Leveraging Tagging to Model User Interests in del.icio.us

Julia Stoyanovich[*]
Columbia University
New York, NY 10025
jds1@cs.columbia.edu

Sihem Amer-Yahia **Cameron Marlow** **Cong Yu**
Yahoo! Research
New York, NY 10018
{sihem, cameronm, congyu}@yahoo-inc.com

Abstract

Social tagging sites such as Flickr, YouTube and del.icio.us are becoming increasingly popular. Users of these sites annotate and endorse content by tagging, and form social ties with other users by including them into their friendship network. The richness of social context raises the users' expectations with respect to the quality of served content, but also presents a unique opportunity for the design of semantically-enriched recommender systems. This paper presents a variety of methods for producing customized hotlists and evaluates their effectiveness on del.icio.us datasets. We model a user's interest in terms of the tags he uses to annotate content, and in terms of his explicitly stated and derived social ties, and demonstrate how such interest can be leveraged to produce hotlists of very high quality. We also discuss possible research directions and outline strategies for the design of a social tagging recommender system.

Introduction

The increasing popularity of social tagging sites such as Flickr, del.icio.us, and YouTube is justified by the benefit of addressing specific communities of interest when searching for information on the Web. The ability to define explicit social ties with other users in the same social content site raises users' expectations with respect to the quality of served content. For example, in del.icio.us, a social bookmarking and tagging site, users can subscribe to their friends' feeds in order to learn about their latest bookmarked URLs. They can also view hotlists (most popular URLs) and browse tags to find related URLs. Similar activities are enabled on photo sites such as Flickr and social networking sites such as Facebook. The unprecedented popularity of these sites is the source of a wealth of user-generated content. The ability to sift through large amounts of content and find the right content to recommend to the right user is a challenging problem with significant impact on the *survival* of these sites. While Information Retrieval relies on the assumption that content is static and user interests are dynamic and expressed using keyword search, Information Filtering techniques have been developed to address dynamic content and static user interests (Adomavicius & Tuzhilin 2005; Konstan 2007). We observe that in del.icio.us both content and user interest are dynamic. Therefore social tagging sites present a unique opportunity for content recommendation. The ability to model users and their interests in this context is a new challenge.

The simplest form of recommendation is a *hotlist* – a list of most popular items among a set of users in a given period of time. Towards the goal of understanding content recommendation in collaborative tagging sites, we propose to study the generation and personalization of hotlists on del.icio.us datasets. To the best of our knowledge, our work is the first to evaluate the quality of content recommendation in social tagging sites.

Our Approach

Hotlists are a mean to expose vital content which has global popularity in the system. While globally popular items usually represent consensus between most users, we experimentally observe that such items only account for a small fraction of any individual user's tagging. We thus look for ways to account for user preferences during hotlists computation.

We first represent the interests of a user by the vocabulary he uses to tag URLs: if a significant portion of the users' tagging actions include the tag *sports*, the user is likely interested in sports-related content. This simple observation allows us to replace a single global hotlist by per-tag lists, and to suggest potentially interesting URLs by drawing from one or more per-tag lists in accordance with the user's preferences. We show that tag-driven customization improves hotlist quality, but its success is still limited by the global aspect. In the second approach, we propose to model interest using social ties. These ties are either explicitly stated or derived. An example of an explicit social tie is a friendship network. We explore the utility of friendship ties in hotlist generation, and demonstrate that such ties can indeed be leveraged to generate hotlists of high quality.

Collaborative Filtering (CF) (Park & Pennock 2007) is a popular method used to determine interest overlap between users based on their behavior such as common ratings of items, or common purchasing and browsing patterns. (Items in CF correspond to URLs in our context.) We adopt a similar approach and construct a *common interest network*

[*]Supported by National Science Foundation grant IIS-0121239 and National Institute of Health grant 5 U54 CA121852-03.
Copyright © 2008, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

that links two users if the sets of URLs they tagged overlap significantly. We demonstrate how such networks can be used to construct personalized hotlists of very high quality. However, we also observe that using the entire set of URLs tagged by a user as basis for discovering social ties only applies to a small subset of the users in our user base.

