Social Wisdom for Search and Recommendation

Ralf Schenkel, Tom Crecelius, Mouna Kacimi, Thomas Neumann,
Josiane Xavier Parreira, Marc Spaniol, Gerhard Weikum
Max-Planck Institute for Informatics
Saarbruecken, Germany
schenkel@mpi-inf.mpg.de

Abstract

Social-tagging communities offer great potential for smart recommendation and "socially enhanced" search-result ranking. Beyond traditional forms of collaborative recommendation that are based on the item-user matrix of the entire community, a specific opportunity of social communities is to reflect the different degrees of friendships and mutual trust, in addition to the behavioral similarities among users. This paper presents a framework for harnessing such social relations for search and recommendation. The framework is implemented in the SENSE prototype system, and its usefulness is demonstrated in experiments with an excerpt of the librarything community data.

1 Introduction

Social networks and online communities provide a great potential for harnessing the "wisdom of crowds", with social interactions of individual users and user groups taken into account. For example, bookmark-sharing services such as del.icio.us can generate collaborative recommendations based on the quality and trust assessment of web pages as well as users. Social-tagging platforms such as flickr, librarything, or lastfm enable community formation, based on common thematic interests, and thus provide ratings and rankings of photos, books, music, etc., based on the social interactions among many users.

These settings resemble the paradigm of collaborative recommendation [5, 12, 17, 19], which applies data mining on customer-product and similar usage data to predict items that users are likely interested in. Such recommendations leverage user-user similarities as well as item-item similarities. For the first aspect, joint behaviour patterns of two users can be exploited, e.g., the number of items purchased by both users. For the second aspect, the overlap in the interests of users in two items can be exploited, e.g., the number of users who purchased both of two items. A popular approach is to apply data-analysis methods (e.g., spectral decomposition) to a user-item matrix.

Social wisdom for searching, ranking, and recommending items differs from such traditional recommender systems in two important ways:

1. There are explicit *friendship* and *trust* relations among users that are orthogonal to similarities of interests and behavior, and these truly social relations can significantly affect the quality of recommendations.

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

In contrast, traditional recommenders consider the user community only in its entirety, whereas social recommenders would discriminate different users based on friendship strengths, mutual trust, etc.

2. There is often a search or discovery *context* like sets of keywords (not necessarily corresponding to the tags of the existing data items) or a real-life task like planning a trip or buying christmas presents, that characterize the ideal results that the user hopes to obtain. This is in contrast to the weakly parameterized nature of traditional recommenders where you can start with a given item to discover new items but not with a vague description that does not match any existing item.

This paper presents a framework for exploiting social wisdom in such a setting, and discusses the ranking of search results and recommendations. Despite the relatively young age of social-tagging communities, there is already a sizable body of literature on a variety of socially enhanced scoring and ranking functions, e.g., [7, 14, 24]. However, most of the prior work has focused on very specific points, such as applying generalized link analysis (i.e., PageRank-style notions of FolkRank, UserRank, SocialRank, etc.) to identify the most central (influential) users or items; and some of the empirical studies have actually raised doubts about the benefit of social tags and friendship relations for improving search [11, 13]. In contrast, this paper aims at a comprehensive framework for scoring, with consideration of both social relations and semantic/statistical relations among items and tags. To this end, we introduce a versatile and richly parameterized scoring model, and we present experiments with librarything data and user-provided quality assessments that demonstrate significant benefits. Throughout the paper we disregard efficiency and scalability issues; these are challenging, too, but out of scope (refer to [4, 8, 20, 23] for efficiency issues).

The paper is organized as follows. Section 2 presents our framework for modeling social-tagging networks and our prototype system coined SENSE (standing for "Socially ENhanced Search and Exploration"). Section 3 presents the socially enhanced scoring model for search and proactive recommendation. Section 4 discusses a user study for experimentally evaluating our approach, based on an excerpt of the librarything community. We conclude with lessons learned and an outlook on further research opportunities.

