Social Ranking: Uncovering Relevant Content Using Tag-based Recommender Systems

Valentina Zanardi
Dept. of Computer Science
University College London
Gower Street, London WC1E 6BT, UK
V.Zanardi@cs.ucl.ac.uk

Licia Capra
Dept. of Computer Science
University College London
Gower Street, London WC1E 6BT, UK
L.Capra@cs.ucl.ac.uk

ABSTRACT

Social (or folksonomic) tagging has become a very popular way to describe, categorise, search, discover and navigate content within Web 2.0 websites. Unlike taxonomies, which overimpose a hierarchical categorisation of content, folksonomies empower end users by enabling them to freely create and choose the categories (in this case, tags) that best describe some content. However, as tags are informally defined, continually changing, and ungoverned, social tagging has often been criticised for lowering, rather than increasing, the efficiency of searching, due to the number of synonyms, homonyms, polysemy, as well as the heterogeneity of users and the noise they introduce. In this paper, we propose *Social Ranking*, a method that exploits recommender system techniques to increase the efficiency of searches within Web 2.0. We measure users' similarity based on their past tag activity. We infer tags' relationships based on their association to content. We then propose a mechanism to answer a user's query that ranks (recommends) content based on the inferred semantic distance of the query to the tags associated to such content, weighted by the similarity of the querying user to the users who created those tags. A thorough evaluation conducted on the CiteULike dataset demonstrates that Social Ranking neatly improves coverage, while not compromising on accuracy.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering; H.3.3 [**Information Search and Retrieval**]: Query formulation; H.3.5 [**Online Information Services**]: Web-based services

General Terms

Algorithms, Performance

Keywords

Tags, Similarity, Web 2.0, Recommender Systems

RecSys'08, October 23–25, 2008, Lausanne, Switzerland.

1. INTRODUCTION

The advent of Web 2.0 has transformed users from passive consumers to active producers of content. This has tremendously increased the amount of information that is available to users (from videos on sites like YouTube and MySpace, to pictures on Flickr, to music on Last.fm, and so on). This content is no longer categorised according to pre-defined taxonomies. Rather, a new trend called *social* (or *folksonomic*) *tagging* has emerged and quickly become the most popular way to describe, categorise, search, discover and navigate content within Web 2.0 websites.

Unlike taxonomies, which overimpose a hierarchical categorisation of content, folksonomies empower end users by enabling them to *personally* and *freely* create and choose the categories (in this case, tags) that best describe a piece of information (a picture, a blog entry, a video clip, etc.). Tag clouds are then widely used to visualise a set of related tags that best describe either individual items or the content of a website as a whole, with the most frequently used tags being given more importance either in font size or color. Other visualisation techniques have been studied, in order to give more importance to tags' relationships rather than popularity [4, 11]. When users want to find content, they navigate, via hyperlinks, from a tag to a collection of items that are associated with that tag.

However, as tags are informally defined, continually changing, and ungoverned, social tagging has often been criticized for lowering, rather than increasing, the efficiency of searching [2]. This is due to the number of synonyms, homonyms, polysemy, as well as the heterogeneity of users, contexts, and the noise that they introduce.

In order to 'connect' users with content that they deem relevant with respect to their interests, efficient searching techniques have to be developed for this novel and unique domain. By efficient, we mean that the searching technique should be both *accurate* (i.e., the returned content does satisfy users' interests), and *complete* (i.e., if there is relevant content in the system, this should be found).

In this paper, we propose a technique called *Social Ranking* that aims to efficiently find, within a potentially huge dataset, content that is relevant to a user's query. In typical Web 2.0 fashion, we assume such content to have been described with an arbitrary number of tags and by an arbitrary number of users. Social Ranking answers a user's query by exploiting traditional recommender system techniques (Section 2): it measures users' similarity based on their past tag activity; it infers tags' relationships based on their association to content; finally, it ranks (recommend) content based

on the inferred distance of the query to the tags associated to such content, weighted by the similarity of the querying user to the users who created those tags. We present the results of an extensive experimental study we have conducted on the CiteULike dataset (http://www.citeulike.org/), demonstrating how Social Ranking neatly improves coverage, without compromising on accuracy (Section 3). We position ourselves with respect to other works in the area in Section 4, before presenting our conclusions and future directions of research (Section 5).

2. MODEL

2.1 Dataset Analysis

In order to understand the key characteristics of the target scenario, and thus develop a query model that is grounded on its peculiarities, we have analysed CiteULike, a typical Web 2.0 website. CiteULike is a social bookmarking website that aims to promote and develop the sharing of scientific references amongst researchers. Similarly to the cataloging of web pages within del.icio.us, and of photographs within Flickr, CiteULike enables scientists to organize their libraries with freely chosen tags which produce a folksonomy of academic interests. CiteULike runs a daily process which produces a snapshot summary of what articles have been posted by whom and with what tags. We downloaded one such archive in December 2007. The archive contained roughly 28,000 users, who had tagged 820,000 papers overall, using 240,000 distinct tags. A pre-analysis of the archive revealed the presence of a vast amount of papers and a vast amount of tags bookmarked/used by *one user only*. In order to make the dataset more manageable, we pruned it so to remove those papers and tags that had been bookmarked/used only once over the entire dataset. We were thus left with roughly 100,000 papers, 55,000 distinct tags, and 28,000 users.