One factor that limits the effectiveness of deriving interest overlap between users in CF is *sparsity*: there are often many more items in the system than any one user is able to tag. (The set of items corresponds to a potentially infinite set of Internet sites.) Another important reason is that people rarely agree on everything: you may agree with your mother on cooking, and with your adviser on research, but your adviser's opinion on food is hardly relevant. We use this idea and demonstrate how tags and item overlap can be combined to construct per-tag common interest networks. Such networks have wider applicability than item-only networks, and can be used to derive hotlists of high quality.

We discuss the quality of a hotlist generation method with respect to its scope and coverage. *Scope* refers to the portion of the user base to which the method can be applied: the larger the scope of a method – the more users can potentially benefit from it. *Coverage* quantifies the average overlap of the hotlist with the user's interests. It is a measure of hotlist quality: the higher the coverage – the more representative a hotlist is of the users' interests. Coverage is related to the concept of *dilution*: the more users cast their votes – the less closely the final top-10 list represents the opinion of each of them individually.

This paper makes the following novel contributions:

- We formalize the problem of customized hotlist generation, and propose evaluation metrics to measure the quality of generated hotlists.
- We present a preliminary quality evaluation of several hotlist generation methods on del.icio.us datasets.
- We discuss open issues in designing a collaborative tagging recommender system.

Data Model and Problem Statement

Given a set of users $\mathcal{U}$, a set of items $\mathcal{I}$, and a set of tags $\mathcal{T}$, we define:

- $\texttt{friends}(u)$ is the set of users in $\mathcal{U}$ defined by u to be his friends. This relationship is directional.
- $\texttt{tags}(u)$ is the set of tags in $\mathcal{T}$ used by user u.
- $\texttt{taggedBy}(i, u, t)$ is true iff user u has tagged item i with tag t. $\texttt{taggedBy}(i, u)$ is defined similarly for any tag.
- $\texttt{items}(u, t) = \{i \in \mathcal{I} \mid \texttt{taggedBy}(i, u, t)\}$ defines all items tagged by u with t. $\texttt{items}(u)$ is defined similarly for any tag.
- $\texttt{taggers}(i, t) = \{u \in \mathcal{U} \mid \texttt{taggedBy}(i, u, t)\}$ denotes all taggers of item $i \in \mathcal{I}$ with tag $t \in \mathcal{T}$. $\texttt{taggers}(i)$ is defined similarly for any tag.

We use the following terminology to describe a hotlist generation method M:

- *Scope* of M is the set of users $\mathcal{U}_{scope} \subseteq \mathcal{U}$ for whom M generates hotlists.
- *Seed* of M is the set of users $\mathcal{U}_{seed} \subseteq \mathcal{U}$ who are used to generate hotlists for $u \in \mathcal{U}_{scope}$.

Given a set of items $\mathcal{I}$, and the seed set $\mathcal{U}_{seed}$, we define the score of an item $i \in \mathcal{I}$ as the number of users in $\mathcal{U}_{seed}$ who tagged item i (each user tags an item at most once):

$$\texttt{score}(i, \mathcal{U}_{seed}) = |\texttt{taggers}(i) \cap \mathcal{U}_{seed}| \qquad (1)$$

The goal of a method M is to produce a hotlist of items *HList* $\in \mathcal{I}$. *HList* includes items that have the highest scores from among those tagged by members of $\mathcal{U}_{seed}$. Without loss of generality, we assume that we generate top-10 hotlists, $|\mathit{HList}| = 10$ for all methods. We quantify the performance of M in terms of *coverage*: a normalized metric that represents the overlap between the hotlist and the items tagged by the user.

$$\texttt{coverage}(\mathit{HList}, u) = \frac{|\mathit{HList} \cap \texttt{items}(u)|}{min(|\texttt{items}(u)|, 10)} \qquad (2)$$

Coverage of a method M is the average of per-user coverage over all users in $\mathcal{U}_{scope}$. We aim to produce hotlists with high average coverage.

del.icio.us Data Description

We evaluate the performance of our hotlist generation methods using del.icio.us datasets. del.icio.us is a social tagging site where users bookmark and annotate URLs with tags, and form social ties by declaring their friendship (also referred to as *network*) with other users.

