How Social is Social Tagging?

Stephan Doerfel
University of Kassel
doerfel@cs.uni-kassel.de

Daniel Zoller
University of Würzburg
zoller@informatik.uni-
wuerzburg.de

Philipp Singer
Graz University of Technology
philipp.singer@tugraz.at

Thomas Niebler
University of Würzburg
niebler@informatik.uni-
wuerzburg.de

Andreas Hotho
University of Würzburg
hotho@informatik.uni-
wuerzburg.de

Markus Strohmaier
GESIS & U. of Koblenz
markus.strohmaier@gesis.org

ABSTRACT

Social tagging systems have established themselves as an important part in today's web and have attracted the interest of our research community in a variety of investigations. This has led to several assumptions about tagging, such as that tagging systems exhibit a social component. In this work we overcome the previous absence of data for testing such an assumption. We thoroughly study the hypothesis of social interaction, leveraging for the first time live log data gathered from the real-world public social tagging system BibSonomy. Our empirical results indicate that sharing of resources constitutes an important and indeed social aspect of tagging.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Information Networks*

Keywords

social tagging; assumptions; folksonomy; behavior

1. INTRODUCTION

Social tagging systems such as Delicious, BibSonomy or Flickr have attracted the interest of our research community for almost a decade. One predominant assumption about social tagging systems is that they are indeed *social* – as their name suggests. The social aspects of tagging have been emphasized early in the history of tagging systems. E.g., in 2005, Weinberger [4] referred to it as one of two aspects that "make tagging highly useful" – i.e., users tag their resources publicly, see the collections of other users and most importantly also might use resources originally posted by others. Nevertheless, there is also the competing hypothesis, that personal information management may be one of the main reasons why people use tagging systems.[1]

---

[1] http://okcancel.com/archives/article/2005/07/tagging-for-fun-and-finding.html

However, testing this assumption[2] on a large scale has been nearly impossible so far, because datasets of detailed usage logs have been unavailable. Thus, most of the previous work analyzing social tagging systems focuses on data displayed on the web, i.e., posts. This data allows for investigating evidence of social behavior, like the number of common resources or tags in user collections. However, only the use of server log data, allows a more detailed study of actual social behavior. In this work we use *usage log data* to investigate social behavior on two levels: First, we examine whether users visit resources tagged by others and second (as even stronger evidence of social interaction) whether they also adopt tagged content of others by copying it.

To the best of our knowledge, this paper presents the first systematic and in-depth empirical study leveraging actual web server log data gathered from a *real-world, public* social tagging system. Our data comes from the social tagging system BibSonomy [1], which allows users to store and share links to websites as well as mostly scientific publications. We use a cleaned dataset (e.g., removing bots, spammers, unsuccessful requests, and requests to resources like CSS, etc.) obtained from BibSonomy's server logs. It consists of around 3.5 million user requests made between 2006 and the end of 2011.

The closest work to the one in this article is by Millen and Feinberg [3], who investigated user logs of the IBM internal social tagging system "dogear" and found that among all user requests to posts, about 74% targeted posts bookmarked by other users. Since their work focuses on a system that is used inside a company and is not publicly accessible, we do not know whether their results hold in general.

2. RESULTS

In the following, we investigate how much social interaction can be detected in the usage data of BibSonomy.

Retrieval. We start by looking at all retrieval-oriented requests. From the server logs we can observe that more than two thirds of all requests (about 68%) from logged-in users go to their own pages. Users visit other pages in about 32% of the requests to look at either general pages containing posts

---

[2] In the literature, several further assumptions have been made about tagging systems, e.g., that their key components (i.e., users, tags and resources) are equally important structural elements of social tagging systems or that popularity in terms of posts means importance. We investigate these other assumptions in greater detail in our technical report [2] and focus here on the social assumption.

of several users (about 16%) or content of individual other users (about 16%). This share of visits to content of others is far below the reported 74% in [3] for a company internal tagging system. In summary, we see that the larger share of interactions in BibSonomy happens with the personal collection. However, the interest in other users' content accounts for a significant part – almost one third of all retrieval requests – of the interaction with the system.

