Understanding Why Users Tag: A Survey of Tagging Motivation Literature and Results from an Empirical Study

Markus Strohmaier

markus.strohmaier@tugraz.at

Graz University of Technology

Inffeldgasse 21 8010 Graz Austria

Christian Körner

christian.koerner@tugraz.at

Graz University of Technology

Inffeldgasse 21 8010 Graz Austria

Roman Kern

rkern@know-center.at

Know Center

Inffeldgasse 21 8010 Graz Austria

Abstract

While recent progress has been achieved in understanding the structure and dynamics of social tagging systems, we know little about the underlying user motivations for tagging, and how they influence resulting folksonomies and tags. This paper addresses three issues related to this question: 1.) What distinctions of user motivations are identified by previous research, and in what ways is user motivation amenable to quantitative analysis? 2.) To what extent does tagging motivation vary across different social tagging systems? and 3.) How does variability in user motivation influence resulting tags and folksonomies? In this paper, we present measures to detect whether a tagger is primarily motivated by categorizing or describing resources, and apply these measures to datasets from seven different tagging systems. Our results show that a) users' motivation for tagging varies not only across, but also within tagging systems, and that b) tag agreement among users who are motivated by *categorizing resources* is significantly lower than among users who are motivated by *describing resources*. Our findings are relevant for 1) the development of tag-based user interfaces 2) the analysis of tag semantics and 3) the design of search algorithms for social tagging systems.

Keywords: social tagging systems, tagging motivation, user motivation, user goals

1. Introduction

Tagging typically describes the voluntary activity of users who are annotating resources with terms - so called "tags" - freely chosen from an unbounded and uncontrolled vocabulary (cf.[1, 2]). Social tagging systems have been the subject of research for several years now. An influential paper by Mika [3] describes, for example, how social tagging systems can be used for ontology learning. Other work has studied the evolution of folksonomies over time, such as Chi and Mytkowicz [4]. Further studies focused on using tags for a variety of purposes, including the use of tags for web search [5, 6], ranking [7], taxonomy development [8], and other tasks.

However, our understanding of folksonomies is

not yet (nor could it be) mature. A question that has recently attracted the interest of the semantic web community is whether the properties of tags in tagging systems and their usefulness for different purposes can be assumed to be *a function of the taggers' motivation or intention behind tagging* [9]. If this was the case, users' motivation for tagging would have broad implications for the design of useful algorithms such as adapting tag recommendation techniques and user interfacesfor social tagging systems. In order to assess the general usefulness of algorithms that aim to, for example, capture knowledge from folksonomies, we would need to know whether these algorithms produce similar results across different user populations driven by different motivations for tagging. Recent research already suggests that different tagging systems afford different motivations for tagging [9, 10]. Further work presents anecdotal evidence that even within the same tagging system, motivation for tagging between individual users may vary greatly [11]. Given these observations, it is interesting to study whether and how the analysis of user motivation for tagging is amenable to quantitative investigations, and whether folksonomies and the tag semantics they contain are influenced by different tagging motivations.

Tagging motivation has remained largely elusive until the first studies on this subject have been conducted in 2006. At this time, the work by Golder and Huberman [1] and Marlow and Naaman [2] have made advances towards expanding our theoretical understanding of tagging motivation by identifying and classifying user motivation in tagging systems. Their work was followed by studies proposing generalizations, refinements and extensions to previous classifications [9]. An influential observation was made by Coates [12] and elaborated on and interpreted in [2] and [9]. This line of work suggests that a distinction between at least two fundamental types of user motivation for tagging is important:

Categorization vs. Description: On one hand, users who are motivated by categorization view tagging as a means to *categorize resources* according to some shared high-level characteristics. These users tag because they want to construct and maintain a navigational aid to the resources for later browsing. On the other hand, users who are motivated by description view tagging as a means to accurately and precisely *describe resources*. These users tag because they want to produce annotations that are useful for later searching. This distinction has been found to be important because, for example, tags assigned by describers might be more useful for information retrieval (because these tags focus on the content of resources) as opposed to tags assigned by categorizers, which might be more useful to capture a rich variety of possible interpretations of a resource (because they focus on user-specific views on resources).

In this paper, we are adopting the distinction between categorizers and describers to study the following three research questions: 1) How can we measure the motivation behind tagging? 2) To what extent does tagging motivation vary across different social tagging systems? and 3) How does tagging motivation influence resulting folksonomies?

The overall contribution of this paper is three-fold: first, our work presents measures capable of detecting the tacit nature of tagging motivation in social tagging systems. Secondly, we apply and validate these measures in the context of seven different tagging systems, showing that tagging motivation significantly varies across and even within tagging systems. Third, we provide first empirical evidence that at least one property of tags ("tag agreement") is influenced by users' motivation for tagging. By analyzing *complete tagging histories* of almost 4000 users from *seven different* tagging systems, to the best of our knowledge this work represents the largest *quantitative* cross-domain study on tagging motivation to date so far.

This article is organized as follows: first we review the state of the art of research on tagging motivation. Then, we introduce and discuss a number of ways to make tagging motivation in social tagging systems amenable to quantitative analysis. This is followed by a presentation of results from an empirical study involving nine datasets from seven different social tagging systems. Subsequently, we evaluate our measures by a combination of different strategies. Finally, we present evidence for a relation between the motivation behind tagging and the level of agreement on tags among users of tagging systems.

2. Related Work

Table 1 gives a chronological overview of past research focusing on user motivation in tagging systems to give an answer to what research was done in the past with regard to tagging motivation.

Table 1: Overview of Research on Users' Motivation for Tagging in Social Tagging Systems

Authors	Categories of Tagging Motivation	Detection	Evidence	Reasoning	Systems investigated	# of users	Resources per user
Coates 2005	Categorization, description	Expert judgment	Anecdotal	Inductive	Weblog	1	N/A
Hammond et al. 2005	Self/self, self/others, others/self, others/others	Expert judgment	Observation	Inductive	9 different tagging systems	N/A	N/A
Golder et al. 2006	What it is about, what it is, who owns it, refining categories, identifying qualities, self reference, task organizing	Expert judgment	Dataset	Inductive	Delicious	229	300 (average)
Marlow et al. 2006	Organizational, social, [and refinements]	Expert judgment	N/A	Deductive	Flickr	10 (25,000)	100 (minimum)
Xu et al. 2006	Content-based, context-based, attribute-based, subjective, organizational	Expert judgment	N/A	Deductive	N/A	N/A	N/A
Sen et al. 2006	Self-expression, organizing, learning, finding, decision support	Expert judgment	Prior experience	Deductive	MovieLens	635 (3,366)	N/A
Wash et al. 2007	Later retrieval, sharing, social recognition, [and others]	Expert judgment	Interviews (semistruct.)	Inductive	Delicious	12	950 (average)
Ames et al. 2007	Self/organization, self/communication, social/organization, social/communication	Expert judgment	Interviews (in-depth)	Inductive	Flickr, ZoneTag	13	N/A
Heckner et al. 2009	Personal information management, resource sharing	Expert judgment	Survey (M. Turk)	Deductive	Flickr, Youtube, Delicious, Connotea	142	20 and 5 (minimum)
Nov et al. 2009	enjoyment, commitment, self development, reputation	Expert judgment	Survey (e-mail)	Deductive	Flickr (PRO users only)	422	2,848.5 (average)

The best known and most influential paper on surveying social tagging systems was done by Golder and Huberman [1] in which an early analysis of folksonomies is conducted. The authors investigate the structure of collaborative tagging systems and found regularities in user activity, tag frequencies, the used tags and other aspects of two snapshots taken from the Delicious system. This work was the first to explore user behavior in these systems and showed that users in Delicious exhibit a variety of different behaviors. Hammond et al.[10] were the first to conduct early analyses of folksonomic systems. The authors examined nine different social bookmarking platforms such as CiteULike, Connotea[1], Delicious, Flickr and others. In this work two dimensions of tagging methodology are differentiated: *tag creation* and *tag usage*. Coates [12] hypothesizes two distinct tagging approaches. The first approach treats tags as a replacement for folders. This way tags describe the category where the annotated resource belongs to. The other approach simply uses tags on resources which make sense to the user and characterize the resource in a detailed manner.

