What Users Actually Do in a Social Tagging System: A Study of User Behavior in BibSonomy

STEPHAN DOERFEL, University of Kassel
DANIEL ZOLLER, University of Würzburg
PHILIPP SINGER, GESIS
THOMAS NIEBLER, University of Würzburg
ANDREAS HOTHO, University of Würzburg and L3S Research Center
MARKUS STROHMAIER, GESIS and University of Koblenz

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. Henceforth, several aspects of social tagging systems have been discussed and assumptions have emerged on which our community builds their work. Yet, testing such assumptions has been difficult due to the absence of suitable usage data in the past. In this work, we thoroughly investigate and evaluate four aspects about tagging systems, covering social interaction, retrieval of posted resources, the importance of the three different types of entities, users, resources, and tags, as well as connections between these entities' popularity in posted and in requested content. For that purpose, we examine live server log data gathered from the real-world, public social tagging system BibSonomy. Our empirical results paint a mixed picture about the four aspects. Although typical assumptions hold to a certain extent for some, other aspects need to be reflected in a very critical light. Our observations have implications for the understanding of social tagging systems and the way they are used on the Web. We make the dataset used in this work available to other researchers.

CCS Concepts: • **Human-centered computing → Social tagging**; **Social tagging systems**; • **Information systems → Web log analysis;**

Additional Key Words and Phrases: Social tagging, assumptions testing, social sharing, folksonomy, bookmarking, behavior

ACM Reference Format:
Stephan Doerfel, Daniel Zoller, Philipp Singer, Thomas Niebler, Andreas Hotho, and Markus Strohmaier. 2016. What users actually do in a social tagging system: A study of user behavior in BibSonomy. ACM Trans. Web 10, 2, Article 14 (May 2016), 32 pages.
DOI: http://dx.doi.org/10.1145/2896821

A small part of this work, specifically part of Section 4.1, was published as an extended abstract in Doerfel et al. [2014a].
This work is funded in part by the DFG through the PoSTS II project.
Authors' addresses: S. Doerfel (corresponding author), Universität Kassel, Knowledge & Data Engineering, Wilhelmshöher Allee 73, 34121 Kassel (Germany); email: doerfel@cs.uni-kassel.de; D. Zoller, T. Niebler, and A. Hotho, Julius-Maximilians-Universität Würzburg, Data Mining and Information Retrieval Group, Am Hubland 97074 Würzburg (Germany); emails: {zoller, niebler, hotho}@informatik.uni-wuerzburg.de; P. Singer and M. Strohmaier, GESIS, Department Computational Social Science, Team Data Science, Unter Sachsenhausen 6-8, 50667 Köln (Germany); emails: {philipp.singer, markus.strohmaier}@gesis.org.

DOI: http://dx.doi.org/10.1145/2896821

1. INTRODUCTION

Social tagging systems have emerged as a way of organizing information and have established themselves as an alternative to more traditional resource directories. In tagging systems, users collect, share, and annotate resources, such as Web sites,[1] photos,[2] or publications.[3] Users of tagging systems have expressed their desire for tagging (assigning keywords (tags) to resources) in other systems as well [Noy et al. 2008], and so the practice of tagging has also found its way as a secondary feature into many Web systems, such as tagging of products in Web stores, articles in blogs, or tweets on Twitter[4] (hashtags).

Social tagging systems have attracted the interest of our research community for almost a decade [Mathes 2004; Golder and Huberman 2006]. Significant advances have been made with regard to our understanding about the emergent, individual, and collective processes that can be observed in such systems [Strohmaier et al. 2010]. Useful algorithms for retrieval [Hotho et al. 2006] and classification [Zubiaga et al. 2011] have been developed that exploit the rich fabric of links between users, resources, and tags in social tagging systems for facilitating information organization, search, and navigation. Other work has focused on the extraction or stabilization of emergent semantics [Golder and Huberman 2006; Cattuto et al. 2007].

While this line of research has significantly increased our ability to describe, model, and utilize social tagging systems, our community has also built its work on certain assumptions about how these systems are used. However, whether—and to what degree—these assumptions hold is still an open research question. In the literature, arguments and evidence regarding the use of tagging systems have been discussed controversially, and researchers have argued for and against them in our community, providing thus all the more reason to evaluate them on real-world usage data. Only a few studies have actually analyzed user behavior in social tagging systems to better investigate these research questions, either by (i) conducting user surveys (e.g., Heckner et al. [2009]) or by (ii) tapping into the rich corpus of tagging data (i.e., the posts) that is available on the Web (e.g., Cattuto et al. [2007]). However, such studies come with certain limitations, such as self-reporting biases or the lack of detailed usage data that reveals how users actually request information. In this article, we overcome these drawbacks by presenting and thoroughly investigating a detailed usage log of the popular real-world, open social tagging system BibSonomy. We thus provide evidence from actual user behavior to shed light on a series of questions from related work regarding the use of a tagging system.

Research questions. We discuss the following four controversial aspects about the use of social tagging systems:

—*The social aspect*: Tagging systems are supposed to be used collaboratively to tag and share resources. We investigate to which degree such sharing actually happens and discuss evidence for the interest of users in the content of others.

—*The retrieval aspect*: The main activities in a tagging system are storing resources and retrieving them later (using the assigned tags). We investigate whether and when users retrieve their resources.

—*The equality aspect*: In the folksonomy model (e.g., Hotho et al. [2006]), users, resources, and tags are modeled as equally important sets of entities. We investigate

[1]For example, BibSonomy: http://www.bibsonomy.org or Delicious: http://www.delicious.com.
[2]For example, Flickr: http://www.flickr.com/.
[3]For example, CiteULike: http://www.citeulike.org/ or BibSonomy.
[4]http://twitter.com.

whether they are indeed equally important in navigation or whether one of the three types of entities is preferred for retrieving and browsing content.
—*The popularity aspect*: The popularity of users, tags, and resources in posts is often seen as an indicator of importance, such as in tag clouds where frequent tags have large font sizes to gain users' attention and to be easily accessible by a mouse click. We investigate whether popularity in posts is matched by popularity in retrieval.

For each of the four research questions, we formulate a (naive) assumption about tagging systems:

—Tagging systems are (as their name suggests) social.
—Users do retrieve their resources after they have stored them.
—Users, resources, and tags are equally important for navigation.
—Popularity in posts implies popularity in requests.

These assumptions are very plain statements, and we do not expect to find them confirmed just like that. However, as they reflect beliefs about tagging systems (evidence from the literature for each aspect is presented in the according sections of Section 4), it is worth investigating to what degree they actually do hold.

Findings. In our analysis of the social tagging system BibSonomy, we find evidence both for and against the assumption that the activities in a tagging system are primarily social. Whereas some user actions indeed indicate social sharing, others are evidence for individual purposes. This suggests that both kinds of activity are relevant in a social tagging system and should therefore be supported in the system's design. We also find that although users post a large number of resources and tags to BibSonomy, they only retrieve a rather small fraction of them later. Next, we find a strong inequality between the use of users, tags, and resources within BibSonomy. User pages are visited much more often than corresponding resource or tag pages, providing clear evidence that assuming tags, users, and resources to be equally important for the navigation in BibSonomy would be wrong. Finally, whereas we observe common usage patterns in post and request behavior on an aggregate level, the patterns are less pronounced on an individual level, suggesting that an entity's popularity in posts is only reflected to certain extents in the requests to that entity.

Contributions. This article makes contributions on three levels. First, on the *methodical* level, we identify several aspects and illuminate a way toward investigating them with log data. Although our findings are limited to a single system (BibSonomy), our method of examining such systems is general. The approach can well be applied to other social tagging systems to investigate these aspects in different contexts and to test assumptions like the ones mentioned previously. Second, on the *empirical* level, we investigate several research questions regarding user behavior in a social tagging system by testing them with actual log data and we report the results exemplarily for the scholarly social tagging system BibSonomy. We discuss and speculate about influences on the behavior in tagging systems of other (non-scholarly) contexts in Section 5. Third, on the *data-driven* level, we make an anonymized BibSonomy server log dataset available to other researchers (see Section 3). This will enable our community to not only verify our results but also investigate similar or new questions on a unique dataset that is not yet available.

Overall, our findings are relevant for researchers interested in user behavior and modeling in the context of social tagging systems and their adoption. In addition, they are relevant to system engineers interested in improving the utility and usefulness of social tagging systems on the Web.

Structure of this article. After the discussion of related work in Section 2, we describe the BibSonomy datasets in Section 3. We then turn our attention to studying and evaluating the aforementioned four aspects on social tagging in Section 4. In Section 5, we discuss differences between BibSonomy and other tagging systems and thus the generalizability of our results. Finally, Section 6 concludes the article.

2. RELATED WORK

In this section, we discuss related literature on the investigation of tagging systems and log file analysis in general. Further related work that is specifically relevant to individual aspects will be discussed in greater detail later in the corresponding context in Section 4.

Inception of tagging. Work on social tagging and emerging folksonomies began in late 2004, when the term *folksonomy* was first mentioned by Vander Wal[5] and continued in 2005 in various blog posts and papers. One of the first reviews about social tagging systems was provided by Mathes [2004]. He noted that social tagging systems allow a much greater variability in organizing content than formal classification can provide. Mathes was also among the first to hypothesize that tag distributions follow power law distributions that can characterize the semantic stabilization of such systems (see Golder and Huberman [2006]). Furthermore, Mathes identified some potentials and uses of tagging systems, such as serendipitous browsing.

User surveys and post analysis. Abrams et al. [1998] already discussed the management of Web site bookmarking long before the rise of social tagging on the Web using a user survey and bookmark files from participants. Their results showed that users are motivated to share bookmarks (still via email back then) and to retrieve them later. Heckner et al. [2009] conducted a survey of tagging systems (namely, Flickr, YouTube, Delicious, and Connotea) with 142 users regarding their motivations. It showed that there are mainly two motivations for users to post content: sharing resources with others and storing them primarily for personal retrieval later on. The strength of these motivations varies from system to system.

Using the post data of tagging systems, several studies analyzed aspects of posting behavior, such as the distributions of users, resources, and tags in posts [Cattuto et al. 2007], behavioral models that explain which tags are chosen [Halpin et al. 2007], or the identification of different types of users—categorizers and describers—regarding their choice of tags [Strohmaier et al. 2010]. However, these studies did not use log data for their analysis to explore the actual retrieval behavior. A review of social tagging, regarding a variety of diverse aspects of such systems—including vocabulary, structure, visualization, motivation, or search and ranking—was presented in Trant [2009].

Web log mining. Predominantly, Web logs have been used to investigate the query behavior in search engines or the use of digital libraries to better understand such a system's users. This can help Webmasters tailor their systems more specifically to the users' needs. A survey on such works about search engines was created by Agosti et al. [2012]. More recently, Thomas [2014] used a combination of controlled user study and Web log analysis to identify signals for situations when users were struggling and found that simple signals, particularly the time spent in a session, are good indicators. Duarte Torres et al. [2014] found significant differences in the search behavior of young and adult users, for instance, regarding the length of queries or the selection of ranked search results. Examples for the analysis of digital libraries can be found in the works of Nicholas et al. [2005].

[5]http://vanderwal.net/random/category.php?cat=153.

Tagging systems exhibit aspects of both search engines and libraries. Although their data is a collection of resources with description and categories, like in a digital library (yet not professionally organized), it is created and organized by users in their individual fashion of assigning tags and entering metadata. Nonetheless, the data is clearly more structured than data on the Web in general, as posts are constructed according to a specified template.

For the analysis of user behavior in social Web systems, request logs have successfully been exploited by Schneider et al. [2009] and Benevenuto et al. [2009]. The gained insights are useful in social studies, and they can help improve a system's design and its traffic distribution over the hardware and can be used for planning viral marketing and advertisement placement. Benevenuto et al. [2009] collected data from a social network aggregator over a period of 12 days. They found that session durations follow heavy-tailed distributions and that users tend to stick with one feature (e.g., photos) within consecutive requests. We conduct a similar analysis in BibSonomy, analyzing the transition probabilities between the retrieval of users, tags, and resources in Section 4.3. Schneider et al. [2009] had access to the click streams of large ISPs and could thus analyze the popularity of individual features over several social networks. They found that the distributions of requests over different features differ from system to system. Similarly, like Benevenuto et al. [2009], they found that especially in the most dominant feature categories (like photos and messaging), users often spend consecutive requests to the same category.

