Analysis of music tagging and listening patterns: Do tags really function as retrieval aids?

Jared Lorince[1], Kenneth Joseph[2], and Peter M. Todd[1]

[1] Cognitive Science Program
Indiana University, Bloomington, Indiana, USA
{jlorince,pmtodd}@indiana.edu
[2] Computation, Organization, and Society Program
Carnegie Mellon University, Pittsburgh, PA, USA
kjoseph@cs.cmu.edu

Abstract. In collaborative tagging systems, it is generally assumed that users assign tags to facilitate retrieval of content at a later time. There is, however, little behavioral evidence that tags actually serve this purpose. Using a large-scale dataset from the social music website Last.fm, we explore here how patterns of music tagging and subsequent listening interact in an effort to determine if there exist measurable signals of tags functioning as retrieval aids. Specifically, we describe our methods for testing if the assignment of a tag tends to lead to an increase in listening behavior. Results suggest that there exists a small but reliable effect of tags increasing listening levels, and also reveal interesting differences in which kinds of tags are most assoicated with future listening.

Keywords: Collaborative tagging, Folksonomy, Music listening, Memory cues, Retrieval aids, Personal information management

1 Introduction

In social tagging systems, users assign freeform textual labels to digital content (music, photos, web bookmarks, etc.). These individual tagging decisions are aggregated into a folksonomy [19], a "bottom-up" classificatory structure developed with little or no top-down guidance or constraints. There are a variety of reasons for which users tag content, but it is overwhelmingly assumed that tagging for future retrieval – assigning a tag to an item to facilitate re-finding it at a later time – is users' principal motivator. But is this a valid assumption?

Collaborative tagging systems are often designed, at least in part, as resource management platforms that expressly facilitate the use of tags as retrieval aids. However, the freeform, and often social, nature of tagging opens up many other possible reasons for which a user might tag a resource. While there is a significant amount of non-controversial evidence for such alternative tagging motivations, (sharing resources with other users, social opinion expression, etc.), the problem with the retrieval aid assumption runs deeper than there simply existing possible alternatives. There is, in fact, almost no behavioral evidence that tags are ever actually used as retrieval aids. While there is much data available on user tagging habits (i.e. which terms are applied to which resources, and when), to our knowledge there is no published research providing behavioral evidence of whether or not tags, once applied to items, actually facilitate subsequent retrieval. This is an issue largely driven by a lack of data: while a web service can in principle track a users' interaction with tags (for instance, if users use tags as search terms to find tagged content), there are no available datasets containing such information, nor can it be crawled externally by researchers.

Despite these issues, this empirical question is not intractable. While detailed information on how existing tags themselves are utilized remains beyond our reach, an alternative approach is to examine how patterns of user interaction with tagged versus untagged content vary. In other words, if tags do serve as retrieval aids, we should expect users to be more likely to interact with resources (e.g. visit bookmarked pages, listen to songs, view photos, etc.) upon the application of a tag.

In the current paper we test this hypothesis using a large-scale dataset consisting of complete listening and tagging histories from more than 100,000 users from the social music website Last.fm. From this dataset, we extract user-artist listening time series, each of which represents the frequency of listening over 90 months of a particular artist by a particular user, and compare time series in which the user has tagged the artist to those that are untagged. Specifically, we address the following two questions:

- **RQ1**: Does comparison of tagged versus untagged time series provide evidence that tagging an artist increases the probability of listening to that artist in the future?
- **RQ2**: Do certain tags prove to be particularly associated with increases in future listening, and if so, can we identify attributes of such "retrieval-targeted" tags as opposed to others?

We describe the various analytic methods we bring to bear on these questions in Section 4, but first present related work (Section 2) and details of our dataset (Section 3). We close in Section 5 with synthesis and interpretation of our results, as well as a plan for future work.