2 SENSE Framework

We studied a variety of social-tagging platforms, most notably, `del.icio.us`, `flickr`, `librarything`, and `lastfm`, in order to come up with a unified set of abstractions that can model the user-provided data and activities in such communities. The resulting model can be cast into a relational schema of the following form (with unique keys underlined):

$$\begin{array}{ll}
\texttt{Users}(\underline{\texttt{username}}, \texttt{location}, \texttt{gender}, \ldots) & \text{//abbreviated: U} \\
\texttt{Friendships}(\underline{\texttt{user1}, \texttt{user2}, \texttt{ftype}}, \texttt{fstrength}) & \text{//abbreviated: F} \\
\texttt{Documents}(\underline{\texttt{docid}}, \texttt{description}, \ldots) & \text{//abbreviated: D} \\
\texttt{Linkage}(\underline{\texttt{doc1}, \texttt{doc2}, \texttt{ltype}}, \texttt{lweight}) & \text{//abbreviated: L} \\
\texttt{Tagging}(\underline{\texttt{user}, \texttt{doc}, \texttt{tag}}, \texttt{tweight}) & \text{//abbreviated: T} \\
\texttt{Ontology}(\underline{\texttt{tag1}, \texttt{tag2}, \texttt{otype}}, \texttt{oweight}) & \text{//abbreviated: O} \\
\texttt{Rating}(\underline{\texttt{user}, \texttt{doc}}, \texttt{assessment}) & \text{//abbreviated: R}
\end{array}$$

We refer to all kinds of data items as *documents*, using IR jargon. These are the items that users explicitly upload (e.g., their own photos) or bookmark and annotate (e.g., web pages, books, songs). Items may be cross-referenced by different types of (possibly weighted) links.

Friendships are user-user relations that come in different forms; this is why we allow multiple types of F relations captured by the *ftype* attribute. *Social friendship* is an explicit, user-provided relation, which can be symmetric or asymmetric; we assume that such a relation exists only if the users know each other by some interaction (in real life or in cyberspace). *Spiritual friendship* among "brothers in spirit", on the other hand, captures similar behavior such as memberships in the same groups or high overlap in tag usage; these are symmetric relations and they do not assume that spiritually related users know each other. The strength of an F relation between two users could be derived from the users' activities such as overlap in tagged documents or trust measures derived from mutual comments and ratings. We treat $fstrength$ as a pluggable building block; it may also be completely absent.

Tagging is a ternary relation between users, documents, and tags. In full generality, it can *not* be decomposed into three binary relations (users-docs, docs-tags, users-tags) without losing information. Nevertheless, binary-relation (or, equivalently, graph or matrix) representations for tagging are very popular in the literature on social networks for convenience. Our approach preserves the full information and feeds it into a scoring model. Independently of tagging activities, the R relation allows users to rate the quality of individual documents. Alternatively, we could aggregate data from the T relation to derive quality measures (e.g., interest or trust in an information source) and keep it as an attribute of R.

Ontology is a light-weight knowledge base that captures different types of "semantic" relations among tags (e.g., synonymy or specialization/generalization). These relations may be provided by domain experts or imported from real ontologies, or they may be built by applying data-mining techniques to the tagging data. The latter case is more realistic for today's types of social tagging communities and is often referred to as "folksonomies" (folklore taxonomies); in this case, the $oweight$ values could be based on tag-usage statistics.

Note that this model is much richer than the datasets in traditional recommender systems. In addition to the shown relations, we can easily add various kinds of aggregation views, for example, document-tag frequencies aggregated over all users. Also note that not all of its aspects apply to every tagging platform (e.g., only few communities would show the users' home locations, some do not facilitate any cross-references among individual items, etc.). In fact, our experimental studies presented in Section 4 utilize only a subset of our model.

