

# Piloted Search and Recommendation with Social Tag Cloud-Based Navigation

Cédric Mesnage  
Faculty of Informatics  
University of Lugano  
cedric.mesnage@usi.ch

Mark Carman  
Faculty of Informatics  
University of Lugano  
mark.carman@usi.ch

## ABSTRACT

We investigate the generation of tag clouds using Bayesian models and test the hypothesis that social network information is better than overall popularity for ranking new and relevant information. We propose three tag cloud generation models based on popularity, topics and social structure. We conducted two user evaluations to compare the models for search and recommendation of music with social network data gathered from "Last.fm". Our survey shows that search with tag clouds is not practical whereas recommendation is promising. We report statistical results and compare the performance of the models in generating tag clouds that lead users to discover songs that they liked and were new to them. We find statistically significant evidence at 5% confidence level that the topic and social models outperform the popular model.

## 1. INTRODUCTION

We investigate mechanisms to explore social network information. Our current focus is to use contextual tag clouds as a mean to navigate through the data and control a recommendation system.

Figure 1 shows the screen of the Web application we developed to evaluate our models. The goal is to find the displayed track using the tag cloud. The tag cloud is generated according to a randomly selected model and the current query. Participants in the evaluation can add terms to the query by clicking on tags which generates a new tag cloud and changes the list of results. Once the track is found, the user clicks on its title and goes to the next task.

Figure 2 shows the principle of our controlled recommendation experiment. The participant sees a tag cloud, by clicking a tag she is recommended with a song. Once the song is rated, a new tag cloud is given according to the previously selected tags.

This paper is structured as follows. We first discuss related work in the area of tag cloud-based navigation. We then detail models for generating context-aware tag clouds

WOMRAD 2010 Workshop on Music Recommendation and Discovery, collocated with ACM RecSys 2010 (Barcelona, SPAIN)  
Copyright ©. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

test, Logout  
Found : 0/12

Your goal in this experiment is to find the following given below by selecting tags from the cloud. Each time you change the query by adding or removing a tag from the query (click on X next to the tag in the query box), the list of relevant tracks is updated. Once you find the track in the list, click on the title to go to the next track. If you can not find it, hit "reject the task" which will give you a new track to find.

80s acoustic **alternative** alternative-rock any mix blues breaks brit  
britpop **classic-rock** classical country dance **electronic** electronica  
experimental favorites female-vocalists folk funk garage-rock glam-rock grunge guitar  
**hard-rock** heavy-metal hip-hop hip-hop indie indie-rock industrial instrumental jazz  
mellow **metal pop** pop-rock powerpop progressive progressive-rock psychedelic  
punk punk-rock rap reggae singer-songwriter soul soundtrack techno trip-hop

Query: rock (X)

Fair Weather Friends by Daedelus  
Stairway to Heaven by Led Zeppelin  
Take Me Out by Franz Ferdinand  
Thunder On The Mountain by Bob Dylan  
Even the Losers by Tom Petty  
Hotel California by Eagles  
Madame George by Van Morrison  
Starlight by Muse  
Smells Like Teen Spirit by Nirvana  
Clocks by Coldplay

reject this task

Figure 1: Searching task.

using both social network and topic modeling based approaches, that we have implemented in our prototype tag cloud-based navigation system. We then describe the data we have collected from the "Last.fm" online music social network, and the evaluation consisting of a pilot user-study, a user survey and a follow up study.

## 2. RELATED WORK

### 2.1 Social Tagging and its Motivations

Research in social tagging is relatively recent with the first tagging applications appearing in the late nineties [12]. The system called Webtagger relied on a proxy to enable users to share bookmarks and assign tags to them. The approach was novel compared to storing bookmarks in the browser's folder in the sense that bookmarks were shared and belonged to multiple categories (instead of being placed in a single folder). The creators argued that hierarchical browsing was tedious and frustrating when information is nested several layers deep.

By 2004, social tagging had reached a point where it was becoming more and more popular, initially on bookmarking sites like Delicious and then later on social media sharing sites such as Flickr and Youtube. Research in social tagging started with Hammond [7] who gave an overview of social bookmarking tools and was continued by Golder *et al.* [5] who provided the first analysis of tagging as a process using tag data from Delicious. They showed that tag data follows a power law distribution, gave a taxonomy of tagging incentives, and looked at the convergence of tag descrip-



Figure 2: Controlled recommendation task.

tions over time for resources on Delicious. The paper led to the first workshop on tagging [21], where papers mainly discussed tagging incentives, tagging applications (in museums and enterprises), tag recommendation and knowledge extraction. Following this workshop, research in tagging has spread in various already established areas namely in Web search, social dynamics, the Semantic Web, information retrieval, human computer interaction and data mining.