2 Background

2.1 The formal study of folksonomies

Collaborative tagging has been considered one of the core technologies of "Web 2.0", and has been implemented for resources as diverse as web bookmarks (Delicious), photos (Flickr), books (LibraryThing), academic Papers (Mendeley), and more. Thomas Vander Wal [19] first coined the term "folksonomy" to describe the emergent semantic structure defined by the aggregation of many individual users' tagging decisions in such a system. These folksonomies have since become the target of much academic research. One of the earliest well-known and involved analyses of a collaborative tagging system is Golder and Huberman's [4] work on the evolution of tagging on Delicious.com. Around the same time, Hotho and colleagues [9] presented a formal definition of a folksonomy: $\mathbb{F} := (U, T, R, Y)$ [1]. The variables U, T, and R represent, respectively, the sets of users, tags, and resources in a tagging system, while Y is a ternary relation between them ($Y \subseteq U \times T \times R$). The "personomy" of a particular user (i.e. the set of annotations generated by an individual), $\mathbb{P} := (T_u, R_u, Y_u)$, can be similarly defined.

[1] This is a slight simplification. For details, see [9]

Since 2006, an extensive literature on *how* people tag has also developed, covering topics like tagging expertise [22], mathematical [2] and multi-agent [12] models of tagging choices, consensus in collaborative tagging [6, 16], and much more. Our understanding of the dynamics of tagging behavior has greatly expanded, but understanding exactly *why* people tag, on the other hand, has proven more elusive.

2.2 Why do people tag?

It is typically assumed that tags serve as retrieval aids, allowing users to re-find content to which they have applied a given tag (e.g. a user could click on or search for the tag "rock" to retrieve the songs she has previously tagged with that term). This assumption is baked into Vander Wal's original defintion of a folksonomy, which he contends"is the result of personal free tagging of information and objects (anything with a URL) *for one's own retrieval*" [19, emphasis added]. This perspective is echoed in many studies of tagging patterns [3, 6, 4].

But while retrieval is the most commonly assumed motivation for tagging, other reasons certainly exist, and various researchers have proposed taxonomies of tagging motivation. Proposed motivational factors in tagging include personal information management (including but not limited to tagging for future retrieval), resource sharing, opinion expression, performance, and activism [8, 23, 1], among others. See [5] for a review.

While the development of motivational theories in tagging is useful, there is almost no work actually grounding them in behavioral observations. The vast majority of existing work either makes inferences about motivation based on design features of a website (e.g. social motivations in tagging require that one's tags be visible to other users, [14]), employs semantic analysis and categorization of tags (e.g. the tags "to read", "classical", and "love" can all be inferred to have different uses, [23, 17]), or directly asks users why they tag using survey methods [1, 15]. The results of such approaches are useful contributions to the field, but none have resulted in testable behavioral hypotheses that can confirm or refute their validity.

is this what you mean to say? Or do you mean none have them have tested the hypotheses they suggest? I can understand it either way but want to be sure

One notable exception is work by Körner and colleagues [10, 11, 24]. They argue that taggers can be classified on a motivational spectrum from categorizers (who use a constrained vocabulary suitable to future browsing of tagged resources) to describers (who use a large, varied vocabulary to facilitate future keyword-based search), and have developed and tested quantifiable signals of these different motivations. The main deficiency of this approach, however, is that their hypotheses are based fully on attributes of user tag vocabularies; they present no way to test whether or not describers actually use tags, once applied, for keyword-based search and that categorizers use them for browsing.

This problem is no fault of the authors, however. Data on how users actually *use* existing tags is simply not available to researchers through any tagging system APIs (or through other methods) that we are aware of. Thus the existing work on tagging motivation is limited to inferring *why* people tag from *how* they tag. In presenting our novel methods, we are aware that they still represent an inferential approach. Our approach is distinct from those described here, however, in that we test a concrete hypothesis about how tagging should affect a behavior on which we *do* have data (interaction with tagged content, in our case music listening).

3 Dataset

Last.fm incorporates two specific features that are presently of interest. First, it implements a collaborative tagging system (a "broad" folksonomy, following Vander Wal's [18] terminology, meaning that multiple users tag the same, publicly available content) in which users can label artists, albums, and songs. Second, the service tracks users' listening habits both on the website itself and on media players (e.g. iTunes) via a software plugin. This tracking process is known as "scrobbling", and each timestamped instance of a user listening to a particular song is termed a "scrobble".