The dataset of tagging actions is very sparse and follows a long tail distribution (Kipp & Campbell 2006; Golder & Huberman 2006): most URLs are tagged by only a handful of users, and many tags are only used by a few users. Sparse datasets are difficult to process efficiently. We applied the following procedure to reduce the size of the datasets, while still preserving the quality of the generated hotlists. First, we observed that URLs which are rarely tagged stand no chance of contributing to a hotlist. Thus, we removed all URLs that were tagged by fewer than 10 distinct users. Additionally, we removed tagging actions that include uncommon tags: only tags used by at least 4 distinct users are included in our dataset. This cleaning procedure resulted in cutting the tail of URLs and tags. As a result, the dataset was reduced to 27% of its original size.

We evaluate the performance of our hotlist generation methods over a random sample[1] of the cleaned del.icio.us data that includes tagging actions during a consecutive one-month period in 2006. Our cleaned dataset contains 116,177 distinct users who tagged 175,691 distinct URLs using 903 distinct tags, for a total of 2,322,458 tagging actions.

Not all registered del.icio.us users tag URLs. However, for the purposes of our evaluation we focus on taggers – users who contributed at least one tagging action to the

[1] We are not able to reveal our subsetting strategy due to the need to preserve sensitive Yahoo! statistics.

cleaned dataset. The terms *users* and *taggers* are used interchangeably in the remainder of this paper.

Evaluating Hotlist Quality

Using Global Popularity

We first consider the quality of hotlists that are based on global popularity of a URL (this is what is referred to as a "hotlist" by most systems). For this method, $\mathcal{U}_{seed} = \mathcal{U}_{scope} = \mathcal{U}$, and the score of a URL i is simply the number of users who tagged that URL: $\texttt{score}(i) = |\texttt{taggers}(i)|$. The top-10 best URLs computed using this method constitute the hotlist, and this hotlist is global, i.e. it is not customized per-user. We refer to this method as `global`.

The average coverage of `global` over all 116,177 users in our experiments is 3%. This amount, while quite small, indicates that there is some correlation between the users' tagging behavior and globally popular URLs. It also argues for improving the method of producing hotlists by accounting for users' interests.

Combining Global Popularity and Tags

We now examine a hotlist customization method that defines the interests of a user in terms of tags used by that user. We define the *interest* of a user u for tag t as the fraction of the user's tagging actions that include t.

$$\texttt{inter}(u,t) = \frac{|\texttt{items}(u,t)|}{|\texttt{items}(u)|} \tag{3}$$

We compute separate top-10 URL lists for each tag $t \in \mathcal{T}$, and experiment with two different ways of using these lists for hotlist generation. In the first approach, which we call `best_tag`, we identify, for each user u, a single tag from among $\texttt{tags}(u)$ for which $\texttt{inter}(u,t)$ has the highest value, and use the global top-10 URLs for that tag as the user's hotlist. Ties are broken arbitrarily. For $t \in \mathcal{T}$,

$$\begin{aligned}\mathcal{U}_{scope} = \{u \in \mathcal{U} | \quad & \forall t' \in \texttt{tags}(u), \\ & \texttt{inter}(u,t) \geq \quad \texttt{inter}(u,t')\} \\ \mathcal{U}_{seed} = \{u \in \mathcal{U} | \quad & t \in \texttt{tags}(u)\}\end{aligned}$$

The method `best_tag` can be used for all taggers $u \in \mathcal{U}$, and we report the average coverage over all users (116,177). Focusing on the single best tag achieves coverage of 9%, a 6% improvement over `global`. A deeper study of the data reveals that `best_tag` is most effective for users with comparatively higher values of $\texttt{inter}(u,t)$ for their best tag.