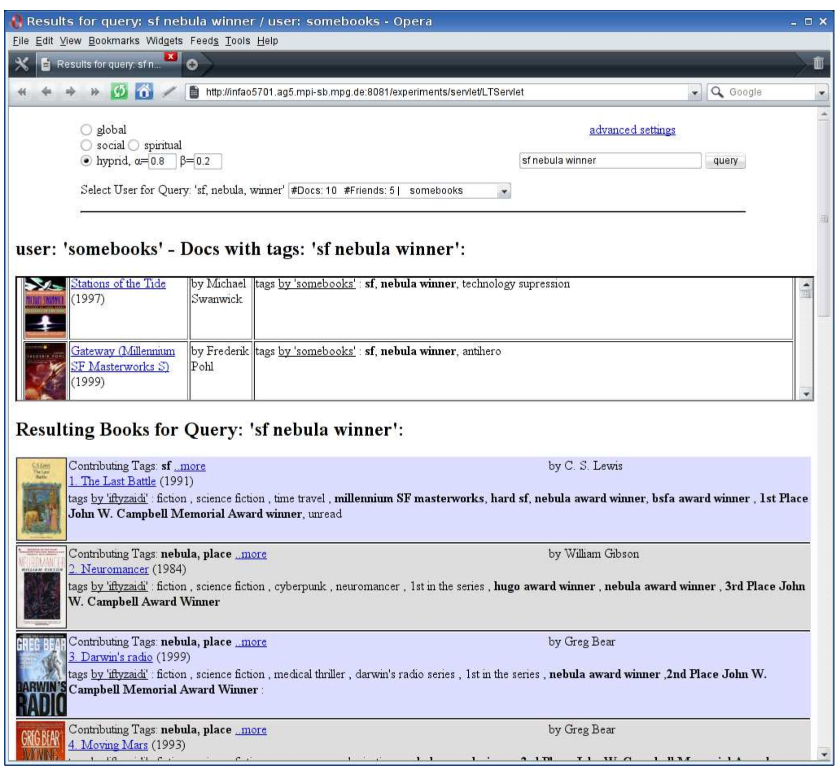

Figure 1: SENSE Screenshot

We have implemented this model based on a relational database system, and populated it with different instances derived from partial crawls of real-life tagging communities. For systematic experimentation with user assessments, we have also built a GUI that supports interactive search and browsing with capabilities for explicit and flexible exploitation of social relations. Figure 1 shows a screenshot of this toolkit, coined SENSE (for "Socially ENhanced Search and Exploration"). The figure shows the ranked results for the query "sf nebula winner" issued by a particular user on an excerpt from the librarything community. The top portion of the results is based on the user's own book collection; the bottom part shows the results obtained by searching other users' collections with consideration of the query initiator's specific friendship relations.

3 Scoring Model

Consider a *query* $Q(u, q_1 \dots q_n)$, issued by a query initiator u with a set of tags $q_1 \dots q_n$. Result documents should contain at least one of the query tags and be ranked according to a *score*. In contrast to standard IR query models, our scoring function can be tuned towards different aspects of social communities. Scores are *user-specific*: they depend on the social and/or spiritual context of the query initiator, according to the configuration of the model. The querying user can decide if her information need is 1) spiritual, 2) social, or 3) global, the latter being agnostic to her social relations.

Friendship Strengths. The core of the scoring model is formed by three different quantizations for user-user affinity or *friendship strength*, corresponding to the three different search modes: spiritual, social, global. Each type of affinity can be implemented in different ways, and our current implementation allows switching between and combining different definitions at run-time. The *spiritual* friendship strength $F_{sp}(u, u')$ of two users u and u', tuned towards spiritual search, is computed based on user-behavior statistics such as overlap of tag usage, bookmarked documents, or commenting and rating activity. The *social* friendship strength $F_{so}(u, u')$, applied for social search, is based on measures like the inverse distance of u and u' in the friendship graph (F relation), but may additionally include other measures. As F_{so} considers also transitive friendships, we have options for different weighting schemes of more distant friends: linearly descending with increasing distance, harmonically descending, geometrically descending, or solely counting immediate friends. The *global* affinity $F_{gl}(u, u') = \frac{1}{|U|}$, used for global searches, gives equal weight to all users. All these measures are normalized such that $\sum_{u' \in U} F(u, u') = 1$ for all u.