Here we utilize an expanded version of a dataset described in earlier work [12, 13] that includes the full tagging histories of approximately 1.9 million Last.fm users, and full listening histories from a subset of those users (approximately 100,000) for a 90-month time window (July 2005 - December 2012, inclusive). Data were collected via a combination of the Last.fm API and direct scraping of publicly available user profile pages. For further details of the crawling process, see [12, 13].

If this is a blind review, you should probably remove these references

For our current purposes, we consider only those users for which we have both tagging and listening histories. For each user, we extract one time series for each unique artist listened to by that user. Each user-artist listening time series consists of a given users' monthly listening frequency to a particular artist for each month in our data collection period, represented as a 90-element vector.

We selected a monthly timescale for listening behavior due to the fact that user tagging histories are only available at monthly time resolution. Furthermore, we perform all analyses here at the level of artists, rather than individual songs. Thus every song scrobbled is treated as a listen to the corresponding artist, and all annotations (which can be applied to songs, albums, or artists) are treated as annotations of the corresponding artist. Our choice to perform all analyses at the level of artists, rather than individual songs, is based on the facts that (a) listening and tagging data for any particular song tend to be very sparse, and (b) the number of time series resulting from considering each unique song listened to by each user would be prohibitively large.

The 2 billion plus individual scrobbles in our dataset generate a total of $\sim$ 95 million user-artist listening time series. In $\sim$ 6 million of these cases, the user has assigned at least one tag to the artist (or to a song or album by that artist) within the collection period (we refer to these as tagged time series), while in the remaining cases ($\sim$ 89 million) the user has never tagged the artist. We summarize these high level dataset statistics in Table 1. Comparison of these tagged and untagged listening time series is the heart of the analyses presented in the next section.

4 Analyses & Results

4.1 RQ1: Comparison of tagged and untagged time series

Our principal research question is whether listening patterns for tagged content are consistent with the expectation that tags serve as memory cues. If this were to be the case, we would expect to see increased listening rates for musical artists once a tag is applied, under the assumption that a tag facilitates retrieval and increases the chances of a user listening to a tagged artist.

Total users	104,829
Total scrobbles	2,089,473,214
Unique artists listened	4,444,119
Unique artists tagged	1,049,263
Total user-artist listening time series	94,875,106
Total tagged time series	5,930,594
Total untagged time series	88,944,512

Table 1: Dataset summary

Unfortunately, several factors combine to make such an analysis difficult. First and foremost, the desired counterfactual of the untagged "version" of a particular tagged series, which would allow a direct testing of how tagging changes listening behavior, does not, of course, exist. We thus must utilize untagged time series in a way that allows them to approximate what a true counterfactual might look like. In searching for such samples, a second difficulty that arises is that listening rates for tagged time series are much greater than for untagged time series (the average number of total listens across time series is 16.9 when untagged and 98.9 when tagged). While suggestive of the importance of tagging, this unbalance also suggests that controls must be instilled in both sample selection and statistical analysis to account for previous listening behavior prior to tagging. Finally, the actual point in time at which tags are expected to increase listening behavior for any given user is unknown, as it is theoretically possible that tagging may affect listening behavior as much three months after the tag has been placed as it does in the immediately following month. Thus, we must formulate our analysis in such a fashion as to account for this possibility.

To alleviate issues with the non-existence of a true counterfactual, we subselect from both the tagged and untagged series using the following formal procedure. We first temporally align the tagged and untagged time series. Both tagged time series are aligned so that they are centered on the month in which they were tagged. If multiple tags were present, we selected the tag within the month which had the most corresponding scrobbles. While there is no analogue to this point in the untagged data, we can partially resolve the issue by noting that tagging is disproportionately likely (approximately 30%, compared to 1.1% if the tagging month were random) to occur in a user's *peak*[2] listening month for a given artist. This provides a basis for aligning the tagged and untagged time series, by selecting only those tagged time series where the tag was applied in the month of peak listening, and then collecting a sample of untagged time series also aligned at the peak of listening. Where the peak was reached in multiple months, we chose one of these at random.