We then analysed this dataset more carefully in terms of users' activity, papers' popularity, and tags' usage. Detailed results are reported in a preliminary version of this paper [27]. With respect to the problem of finding and recommending content in Web 2.0 websites, the following insights can be drawn:

Long Tail of Tags: a power law distribution curve emerges for tags' usage, identifying a small portion of frequently used tags, and a long tail (roughly 70%) of tags being used by 20 users (i.e., 0.08% of the whole user set) or less instead. Moreover, papers were described by no more than ten different tags (and usually less then five). This suggests that finding content using standard keyword based searches is likely to fail, due to empty overlaps between the tags used in the query and those associated to papers.

Long Tail of Papers: a rather steep power law distribution curve emerges for papers popularity too, identifying a small portion of papers being bookmarked (and tagged) by at least five different users, and a huge tail (more than 85%) of papers being bookmarked by less than five users instead (i.e., 0.02% of the whole user set). This suggests that standard recommender systems techniques would likely perform poorly in terms of accuracy and coverage, because of almost-empty overlaps of users' profiles.

A content search/recommender technique for Web 2.0 websites should thus be developed, taking into account these intrinsic characteristics of the target scenario. We found the following two properties to be promising to tackle both accuracy and coverage:

Clustering of Users for Improved Accuracy: although users vary a lot in terms of activity, even the most active users bookmark a rather tiny portion of the whole paper set. This suggests that users have clearly defined interests that map to a small proportion of the whole CiteULike content. This is confirmed by tags' usage: each user masters a small subset of the whole folksonomy, and users sharing part of the folksonomy form fairly small clusters. We formulate the hypothesis that, by looking at users' tag activity, *users' similarity* can be quantified and exploited to answer content searches more accurately.

Clustering of Tags for Improved Coverage: despite the emergence of a rather broad folksonomy, each paper was described by just a handful of tags. This would suggest that there is a core of shared knowledge about tags within the communities who use them, and these are recurrently used to describe related papers. We formulate the hypothesis that, by looking at what tags were associated to what papers, *tags' similarity* (or, rather, 'relationship') can be quantified and exploited to uncover relevant items during content searches.

Based on these observations, we have developed a content search and recommendation technique called *Social Ranking*.

2.2 Social Ranking

Let us consider a user $\overline{u}$ who is interested in retrieving some content of interest (in our specific case, papers). User $\overline{u}$ could explicitly submit a query $q_{\overline{u}}$ consisting of query tags $t_1, t_2, \ldots, t_n$; alternatively, in a more typical recommender system fashion, the system could implicitly run a query, using the set of tags $t_1, t_2, \ldots, t_n$ associated by the user to his latest bookmarked paper, or the set of his most frequently used tags overall, etc. In both cases, the system answering the query would normally rank results according to the following two criteria: the higher the number of query tags associated to the resource, the higher its ranking; and, the higher the number of users u_i who tagged the resource using (some of the) query tags, the higher its ranking. Intuitively speaking, the first criterion caters for accuracy of the result, the second caters for confidence in it. The formula is:

$$R(p) = \sum_{u_i} (\#t_x \text{ used by } u_i \text{ on } p \mid t_x \in q_{\overline{u}}) \ , \qquad (1)$$

that is, the ranking of paper p is computed as the number of tags t_x that users u_i who bookmarked p used and that belonged to the query set $q_{\overline{u}}$.

As we shall demonstrate experimentally in Section 3, while this simple technique works well to find popular content described with popular tags, it fails to address queries that look for the very long tail of medium-to-low popularity content, as a large amount of low-score results are returned. Accuracy is not the only problem: if the user running the query also uses tags that belong to the long tail of tags, chances are that relevant content is not found at all, and coverage then becomes the most pressing issue.

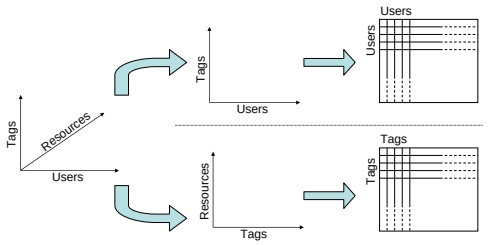

Figure 1: Transformation of the dataset

To address these problems, we propose Social Ranking, a technique inspired by traditional Collaborative Filtering mechanisms [22]: first, we identify the users with similar interests to the querying user $\overline{u}$; according to our analysis, such community should be easily identified by studying users' tag activity. Content tagged by these users should be scored higher in a way that is proportional to the quantified similarity. Second, even though tags can be broadly clustered in domains of knowledge, people tend to use slightly different subsets of them within each domain. We thus identify the tags that are similar (or, rather, related) to the query tags, thus expanding the query to this enlarged set. We believe, and our evaluation will confirm, that *users' similarity improves accuracy* of the results, while *tags' similarity (i.e., query expansion) improves coverage*.

In the remainder of this section, we illustrate how we compute users' similarity (Section 2.2.1), how we compute tags' similarity (Section 2.2.2), and how we combine these two techniques together (Section 2.2.3).