Marlow et al. [2] show two high level types of categorization for motivation: *organizational* and *social* practices. The authors further elaborate a list of incentives by which users can be motivated. Examples are *future retrieval* where tags are used to make the resource easier to find by the annotator himself and *contribution and sharing* in which keywords are facilitated in order to create clusters of resources to make them retrievable by other users.

Xu et al. [13] created a taxonomy of tags for the creation of a tag recommendation algorithm. These five categories are: *Content-based* tags which give insight into the content or the categories of an annotated object (e.g. names, brands etc.) *Context-*

---

[1] http://www.connotea.org

based tags show the context under which the resource is stored (examples are: location, time, etc.) *Attribute* tags tell about properties of a resource. *Subjective* tags explain the user's opinion of a given resource. *Organizational* tags allow a user to organize her library.

[14] introduce a user-centered model for vocabulary evolution to answer the questions of how strongly a user's investment and habit as well as the community of a system influence tagging behavior. The authors combine the seven tag classes by [1] into three classes of their own: *Factual Tags* describe properties (like place, time, etc.) of the movie. *Subjective Tags* expressing users' opinions about the annotated movie. *Personal Tags* represent tags which are used to organize a user's library.

Further work by [11] focuses on the incentives users have when they utilize a social computing systems. By analyzing interviews conducted with twelve users from the Delicious system the authors the primary answers were the reuse of tags that were previously assigned to resources as well as the usage of terms a user expects to search on.

Heckner et al. [15] perform a comparative study on four different tagging systems (YouTube, Connotea, Delicious and Flickr) and observe differences in tagging behavior for different digital resources. General trends the authors identify were amongst others: photos are tagged for content and location. Videos are tagged for persons and scientific articles are tagged for time and task.

Several interesting observations can be made: first, Table 1 shows that a vast variety of categories for tagging motivation has been proposed in the past. While these categories give us a broad and multifaceted picture of user motivations in tagging systems, the total set of categories suffers from a number of potential problems, including overlap and incompleteness of categories, as well as differences in terms of coverage and scope. A common understanding, or even a stable set of approaches has yet to emerge. This suggests that research in this direction is still in an early stage. Secondly, we can observe that while earlier work focused on anecdotal evidence [12] and theory development [1], empirical validation has become increasingly important. This is evident in the recent focus on interview- and survey-based research ([2, 9, 11, 16, 17]). At the same time the number of users under investigation increased from a single user (in 2005) up to several hundred in later work (2009). Third, to the best of our knowledge,

no quantitative approaches for measuring tagging motivation exist; all previous research requires expert judgment to assess the appropriate category of tagging motivation based on observations, datasets, interviews or surveys. In a departure from past research, this paper sets out to explore *quantitative* measures for tagging motivation. Additional papers focused on other aspects of tagging, such as the development of tag classification taxonomies [18] or experimental studies on latent influences during tagging [19, 20].

3. Tagging Motivation

Our study was driven by the desire to identify categories of tagging motivation that lend themselves to quantification. As we have discussed in the previous Section 2, most existing categories of tagging motivation are rather abstract and high level, and do not lend themselves naturally to quantification. For example, it is not obvious or self-evident how an automatic distinction between motivations such as self-expression [21], enjoyment [16] or social-recognition [11] can be made. For these reasons, this paper focuses on a particularly promising distinction between *categorizers* and *describers*, inspired by Coates [12] and Heckner [9] and further refined and discussed in [22]. Several intuitions about this distinction (such as the kinds of tagging styles different types of users would adopt) make it a promising candidate for future investigations. Figure 1 illustrates this distinction with tag clouds [23] produced by a typical a categorizer vs. a typical describer.

While the example on the left illustrates that some categorizers even use special characters to build a pseudo-taxonomy of tags (e.g. `fashion_blog`, `fashion_brand`, `fashion_brand_bags`), what we see here is an extreme example. What separates categorizers from describers is the fact that they use tags as categories (often from a controlled vocabulary), as opposed to descriptive labels for the resource they are tagging. Details about this distinction are introduced next.

3.1. Using Tags to Categorize Resources

Users who are motivated by categorization engage in tagging because they want to construct and maintain a personal navigational aid to the resources (URLs, photos, etc) being tagged. This

a. Typical Categorizer

b. Typical Describer

Figure 1: Examples of Tag Clouds Produced by Different Users: Categorizer (left) vs. Describer (right)

implies developing a limited set of tags (or categories) that is rather stable over time. The tags that are assigned to resources are regarded as an investment into a structure, and changing this structure is typically regarded to be costly to a categorizer (cf. [21]). Resources are assigned to tags whenever they share some common characteristic important to the mental model of the user (e.g. 'photos', 'trip_to_Vienna' or 'favorites'). Because the tags assigned by categorizers are very close to the mental models of users, they can act as suitable facilitators for navigation and browsing.

3.2. Using Tags to Describe Resources

Users who are motivated by description engage in tagging because they want to appropriately and relevantly describe the resources being tagged. This typically implies an open set of tags, with a rather dynamic and unlimited tag vocabulary. Tags are not viewed as an investment into a tag structure, and changing the structure continuously is not regarded as costly. The goal of tagging is to identify those tags that match the resource best. Because the tags assigned are very close to the content of the resources, they can be utilized for searching and retrieval.

Table 2: Differences Between Categorizers and Describers

	Categorizer (C)	Describer (D)
Goal	later browsing	later retrieval
Change of vocab.	costly	cheap
Size of vocab.	limited	open
Tags	subjective	objective
Tag reuse	frequent	rare
Tag purpose	mimick taxonomy	descriptive labels

Table 2 illustrates key differences between the two identified types of tagging motivation. While these two categories make an ideal distinction, tagging in the real world is likely to be motivated by a combination of both. A user might maintain a few categories while pursuing a description approach for the majority of resources and vice versa, or additional categories might be introduced over time. In addition, the distinction between categorizers and describers is not about the *semantics* of tagging, it is a distinction based on the *motivation* for tagging. One implication of that is that it would be perfectly plausible for the same tag (for example 'java') to be used by both describers and categorizers, and serve both functions at the same time. In other words, the same tag might be used as a category or a descriptive label.

So the function of each tag is determined by *pragmatics* (how tags are used) rather than *semantics* (what they mean). In that sense, tagging motivation presents a new and additional perspective on studies of folksonomies. Taking this new perspective allows for studying whether tagging motivation has an impact on tag semantics. For example, it is reasonable to hypothesize that the tags produced by describers are more descriptive than tags produced by categorizers. If this was the case, considering tagging motivation in studies focused on utilizing tags for information retrieval or ontology learning would benefit from knowledge about the users' motivation for tagging.

4. Methodology and Setup

In social tagging systems, the structure of data captured by tagging can be characterized by a tripartite graph with hyper edges. The three disjoint, finite sets of such a graph correspond to 1) a set of users $u \in U$ 2) a set of objects or resources $r \in R$ and 3) a set of annotations or tags $t \in T$ that are used by users U to annotate resources R.

A *folksonomy* is thus a tuple $\mathbb{F} := (U, T, R, Y)$ where Y is a ternary relation between those three sets, i.e. $Y \subseteq U \times T \times R$ (cf. [3, 7, 24]). A personomy F_u of a user $u \in U$ is the reduction of F to u [25].

To study whether and how we can measure tagging motivation, and to what extent tagging motivation varies across different social tagging systems, we develop a number of measures and apply them to a range of personomies that exhibit different characteristics. We assume that understanding how an extreme describer would behave over time in comparison to an extreme categorizer helps in assessing the usefulness of measures to characterize user motivation in tagging systems.

In this work, we adopt a combination of the following strategies to explore the usefulness of the introduced measures: first we apply all measures to both synthetic and real-world tagging datasets. After that we analyze the ability of measures to capture predicted (synthetic) behavior. Finally, we relate our findings to results reported by previous work.