After aligning all tagged and untagged samples in this fashion, we further limited our analysis to a 13 month period extending from 6 months prior to the peak month to 6 months after the peak. This allows us to consider a variety of ways in which listening prior to the tag may affect future behavior while still admitting that there is likely a reasonable span in which tagging and past listening behavior have an effect on future actions. Finally, we further constrain our sampling to time series with:

[2] The month in which they listen the most times overall

- more than 25 total listens;
- a peak in listening at least 6 months from the edges of our data collection period (i.e. ensuring that the period from 6 months before to 6 months after the peak does not extend beyond the limits of our data range); and
- at least one listen 6 months prior to and after the peak (i.e. if the peak occurs in July, there should be at least one listen between January and June, and one between August and the following January).

Constraining our time series in this manner, we are left with a total of 206,140 tagged time series. We then randomly sampled from the 4.1M untagged time series an equal number meeting the same three criteria. All results below have been verified with multiple random samplings of the untagged data.

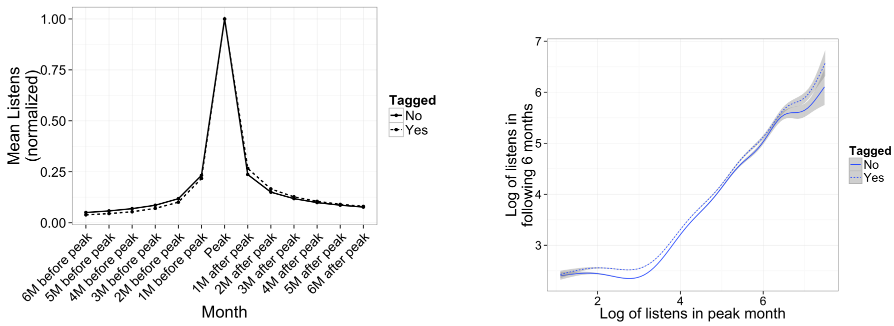
In Figure 1a we plot mean playcounts, with 95% normal confidence intervals, for each month across all tagged and untagged time series in the subsampled data. All values are normalized by the peak, and thus values at the peak month for both the tagged and untagged lines are unity. By visually comparing the line heights before and after the peak, Figure 1a shows that the mean normalized listening rate increases in the months after the peak for both tagged and untagged time series. However, we also see a small but reliable effect wherein tagged time series show proportionally higher mean normalized listening rates after the peak month (in which the the tag was applied) as compared to untagged time series. This is suggestive of an increase in listening as a result of tagging.

While Figure 1a thus gives evidence that supports our hypothesis, there are two important caveats to the data in the plot. First, because listening distributions are heavily skewed for any given month, the mean is not necessarily representative of the data. Though qualitative plots of transformed variables proved to be similar, further statistical analysis uses a log transformed version of the listening counts to account for these deviations. Second, while normalization controls for differences in listening counts between tagged and untagged data to some extent, we would prefer a method that explicitly accounts for listening behavior prior to and including the peak month on future behavior.

To more robustly test our hypothesis, we thus utilize a regression model relating post- and pre-peak listening behavior. Due to the volume of data we are dealing with, it was unreasonable make the assumption of linear dependence of the dependent variable on the independent variables. We therefore opted for a Generalized Additive Model (GAM, [7]), for which we utilized the R package mgcv [20]. Our dependent variable in the regression is the logarithm of the sum of all listens in the six months after a tag has been applied, as we are unsure at what point post-tagging the effect of a tag may be relevant.[3] Our independent variables are an indicator of whether or not the time series has been tagged, as well seven continuous-valued predictors, one each for the logarithm of the sum of listens in the peak month and the six previous months.

The regression model, which explained approximately 30% of the variance in the data (adj. R-sq.), indicated that smoothed (using thin-plate regression splines) parameters for all seven previous months had a signficant effect on post-peak listening behavior. As we cannot show the form of this effect for all model

[3] Qualitatively, our results hold when testing listening for each individual month as well. Also of note is our choice of using the log of the dependent variable rather than a count-based regression model (e.g. a Negative Binomial regression). The model used here appeared to fit the data better based on a variety of statistical and visual goodness-of-fit tests.



