Imitation and Quality of Tags in Social Bookmarking Systems – Collective Intelligence Leading to Folksonomies

Fabian Floeck, Johannes Putzke, Sabrina Steinfels, Kai Fischbach, Detlef Schoder

University of Cologne, Department of Information Systems and Information Management, Pohligstr. 1, 50969 Cologne (Köln), Germany
{Putzke, Steinfels, Fischbach, Schoder}@wim.uni-koeln.de

Abstract. Social bookmarking platforms often allow users to see a list of tags that have been used previously for the webpage they are currently bookmarking, and from which they can select. In this paper, the authors analyze the influences of this feature on the tag categorizations resulting from the collaborative tagging effort. The main research goal is to show how the interface design of social bookmarking systems can influence the quality of the collective output of their users. Findings from a joint research project with the largest Russian social bookmarking site *BobrDobr.ru* suggest that if social bookmarking systems allow users to view the most popular tags, the overall variation of keywords used that are assigned to websites by all users decreases.

Keywords: Collective Intelligence, Collaborative Tagging, Folksonomies, Shared Knowledge, Social-Bookmarking-Systems

1 Introduction

In recent years, social bookmarking services such as *BobrDobr.ru*, *citeulike.org*, *del.icio.us*, and *mister-wong.de* have gained an increasingly large user base [e.g. 1]. By using a social bookmarking service, users can bookmark objects on the World Wide Web – identified by their Unified Resource Locators (URLs) – and annotate each object with metadata, or so-called "tags". A tag is a keyword that describes the annotated object from the user's point of view. The process of many users assigning arbitrary tags to shared objects is often called "collaborative tagging," and the set of tags that results is typically denoted "folksonomy" [2, 3].[1]

Considerable research has been devoted to the suitability of folksonomies for content classification, and particularly to the tradeoff between the users' bottom-up approach of assigning free keywords for classification and the quality of top-down-

[1] Since the term "folksonomy is said to have been coined by Vander Wal, the authors reference some of Vander Wal's blog entries. However, there are many other (academic) definitions of folksonomies and their characteristics [e.g. 4].

defined classifications created by experts [e.g. 5]. However, rather less attention has been paid to whether the quality of folksonomies depends on the functionalities offered by the bookmarking system used to create them. In this paper, the authors analyze empirically the influence of functionalities that allow users to see which tags others have assigned to a URL during the process of bookmarking on the resulting set of tags of these URLs.

The remainder of this paper is structured as follows. Section 2, Related Work, reviews the literature that examines the quality of folksonomies and their suitability for content classification, as well as works that deal with the characteristic and effects of imitation. Through this, possible scenarios for the effects of imitation in a social bookmarking system are drawn. In Section 3, Hypotheses, the authors develop four hypotheses to test the impact of the visibility of the most popular tags for a certain URL on the variation of tags assigned by the users for this URL. Section 4, Analyses, describes the dataset, the methodology, and the operationalization of variables. In Results (Section 5), the authors describe the procedures used for the empirical analyses and highlight the findings. Finally, Section 6, Summary and Conclusions, discusses the implications of the findings, notes their limitations, and provides some suggestions for further research.

2 Related Work

Folksonomies are one of the current research trends in a variety of academic disciplines such as information systems [e.g. 6, 7], computer science [e.g. 8, 9], physics [e.g. 10, 11], anthropology, and sociology [e.g. 12]. Hence, a search for “social bookmarking”, “social bookmark”, “folksonomy” or “social tagging” in the title / keywords or abstract fields in “Web of Science”[2] yielded 306 articles published between 2005 and 2009.[3] Since it is not possible to review the entire related literature here, the literature review will focus on studies that examine the emergence of folksonomies in social bookmarking systems [e.g. 13]. Before taking this specific focus, the authors also provide some insights into the applications and uses of social bookmarking services.

There are different potential usage scenarios for tags and social bookmarking services. The most evident applications might be the use of folksonomies for web search optimization [e.g. 14, 15] and knowledge organization [e.g. 16], since folksonomies can facilitate detection of non-explicit properties of web objects. For example, a user can characterize a web page by annotating it with the tag “funny” [13]. An automated system that extracts existing metadata from web pages cannot achieve this, as the system cannot comprehend this property. Therefore, folksonomies might also be a useful approach for the semantic web [e.g. 17].

Social bookmarking services are typically free of charge and require little knowledge to use. This increases active participation by many users [18], who not

[2] http://apps.isiknowledge.com/.

[3] Note that this keyword search did not include the terms “ontology” or “classification”.

only manage their own bookmarks with social bookmarking services, but also use the service to search in other user's public bookmarks for web objects with specific properties, represented by tags that serve as a filter. This allows all users of the service to benefit from all other bookmarks [19] in addition to the individual benefits of users from their own bookmarks [18].

Folksonomies resemble a "bottom-up" approach for generating a vocabulary. Some researchers argue that tags, unlike controlled vocabularies or hierarchical taxonomies, more accurately reflect users' conceptual models. The idea is that every user adds his or her individual conceptual model of a piece of content to the pool of tag descriptions, which in turn is accounted for in the aggregated description of the content. Thus in social bookmarking systems, a large number of users – each investing their own cognitive resources – is collectively making sense of specific web contents. The result might be considered a "democratically agreed upon" description of content.

For the social bookmarking system *del.icio.us*, Golder and Huberman [13] found that after a certain number of bookmarks has been created for a specific URL, a stable power law pattern with fixed proportions of frequency for each tag can be observed, no matter how many bookmarks are stored for that particular URL after that point. They relate the emergence of this stable pattern to two basic explanations.