Jiang et al. [2013] presented an analysis of the Web logs of the Chinese social network Renren. They looked at so-called latent interactions (i.e., visits to a pages). They found that such latent interactions account for the majority of activities in the network and that there are more users who passively consume the content of the network than there are users who actively engage in interaction with others. Further experiments reveal that visits to strangers are rarely reciprocated (even though Renren users can see who visited their content). By and large, latent interactions are "less limited by constraints, such as time and energy, but more meaningful (...) than the social graph" [Jiang et al. 2013]. For social bookmarking systems, such findings raise the following question: how strong is the relation between the active contribution of tags and resources and their consumption (in terms requests to them)? We therefore analyze the popularity in retrieval requests (representing the latent interaction or consumption) and in posts (representing the active contribution) in Section 4.4.

Web log mining in social tagging systems. Only very few studies have used Web logs in their analysis of tagging systems. Carman et al. [2009] combine tagging data with log data from search engines and compare the distribution of tags to that of query terms in search. They find a large overlap in the systems' vocabularies and correlations between the frequency distributions of queries and tags to the same URLs. However, they also provide evidence that both tag and query term samples do not come from the same distribution.

Although a large amount of literature on tagging systems exists, to the best of our knowledge the only work utilizing and analyzing log data from a tagging system are that of Millen and Feinberg [2006], Millen et al. [2007], and Damianos et al. [2007]. Millen and Feinberg [2006] investigated user logs of the social tagging system *Dogear* (internally used at IBM) with a focus on social navigation in the system. They found strong evidence that social navigation—users who are regularly looking at bookmark collections of other people—is a fundamental part of the social tagging system. They also found a positive correlation between the assignment frequency of a tag in posts and the frequency of it being used for browsing. These findings have been highly relevant for the understanding of tagging behavior, as they provide actual evidence of how users make use of a tagging system's content. Millen et al. [2007] combined log analysis and user interviews to investigate the way users retrieve resources. They observe diverse behavior patterns for different users and find that heavy users tend to spend more time with their own collections than users with only few bookmarks.

Damianos et al. [2007] introduced a tagging prototype called *onomi* to the organization MITRE. They use log data to determine how well the system was accepted and present several usage statistics from a 10-month test period. They found that their users can be categorized into information providers and information consumers depending on their individual ratio of browsing and bookmarking activities.

We compare findings in our experiments to the preceding analyses where possible. However, all three works focus on local social tagging systems located inside the network of a particular company. Therefore, they represent private systems where users only tag resources inside the company's field of interest, and hence the context in which the results were obtained compares only to a certain degree to that of an open, public tagging system. Millen et al. [2007] already note that the company's internal services require their users to use corporate identities instead of pseudonyms, which is typically not the case in public systems. Contrarily, in this work, we focus on the publicly available system BibSonomy to overcome this limitation. This leads to some interesting deviating insights that are discussed in Section 4.4 regarding the social and popularity aspects. We extend the analyses in Millen and Feinberg [2006], Millen et al. [2007], and Damianos et al. [2007] by investigating a series of aspects of social tagging systems and also benefit from long-time log data allowing us to get a clearer overview of actual user behavior in an already established social tagging system.

Finally, a recent study by Lorince et al. [2015] analyzed aspects of retrieval in the tagging system Last.fm. They did not explicitly use log file analysis but instead profited from usage information that is made publicly available through the Web interface of Last.fm. Comparing the use of tagged and untagged content, they conclude that tags increase retrieval only to a minor extent. Since not tagging resources is rarely an option in a tagging system, their analyses are not directly comparable to ours and are rather specific to Last.fm. However, where possible (although the analyses here are different), we compare our results to theirs and come to similar conclusions.

The social bookmarking system BibSonomy. BibSonomy is a social bookmarking system where users tag publication references and Web site bookmarks. The service started publicly in 2006 and has since been continuously developed. BibSonomy has been subject to many diverse scientific studies of which we briefly mention a few in the following. Noy et al. [2008] used BibSonomy as one candidate in the Collaborative Knowledge Construction (CKC) Challenge with the goal to infer users' expectations toward different tools "for collaborative construction of structured knowledge." In that challenge, it was found that users share several expectations toward such systems, such as a suitable Web interface, private and public spaces, or export facilities. Several users expressed their wish that tagging should also be integrated into further services (other than specific tagging systems). BibSonomy has also been used to evaluate and test various recommender systems for tag recommendations (e.g., Jäschke et al. [2008]) or for resource recommendations (e.g., Bogers [2009]). Borrego and Fry [2012] used the publicly visible publication posts of BibSonomy to analyze the use of publications by scholars. They found that in BibSonomy, articles that appeared in commercially published journals significantly outnumber articles from others sources like science-minded repositories or open access journals. Other studies have investigated multilingual tagging [Stiller et al. 2011], the consistency of community structures extracted from of various interaction networks between users (evidence networks) [Mitzlaff et al. 2011], spam detection [Krause et al. 2008], or privacy protection in tagging systems [Doerfel et al. 2013]. BibSonomy's architecture and several analyses of its posts are summarized in Benz et al. [2010].

3. DATASET

The datasets used in this article are based on Web server logs and database contents of the social tagging system BibSonomy. This service allows users to store, tag, and share links to Web sites and (scientific) publications. Storing a resource (Web site or publication) creates a post—that is, a structure containing the data of the resource, the user who created it, and the tags that were assigned to this resource by that user. BibSonomy offers for example the following options to query for posts:[6] Users can request to see all posts with one or several tags[7] or all posts of a specific user or group,[8] and they can also use a combination of user and tag restrictions.[9] Figure 1(a) shows the bookmarks and publications of the user "hotho" with the tag restrictions "web" and "mining". Users can also use a full text search for retrieval.[10]

For each resource, BibSonomy has a page that lists its tags and users from all posts.[11] Publication posts have a *details* page[12] (for an example, see Figure 1(b)) that shows the metadata of the publication (as entered by the user who created the post) and offers export options.

Posts of bookmarked Web sites can also contain metadata (like a description of the Web site), but requests to a bookmark post are usually directly to the bookmarked Web site and thus external requests. Such requests are not recorded in BibSonomy's server logs, and therefore we must restrict some experiments exclusively to publication requests.

In BibSonomy, users can form groups or declare friendships to other users. Both friendships and groups are used in the visibility concept of posts. We will take a look at both in our discussion of the social aspect in Section 4.1. BibSonomy offers many further features, like discussion forums that exceed the usual tagging system functionality. Therefore, such features have been excluded from our experiments.

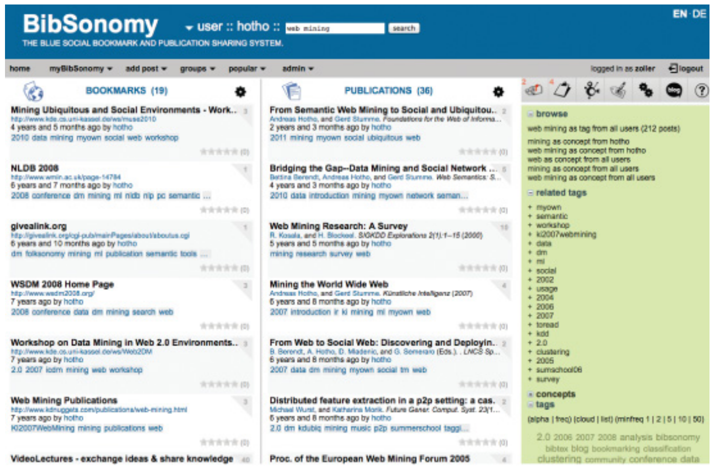
Due to its high rank in search engines, BibSonomy is a popular target for spammers. Spammers are users who store links to advertisement sites to increase their visibility on the Web. BibSonomy uses a learning classifier [Krause et al. 2008] and manual classification by the system's administrators to detect spam. In all experiments, we only used data of users who have been classified as nonspammers.

We restricted the datasets to data that had been created between the start of BibSonomy in 2006 and the end of 2011, as in early 2012 the login mechanism was modified, which introduced significant changes to the logging infrastructure. With this article, we make anonymized datasets of logs and posts available to researchers.[13]

3.1. User and Content Dataset

We use tagging data from BibSonomy's database (i.e., users with their posts, containing resources and tags) and all data about groups and friendships. In the considered time frame, 852,172 people registered a user account, of which 17,932 were classified as

[6]For details on the BibSonomy URL schema, see http://www.bibsonomy.org/help_en/URL-Syntax.
[7]http://www.bibsonomy.org/tag/web+mining.
[8]http://www.bibsonomy.org/user/hotho.
[9]http://www.bibsonomy.org/user/hotho/web+mining.
[10]http://www.bibsonomy.org/search/Semantic+Web.
[11]http://www.bibsonomy.org/bibtex/157fe43734b18909a24bf5bf6608d2a09.
[12]http://www.bibsonomy.org/bibtex/157fe43734b18909a24bf5bf6608d2a09/hotho.
[13]http://www.kde.cs.uni-kassel.de/bibsonomy/dumps/.



(a) User page with tag restrictions (tags: “web” and “mining”)

(b) Publication details page

Fig. 1. Screenshots of BibSonomy’s Web interface. Typical resource list layout showing both bookmarks and publications side by side (a) and a publication details page that lists all information entered by a user for a specific publication (b). The screenshots were taken in September 2014. In late November 2014, a new layout was introduced.

nonspammers. They created 551,606 bookmark posts and 2,391,721 publication posts using 250,344 tags.

3.2. Request Log Dataset

The BibSonomy log files include all HTTP requests (caching is disabled) to the system including common request attributes like IP address, user agent, date, and referer, as well as a session identifier and a cookie containing the name of the logged-in user. Out of the more than 2.5 billion requests, we used only those from logged-in nonspammers and additionally filtered out requests to extra resources including CSS, JavaScript, and image files, as well as requests from Web bots (using a heuristics comparing

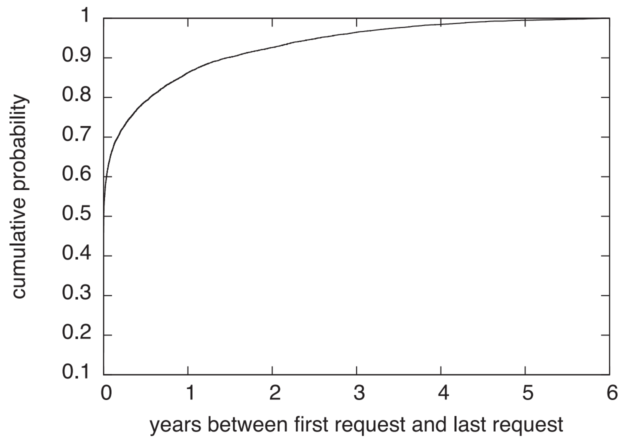


Fig. 2. The cumulative distribution function of the time difference between the first request to BibSonomy and the last request for all users classified as nonspammers.

user agents to those of known bots in various online databases). Furthermore, we removed pages that were irrelevant to our study (like help pages and administration pages). Additionally, to ensure capturing of only actual user behavior, we used a simple heuristic based on the status code of a request's referer to filter automatic redirects caused by the system instead of by choice of the user (e.g., redirects to the personal user page after editing a post).[14] The remaining dataset contained about 3.1 million requests.

For each considered user in the dataset, we calculated the elapsed time between the first request to BibSonomy and the last request made by a user. We only looked at requests to pages analyzed in at least one of our experiments. The resulting cumulative distribution function of the time differences is shown in Figure 2. Although more than half of the users (about 54%) stopped using BibSonomy after 1 month, more than 20% of the users interacted with the system for more than 6 months. About 14% of the users remained active for more than 1 year.

4. FOUR ASPECTS OF USER BEHAVIOR

In this section, we present our results. For each aspect, we (i) make the research question behind it explicit, (ii) review evidence and arguments related to that aspect from the literature, (iii) present the results of our research, and (iv) discuss our findings.

4.1. The Social Aspect

With the *social aspect*, we investigate whether users of a social tagging system (re)use resources that have been shared and tagged by others, either by viewing them or by copying them into their own collection.

4.1.1. Debate in the Literature. The social aspect of tagging has been subject to controversial discussion in the past. It has been praised and disputed already early in the history of tagging systems. Mathes [2004] stated that folksonomies could "lower the barriers to cooperation," and Weinberger [2005] named it as one of two aspects that "make tagging highly useful." Marlow et al. [2006] presented an early model for social tagging systems where they argued that social relations between users are a critical

[14]This is an improvement over the preprint version of this article [Doerfel et al. 2014b] explaining small quantitative (but never qualitative) differences in some results.

element. The authors point out that social interaction connects bookmarking activities of individuals with a rich network of shared tags, resources, and users. Furthermore, Millen and Feinberg [2006] found out that approximately 74% of all page requests in Dogear—an internal social tagging service at IBM—refer to content that was bookmarked by other users. In contrast to that, Damianos et al. [2007] noticed that users are looking more at their own (70%) rather than other users' collections in their system onomi. Yet it is not self-evident that similar observations can be made for public tagging systems, where users use the system without direct company guidance that might influence their behavior. However, users may choose to use such systems for individual purposes only, such as to create their own collections, and thus ignore the resources of other users. Vander Wal [2005] already pointed out that personal information management may be one of the main reasons people use social tagging systems, which was also emphasized by Terdiman [2005]. Porter [2005] claimed that "Personal value precedes network value: Selfish use comes before shared use." Halpin et al. [2007] list difficulties in users' vocabularies (ambiguity, synonyms, etc.) that might hinder the retrieval of other users' content. A user survey by Heckner et al. [2009] found that approximately 70% of all users store resources in tagging systems comparable to BibSonomy mainly to retrieve them themselves, not particularly to share them (in contrast to systems where images or videos are shared). However, it is also noted that "even users of systems who claim that personal information management is very important for them, state that sharing is also part of their motivation of using the systems" [Heckner et al. 2009]. Although this survey takes the perspective of *motivation* for posting, we will rather take the viewpoint of the *use* of posts. We conducted a first evaluation of social behavior in Doerfel et al. [2014a], which we extend and detail in the following.