Sen *et al.* [19] examine factors that influence the way people choose tags and the degree to which community members share a vocabulary. The three factors they focus on are personal tendency, community influence and the tag selection algorithm (used to recommend tags). Their study focuses on the MovieLens system that consists of user reviews for movies. They categorize tags into three categories: factual, subjective and personal. They then divided users of the system into four groups each with a different user interface: the unshared group didn't see any community tags; the shared group saw random tags from their group; the popular group saw the most popular tags; and the recommendation group used a recommendation algorithm (that selected tags most commonly applied to the target movie and to similar movies). They find that habit and investment influence the users' tag applications, while the community influences a user's personal vocabulary. The shared group produced more subjective tags, while the popular and recommendation group produced more factual tags. The authors also conducted a user survey in which they asked users whether they thought tagging was useful for different tasks: self-expression (50%), organizing (44%), learning (23%), finding (27%), and decision support (21%).

Marlow *et al.* [14, 15] define a taxonomy of design aspects of tagging systems that influence the content and usefulness of tags, namely tagging rights (who can tag), tagging support (suggestion algorithms), aggregation model (bag or set), resource type (web pages, images, etc.), source of content (participants, Web, etc.), resource connectivity (linked or not), and social connectivity (linked or not). They also propose aspects of user incentives expressing the different motivations for tagging: future retrieval, contribution and

sharing, attracting attention, playing and competition, self presentation, opinion expression.

Cattuto *et al.* [2, 1] perform an empirical study of tag data from Delicious and find that the distribution of tags over time follows a power law distribution. More specifically they find that the frequency of tags obeys a Zipf's law which is characteristic of self-organized communication systems and is commonly observed in natural language data. They reproduced the phenomenon using a stochastic model, leading to a model of user behavior in collaborative tagging systems.

## 2.2 Browsing with Tags

Fokker *et al.* [4] present a tool to navigate Wikipedia using tag clouds. Their approach enables the user to select different views on the tag cloud, such as recent tags, popular tags, personal tags or friends tags. They display related tags when the user "mouses over" a tag in the cloud. They do not, however, generate new contextually relevant tag clouds when the user clicks on a tag.

In [16], Millen *et al.* investigate browsing behavior in their Dogear social bookmarking application. The application allows users to browse other peoples' bookmark collections by clicking on their username. They find that most browsing activity of the web site is done through exploring peoples' bookmarks and then tags. They compare the 10 most browsed tags with 10 most used tags applied and find that there is a strong correlation. While their findings do not show that tagging improves social navigation in general, they do show that browsing tags helps users to navigate the bookmark collections of others. Following on from this, Ishikawa *et al.* [10] studied the navigation efficiency when browsing other users' bookmarks. The idea is to decide which user to browse first in order to discover faster the desired information. While relevant to tag-based navigation, this study does not deal with the problem of how best to rank tags in order to improve cloud-based navigation in general.

In [13], Li *et al.* propose various algorithms to browse social annotations in a more efficient way. They extract hierarchies from clusters and propose to browse social annotations in a hierarchical manner. They also propose a way to browse tags based on time. As discussed by Keller *et al.* [12] a single taxonomy is not necessarily the best way to navigate a corpus, however.

A more comprehensive study was performed by Sinclair *et al.* [20] to examine the usefulness of tag clouds for information seeking. They asked participants to perform information seeking tasks on a folksonomy like dataset, providing them with an interface consisting of a tag cloud and a search box. The folksonomy was created by the same participants who were asked to tag ten articles at the beginning of the study, leading to a small scale folksonomy. The tag cloud displayed 70 terms in alphabetical order with varying font size proportional to the log of its frequency. The authors give the following equation for the font size:

$$TagSize = 1 + C \frac{\log(f_i - f_{min} + 1)}{\log(f_{max} - f_{min} + 1)} \quad (1)$$

where  $C$  corresponds to the maximum font desired,  $f_i$  to the frequency of the tag to be displayed,  $f_{min}$  and  $f_{max}$  to the minimum and maximum frequencies of the displayed tags. Clicking on a tag in the cloud brings the user to a

new page listing articles annotated with that tag and a new tag cloud of co-occurring tags. Clicking again on a tag restricts the list to the articles tagged with both tags and so on. The search is based on a TF-IDF ranking. Participants were asked 10 questions about the articles and then to tell if they preferred using the search box or the tag cloud and why. They found that the tag cloud performed better when people are asked general questions, for information-seeking, people preferred to use the search box. They conclude that the tag cloud is better for browsing, enhancing serendipity. The participants commented that the search box allows for more specific queries. While similar to our study on tag cloud-based navigation, the work of Sinclair *et al.* [20] differs in a number of important ways: (i) Their aim was to compare tag-based navigation directly with search, while ours is to compare different tag cloud generation methods, based on social network information and topic modeling techniques. (ii) In their study the folksonomy was generated by the participants and is quite small as result, while we rely on an external folksonomy for which scaling becomes an important issue.