The first explanation is "shared knowledge". Golder and Huberman [13] argue that sharing the same experiences that may be universal within a culture or community leads to similar ways of sensemaking. As a consequence, categories emerge that are widely agreed upon, co-existing with personal categories that are rarely reproduced. Accordingly, experiments showed that basic-level categories are the most probable classifications made when objects are perceived for the first time [20], although an individual's expertise in a specific field influences what she or he considers to be a basic category [21]. As the variation of basic-level categories is lower the more general they are, and more general descriptions are preferred, users agree on general descriptions but differ on more specific ones. Other researchers trace the users' behavior of preferring broad and simple categories over specific ones back to the aim of investing the "least cognitive effort" [e.g. 4]. Thus, following the idea of a shared conceptual model, the power law tag pattern for given content may, over time and with more users bookmarking, come closer to a realistic representation of what the user base collectively "thinks" that content is about.

The second explanation for this pattern is the presence of *imitation* behaviour. Some tags (e.g., basic categories) are popular; by being imitated, they become even more popular over time, eventually forming and amplifying the power law pattern. Imitation behaviour can also be traced back to the need to save cognitive ressources [4] and is facilitated by a feature common to many bookmarking systems that makes it possible to see the most popular tags assigned to a URL. Users are looking for the best tag choices without having to think too much about them, and they are often indecisive about the right tagging choice. Therefore, they trust the "social proof" [22] offered by the decisions of the majority. It is not a wish to conform that explains this "rational imitation" [23], but rather the need of the individual user to arrive at a better tagging decision.

What results when users "follow the behaviour of the preceding individual" [24] or individuals is an "information cascade". With every imitated decision, the individual puts aside her or his individual tagging choices and the stability of the information cascade grows as the popular tags (e.g., basic categories) become even more popular. As Bikhchandani et al. [24, pp. 1006, 1009] put it: "Intuitively, cascades aggregate the information of only a few early individuals' actions." The result: "The social cost of cascades is that the benefit of diverse information sources is lost." This means that only the conceptual models of a few early individuals are the basis of the emerging collective tag description. Individual viewpoints expressed or mistakes made in the early bookmarks are consequently represented at a disproportionately high rate in the eventual distribution.

Note, however, that an information cascade in a social bookmarking system can take two forms, the latter one being different from the described process:

1. Replacement of individual tags by popular tags: the overall tag variation for a URL decreases because users do not think about tags of their own or neglect their own tag choices in favor of the tag choices of the majority. Diverse information is lost.
2. Extension of individual tags by popular tags: When opting for a popular tag or set of tags, users still assign their own full sets of tags to the URL. Here, no information is lost; there is, however, an overrepresentation of the early assigned tags that effects, for example, the display of tag clouds for navigation purposes.

These two scenarios (see Fig. 1) are not mutually exclusive but rather occur simultaneously and interfere with each other even in one single bookmarking process.

The goal of this study was to find empirical evidence 1) that there is overrepresentation of early-assigned tags because they are imitated, and 2) that this imitation leads to a loss of information in the collective description because early tags suppress and eliminate individual tags.

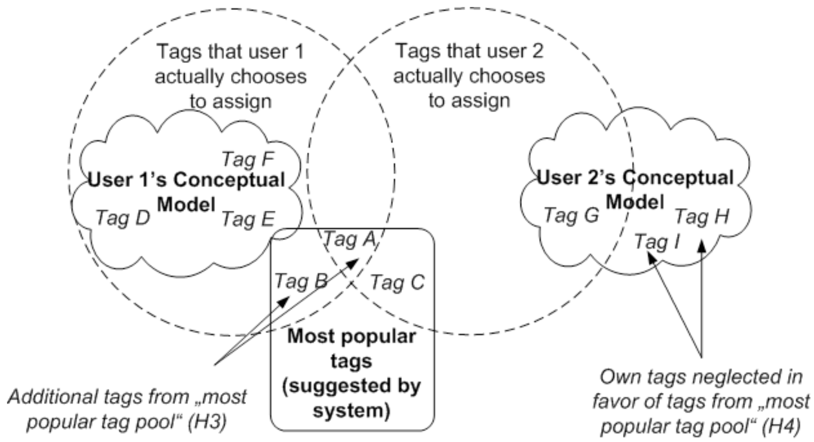

Fig. 1. Tag suggestions leading to more tags (H3) vs. suggestions leading to less tags (H4)

3 Hypotheses

Looking at the empirical findings and knowledge regarding shared categories highlighted in the previous section, the authors assume that there must exist categories or tags that users share and agree upon widely, as well as some they agree upon less – and the less they agree upon these, the more specific the tags become. The authors further assume that the tags that the users share and agree upon, therefore, will not only be assigned more often, but also will have a high probability of appearing quite early (compare [13]).[4] Hence,

H1*: There is a positive correlation between the time when a tag is first assigned to a URL and its overall frequency for that particular URL.*

The authors suppose that ***H1*** is true whether or not users imitate other users' tags. However, the effect postulated in hypothesis ***H1*** should be stronger when imitation is made possible by the user interface. Users then agree upon the same tags, but also, tags that are assigned frequently and early already are imitated because they are suggested as "most popular". As pointed out above, it is likely that an information cascade will ensue, resulting in a higher copy rate and thus an eventual overrepresentation of early-assigned tags. Therefore,

H2: *The positive correlation between the time when a tag is first assigned to a URL and its overall frequency for this URL is stronger if users can see tags already assigned by other users.*

Under the assumption that different users have different associations regarding a URL, it is very likely that the most popular tags contain tags a user had not thought of when he or she decided to bookmark a URL. On average, then, the user has a wider choice of tags when the most popular tags are displayed than without that functionality. The user may decide to choose some of the displayed tags, sometimes in addition to and sometimes instead of his or her own tags, as discussed in the previous section. Overall, users should assign more tags to a URL when they have the option to imitate (see Fig. 1). Hence,

H3*: A URL receives more tags from a single user if that user can see other users' previously assigned tags.*

A user provided with the most popular tags might, according to the concepts of the least cognitive effort and social proof, neglect some of his or her own conceptual models and the associated tags. Instead, the user will settle for some of the offered popular tags and regard them as a good choice. This leads to a lower degree of variation in the set of tags a URL receives overall. Therefore,

H4*: A URL receives fewer different keywords in the form of tags overall if that user can see other users' previously assigned tags.*

In the next section, we describe how we tested these hypotheses empirically based on the complete dataset of a social bookmarking service.