Assuming that the different motivations for tagging produce different personomies (different tagging behavior over time), we can use synthetic data from extreme categorizers and describers to find upper and lower bounds for the behavior that can be expected in real-world tagging systems.

4.1. Synthetic Personomy Data

To simulate behavior of users who are mainly driven by description, data from the ESP game dataset[2] was used. This dataset contains a large number of inter-subjectively validated, descriptive tags for pictures useful to capture describer behavior.

To contrast this data with behavior of users who are mainly driven by categorization, we crawled data from Flickr, but instead of using the tags we used information from users' *photo sets*. We consider each photo set to represent a tag assigned by a categorizer for all the photos that are contained within this set. The personomy then consists of all photos and the corresponding photo sets they are assigned to. We use these two synthetic datasets to simulate behavior of 'artificial' taggers who are mainly motivated by description and categorization.

4.2. Real-World Personomy Data

In addition to the synthetic datasets, we also crawled popular tagging systems. The datasets needed to be *sufficiently large* in order to enable us

Table 3: Overview and Statistics of Social Tagging Datasets. Asterisks (*) indicate synthetic personomies of extreme categorization/description behavior.

Dataset	$|U|$	$|T|$	$|R|$	$|R_u|_{min}$	$|T|/|R|$
ESP Game*	290	29,834	99,942	1,000	0.2985
Flickr Sets*	1,419	49,298	1,966,269	500	0.0250
Delicious	896	184,746	1,089,653	1,000	0.1695
Flickr Tags	456	216,936	965,419	1,000	0.2247
Bibsonomy Bookmarks	84	29,176	93,309	500	0.3127
Bibsonomy Publications	26	11006	23,696	500	0.4645
CiteULike	581	148,396	545,535	500	0.2720
Diigo Tags	135	68,428	161,475	500	0.4238
Movielens	99	9,983	7,078	500	1.4104

to observe tagging motivation across a large number of users and they needed to be *complete* because we wanted to study a user's complete tagging history over time - from the users first bookmark up to the most recent bookmarks. Because many of the tagging datasets available for research focus on sampling data on an aggregate level rather than capturing *complete* personomies, we had to acquire our own datasets.

Table 3 gives an overview of the datasets acquired for this research. To the best of our knowledge, the aggregate collection of these datasets itself represents the most diverse and largest dataset of complete and very large personomies (>500 tagged resources, recorded from the user's *first bookmark* on) to date. The restriction a minimum of 500 tagged resources was set to capture users who have actively used the system. Because assessing tagging motivation for users who are active or less active in the system is a harder problem, we defer it to future work. However in the Section 7 we will see that measuring below 500 tagged resources might be feasible as well.

Acquiring this kind of data in itself represents a challenge. To give an example, from about 3.600 users on Bibsonomy, only ˜100 users had more than 500 annotated resources according to a database dump made available at the end of 2008. The number of large and complete personomies acquired for this work totaled 3,986 users. The data was stored in an XML-based data schema to capture personomies from different systems in a uniform manner. This allows easy and consistent processing of statistical calculations. In the following, we briefly de-

---

[2]http://www.cs.cmu.edu/~biglou/resources/

scribe the data acquisition strategies for selected datasets. The other datasets were acquired in a similar fashion.

- **ESP Game**: Because temporal order can be assumed to be of little relevance for describers, a synthetic list of users has been created by randomly selecting resources from the ESP game dataset in order to simulate personomy data. The same thresholds as for the Flickr and Bibsonomy datasets were used. A number of virtual user personomies were generated by randomly choosing sets of resources and corresponding descriptive tags from the available corpus of real-world image labeling data. An almost arbitrary large number of personomies can be generated with this approach.

- **Flickr**: Using Flickr's web page that shows recently uploaded photos, a list consisting of several thousand user ids has been gathered. Because on Flickr there is a photo limit of 200 for regular users (i.e. users without a premium account) only public accounts were considered during crawling. For all users in our list, we retrieved all stored photos that are publicly accessible. The lower bound of tagged resources in our dataset is 500. The upper bound of tagged resources was 4000. For the second dataset that focuses on Flickr sets data, the same thresholds have been applied. For users that satisfied our requirements, the complete tagged photo stream and / or the photos belonging to photo sets have been retrieved using the Flickr Rest API.

- **Delicious**: Starting from a set of popular tags, a list consisting of several thousand usernames that have bookmarked any URLs with those tags has been recursively gathered. Because there is a limit for public access of the stored bookmarks to the most recent 4,000 entries of a user, only users having fewer bookmarks were used for our analysis. The minimum number of resources for each user was 1,000. We acquired the complete personomy data for users satisfying these criteria.

- **Bibsonomy**: We used a database dump of about 3,600 Bibsonomy users (from the end of 2008) that was provided by the tagging research group at University of Kassel. The lower and upper bound for both datasets (bookmarks and publications) are the same as for Flickr. The bookmark and publication datasets have been extracted from the database dump and were directly encoded into the XML-based personomy schema.

5. Measuring Tagging Motivation

In the following, we develop and apply different measures that aim to provide useful insights into the fabric of tagging motivation in social tagging systems. The measures introduced below focus on statistical aspects of users' *personomies only* instead of using entire folksonomies. For each measure, the underlying intuition, a formalization and the results of applying the measure to our datasets are presented next.

5.1. Growth of Tag Vocabulary - M_{vocab}

Intuition: Over time, the tag vocabulary (i.e. the set of all tags used in a users' personomy) of an ideal categorizer's tag vocabulary would be expected to reach a plateau. The rationale for this idea is that only a limited set of categories is of interest to a user mainly motivated by categorization. On the other hand, an ideal describer would not aim to limit the size of her tagging vocabulary, but would likely introduce new tags as required by the contents of the resources being tagged.

Description: A measure to distinguish between these two approaches would examine whether the cumulative tag vocabulary of a given user plateaus over time.

$$M_{vocab} = \frac{|T|}{|R|} \qquad (1)$$

Figure 2, left-most column, depicts the growth of tag vocabulary over 3,000 bookmarks across the different tagging datasets used in our study. The differences between extreme describers (synthetic ESP game data in the top row) and extreme categorizers (synthetic Flickr sets data in the bottom row) is stark. Data from Delicious users (in the middle, left) shows that real word tagging behavior lies in between the identified extremes.

Discussion: What can be observed from our data is that, for example, categorizers in tagging systems do not start with a fixed vocabulary, but introduce categories over time. For the first bookmarks of users, M_{vocab} exhibits only marginal differences between categorizers and describers, since
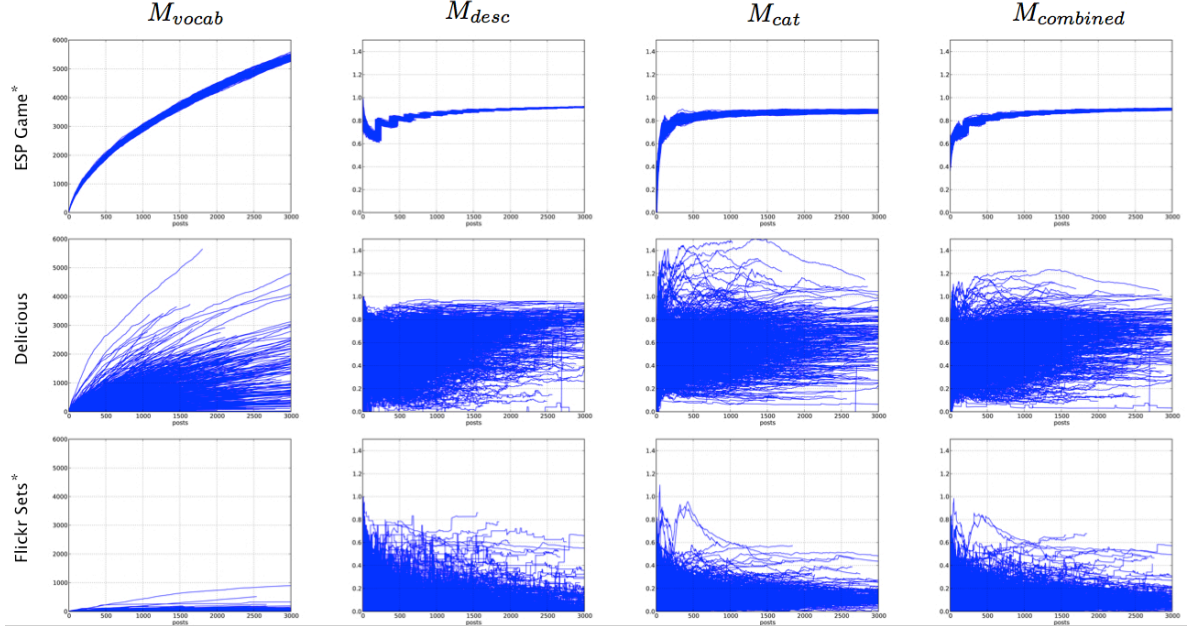
Figure 2: Overview of the introduced measures (from left to right: M_{vocab}, M_{desc}, M_{cat} and $M_{combined}$) over time for the two synthetic datasets (top and bottom row) and the Delicious dataset (middle row). The synthetic datasets form approximate upper- and lower bounds for "real" tagging datasets.