In [8], Hassan-Montero *et al.* propose an improvement to tag clouds by ordering the tags according to similarity rather than alphabetically. They use the Jaccard coefficient to measure similarity between tags, which is the ratio between the number of resources in which the two tags both occur and the number in which either one occurs. If  $D(w)$  denotes all resources (documents) annotated with tag (word)  $w$ , then the similarity is given by:

$$RC(w_1, w_2) = \frac{|D(w_1) \cap D(w_2)|}{|D(w_1) \cup D(w_2)|} \quad (2)$$

The authors then define an additional metric to select which tags to display in each cloud (so as to maximize the number of resources “covered by the cloud”). Their method provided, however, little improvement on the coverage of the selected tags. The tag cloud layout is based on the similarity coefficient. The authors also do not provide a user evaluation of the tag cloud generated.

Kaser *et al.* [11] propose a different algorithm for displaying tag clouds. Their methods concern how to produce HTML in various situations. They also give an algorithm to display tags in nested tables. They do not provide an evaluation regarding the usefulness of the new visual representations.

In [18], Sen *et al.* investigate the question tag quality. Tagging systems must often select a subset of available tags to display to users due to limited screen space. Knowing the quality of tags helps in writing a tag selection algorithm. They conduct a study on the MovieLens movie reviews system, adding to the interface different mechanisms for users to rate the quality of tags. All tags can not be rated, therefore they look for ways of predicting tag quality, based on aggregate user behavior, on a user’s own ratings and on aggregate users’ ratings. They find that tag selection methods that normalize by user, such as the numbers of users who applied a tag, perform the best.

In [9], Heymann *et al.* investigate the social tag prediction problem, the purpose of which is to predict future tags for a particular resource. The ability to predict tag applications can lead to various enhancements, such as increased recall, inter-user agreement, tag disambiguation, bootstrapping and system suggestion. They collected tag data from

Delicious and fetched the web pages for each bookmark. They analyze two methods: The first applies only when the bookmarked items are web pages (and not images, songs, videos, etc.). They develop an entropy based metric which measures how much a tag is predictable. They then extract association rules based on tag co-occurrence and give measurements of their interest and confidence. They find that many tags do not contribute substantial additional information beyond page text, anchor text and surrounding hosts. Therefore this extra information are good tag predictors. In the case of using only tags, predictability is related to generality in the sense that the more information is known about a tag (*i.e.* the more popular it is), the more predictable it is. They add that these measures could be used by system designers to improve system suggestion or tag browsing.

Ramage *et al.* [17] compare two methods to cluster web pages using tag data. Their goal is to see whether tagging data can be used to improve web document clustering. This work is based on the *clustering hypothesis* from information retrieval, that “the associations between documents convey information about the relevance of documents to requests”. The document clusters are used to solve the problem of query ambiguity by including different clusters in search results.

All of the above mentioned work differs from our current study of tag cloud-based navigation in the following ways: (i) Previous studies have investigated the usefulness of tag clouds primarily from the basic visualization rather than the navigation standpoint. (ii) Those studies explicitly investigating tag cloud based navigation, have concentrated on simple algorithms for generating tag clouds. (iii) Previous studies investigating more sophisticated algorithms for tag prediction have evaluated those algorithms by assessing prediction accuracy on held-out data rather than “in situ” evaluation with real users for a particular application (tag cloud based navigation).

### 3. TAG CLOUD BASED NAVIGATION

In this section we describe algorithms for generating context-aware tag clouds and query results list for tag cloud based navigation. Generating a tag cloud simply involves selecting the one hundred tags which are the most probable (to be clicked on by the user) given the current context (query). Estimating which terms are most probable depends on the model used as we discuss below.

#### 3.1 Generating Context Aware Tag Clouds

We now investigate three different models for generating context-aware tag clouds. For each model we describe first how an initial context-independent cloud is generated. We then describe how the context dependent cloud is generated in such a way as to take the current query (context) tags into account.

##### 3.1.1 Popularity based Cloud Generation Model

The first and simplest tag cloud generation model is based on the popularity of the tags across all documents in the corpus. We first describe a query independent tag cloud, which can be used as the initial cloud for popularity based navigation.

Ranking tags by popularity on the home page gives users a global access point to the most prolific sections of the portal. The most popular tags are reachable from the popular

tag cloud and displayed with a font size proportional to the amount of activity on that tag. A measure of the popularity of a tag across the corpus is given in the following:

$$p(w) = \frac{\sum_{d \in D} N_{w,d}}{\sum_{d \in D} N_d} \quad (3)$$

where  $N_{w,d}$  is the count of occurrences of tag  $w$  for resource (document)  $d$  and  $N_d = \sum_{w \in V} N_{w,d}$  is the total count for the document.