(a) Mean normalized playcount by month.

(b) Regression results, with 95% confidence interval.

Fig. 1: Comparison of tagged and untagged listening time series

variables, Figure 1b instead displays a similar model which considers only the effect of listening in the peak month on post-peak listening. As this plot suggests and the full model confirms, we can conclude that, controlling for all previous listening behavior, a tag increases the logarithm of post-peak listens by .147 [.144,.150]. This indicates that the effect of a tag is associated with around 1.15 more listens, on average, than if it were not to have been applied.

4.2 Tag analysis

To examine if and how different tags are associated with increased future listening, we ran a regression analysis similar to that described above. Four important changes were made. First, as would be expected, we considered only tagged time series. Second, instead of a single tagged/untagged indicator, we included binary (present / not present) regressors for all unique tags that had at least 25 occurrences in our subsample. Third, due to the data-hungry nature of the GAM and the large number of additional variables introduced by utilizing all tags as unique predictors, we chose to only control for listening in the peak month. This decision limited the computational difficulties associated with estimating a model of this size and did not appear to affect model fit substantially in subsamples of the data. Finally, we eliminated the constraint that a tag must occur in the peak month of a time series, as there is no meaningful comparison to be made with untagged data in this analysis. This allowed us to include additional tagged time series. If users had multiple tags for a particular artist, we again selected one tag randomly rather than include the same user artist combination twice in our analyiss.

This expanded sample consisted of XXX,XXX tagged time series, with 6,060 unique tags. After running the model, which explains ∼16.7% of the variance in the data (adj. R-sq.), 321 unique tags proved to be statistically significant predictors at $p < .001$. While we only have sufficient evidence to make claims about these 321 tags, qualitative examination of which tags are relatively strong predictors in the model proves informative.

The most telling observation is that commonly-used genre tags (e.g. "pop", "jazz", and "hip-hop") – which are the most common tags overall in our full dataset – tend to be weak, negative predictors of future listening. In contrast,

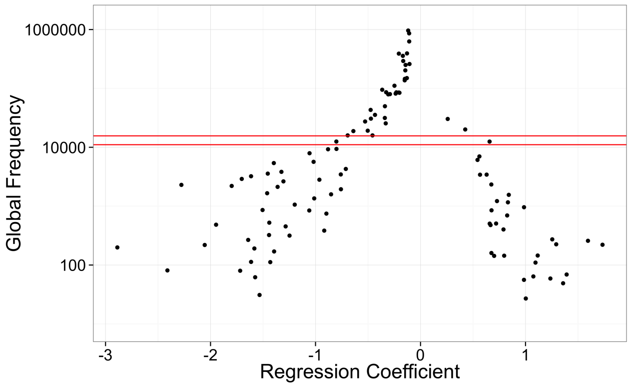


Fig. 2: Logarithm of tags' global popularity as a function of regression coefficient.

relatively strong predictors (both positive and negative) appear to be comparatively obscure, possibly idiosyncratic tags ("arguman-loved tracks", "mymusic", "leapsandbounds cdcollection")[4] . To examine this trend quantitatively, we plot in Figure 2 global tag popularity (i.e. the total number of uses of a tag in our full dataset, which consists of ∼ 50 million annotations) as a function of the tag's coefficient in the regression model for all tags that had a statistically significant coefficient. The red bands marked the upper and lower limits of a bootstrapped 95% confidence interval on the popularity of the 5,739 remaining tags that were *not* significant in the regression model. The result is a clear trend which suggests that the most popular tags are significant, weakly negative predictors of future listening, while both positive and negatively strong predictors tend to be relatively unpopular. Tags which were not significant in the model tend to be of moderate to high popularity.

footnote also needs to be removed for blind review

These data suggest that, at least for the small number of tags about which we can make statistically meaningful claims, those that are globally popular and well-known have relatively little effect on future listening, and are generally associated with small *decreases* in post-taggging listening rates. The tags that seem to "matter" (i.e. those that are relatively strong predictors of whether or not a user will listen to an artist after tagging it) are generally much less popular.