The actual friendship strength used to evaluate a query is a linear combination of these three measures:

$$F_u(u') = \alpha \cdot F_{so}(u, u') + \beta \cdot F_{sp}(u, u') + (1 - \alpha - \beta)\frac{1}{|U|}$$

The parameters α and β, $0 \leq \alpha, \beta \leq 1$, can be configured and dynamically adjusted by the user (or an agent on behalf of the user). Extreme choices would be purely spiritual ($\alpha = 0, \beta = 1$), purely social ($\alpha = 1, \beta = 0$), or purely global ($\alpha = 0, \beta = 0$) search; other combinations are reasonable and more interesting.

Score for Tags. To compute the score $s_u(d, t)$ of a document d with respect to a single tag t relative to the querying user u, we use a scoring function in the form of a simplified BM25 score [18]:

$$s_u(d, t) = \frac{(k_1 + 1) \cdot |U| \cdot sf_u(d, t)}{k_1 + |U| \cdot sf_u(d, t)} \cdot idf(t)$$

where k_1 is a tunable coefficient (just like in standard BM25), $|U|$ is the total number of users, $sf_u(d, t)$ is a user-specific tag frequency explained below, and $idf(t)$ is the inverse document frequency of tag t, instantiated as

$$idf(t) = \log \frac{|D| - df(t) + 0.5}{df(t) + 0.5}$$

with $df(t)$ denoting the number of documents that were tagged with t by at least one user. BM25 is a probabilistic IR model that has been very successful and popular in text retrieval. Unlike the original BM25 formula, our model has no notion of document lengths; the number of tags assigned to a document does not vary as much as the length of text documents.

The *socially-enhanced tag frequency* $sf_u(d,t)$, our replacement for the standard term frequency (tf) known from text IR, weights tags by the friendship strength of the query initiator and the user who added the tag to the document. More formally, denoting by $tf_u(d,t)$ the number of times user u used tag t for document d, we define the socially-enhanced tag frequency $sf_u(d,t)$ for a tag t and a document d, relative to a user u, as

$$sf_u(d,t) = \sum_{u' \in U} F_u(u') \cdot tf_{u'}(d,t).$$

For example, if Alice has four (out of her many) friends who have tagged d with t (each once, as it is the norm in social tagging) and each of these immediate friends has weight $1/8$ and only immediate friends matter, then $sf_{Alice}(d,t)$ would be $4 \cdot 1/8 = 1/2$.

Tag Expansion. Even though related users are likely to have tagged related documents, they may have used different tags to describe them. It is therefore essential to allow for an expansion of query tags to “semantically” related tags. To avoid topic drift problems, we adopt the *careful expansion* approach proposed in [22] which considers, for the score of a document, only the best expansion of a query tag, not all of them. More formally, we introduce the *tag similarity* $tsim(t_1,t_2)$ (an instantiation of the $oweight$ attribute of the O relation) for a pair of tags t_1 and t_2, $0 \leq tsim(t_1,t_2) \leq 1$. The final score $s^*_u(d,t)$ of a document d with respect to a tag t and relative to a querying user u, considering tag expansion, is then defined as

$$s^*_u(d,t) = \max_{t' \in T} tsim(t,t') \cdot s_u(d,t')$$

A good source for high-quality tag expansions would be human-made ontologies, however for most applications, it is unlikely that they will be available. Our implementation therefore provides several alternatives to compute the similarity between two tags, none of which requires that an explicit ontology is available. The currently preferred one is based on the co-occurrence of the tags in the entire document collection by estimating conditional probabilities:

$$tsim(t,t') = P[t|t'] = \frac{df(t)}{df(t \wedge t')}$$

where $df(t \wedge t')$ is the number of documents that have been tagged by both tags (but possibly by different users). This asymmetric measure (as opposed to symmetric similarities such as Dice or Jaccard coefficients) aims to identify good *specialization* or *instantiation* tags rather than synonymy or generalization. For example, someone searching for “snake” may be happy to see results that contain the specialized tags “Black Mamba”, “Cobra”, etc., but is not interested in documents that feature more general tags such as “vertebrate” or “animal” as they will probably lead to results that are too general as well. In fact, one would expect a much higher probability, in the underlying dataset, that a document tagged “Cobra” also has the “snake” tag than, conversely, a document tagged “snake” also having the tag “Cobra” (simply because Cobras are only one of many types of snakes). Similar techniques for mining asymmetric tag relations have been used in different contexts (e.g., [2, 6, 9]). Note that the same similarity measure cculd be applied to measure the strength of relationships in an ontology; here, the pairs of tags under consideration would be those connected by an edge in the ontology.

The above form of tag expansion captures “semantic” associations, but disregards the social relations among users. For *socially-enhanced tag expansion* we compute the similarities between tags in a way that gives a co-occurrence higher weight if the two tags were given by a close friend of the current user u. This idea leads to the formula:

$$tsim_{so}(u,t,t') = \sum_{u' \in U} F_u(u') \cdot \frac{df_{u'}(t')}{df_{u'}(t \wedge t')}$$

where $df_{u'}(t \wedge t')$ is the number of documents tagged by the same user u' with both t and t'. Intuitively, we postulate that user u is interested in seeing Ferraris as a result to a query about sports cars if her friends prefer Ferraris and show this in their tagging activities.

Score for Queries. Finally, the score for an entire query with multiple tags $q_1 \ldots q_n$ is the sum of the per-tag scores:

$$s_u^*(d, q_1 \ldots q_n) = \sum_{q_1 \ldots q_n} s_u^*(d, q_i)$$

Note that this score assumes an IR-style "andish" query evaluation: not all query tags must be matched, but more matches typically lead to higher scores. However, the model can easily be extended to conjunctive evaluation by setting $s_u^*(d, q_1 \ldots q_n) = 0$ when at least one of the $s_u^*(d, q_i) = 0$.

4 Experiments

4.1 Setup

To study the effectiveness of the socially-enhanced scoring model, we performed experiments with data extracted from partial crawls of the `del.icio.us`, `flickr`, and `librarything` sites. We concentrate on the librarything data here, for lack of space and also because this is the most interesting of the three scenarios. We found the social aspects in del.icio.us to be rather marginal, as most bookmarked pages are of fairly high quality anyway; so a user does not benefit from his friends' recommendations more than from the overall community. Flickr has recently grown so much that the tagging quality seems to be gradually degrading; only the owner of a photo provides tags, and these are sometimes relatively unspecific annotations that are given to all photos of an entire series (e.g. vacation July 2007). Librarything, on the other hand, features intensive tagging of a quality-controlled set of items, namely, published books, and its users have built up rich social relations. Finally, book recommendation is a matter of subjective taste, so that social relations do indeed have high potential value. You trust your friends' taste, not necessarily their "technical" expertise.

We extracted the following data from the librarything site: 11,717 users who together own or have read 1,289,128 distinct books with a total of 14,738,646 tagging events (including same tags for the same book by different users), and 17,915 explicit friendships. For the latter, we used the librarything notion of friends (where users mutually agree on being friends) and the notion of referring to an "interesting library". The users included 6 users from our institute who have been contributing to librarything for an extended time period and have made various social connections. These 6 users ran recommendation queries and assessed the quality of the results in our study. Note that such human assessment is indispensable for this kind of experiments, and in our setting it was crucial that a query result was assessed by the same user who posed the query. Altogether, our 6 test users ran 49 queries, shown in Table 1.