We can now compute a context sensitive version of the popular tag cloud quite simply as follows:

$$p(w|Q) = \frac{\sum_{d \in D(Q)} N_{w,d}}{\sum_{d \in D(Q)} N_d} \quad (4)$$

Where  $D(Q) = \cup_{w \in Q} D(w)$  is the union of all resources that have been tagged with words from the query  $Q$ .

### 3.1.2 Social Network Structure based Cloud Generation Model

We are interested in taking advantage of additional information contained in the social network of users (friendships) in order to improve the quality of the tag cloud. We assume that the friends of a user are likely to share similar interests and thus we can use the tag description of a user's friends to smooth the tag description of the user.

We calculate an entry (context independent) social tag cloud as follows:

$$p(w) = \sum_{u \in U} \sum_{u' \in f(u)} \frac{N_{w,u'}}{\sum_{w \in W} N_{w,u'}} \quad (5)$$

where  $f(u)$  is the set of friends of user  $u$  and  $U$  denotes the set of all users in the social network.

We apply a slightly different derivation to calculate the context dependent social tag cloud. We estimate the probability  $p(w|w')$  given the context tag  $w'$ . These probabilities are precomputed and combined depending on the query at run time. We hypothesize that users who are friends on a social tagging website are likely to have similar interests (likes & dislikes) and that we can use the social network structure to improve contextual tag cloud generation. We can leverage the social network (by marginalizing out the user  $u$ ) as follows:

$$p(w|w') = \sum_{u \in U} p(w, u|w') \quad (6)$$

$$= \sum_{u \in U} p(w|u) \frac{p(w'|u)p(u)}{p(w')} \quad (7)$$

Calculating  $p(w')$  and  $p(u) = N_u / \sum_{u' \in U} N_{u'}$  is straightforward. We compute  $p(w|u)$  by summing over tag counts  $N_{w,u'}$  for users in the social network of the user  $u$ :

$$p(w|u) = \frac{\sum_{u' \in f(u)} N_{w,u'}}{\sum_{u' \in f(u)} N_{u'}} \quad (8)$$

Note that since the summation in Equation 7 over all users involves a very large computation, we perform the summation only over the top 200 users as ranked according to the frequency  $p(w|u)$ .

### 3.1.3 Topic Model based Cloud Generation Model

Another way to smooth the relative term frequency estimates and thereby improve the quality of the tag clouds generated is to rely on latent topic modeling techniques [6]. Using these techniques we can extract semantic topics representing user tagging behavior (aka user interests) from a matrix of relationships between tags and people. Topic models are term probability distributions over documents (in this case users) that are often used to represent text corpora. We apply a commonly used topic modeling technique called latent Dirichlet allocation (LDA) [6] to extract 100 topics by considering people as documents (and tags as their content).

The entry (context independent) tag cloud based on topic modeling is defined as follows:

$$p(w) = \sum_{z \in Z} p(w|z)p(z) \quad (9)$$

Where  $p(w|z)$  denotes the probability of the tag  $w$  to belong to (being generated by) topic  $z$ , its value is given as an output of the LDA algorithm.  $p(z)$  is the relative frequency of the topic  $z$  across all users in the corpus.

To compute the context aware tag cloud based on topic modeling, we simply marginalize over topics (instead of users):

$$p(w|w') = \sum_{z \in Z} p(w|z)p(z|w') \quad (10)$$

$$= \sum_{z \in Z} \frac{p(w|z)p(w'|z)p(z)}{p(w')} \quad (11)$$

## 3.2 Ranking Resources

We follow a standard Language Modeling [3] approach to ranking resources (documents) according to a query. Thus we rank resources according to the likelihood that they would be generated by the query, namely the probability  $p(d|Q)$ , where  $d$  is a resource and  $Q$  the query as a set of tags. We give here the derivation of  $p(d|Q)$  by applying Bayes' rule.

$$p(d|Q) = \frac{p(Q|d)p(d)}{p(Q)} \quad (12)$$

For ranking we can drop the normalization by  $p(Q)$  as it is the same for each resource  $d$ , which gives us:

$$score(d|Q) = p(Q|d)p(d) \quad (13)$$

We apply the naive Bayes assumption and consider the words in the query to be independent given the document  $d$ . Thus  $p(Q|d)$  factorizes into the product of word probabilities  $p(w|d)$ :

$$score(d|Q) = p(Q|d)p(d) \approx p(d) \prod_{w \in Q} p(w|d) \quad (14)$$

This product is equivalent in terms of ranking to the sum of the corresponding log probabilities. Thus we compute the score for a particular tag as follows :

$$score(d|Q) =_{ranking} \log p(d) + \sum_{w \in Q} \log p(w|d) \quad (15)$$

Computing  $p(d)$  is straightforward, we can either use the length of the tag description of the resource  $d$  or the uniform

distribution  $p(d) = 1/D$  where  $D$  is the count of documents in the corpus.