5 Conclusion

In this paper we set out to test the oft-cited assumption that tags serve as retrieval aids for indivduals in collaborative tagging systems. We did so via a novel methodology, testing for evidence that tagging an artist increases a user's future listening to that artist in comparison to a carefully selected set of untagged time series. Results suggest that tagging an artist does lead to an increase in listening, but that this increase is, on average, quite small (amounting to only 1 or 2 listens over a 6 month period). Given the various possible motivations for tagging,

[4] For a full listing of the regression coefficients across all tags in the model, see `https://dl.dropboxusercontent.com/u/625604/papers/lorince.joseph.todd.2015.sbp.supplemental/regression_coefficients.txt`

however, we expect only some tags to serve as retrieval cues, and thus tested the relative predictiveness of future listening for different tags. This analysis revealed systematic differences in how predictive the presence or absence of different tags was for future listening as a function of tag popularity. Specifically, we found that the most popular tags tend to have a small or non-significant effect on future listening, while less popular tags appear to be those that "matter", both as positive and negative predictors of future listening.

Based on such a small sample, we are at this point tentative to make strong claims about what specifically differntiates those unpopular tags that are strong negative versus strong positive predictors. The evidence is, nevertheless, suggestive of relatively uncommon (and likely, in many cases, to be idiosyncratic) tags being those most predictive of future listening behavior. This raises the intriguing possibility that the descriptive, popular tags that are arguably most useful to the community at large (i.e. genre labels and related tags), are not particularly associated with increases in listening, and thus are likely not functioning as memory cues.

This suggests that, while on average tagging an artist has a small positive effect on future listening, the most common tagging activities are *not* strong predictors of future retrieval. We cannot be sure which of the many other possible tagging motivations are at play here, nor can we know if and when a tag is applied with the intention of being used for retrieval, while ultimately not being used for this purpose. That said, these results do suggest that descriptive, relatively well-known genre classifiers do not show evidence of use as retrieval aids, but are nonetheless the most commonly applied tags. This may indicate that the primary motivation for tagging on Last.fm is not for personal information management (tagging a resource for one's own retrieval), but rather is socially-oriented, resulting in tags that are useful for the community at larger. This leads to the interesting possibility that a folksonomy can generate the useful, crowdsourced classification of content that proponents of collaborative tagging extol, but that this process is not strongly driven by the self-directed, retrieval-oritented tagging that is typically assumed in such systems.

While our results provide clues as to whether tags really function as retrieval aids, this remains early stage work addressing a hitherto unstudied research question. There is certainly room to refine and build upon the methods we present here for testing if and when tagging increases listening rates. In particular, our analysis at the level of artists (rather than the individual resources tagged) may be problematic, and we hope to eventually develop models that operate directly at the level of the content tagged (though data sparsity issues will make this a challenge). It will also be critical to expand on methods for understandings of which tags serve as memory cues and under what circumstances. It is clearly the case that not *all* tags function as memory cues (and indeed, some curiously have an almost opposite effect), so more robustly identifying which tags do serve as memory retreival items is fruitful direction for future work. Incorporating research on human memory from the cognitive sciences can also further inform hypotheses and analytic approaches to these questions, something we are actively pursuing in ongoing research. A final limitation is that we are exploring tagging in a particular collaborative tagging system, which operates in the possibly idiosyncratic domain of music. Tagging habits may vary systematically in different cotent domains, but until usable data becomes avaialble, we can only speculate as to exactly how.

In closing, to address the titular question of whether or not tags function as retrieval aids, the best answer would appear to be "sometimes". While there is much work to be done on when and why particular tags serve this function and others do not, it is clear that the overarching retrieval assumption is far from valid: Tags certainly do not always function as memory cues, and our results suggest that retrieval may actually be among one of the least common of tagging motivations.