---

[4] This may sound trivial, but it is the basic correlation that must be shown to prove the assumptions the authors make for the imitation model.

4 Analyses

4.1 Dataset

To test the proposed hypotheses, the authors conducted a joint research project with the Russian social bookmarking platform *BobrDobr.ru*, which was selected as the data source for four main reasons:

First, *BobrDobr.ru* offers users two different possibilities for adding a bookmark to the data base: one allowing for imitation and one that leaves bookmarking up to users themselves, individually. These different procedures were crucial for the research design of this study. The two bookmarking methods work as follows. The first ("internal") method enables the user to copy a bookmark for a certain URL that has already been bookmarked by other users within the system. When a user chooses to copy and add the bookmark to her or his own library, the system displays to the bookmarking user the five tags used most often (to date) for that URL. With the second ("external") method, the user can either click on a special button installed in the web browser or a *BobrDobr.ru*-Link implemented by the website itself to bookmark the URL the user is currently visiting outside the system. No tags of other users are shown if URLs are bookmarked "externally" via the browser button. This setting allowed the authors to compare the sets of tags a URL had received from both methods: one set resulting from the "internal" method, that is, imitation of the most popular tags; and the other set resulting from the "external" method, which gave no insight into popular tags.[5]

Second, the authors wanted to analyze a social bookmarking service that is visited primarily by users from one language area only (unlike, for example, *del.icio.us*). *BobrDobr.ru* is the leading Russian social bookmarking platform[6] and has a relatively homogenous user base in terms of language due to the Cyrillic alphabet. According to the operators of *BobrDobr.ru*, 90 percent of users reside in Eastern Europe and 80 percent using a Russian system.

Third, except than the language, the platform has neither a specific target user group nor restrictions in terms of content (like, for example, *citeulike.org*).

Fourth, the operators of *BobrDobr.ru* were willing to conduct a joint research project and provided the researchers full access to their database. Furthermore, the operators allowed the researchers to determine which data they would like collected, and the operators stored these data persistently for the duration of the research (e.g. the time stamps for each action).

The data collection began on the date of the relaunch of *BobrDobr.ru* on 1 May 2008, when the two different bookmarking procedures ("external" and "internal")

---

[5] Users of *BobrDobr.ru* also have a third option to add a bookmark to the database: they can import their collections of browser bookmarks. The system then automatically creates tags from the directory names in which the imported URLs were formerly stored in the browser. The authors explain how they dealt with the analysis of imported bookmarks and their related tags in the following paragraphs of this section.

[6] For example, *Bobdobr.ru*'s traffic rank on *alexa.com* is 471 (accessed on 3/12/2009).

were introduced, and ended on 20 August 2008. The dataset contained logs for all user activity during the observation period. At the end of the data collection period, the platform had about 61,000 registered users who created about 4,000 bookmarks per day. As the authors only analyzed URLs that were first bookmarked after the relaunch date, they were able to observe the development of tag sets assigned to URLs right from the first bookmark they ever received and still analyze a large system with many users. Before analyzing the imitation of tags, the researchers also had to account for the fact that *BobrDobr.ru* offers users the option of saving bookmarks as either "public" or "private", the latter resulting in saved bookmarks not being displayed to other users nor being included among the "most popular tags" feature. Private bookmarks and their tags were, therefore, removed from the dataset, as their analysis would not contribute to the research questions. After their removal, a total of 110,740 stored URLs remained.

The authors also had to consider how to deal with imported bookmarks that were automatically tagged by the system (option 3). The automatic tagging process works as follows. If a URL has not been stored in a specific browser folder, the bookmark receives a single tag when imported to *BobrDobr.ru*, indicating that the URL belongs to "no category". If a URL has been stored in a specific browser folder, the directory and sub-directory names are assigned as tags. Since Veres [25] and Golder and Huberman [13] showed that hierarchical and taxonomical descriptions for URLs – equivalent to those that result from the automated tag creation taking the directory names of browser bookmarks as a basis – correspond to common tag choices, and as the authors found no indicators for assuming that these tags have a lower probability to be imitated, the authors decided to keep those bookmarks and tags in the dataset. Since URLs that were *solely* tagged automatically by the import mechanism of the system do not allow any statements about imitation of tags – that is, since no tags were created manually for these users by any user inside the system – the authors removed all URLs from their further analyses that solely received imported bookmarks. Furthermore, all "no category" tags were removed, because they were never assigned by hand – and, accordingly, never imitated – in the dataset and only appeared in imported bookmarks. In total, the authors removed 32,739 URLs; the resulting dataset contained 78,001 URLs and 299,786 tags.