For the browsing experiment, the log probabilities within the summation are exponentially weighted so as to give preference to the most recently clicked tags, as follows:

$$browsing\_score(d|Q) = \log p(d) + \sum_{i=1}^{|Q|} \alpha^{i-1} \log p(w_i|d) \quad (16)$$

Here  $w_i$  denotes the  $i^{\text{th}}$  most recent term in the query  $Q$ , and  $\alpha$  is a decay parameter set to 0.8 in our experiments.

### 3.3 Precomputation

For each model we precompute the values for  $p(w|w')$  which gives us three matrices of relations between tags. At run time we rank the tags to generate a contextual tag cloud according to a query of multiple tags as follows:

$$p(w|Q) = \beta \log p(w) + \sum_{w' \in Q} \log p(w|w') \quad (17)$$

In our experiments we set the parameter  $\beta$  to 0.5.

## 4. EMPIRICAL SETUP

We choose "Last.fm" to fetch our experimental dataset. "Last.fm" is a music sharing online social network which allows one to get social network data and tagging data from their application programming interface (API). To our knowledge it is the only network which enables researchers to fetch the friends of any user in the system. Fetching the social network is essential for experiments with social tag clouds.

We gather tag data by crawling users via their friend relationships. Once a new user is fetched, we download her own tags and then recursively fetch her friends and so on. We start by fetching the network of the author. In order to get a complete subset of the social network of "Last.fm", we apply a breadth first search by exploring recursively the relations of each user. Once we have a substantial subset of the social network and tags, we fetch the tracks assigned to the tags. For each tag fetched, we get the 50 top tracks annotated with this tag.

Table	Size
People	126035
Friends	3523626
Tags	343681
Tracks	435257
Usages	900259
Tag applications	4236024

Table 1: Dataset size

Table 1 reports the size of the main tables of the database. The database accounts for more than 120 thousand people having 3.5 million friend relationships which makes an average of 27 friends per person. These individuals have used more than 340 thousands unique tags a total of more than 4 million times, which makes an average usage of 12 times per tag. The total number of usages is over 900 thousand which makes an average of 3 people using each tag.

Figure 3 shows the degree distribution of the number of friends. It shows the frequency of people with respect to the number of friends they have. The plot axes are the log of the values for better visualization. The plot shows a power law distribution in the number of friends per person with a number of friends superior to 10. Below ten friends, we have not seen enough data to have a good estimation of the distribution of the number of people with that many friends, so the distribution is curved. Power laws have been observed in other social networks and show that social networks are scale-free. Tag usage also shows a power law distribution.

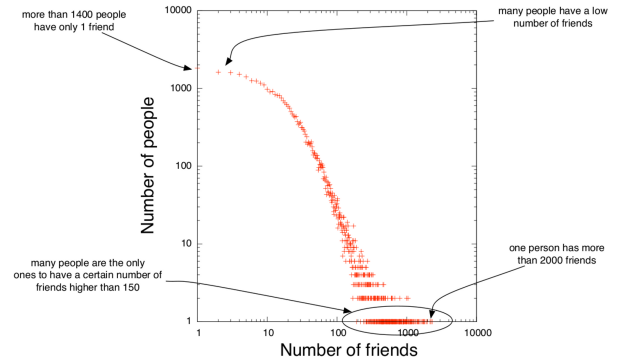


Figure 3: Plot of the distribution of friends.

Once the data is fetched by the ruby scripts via the "Last.fm" Web API, we migrate it to a MySQL database for processing. We precompute various tables to store data that will be used multiple times in the calculations. For instance we compute the term frequency of each tag, the term frequency for each tag and each user, the frequency of the friends of a user for a tag. From these tables we can then compute similarity tables between the probability of one tag given another for each model which corresponds to  $p(w|w')$ , we do this only for the tags used by at least 5 people which accounts for about twenty thousand tags.

## 5. EVALUATION

We built a web application to evaluate our models in a user study. We conducted a pilot study where tag clouds are used to search tracks, a user survey and a follow-up study with the search task and a browsing task where participants used the tag cloud to pilot a recommendation system. We find statistically significant evidence that the topic model and the social model perform better to generate tag clouds that lead to recommend songs that were liked and unknown by the participants than our base line, the popular model.

### 5.1 Pilot Study

The pilot study took place at the university of Lugano. We gathered 17 participants from our Bachelor, Master and Phd programs. Participants registered on an online form before the evaluation. They were asked to fill in an entry form and an exit form to answer general questions. The participants are asked to perform 20 tasks in which they must find a particular track. Tracks are selected randomly from a pool of the 200 most popular tracks. The tag generation method is also selected randomly for each task.