2.2.1 Users' Similarity

Social tagging typically provides a 3-dimensional relationship between users, resources and tags (users bookmark resources with a certain number of tags). Different definitions of users' similarity can be derived; here we consider a simple yet effective one: the more tags two users have used in common, the more similar they are, regardless of what resources they used it on. This definition projects our 3-dimensional space onto a 2-dimensional one, throwing away information about 'resources', and keeping only information about what tags a user has used and how often (Figure 1, top). While one may argue that, in so doing, we discard important information, we believe that, in scenarios where tags are clustered around topics, the information lost is not significant.

We thus describe each user u_i with a vector v_i where $v_i[j]$ counts the number of times that users u_i used tag t_j. Given two users u_i and u_j, we then quantify users' similarity $sim(u_i, u_j)$ as the cosine of the angle between their vectors:

$$sim(u_i, u_j) = cos(v_i, v_j) = \frac{v_i \cdot v_j}{||v_i|| * ||v_j||}$$

Various similarity measures can be used other than the cosine-based similarity [5]. For example, concordance-based similarity [1] could be used, so that the more tags two users share, the more similar they are (regardless of how many times they have used them). However, we believe tag frequency to be an important piece of information to determine a user's interests. Alternatively, Pearson Correlation (and its variations - e.g., weighted Pearson [19, 5]) could be used; as shown in [14], different similarity measures perform differently, both in terms of accuracy and coverage; we chose cosine-based similarity for its constantly good performance, although we plan to study the impact of other similarity measures in the future.

2.2.2 Tags' Similarity

We define tags' similarity as follows: the more resources have been tagged with the same pair of tags, the more similar (related) these tags are, regardless of the users who used them. This definition projects our 3-dimensional space onto a 2-dimensional one, as shown in Figure 1, bottom part. Similarly to what we said before, in scenarios where users' interests are a rather small and consistent subset of the broader range of topics in the whole website, we believe that the information thrown away during the projection is not significant.

We thus describe each tag t_i with a vector w_i where $w_i[j]$ counts the number of times that tag t_i was associated to paper p_j. Given two tags t_i and t_j, we then quantify tags' similarity $sim(t_i, t_j)$ as the cosine of the angle between their vectors:

$$sim(t_i, t_j) = cos(w_i, w_j) = \frac{w_i \cdot w_j}{||w_i|| * ||w_j||}$$

2.2.3 Two-Step Query Model

The query model we propose exploits the two similarity measures discussed above (on users and on tags) in the following way. When user $\overline{u}$ submits a query $q_{\overline{u}} = \{t_1, t_2, \ldots, t_n\}$ to discover content that can be described by query tags $t_1, t_2, \ldots, t_n$, two steps take place:

1. **Query Expansion**: the set of query tags $q_{\overline{u}}$ is expanded so to include, besides $\{t_i \mid t_i \in q_{\overline{u}}\}$ (for which $sim(t_i, t_i) = 1$), those tags $t_{n+1}, \ldots, t_{n+m}$ that are deemed most similar to the query tags (for which $0 < sim(t_i, t_j) \leq 1$, with $i \in [1, n]$ and $j \in [n+1, n+m]$). This set, which we call q^*, is constructed so to include, for each $t_i \in q_{\overline{u}}$, its top k most similar tags, in a fashion similar to the top k Nearest Neighbour (kNN) strategy in recommender systems. A thorough analysis of the impact of k on both accuracy and coverage will be presented in Section 3.

2. **Ranking**: all resources that have been tagged with at least one tag from the extended query set are retrieved. Their ranking depends on a combination of: the relevance of the tags associated to the paper with respect to the query tags (papers tagged with $t_i, i \in [1, n]$ should count more than those tagged with $t_j, j \in [n+1, n+m]$); and, the similarity of the taggers with respect to the querying user $\overline{u}$ (papers tagged by similar users should be ranked higher, as these users are more likely to share interests with $\overline{u}$ than others, and thus are in a better position to recommend relevant content).

The ranking of a paper p would then be computed as:

$$R(p) = \sum_{u_i} \left(\sum_{\substack{\{t_x | u_i \text{ tagged } p \text{ with } t_x\}, \\ t_j \in q^*}} sim(t_x, t_j) \right) * (sim(\overline{u}, u_i) + 1) \quad (2)$$

where, for each user u_i who tagged p, $\sum sim(t_x, t_j)$ quantifies how relevant the tags t_x associated by u_i to p are with respect to the tags t_j belonging to the expanded query set q^*; note that, in the basic case of formula 1, this simply meant counting how many tags from $q_{\overline{u}}$ user u_i associated to p. Moreover, the relevance is then magnified (i.e., papers are pushed higher up in the ranking) in a way that is proportional to user's similarity $sim(\overline{u}, u_i)$.

Assuming that users' similarity $sim(u_i, u_j)$ and tags' similarity $sim(t_i, t_j)$ are computed offline (i.e., daily, weekly, etc.), then the complexity of answering a query containing T tags is $O(k \cdot T \cdot P \cdot N)$, where P is the number of papers in the system and N is the number of users. However, this is a gross overestimation: as our dataset pre-analysis has shown, each tag is used on average on at most 40 papers (with $40 << P$), and each paper has been tagged on average by less than 5 users (with $5 << N$), so that the time to answer a query is simply proportional to the number of tags in the expanded query set (i.e., $k \cdot T$).

