Exploiting Social Relations for Query Expansion and Result Ranking

Matthias Bender, Tom Crecelius, Mouna Kacimi, Sebastian Michel,
Thomas Neumann, Josiane Xavier Parreira, Ralf Schenkel, Gerhard Weikum
Max-Planck Institute for Informatics
Saarbrücken, Germany
{mbender, tcrecel, mkacimi, smichel, neumann,
jparreir, schenkel, weikum}@mpi-inf.mpg.de

***Abstract*— Online communities have recently become a popular tool for publishing and searching content, as well as for finding and connecting to other users that share common interests. The content is typically user-generated and includes, for example, personal blogs, bookmarks, and digital photos. A particularly intriguing type of content is user-generated annotations (*tags*) for content items, as these concise string descriptions allow for reasonings about the interests of the user who created the content, but also about the user who generated the annotations.**

This paper presents a framework to cast the different entities of such networks into a unified graph model representing the mutual relationships of users, content, and tags. It derives scoring functions for each of the entities and relations. We have performed an experimental evaluation on two real-world datasets (crawled from deli.cio.us and Flickr) where manual user assessments of the query result quality show that our unified graph framework delivers high-quality results on social networks.

I. Introduction

The advent of online social community platforms (e.g., Flickr, del.icio.us, MySpace, Facebook, or YouTube) has changed the way users interact with the Internet. While previously most users were mere information *consumers*, those platforms are offering an easy and hassle-free way for typical users to also publish their own content, making the users also information *producers*. As a matter of fact, it has been shown that (in particular teenage) users today spend the majority of their online time on such platforms — as of December 2006, US Internet users spent 11.9% of their total online time on MySpace [16]. On these social platforms, users are encouraged to share photos, videos, opinions, to rate content, but also to explore the online community and to find people with similar interest profiles. In this sense, online social community platforms not only change the way people interact with the Internet, but also the way users interact with each other.

While differing in the type of content that they focus on (e.g., blog entries, photos, videos, bookmarks), almost all online social community platforms work similarly. Initially, *users* must register in order to join the community. Once registered, they start to produce information, ideally by publishing their own *documents*[1] and by adding *tags* (or ratings, comments, etc) to other content already available in the community. The platforms also offer a way to maintain a list of *friends* and means to keep friends informed about your latest content items. The size of your friend network is often considered as your reputation in the network; making new "friends" often seems at least as important as publishing new content. While initially, many users populate the list of friends with people they already know from the offline world or other online communities, as time goes by they typically identify previously unknown users that they share common interests with and also add those users to the friends list.

One particularly interesting feature of these communities is the widely-used opportunity to attach manually generated annotations (so-called *tags*) to content items. In this context, tags can be considered precise descriptions of content items, flavored with the respective personal interest of the user who generated the tag. Most online communities offer comfortable and intuitive ways to explore new content items based on these tags, e.g., via tag clouds. Thus, tagging has emerged as an important asset to explore the fast-growing communities in order to identify interesting content and users.

The typically high quality of user-generated tags suggests to leverage this "wisdom of the crowds" for identifying and ranking high-quality and high-authority content. However, the existing, traditional algorithms for searching on the Web fall short of being effective in social networks, as they disregard the social component and focus on the content quality only. This makes a strong case for novel methods that exploit the different entities present in social networks (users, documents, tags) and their mutual relationships.

This paper proposes a framework to cast the different entities of such networks into a unified graph model representing the mutual relationships of users, content, and tags. It derives scoring functions for each of the entities and their relations. We have have performed an experimental evaluation of the qualitiy of these scoring functions on two real-world datasets (crawled from deli.cio.us and Flickr). Manual user assessments of the result quality suggest that our unified graph framework delivers high-quality results on social networks.

[1]For the remainder of this paper, we will use the familiar IR-jargon term "document", while in truth referring to a more general notion of content items.

II. Related Work

Despite the short time period of research interest, social networks have already been addressed in a large number of previous works. In [1], the authors present a survey of the nature of social networks together with the challenges of searching and ranking in such environments. [11] proposes *FolkRank*, an algorithm inspired by the well-known PageRank method [5] that exploits the structure of the folksonomy [15] for assigning authority scores to elements on the network, that are subsequently used to improve the result ranking. Bao et al. [2] introduce *SocialPageRank*, an authority measure for documents that evaluates the quality of a page based on its annotations, and *SocialSimRank*, which is a similarity measure for tags.