The evaluation is designed as a within subject study. Each participant is her own control group as a model is randomly

selected for each task and the participant is not directly informed of which model is used. Each action of the participants are stored in a log in the database.

Most participants had fun during the experiment. Probably listening to the music and discovering new music helps with this fun aspect and keeps the participants motivated. A participant noticed that quickly he was selecting popular tags and quickly browsing for the “red link” to stop the task. This technique had him finish with the second place, we believe the first finishing participant had the same technique and was rejecting tasks faster if he couldn’t find it with popular tags. From the comments given, a participant gives as advantages “you don’t have to think about the search terms, you can just pick one”, another one adds “relief from typing”. It seems to be the major advantage of tag navigation, it is hard for a person to come up with search terms from the vocabulary he has in mind, whereas when presented with a vocabulary, it is simple for him to choose what terms to use. Multiple participants think it would be simpler for them to type search keywords when they know before hand what terms they would use rather than browsing the tag cloud to find the term they are looking for. Again it seems tag clouds are good to help remembering terms and when the participant does not know what terms to use, but in the case the participant has knowledge of what he is looking for it is easier for her to type. A participant note “if a tag is not in the list, I can not use it. Free search would be better from this point of view”.

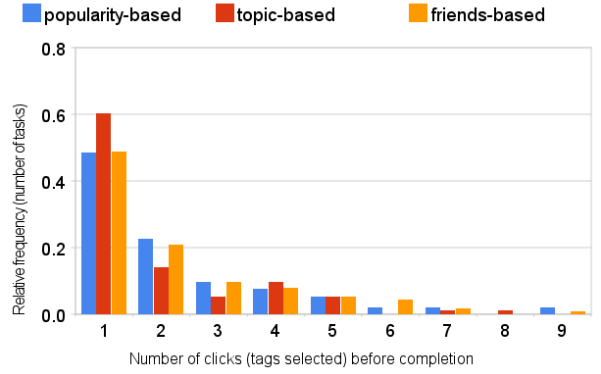
Some participants mentioned as an advantage “discovering new music”. Probably the evaluation process by itself makes the participant discover new music by selecting randomly a track from the 100 most popular tracks. Also people discovered new music by reading the list of tracks when they clicked on tags. A participant mentioned that he would like a tag cloud to navigate pages from his browsing history in his web browser. A tag cloud would help remembering topics he has seen in his browsing life.

Model	Started	Completed	Rel. Frequency (%)
Popular	132	94	71.2 ±3.9
Topic	131	93	71.0 ±4.0
Social	158	116	73.4 ±3.5

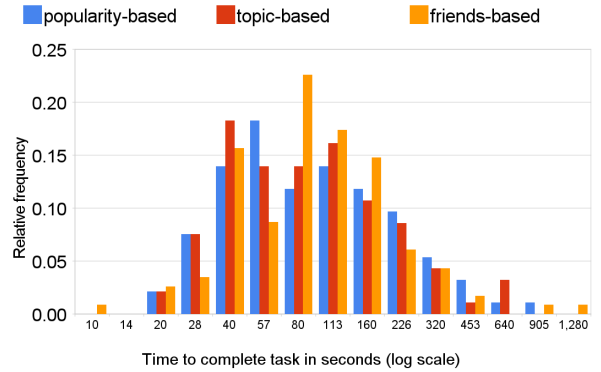
**Table 2: Completed tasks per model. The rate of task completion along with the standard error in the estimate is given in the last column. The models are not found to be statistically significantly different from one another.**

A total of 302 tasks were completed and 101 were rejected. Each time a new task is given the model used to generate the tag cloud is selected randomly from the three models available. 94 tasks were completed for the popular tag cloud and 94 as well for the tag cloud based on topic models. The tag cloud based on social network lead to 116 completed tasks. Participants completed more often tasks involving the social tag cloud rather than the two other tag clouds. Table 2 summarises the number of started and completed tasks and gives the relative frequency in percentage for each model. The relative frequency of completed tasks regarding the number of started tasks for each model is similar.

Figures 4 and 5 give an overview of the results. Figure 4 represents the relative frequency, the number of tasks com-



**Figure 4: Histogram of different navigation path lengths across the three cloud generation models.**



**Figure 5: Histogram of time taking to complete tasks for different models.**

pleted with that number of tags clicked relative to the total number of clicks for each model. We see that most of the tasks were completed after the first click. The tracks to find were selected from the top 100 popular tracks in our dataset. These tracks have a high probability of containing a popular tag.

We have graphed the data to show differences in the distribution of click-counts (navigation path lengths) and time to completion (time to find a song). On average, the time taken to complete a task is slightly shorter for topic-based tag clouds than the popular one (390 seconds against 400 seconds) and a bit better for the social based tag cloud (320 seconds against 400 seconds). While the distributions do vary slightly: the topic based model appears to have slightly lower navigation path lengths, and time to success values, the differences are minimal and the results are not considered conclusive nor statistically significant.