We call this approach Social Ranking, as it exploits information coming from the emergent social network of users and social network of tags to rank content in a way that is meaningful to the querying user. In the next section, we present the results obtained when evaluating this approach.

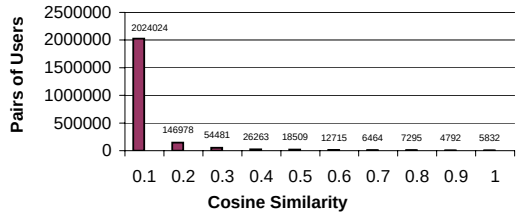

Figure 2: Distribution of users' similarity

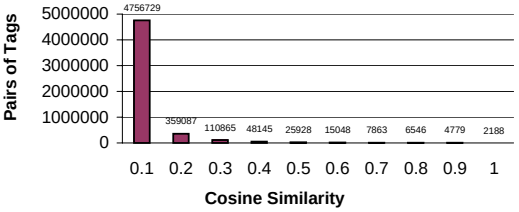

Figure 3: Distribution of tags' similarity

3. EVALUATION

We have thoroughly analysed the performance of Social Ranking on the CiteULike dataset, both in terms of accuracy and coverage (Section 3.3). Before discussing these results, we briefly illustrate the portion of the dataset we have been experimenting with (Section 3.1), and describe how we have conducted the experiments (Section 3.2).

3.1 The Dataset

Based on our pre-analysis of the CiteULike dataset (Section 2.1), we have performed a cut, in order to obtain a small yet meaningful subset to experiment with. In particular, we have considered only those tags that have been used on at least 15 different papers, and by at least 20 users. This has left us with a dataset consisting of roughly 12,000 users, 83,000 papers, and 16,000 tags. Note that the long tail phenomenon still vastly dominates in the pruned dataset:

Long tail of users' similarity: as shown in Figure 2, the vast majority of users' pairs have very low value of similarity (below 0.1), while there exists a long tail of higher similarity pairs. This would suggest users are highly focused (and clustered) around topics, and thus only a relatively small portion of users are indeed good recommenders to each other.

Long tail of tags' similarity: as shown in Figure 3, each tag is related to only a very small subset of other tags, again suggesting that only a relatively small portion of tags are used (and thus need to be learned) to describe specific categories of content.

We believe that the results we are going to present in this section generally hold for datasets that exhibit similar characteristics.

3.2 Simulation Setup

In order to quantify accuracy and coverage of Social Ranking, we have conducted the following basic experiment: we picked up a user $\overline{u}$, "hid" one of his bookmarked papers p as well as the tags that $\overline{u}$ had associated to p; we then performed a query q with such tags. Since p was bookmarked by $\overline{u}$ (before we hid it), $\overline{u}$ is obviously interested in it, so a recommender system should be able to return p (coverage). Note that, in our pruned dataset, it was always the case that, even after hiding $\overline{u}$'s bookmark for p, at least another bookmark made by a user u' for p existed, as we only kept in the dataset those papers that had been bookmarked by more than one user; it should thus be possible, in principle, to locate and return p. Moreover, the highest the ranking of p in the list of returned papers (i.e., the closest to the top), the better the accuracy of the ranking algorithm.

Given the high variability of users' behaviour and papers' popularity in the dataset, we have identified 6 different categories of experiments, based on:

- the level of activity of the querying user, distinguishing heavy taggers HT (users who tagged more than 50 papers), medium taggers MT (users who tagged between 10 and 50 papers), and low taggers LT (users who tagged less than 10 papers);
- the level of popularity of the hidden bookmark, distinguishing popular papers PP (those that had been bookmarked by at least 5 users), and unpopular ones UP (those that had been bookmarked by less than 5 users).

For each user in each group (heavy/medium/low taggers), three bookmarks where chosen at random within each paper category (popular/unpopular), hidden and their corresponding tags searched. Since the number of users in each group varies, so does the total number of queries performed (from 1,800 for the small group of HT/PP, to 13,100 for the much larger group of LT/UP). Results are reported for each category. In all experiments, we compare the output of our Social Ranking algorithm (formula 2) with the simple benchmark presented in Section 2.2 (formula 1).