4.1.2. Results. We investigate the different forms of interest in the content of other users through three actions. First, visiting content is a sign of interest in the material of others. Second, similarly, copying resources shows a stronger and less casual interest, as it means actively integrating the content into one's own collection. Third, copying tags when copying a post is another sign of interest in the actions of other users. Not only post's resource is appreciated, but also the way it was annotated by the owner of the copied post.

Visiting content. First, we analyze the ownership of visited (retrieved) content. We distinguish between four different ownership categories: (i) *user's own*, meaning that a user retrieved content (posts) explicitly from his own collection; and (ii) *groups and friends*, meaning that users retrieved content explicitly from a group they are a member of or from a user to whom they had declared friendship (and thus made use of an excplicit social tie); and (iii) *other users*, meaning that a user retrieved content from a specific user who was neither the user himself nor a member of any of the user's group nor a friend; and (iv) *general*, meaning that a user retrieved content without specifying a particular user (e.g., a request by tag).[15] Table I shows the number of requests (and their shares) in each of these ownership categories during the years 2006 through 2011. We can observe that about two thirds of all requests of logged-in users target their own pages. Users visit other pages in about 35% of the requests to look at either general pages (i.e., pages containing posts of several users (about 17%)) or content of individual other users or groups (about 17%). Thus, requests to groups and friend pages are both rather infrequent (only about 3%), indicating that these particularly social features (in Bibsonomy, they are used to control the visibility of posts) play only a minor role. Further, the share of visits to content of others is below 74%, reported

[15]The BibSonomy landing page was considered separately, although it lists the most recent posts in the system, and thus could be considered in the category *general*. However, many users just visit that page to start their session using the input fields provided on that page. Thus, they ignore the displayed resources.

Table I. Content Visits

	Requests (#)	Requests (%)
User's Own	884,525	65.47
Groups and Friends	44,694	3.31
Other Users	188,057	13.92
General	233,710	17.30
Landing Page	296,090	—

Note: Requests are counted in four categories of ownership: requests to the logged-in user's own content, to content from group members or friends, to content from other users, or to general (non–user-specific) pages. Requests to the landing page (see footnote 15) are excluded from the calculation of the shares.

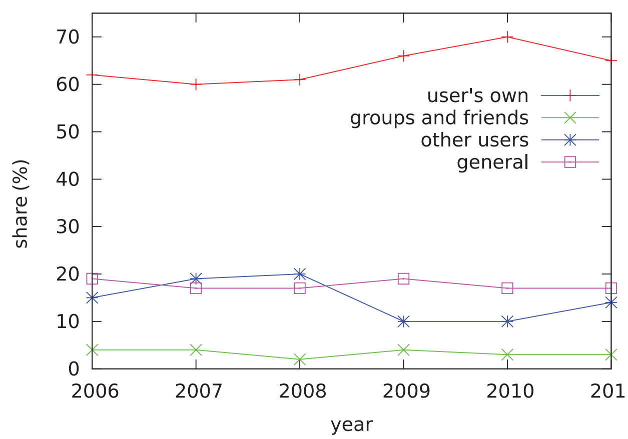
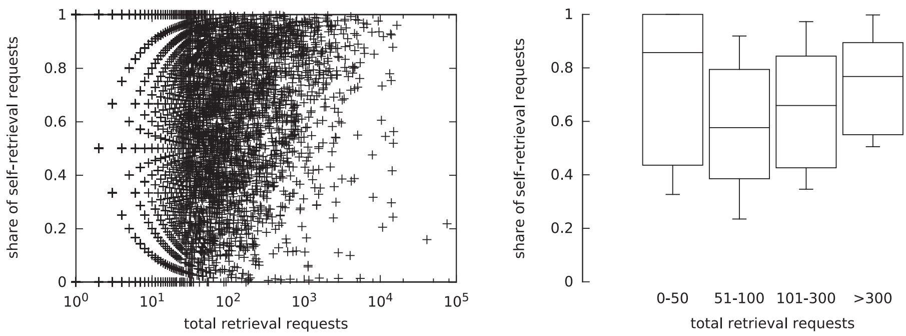


Fig. 3. Content visits over the years. Per year, shown are the shares of requests to content of different ownership categories.

by Millen and Feinberg [2006] for a company's internal tagging systems, but similar to the reported share by Damianos et al. [2007]. 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—more than one third of all retrieval requests—of the interaction with the system. Thus, explicit social ties between users (groups or friends) play only a minor role.

The previous results are aggregated both over time and over the set of all users. Therefore, next we examine them first over time and then as distribution over the users. Figure 3 shows the shares of requests to content in the four different ownership categories per year. We can observe that requests to content in the user's own collection account for the largest share in every year, roughly between 60% and 70%, thus dominating the other three categories. The shares of requests to general pages and of requests to groups or friend pages fluctuate only a little over the years.

The share of requests to other users exhibits a sharp drop—the share roughly cuts in half—from 2008 to 2009. This drop coincides with slight increases in all other shares, most noticeably that of requests to a user's own content. We can only speculate about possible reasons for that effect. One possible reason could be changes in the user interface or added features. BibSonomy is continuously developed, and thus users might prefer some new features over previous ones. Another plausible hypothesis is that since the system had been available for 3 years, users had the time to create large collections. Thus, such users spend more and more requests on their own collection to



(a) Scatter plot with logarithmic axes (b) Aggregated box plot

Fig. 4. Retrieval intensity versus self-retrieval share. Plotted are the total number of retrieval requests versus the share of requests to own content among those requests for each user (a) and with users grouped into four buckets (b) by their total number of requests.

retrieve its content or to navigate through it before they find the resources they were looking for.

As this hypothesis concerns the motivation of users, one would actually have to ask the users to verify it, which is beyond the scope of this work. However, we can look for further evidence by examining the behavior of individual users. We compare a user's share of requests to his own content to the intensity of this user's retrieval. Therefore, we determine for each user his total number of retrieval requests (measuring the intensity in which BibSonomy is used by that user to gather content) and the share of requests to his own content among those requests. We can now determine the correlation between these two quantities over the set of all users using Spearman's rank correlation coefficient ρ. We yield $\rho = -0.42$, indicating a negative correlation: users who spend fewer requests on retrieval in total spend a larger share of those request on self-retrieval.

The scatter plot of the two quantities, shown in Figure 4(a), paints a partially different picture. There are users with approximately 50 requests or less who show various individual shares of self-retrieval. However, among the users with more than 50 retrieval requests, we can observe the tendency that with a rising total number of retrieval requests, the focus turns rather on self-retrieval—contrary to the negative correlation that we observed on the full set of users. The plot does not show the number of users that each dot represents, but in fact the set of users with fewer than 50 requests accounts for about 75% of all users. Thus, the left part of the plot represents most users.

For the box plot in Figure 4(b), we grouped the users by their total number of retrieval requests into four buckets. The first bucket contains all users with fewer than 50 requests, the second contains those with 51 to 100 requests, the third contains those with 101 to 300 requests, and the last bucket contains all users with more than 300 retrieval requests. Thus, the first bucket represents the majority of users, whereas the other users are almost equally distributed over the remaining buckets with 8% in the second and fourth bucket each and 9% in the third bucket. Although there is a lot of divergence in all buckets, we can observe two tendencies. Many users with a low total number of retrieval requests (fewer than 50) use the system more frequently for self-retrieval than other users. Among the users with more than 50 requests, we observe the tendency that the share of self-retrieval grows with the number of requests. Indeed, if we measure the correlation only among those users with more than 50 requests, we now yield a positive correlation of $\rho = 0.22$. These findings strengthen the hypothesis

Table II. Copying Resources

Share Category	Bookmarks	Publications	Total
Copied posts	3.5%	17.6%	10.7%
Posts that could have been copied	15.6%	36.9%	26.5%
Copied posts among posts that could have been copied	22.2%	47.7%	40.4%

Note: The table represents shares of posts that were and could have been created as copies of other users' content.

that many users who use the system more than only casually tend to spend a larger share of their time on their own collections. Therefore, long-time users will be more likely to request their own content than that of other users. In addition, the tendency of more active users toward visiting more of their own than other content conforms qualitatively with Millen et al. [2007], who similarly observed that a stronger use of the tagging system usually means that more time is spent on one's own collection.

Copying resources. When users added new posts to their collections, in 10.7% of all cases a bookmark or a publication was copied from another user, as we can see in Table II.[16] Users copied publications (17.6%) more often than bookmarks (3.5%). One reason for this difference might be the fact that users leave the system when they follow a bookmarked link, whereas they stay within BibSonomy when they check out details of a publication. Thus, using a bookmarklet (provided by BibSonomy), while visiting the Web page to be bookmarked, is the easiest way to post a Web site. In contrast, for a publication, clicking the copy button on its details page is the easiest option. We note that the share of 3.5% of copied bookmarks is close to the 2.2% share reported by Millen and Feinberg [2006] for the IBM internal system Dogear, whereas the share for publications (17.6%) exceeds that value by a factor of eight.

Since a resource could only be copied if another user had already posted that resource in BibSonomy, we have to take into account whether posted resources were already present in the system when a user posted them. The last line of Table II shows the share of actually copies among possible copies—that is, the number of posts that have been created as copies divided by the number of posts where the posted resource had already been available in the system. Of all posts that could have been created through copying at the time of their posting, a share of roughly 40.4% (and even 47.7% for publications) has indeed been copied. This can be regarded as a relatively large share, as looking up publications or Web sites in BibSonomy is only one out of many possible ways to find interesting bookmarks and publications on the Web or elsewhere.

Copying tags. Finally, we study whether not only resources but also tags are copied. For 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 user's own vocabulary was used. In 42% of all copies, at least one of the original post's tags was adopted. In the other copy events, 44% of the original posts had only special tags like "imported," which are probably not meaningful to the user copying the post. Similarly to the copying of resources, we thus find that users copy tags in a large number of cases, although in the majority of cases, one's own tags were used.

4.1.3. Discussion. We found evidence for both personal information management and social interaction. In general, the findings fit well to the result from Heckner et al. [2009] that the motivation for posting Web sites and publications is not predominantly

[16]We ignore imports of bookmark or publication lists (e.g., browser bookmark or BibTeX files) because during such transfers of one's own collections to BibSonomy, it would not be meaningful to look for resources in other users' collections.

social. We found only a relatively low share of visits to groups and friends, showing that explicit social features, which allow users to network, play only a minor role. Furthermore, the majority of requests targets content from the logged-in user's own collection. However, although users might not contribute content while particularly intending to share it (like in social networks), we could yet observe evidence that they do profit from the availability of other users' content. 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 exploited.

Fu et al. [2010] showed that users who see the tags of others are biased in their own choice of tags toward the tags they saw or at least to semantically similar tags (semantic imitation). Thus, even the awareness of other users' content can change someone's behavior. In our study, we found that users target the pages of specific other users. Hence, a question for future work will be the motivation to select particularly these users and to find the underlying mechanisms, such as relations between or biases toward users, similarly to the semantic relatedness of tags.

For Webmasters of such systems, our results indicate that it is reasonable to assist users in discovering the content of others, such as through search functionality or recommendations.