## 5.2 User survey

We conducted a short user survey together with the pilot study. Table 3 gives the statements that were asked to be ranked on a likert scale. Figure 6 represents the answers of the participants for each question.

The answers to question 1 clearly shows that our users are heavy internet users which you would expect when conducting a survey in a computer science faculty. Eleven partic-

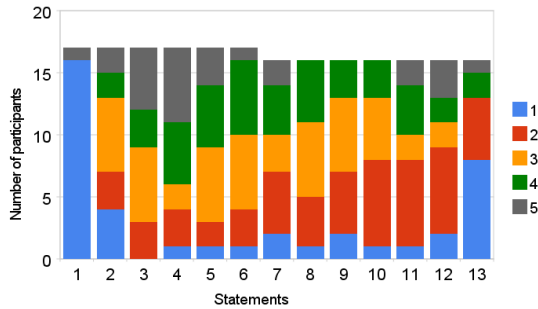


Figure 6: Number of participants per statement (best viewed in colour).

Entry	
1.	I use the internet regularly
2.	I regularly search for music online
3.	I often use tagging systems to search for information
4.	I often tag items in tagging systems
5.	I prefer to navigate tagging systems by clicking on tags rather than searching (via keyword queries)
6.	I am interested in popular music
Exit	
7.	I like navigating the tag cloud
8.	I think it is easy to find items by navigating the tag cloud
9.	I find that managing the selected keywords is easy
10.	I think I can find items quickly with the tag cloud
11.	I would use the tag cloud to navigate the web
12.	I would use the tag cloud to navigate files on my personal computer
13.	I think that tag cloud navigation helps with discovering new music

Table 3: Study statements

Participants mostly disagree with statement 4 and 8 with statement 3 which are both statements about the usage of tagging systems, which shows that tagging is still a feature that is not broadly used by people even in a computer science department. Answers to statements 5 to 9 are inconclusive, participants are mostly undecided. No participant strongly disagree with statement 8 but only 5 mostly agree, finding items by navigating a tag cloud is a hard task for a human which shows that improvements regarding searchability are needed. Eight participants agree with statements 10 and 11 and 9 with statement 12. These three statements are about using the tag cloud to navigate various resources.

Most participants find it easy to navigate the tag cloud and would use a tag cloud to navigate the Web or their personal files. Eight participants out of 17 agree with the 13th statement, 13 mostly agree. This confirms the fact that tag-based navigation improves discovery of new resources.

### 5.3 Follow-up study

We conducted a second study for which we adapted the system based on the comments we received in the pilot study. We improved the efficiency of the system by precomputing term relational matrices ( $p(w|w')$ ). For this evaluation we had 20 participants. None of the participants finished the evaluation, since the search task was harder than in the pilot study. Less results were given per query which forced people to use more precise queries.

Model	Started	Completed	Rel. Frequency (%)
Popular	144	30	20.8 $\pm$ 3.4
Topic	160	32	20.0 $\pm$ 3.2
Social	148	37	25.0 $\pm$ 3.6

Table 4: Number of completed tasks per model. While the social model appears to slightly outperform the other models, the difference is not statistically significant at the 5% confidence level.

Results in Table 4 show our social model slightly outperforming the popular and topic models. The results are not statistically significant.

To complete the tasks participants used multiple tags in their queries, a total of 54 for the popular model, 66 for the topic model and 68 for the social model. This suggests that the social model proposes tags that are more closely related to each other and therefore enables the user to make longer queries.

### 5.4 Experimenting with recommendation

The recommendation experiment consisted of tasks in which participants had to select a tag from the tag cloud and then listen to a song recommended from the current query (the query being composed of the tags selected so far), participants would rate the song (whether they like it or not) and then go back to the new tag cloud generated according to the query and the model.

Model	Rated	Liked	Rel. Frequency (%)
Popular	131	90	<b>68.7</b> $\pm$ 4.1
Topic	104	60	57.7 $\pm$ 4.8
Social	148	75	50.7 $\pm$ 4.1

Table 5: Relative frequencies of liked ratings. The popular model significantly outperforms the other models at the 5% confidence level (according to the two-proportion unpooled one-sided z-test).

Table 5 shows that the popular model outperforms the topic model and social model to generate tag clouds that lead participants to recommended songs that they like. This can be simply explained. Popular items are liked by the majority of people. It is most likely that if we recommend a popular song, it will be liked.

Model	Liked	Unkown&Liked	Rel. Frequency (%)
Popular	90	16	17.8 $\pm$ 4.0
Topic	60	22	<b>36.7</b> $\pm$ 6.2
Social	75	23	<b>30.7</b> $\pm$ 5.3

Table 6: Relative frequencies of unknown resources within liked ratings. Both the topic and social models tend to lead the user to find more unknown music that they like than the popular model. Results are statistically significant at the 5% confidence level.