not a huge variety of different behavior can be examined on the y-scale. For the other measures the variety is reflected immediately in the figures. In addition, tag ratio is influenced by the number of tags that users assign to resources, which can vary from user to user regardless of their tagging motivation. To address some of these issues, we will present more sophisticated measures that follow different intuitions about categorizers and describers next.

5.2. Orphaned Tags - M_{desc}

Intuition: Categorizers would have a low interest in introducing orphaned tags, i.e. tags (or categories) that are only used for a single resource, and are then abandoned. Introducing categories of size one seems counterintuitive for categorizers, as it would prevent them from using their set of tags for browsing efficiently. On the other hand, describers would not refrain from introducing orphaned tags but would rather actively introduce a rich variety of different tags to resources so that re-finding becomes easier when searching for them later. Following this intuition, describers can be assumed to have a tendency to produce more orphaned tags than categorizers.

Description: The number of tags that are used only once within each personomy in relation to the total number of tags would be a simple measure to detect this difference in motivation. A more robust approach to measuring the influence of orphaned tags is described as follows:

$$M_{desc} = \frac{|\{t : |R(t)| \le n\}|}{|T|}, n = \lceil \frac{|R(t_{max})|}{100} \rceil (2)$$

We introduce a measure M_{desc} that aims to capture ideal describer behavior based on orphaned tags. In the formula $R(t)$ denotes the number of resources a tag t annotates. The measure ranges from 0 to 1, where a high value reflects a large number of tags that are used rarely. The threshold n for rare tags is derived from the individual users' tagging style, where t_{max} is the most frequently used tag of the user. The M_{desc} measure can be interpreted as the proportion of tags within the first percentile of the tag frequency histogram. It is more robust than a measure focusing on tags with $frequency = 1$ because it is not only influenced by those tags, but by its relation to the most popular tag of a given user. A high M_{desc} value can be expected to indicate that the main motivation of a user is to describe the resources.

8

The second column from the left in Figure 2 shows that M_{desc} (orphaned tags) indeed captures a difference between extreme describers (ESP Game on top) and extreme categorizers (Flickr photo sets on the bottom). Again, data from Delicious show a vast variety of different rates of orphaned tags, suggesting that user motivation varies significantly within this system.

Discussion: The M_{desc} ranges from 0 to 1 where 0 represents extreme categorizing behavior and 1 indicates extreme descriptive behavior. The resulting value gives insight into how a user re-uses tags in her library.

5.3. Tag Entropy - M_{cat}

Intuition: For categorizers, useful tags should be maximally discriminative with regard to the resources they are assigned to. This would allow categorizers to effectively use tags for navigation and browsing. This observation can be exploited to develop a measure for tagging motivation when viewing tagging as an encoding process, where entropy [26] can be considered a measure of the suitability of tags for this task. A categorizer would have a strong incentive to maintain high tag entropy (or information value) in his tag cloud. In other words, a categorizer would want the tag-frequency distribution as equally distributed as possible in order for her to be useful as a navigational aid. Otherwise, tags would be of little use in *browsing*. A describer on the other hand would have little interest in maintaining high tag entropy as tags are not used for navigation at all.

Description: In order to measure the suitability of tags to navigate resources, we develop an entropy-based measure for tagging motivation, using the set of tags and the set of resources as random variables to calculate conditional entropy. If a user employs tags to encode resources, the conditional entropy should reflect the effectiveness of this encoding process. The measure uses a user's tags T_u and her resources R_u as input respectively:

$$H(R|T) = -\sum_{r \in R}\sum_{t \in T} p(r,t) log_2(p(r|t)) \qquad (3)$$

The joint probability $p(r,t)$ depends on the distribution of tags over the resources. The conditional entropy can be interpreted as the uncertainty of the resource that remains after knowing the tags. The conditional entropy is measured in bits and is influenced by the number of resources and the tag vocabulary size. To account for individual differences

in users, we propose a normalization of the conditional entropy so that only the encoding quality remains. As a factor of normalization we can calculate the conditional entropy $H_{opt}(R|T)$ of an ideal categorizer, and relating it to the actual conditional entropy of the user at hand. Calculating $H_{opt}(R|T)$ can be accomplished by modifying $p(r,t)$ in a way that reflects a situation where all tags are equally discriminative while at the same time keeping the average number of tags per resource the same as in the user's personomy. Based on this approach, we define a measure for tagging motivation by calculating the difference between the observed conditional entropy and the conditional entropy of an ideal categorizer put in relation to the conditional entropy of the ideal categorizer:

$$M_{cat} = \frac{H(R|T) - H_{opt}(R|T)}{H_{opt}(R|T)} \qquad (4)$$

In Figure 2 the conditional tag entropy M_{cat} is shown in the second column from the right. It captures the differences between the synthetic, extreme categorizers and describers well, evident in the data from the ESP game (top) and Flickr sets (bottom). Again, real world tagging behavior shows significant variation.

Discussion: Conditional Tag Entropy ranges from 0 for users that behave exactly like an ideal categorizer, and can reach values greater than 1 for extreme non-categorizing users. The resulting value is thus inversely proportional to users' encoding effectiveness. The results presented in Figure 2 show that Delicious users exhibit traits of categorizers to varying extent, with some users showing scores that we calculated for ideal categorizers (lower part of the corresponding diagram). Further, it can be seen that the M_{cat} measure identifies users of the ESP game as extreme describers whereas users of the Flickr sets dataset are found to be extreme categorizers. An interesting aspect is that in the Delicious dataset users are included which outperform the extreme synthetic describers.

5.4. A Combination of Measures - $M_{combined}$

The final proposed measure is a combination of both M_{desc} and M_{cat} because (i) these measure both capture assumptions about intuitions that we except to see in describers and categorizers and (ii) they are not dependent on tagging styles (such as

the number of tags per resource) that might vary across describers and categorizers.

$$M_{combined} = \frac{M_{desc} + M_{cat}}{2} \qquad (5)$$

The right-most column of Figure 2 depicts the temporal behavior of $M_{combined}$ for the two synthetic datasets and the Delicious users.

5.5. Evaluation: Usefulness of Measures

The introduced measures have a number of useful properties: they are content-agnostic and language-independent, and they operate on the level of individual users. An advantage of content-agnostic measures is that they are applicable across different media (e.g. photos, music and text). Because the introduced measures are language-independent, they are applicable across different user populations (e.g. English vs. German). Because the measures operate on an personomy level, only tagging records of individual users are required (as opposed to comparing a users' vocabulary to the vocabulary of an entire folksonomy).