4.2. The Retrieval Aspect

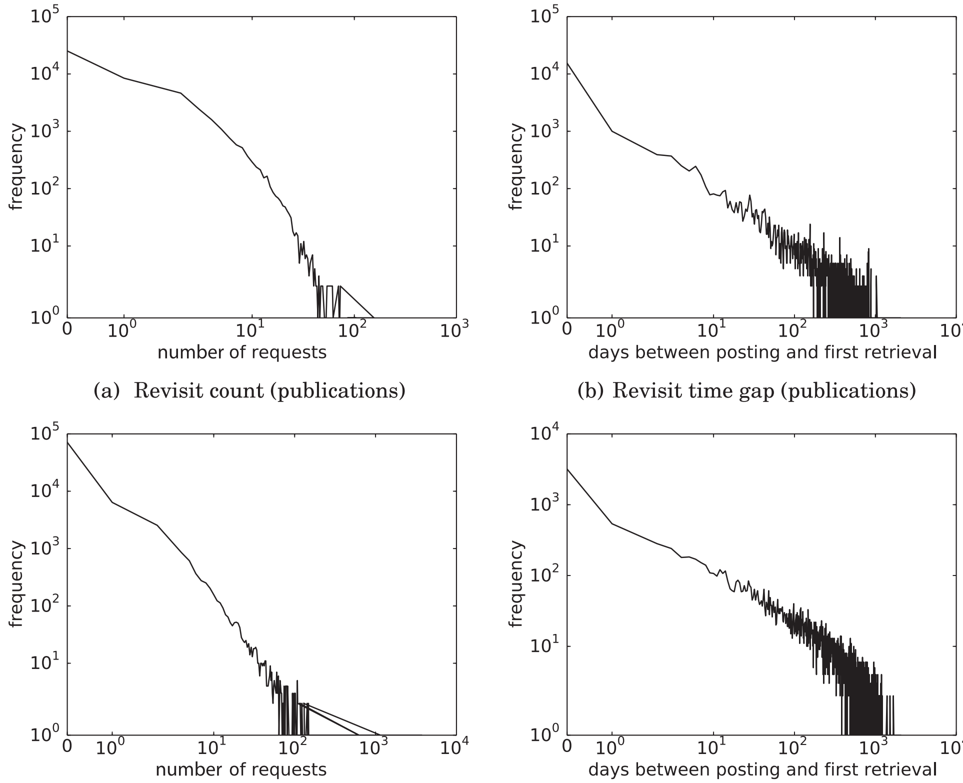
With the *retrieval aspect*, we refer to the notion that tagging systems are used to manage personal collections of resources for their retrieval later on. We investigate to what degree users make use of their resources and tags after they have stored them.

4.2.1. Debate in the Literature. In a study on the use of browser bookmarks by Abrams et al. [1998], it was found that users revisit about 96% of their own bookmarks within 1 year. Since the idea of social bookmarking, in a way, is an advancement of the classic browser-based bookmarking, the question arises whether the retrieval behavior in tagging systems is similar to that reported for browser bookmarks. It was hypothesized already at the beginning of social tagging research that personal information management may be one of the main reasons people use social tagging systems (e.g., Vander Wal [2005]). Furthermore, as mentioned in the previous section, the user survey by Heckner et al. [2009] identified personal management as the main motivation to post resources to systems like BibSonomy. Additionally in the preceding section, we saw that the major part of all retrieval requests targets the respective user's personal collection and that such use of one's own content is part of personal information management.

Regarding tags, the assumption that they are used to retrieve content later has been made several times by Vander Wal [2007], Golder and Huberman [2006], and Glushko et al. [2008]. A survey by Ames and Naaman [2007] found that for several interviewed users, self-organization was a primary incentive for using tags when annotating photos, but they also noticed social aspects as an important influence for tagging as well. Recently, Lorince et al. [2015] observed for the music tagging system Last.fm that using tags only rarely increases retrieval rates: tagged content was listened to about 1.15 times as often as untagged content. This is an indication that tags might not play the role in retrieval that was assumed.

4.2.2. Results. We present statistics about revisiting patterns obtained for both publication posts[17] and assigned tags. More precisely, we investigate how many times users revisit their own posts and tags and also the time difference between the posting of

[17]Requests to bookmarks could not be analyzed, as they target pages outside BibSonomy; therefore, requests for such pages are not recorded in the logs (see Section 3).



(a) Revisit count (publications)
(b) Revisit time gap (publications)
(c) Revisit count (tags)
(d) Revisit time gap (tags)

Fig. 5. Revisitation behavior of users. (a) Number of times users revisit their own publications. (b) Number of days elapsed between the posting of a publication and its first retrieval by its owner. (c) Revisit count for tags. (d) Elapsed days for tags. All four figures are visualized on a log-log scale. Note that in all four diagrams, a zero has been artificially added to the x-axis to visualize the large number of resources and tags that are never visited.

a resource or tag and its first retrieval, counted in days. To give users a reasonable amount of time for revisits, we captured all posts until the end of 2010 and all requests until 2011. This means that each user had at least a whole year to revisit their posted resources and tags. A year is also the time frame for which Abrams et al. [1998] reported that most of the bookmarks kept by users in their browser are revisited at least once.

The results are shown in Figure 5. Approximately 49% of all publications were revisited by their owner at least once. If a publication has been revisited at all, it most often was revisited only once (see Figure 5(a)). Furthermore, we can observe in Figure 5(b) that most of the first revisits to a page took place shortly after the resource had been posted—often on the same day. These visits could well be control visits to check the created post; however, it could also mean that users posted a publication immediately before they used it (e.g., a cited reference from a paper on which they were working). The revisit investigations of tags show a more drastic picture. Only around 17% of tags are used in queries at least once by a user who previously assigned them to a post. In Figure 5(c) and (d), we can observe similar patterns as for publications. If revisited, tags most often have only been revisited once and often shortly after the assignment.

4.2.3. Discussion. In the previous section, we saw that interactions with the personal collection account for the dominant share of users' retrieval requests. Although, according to Heckner et al. [2009], users use the system for later retrieval, we now find that only about half of all publications are revisited. Particularly interesting about this observation is that it does not agree with the work by Abrams et al. [1998] on browser bookmarks, where 96% of all bookmarked resources were revisited in the time span of 1 year. The difference might be a result of several factors. First of all, using a publication is different from revisiting a Web site—many Web sites often renew their content frequently and usually are easier to consume than scientific publications. Moreover, the user survey reported the difficulty of creating and organizing bookmarked resources, whereas tagging systems aim to simplify the process of creating and ordering bookmarks as much as possible. This could imply that users tend to store more simply because the effort is low. Another reason for the lower retrieval rate is certainly that the retrieval of single posts is only one way to make use of one's own collection. Another reasonable way of using stored publications for citation is to mass export (e.g., simply all or many publications in the collection) them into a suitable citation format and selecting the actually used publications offline. Finally, the study on browser bookmarks was conducted in 1998, and it can be expected that user behavior has since changed. Browsers have changed significantly, and they now allow tagging for better structuring of bookmarks. However, the differences are noteworthy, as social bookmarking (particularly of Web sites) can be seen as an advancement of browser bookmarks. In fact, the availability of social bookmarking would surely be a factor in today's browser bookmark collecting habits.

More surprising is the small share of one's own tags used for retrieval. An explanation for this observation might be that it is reasonable to use many tags for a resource to increase the chance of successful retrieval later on. However, for the retrieval itself, only a fraction of these tags is sufficient. Furthermore, in the next section, we will show that using tags is not the dominant way to query BibSonomy. Finally, we note that our results are in line with the conclusions from Lorince et al. [2015], who found that in Last.fm, tagged content is only slightly more often retrieved (1.15 times) than untagged content. Considering that BibSonomy (like most other tagging systems) does not allow untagged resources, these results are not directly comparable. Still, both findings are evidence that the role of tags in retrieval is not as strong as previously assumed.

For Webmasters of a tagging system, our results indicate that visits, and thus the actual retrieval of resources or tags, are an important quantity and should be considered in the assessment of the importance of resources or tags (i.e., in features like ranking of search results or tag recommendations). Until now, the dedicated algorithms typically focused on the number of posts in which a resource or tag occurs (e.g., Jäschke et al. [2008]) than on their retrieval.

4.3. The Equality Aspect

With the *equality aspect*, we focus on the question of whether the three entity sets in a tagging system—the sets of users, resources, and tags—are equally important for navigation and retrieval.

4.3.1. Debate in the Literature. A folksonomy—the structure underlying tagging systems—has been defined as a quadruple $\mathbb{F} = (U, T, R, Y)$ consisting of the sets of users U, resources R, and tags T together with the tag-assignment-relation $Y \subseteq U \times R \times T$ (compare Hotho et al. [2006]). In that model, users, resources, and tags are treated equally and in fact even symmetrically. The folksonomy model has been widely accepted, and many algorithms build on it, such as the FolkRank by Hotho et al. [2006] or the tensor factorization method by Rendle et al. [2009]. Since tag

Table III. Entity Request Shares in BibSonomy

	User	Resource	Tag
Requests (# per entity)	30.33	0.14	1.08
Requests (# total)	543,837	316,582	269,212
Requests (% total)	48.14	28.03	23.83
Requests (# to self)	435,513	217,587	192,737
Requests (% to self)	51.49	25.72	22.79
Requests (# to others)	108,324	98,995	76,475
Requests (% to others)	38.17	34.88	26.95

Note: For each set of folksonomic entities, we report the average number of requests per entity in that set (i.e., dividing the total number of requests to tags by the total number of tags) and the total number and relative share of requests to entities of that set among all requests (total) and among requests targeting content outside one's own collection (to others), such as a request to User X by User Y.

assignments link entities of all three sets together, the idea of the typical folksonomy navigation is that these entities can be navigated by following these links back and forth (e.g., clicking a tag to request all posts to which that tag is assigned). A counterargument to the symmetry of tags, resources, and users is the fact that tag assignments usually occur in groups, which are represented by the posts of a tagging system. Each post is created by *one* user who assigns several tags to *one* resource. Thus, one post usually provides links to the one user and to the one resource, but to more than one tag.

4.3.2. Results. We discuss this aspect in two parts. First we analyze the shares among all requests with respect to the entities they target (either users, resources, or tags). Second, we investigate the probability of transitions between entities of different classes in the users' navigational paths.

Request shares: users versus resources versus tags. As in previous assumptions, we analyze retrieval requests. We split them into requests querying specifically for users, tags, or resources.[18] Requests that did not query a user, a resource, or a tag are ignored. In BibSonomy, requests can contain more than one queried entity. For example, a request can query for a user u and a resource r to retrieve the post that user u created for r. To assign such requests to one of the three groups, we determined the *dominating* entity. In the preceding example, the target is clearly the resource rather than the user. A request containing a user u and a tag t (retrieving posts that u annotated with t) has been categorized as a tag request. The third possible combination (resource plus tag) does not exist in BibSonomy. Through this rule, we rather underestimate the significance of users in the users' navigation behavior. For comparison, next we address another option of counting requests to entities.

We group the set of all requests into three groups: requests to users, requests to tags, and requests to resources. For each of these groups, the respective column of Table III presents a different aggregation of these requests. The first row shows the average number of requests to an entity (e.g., the number of all requests to any tag divided by the total number of tags in the system). The second row shows the actual number of requests in a group, and the third row shows the percentage that these requests account for among all retrieval requests of the three groups combined. The other rows of the table are similar to rows two and three; however, the requests considered are restricted to either (a) requests to the user's own collection (to self), in rows four and

[18]Note that requests to resources are generally underrepresented due to the lack of recorded requests to bookmarks (see Section 3).

five (e.g., a user requested all of his own posts with a particular tag), or (b) requests to content outside one's own collection (to others), in rows six and seven (e.g., requests to other users or requests without a specific user).

From the second row of Table III, we can clearly see that the total request numbers are not equally distributed. There are about 2.1 times more requests to specific users than to specific tags. The share of resources (28.03%) is slightly higher than that of tags (23.83%). One might assume that this imbalance toward users is due to a similar imbalance in the numbers of all users, all resources, and all tags. However, from the average requests per entity, we can deduce that this is not the case. On average, an individual user is requested many more times than a tag, which in turn is requested much more often than an individual resource.

Using the restrictions (to self) and (to others) of the set of requests, we can distinguish between the use case of retrieving one's own content (personal management) and browsing the folksonomy.

From the figures in Table III, we can observe that for requests to one's own collection (to self), the share of requests by user increases slightly compared to the share among all requests (total): from 48.14% to 51.49%. This is not surprising, as all requests to one's own collection must necessarily be requests to a user: to oneself. Thus, among these, tag or resource requests are those that have two targets (a user and a tag or resource) and have been classified into one of either category by the rule mentioned earlier. Looking at the requests to content outside one's own collection (to others), we observe that the share of user requests drops compared to the full set of requests: from 48.14% to 38.17%. Nevertheless, the queries for users still outnumber those for tags but to a lesser extent. Thus, we find that users are the most frequently used means of navigation in the folksonomy. This effect is visible in both use cases—personal management and browsing of other users' content.

It is also interesting to note that the ratio between requests to tags and resources is roughly comparable over all three request restrictions: 1.18 (total) versus 1.13 (to self) versus 1.29 (to others). This indicates a comparable user behavior within one's own collection and within the content of other users.