If we look at the relative frequencies of songs that were new to the participants within the songs that they liked, we find that the popular model is the least efficient, intu-

itively popular items are liked and already known, which is why they are popular because so many people know them. Table 6 shows that the topic model is the best model followed closely by the social model, both models outperform quite significantly the popular model. These results support our thesis that using social relationships enhances the recommendation of new and relevant information. The topic model performs better than the social model, we believe that once the social model is personalized, *i.e.* uses the actual social network of the participant instead of an overall probability from a social network, the social model would perform even better.

## 6. CONCLUSION AND FUTURE WORK

Our work has some limitations, the number of participants of the pilot study and follow-up study is relatively small (17 and 20 participants) which does not allow us to draw strong conclusions. We focused our attention on only one dataset from "Last.fm" with online music data, the conclusions can not be generalised to tag cloud based navigation of other corpora.

Our survey shows that search is not practical with tag clouds whereas recommendation and discovery of new information is. Our follow-up study shows that in the case of recommendation of items that people liked and were new to them, the topic and social models perform much better than the popularity model.

### 6.1 Future Work

We are working on a new evaluation methodology to leverage the social model with social network data from the participants. The rest of the evaluation works as the one described in this paper. We believe that this personalized social model will outperform the topic model.

## 7. REFERENCES

- [1] C. Cattuto. Semiotic dynamics in online social communities. *The European Physical Journal C - Particles and Fields*, 46(0):33–37, aug 2006.
- [2] C. Cattuto, V. Loreto, and L. Pietronero. From the Cover: Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences*, 104(5):1461, 2007.
- [3] P. R. Christopher D. Manning and H. Schutze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [4] J. Fokker, J. Pouwelse, and W. Buntine. Tag-Based Navigation for Peer-to-Peer Wikipedia. In *Collaborative Web Tagging Workshop at WWW2006, Edinburgh*, 2006.
- [5] S. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, April 2006.
- [6] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 2004.
- [7] T. Hammond, T. Hannay, B. Lund, and J. Scott. Social Bookmarking Tools (I). *D-Lib Magazine*, 11(4):1082–9873, 2005.
- [8] Y. Hassan-Montero and V. Herrero-Solana. Improving tag-clouds as visual information retrieval interfaces. In *InScit2006: International Conference on Multidisciplinary Information Sciences and Technologies*, 2006.
- [9] P. Heymann, D. Ramage, and H. Garcia-Molina. Social tag prediction. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 531–538, New York, NY, USA, 2008. ACM.
- [10] T. Ishikawa, P. Klaisubun, and M. Honma. Navigation efficiency of social bookmarking service. pages 280–283, Nov. 2007.
- [11] O. Kaser and D. Lemire. Tag-cloud drawing: Algorithms for cloud visualization. In *Tagging and Metadata for Social Information Organization Workshop, WWW07*, 2007.
- [12] R. M. Keller, S. R. Wolfe, J. R. Chen, J. L. Rabinowitz, and N. Mathe. A bookmarking service for organizing and sharing urls. In *Selected papers from the sixth international conference on World Wide Web*, pages 1103–1114, Essex, UK, 1997. Elsevier Science Publishers Ltd.
- [13] R. Li, S. Bao, Y. Yu, B. Fei, and Z. Su. Towards effective browsing of large scale social annotations. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 943–952, New York, NY, USA, 2007. ACM.
- [14] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *HYPertext '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40, New York, NY, USA, 2006. ACM.
- [15] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*, May 2006.
- [16] D. R. Millen and J. Feinberg. Using social tagging to improve social navigation. In *Workshop on the Social Navigation and Community based Adaptation Technologies*, 2006.
- [17] D. Ramage, P. Heymann, C. D. Manning, and H. Garcia-Molina. Clustering the tagged web. In *Second ACM International Conference on Web Search and Data Mining (WSDM 2009)*, November 2008.
- [18] S. Sen, F. M. Harper, A. LaPitz, and J. Riedl. The quest for quality tags. In *GROUP '07: Proceedings of the 2007 international ACM conference on Supporting group work*, pages 361–370, New York, NY, USA, 2007. ACM.
- [19] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl. tagging, communities, vocabulary, evolution. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 181–190, New York, NY, USA, 2006. ACM Press.
- [20] J. Sinclair and M. Cardew-Hall. The folksonomy tag cloud: when is it useful? *J. Inf. Sci.*, 34(1):15–29, 2008.
- [21] F. Smadja, A. Tomkins, and S. Golder. Collaborative web tagging workshop. In *WWW2006, Edinburgh, Scotland*, 2006.