In the next step, the authors selected all URLs that had received both types of bookmarks, "internal" and "external" ones. Some 661 URLs (.85%) met this requirement. The majority of URLs (77,282, or 99.08%) had received only "external" bookmarks. A small number of 58 URLs (.07%) had received no other bookmarks except for "internal" ones. Bookmarks that were created only internally or only externally were excluded from further analysis, because they would have distorted the results due to the fact that they are not two randomly assembled groups but rather had the risk of a self-selection bias by containing specific types of URLs linked to specific tagging behaviour. The composition of these groups and the resulting differences in tag sets would have always mixed with the effects of the bookmarking method. The authors, therefore, confined themselves to analyzing the 661 URLs (and their associated 5,799 unique tags) that provided clear insights into these effects.[7]

Table 1. Distribution of bookmarks and tags within the URLs[8]

	Internal bookmarks (copy method)		External bookmarks (browser button method)	
No. *m* of received bookmarks for a URL	No. of URLs with *m* bookmarks	No. of tags that had been assigned to these bookmarks (through copy method)	No. of URLs with *m* bookmarks	No. of tags that had been assigned to these bookmarks (through browser button method)
1	492	1320	485	2009
2-10	75	367	146	1203
11-20	46	405	3	56
21-30	48	505	3	102
31-40	0	0	2	115
41-50	0	0	19	979
51-60	0	0	3	184
Total	661	2597	661	4648

Table 1 depicts the distribution of bookmarks and tags for these 661 URLs. Each URL is represented in both columns, in the internal bookmarks as well as in the external bookmarks, as the URLs had received both types of bookmarks. However, a URL could have received 20 internal bookmarks, for example, but 30 external bookmarks and was therefore to be found in a different row for each column. A high proportion of URLs had only one or very few bookmarks. This is typical of the power law distribution of bookmarks over URLs [e.g. 13]. The table also shows that many more external bookmarks than internal ones were created.

[7] Although the number of 661 URLs might seem low in comparison to the total of 110,740 URLs in the original dataset, the sample size is still sufficiently large enough to draw meaningful statistical inferences about imitation behaviour and, hence, do not limit the analysis. Rather, this was a necessary step to ensure high data quality.

[8] Table 1 refers to the remaining dataset with both types of bookmarks (after eliminating private bookmarks and bookmarks for URLs that had received only imported bookmarks), examined in separate groups. One URL with 32 bookmarks was moved to the group with 21-30 bookmarks, as there was only one with more than 30 internal bookmarks.

4.2 Methodology and Operationalization

The proposed hypotheses were tested using bivariate correlation analysis and the non-parametric Wilcoxon signed rank test. The variables used for these analyses were operationalized as follows.

For the test of ***H1***, the authors had to correlate the temporal order of tag assignment and the tag assignment frequencies of a URL.[9]

Concerning the operationalization of *overall tag frequency*, the authors had to consider that the absolute tag frequencies of a certain tag could not be compared between different URLs, because each had been assigned a different number of tags and a different number of unique tags. Ignoring the different group means could have led to false conclusions regarding the overall relationship, on the aggregated level, of all URLs [26]. Thus, the absolute "internal" and "external" tag frequencies were rescaled as follows (for a detailed rationale for rescaling, such as "avoiding regression to the mean", see, e.g., [27]). First, the frequencies were transformed into ranks. The tag most often assigned to a URL received rank "1". When tags had been assigned in equal numbers, the resulting rank numbers were averaged for tied observations. Afterwards, percentile ranks were calculated using the formula $(r-1/2)/w$, where w is the number of observed tags for the underlying URL, and r denominates the rank of the value, with values from 1 to w. Finally, to obtain normal scores, the z-scores of the percentile ranks were computed, which result from standardizing the percentile ranks with respect to the mean and standard deviation of all the tags assigned to the underlying URL.

The *point in time* in which a tag t has been assigned to a specific URL u for the first time was measured as date in minutes. These timestamps were rank-transformed and standardized just as described for the frequencies, with the first assigned tag getting rank "1". This way, both of the rank orders were relative to the remaining ranks within the same URL and were then analyzed as two variables of a bivariate distribution.[10]

---

[9] The bivariate distribution of the variables showed an unknown, non-linear relationship that differed between URLs. Other authors who have discussed social bookmarking with a focus on statistics [e.g., 28] have proposed transforming tag variables using the natural logarithm; this recommendation is based on many observations that tag distributions typically follow a power law / scale free distribution. The result would be linear relationships between the transformed variables and the applicability of linear models. For most of the URLs in the dataset analyzed, though, a logarithmic transformation was not viable, because no consistent scale free distribution could be observed for the URLs. This problem was then bypassed through the eventually used non-parametric transformation of the variables.

[10] Due to space limitations, the univariate distributions of the internal tag frequencies are not discussed here in detail. The univariate distribution (also for various groups as highlighted in the next section) showed a high concentration around the normal rank "0", similar to the distribution of the temporal ranks. This is plausible when considering that a large number of URLs had received only one bookmark with n tags, so that each of the tags had a total internal tag frequency of "1" and a normal rank of tag frequency (internal method) of "0"

For the test of ***H2***, a variable was needed that would represent the difference of the internal and external assignment frequency of each tag for each URL. This variable was calculated by subtracting the absolute external tag frequency from the internal tag frequency. The differences were then rank transformed. High ranks were given to tags that were assigned internally much more often in absolute numbers than externally.

For the tests of ***H3*** and ***H4***, the variables t_{int} and t_{ext} indicate the average amount of tags assigned to a specific URL per bookmark, and dt_{int} and dt_{ext} indicate the average amount of unique tags for that same URL per bookmark; the index always shows whether tags assigned internally or externally are considered.

5 Results

Hypothesis ***H1*** postulates a positive correlation between the rank of temporal order in which a tag *t* has been assigned to a specific URL *u* for the first time and how many times *t* has been assigned to *u* overall in respect to all other tags of that URL.