We draw two conclusions from our observation: that accounting for tagging behavior improves hotlist quality, and that a more general method for hotlist selection is needed, particularly for cases where no tag can be identified that clearly dominates a user's interest. We thus propose another method, `dom_tags`, where we identify taggers who have a strong interest in one or several tags, and then combine the best URLs from the per-tag top-10 lists into a single custom top-10 hotlist. For a tag $t \in \mathcal{T}$,

# dominant tags	$\lvert\mathcal{U}_{scope}\rvert$	coverage
1	36,736	10%
2	16,452	14%
3	6,466	18%

Table 1: Effect of the number of dominant tags on the performance of `dom_tags`

interest	$\lvert\mathcal{U}_{scope}\rvert$	coverage
30%	36,736	10%
40%	31,391	11%
50%	25,703	13%
60%	23,927	13%
70%	20,943	14%
80%	19,704	14%
90%	19,392	14%
100%	19,347	14%

Table 2: Effect of interest on the performance of `dom_tags`

$$\begin{aligned}\mathcal{U}_{scope} = \quad & \{u \in \mathcal{U} | t \in \texttt{tags}(u) \wedge \texttt{inter}(u,t) \quad > thresh\} \\ \mathcal{U}_{seed} = \quad & \{u \in \mathcal{U} | t \in \texttt{tags}(u)\}\end{aligned}$$

In the current set of experiments, we say that a user u has a strong interest in a tag t if $\texttt{inter}(u,t) > 0.3$. This threshold was determined empirically, and needs further validation in a future study. With an interest threshold set to 0.3, a user can have at most 3 dominant tags. If more than one tag passes the threshold, we draw an equal number of items from the top-10 list that corresponds to each dominant tag. If two tags t_1 and t_2 both pass the interest threshold for user $u \in U_{scope}$, the final list *HList* consists of the top-5 entries from *HList*$_1$ and the top-5 entries from *HList*$_2$. For users with 3 dominant tags, we choose the top-3 URLs from each *HList*$_1$, *HList*$_2$, and *HList*$_3$ and build a top-9 hotlist.

Table 1 lists the partition of the users in our dataset by the number of tags for which they have strong interest, and presents performance of `dom_tags` for these users. Note that `dom_tags` has less than perfect scope: the total number of users partitioned this way is smaller than the entire user base; 49% of del.icio.us users in our dataset have no dominant tags. The first row of Table 1 reports coverage as an average over the 36,736 users who have one dominant tag. For those users, the generated hotlist constitutes 10% of all URLs tagged by each user, on average. This number increases when users have 2 and 3 dominant tags. Clearly, the more dominant tags there are for a user – the better the coverage. However, the higher the number of dominant tags – the more limited the scope of applicability.

We now consider the relationship between strength of a user's interest for a tag, and the coverage of `dom_tags`. We do this for users with a single dominant tag (row 1 in Table 1). Table 2 summarizes our findings: as expected, the stronger the user's affinity for a tag, the better the coverage, but also the more limited the scope.

We argue here that Table 2 motivates a richer hotlist generation strategy: while coverage increases with increasing affinity for a tag, a plateau is reached as interest reaches 60%. Hotlists generated by `dom_tags` take into account a user's interests, as derived from his linguistic choices. However, items in the tag-specific top-10 lists still represent consensus of a very large group of users: $|\mathcal{U}_{seed}|$ is as large as 29,712 for the most popular tag in our experiments, i.e. 29,712 distinct users in our sample used this tag. For users with two or three dominant tags, multiple top-10 lists are combined, further increasing the size of $\mathcal{U}_{seed}$.