Figure 3 depicts M_{desc} and M_{cat} measures for all tagging datasets at $|R_u| = 500$, i.e. at the point where all users have bookmarked exactly 500 resources. We can see that both measures identify synthetic describer behavior (ESP game, upper left) and synthetic categorizer behavior (Flickr photosets, lower right) as extreme behavior. We can use the synthetic data as points of reference for the analysis of real tagging data, which would be expected to lie in between these points of reference. The diagrams for the real-world datasets show that tagging motivation in fact mostly lies in between the identified extremes. The fact that the synthetic datasets act as approximate upper and lower bounds for real-world datasets is a first indication for the usefulness of the presented measures. We also calculated Pearson's correlation between M_{desc} and M_{cat}. The results presented in Figure 3 are encouraging because the measures were independently developed based on different intuitions and yet they are highly correlated.

Figure 4 presents a different visualization for selected datasets to illustrate how tagging motivation varies within and across social tagging systems. Each row shows the distribution of users for a particular dataset according to $M_{combined}$. Again, we see that the profiles of Delicious, Diigo and Movielens as well as other datasets largely lie in between

these bounds. It can be seen that users from the Diigo system tend to be more driven by description behavior whereas users from Movielens seem to use their tags in a categorization manner. An interesting aspect that was not evaluated during our experiments was the influence on user interface characteristics on the behavior of users. The characteristic distribution of different datasets provides first *empirical* insights into the fabric of tagging motivation in different systems, illustrating a broad variety within these systems. Each row shows the distribution of users of a particular dataset according to the mean of the two measures M_{desc} and M_{cat}, binned into 5 equal bins in the interval [0..1]. From the diagram we can see that two profiles - the profiles for the two synthetic datasets - the ESP Game (back row) and the Flickr Photosets (second back row) - stand out drastically in comparison to all other datasets. While almost all users of the Flickr photoset dataset fall into the first bin indicating that this bin is characteristic for categorizers (second back row), almost all users of the ESP Game fall into the fifth bin indicating that this bin is characteristic for describers (back row). Using this data as a point of reference, we can now characterize tagging describermotivation in the profiles of the remaining datasets. The profiles of Flickr tags, Delicious and Bibsonomy largely lie in between these bounds, and thereby provide insights into the fabric of tagging motivation in those systems.

In addition, we want to evaluate whether individual users that were identified as extreme categorizers / extreme describers by $M_{combined}$ were also confirmed as such by human subjects. In our evaluation, we asked one human subject (who was not related to this research) to classify 40 exemplary tag clouds into two equally-sized piles: a categorizer and a describer pile. The 40 tag clouds were obtained from users in the Delicious dataset, where we selected the top20 categorizers and the top20 describers as identified by $M_{combined}$. The inter-rater agreement kappa between the results of the human subject evaluation and $M_{combined}$ was 1.0. This means the human subject agrees that the top20 describers and the top20 categorizers (as identified by $M_{combined}$) are good examples of extreme categorization / description behavior. The tag clouds illustrated earlier (cf. Figure 1) were actual examples of tag clouds used in this evaluation.

Finally, it is interesting to investigate the minimum number of tagging events that is necessary to approximate a users' motivation for tagging at
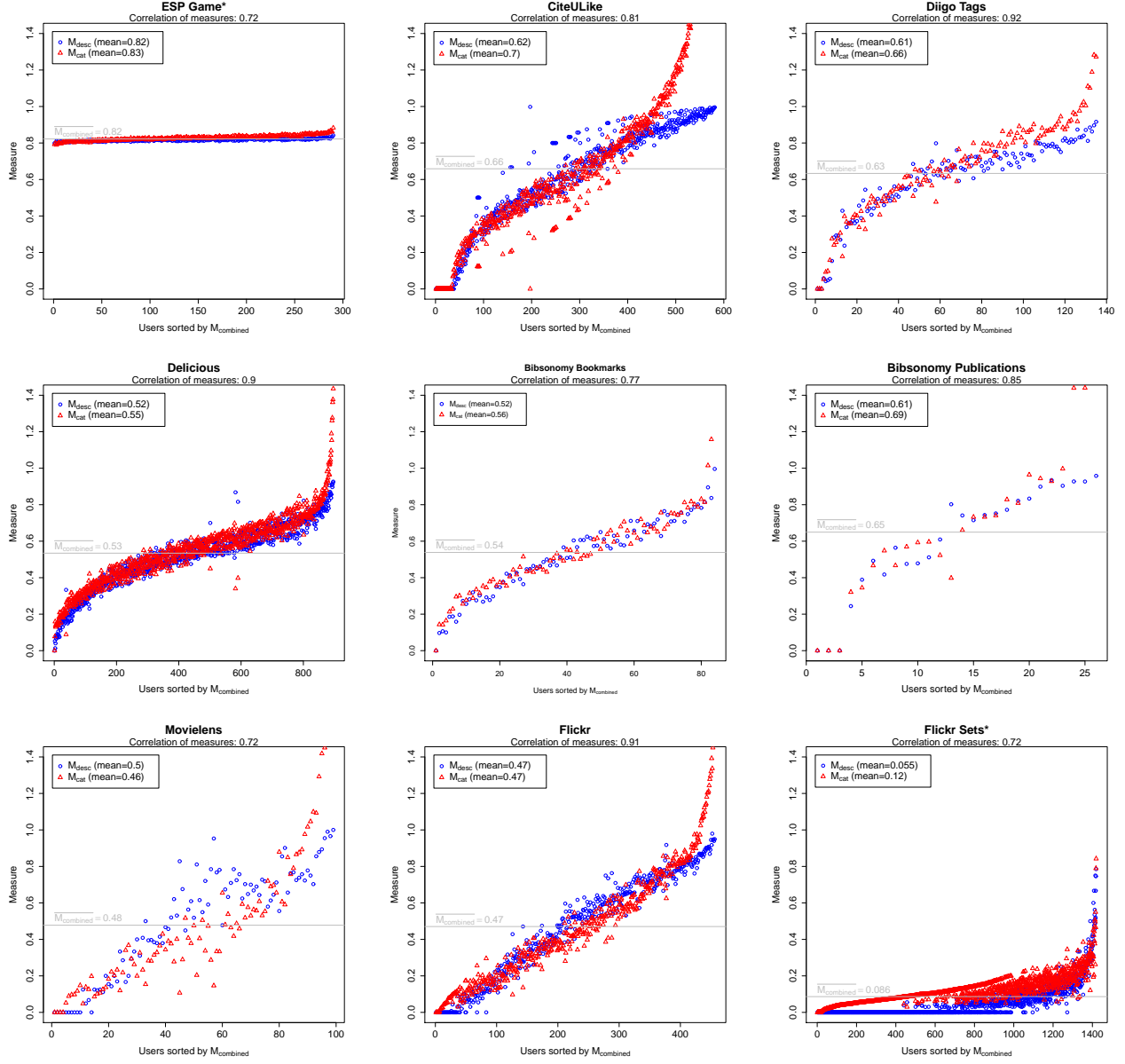
Figure 3: M_{desc} and M_{cat} at $|R_u| = 500$ for 9 datasets, including Pearson correlation and the mean value for $M_{combined}$. The top-left and the lower-right figure show the reference datasets for describers and categorizers respectively (designated with an asterisk).

Table 4: Classification results for <500 resources according to the $M_{combined}$ measure on the delicious dataset. The identified user behavior at 500 resources was used as a ground truth for the previous steps.