For the hypothesis test, the authors first analyzed the "internal bookmarks". In a first step, the dataset was split into the four groups highlighted in Table 1 (i.e., one group for URLs with one bookmark, one group for URLs with 2-10 bookmarks, and so on). This decision was taken because the proportions of the tags frequencies of a URL shift as the URL gets more and more bookmarks (compare [13]). For example, the postulated effect of H1 cannot occur at all if a URL had received only a single bookmark. Therefore, the pooled analysis of a URL with one or two bookmarks with a URL with 40 bookmarks could not give meaningful insights, because the emerging descriptions for each URL were in different "stages of development". To reduce the number of groups, URLs with similar amounts of bookmarks were combined, as shown in Table 1.

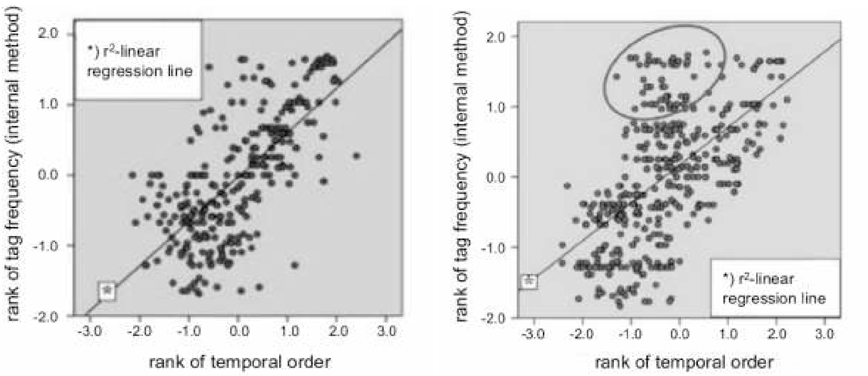

Fig. 2. Bivariate distribution "normalized rank of temporal order – normalized rank of tag frequency" for groups of internal tags. From left to right: Group with 11-20 and group with 21-30 bookmarks.

In line with the authors' expectations, for the group with one bookmark only, there was no statistically significant correlation between the normalized rank of tag frequency and the normalized rank of temporal order ($r(1320)=.02$; $p=.50$). A broader spectrum of tag frequencies was observed for the group with 2-10 bookmarks, showing, as expected a slightly positive correlation between temporal rank order and tag frequency ($r(367)=.24$; $p=.00$), but also a broad dispersion with no obvious linearity. For the group with 11-20 bookmarks, the resulting correlation came closer to the expected linear relationship (see the left graph of Figure 2) with a relatively high coefficient ($r(405)=.71$; $p=.00$). This group also contained a number of outliers, tags that had been assigned seldom but had been assigned relatively early to the observed URL. The results of the group with 21-30 bookmarks were similar ($r(505)=.59$; $p=.00$), only that outliers occurred more often, and in another form as well: there were some tags that were assigned relatively late but were still quite common; the oval in the second graph of Figure 2 highlights them. In summary, ***H1*** is supported for the internal bookmarks.

In a second step, the authors tested ***H1***, analyzing the external bookmarks for the groups from Table 1 (see Figure 3).

As for the internal bookmarks, the result for group 1 (URLs with only one bookmark) of the external bookmarks did not indicate a correlation between the rank of temporal order in which a tag t has been assigned to a specific URL u for the first time and how many times t has been assigned to u overall ($r(2009)= -.00$; $p=.84$). The distribution of the group with 2-10 bookmarks showed no linear relationship (see Figure 3) although Pearson's r was significant ($r(1203)=.32$; p=.00). Many tags received high frequency ranks (i.e., they were rarely assigned), independently from the point in time they had been assigned. A considerable amount of tags had only been assigned once (marked by the dotted-line brackets in the figures). They may be very individual tags, so specific that no other user would associate them with the underlying URL, independently from the point in time they were assigned. We call these values "*tf1*-values" for "tag frequency 1". When excluding these values from the analysis, the correlation coefficient rose to $r(403)=.39$, $p=.00$, which could be an indicator for the appropriateness of the authors' interpretation. For the group with 11-20 bookmarks, the correlation was positive ($r(56)=.29$; $p=.03$) and the exclusion of the *tf1*-values did not lead to a higher coefficient. In group 21-30, *tf1*-values were highly represented and their exclusion, therefore, had a high impact: *before* their exclusion, the correlation coefficient was $r(102)=.29, p=.00$; *after* their exclusion, the correlation coefficient was $r(33)=.52$, $p=.00$. For all the remaining groups, the analysis was run twice as for the other groups. Table 2 presents the results of these analyses.

Table 2. Correlation between "rank of temporal order" and "frequency"

Group	Results before exclusion of *tf1*	Results after exclusion of *tf1*
31-40	$r(115)=.240, p=.01$	$r(35)=.59, p=.00$
41-50	$r(979)=.46, p=.00$	$r(372)=.61, p=.00$
51-60	$r(184)=.35, p=.00$	$r(57)=.82, p=.00$

In summary, ***H1*** is also supported for the external bookmarks.