In large seed sets, opinions and interests of individual users are approximated. The larger the seed – the coarser the approximation. We call this effect *dilution*. To minimize dilution, given a user $u \in \mathcal{U}_{scope}$, a hotlist generation strategy needs to identify the seed set $\mathcal{U}_{seed}$ that is both representative of the interests of the user u, and focused on that user's interests. In the remainder of this section, we seek to reduce dilution by considering explicit and implicit social ties between users during hotlist generation.

Computing Hotlists using Friendship

The goal of this experiment is to explore the utility of the friendship network in del.icio.us for hotlist generation. Of 116,177 taggers in our experiments, 36,248 (31%) also participate in the friendship network, and we term such users *friendly taggers*. Choosing the friendship network as $\mathcal{U}_{seed}$ is justified by the fact that del.icio.us users tend to pay attention to their friends' tagging actions, which influence their own: users can subscribe to their friends' feeds and get notified whenever one of their friends tags a new URL.

For each friendly tagger, we draw 10 URLs with highest popularity from among URLs tagged by his friends. We refer to this hotlist generation method as `friends`. The scope is the set of friendly taggers, and the seed is defined for a fixed $du \in \mathcal{U}_{scope}$:

$$\begin{aligned} \mathcal{U}_{scope} &= \{u \in \mathcal{U} | \exists f \in \mathcal{U} \wedge f \in \mathtt{friends}(u)\} \\ \mathcal{U}_{seed} &= \{f \in \mathtt{friends}(du)\} \end{aligned}$$

We focus on a *random subset* of friendly taggers, 4,644 in all, for our experiments in this section, and find the coverage of `friends` to be 43%, a significant improvement over `global`, which was 3% for the sample of friendly taggers in our experiments. Note that $avg(|\mathcal{U}_{seed}|) = 4$ for our sample, multiple orders of magnitude less than in the previous sections!

We now explore how tagging can be used to deduce interest overlap among users, and how such overlap can be used to generate hotlists of very high quality. We report two experiments: in the first, a tie is derived between two users if the sets of URLs they tag overlap significantly; in the second, we enrich the set of derived ties by considering tag-specific overlap in URLs.

Interest as Overlap in URLs

In this experiment we compute a *URL Interest Network* by considering the overlap in tagged URLs between users. We quantify *agreement* between users u_1 and u_2 as the fraction of URLs tagged by u_1 that were also tagged by u_2.

agreement	$\|\mathcal{U}_{scope}\|$	$avg(\|\mathcal{U}_{seed}\|)$	coverage
30%	1382	169	61%
50%	913	137	73%

Table 3: Effect of the agreement threshold on the effectiveness of `url_interest`

$$\mathtt{agr}(u_1, u_2) = \frac{|\mathtt{items}(u_1) \cap \mathtt{items}(u_2)|}{|\mathtt{items}(u_1)|} \qquad (4)$$

Note that agreement is directional. If $\mathtt{agr}(u_1, u_2)$ is above a certain *agreement threshold*, we will use URLs tagged by u_2 to derive the hotlist for u_1. We refer to this method as `url_interest`. The scope is the set of users whose agreement with at least one other user is above the threshold. For a fixed user $du \in \mathcal{U}_{scope}$, we define the seed as the set of users with whom du agrees.

$$\begin{aligned} \mathcal{U}_{scope} &= \{u \in \mathcal{U} | \exists f \in \mathcal{U} \wedge \mathtt{agr}(u, f) > thresh\} \\ \mathcal{U}_{seed} &= \{f \in \mathcal{U} \wedge \mathtt{agr}(du, f) > thresh\} \end{aligned}$$

Table 3 summarizes the effectiveness of `url_interest` in terms of scope and coverage. We observe that while the method achieves very good coverage, it is very limited in its scope: only 1382 users can benefit from customized hotlists if 30% agreement is required. The scope is further limited to 913 users for a minimum agreement of 50%. Note also that $avg(|\mathcal{U}_{seed}|)$ is lower for the 50% threshold. We believe this to be a case of lower dilution (fewer users in the $\mathcal{U}_{seed}$) leading to higher coverage. However, further experimental validation is needed to better understand the effect of dilution on scope and coverage of `url_interest`.

