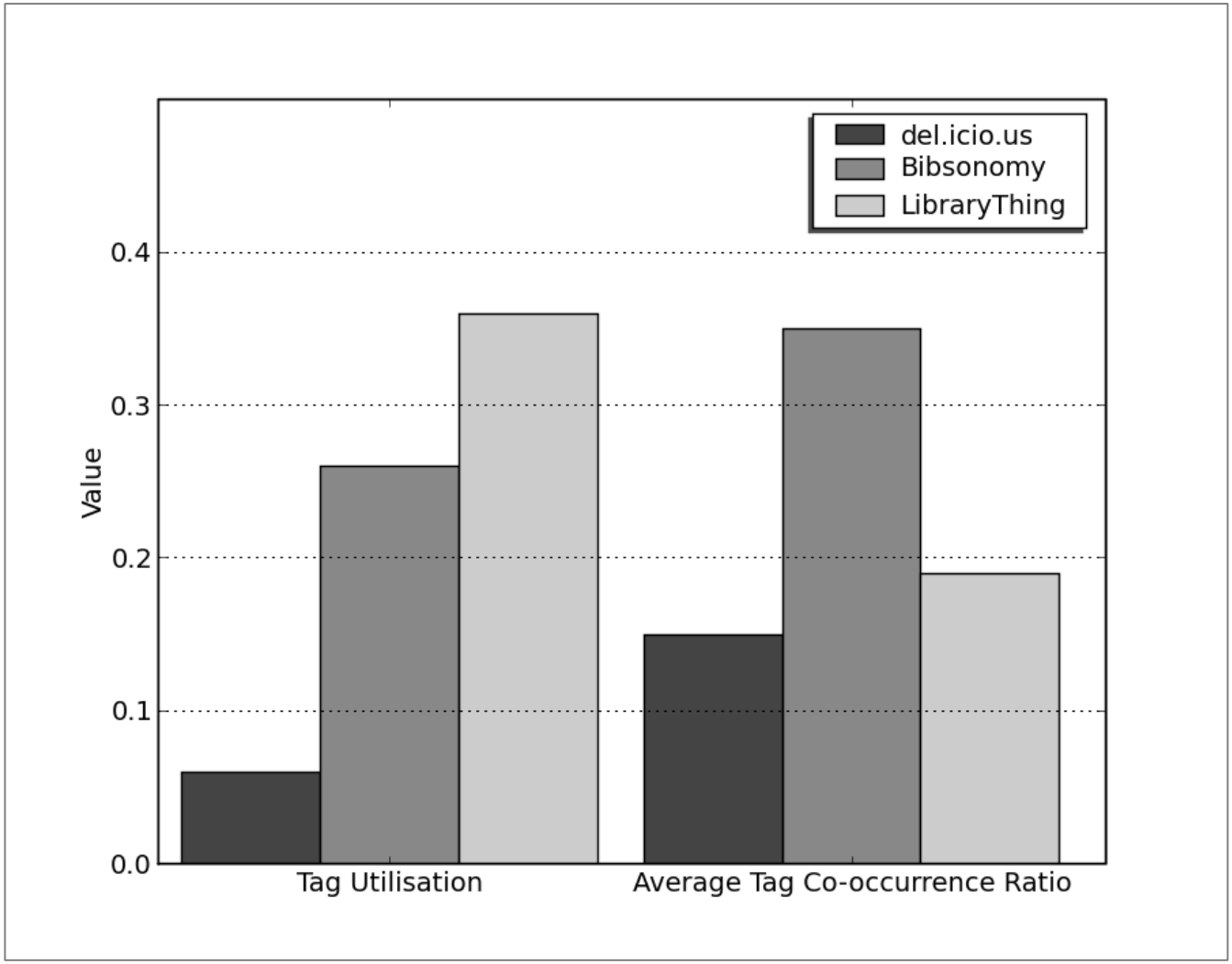
Average Tag Co-occurrence Ratio

$$Avg\_Tag\_Co(u) = \sum_{t_i, t_j \in T_u, t_i \neq t_j} \frac{Co(t_i, t_j)}{2 \times C_2^{|T_u|}}$$