	0	100	200	300	400
Correctly identified user behavior	50	77	84	90	91
Incorrectly identified user behavior	50	23	16	10	9

later stages. Figure 5 shows that for many users, $M_{combined}$ provides a good approximation of users' tagging motivation particularly at early phases of a user's tagging history. This knowledge might help
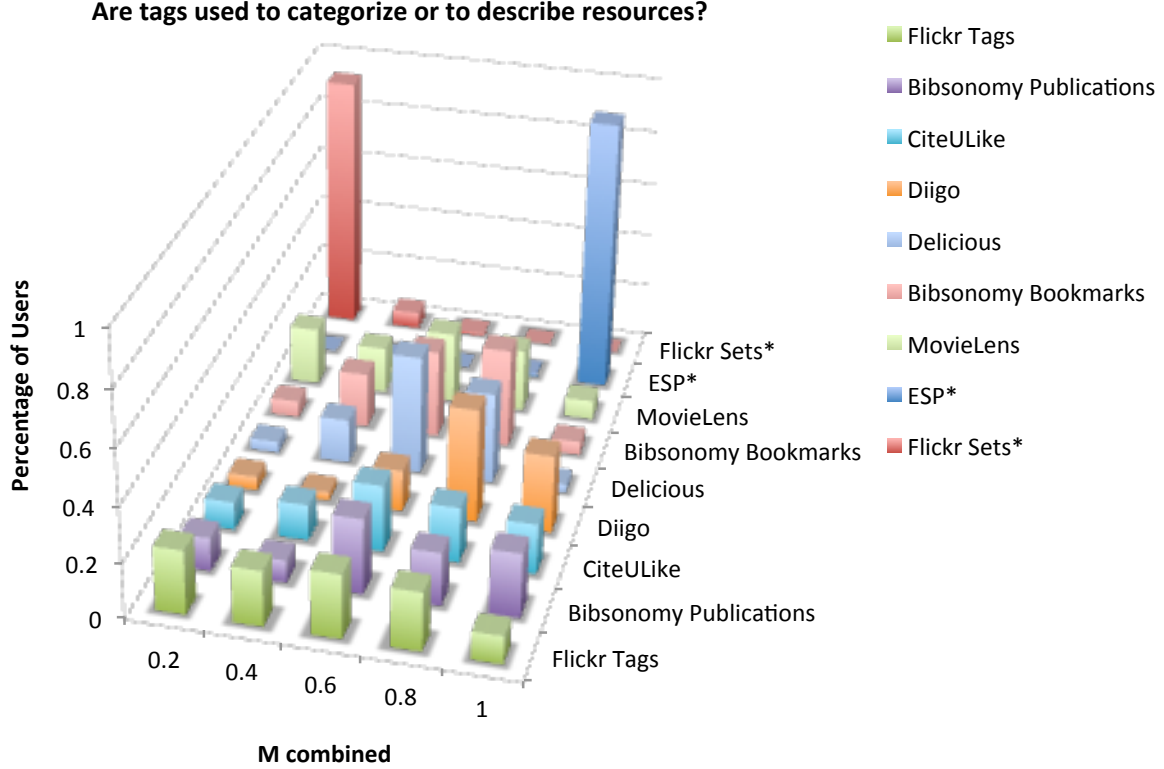
11

Figure 4: $M_{combined}$ at $|R_u| = 500$ for 9 different datasets, binned in the interval [0.0 .. 1.0]. The synthetic Flickr Sets dataset (back row) indicates extreme categorizer behavior, the synthetic ESP game dataset (second back row) indicates extreme describer behavior (designated by an asterisk).

to approximate users' motivation for tagging over time. In Table 4, we show the classification accuracy of our approach using the classification of users at 500 resources as our ground truth. Our results provide first insights into how reducing the number of required resources might influence classification accuracy. The results presented suggest that using 100 instead of 500 tagged resources would introduce 23 false classifications. However, more work is required to address this question in a more comprehensive way.

5.6. Other Potential Measures

The three introduced measures for tagging motivation were developed independently, using different intuitions. While the growth of the tag vocabulary is a measure that can be regarded to be influenced by many latent user characteristics (e.g. the rate of tags/resource), M_{desc} and M_{cat} are intended to normalize behavior and account for individual differences of tagging behavior not related

to tagging motivation. M_{desc} was motivated by the intention to identify describer behavior while M_{cat} was motivated by the intention to identify categorizer behavior. For M_{desc} and M_{cat}, small values (close to 0) can be seen as indicators for a tagging behavior mainly motivated to categorize resources and high values can be seen as an indicator for describers.

Other ways of measuring user motivation include tapping into the knowledge contained in resources themselves or relating individual users' tags to tags of the entire folksonomy. To give an example:

A conceivable measure to identify describers would calculate the overlap between the tag vocabulary and the words contained in a resource (e.g. a URL or scientific paper), as proposed by [27]. According to our theory, tags produced by describers would be expected to have a higher vocabulary overlap than the ones of categorizers. One problem of such an approach is that it would rely on resources being available in the form of textual
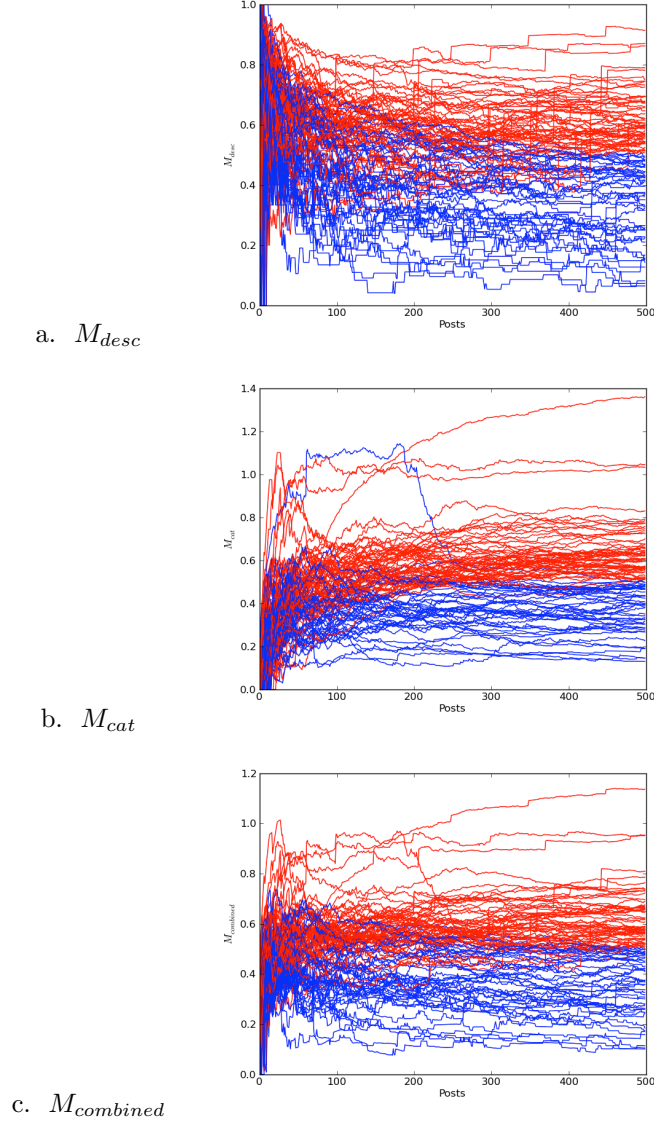
a. M_{desc}

b. M_{cat}

c. $M_{combined}$

Figure 5: M_{desc} (a.), M_{cat} (b.) and $M_{combined}$ (c.) in the interval $|R_u| = [0..500]$ for 100 random users obtained from the delicious dataset. The 100 users are split into two equal halves at $|R_u| = 500$ according the the corresponding measure, with the upper half colored red (50 describers) and the lower half colored blue (50 categorizers). Due to our particular approach, it is obvious that all measures exhibit perfect separation at $|R_u| = 500$, but $M_{combined}$ (right) appears to exhibit faster convergence and better separability in early phases of a user's tagging history, especially for small $|R_u|$ (such as $|R_u| < 50$).

data, which limits its use to tagging systems that organize textual resources and prohibits adoption for tagging systems that organize photos, audio or video data.

Another measure might tap into the vocabulary of folksonomies: if a user tends to use tags that are frequently used in the vocabulary of the entire social folksonomy, this might serve as an indicator for her to be motivated by description, as she orients her vocabulary towards that of the crowd consensus. If she does't, this might serve as an indicator for Categorization. Problems of such an approach include that it is a) language dependent, and b) it requires global knowledge of the vocabulary of a folksonomy.

By operating on a *content-agnostic*, *user-specific* and *language-independent* level, the measures introduced in this paper can be applied to the tagging

records of a single user, without the requirement to crawl the entire tagging system or to have programmatic access to the resources that are being tagged.