The limited scope of `url_interest` is due to the fact that strong agreement of the kind required by this method is uncommon. To include an edge between two users in the interest network, we require that they agree on at least 30% of the tagged URLs over-all. We observe that, while agreement of this kind over all interests may be rare, people more commonly agree with others on only a part of their interests. We explore this idea in the final experimental section.

Interest as Overlap in URLs and in Tags

In this section, we propose to combine interest in a tag, as explored in `dom_tags`, with agreement based on URL overlap, as in the previous section, to construct a *tag-URL-interest network*. We use tag interest to generate a global partitioning of tagging actions. We then search for URL-agreement within these partitions. We call this method `tag_url_interest`, and propose to use it for the taggers who show strong interest in one or more tags (see Equation 3). We first define a tag-specific version of agreement as:

$$\mathtt{agr}(u_1, u_2, t) = \frac{|\mathtt{items}(u_1, t) \cap \mathtt{items}(u_2, t)|}{|\mathtt{items}(u_1, t)|} \qquad (5)$$

We define the scope and seed for a fixed tag $t \in \mathcal{T}$. The scope of `tag_url_interest` is the set of taggers with

method	$\|\mathcal{U}_{scope}\|$	$avg(\|\mathcal{U}_{seed}\|)$	coverage
`dom_tags`	1235	26,856	17%
`tag_url_inter`	1235	227	82%
`url_inter`	205	203	85%

Table 4: Relative performance of `dom_tags`, `tag_url_interest`, and `url_interest`

strong interest in t, and with URL agreement, *for that tag*, with at least one other user. The seed is defined for a fixed $du \in \mathcal{U}_{scope}$, and for a fixed tag t, as the set of users who are in tag-url-agreement with du and who have used t.

$$\begin{aligned}
\mathcal{U}_{scope} = \quad & \{u \in U| \quad t \in \texttt{tags}(u) \wedge \\
& \qquad\qquad \texttt{inter}(u,t) > int_thresh \wedge \\
& \qquad\qquad \exists f \in \mathcal{U} \wedge \texttt{agr}(u,f,t) > agr_thresh\} \\
\mathcal{U}_{seed} = \quad & \{f \in \mathcal{U}| \quad t \in \texttt{tags}(f) \wedge \\
& \qquad\qquad \texttt{agr}(du,f,t) > agr_thresh\}
\end{aligned}$$

We evaluated the effectiveness of `tag_url_interest` on a subset of users in our experiments: we choose users with strong interest in exactly 2 tags. Out of 16,452 users with strong interest in 2 tags, 1235 were in the scope of `tag_url_interest`. We note that, while the scope of the current method is still limited, it greatly exceeds the scope of `url_interest`: only 205 users in the scope of `tag_url_interest` where also in the scope of `url_interest`. Table 4 summarizes the relative performance of `dom_tags`, `url_interest` and `tag_url_interest` for users in $\mathcal{U}_{scope}$ of `tag_url_interest`. Because $\mathcal{U}_{scope}$ of `tag_url_interest` is a subset of $\mathcal{U}_{scope}$ of `dom_tags`, we report performance of `dom_tags` for all users in `tag_url_interest`. In this table, we use 30% as the threshold for both interest and URL agreement.

`tag_url_interest` significantly outperforms `dom_tags` in terms of coverage. However, for users who are in both `tag_url_interest` and `url_interest`, the latter does better. This represents 205 users for whom `tag_url_interest` achieves an 82% coverage while `url_interest` achieves an 85% coverage. This reinforces the idea that accounting for agreement over all URLs is stronger than agreement on one tag at a time, and supports our dilution hypothesis: $avg(|\mathcal{U}_{seed}|)$ is smallest for `url_interest`, followed by `tag_url_interest`.