With the preceding assignment of each request to one dominating entity, we chose a rather conservative approach that tends to *underestimate* the relevance of requests to users. As mentioned earlier, we counted each request with multiple queried entities only once, for the dominant entity in that request, which in all cases was either a tag or a resource (never a user). Therefore, in a similar experiment, we directly counted the requested entities, meaning that a request with a requested resource and requested user was counted for both user and resource. The results are qualitatively comparable and show an even stronger imbalance toward users—about 63% of the requested entities were users.

Request shares over time. The figures in Table III indicate that users are the main means of navigation in BibSonomy rather than tags, as one might have expected in a *tagging* system. To gain deeper insights, in the following we look at similar figures over time. We investigate the full set of retrieval requests—that is, next to requests to users, resources, or tags, we also count all other retrieval requests (e.g., requests to the full-text search,[19] requests by author or by a publication's BibTeX key, or requests to pages listing the recently most popular resources or tags). Figure 6[20] shows how these shares

[19]The full text of a publication in BibSonomy is its collection of metadata.

[20]The shares in Table III are computed only of those retrieval requests that target either users, resources, or tags. Through the inclusion of the "other" requests, the shares in Figure 6 are thus not directly comparable to those in the table.

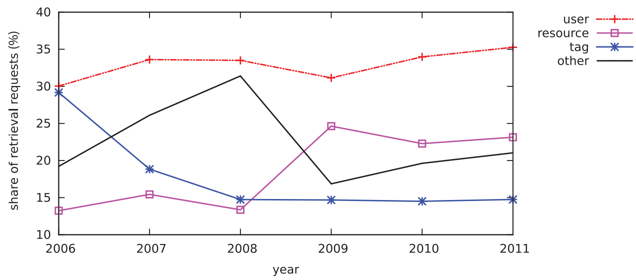


Fig. 6. Request shares over time. Plotted are the shares of retrieval requests each year to the different classes of entities: users, resources, and tags. All remaining retrieval requests that target neither entity are summarized as "other."

Table IV. Tag-Based Retrieval Versus Search in BibSonomy

	Tag	Search
Requests (# total)	269,212	141,843
Requests (% total)	65.49	34.51
Requests (# to self)	192,737	20,663
Requests (% to self)	90.32	9.68
Requests (# to others)	76,475	121,180
Requests (% to others)	38.69	61.31

Note: We report the total number and relative share of requests to either tag pages or to search among all requests (total), among requests targeting one's own content (to self), and among requests targeting content outside one's own collection (to others), such as a request to User Y by User X.

develop over time. We can observe that in the first year, the number of requests to tags and to users were almost equal. The share of requests to users stayed roughly the same over the years. However, within 2 years, the (relative) share of requests to tags drops significantly and stays almost constant afterward. At the same time, the share of other requests increases. This indicates that users found other means of navigation rather than using tags for retrieval. Since BibSonomy has constantly been extended with new features, it is natural that users would use these new features (e.g., a full-text search) and therefore others like tags to a lesser extent. Rather surprising, however, is that only the share of requests to tags shrinks.

Request shares: tags versus search. One feature that is particularly suitable to retrieve resources is the full-text search. As it is not part of the folksonomy structure, which underlies a tagging system (see earlier), we have omitted it in the previous analyses. However, since we saw that the role of tagging is not as dominant as one would expect, in the next analysis we compare the requests using tags to those to the full-text search. Table IV shows, similarly to Table III, the absolute number of requests together with the shares (among requests to either search or tags). Again, we distinguish between such requests in general (total), requests to one's own collection (to self), and requests to content of others (to others). We can observe that overall, requests using tags outnumber those using the search by roughly 2:1. However, the choice of either means depends clearly on the scope of the search. If one's own collection is targeted, requests using tags outnumber the full-text search by roughly 9:1. Yet requests to content outside one's own collection are more often conducted with the full-text search (about 1.6 as many requests as with tags). These figures indicate that users indeed make use of tags but rather when they retrieve their own resources; after all, it is

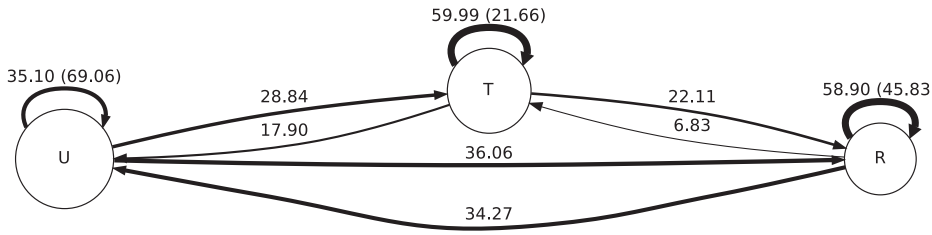


Fig. 7. Transition probabilities between the three entities of the folksonomy. The nodes user (U), resource (R), and tag (T) correspond to the columns in Table III, and their size reflects the total number of requests to entities of these sets. The edges represent transition probabilities from a page of one entity set to another. The percentage in brackets on the self-loop edges describe the fraction of pagination or reload effects on navigation.

"their" tags they used to annotate them with in the first place. For retrieving content that was annotated by others, the full-text search is more often the preferred option.

Transition probabilities. Next we look at navigational transition probabilities between the entity categories users, resources, and tags (e.g., the probability of requesting a tag after requesting a user). We determine the transitions using each requests' HTML referer attribute and compute first-order Markov chain probabilities from one entity set to another. Figure 7 shows the results as a graph. We can observe that self-transitions are dominant, suggesting that users tend to stay with the same type of (requested) entity in their navigational paths through BibSonomy. This observed share of self-transitions is consistent with findings in social networks by Schneider et al. [2009], who also observed that users tend to stick with the same feature in consecutive requests.

Aside from that, there are a lot of transitions from user pages to resource pages and tag pages. This is not surprising, as user pages consist of listings of a user's resources, which can be reached with a single click. This also explains the transitions back to user pages and symbolizes the "browsing" in the system. The exception to that is that there are few transitions from a resource page to tag pages, meaning that users only rarely seem interested in resources with the same tags as the resource at hand.

We also looked at the fraction of actual page self-transitions, where the referer and actual URL of a request are the same. This effect occurs typically from page reloading or pagination effects, such as when a user views the next 20 elements of a longer list of publications. In Figure 7, these fractions (of the category self-transitions) are represented by the percentages in brackets. For example, about 69% of the requests that lead from a user page to a user page actually lead from a user page to itself again. This might be explained by the fact that most users' collections exceed the amount of items that are displayed at the same time, so users have to "turn the page" to view the next items. Interestingly, this effect is greatly diminished on tag pages, where only 22% of tag-to-tag page transitions are actual self-transitions. This could mean that many transitions lead from a tag page to a more refined tag page (e.g., by selecting an additional tag in the following request), with only a minor amount of pagination-related navigation. The relatively high amount of self-transitions on resource pages (about 46%) mainly stems from exporting the resource in a given format. This means that users tend to visit other publications directly after visiting a resource page in only roughly one of four cases.

4.3.3. Discussion. We have observed a strong *inequality* between the use of the three folksonomy entities of users, tags, and resources. Although the numbers of requests to tags and to individual resources are similar, they are dominated by the requests to user pages. This is surprising, as there are fewer user pages than tag or resource pages available in BibSonomy. When discussing navigation within folksonomies, resources are usually regarded as targets of queries. As navigational means to find or retrieve theses resources, often tags—rather than users—receive the larger interest, as they can function as resource descriptors. In BibSonomy, however, it seems that the user pages are the main means of navigation, and tags are mainly used to navigate through one's own collection rather than through the system. These findings are also additional evidence for the social aspect, which we discussed in Section 4.1. Not only do users show interest in the content of others, they also use the other users as an important means of navigation. From the transition probabilities, we saw that especially navigation from resources to tags (and thus to potential further resources to the same tag/topic) is rather rare. The unexpected observation that tags do not play the main role in users' navigation behavior has consequences for Webmasters who run and design such systems. Algorithms like FolkRank, which model the transitions between entities, need to be revisited. There, transitions between users, tags, and resources are modeled with equal probabilities, which—as we found out—does not reflect actual user behavior properly.

4.4. The Popularity Aspect

The *popularity assumption* concerns the practice of measuring an entity's popularity by counting the number of posts in which it occurs. We investigate to what extent this popularity of folksonomic entities—the number of posts in which a user, a resource, or a tag occurs or its frequency distributions—matches similar properties in requests.

4.4.1. Debate in the Literature. In tagging systems, the notion of popularity is exploited in several ways: (i) special "popular" pages summarize the most frequently posted resources or tags; (ii) next to a resource, the number of posts it which it occurs is shown; (iii) users' profile pages often show the number of their posts; and (iv) several algorithms for the recommendation of tags [Jäschke et al. 2008] and resources [Bogers 2009] suggest the most frequently used entities. Perhaps most prominently, tag frequency is exploited in tag clouds, where the frequency of a tag corresponds to its font size and, particularly, rare tags sometimes are not displayed at all. Brooks and Montanez [2006] point out that it is taken for granted that the tags a user assigns are the same as those a reader would select. Hence, the authors identified the relationship between the task of article tagging and information retrieval as an open question to investigate. In a user study by Sinclair and Cardew-Hall [2008], it was found that tag clouds are perceived as visual summaries of resources and that clicking in tag clouds requires less cognitive effort than entering search queries. This indicates that the size of a tag is indeed relevant for users in their query behavior, but to the best of our knowledge, the correlation between tag usage in posts and requests has not yet been investigated in a large-scale scenario other than for the company system Dogear [Millen and Feinberg 2006], for which a correlation of 0.67 between the frequencies of a tag in posts and in requests is reported. Contrary to the often assumed connection between tags' popularity in posts and their importance in retrieval, Lorince et al. [2015] found that it is rather the more idiosyncratic, less often used tags that lead to higher retrieval rates.

Regarding overall behavior, Cattuto et al. [2007] noted that frequencies of entities in posts follow a heavy-tailed distribution—mostly clean power law fits. Power law functions are known to exhibit scale invariance and are mostly explained by the Yule process, which is also known as preferential attachment.

4.4.2. Results. Since tag clouds are one of the most popular applications of popularity in tagging systems, we begin the investigation of the popularity assumptions by looking at tags. Afterward, we analyze the same questions for resources and users.

Tags. We start the discussion of popularity of tags by analyzing their distributions of frequencies in the request logs ($D^{req}_{F,Tag}$) and in the posts ($D^{post}_{F,Tag}$).[21] More precisely:

—$D^{req}_{F,Tag}(k)$ counts how many tags have been *requested*[22] exactly k times (e.g., $n = D^{req}_{F,Tag}(k)$ means that exactly n tags have been requested exactly k times), and

—$D^{post}_{F,Tag}(k)$ counts how many tags have been *assigned* to exactly k posts (and thus constitutes the usual node degree distribution described in Cattuto et al. [2007]).

Both distributions are shown in Figure 8(a).[23]

The first observation is that $D^{post}_{F,Tag}$ dominates $D^{req}_{F,Tag}$, meaning that in total there are more tag assignments than requests for tags. Since tag frequency distributions in posts ($D^{post}_{F,Tag}$) are known to be heavy tailed [Cattuto et al. 2007]—mostly power law—it was to be expected that the distribution of tag frequencies in the request logs ($D^{req}_{F,Tag}$) has similar properties. To confirm this, we first fitted the power law function ($y = cx^{-\alpha}$ where $x > x_{min}$) to the empirical data using the methods of Clauset et al. [2009]. Next we compared the corresponding fit to the exponential function as a lower barrier for heavy-tailed distributions, as well as other heavy-tailed probability distributions, namely the log-normal function and the power law function with an exponential cutoff (which means that for large x values, the function deviates from the typical power law function). We visualize the empirical distributions, the best power law x_{min} values (vertical lines), and the corresponding fits in Figure 8(b) for both $D^{post}_{F,Tag}$ and $D^{req}_{F,Tag}$.[24] For the fits of the power law function, we obtained $\alpha = 1.98$ and $x_{min} = 44$ for $D^{post}_{F,Tag}$, and $\alpha = 1.89$ and $x_{min} = 2$ for $D^{req}_{F,Tag}$. The distributions are similar with regard to their slopes α. Noteworthy is the higher result of x_{min} for $D^{post}_{F,Tag}$ (in contrast to the small value for $D^{req}_{F,Tag}$), indicating that the power law fit only holds for a smaller portion of the distribution (the tail). Visual inspection suggests that there are slightly fewer tags with low frequencies than one would expect in a power law distribution. Although an in-depth analysis of this phenomenon is beyond the scope of this work, we can speculate that it might be a consequence of the use of tag recommenders that typically suggest tags that are already frequently used, leading to an ignorance of low-frequency tags.