6. Influence on Tag Agreement

Given that tagging motivation varies even within social tagging systems, we are now interested in whether such differences influence resulting folksonomies. We can assert an influence if folksonomies primarily populated by categorizers would differ from folksonomies primarily populated by describers in some relevant regard. We investigate this question next.

In order to answer this question, we selected the Delicious dataset for further investigations. This was done for the following reasons: firstly the dataset exhibits a broad range of diverse behavior with regard to tagging motivation. Secondly, Delicious is a broad folksonomy[3] allowing multiple users to annotate a single resource and therefore generating a good coverage of multiply tagged documents. Other datasets represent narrow folksonomies - such as flickr - where most resources are tagged by a single user only, or do not exhibit sufficient overlap to conduct meaningful analysis.

To explore whether tagging motivation influences the nature of tags, we split the Delicious user base into two groups of equal size by selecting a threshold of 0.5514 for the $M_{combined}$ measure: the first group is defined by users who score lower than the threshold. We refer to this group as *Delicious categorizers* from now on. The second group is defined as the complementary set of users, with an $M_{combined}$ greater than 0.5514. We refer to this group as *Delicious describers*. This allows us to generate two separate tag sets for each of the 500 most popular resources contained in our dataset, where one tag set is constructed by Delicious describers and the other by Delicious categorizers. For each tag set, we can calculate tag agreement, i.e. the number of tags that k percent of users agree on for a given resource.

We define agreement on tags in the following way:

$$T_a = t \in T_r, \frac{|U_t|}{|U_r|} > k \qquad (6)$$

In this equation U_t denotes the number of users that have a resource annotated with the tag t and U_r represents the number of users that have the resource in their library. $|T_a|$ describes the number of tags that k percent of users agree on. We restrict our analysis to $|U_t| > 3$ to avoid irrelevant high values of tag agreement when only few users are involved. From a theoretical point of view, we can expect that describers would achieve a higher degree of agreement compared to categorizers, as describers focus on describing the content of resources. This is exactly what can be observed in our data.

Table 6 shows that among the 500 most popular resources in our dataset, describers achieve a significantly higher tag agreement than categorizers ('describer wins' in the majority of cases). For $k=30\%$ we can see that describers win in 94.2% of cases (471 out of 500 resources), categorizers win 1% of cases (5 resources) and 4.8% of cases (24 resources) are a tie. Other k produce similar results. We have conducted similar analysis on the CiteULike dataset, which has a much smaller number of overlapping tag assignments, and therefore is less reliable. The results from this analysis is somewhat conflicting - further research is required to study the effects more thoroughly. However, we believe that our work has some important implications: the findings suggest that when considering the users' motivation for tagging, not all tags are equally meaningful.

Depending on a researcher's task, a researcher might benefit from focusing on data produced by describers or categorizers only. To give an example: for tasks where high tag agreement is relevant (e.g. tag-based search in non-textual resources), tags produced by describers might be more useful than tags produced by categorizers. In related work, we showed that the emergent semantics that can be extracted from folksonomies are influenced by our distinction between different motivations for tagging (categorizers vs. describers) [28].

7. Limitations

Our work focuses on understanding tagging motivation based on a distinction between two categories, i.e. categorization vs. description, and does not yet deal with other types of motivations, such as *recognition*, *enjoyment*, *self-expression* and other categories proposed in the literature (cf. [16, 17, 29, 18]). Future work should investigate

Table 5: Tag agreement among Delicious describers and categorizers for 500 most popular resources. For all different k, describers produce more agreed tags than categorizers. Results for $k > 70$ are dominated by ties.

k	10	20	30	40	50	60	70	80	90
Desc. Wins	**378**	**463**	**471**	**452**	**380**	**286**	172	74	23
Cat. Wins	56	12	5	7	5	3	4	4	0
Ties	66	25	24	41	115	211	**324**	**422**	**477**

the degree to which other categories of tagging motivation are amenable to quantitative analysis. This work is the first attempt towards a viable way of automatically and quantitatively understanding tagging motivation in social tagging systems.

Because we distinguish between describers and categorizers based on observing their tagging behavior over time, a minimum number of resources per user is required to observe sufficient differences. In our case, we have required a $|R_u|$ min of 500 respectively 1,000 resources. In some tagging applications, this requirement is likely to be too high (e.g. youtube.com). However, from our data - especially from our analysis presented in Figure 5 - we feel that measuring tagging motivation for $|R_u| < 500$ appears feasible as well, but the introduced measures might produce results that are of lower accuracy, making analysis of tagging motivation more challenging. While we regard this issue to be important, we leave it to future work.

Furthermore, all analyses on "real" social tagging datasets conducted in this work focused on analyzing publicly available tagging data only - private bookmarks are not captured by our datasets. To what extent private bookmarks might influence investigations of tagging motivation remains an open question.

8. Contribution and Conclusions

The contribution of this paper is twofold: firstly the paper presents the first thorough literature survey on the topic of tagging motivation. Secondly, we show a preliminary study presenting measures for approximating tagging motivation.

This paper presented results from a large empirical study involving data from seven different social tagging datasets. We introduce and apply an approach that aims to measure and detect the tacit nature of tagging motivation in social tagging systems. We have evaluated the introduced measures with synthetic datasets of extreme behavior as points of reference, via a human subject study

and via triangulation with previous findings. Based on a large sample of users, our results show that 1) tagging motivation of individuals not only varies across but also significantly within social tagging systems, and 2) that users' motivation for tagging has an influence on resulting tags and folksonomies. By analyzing the tag sets produced by Delicious describers and Delicious categorizers in greater detail, we showed that agreement on tags among categorizers is significantly lower compared to agreement among describers. These findings have some important implications:

Usefulness of Tags: Our research shows that users motivated by categorization produce fewer descriptive tags, and that the tags they produce exhibit a lower agreement among users for given resources. This provides further evidence that not all tags are equally useful for different tasks, such as search or recommendation. Rather the opposite seems to be the case: to be able to assess the usefulness of tags on a content-independent level, deeper insight into users' motivation for tagging is needed. The measures introduced in this paper aim to illuminate a path towards understanding user motivation for tagging in a quantitative, content-agnostic and language-independent way that is based on local data of individual users only. In related work, the introduced distinction between categorizers and describers was used to demonstrate that emergent semantics in folksonomies are influenced by the users' motivation for tagging [28]. However, the influence of tagging motivation on other areas of work, such as taxonomy learning from tagging systems [30], simulation of tagging processes [31] or tag recommendation [32], remains to be studied in future work.

Usage of Tagging Systems: While tags have been traditionally viewed as a way of freely describing resources, our analysis reveal that the motivation for tagging across different real world social tagging systems such as Delicious, Bibsonomy and Flickr varies tremendously. Our analysis corroborates previous qualitative research and pro-

vides first quantitative evidence for this observation. Moreover, our data shows that even within the same tagging systems the motivation for tagging varies strongly. Our research also highlights several opportunities for designers of social tagging systems to influence user behavior. While categorizers could benefit from tag recommenders that recommend tags based on their individual tag vocabulary, describers could benefit from tags that best describe the content of the resources. Offering users tag clouds to aid the navigation of their resources might represent a way to increase the proportion of categorizers, while offering more sophisticated search interfaces and algorithms might encourage users to focus on describing resources.

Studying the extent to which the motivation behind tagging influences other properties of social tagging systems such as efficiency [4], structure [33], emergent semantics [34] or navigability [35] represent interesting areas for future research.

9. Acknowledgements

References

[1] S. Golder, B. Huberman, Usage patterns of collaborative tagging systems, Journal of Information Science 32 (2) (2006) 198.

[2] C. Marlow, M. Naaman, D. Boyd, M. Davis, Ht06, tagging paper, taxonomy, flickr, academic article, to read, in: HYPERTEXT '06: Proceedings of the 17th Conference on Hypertext and Hypermedia, ACM, New York, NY, USA, 2006, pp. 31–40.

[3] P. Mika, Ontologies are us: A unified model of social networks and semantics, Web Semantics: Science, Services and Agents on the World Wide Web 5 (1) (2007) 5–15.