Related Work

The idea of motivating participation by displaying the value of contribution is characteristic of collaborative reviewing sites (Rashid *et al.* 2006) but has received little attention in collaborative tagging sites. Impact of reviews has been studied extensively in the e-commerce arena, and it has been shown that reviews impact sales (Chen, Dhanasobhon, & Smith 2007; Chevalier & Mayzlin 2006; Ghose & Ipeirotis 2006). Most of the current scientific literature dealing with user-contributed reviews concerns text analysis to distill reviews (Popescu & Etzioni 2005), sentiment detection (Pang & Lee 2004; 2005), and the impact of reviews on product sales (Bickart & Schindler 2001; Ghose & Ipeirotis 2007; Kim *et al.* 2006; Jingjing *et al.* 2007). In contrast, very little has been done to extract value in collaborative tagging systems and provide participation incentives to users.

In (Wu, Zhang, & Yu 2006), a probabilistic generative model is defined that uses tagging to obtain the emergent semantics in social annotations. The authors study the relationship between items, tags, and users by means of co-occurrence analysis, and map these elements to a multi-dimensional *conceptual space*, where each dimension represents a category of knowledge. The authors demonstrate how their model may be learned and subsequently used to derive tag ambiguity information. They go on to show how their probabilistic model may be used for semantic search and discovery in social tagging sites like del.icio.us. A model such as that outlined in (Wu, Zhang, & Yu 2006) may be used as basis for an alternative hotlist generation strategy. Quality of this strategy in terms of scope and coverage may be evaluated in a framework such as ours.

Collaborative Filtering is a popular method which compares user profiles in order to determine interest overlap (Adomavicius & Tuzhilin 2005). A user profile is built from explicit or implicit data. The system can explicitly ask users to rate an item on a numerical scale, or to rank a collection of items from most to least favorite. The system can also record items that a user browsed or purchased, and analyze item viewing times. CF compares user profiles and calculates a list of recommended items for the user. Several methods have been developed to address data sparsity. Most of them (item-based and user-based) rely on statistical approximation (Bell, Koren, & Volinsky 2007; Park & Pennock 2007).

In our approach, we express interest qualitatively, using tags and derived ties. In (Agichtein, Brill, & Dumais 2006), it is shown that ranking in Web search can be improved by incorporating user behavior. Similarly, we show that incorporating tagging behavior improves hotlist generation. This motivates the need to better understand the principles behind designing a recommender tagging system, as was briefly discussed in (Golder & Huberman 2006). According to a study of del.icio.us tagging practices described in (Kipp & Campbell 2006), tagging exhibits self-organizing patterns. In our work we explore how users can be classified into groups based on their tagging behavior, and how such groups can be used to improve the quality of recommended hotlists.

Summary and Discussion

We showed that users' tagging behavior can be leveraged to derive implicit social ties, and that such ties serve as a good indicator of users' interests. Clearly, hotlists generated by observing social ties are of higher relevance than those produced by global hotlist generation strategies. In this section, we discuss the challenges towards the design of a *social tagging recommender system.*

Interaction with Users

It has been shown that providing an explanation for recommendations helps inspire user trust and loyalty, making it easier for users to identify interesting items (Konstan 2007; N.Tintarev & J.Masthoff 2007). While numerical values (average item ratings) have been used in the past by recommenders such as Amazon and Launch, they only provide a coarse explanation of recommendations. Some systems support their recommendations by statements about previous performance such as "MovieLens has predicted correctly 80% of the time for you". Complex explanations such as two dimensional statistics, e.g., correlation and variance, often do not work. Instead of stating "People who have tagged this URL, also tagged these others", we propose to make explanations more meaningful by providing semantic information such as in "These URLs have been tagged by people who share the same interest as you in Sports".

Core Recommendation Strategy

Users Our experiments demonstrate that there is significant heterogeneity in users' tagging behavior, and thus in the types of social ties that can be derived for different users. Applicability and effectiveness of a recommendation strategy to a user is directly influenced by these social ties. We argue for partitioning the users into categories that fit their behavior and interests for purposes of recommendation.