References

1. Ames, M., Naaman, M.: Why we tag: motivations for annotation in mobile and online media. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 971–980. ACM (2007)
2. Cattuto, C., Loreto, V., Pietronero, L.: Semiotic dynamics and collaborative tagging. Proceedings of the National Academy of Sciences 104(5), 1461–1464 (2007)
3. Glushko, R.J., Maglio, P.P., Matlock, T., Barsalou, L.W.: Categorization in the wild. Trends in Cognitive Sciences 12(4), 129–135 (2008)
4. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of information science 32(2), 198–208 (2006)
5. Gupta, M., Li, R., Yin, Z., Han, J.: Survey on social tagging techniques. ACM SIGKDD Explorations Newsletter 12(1), 58–72 (2010)
6. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: Proceedings of the 16th international conference on World Wide Web. pp. 211–220. ACM (2007)
7. Hastie, T.J., Tibshirani, R.J.: Generalized additive models (1990)
8. Heckner, M., Heilemann, M., Wolff, C.: Personal Information Management vs. Resource Sharing: Towards a Model of Information Behavior in Social Tagging Systems. In: ICWSM (2009)
9. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: Proceedings of 3rd European Semantic Web Confernece (ESWC). pp. 411–426. Springer (2006)
10. Körner, C., Benz, D., Hotho, A., Strohmaier, M., Stumme, G.: Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In: Proceedings of the 19th international conference on World wide web. pp. 521–530. ACM (2010)
11. Körner, C., Kern, R., Grahsl, H.P., Strohmaier, M.: Of categorizers and describers: An evaluation of quantitative measures for tagging motivation. In: Proceedings of the 21st ACM conference on Hypertext and hypermedia. pp. 157–166. ACM (2010)
12. Lorince, J., Todd, P.M.: Can simple social copying heuristics explain tag popularity in a collaborative tagging system? In: Proceedings of the 5th Annual ACM Web Science Conference. pp. 215–224. ACM (2013)
13. Lorince, J., Zorowitz, S., Murdock, J., Todd, P.M.: "Supertagger" behavior in building folksonomies. In: Proceedings of the 6th Annual ACM Web Science Conference. pp. 129–138. ACM (2014)
14. Marlow, C., Naaman, M., Boyd, D., Davis, M.: HT06, tagging paper, taxonomy, Flickr, academic article, to read. In: Proceedings of the seventeenth conference on Hypertext and hypermedia. pp. 31–40. ACM (2006)
15. Nov, O., Naaman, M., Ye, C.: What drives content tagging: the case of photos on Flickr. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 1097–1100. ACM (2008)
16. Robu, V., Halpin, H., Shepherd, H.: Emergence of consensus and shared vocabularies in collaborative tagging systems. ACM Transactions on the Web (TWEB) 3(4), 14 (2009)
17. Sen, S., Lam, S.K., Rashid, A.M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M., Riedl, J.: Tagging, communities, vocabulary, evolution. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work. pp. 181–190. ACM (2006)

18. Vander Wal, T.: Explaining and Showing Broad and Narrow Folksonomies (2005), `http://www.vanderwal.net/random/entrysel.php?blog=1635`
19. Vander Wal, T.: Folksonomy Coinage and Definition (2007), `http://vanderwal.net/folksonomy.html`
20. Wood, S.N.: mgcv: GAMs and generalized ridge regression for R. R news 1(2), 20–25 (2001)
21. Yeung, A., Man, C., Noll, M., Gibbins, N., Meinel, C., Shadbolt, N.: On measuring expertise in collaborative tagging systems (2009)
22. Yeung, C.m.A., Noll, M.G., Gibbins, N., Meinel, C., Shadbolt, N.: SPEAR: Spamming-Resistant Expertise Analysis and Ranking in Collaborative Tagging Systems. Computational Intelligence 27(3), 458–488 (2011)
23. Zollers, A.: Emerging motivations for tagging: Expression, performance, and activism. In: Workshop on Tagging and Metadata for Social Information Organization, held at the 16th International World Wide Web Conference (2007)
24. Zubiaga, A., Körner, C., Strohmaier, M.: Tags vs shelves: from social tagging to social classification. In: Proceedings of the 22nd ACM conference on Hypertext and hypermedia. pp. 93–102. ACM (2011)