[4] E. Chi, T. Mytkowicz, Understanding the efficiency of social tagging systems using information theory, in: HYPERTEXT '08: Proceedings of the 19th ACM Conference on Hypertext and Hypermedia, ACM New York, NY, USA, 2008, pp. 81–88.

[5] Y. Yanbe, A. Jatowt, S. Nakamura, K. Tanaka, Can social bookmarking enhance search in the web?, in: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, ACM, 2007, p. 116.

[6] P. Heymann, G. Koutrika, H. Garcia-Molina, Can social bookmarking improve web search?, in: WSDM '08: Proceedings of the int'l Conference on Web Search and Web Data Mining, ACM, 2008, pp. 195–206.

[7] A. Hotho, R. Jäschke, C. Schmitz, G. Stumme, Information retrieval in folksonomies: Search and ranking, Lecture Notes in Computer Science 4011 (2006) 411.

[8] P. Heymann, H. Garcia-Molina, Collaborative creation of communal hierarchical taxonomies in social tagging systems, Tech. rep., Computer Science Department, Standford University (April 2006).
URL dbpubs.stanford.edu:8090/pub/2006-10

[9] M. Heckner, M. Heilemann, C. Wolff, Personal information management vs. resource sharing: Towards a model of information behaviour in social tagging systems, in: ICWSM '09: Int'l AAAI Conference on Weblogs and Social Media, San Jose, CA, USA, 2009.

[10] T. Hammond, T. Hannay, B. Lund, J. Scott, Social bookmarking tools (I), D-Lib Magazine 11 (4) (2005) 1082–9873.

[11] R. Wash, E. Rader, Public bookmarks and private benefits: An analysis of incentives in social computing, in: ASIS&T Annual Meeting, Citeseer, 2007.

[12] T. Coates, Two cultures of fauxonomies collide, http://www. plasticbag. org/archives/2005/06/two_cultures_of_fauxonomies_collide/. Last access: May 8 (2005) 2008.

[13] Z. Xu, Y. Fu, J. Mao, D. Su, Towards the semantic web: Collaborative tag suggestions, in: Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland, Citeseer, 2006.

[14] S. Sen, S. Lam, A. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. Harper, J. Riedl, Tagging, communities, vocabulary, evolution, in: Proceedings of the 2006 20th anniversary Conference on Computer Supported Cooperative Work, ACM, 2006, p. 190.

[15] M. Heckner, T. Neubauer, C. Wolff, Tree, funny, to_read, google: what are tags supposed to achieve? a comparative analysis of user keywords for different digital resource types, in: Proceedings of the 2008 ACM workshop on Search in social media, SSM '08, ACM, New York, NY, USA, 2008, pp. 3–10. doi:10.1145/1458583.1458589.
URL http://doi.acm.org/10.1145/1458583.1458589

[16] O. Nov, M. Naaman, C. Ye, Motivational, Structural and Tenure Factors that Impact Online Community Photo Sharing, in: ICWSM 2009: Proceedings of AAAI Int'l Conference on Weblogs and Social Media, 2009.

[17] M. Ames, M. Naaman, Why we tag: motivations for annotation in mobile and online media, in: CHI '07: Proceedings of the SIGCHI Conference on Human factors in Computing Systems, ACM, New York, NY, USA, 2007, pp. 971–980.
URL http://portal.acm.org/citation.cfm?id=1240624.1240772

[18] K. Bischoff, C. S. Firan, W. Nejdl, R. Paiu, Can all tags be used for search?, in: CIKM '08: Proceeding of the 17th ACM Conference on Information and Knowledge Management, ACM, New York, NY, USA, 2008, pp. 193–202.
URL http://portal.acm.org/citation.cfm?id=

1458112

[19] E. Rader, R. Wash, Influences on tag choices in del.icio.us, in: CSCW '08: Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work, ACM, New York, NY, USA, 2008, pp. 239–248.
URL http://portal.acm.org/citation.cfm?id=1460563.1460601&coll=Portal&dl=GUIDE&CFID=31571793&CFTOKEN=71651308

[20] C. Körner, R. Kern, H. P. Grahsl, M. Strohmaier, Of categorizers and describers: An evaluation of quantitative measures for tagging motivation, in: 21st ACM SIGWEB Conference on Hypertext and Hypermedia (HT 2010), Toronto, Canada, ACM, Toronto, Canada, 2010.

[21] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, J. Riedl, tagging, communities, vocabulary, evolution, in: CSCW '06: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, ACM, New York, NY, USA, 2006, pp. 181–190.
URL http://portal.acm.org/citation.cfm?id=1180904

[22] M. Strohmaier, C. Körner, R. Kern, Why do users tag? Detecting users' motivation for tagging in social tagging systems, in: International AAAI Conference on Weblogs and Social Media (ICWSM2010), Washington, DC, USA, May 23-26, 2010.

[23] J. Sinclair, M. Cardew-Hall, The folksonomy tag cloud: when is it useful?, Journal of Information Science 34 (1) (2008) 15.

[24] R. Lambiotte, M. Ausloos, Collaborative tagging as a tripartite network, Lecture Notes in Computer Science 3993 (2006) 1114.

[25] A. Hotho, R. Jäschke, C. Schmitz, G. Stumme, Bib-Sonomy: A social bookmark and publication sharing system, in: Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th Int'l Conference on Conceptual Structures, Citeseer, 2006, pp. 87–102.

[26] T. M. Cover, J. A. Thomas, Elements of information theory, Wiley-Interscience, New York, NY, USA, 1991.

[27] U. Farooq, T. Kannampallil, Y. Song, C. Ganoe, J. Carroll, L. Giles, Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics, in: Proceedings of the 2007 int'l ACM Conference on Supporting Group Work, ACM, 2007, pp. 351–360.

[28] C. Körner, D. Benz, M. Strohmaier, A. Hotho, G. Stumme, Stop thinking, start tagging - tag semantics emerge from collaborative verbosity, in: Proceedings of the 19th International World Wide Web Conference (WWW 2010), ACM, Raleigh, NC, USA, 2010.

[29] Z. Xu, Y. Fu, J. Mao, D. Su, Towards the semantic web: Collaborative tag suggestions, in: Proceedings of the Collaborative Web Tagging Workshop at the WWW 2006, Edinburgh, Scotland, 2006.
URL http://.inf.unisi.ch/phd/mesnage/site/Readings/Readings.html

[30] T. Eda, M. Yoshikawa, T. Uchiyama, T. Uchiyama, The effectiveness of latent semantic analysis for building up a bottom-up taxonomy from folksonomy tags, World Wide Web 12 (4) (2009) 421–440.
URL http://dx.doi.org/10.1007/s11280-009-0069-1

[31] P. A. Chirita, S. Costache, W. Nejdl, S. Handschuh, P-tag: large scale automatic generation of personalized annotation tags for the web, in: WWW '07: Proceedings of the 16th international conference on World Wide Web, ACM, New York, NY, USA, 2007, pp. 845–854.
URL http://portal.acm.org/citation.cfm?id=1242686&dl=GUIDE,

[32] A. Shepitsen, J. Gemmell, B. Mobasher, R. Burke, Personalized recommendation in social tagging systems using hierarchical clustering, in: Proceedings of the 2008 ACM conference on Recommender systems, ACM, 2008, pp. 259–266.

[33] S. Golder, H. B.A., Usage patterns of collaborative tagging systems, Journal of Information Science 32 (2) (2006) 198–208.
URL http://jis.sagepub.com/cgi/doi/10.1177/0165551506062337

[34] C. Körner, D. Benz, A. Hotho, M. Strohmaier, G. Stumme, Stop thinking, start tagging: Tag semantics arise from collaborative verbosity, in: 19th International World Wide Web Conference (WWW2010), Raleigh, NC, USA, April 26-30, ACM, 2010.

[35] D. Helic, C. Trattner, M. Strohmaier, K. Andrews, On the navigability of social tagging systems, in: The 2nd IEEE International Conference on Social Computing (SocialCom2010), Aug 20-22, 2010, Minneapolis, Minnesota, USA, 2010.