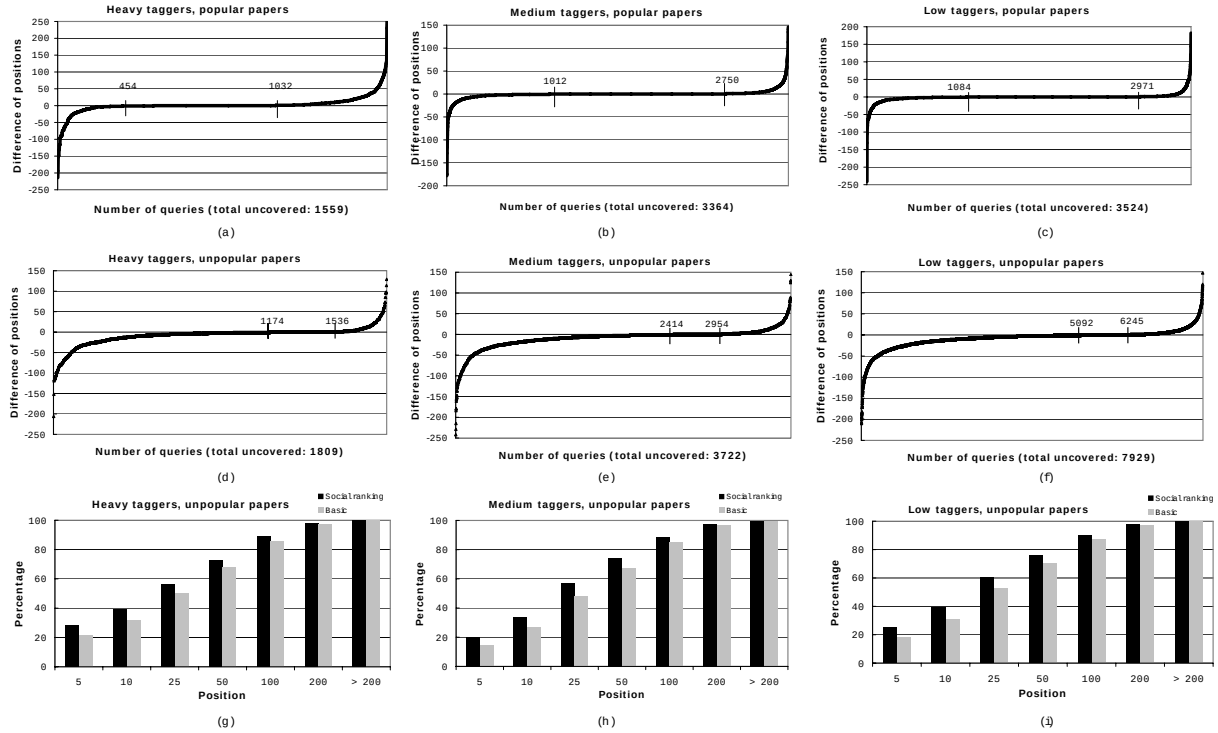

Figure 4: Social Ranking (without query expansion) vs. Basic Model

3.3 Results

3.3.1 Impact of Users' Similarity on Accuracy

The first set of experiments we conducted aimed to analyse the impact of users' similarity alone on the ranking of results. We thus compared the basic query model with the advanced query model where tag expansion had been disabled. For each query, the list of returned papers is thus the same (i.e., the search happens using the same query tags), but ordered differently (i.e., users' similarity in Social Ranking causes reshuffles). For each query that uncovered the hidden paper, we have computed the position of such paper in the ranked list of results produced by Social Ranking *minus* its position in the ranked list of results produced by the basic model: the lower the difference, the better the accuracy of Social Ranking, and viceversa. Figure 4 plots the results, sorted by the measured difference, for all six categories (first row for *popular papers* and second row for *unpopular papers*); the two x values highlighted in each chart represent the first and last query for which the two approaches perform the same (i.e., the difference in ranking is zero).

As shown, the ranking of results is slightly better when using the basic query model in the first scenario: when focusing on mainstream content (i.e., the hidden paper has been tagged many times by different users), simple searches based on exact tag matching work well enough. However, in all other scenarios, the advanced query model outperforms the basic one (i.e., it returns the hidden paper at a higher ranking in the vast majority of cases). The improvement is more dramatic when considering unpopular papers (second row), thus confirming the importance of weighting the recommendations coming from similar users more, when looking for less 'mainstream' content. If we take a closer look at the 'unpopular papers' set of results, we can notice that, on heavy taggers (Figure 4 (d)), 25% of the hidden papers are returned at positions that are between 10 and 205 positions better using Social Ranking than when using the basic model, against only 7% of cases where the basic model result is better ranked (between 10 and 130 positions gap); on medium taggers (Figure 4 (e)), 31% of results are better ranked (with a gap between 10 and 242) against 8% (with gap [10, 144]); finally, on low taggers (Figure 4 (f)), the ratio is 28% against 8%, and similar ranking gap.

In order to better appreciate the improvement obtained in terms of accuracy, we have also plotted the cumulative distribution of the ranking of the "hidden" papers, using the advanced model without query expansion and the basic model, for unpopular papers. Figures 4 (g) (h) and (i) (third row) illustrate the results: as shown, Social Ranking neatly improves the absolute ranking of the hidden paper, and it does so more evidently for heavy and medium taggers, that is, for users whose similarity can be better assessed thanks to their activity within the system. For example, about 30% of the hidden papers are found in the top 5 positions using Social Ranking on heavy taggers, against 20% using the basic model. This first set of experiments thus demonstrates our hypothesis that users' similarity is effectively exploited by Social Ranking to improve accuracy, and this is particularly important when trying to dig out unpopular content.

Let us now focus on coverage. The column labeled '$k = 0$' in Table 1 summarises the percentage of papers that remained hidden when tag expansion was not used. As shown,

Category	Queries	Not Found				
		$k=0$	$k=5$	$k=10$	$k=20$	$k=50$
HT/PP	1882	17%	8%	6%	4%	2%
MT/PP	4094	18%	8%	6%	5%	2%
LT/PP	4171	16%	8%	6%	4%	3%
HT/UP	2400	24%	14%	12%	9%	5%
MT/UP	5835	36%	22%	17%	14%	8%
LT/UP	13130	40%	26%	23%	18%	13%

Table 1: Percentage of queries remaining hidden.

this percentage is approximately 16-18% for popular papers, and it quikcly increases up to 40% for unpopular ones. Given that all papers in our dataset have been bookmarked by more than one user, low coverage is an indication that *different users bookmark the same resources differently.* Searching techniques based on user-specified query-tags only are thus unable to uncover unpopular yet relevant resources; in the next section, we demonstrate how coverage can be improved by expanding user-defined query tags to include semantically related ones.