Copying Resources. Next, we investigate to what extent users not only visit content from others, but copy it (instead creating new posts themselves). There are different ways to post a single resource to BibSonomy: (a) by manually filling in a form with all metadata of a publication or bookmark, (b) by using a bookmarklet in the browser that extracts all metadata from the currently visited website or publication, (c) by copying existing entries from other users while browsing in the system, or (d) using other methods like the extraction of metadata using the ISBN of a publication. While copying resources from others (c) would be an instance of sharing, the other options (a,b, and d) are rather examples of personal information management. Figure 1 illustrates that the bookmarklet feature is used most frequently (64%), followed by entering the metadata manually (25%), and only about 11% of the resources are copied from other users. Less than 1% were stored using one of the other methods. When we look at publications and bookmarks separately, we observe that the ratios between the distinct ways of posting resources differ. Publications are copied more often than bookmarks (18% vs. 3%) and users use the bookmarklets for bookmarks (81%) more frequently than for publications (48%). One reason for these differences might be the fact that users leave the system when they follow a bookmarked link, while they stay within BibSonomy when they check out details of a publication. Thus, using the bookmarklet is the easiest way to post a website and clicking the copy button is a more convenient option for copying a visited publication. We also note that the share of 3% of copied bookmarks is close to the 2.2% share reported in [3] for the IBM-internal system dogear, while the share for publications (18%) exceeds that value by a factor of eight.

Since a resource can only be copied if another user had already posted that resource in BibSonomy, we have to take into account whether posted resources were already in the system when a user posted them. Among the posting requests where posts were not copied from another user, only about 17% had a corresponding resource already present in the database and thus could have been copied. In 42% of the postings where the resources had previously been posted, the new post was created through copying. This can be regarded as a relatively large share, since looking up publications and bookmarks in BibSonomy is only one out of many possible ways to find interesting resources on the web or elsewhere.

Copying Tags. Finally, we study whether not only resources, but also tags are shared. To that purpose we counted how often users who copied a resource used tags from their own vocabulary or tags of the original post to describe their new post. In 87% of all copy requests, at least one tag from the own vocabulary has been used. In 42% of all copies, at least one of the original post's tags has been adopted. In the other copy events, 44% of the original posts had only special tags like "imported" that are probably not meaningful for the user copying the post. Similarly to the copying of resources we find evidence for social interaction: users share tags, i.e.,

although in the majority of cases own tags were used, in a large number of cases, tags from others are adopted as well.

3. CONCLUSIONS

In this work we empirically focused on finding answers to the question *how social is social tagging?* We tackled this question via a study of a large-scale usage log dataset obtained from the public social tagging system BibSonomy. We found evidence for both personal information management and social interaction: Although more requests are spent on a user's own collection than on other users' posts, the interest in other users' content is still significant as posts of other users are frequently visited and copied. The shares of visited posts and copied resources and tags are evidence of social interaction and demonstrate, that the collaborative aspect of the tagging system is recognized and used.

In the future, we want to address further assumptions and we plan to analyze individual user behavior like personal ratios of contribution and retrieval, or preferences to tag querying or text search, as well as the influence of recommender systems and the user interface.

4. REFERENCES

[1] D. Benz, A. Hotho, R. Jäschke, B. Krause, F. Mitzlaff, C. Schmitz, and G. Stumme. The social bookmark and publication management system BibSonomy. *The VLDB Journal*, 19(6):849–875, Dec. 2010.

[2] S. Doerfel, D. Zoller, P. Singer, T. Niebler, A. Hotho, and M. Strohmaier. Of course we share! Testing assumptions about social tagging systems, 2014. cite arxiv:1401.0629.

[3] 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.

[4] D. Weinberger. Tagging and why it matters. *SSRN eLibrary*, 2005.