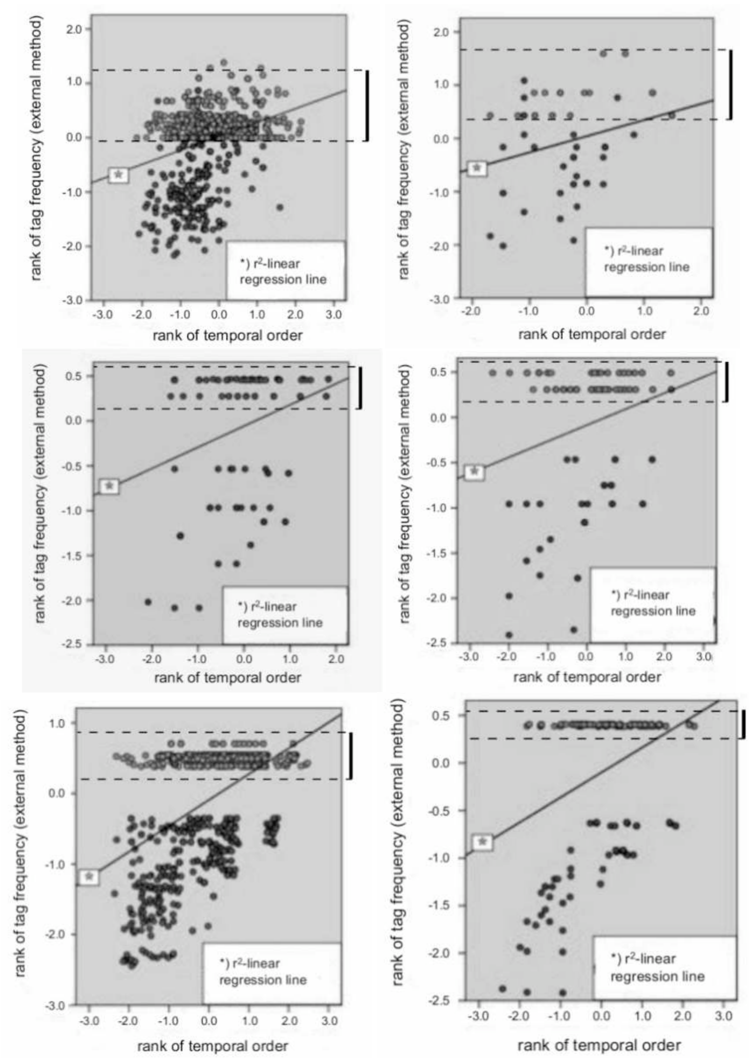

Fig. 3. Bivariate distribution "rank of temporal order – rank of tag frequency (external method)" for the groups of external tags. From left to right, from above to below: Group with a) 2-10, b) 11-20, c) 21-30, d) 31-40, e) 41-50, f) 51-60 bookmarks.

The authors now turn to hypothesis ***H2***, which assumes the correlation postulated in ***H1*** to be stronger for the bookmarking method allowing for imitation. To test this hypothesis, the authors calculated the correlation between the rank of temporal tag order and the rank of differences between absolute tag frequencies of external and internal bookmarks.

Figure 4 does not show any linear relationship; instead, it shows a broad, regular distribution. Also, the exclusion of the *tf1*-values leads to a slightly higher negative correlation coefficient (r(687)=-.09; *p*=.02), but changes in the distribution or a particular pattern could not be observed. Hence, the authors cannot conclude whether the results argue for or against the presence of imitation and hence cannot decide whether ***H2*** is supported.

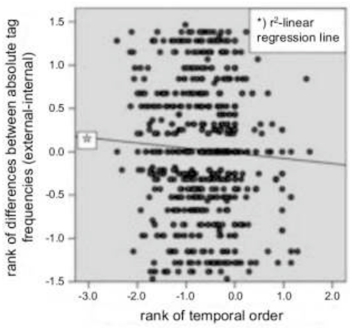

Fig. 4. Bivariate distribution "rank of temporal order – rank of differences between absolute tag frequencies (external-internal)"

H3 postulates that the average amount of tags a user assigns to a URL increases when the system allows that user to see other users' tags, while ***H4*** postulates that the average amount of *different* tags assigned to a URL by all users decreases.

A Wilcoxon signed rank test was conducted to compare the 661 URLs that received both the internal and external types of bookmarks. For this test, the authors calculated the differences between the values "number of tags" and "number of different tags" resulting from the internal and the external method for each URL. The Wilcoxon signed rank test showed that fewer tags were assigned through the internal than through the external bookmarking method (*z*=-7.80; *p*=.00), that is, more URLs received more tags from the external method than the other way around. Hence, when imitation was possible, users did not assign more tags to a URL, even if they may have had more choices. ***H3***, therefore, is not supported. In contrast, the test results of the second Wilcoxon signed rank test (*z*=-13.54; *p*=.00) point favourably to hypothesis ***H4***, which assumes that the variation of assigned tags decreases when imitation is possible. More URLs received less diversified tags from the internal method than the other way around. ***H4***, therefore, is supported. Of course, the lower tag variation for the internal method is due in large part to the smaller number of tags assigned. Note, however, that the negative value of *z* is considerably larger for the different tags assigned, which speaks to a second cause, namely imitation.

6 Summary and Conclusions

The aim of this study was to analyze the impact of a feature allowing for imitation provided by a social bookmarking system on its resulting metadata. The authors were able to show that tags that were assigned early were assigned more often, coming closer to a linear relationship the more bookmarks were assigned to a URL. This was true for internally as well as for externally assigned tags (***H1***). The authors were not able to show a clear increase in the correlation of the time of assignment and the frequency of a tag when the imitation feature was present, meaning that popular tags get more popular when imitation is possible (***H2***). Nonetheless, comparing the bivariate distributions in the internal and external case, it became clear that with the internal method tags that were assigned only once were largely lacking. This is an indicator that users neglect some of their personal, very individual tags when tag recommendations are available.

The authors' assumption that users assign more tags when provided with other users' tag choices clearly had to be rejected (***H3***), since the results indicated the opposite. One reason for this may be that users see little or no need to think about tags of their own when presented with a choice of popular tags from which to choose conveniently. This would be in line with the principle of least cognitive effort described by Munk et al. [4]. Consequently, users adopt only the suggested set of tags. This explanation fits with the findings that, when imitation is possible, the number of different tags assigned to a URL decreases partly due to the drop in the number of tags assigned overall (***H4***). The other element of the decrease of different tags could then be explained by users constraining themselves to the sole use of suggested tags.