- Example

User	BM	Tags	Measures
u_1	d_1	web2.0, semanticweb, ontology, notes	$TagUtil(u_1)$ = 0.60 $Avg\_Tag\_Co(u_1)$ = 0.80
	d_2	semanticweb, ontology	
	d_3	semanticweb, ontology, rdf	
u_2	d_4	semanticweb, folksonomy, tagging	$TagUtil(u_2)$ = 0.33 $Avg\_Tag\_Co(u_2)$ = 0.27
	d_5	toread, cooking, recipe, food	
	d_6	sports, football, news	

Analysis of Personomies

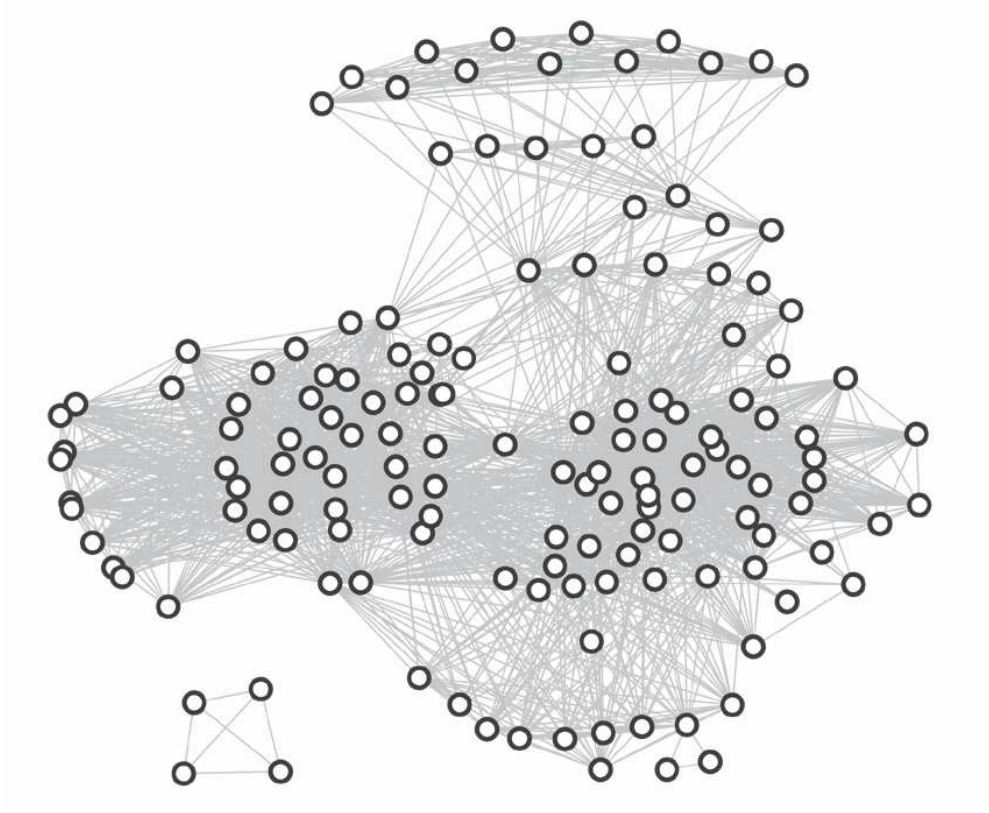


Constructing User Profiles

- User interests are very diverse, single term vectors not able to reflect such diversity
- How can we identify the interests of a user?
- Proposed method:
  1. Perform clustering of bookmarks of a user based on the tags assigned to them
  2. Extract tags from the clusters to represent the interests of the user