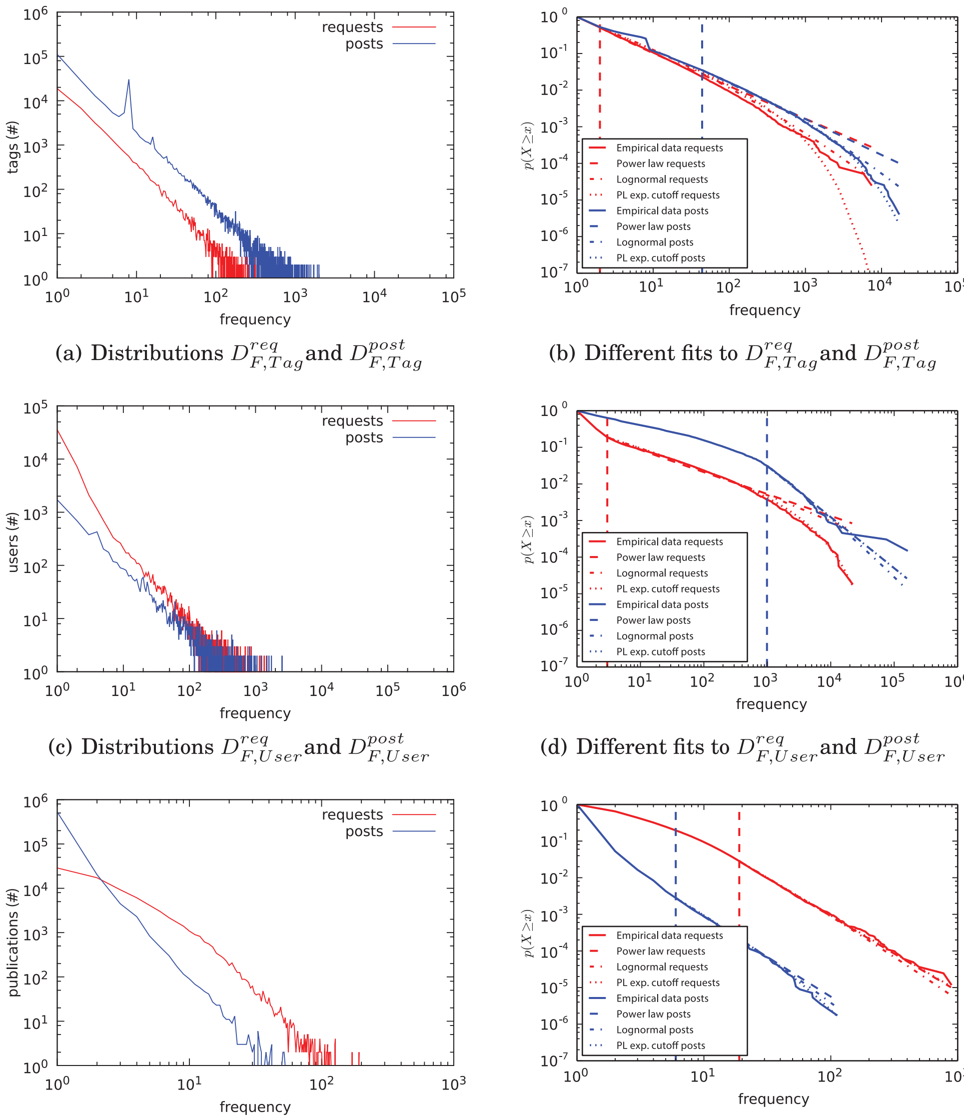
A comparison between the fits to the different distributions showed that the power law function is a statistically significantly better fit to the data than the exponential fit. Both the log-normal and the power law function with an exponential cutoff are also good fits to the data, confirming our assumption about heavy-tailed distributions, and they are even slightly better fits to the data compared to the pure power law function, as one can see in Figure 8(b). This can be explained by the slight decay in the distributions—visible where the line of the empirical distribution ($D^{req}_{F,Tag}$ at $\approx 10^2$ and $D^{post}_{F,Tag}$ at $\approx 10^3$) falls below the straight line of the respective power law fit. Similar to the explanations by Mossa et al. [2002], which are also discussed by Cha et al. [2009], this may be reasoned due to information filteringm which might hinder preferential attachment.

[21]We ignore posts from two users who are known to only automatically create posts from publication catalogues to provide more content in the system.

[22]In the request distribution, we do not distinguish between requests made by clicking on tags (e.g., in a tag cloud or next to a post) and those made by entering them directly into the tag search field, as these types of requests are indiscernible in the logs.

[23]A close investigation of the notable peak in the distribution $D^{post}_{F,Tag}$ at frequency 8 reveals that this anomaly is due to the activities of one single user, who used 28,989 tags exactly eight times. We therefore ignore the peak in the following discussion.

[24]For better visibility, we omitted the (weak) exponential fit.



(a) Distributions $D^{req}_{F,Tag}$ and $D^{post}_{F,Tag}$

(b) Different fits to $D^{req}_{F,Tag}$ and $D^{post}_{F,Tag}$

(c) Distributions $D^{req}_{F,User}$ and $D^{post}_{F,User}$

(d) Different fits to $D^{req}_{F,User}$ and $D^{post}_{F,User}$

(e) Distributions $D^{req}_{F,Pub}$ and $D^{post}_{F,Pub}$

(f) Different fits to $D^{req}_{F,Pub}$ and $D^{post}_{F,Pub}$

Fig. 8. Frequency distributions in requests and posts. In log-log scale, displayed are the frequency distributions for tags in requests ($D^{req}_{F,Tag}$) and for tags in posts ($D^{post}_{F,Tag}$) (a), and fits of the respective complementary cumulative probability distributions to different standard cumulative probability distributions (the vertical lines indicate the corresponding x_{min} values) (b). Accordingly, (c) shows the frequency distributions of users ($D^{req}_{F,User}$ and $D^{post}_{F,User}$) and (d) shows the corresponding fits, whereas for resources, the distributions $D^{req}_{F,Pub}$ and $D^{post}_{F,Pub}$ are shown in (e) and their fits in (f).

Table V. Correlation and Divergence of Request and Tag Distributions

Requests	Posts	r	ρ	JS_2
$D^{req}_{F,Tag}$	$D^{post}_{F,Tag}$	0.968	0.596	0.052
D^{req}_{Tag}	D^{post}_{Tag}	0.420	0.059	0.440
$^{\emptyset}D^{req}_{Tag}$	$^{\emptyset}D^{post}_{Tag}$	0.414	0.517	0.271
$D^{req}_{F,User}$	$D^{post}_{F,User}$	0.942	0.242	0.197
D^{req}_{User}	D^{post}_{User}	0.092	0.548	0.492
$^{\emptyset}D^{req}_{User}$	$^{\emptyset}D^{post}_{User}$	0.081	0.712	0.471
$D^{req}_{F,Pub}$	$D^{post}_{F,Pub}$	0.823	0.803	0.329
D^{req}_{Pub}	D^{post}_{Pub}	0.554	0.032	0.707
$^{\emptyset}D^{req}_{Pub}$	$^{\emptyset}D^{post}_{Pub}$	0.609	0.252	0.152

Note: Pearson's correlation coefficient r, Spearman's rank correlation coefficient ρ, and the Jensen-Shannon divergence JS_2 for pairs of distributions are shown. In each row, a distribution D^{req}_{Entity} (Entity is either Tag, User, or Publication (Pub)) of requests (or their frequencies ($D^{req}_{F,Entity}$)) is compared to a distribution D^{post}_{Entity} of posts (or their frequencies ($D^{post}_{F,Entity}$)).

However, we need to keep in mind that there is only a slight decay visible. Nevertheless, detailed investigations regarding this cutoff are necessary for a better understanding of this behavior. By and large, we can observe similar processes of how users post tags and how they request them—processes yielding heavy-tailed distributions.

Further, we directly compare $D^{post}_{F,Tag}$ and $D^{req}_{F,Tag}$ to each other using Pearson's correlation coefficient r and Spearman's ρ.[25] From the first row in Table V, we can observe that the Pearson and Spearman correlations are high. An explanation for the smaller Spearman's ρ value is the fluctuation in the distributions (see Figure 8(a)), where the number of tags no longer decreases monotonously with increasing frequency. Finally, a comparison of the distributions using the Jensen-Shannon divergence JS_2 confirms similarity.

In the tag frequency distributions, we found similarity in the way users use and request tags. As a next step, we analyze the tag popularity on the level of individual tags to see whether there are similarities regarding which tags users assign and request. Particularly, we look at the distributions D^{req}_{Tag} and D^{post}_{Tag}, where

—$D^{req}_{Tag}(t)$ is the number of *requests* to a tag t (e.g., $n = D^{req}_{Tag}$("web") means that the tag "web" has been requested exactly n times), and
—$D^{post}_{Tag}(t)$ is the number of *posts* in which tag t occurs.

Figure 9 shows the scatter plot of these two tag distributions, where each point in the diagram denotes one tag t with its number of requests $D^{req}_{Tag}(t)$ and its number of posts $D^{post}_{Tag}(t)$ as coordinates. We can immediately see that despite the similarity in the behavior of tag frequencies, there are enormous differences on the level of individual tags. Only for very frequent tags (more than 100 requests) could one presume a correlation between both frequency counts. To quantify the effect, the second row of Table V shows the correlation coefficients and the Jensen-Shannon divergence for the two distributions $D^{req}_{Tag}(t)$ and $D^{post}_{Tag}(t)$. Different from previous distributions, we can

[25]Note that all correlation results in this section are statistically significant with a p-value below 0.05, which is why we do not directly report it explicitly for each calculation.

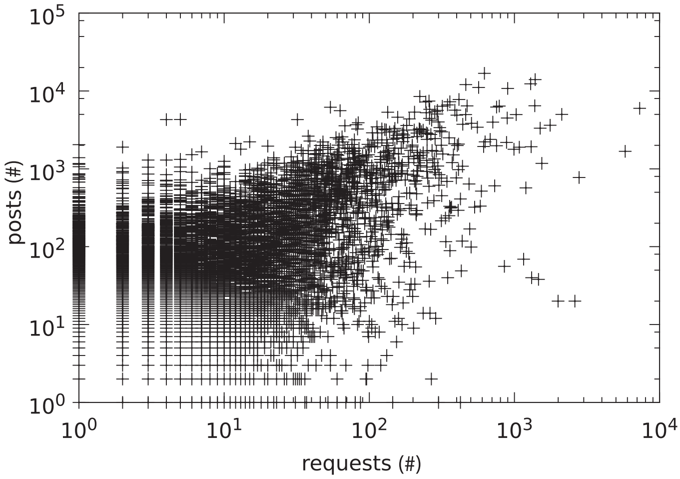


Fig. 9. Occurrences of tags in requests and in posts. The scatter plot in log-log scale of the numbers of requests to a tag t versus the number of posts in which a tag t occurs. Only for higher frequencies do the number of requests and posts of a tag appear to be correlated.

observe rather low correlation and a much higher divergence. This means—contrary to popularity assumption—that the number of posts to which a tag is assigned and the number of times a tag is queried are only mildly correlated. The found correlation of $r = 0.42$ is also lower than the one reported for the company system Dogear (0.67).

A closer look at the log data revealed that many tags used in posts were never queried at all, and several tags had been queried but were never assigned to any post. Therefore, we look at similar distributions as before, but we specifically ignore tags that only occur in one of the two tag distributions. We yield distributions ${}^{\emptyset}D^{req}_{Tag}$ and ${}^{\emptyset}D^{post}_{Tag}$, reducing the number of considered tags significantly to only 11%. Their distributions' correlations and divergence can be found in the third row of Table V. We can observe that the limitation to such "active" tags yields a higher Spearman correlation and less divergence, as the active tags' rankings exhibit far less ties than the full set of tags.

Users and publications. As with tags, we investigate similar distributions of both users and resources: D^{req}_{User} counts the requests to specific users, D^{post}_{User} counts a user's posts, D^{req}_{Pub} counts the requests to a particular publication, and D^{post}_{Pub} counts the posts containing a publication. Similarly, we have the according frequency distributions (e.g., $D^{req}_{F,User}$) and the restricted distributions to active entities ignoring those that occur either only in posts or only in requests (e.g., ${}^{\emptyset}D^{req}_{Pub}$). Here again, we restrict resources to publications (and thus omit bookmarks), as visits of bookmarks are not recorded in the log files (see Section 3). The correlation results are depicted in rows four through nine in Table V, and the frequency distributions are illustrated in Figure 8(c) and (e).