Overall, these findings point to strong effects of imitation on the emerging folksonomies, leading to a less pluralistic collective description of content. This might be interpreted as an overrepresentation of early tags leading to the elimination of individual information and thus a decreasing quality of the folksonomy. But, as this effect could not be shown unequivocally, the less different keywords can also mean a unification of synonyms or different spelling. In such a case, users who used different tags for the description of exactly the same concept would align their tags with the "standard" for that concept set in the most popular tags. This could be seen as a positive effect of imitation.

Still, these findings have important implications for researchers and practitioners who analyze or intend to use folksonomies for content classification. The feature of the system analyzed in this study, for example, aimed only at increasing convenience for users during the tagging process. However, this feature also had a large impact on the imitation of tags, and hence also on the variety of tags and the resulting folksonomy as a whole. Therefore, the authors recommend that system designers always consider how features they implement may affect the quality of the resulting folksonomies, particularly when they would like to utilize full cognitive input of all users.

The simplest solution to ensure a higher quality of collective tagging would be the general omission of tag suggestions based on aggregated taggings. This, however, is not viable in systems that face tough competition and that rely on simple, effortless usage to attract users. Another option would be a “type-ahead” feature that requires users to begin typing two to three letters and then offers suggestions for completion from the pool of most popular tags. In this way, users would have to tap into their own cognitive models and would not be guided in one direction or the other before they think themselves about possible tags. To benefit from positive imitation effects, the suggestion for a tag should be the most popular synonym and spelling form. More elaborate mechanisms produce tag suggestions based on the content of bookmarked websites or files; others draw on the tri-partite links between resources, tags, and users (i.e, graph-based tag suggestion). These solutions rely on the most popular tags only to a small part or not at all and are therefore not likely to produce information cascades.

In addition to this managerial insight, this study provides some insights for future scientific work on the subject. The authors are not aware of any study that tests statistically any formal hypotheses of imitation of tags in social bookmarking systems with empirical field data.

As with any empiricial study, this work is subject to limitations. The authors do not consider these limitations to void any results so long as the reader remains aware of them when interpreting the results. In fact, they suggest either some future research that examines collective intelligence in social bookmarking systems or provide additional insights about user behavior in social bookmarking systems.

First, indicators speak to the presence of spam in the raw data (cf. also [13]); for example, a large number of tags were never copied with respect to one URL but assigned to a great number of different URLs. No method to avoid spam completely in social bookmarking systems exists, so only an experimental design or a system that can guarantee a human-only userbase could provide a folksonomy created completely absent the influence of spam.

Second, the small number of internally created bookmarks and the resulting small number of URLs with both types of bookmarks (661 URLs; 0.85%) was surprising. The majority of URLs (77,282 URLs, i.e. 99.08%) had received only “external” bookmarks. The low affinity to the copying of other users’ bookmarks could be explained by the users’ preference to inspect website content themselves before tagging it, and therefore having a preference for the external bookmarking method. It seems that far fewer users than expected browse the system to "stumble upon" bookmarks that might interest them, but use the external bookmarking button only to save addresses they found for themselves on the web.

Third, considering these findings, it is also legitimate to raise objections to the implicit assumption that the two bookmarking methods do not just differ only with respect to suggesting or not suggesting tags. For example, one could argue that some people browse for bookmarks internally and then click on a link to inspect the website, after which they bookmark the website using the external method. Still, in this case, the user would not see the most popular tags either, as these are only displayed when bookmarking internally. One might also argue that these users would remember those tags they saw beforehand and assign them as well when bookmarking URLs with the external method. Both threats to reliability would lead only to a higher than postulated similarity rate between externally and internally assigned tags. This would reduce the power of the tests employed, and hence it would be more difficult to find support for the hypothesized effects of imitation. However, if significant effects of imitation are found (as in this study), the reader can have confidence in these findings, because the reliability threat does not threaten validity when a difference is found.[11]

Fourth, it should be mentioned that the cultural background of the userbase of *BobrDobr.ru* could have influenced the results. It cannot be precluded that a different affinity for imitation is present in different cultural contexts.[12] Still, as none of the literature on information cascades or rational imitation reviewed by the authors offers any hints as to cultural influences on the postulated effects, we assume that the basic mechanism of imitative behavior is universal in such systems even if the extent of imitation can be different.

Fifth, one might argue that users are heterogeneous regarding their preferred bookmarking method. In this case, imitation behavior linked to the internal bookmarking method could be an effect of the user type preferring this method and their affinity to imitation rather than an effect of the bookmarking method itself. Future research should analyze whether preferences for a certain bookmarking method are linked to a certain type of user. The best way to factor out such effects unambiguously would be an experimental design with randomly assembled groups of users, one bookmarking with and one without suggestions. The authors suppose that such a lab experiment would provide additional confidence in the findings from this field experiment.

The authors hope that this research will assist other researchers in conducting these types of studies and form the basis for substantial future research into imitation of tags and information cascades in social bookmarking services. Further, the authors hope that this research provides useful insights for managers, librarians, and other practitioners who use folksonomies for content classification and who must design social bookmarking systems that will lead to high-quality folksonomies.