Query results were computed for a variety of scoring models: different values of α and β and different strategies for tag expansion. The results from all runs for the same query were pooled; all of them together were shown to the corresponding user in random order (in a browser-based GUI), and the user assessed the quality of each result by assigning one of three possible ratings: 0 = irrelevant or uninteresting, 1 = relevant and interesting, 2 = super-relevant and very appealing. Results that the user already knew, that is, books that she has in her own library, are always discounted.

As for quality measures, we computed, for each run separately:

- the *precision* for the top-10 results, treating both ratings of 2 and 1 as relevant,
- the *normalized discounted cumulative gain (NDCG)* [15] for the top-10 cutoff point. DCG aggregates the ratings (2, 1, or 0) of the results with geometrically decreasing weights towards lower ranks ($DCG \propto$

user 1	user 2	user 3	user 4	user 5	user 6
thailand travel	web learning	time traveler	religion god world	information retrieval	sf nebula winner
asia guide travel	mountain climbing	leonardo vinci	challenge theory	probability statistics	fantasy politics
technology enhanced learning knowledge management	kali death	english grammar	imagination fantasy science	database system	fantasy dragaera
multimedia metadata standards	buddha	romance prague	drama story novel	transaction management	sf nuclear war
knowledge management media theory	houdini	brazilian literature	magic fantasy	data mining	fantasy malazan
social network analysis theory	science illusion magic	shalespeare play	india philosophy	software development	
multimedia social software	mystery magic	stephanie plum	fantasy story		
	religion irony humor	search engines	novel family life		
	yakuza	spanish literature	science fiction future		
	hitman	portuguese literature			
		harry potter			
		wizard			

Table 1: Queries of the user study

α \ β	0.0	0.2	0.5	0.8	1.0
0.0	0.666	**0.698**	**0.688**	**0.682**	0.680
0.2	0.661	0.678	0.686	**0.690**	n/a
0.5	0.637	0.657	0.663	n/a	n/a
0.8	0.612	0.647	n/a	n/a	n/a
1.0	0.549	n/a	n/a	n/a	n/a

Table 2: Precision[10] for all users

α \ β	0.0	0.2	0.5	0.8	1.0
0.0	0.546	**0.572**	**0.568**	0.565	0.565
0.2	0.564	**0.572**	**0.579**	**0.581**	n/a
0.5	0.539	0.552	0.559	n/a	n/a
0.8	0.515	0.546	n/a	n/a	n/a
1.0	0.465	n/a	n/a	n/a	n/a

Table 3: NDCG[10] for all users

$\sum_{rank\ i} \frac{2^{rating(i)}-1}{log_2(1+i)}$) and is then normalized into NDCG by dividing by the DCG of an ideal result (first all results with rating 2, followed by all results with rating 1, followed by results with rating 0). NDCG is a widely adopted standard measure in IR.

4.2 Results

Tables 2 and 3 show the precision and NDCG values for different choices of the configuration parameters α and β, without any form of tag expansion. These are micro-averaged results over all test users. Values printed in boldface are results that were significantly better than the baseline case ($\alpha = \beta = 0$) according to a statistical t-test with test level 0.1.

The results show that both social (increasing α) and spiritual (increasing β) processing can improve the result quality. This holds for each of these two directions individually, and the combined effect is even better with a typical maximum at $\alpha = 0.2$ and $\beta = 0.8$. It may seem that the improvements, for example, from an NDCG value of 0.546 for the baseline to 0.581 for the best case, is not impressive. However, one has to keep in mind that differences in such effectiveness measures generally tend to be small in IR experiments as opposed to efficiency differences (e.g., response times) in the DB literature; we emphasize that the gains are statistically significant. Moreover, it is worth pointing out that for some individual users (i.e., micro-averaging over the queries of one user only) or for individual queries the gains are higher. As anecdotic evidence, the query “science illusion magic” posed by User 2 strongly benefited from the user’s social relations: with global scoring alone, many good results were missed; with spiritual scoring alone, the results drifted towards a big “Harry Potter” cluster which was not what the user wanted; only the combination of social and spiritual similarity gave the excellent results that the user appreciated (which included novels such as “Prestige”, “Labyrinths”, “Invisible Cities”).