3.3.2 Impact of Tags' Similarity on Coverage

The second set of experiments we have conducted aimed at comparing the full Social Ranking model against the basic one. During query expansion, Social Ranking extends each query tag with the top kNN tags. We have been experimenting with different values of $k = 5, 10, 20, 50$; we have been measuring the impact of the full model on both accuracy and coverage. Our goal was to neatly improve coverage, especially when dealing with unpopular content, without severely impacting on accuracy.

Table 1 reports the percentage of queries for which the target paper still remained hidden, across all values of k (including $k = 0$, that is, no query expansion). As shown, even small values of k cause the number of not-found items to quickly drop. For example, when $k = 5$, the number of *unpopular* items not found falls from 24% for heavy taggers, 36% for medium taggers, and 40% for low taggers, *down to* 14%, 22%, and 26% for the three users' categories respectively. For $k = 10$, there is an average 50% reduction of not-found items, with respect to the case of no query expansion ($k = 0$). Coverage keeps improving, although less dramatically, for higher values of k.

In order to assess the impact of query expansion on accuracy, we report two separate sets of results. For those queries that were uncovered by both Social ranking and the basic query model, we have computed the *percentiles* of the ranking of the "hidden" paper. Table 2 shows results across all 6 scenarios (Social Ranking positions on the left of each cell, and basic query model positions on the right). We only report results for $k = 10$ and $k = 20$ for space reasons. When both approaches uncover a paper, accuracy is comparable: for instance, 10% of the hidden papers are found in the top 5 positions; half of the hidden papers (50th percentile) are found in the top 10 positions in the case of popular papers, and in the top 40 positions in the case of unpopular papers, by both approaches. This confirms that the improvement obtained on coverage via query expansion does not compromise accuracy for values of k up to 20; this is aligned with our pre-analysis of the dataset, which revealed that the vast majority of papers were tagged with no more than 10 different tags: increasing k much beyond that value increases noise instead (with only a small improvement on coverage, as confirmed by Table 1).

Finally, for the set of queries uncovered by Social Ranking only, we have computed the cumulative distribution of their ranking. Once again, for space concerns, we only display the results for the critical case of *unpopular papers* and for $k = 10$ (Figure 5). As the charts illustrate, more than 40% of the papers that could not be found using the basic model, are now returned in the top 100 positions (and between 20% and 30% of them in the top 50). This second set of experiments thus demonstrates that tags' similarity can indeed be exploited, not just to uncover relevant content, but also to recommend it highly, so to bring it to the attention of the end user.

4. RELATED WORK

Research in the area of social tagging has proliferated in recent years, due to the increasing popularity of such systems. Studies have been conducted both to understand tag usage and evolution (e.g., [23, 3]), and to learn and exploit their hidden semantics. In [7], a large study of social tagging on the popular del.icio.us bookmarking system is presented, aimed at characterizing users' activity, pages' popularity, and tags' distribution; the knowledge base (in this case, the whole Web) is so large and dynamic that the authors are quite pessimistic on the benefits that social bookmarking can bring to web searches. In [6], the same authors have shown how searches on del.icio.us can be improved if a navigable hierarchical taxonomy of tags is derived from tag usage, to help users broadening/narrowing the set of tags that best describe their interests. Our approach takes a different stance, and rather than offering users an organised tag navigation system, it aims to transparently improve users' searches based on emergent tags semantics and query expansion. In [18], tags are related back to a fixed ontology of concepts, thus exploiting both techniques to enhance information retrieval capabilities. Differently from this approach, our goal is to autonomically derive tags' relationships, which can then be fitted into an effective query search algorithm, without relying on a prefixed ontology. In [20], semantics that specifically relate to places and events are inferred for resources within the Flickr dataset; their approach is highly tied to location information, and thus not easily generalizable to other domains. In [25], a probabilistic generative model is proposed to describe users' annotation behaviour, and to automatically derive tags emergent semantics; during searches, their approach is capable of grouping together synonymous tags, while it calls for user's intervention when highly ambiguous tags are found. Very early work, but with similar goals, is presented in [26], where a simpler technique, based on an analysis of the relationship between users, tags and resources, is proposed to disambiguate tags. Tag systems have recently revealed their susceptibility to tag spam, that is, malicious annotations generated to confuse users. The problem has been well analysed in [13], where the authors tried to identify misused tags, and quantify the extent to what tagging systems are robust against spam. Robust solutions to tag spamming are still being investigated.