User Profiles

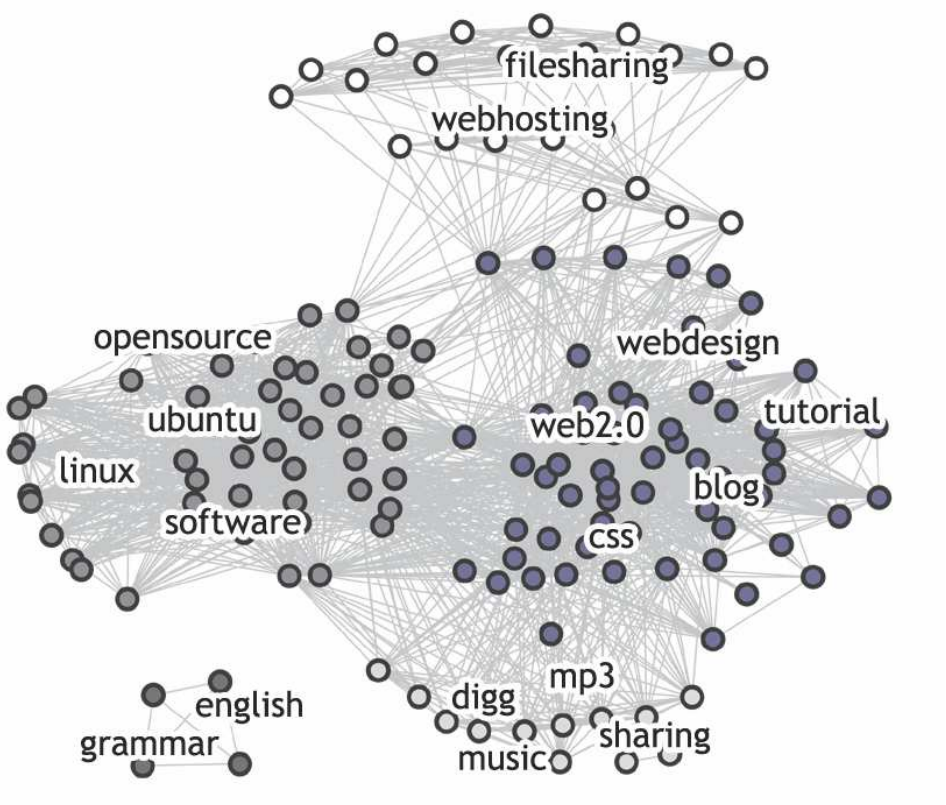
- Represent a personomy as a network of documents



- Nodes represent bookmarks of the user
- Edges are weighted by the number of common tags on the two bookmarks

User Profiles

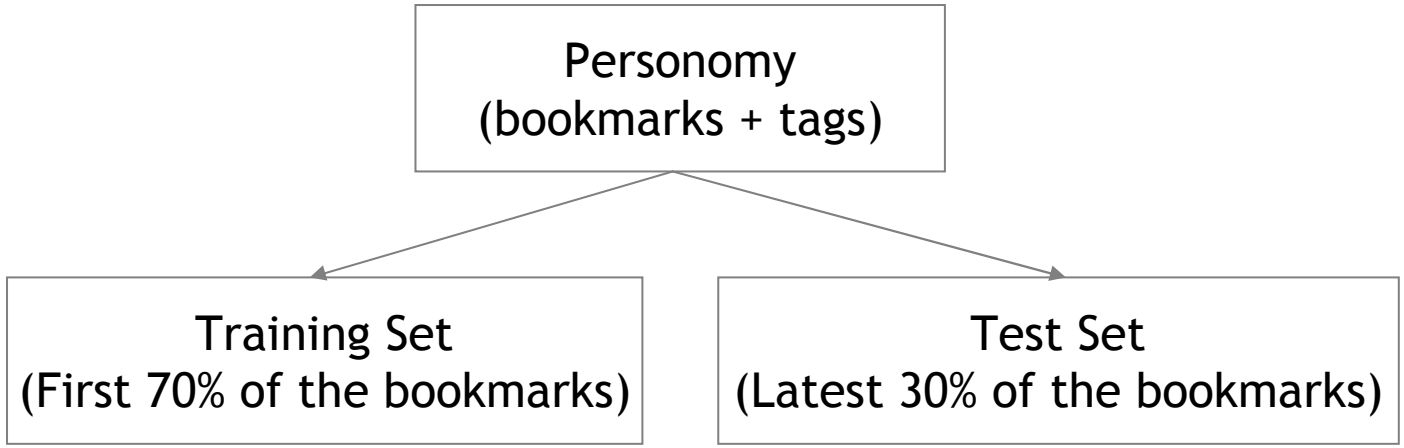
- Clustering by using the fast greedy algorithm for modularity optimisation