α \ β	0.0	0.2	0.5	0.8	1.0
0.0	0.545	0.565	0.565	0.563	0.565
0.2	0.561	0.573	0.581	0.582	n/a
0.5	0.538	0.550	0.554	n/a	n/a
0.8	0.506	0.540	n/a	n/a	n/a
1.0	0.459	n/a	n/a	n/a	n/a

Table 4: NDCG[10] with expansion, up to 5 expansions per tag

α \ β	0.0	0.2	0.5	0.8	1.0
0.0	0.537	0.560	0.558	0.556	0.556
0.2	0.535	0.550	0.567	0.564	n/a
0.5	0.515	0.536	0.545	n/a	n/a
0.8	0.487	0.522	n/a	n/a	n/a
1.0	0.454	n/a	n/a	n/a	n/a

Table 5: NDCG[10] with social expansion, up to 5 expansions per tag

Tables 4 and 5 show the NDCG results with tag expansion enabled, aggregated over all 49 queries of the user study. We compared the purely semantic expansion that ignores social relations against the socially-enhanced tag expansion that prefers tags used by friends. Across the entire query mix of all users, neither of the two expansion methods achieved significant improvements, but again, for individual users such as User 2 there were noticeable gains. For example, the query “Yakuza” created the expansion tags “Cosa Nostra”, “Triads”, and “nightclub” (among the top-5 expansions); the first and second expansion could have been expected (and created also by an ontology-based method), but the third expansion really reflected tag co-occurrences and implicitly the contents of the kinds of novels that the user wished to discover. Social expansion, on the other hand, did not improve results; in fact, it sometimes reduced the quality. For example, by considering the friendships of User 2, the “Yakuza” query ended up with the expansion “Ninjas” and led to poorer query results.

5 Lessons Learned and Open Issues

In this paper, we have developed a comprehensive framework for socially enhanced search, ranking, and recommendation. Our experimental evaluation exhibits interesting results and indicates the potential of exploiting social-tagging information for scoring and ranking. However, the results reveal mixed insights, and thus also underline the need for further investigating this line of research.

The combination of social and spiritual scoring nicely improved the results of certain queries or users, but also led to result degradation in other cases. On average, there is a significant gain but it is not as impressive as one could have hoped for. It seems that categorizing queries and identifying the query types that can benefit from social and spiritual relations is the key to a robust solution that would choose non-zero values for α and β only when benefits can be expected. In our user study, the queries seem to fall into the following four categories:

1. Queries with a purely *global* information need that perform best when $\alpha = \beta = 0$; examples are “Houdini”, “search engines”, “English grammar”, all fairly precisely characterized topics with objectively agreeable high-quality results.

2. Queries with a subjective-taste and thus *social* aspect that perform best when $\alpha \approx 1$; an example is the query “wizard”. This query produces a large number of results but the user may like only particular types of novels such as “Lord of the Rings”, for which “wizard” is a relatively infrequent tag overall but was frequent among that user’s friends.

3. Queries with a spiritual information need that perform best when $\beta \approx 1$; an example is the query “Asia travel guide” where one can harness the aggregated expertise of the entire user community without consideration of social relations.

4. Queries with a mixed information need that perform best when $\alpha, \beta \approx 0.5$; an example is the query “mystery magic”.

Obviously, these lessons are still very preliminary. Our future work aims at developing a principled understanding of query properties and their potential for socially-enhanced recommendation. Other issues that are worthwhile addressing include the *temporal evolution* of tagging and social relations (see, e.g., [3, 10]) and the notion of *diversity* in query results and recommendations (see, e.g., [16]). For interesting and surprising discoveries, you want to benefit from the natural diversity of cultures and tastes in your social network. (Even computer geeks should have some friends who are not in the IT business or in computer science.) Finally, efficiency and scalability in indexing and query processing pose major research challenges as well, and are being addressed in ongoing work such as [1, 20, 23].