The distributions of user (publication) frequencies in requests $D^{req}_{F,User}$ ($D^{req}_{F,Pub}$) and in posts $D^{post}_{F,User}$ ($D^{post}_{F,Pub}$) are similar and yield a relatively high correlation according to Pearson's r (rows four and seven in Table V). Their Jensen-Shannon divergences JS_2 are higher than for tags, but still the distributions are relatively similar. Since the distributions $D^{post}_{F,Pub}$ and $D^{req}_{F,Pub}$ are for the most part monotonically decreasing (Figure 8(e)), their rank correlation is high, unlike for the frequencies of users (Figure 8(c)). Notable in both cases (users and publications) is that the distributions of frequencies in posts and requests are no longer "parallel" as they were in the case of tags (compare Figure 8(a) to Figure 8(c) and (e)).

Power law fits for the publication frequency distributions of both posts $D^{post}_{F,Pub}$ ($\alpha = 3.17$, $x_{min} = 5$), and requests $D^{req}_{F,Pub}$ ($\alpha = 3.04$, $x_{min} = 22$) are decent fits with relatively low x_{min} values (see Figure 8(f)). Not surprisingly, the fits of the power law function are statistically significantly better than those of the exponential function. However, it is extremely difficult to distinguish the fits of the log-normal function and the power law function with exponential cutoff from the power law fit—a strong indicator for the presence of heavy-tailed distributions. For user frequencies, our results also indicate a good power law fit for both $D^{post}_{F,User}$ ($\alpha = 2.39, x_{min} = 988$) and $D^{req}_{F,User}$ ($\alpha = 1.60, x_{min} = 3$). The fits are shown in Figure 8(d). Similar to our investigations on tag frequencies, we obtain a higher x_{min} value for the frequencies in posts than for those in requests. For $D^{req}_{F,User}$, all candidate functions are better fits than the exponential function; both the log-normal as well as the power law function with exponential cutoff are better fits to the data than the pure power law function. The power law with cutoff is even better than the log-normal. For $D^{post}_{F,User}$, the power law fit is better than the exponential function, and it is difficult to distinguish from the other candidate distributions.

Regarding individual entities, we again measure correlations between the respective distributions in Table V (for users in rows five and six and for publications in rows eight and nine). For the resources (D^{req}_{Pub} and D^{post}_{Pub}), we obtain similar results as previously for tags. Pearson's correlation is moderate, the divergence is even higher than for tags, there is almost no rank correlation, and removing "inactive" publications (occurring either only in posts or in requests) yields higher rank correlation and lower divergence. The elimination of such publications leaves only about 12% of the original set of publications. By and large, we find only moderate correlation even among the actively posted and requested publications. A possible explanation for the correlation results might be based on the large number of publications that only get posted and requested infrequently. Slight changes in the post or request counts (e.g., once vs. twice) only change Pearson's correlation slightly but have a large influence on Spearman's correlation. For users (D^{req}_{User} and D^{post}_{User}), we find different behavior: almost no correlation according to Pearson's r, moderate rank correlation ρ (higher than for tags and publications), and divergence JS_2. This indicates that users with many posts indeed tend to be requested more but not proportionally more.

4.4.3. Discussion. The obtained results do not clearly support the initial assumption. The overall behavior of tag (and to a smaller degree of user and resource) frequencies is similar in requests and posts, and they are heavy tailed as expected. In all examples, we can find a good power law fit. However, in some occasions, the distribution decays from the straight power law function, which indicates the presence of other heavy-tailed distributions that might be based on distinct processes creating these distributions. This warrants further detailed investigation in the future.

On the level of individual entities, we observe weaker correlations—and only among the more actively used entities. It is surprising that despite the fact that tag clouds are displayed in BibSonomy and users can click tags to find according resources, the choice of tags in requests is not more strongly correlated to their popularity in posts. Additionally, we noted a strong difference in the company system Dogear, where much stronger correlations could be observed for tags. For operators of a tagging system, the results indicate that it is reasonable to exclude rarely requested tags completely from tag clouds or to use request frequencies instead or in addition to post frequencies in tag clouds. These could even be personalized to a user's query behavior.

5. DISCUSSION

In the previous section, we investigated four aspects of social tagging using the BibSonomy system as our showcase. Although our findings in this work are limited to BibSonomy, our approach is directly applicable to other tagging systems, and we briefly discuss some aspects of such a transfer here. As shown in the user study by Heckner et al. [2009], different tagging systems yield different characteristics (in their case, regarding the users' tagging motivation). Likewise, Yamaguchi et al. [2015] compared different statistics (node degree distributions, average number of tags per resources) and found differences between Twitter (where users tag other users), Flickr, and Delicious. In the following, we speculate about possible influences and results of our studies in other tagging systems.

We have already mentioned the influence of the *degree of openness*. In contrast to public, openly available systems, a company's internal systems can impose certain requirements on their users, such as the use of real names instead of pseudonyms or boundaries for the tags and resources in the system. For example, the knowledge of whose resources one browses could be a strong influence for the social behavior of sharing and visiting. Indeed, we have found similarity but also pronounced differences between the usage behavior in BibSonomy compared to that in Dogear [Millen and Feinberg 2006] in our investigation of the social aspect in Section 4.1 and also in the popularity aspect in Section 4.4.

Another influence is surely the *type of resources* that are bookmarked. Heckner et al. [2009] have shown that motivations for tagging (sharing or personal information management) were different in the systems YouTube (resources are videos) and Flickr (images) compared to Delicious (Web links) and Connotea (publication references). A major difference between those two pairs of systems is that resources like links and publication references are taken from other available sources, whereas images and videos are often published in the respective system for the first time. We thus expect that with regard to the social aspect, we would find similar results on Delicious and Connotea, because BibSonomy allows users to tag Web links (like Delicious) and references to publications (like Connotea). On the other hand, we can speculate that systems like YouTube and Flickr would show different results, such as a much stronger interest in the content of other users.

The *age* of the system is another influence. All three previous log file analyses [Millen and Feinberg 2006; Damianos et al. 2007; Millen et al. 2007] report results from periods of 8, 10, and 12 months, respectively, shortly after the systems' creation in 2005. In contrast to that, our log dataset covers a period of 6 years.

Finally, the *navigation concept* and the *graphical user interface* can play a role. BibSonomy offers the typical folksonomy navigation by always presenting users, resources, and tags as linked entities. Different tagging systems may make different design choices regarding the visibility and accessibility of individual entities.

To investigate these questions further, one would have to conduct experiments on logs of other systems as well. However, the bottleneck is the availability of such datasets. Therefore, our study is a first step toward analyzing user behavior using log files. We encourage other researchers and Webmasters of tagging systems to conduct similar studies, using the methods presented here, on their tagging systems and to compare their results to ours.

6. CONCLUSIONS

In this work, we tackled several prominent research questions about social tagging systems using a Web server log dataset from BibSonomy containing data on both posts and requests. We have thus supplemented previous work that has tapped into surveys and post data by also reflecting on actual user behavior leveraging request data.

6.1. Lessons Learned

Our findings paint a rather mixed picture about the four aspects we studied.

The social aspect. In our analysis of the social tagging system BibSonomy, we found evidence both for and against the assumption that the activities in a tagging system are primarily *social*. Whereas some user actions indeed indicate social sharing, others are evidence for individual purposes (see Table I and Figure 3). Furthermore, we could observe that resources are reused by others; especially publications are copied often (see Table II). This suggests that both kinds of activity are relevant in a social tagging system and should therefore be supported in the system's design. In addition, it is encouraging news for tagging in general, as it fits to the idea that users contribute to the system for their own purposes, but they can still profit from the contributions of others, which justifies the collaborative nature of social bookmarking.

We saw that users who just started with the system tend to spend more requests on the retrieval of one's own content. Given time, users start showing interest in the resources of others; however, with a rising number of requests, the share of requests to one's own collection rises (see Figure 4). A possible interpretation of this result could be that long time users need more support to navigate their own collections. One such approach could be recommenders for one's own content (traditionally, folksonomic recommendations have focused on recommending new resources (e.g., Doerfel et al. [2016]), or cleanups, removing resources that have never been revisited or that have not been used in a long time, to reduce the size of one's collection and thus make self-retrieval more efficient.

Finally, we saw that explicit social ties, like groups and friendship relations, play only a minor role for retrieval in BibSonomy. One reason might be that the main advantages of these features are related to BibSonomy's visibility concept. To encourage users to network, more features exploiting such ties could be helpful. In BibSonomy (meanwhile), users are assisted during the posting process. When the resource that is about to be posted can be found in the system, the respective metadata is suggested to the active user. A possible extension of this feature, relying on social ties, would be to highlight metadata that has been entered by friends or members of the active user's groups. Another feature that relies on user networks is a discussion feature where users can discuss online with their peers. Often, users will rather not make their opinion about their colleagues' work completely public. However, when they review publications, they might be willing to share them with persons from their own network. Moreover, users can send interesting publications to other users. This enables another form of retrieval in tagging systems that is driven by actively sharing (in contrast to just posting a resource, which can be seen as a more passive form of sharing).

The personal management aspect. We observed that users did not retrieve their own resources and tags as much as one would expect (see Figure 5). A consequence that we mentioned earlier could be reminders of unvisited resources or cleanup recommendations to remove unused resources or tags and thus to keep one's own collection manageable. Moreover, the observation suggests that visits to resources or tags could be valuable indicators for their importance to the user or generally to the system. In altmetrics—the study of measuring publications' impact through use on the Web—it is often assumed that visits, downloads, and so on are indicators for a publication's importance, and it has been found that they correlate with citations (e.g., Thelwall and Wilson [2015] or Zoller et al. [2015]). A consequence for Webmasters of tagging systems would be to make use of these statistics (i) by showing them to the users in the system and (ii) by exploiting them in ranking and recommendation algorithms. Particularly the latter opens the field for new studies of recommender systems, as almost any known folksonomic recommendation algorithm could be revisited and extended to make use of usage data.

The equality aspect. We found a strong *inequality* between the use of users, tags, and resources for navigation within BibSonomy (see Table III and Figure 6). User pages are visited much more often than resource or tag pages, providing clear evidence that assuming tags, users, and resources to be equally important for the navigation in BibSonomy would be wrong. This observation gives rise to a series of further investigations regarding the choice of the visited users during navigation. One could learn relations between the users, and this information could be used to recommend users to visit, or even to recommend these users' resources directly, such that the active users no longer have to visit user pages to retrieve resources.

We also noticed a pronounced difference between retrieval of one's own content and content of others. Users tend to use the full-text search for the latter task, whereas they use tags to retrieve their own resources (see Table IV). This suggests that users know their way within their own vocabulary. On the other hand, since the full-text search in BibSonomy includes the full metadata of a publication as well as the tags, it makes sense for users to use it for finding new content, allowing hits to match the queries terms in the metadata as well as in tags. An idea for making tag search more successful for finding other users' content might be to extend the tag search beyond the Boolean approach of returning only posts that have the queried tag: given a tag, one could return posts with similar tags. Thus, similarity can be obtained from known word ontologies like WordNet,[26] from semantics that are extracted from the tagging data in the system, from the "feedback cycle" of tagging—users find a post and then use their own tags when they copy it [Halpin et al. 2007]—or from relations found in the log data (users who queried for tag X clicked on posts with tag Y, etc.). To gather further evidence for the behavior of users with tags of others, it would be helpful to repeat this study on a tagging system where the resources are not suitably describable by full-text, and thus a full-text search would be less helpful, for instance, on a tagging system for images like Flickr.[27]

Transition probabilities showed that users often tend to stay with the same type of retrieval (e.g., one tag request after another). Requests to users are those with the highest share of different follow-up requests. Only 35% of the requests following a user request are again user requests (see Figure 7). As with usage statistics, transition probabilities could also be used as additional information in recommendation and ranking algorithms.

The popularity aspect. Finally, we compared popularity of entities in posts and in requests. We observed common usage patterns on an aggregate level (see Figure 8), yet the patterns are less pronounced on an individual level (see Figure 9 and Table V), suggesting that an entity's popularity in posts is only reflected to a certain extent in the requests to that entity. Such information could be valuable in the visualization of folksonomy information. Until now, tag clouds are the most popular means of displaying tags, and usually tags are ordered by their number of occurrences in posts. Our results suggest that including the occurrences in requests could be helpful.

Overall, this article contributes a stepping stone for studies of social tagging systems by using actual traces of user behavior that can be found in request log data, a basis for comparative studies, exploring the extent to which these different aspects are pronounced in different tagging systems, and new insights about the use of literature in a publication management system.

[26]https://wordnet.princeton.edu/.
[27]https://www.flickr.com/.

6.2. Future Research

It is reasonable to assume that different tagging systems (Flickr, Delicious, BibSonomy and others) exhibit unique characteristics and dynamics that make them amenable to different uses and purposes. Further studies of request log data in other tagging systems would be helpful in uncovering these differences. In addition, we provide new insights about the relative importance of users, tags, and resources in social tagging systems. Finding that the equality assumption does not hold generally has important implications for the layout of tagging systems and for the design and implementation of algorithms that address search and retrieval. For example, the FolkRank [Hotho et al. 2006] algorithm might profit from the inclusion of weights reflecting popularity or transition probability in requests.