Time Sensitivity There are different reasons why a user would bookmark a URL and use specific tags to label it. Our evaluation of hotlist recommendation strategies shows that a user's interest in a URL is influenced by his social ties, many or all of which may be external: formed in real life, in another social content site, etc. Another external influence over a user's interest in a URL or in a tag is the general popularity of that URL or tag. Collaborative Filtering is based on an assumption that people who agreed in the past tend to agree again in the future, and that items that are similar will continue to be similarly liked or disliked by a given user. Consequently, CF systems do little to detect or predict changes in user preferences. During our experimental evaluation we found that the assumption of persistence of interest and agreement does not hold in del.icio.us: derived social ties, particularly those based on item overlap, tend to evolve. We identified the short lifespan of URLs as the main reason: most URLs are only tagged actively for a period of 1-2 weeks. Similarly, bursts in interest have been observed and studied in the past (Cohen & McCallum 2003). Indeed, some tags appear at certain periods of time and indicate a trend in the general public or among a community of users, e.g., people tend to visit travel sites more often during holiday time. It is thus essential to incorporate time into any successful social tagging recommendation strategy.

References

Adomavicius, G., and Tuzhilin, A. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* 17(6).

Agichtein, E.; Brill, E.; and Dumais, S. 2006. Improving web search ranking by incorporating user behavior information. In *SIGIR*.

Bell, R.; Koren, Y.; and Volinsky, C. 2007. Modeling relationships at multiple scales to improve accuracy of large recommender systems. In *KDD*.

Bickart, B., and Schindler, R. 2001. Internet forums as influential sources of consumer information. *J. of Interactive Marketing* 15(3).

Chen, P.-Y.; Dhanasobhon, S.; and Smith, M. 2007. All reviews are not created equal: The disaggregate impact of reviews and reviewers at amazon.com. In *ICIS*.

Chevalier, J. A., and Mayzlin, D. 2006. The effect of word of mouth on sales: Online book reviews. *J. of Marketing Research*.

Cohen, W., and McCallum, A. 2003. Information extraction from the world wide web. In *KDD*.

Ghose, A., and Ipeirotis, P. 2006. Designing ranking systems for consumer reviews : The impact of review subjectivity on product sales and review quality. In *WITS*.

Ghose, A., and Ipeirotis, P. 2007. Designing ranking systems for consumer reviews : The economic impact of customer sentiment in electronic markets. In *ICDSS*.

Golder, S. A., and Huberman, B. A. 2006. The structure of collaborative tagging systems. *Information Dynamics Lab, HP Labs*.

Jingjing, L.; Cao, Y.; Lin, C.; Huang, Y.; and Zhou, M. 2007. Low-quality product review detection in opinion summarization. In *EMNLP-CoNLL*.

Kim, S.-M.; Pantel, P.; Chklovski, T.; and Pennacchiotti, M. 2006. Automatically assessing review helpfulness. In *EMNLP*.

Kipp, M. E., and Campbell, D. G. 2006. Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. *Faculty of Information and Media Studies University of Western Ontario*.

Konstan, J. A. 2007. Introduction to recommender systems. In *SIGIR*.

N.Tintarev, and J.Masthoff. 2007. A survery of explanations in recommender systems. In *WPRSIUI*.

Pang, B., and Lee, L. 2004. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In *ACL*.

Pang, B., and Lee, L. 2005. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In *ACL*.

Park, S.-T., and Pennock, D. 2007. Applying collaborative filtering techniques to movie search for better ranking and browsing. In *KDD*.

Popescu, A.-M., and Etzioni, O. 2005. Extracting product features and opinions from reviews. In *HLT*.

Rashid, A. M.; Ling, K.; Tassone, R. D.; Resnick, P.; Kraut, R.; and Riedl, J. 2006. Motivating participation by displaying the value of contribution. In *CHI*.

Wu, X.; Zhang, L.; and Yu, Y. 2006. Exploring social annotations for the semantic web. In *WWW*.