References

[1] S. Amer-Yahia, M. Benedikt, P. Bohannon. Challenges in Searching Online Communities. *IEEE Data Eng. Bull.* 30(2), 2007.

[2] R. A. Baeza-Yates, A. Tiberi. Extracting Semantic Relations from Query Logs. KDD 2007.

[3] N. Bansal, N. Koudas. Searching the Blogosphere. WebDB 2007.

[4] S. Bao, G.-R. Xue, X. Wu, Y. Yu, B. Fei, Z. Su. Optimizing Web Search Using Social Annotations. WWW 2007.

[5] R.M. Bell, Y. Koren, C. Volinsky. Modeling Relationships at Multiple Scales to Improve Accuracy of Large Recommender Systems. KDD 2007.

[6] P. Cimiano, A. Hotho, S. Staab. Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis. J. Artif. Intell. Res. (JAIR) 24, 2005.

[7] A. Damian, W. Nejdl, R. Paiu. Peer-sensitive Objectrank - Valuing Contextual Information in Social Networks. WISE 2005.

[8] A. Das, M. Datar, A. Garg, S. Rajaram. Google News Personalization: Scalable Online Collaborative Filtering. WWW 2007.

[9] J. Diederich, W.-T. Balke. The Semantic GrowBag Algorithm: Automatically Deriving Categorization Systems. ECDL 2007.

[10] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, A. Tomkins. Visualizing Tags over Time. *ACM Transactions on the Web*, 1(2), 2007.

[11] S. Golder, B. A. Huberman. Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2), 2006.

[12] J. L. Herlocker, J. A. Konstan, L. G. Terveen, J. Riedl. Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*, 22(1), 2004.

[13] P. Heymann, G. Koutrika, H. Garcia-Molina. Can Social Bookmarking Improve Web Search? WSDM 2008.

[14] A. Hotho, R. Jaeschke, C. Schmitz, G. Stumme. Information retrieval in folksonomies: Search and ranking. ESWC 2006.

[15] K. Järvelin, J. Kekäläinen. IR Evaluation Methods for Retrieving Highly Relevant Documents. SIGIR 2000.

[16] J.A. Konstan, S.M. McNee, C.-N. Ziegler, R. Torres, N. Kapoor, J. Riedl. Lessons on Applying Automated Recommender Systems to Information-Seeking Tasks. AAAI 2006.

[17] G. Linden, B. Smith, J. York. Amazon.com Recommendations: Item-to-item Collaborative Filtering. *IEEE Internet Computing*, 7(1), 2003.

[18] S.E. Robertson, S. Walker. Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval. SIGIR 1994.

[19] J. B. Schafer, D. Frankowski, J. L. Herlocker, S. Sen. Collaborative Filtering Recommender Systems. In *The Adaptive Web*, 2007.

[20] R. Schenkel, T. Crecelius, M. Kacimi, S. Michel, T. Neumann, J.X. Parreira, G. Weikum. Efficient Top-k Querying over Social-Tagging Networks. SIGIR 2008.

[21] J. Stoyanovich, S. Amer-Yahia, C. Marlow, C. Yu. Leveraging Tagging to Model User Interests in del.icio.us. AAAI-SIP 2008.

[22] M. Theobald, R. Schenkel, G. Weikum. Efficient and Self-tuning Incremental Query Expansion for Top-k Query Processing. SIGIR 2005.

[23] M.V. Vieira, B.M. Fonseca, R. Damazio, P.B. Golgher, D. de Castro Reis, B. Ribeiro-Neto. Efficient Search Ranking in Social Networks. CIKM 2007.

[24] S. Xu, S. Bao, Y. Cao, Y. Yu. Using Social Annotations to Improve Language Model for Information Retrieval. CIKM 2007.