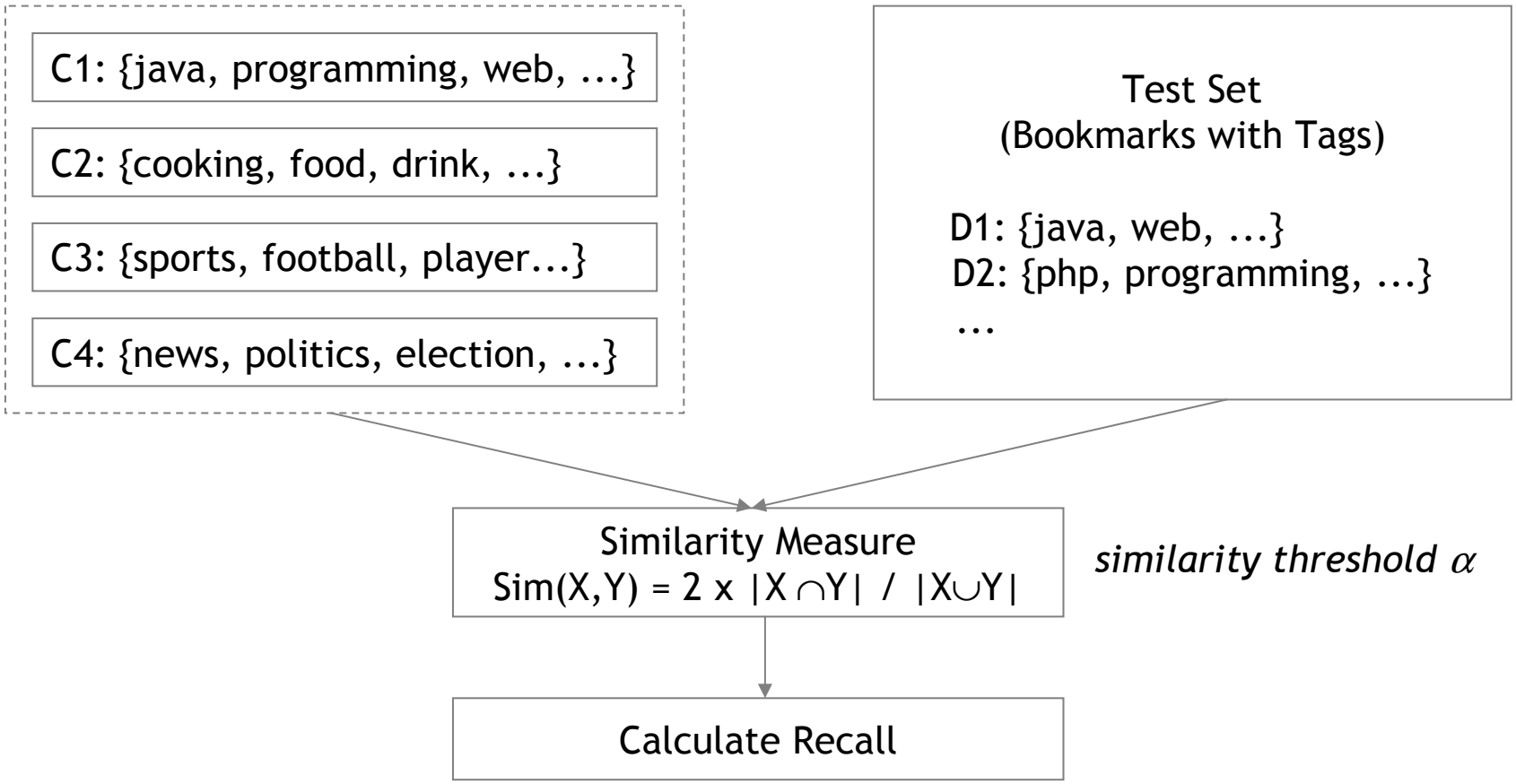
- For each cluster, extract a set of most frequently used tags *(the top f%)*
- The sets of tags act as a *signature* of the cluster
- They also represent the *multiple interests* of the user