It is our hope that our work triggers a new line of research on social tagging systems that utilizes traces of actual user behavior to test and challenge our existing body of knowledge about these systems gained from other inquisition methods, such as surveys or post data.

REFERENCES

David Abrams, Ron Baecker, and Mark Chignell. 1998. Information archiving with bookmarks: Personal Web space construction and organization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'98)*. ACM, New York, NY, 41–48. DOI:http://dx.doi.org/10.1145/274644.274651

Maristella Agosti, Franco Crivellari, and GiorgioMaria Di Nunzio. 2012. Web log analysis: A review of a decade of studies about information acquisition, inspection and interpretation of user interaction. *Data Mining and Knowledge Discovery* 24, 3, 663–696. DOI:http://dx.doi.org/10.1007/s10618-011-0228-8

Morgan Ames and Mor Naaman. 2007. Why we tag: Motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, New York, NY, 971–980. DOI:http://dx.doi.org/10.1145/1240624.1240772

Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. 2009. Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement (IMC'09)*. ACM, New York, NY, 49–62. DOI:http://dx.doi.org/10.1145/1644893.1644900

Dominik Benz, Andreas Hotho, Robert Jäschke, Beate Krause, Folke Mitzlaff, Christoph Schmitz, and Gerd Stumme. 2010. The social bookmark and publication management system BibSonomy. *VLDB Journal* 19, 6, 849–875. DOI:http://dx.doi.org/10.1007/s00778-010-0208-4

Toine Bogers. 2009. *Recommender Systems for Social Bookmarking*. Ph.D. Dissertation. Tilburg University, Tilburg, The Netherlands.

Ángel Borrego and Jenny Fry. 2012. Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy. *Journal of Information Science* 38, 3, 297–308. DOI:http://dx.doi.org/10.1177/0165551512438353

Christopher H. Brooks and Nancy Montanez. 2006. Improved annotation of the blogosphere via autotagging and hierarchical clustering. In *Proceedings of the 15th International Conference on World Wide Web (WWW'06)*. ACM, New York, NY, 625–632. DOI:http://dx.doi.org/10.1145/1135777.1135869

Mark J. Carman, Mark Baillie, Robert Gwadera, and Fabio Crestani. 2009. A statistical comparison of tag and query logs. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'09)*. ACM, New York, NY, 123–130. DOI:http://dx.doi.org/10.1145/1571941.1571965

Ciro Cattuto, Christoph Schmitz, Andrea Baldassarri, Vito D. P. Servedio, Vittorio Loreto, Andreas Hotho, Miranda Grahl, and Gerd Stumme. 2007. Network properties of folksonomies. *AI Communications Journal* 20, 4, 245–262.

Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. 2009. Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Transactions on Networking* 17, 5, 1357–1370.

A. Clauset, C. Rohilla Shalizi, and M. E. J. Newman. 2009. Power-law distributions in empirical data. *SIAM Review* 51, 4, 661–703. DOI:http://dx.doi.org/10.1137/070710111

Laurie E. Damianos, Donna Cuomo, John Griffith, David M. Hirst, and James Smallwood. 2007. Exploring the adoption, utility, and social influences of social bookmarking in a corporate environment. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. IEEE, Los Alamitos, CA, 86–95. DOI:http://dx.doi.org/10.1109/HICSS.2007.219

Stephan Doerfel, Andreas Hotho, Aliye Kartal-Aydemir, Alexander Roßnagel, and Gerd Stumme. 2013. *Informationelle Selbstbestimmung Im Web 2.0 Chancen Und Risiken Sozialer Verschlagwortungssysteme*. Vieweg & Teubner Verlag.

Stephan Doerfel, Robert Jäschke, and Gerd Stumme. 2016. The role of cores in recommender benchmarking for social bookmarking systems. *ACM Transactions on Intelligent Systems and Technology* 7, 3, 40:1–40:33. http://doi.acm.org/10.1145/2700485

Stephan Doerfel, Daniel Zoller, Philipp Singer, Thomas Niebler, Andreas Hotho, and Markus Strohmaier. 2014a. How social is social tagging? In *Proceedings of the 23rd International World Wide Web Conference (WWW'14)*. ACM, New York, NY.

Stephan Doerfel, Daniel Zoller, Philipp Singer, Thomas Niebler, Andreas Hotho, and Markus Strohmaier. 2014b. Of Course We Share! Testing Assumptions about Social Tagging Systems. Available at http://arxiv.org/abs/1401.0629.

Sergio Duarte Torres, Ingmar Weber, and Djoerd Hiemstra. 2014. Analysis of search and browsing behavior of young users on the Web. *ACM Transactions on the Web* 8, 2, Article No. 7. DOI:http://dx.doi.org/10.1145/2555595

Wai-Tat Fu, Thomas Kannampallil, Ruogu Kang, and Jibo He. 2010. Semantic imitation in social tagging. *ACM Transactions on Computer-Human Interaction* 17, 3, Article No. 12. DOI:http://dx.doi.org/10.1145/1806923.1806926

Robert J. Glushko, Paul P. Maglio, Teenie Matlock, and Lawrence W. Barsalou. 2008. Categorization in the wild. *Trends in Cognitive Sciences* 12, 4, 129–135. DOI:http://dx.doi.org/10.1016/j.tics.2008.01.007

Scott A. Golder and Bernardo A. Huberman. 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science* 32, 2, 198–208. DOI:http://dx.doi.org/10.1177/0165551506062337

Harry Halpin, Valentin Robu, and Hana Shepherd. 2007. The complex dynamics of collaborative tagging. In *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*. ACM, New York, NY, 211–220. DOI:http://dx.doi.org/10.1145/1242572.1242602

Markus Heckner, Michael Heilemann, and Christian Wolff. 2009. Personal information management vs. resource sharing: Towards a model of information behaviour in social tagging systems. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM'09)*.

Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. 2006. Information retrieval in folksonomies: Search and ranking. In *The Semantic Web: Research and Applications*. Lecture Notes in Computer Science, Vol. 4011. Springer, 411–416.

Robert Jäschke, Leandro Marinho, Andreas Hotho, Lars Schmidt-Thieme, and Gerd Stumme. 2008. Tag recommendations in social bookmarking systems. *AI Communications* 21, 4, 231–247. DOI:http://dx.doi.org/10.3233/AIC-2008-0438

Jing Jiang, Christo Wilson, Xiao Wang, Wenpeng Sha, Peng Huang, Yafei Dai, and Ben Y. Zhao. 2013. Understanding latent interactions in online social networks. *ACM Transactions on the Web* 7, 4, Article No. 18. DOI:http://dx.doi.org/10.1145/2517040

Beate Krause, Christoph Schmitz, Andreas Hotho, and Gerd Stumme. 2008. The anti-social tagger—detecting spam in social bookmarking systems. In *Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb'08)*. ACM, New York, NY, 61–68. DOI:http://dx.doi.org/10.1145/1451983.1451998

Jared Lorince, Kenneth Joseph, and Peter M. Todd. 2015. Analysis of music tagging and listening patterns: Do tags really function as retrieval aids? In *Social Computing, Behavioral-Cultural Modeling, and Prediction*. Lecture Notes in Computer Science, Vol. 9021. Springer, 141–152. DOI:http://dx.doi.org/10.1007/978-3-319-16268-3_15

Cameron Marlow, Mor Naaman, Danah Boyd, and Marc Davis. 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proceedings of the 17th Conference on Hypertext and Hypermedia (HYPERTEXT'06)*. ACM, New York, NY, 31–40. DOI:http://dx.doi.org/10.1145/1149941.1149949

Adam Mathes. 2004. Folksonomies—Cooperative Classification and Communication Through Shared Metadata. Retrieved April 3, 2016, from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html.

David R. Millen and Jonathan Feinberg. 2006. Using social tagging to improve social navigation. In *Proceedings of the Workshop on Social Navigation and Community Based Adaptation Technologies*.

David R. Millen, Meng Yang, Steven Whittaker, and Jonathan Feinberg. 2007. Social bookmarking and exploratory search. In *ECSCW 2007*, L. J. Bannon, I. Wagner, C. Gutwin, R. H. R. Harper, and K. Schmidt (Eds.). Springer, London, 21–40. DOI:http://dx.doi.org/10.1007/978-1-84800-031-5_2

Folke Mitzlaff, Martin Atzmueller, Dominik Benz, Andreas Hotho, and Gerd Stumme. 2011. Community assessment using evidence networks. In *Analysis of Social Media and Ubiquitous Data*. Lecture Notes in Computer Science, Vol. 6904. Springer, 79–98. DOI:http://dx.doi.org/10.1007/978-3-642-23599-3_5

Stefano Mossa, Marc Barthélémy, H. Eugene Stanley, and Luís A. Nunes Amaral. 2002. Truncation of power law behavior in "scale-free" network models due to information filtering. *Physical Review Letters* 88, 13, 138701.

David Nicholas, Paul Huntington, and Anthony Watkinson. 2005. Scholarly journal usage: The results of deep log analysis. *Journal of Documentation* 61, 2, 248–280.

Natalya Fridman Noy, Abhita Chugh, and Harith Alani. 2008. The CKC challenge: Exploring tools for collaborative knowledge construction. *IEEE Intelligent Systems* 23, 1, 64–68. DOI:http://dx.doi.org/10.1109/MIS.2008.14

Joshua Porter. 2005. Learning More about Structured Blogging. Retrieved April 3, 2016, from http://bokardo.com/archives/learning-more-about-structured-blogging/.

Steffen Rendle, Leandro Balby Marinho, Alexandros Nanopoulos, and Lars Schmidt-Thieme. 2009. Learning optimal ranking with tensor factorization for tag recommendation. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 727–736.

Fabian Schneider, Anja Feldmann, Balachander Krishnamurthy, and Walter Willinger. 2009. Understanding online social network usage from a network perspective. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference (IMC'09)*. ACM, New York, NY, 35–48. DOI:http://dx.doi.org/10.1145/1644893.1644899

James Sinclair and Michael Cardew-Hall. 2008. The folksonomy tag cloud: When is it useful? *Journal of Information Science* 34, 1, 15–29. DOI:http://dx.doi.org/10.1177/0165551506078083

Juliane Stiller, Maria Gäde, and Vivien Petras. 2011. Is tagging multilingual? A case study with BibSonomy. In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries (JCDL'11)*. ACM, New York, NY, 421–422. DOI:http://dx.doi.org/10.1145/1998076.1998165

Markus Strohmaier, Christian Körner, and Roman Kern. 2010. Why do users tag? Detecting users' motivation for tagging in social tagging systems. In *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM'10)*.

Daniel Terdiman. 2005. Folksonomies Tap People Power. Retrieved April 3, 2016, from http://www.wired.com/science/discoveries/news/2005/02/66456?currentPage=all.

Mike Thelwall and Paul Wilson. 2015. Mendeley Readership Altmetrics for Medical Articles: An Analysis of 45 Fields. Retrieved April 3, 2016, from http://www.onlinelibrary.wiley.com/doi/10.1002/asi.23501/full.

Paul Thomas. 2014. Using interaction data to explain difficulty navigating online. *ACM Transactions on the Web* 8, 4, Article No. 24. DOI:http://dx.doi.org/10.1145/2656343

Jennifer Trant. 2009. Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information* 10, 1, 1–44.

Thomas Vander Wal. 2005. Tagging for Fun and Finding. Retrieved April 3, 2016, from http://okcancel.com/archives/article/2005/07/tagging-for-fun-and-finding.html.

Thomas Vander Wal. 2007. Folksonomy. Retrieved April 3, 2016, from http://vanderwal.net/folksonomy.html.

David Weinberger. 2005. Tagging and Why It Matters. Retrieved April 3, 2016, from http://cyber.law.harvard.edu/wg+home/uploads/507/07-WhyTaggingMatters.pdf. DOI:http://dx.doi.org/ 10.2139/ssrn.870594

Yuto Yamaguchi, Mitsuo Yoshida, Christos Faloutsos, and Hiroyuki Kitagawa. 2015. Patterns in interactive tagging networks. In *Proceedings of the International AAAI Conference on Web and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10556.

Daniel Zoller, Stephan Doerfel, Robert Jäschke, Gerd Stumme, and Andreas Hotho. 2015. On publication usage in a social bookmarking system. In *Proceedings of the 2015 ACM Conference on Web Science*. http://websci15.org/accepted-submissions.

Arkaitz Zubiaga, Christian Körner, and Markus Strohmaier. 2011. Tags vs shelves: From social tagging to social classification. In *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia (HYPERTEXT'11)*. ACM, New York, NY, 93–102. DOI:http://dx.doi.org/10.1145/1995966.1995981

Received June 2015; revised December 2015; accepted February 2016