References

1. Marlow, C., Naaman, M., Boyd, D., Davis, M.: HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead. In: 17th conference on Hypertext and Hypermedia, Odense, Denmark, pp. 31--40. Association for Computing Machinery, New York, NY (2006)
2. Vander Wal, T.: Folksonomy Definition and Wikipedia. Blog entry, November 2005, available at http://www.vanderwal.net/random/entrysel.php?blog=1750, accessed on 2009-09-09 (2005)
3. Vander Wal, T.: Folksonomy: Coinage and Definition. Blog entry, February 2007, available at http://vanderwal.net/folksonomy.html, accessed on 2009-09-09 (2007)
4. Munk, T.B., Mørk, K.: Folksonomy, the Power Law & the Significance of the Least Effort. Knowledge Organization, 34(1), pp. 16--33 (2007)

[11] The real impact of imitation should be rather stronger.

[12] One could argue, for example, that the states of the former U.S.S.R are more prone to collectivism, which, in turn, favors imitative behavior [compare e.g. 29].

5. Al-Khalifa, H. S., Davis, H. C.: Replacing the Monolithic LOM: A Folksonomic Approach. In: 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007), Niigata, Japan (2007)
6. Tsui, E., Wang, W. M., Cheung, C. F., Lau, A. S. M.: A Concept-Relationship Acquisition and Inference Approach for Hierarchical Taxonomy Construction From Tags. Information Processing & Management, 46(1), in print (2010)
7. Van Damme, C., Coenen, T., Vandijck, E.: Turning a Corporate Folksonomy into a Lightweight Corporate Ontology. In: Abramowicz, F. (ed.) BIS 2008, Lecture Notes in Business Information Processing, vol. 7, pp. 36--47 (2008)
8. Farooq, U., Song, Y., Carroll, J. M., Giles, C. L.: Social Bookmarking for Scholarly Digital Libraries. IEEE Internet Computing, 11(6), pp. 29--35 (2007)
9. Li, X. R., Snoek, C. G. M., Worring, M.: Learning Social Tag Relevance by Neighbor Voting. IEEE Transactions on Multimedia, 11(7), pp.1310—130 (2009)
10. Goshal, G., Zlatic, V., Caldarelli, G., Newman, M. E. J.: Random hypergraphs and their applications. Physical Review E, 79(6), in print (2009)
11. Zhang, Z. K., Zhou, T., Zhang, Y. C.: Personalized Recommendation via Integrated Diffusion on User-Item-Tag Tripartite Graphs. Physica A - Statistical Mechanics and Its Applications, 389(1), pp. 179--186 (2010)
12. Boast, R., Bravo, M., Srinivasan, R.: Return to Babel: Emergent Diversity, Digital Resources, and Local Knowledge. Information Society, 23(5), pp. 395--403 (2007)
13. Golder, S.A., Huberman, B.A.: Usage Patterns of Collaborative Tagging Systems. Journal of Information Science, 32(2), pp. 198--208 (2006)
14. Heymann, P., Koutrika, G., Garcia-Molina, H.: Can Social Bookmarking Improve Web Search? In: International Conference on Web Search and Web Data Mining (WSDM 2008), Palo Alto, California, USA, pp. 195--206 (2008)
15. Bao S., Wu, X., Fei, B., Xue, G., Su, Z., Yu, Y.: Optimizing Web Search Using Social Annotations. In: 16th International Conference on World Wide Web (WWW 2007), May 8-12, Banff, Alberta, Canada (2007)
16. Macgregor, G., McCulloch, E.: Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. Library Review, 55(5), pp. 291--300 (2006)
17. Specia, L., Motta, E.: Integrating Folksonomies with the Semantic Web. In: 4th European Semantic Web Conference (ESWC 2007), June 3-7, Innsbruck, Austria, pp. 624--639 (2007)
18. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information Retrieval in Folksonomies: Search and Ranking. In: 3rd European Semantic Web Conference (ESWC 2006), June 11-14, Budva, Montenegro, pp. 411--426 (2006)
19. Arakji, R., Benbunan-Fich, R., Koufaris, M.: Exploring Contributions of Public Resources in Social Bookmarking Systems. Decision Support Systems, 47(3), June, pp. 245--253 (2009)
20. Rosch, E., Mervis, C.B., Gray, W., Johnson, D. and Boyes-Braem, P.: Basic Objects in Natural Categories. Cognitive Psychology, 8 (3), pp. 382--439 (1976)
21. Tanaka, J.W., Taylor, M.: Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder? Cognitive Psychology, 23(3), pp. 457--482 (1991)
22. Cialdini, R.B.: Influence: Science and Practice. Harper Collins College Publishers, New York, USA (1993)
23. Hedström, P.: Rational Imitation. In: Social Mechanisms: An Analytic Approach to Social Theory, Cambridge University Press, Cambridge, UK, pp. 306--327 (1998)
24. Bikhchandani, S., Hirshleifer, D.,Welch, I.: A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. The Journal of Political Economy, 100(5), October, pp. 992--1026 (1992)

25. Veres, C.: The Language of Folksonomies: What Tags Reveal About User Classification. In: Kop, C., Fliedl, G., Mayr, H.C., Métais, E. (eds.) NLDB 2006, LNCS, vol. 3999, pp. 58--69, Springer (2006)
26. Boyd, L.H., Iversen, G.R.: Contextual Analysis: Concepts and Statistical Techniques. Wadsworth, Belmont, California, USA (1979)
27. Jones, Q., Ravid, G., Rafaeli, S.: Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration. Information Systems Research, 15(2), pp. 194--210 (2004)
28. Raban, D.R., Rabin, E.: Statistical Inference from Power Law Distributed Web-Based Social Interactions. Internet Research, 19(3), pp. 266--278 (2009)
29. Marron, D.B., Steel, D.G.: Which Countries Protect Intellectual Property? The Case of Software Piracy. Economic Inquiry, 38(2), pp. 159-174 (2000)