Research has been very active also in relating tag activity to users, in order to discover their interests and consequently users' communities. Work within the Semantic Web domain has tried to classify users into categories and describe the key features of such categories [15]. More recently, users

Category	Percentiles ($k = 10$)										Percentiles ($k = 20$)									
	5		10		25		50		75		95		5		10		25		50		75		95	
HT/PP	1	1	1	1	2	2	6	5	27	23	102	88	1	1	1	1	2	2	7	5	29	23	112	88
MT/PP	1	1	1	1	1	1	4	3	15	13	70	64	1	1	1	1	2	1	5	3	17	13	80	64
LT/PP	1	1	1	1	1	1	3	3	12	10	82	71	1	1	1	1	1	1	4	3	15	10	88	71
HT/UP	2	2	3	3	10	7	35	25	80	67	186	162	2	2	4	3	12	7	39	25	86	67	215	162
MT/UP	3	2	5	4	12	9	31	27	76	67	207	170	3	2	5	4	13	9	35	27	85	67	245	170
LT/UP	2	2	4	4	9	8	26	23	71	61	209	169	2	2	4	3	10	8	30	23	80	61	257	169

Table 2: Percentiles of the ranking of results, for Social Ranking vs. Basic Model

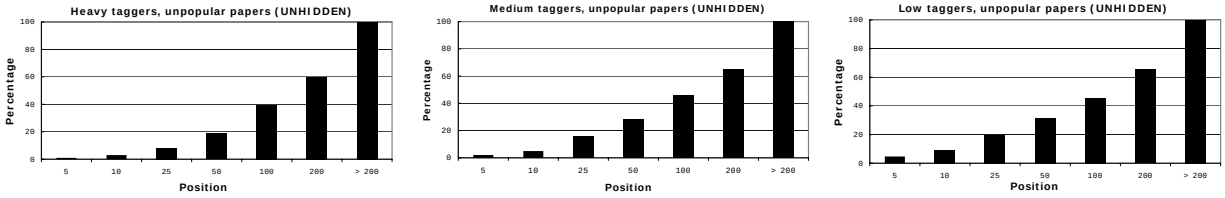

Figure 5: Accuracy of queries uncovered by Social Ranking ($k = 10$)

have been classified according to their explicitly stated profile [9], based on a probabilistic model which takes into account users's interest to topics [28], and based on their level of tagging activity and breadth of interests [12]. In [16], users' common interests are discovered based on patterns of frequently co-occurring tags, using a classical association rule algorithm, which however does not take into account considerations about user's activity. All these works, including our attempt to find similar users, are based on the observation that real world networks exhibit a so-called community structure [21]; defining the set of characteristics that would enable the best fitting and natural clustering of taggers and tags is an open research question.

In this paper, we have been combining the two research streams highlighted above (i.e., automatic learning of tag semantics and users' interests) in order to improve query searches and ranking. Other research groups have been conducting research in the same area. In [24, 17], the integration of tag information within standard recommender system's algorithms has been proposed, in order to give better recommendations to users; although very promising, at present such works do not take into account the 'activity' of users, in terms of amount of resources being tagged, and number of tags being used. We believe this information to be crucial to extract users' interests and thus improve the efficiency of searches. Tag activity has been combined with a PageRank-like algorithm, in order to improve the ranking mechanism, in situations where resources are not linked together as in a typical web graph structure [8]; their approach, called FolkRank, provides good results when querying the folksonomy for topically related elements, while it is easily subverted if less related/popular tags are being used, due to the size and sparsity of folksonomies on the web. In [10], users' similarity is exploited first to generate a set of tags of relevance to the target user, then to recommend him items described by such tags; as for FolkRank, this approach is mostly tailored to scenarios with high users' activity and low tag noise. Social Ranking focuses on scenarios where there is only little information shared between users instead. In these settings, we have demonstrated how a combination of users' and tags' similarity can ameliorate the sparsity problem. Further improvements could be achieved by clustering users within better scoped communities; we intend to explore this aspect next.

5. CONCLUSION

In this paper we have presented Social Ranking, a technique that aims to improve content searches in Web 2.0 scenarios, by exploiting users' similarity and tags' similarity. The former is used to improve accuracy: the higher the similarity between the querying user and the user that has bookmarked it, the higher the chances that the paper is of relevance, thus reducing the amount of uninteresting content being presented to users. The latter is used to improve coverage: by implicitly learning tags' similarity from their usage, a larger amount of relevant yet unpopular content can be uncovered, thus ameliorating the problem of heterogeneity, sparsity and lack of structure in folksonomy.

Our ongoing work spans different directions. First, we are refining the techniques we use to find similar users and similar tags. For the former, we have started analysing the impact of a variety of clustering techniques to identify communities of users with shared interests; beyond similarity in the tags' usage, there exist other parameters of relevance, including level of activity (to distinguish active users who contribute to the knowledge base, from passive consumers), variety of tags used (unpopular tags may reveal more about a user's interests than popular ones), and so on. For the latter, we are enriching query expansion with words that are semantically close, as defined by dictionary-based approaches like WordNet (http://wordnet.princeton.edu/). Further evaluation is also required, using different Web 2.0 datasets (e.g., Last.fm, Bibsonomy, del.icio.us, etc.), different similarity measures (e.g., Pearson, concordance, etc.), and comparing against less naive approaches (e.g, [8]).

Acknowledgments. The authors would like to thank Sonia Ben Mokhtar, Franco Raimondi, Neal Lathia and Matteo Dell'Amico for their continuous help and the useful discussions which lead to the publication of this work.