Usefulness of the User Profiles

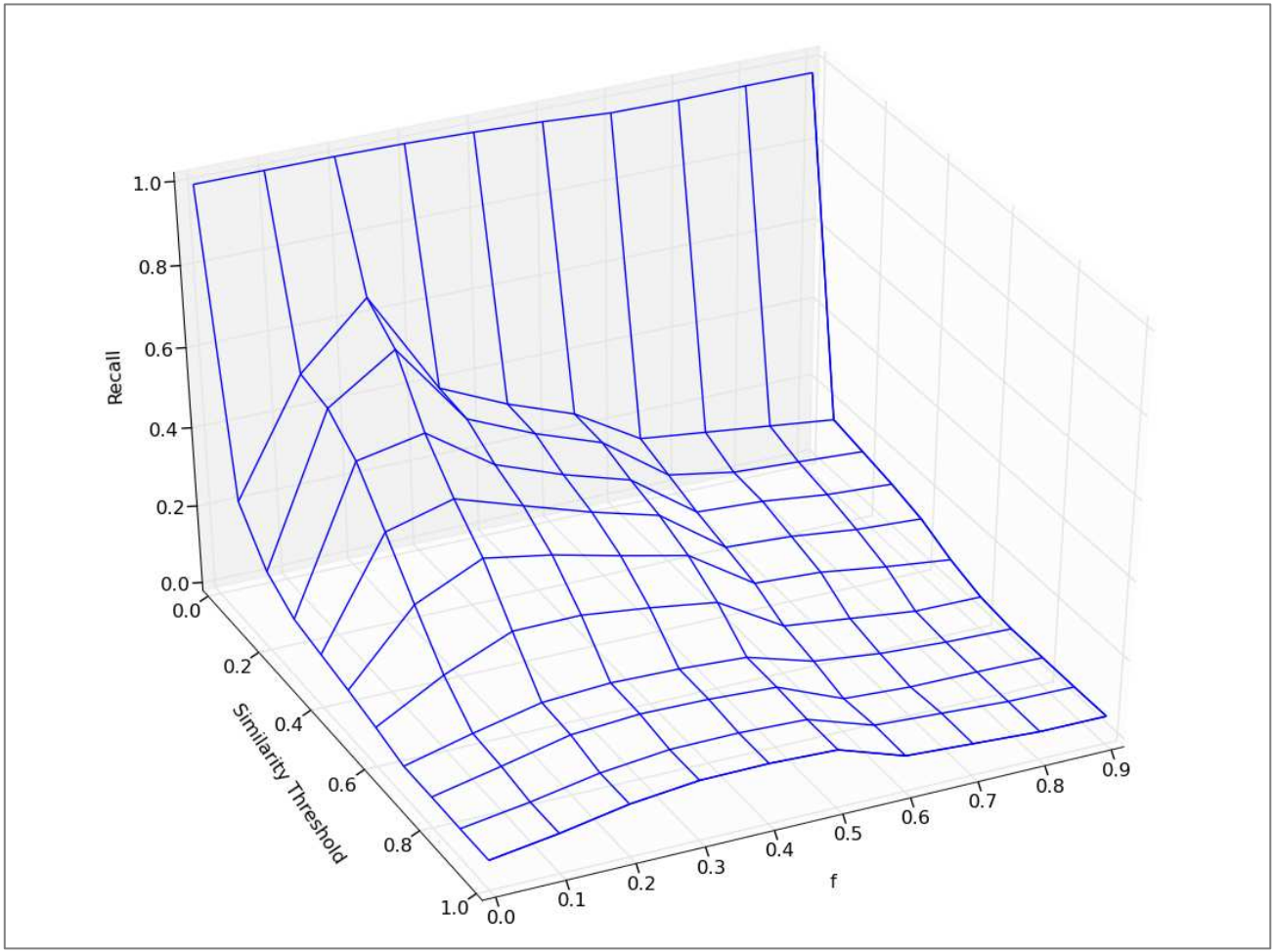
- Study whether the user profiles help the recommendation process
- With the Delicious dataset, we divide the personomy of a user into two sets:



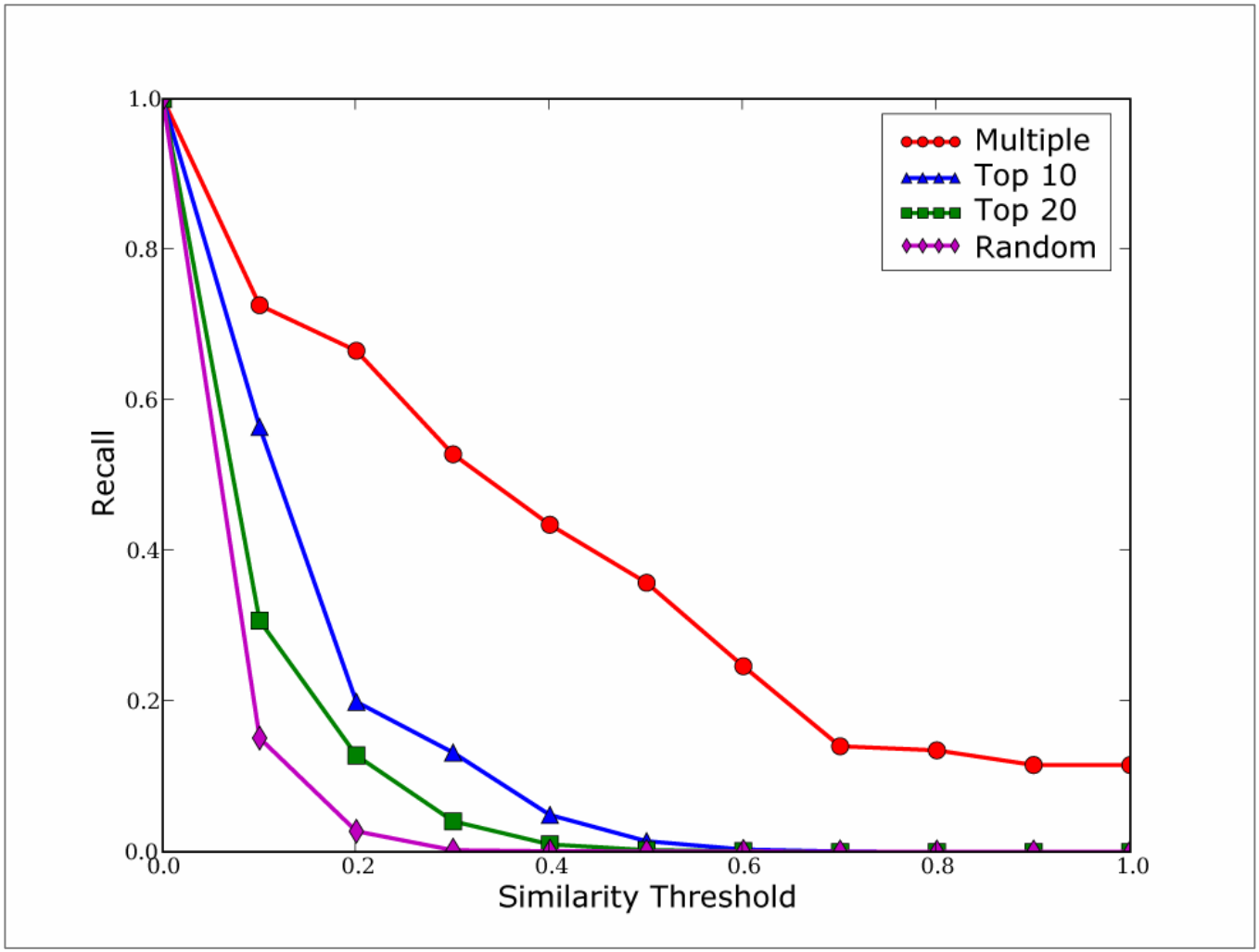
Retrieving Bookmarks



Experiments and Discussions



Experiments and Discussions



Summary

- The profiles generated are able to retrieve more relevant bookmarks at a particular threshold
- User profiles are more fine-grained, allow better judgement of the relevance of a document
- The clustering process produce meaningful clusters which represent the multiple interests of the users

Future Research Directions

- Use other clustering methods
- Introduce weights to the tags, distinguish between less and more important tags
- Study how the generated profiles can be used in a recommendation system

Thank You!

Albert Au Yeung
cmay06r@ecs.soton.ac.uk
http://users.ecs.soton.ac.uk/cmay06r/