6. REFERENCES

[1] A. Agresti. Analysis of Ordinal Categorical Data. John Wiley and Sons, 1984.

[2] S. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.

[3] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *Proc. of the 16th Intl. Conference on World Wide Web*, pages 211–220, New York, NY, USA, 2007.

[4] Y. Hassan-Montero and V. Herrero-Solana. Improving tag-clouds as visual information retrieval interfaces. In *Intl. Conference on Multidisciplinary Information Sciences and Technologies*, Merida, Spain, October 2006.

[5] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An Algorithmic Framework for Performing Collaborative Filtering. In *Proc. of the 22nd Intl. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 230–237, 1999.

[6] P. Heymann and H. Garcia-Molina. Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. Technical Report 2006-10, Stanford University, April 2006.

[7] P. Heymann, G. Koutrika, and H. Garcia-Molina. Can Social Bookmarking Improve Web Search? *Resource Shelf*, November 2007.

[8] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information Retrieval in Folksonomies: Search and Ranking In *Proc. of the 3rd European Semantic Web Conference*, pages 411–426, 2006.

[9] W. H. Hsu, J. Lancaster, M. S. Paradesi, and T. Weninger. Structural Link Analysis from User Profiles and Friends Networks: A Feature Construction Approach. In *Intl. Conference on Weblogs and Social Media*, March 2007.

[10] A.-T. Ji, C. Yeon, H.-N. Kimand, and G.-S. Jo. Collaborative Tagging in Recommender Systems. In *Advances in Artificial Intelligence*, March 2007.

[11] O. Kaser and D. Lemire. Tag-Cloud Drawing: Algorithms for Cloud Visualization, Tagging and Metadata for Social Information Organization. In *Intl. Conference on the World Wide Web*, Alberta, Canada, October 2007.

[12] S. Kelkar, A. John, and D. Seligmann. An Activity-based Perspective of Collaborative Tagging. In *Intl. Conference on Weblogs and Social Media*, March 2007.

[13] G. Koutrika, F. A. Effendi, Z. Gyöngyi, P. Heymann, and H. Garcia-Molina. Combating spam in tagging systems. In *Proc. of the 3rd Intl. Workshop on Adversarial Information Retrieval on the Web*, pages 57–64, New York, NY, USA, 2007. ACM Press.

[14] N. Lathia, S. Hailes, and L. Capra. The effect of correlation coefficients on communities of recommenders. In *Proc. of 23rd ACM Symposium on Applied Computing*, 2008.

[15] K. F. Lawrence and M. C. Schraefel. Bringing Communities to the Semantic Web and the Semantic Web to Communities. In *Proc. of the 15th Intl. Conference on World Wide Web*, 2006.

[16] X. Li, L. Guo, and Y. E. Zhao. Tag-based Social Interest Discovery. In *Proc. of the 17th Intl. World Wide Web Conference*, 2008.

[17] R. Nakamoto, S. Nakajima, J. Miyazaki, and S. Uemura. Tag-based Contextual Collaborative Filtering. In *18th IEICE Data Engineering Workshop*, 2007.

[18] A. Passant. Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs. In *Proc. of Intl. Conference on Weblogs and Social Media*, 2007.

[19] H. Polat and W. Du. Privacy-Preserving Collaborative Filtering using Randomized Perturbation Techniques. In *The 3rd IEEE Intl. Conference on Data Mining (ICDM)*, Melbourne, FL, November 2003.

[20] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *Proc. of the 30th Intl. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 103–110, New York, NY, USA, 2007.

[21] J. Ruan and W. Zhang. Identifying network communities with a high resolution. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 77(1), 2008.

[22] J. Schafer, J. A. Konstan, and J. Riedl. Recommender Systems in E-commerce. In *ACM Conference on Electronic Commerce*, pages 158–166, 1999.

[23] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, M. F. Harper, and J. Riedl. tagging, communities, vocabulary, evolution. In *Proc. of the 20th Conference on Computer Supported Cooperative Work*, pages 181–190, New York, NY, USA, 2006.

[24] K. H. L. Tso-Sutter, L. B. Marinho, and L. Schmidt-Thieme. Tag-aware Recommender Systems by Fusion of Collaborative Filtering Algorithms. In *Proc. of 23rd ACM Symposium on Applied Computing*, pages 16–20, 2008.

[25] X. Wu, L. Zhang, and Y. Yu. Exploring social annotations for the semantic web. In *Proc. of the 15th Intl. Conference on World Wide Web*, pages 417–426, New York, NY, USA, 2006.

[26] C. M. A. Yeung, N. Gibbins, and N. Shadbolt. Mutual Contextualization in Tripartite Graphs of Folksonomies. In *Proc. of the 6th Intl. Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), Busan, South Korea*, volume 4825 of *LNCS*, pages 960–964, November 2007.

[27] V. Zanardi and L. Capra. Social Ranking: Finding Relevant Content in Web 2.0. In *Intl. Workshop on Recommender Systems. In conjunction with the 18th European Conference on Artificial Intelligence (ECAI)*, Patras, Greece, July 2008.

[28] D. Zhou, E. Manavoglu, J. Li, L. C. Giles, and H. Zha. Probabilistic models for discovering e-communities. In *Proc. of the 15th Intl. Conference on World Wide Web*, pages 173–182, New York, NY, USA, 2006.