Golder and Huberman [8] provide a study of the tagging behavior and tag usage. Tag usage is also addressed in [10], where the authors propose an algorithm to generate a tag hierarchy based on a measure of tag similarity. Dmitriev et al. [7] demonstrate that explicit user tagging can help to improve precision of queries for Intranet search.

Several other works have looked at social networks from a decentralized point of view, for instance in the context of peer-to-peer networks. Some of these approaches exploit social links to propose new strategies for content searching. Bender et al. [3] propose social and spiritual routing strategies based on explicit friendship relationships and behavioral affinity. The authors studied the effectiveness of these strategies for delivering high-quality results using the social bookmark community del.icio.us. Mislove et al. [17] propose PeerSpective, a social network-enhanced Web search engine that takes into account the content that has previously been accessed by users, but may not have been indexed by search engines. Additionally, the authors discuss a ranking strategy that takes into account both the hyperlinks of the Web and the social links of the user community. [6] proposes the notion of Peer-Sensitive ObjectRank, where peers receive resources from their friends and rank them using different trust values for each peer, i.e., assigning higher score values to resources from trusted friends. Borch et al. [4] present a social P2P search infrastructure which groups peers based on the similarity of their keyword searches. The basic idea is to send queries to peers likely to have interesting resources. Khambatti et al. [13] introduce the notion of peer communities, consisting of active peers involved in sharing, communicating, and promoting common interests. These communities are self-organizing using distributed formation and discovery algorithms.

III. Network Graph Model

This section casts the entities that occur in social networks into a common unified graph model, representing the different elements of a social network and their mututal relationships. Such a graph will eventually allow well-founded query execution schemes that go far beyond ad-hoc retrieval models for social networks that often include many hard-to-tune parameters.

We identify the following three major types of entities of social networks that are to be represented by nodes in our graph model:

- **User:** people that produce content either by publishing own documents or by tagging existing content
- **Tag:** keyword used to describe a particular content item
- **Document:** content item that is published by a user (e.g., blog entry, bookmark, photo)

Additionally, social networks exhibit various relationships, both within the nodes of each type (*intra-node type*) and between nodes of different types (*inter-node type*) that are represented by edges in our graph. The idea to represent the social network as a set of ternary relations of the form *(user,tag,document)* is not representable by means of a graph. Therefore, we break this ternary relation into a set of binary relations and additionally extend the tuples to also consider scores, i.e., link weights to indicate the degree of relationship between two nodes. It is important to note that link weights can be defined in many different ways. The examples we provide in the following paragraphs present some of the alternatives, but are not meant to be exhaustive. Our model and the upcoming scoring and ranking models are independent from changes to the weighting functions.

We will go through all different types of relations in detail in the following paragraphs.

Intra Node Type Relations

Each of the three node types exhibits some sort of relation between the nodes of the same type.

- **Friends(User1, User2, Friendship):** social networks allow users to maintain a list of friends, making it straightforward to create a *Friends* relation with instances of the form that *User1* considers *User2* as a friend by the explicit act of adding *User2* to its friend list. Additionally, implicit ways of creating edges between users could be based on similar tags being issued to documents or personal content items receiving similar tags from third parties (both as indicators that such users share common interests). Thus, the Friendship strength of a link between two users can reflect the degree of trustiness, the degree of mutual interest overlap, or can simply be inversely proportional to the total number of friends of the user.
- **Similarity(Tag1, Tag2, TagSim):** users frequently use more than one tag to describe a particular document, also documents can be tagged by more than one user, and the same tag can be used more than once. There is no restriction on which terms can be used as tags, and given the diversity of the users on the network, it is often the case that different words or synonyms are used to describe the same content item. Similarity is a way of clustering the tags with respect to their meaning. The similarity measure can be derived, for instance, from the overlap in the tagged pages or tagging users, in the spirit of folksonomies. Note that these are the only undirected edges in our graph.

- **Linkage(Document1, Document2, Weight):** documents can also exhibit connections among them. In case of web pages, this linkage is obvious and is given by the hyperlink structure, and weights are often chosen proportional to the number of outlinks. For other types of documents, a different notion of link and weights needs to be defined.

Inter Node Type Relations

To this end, we would end up with separate subgraphs for users, tags, and documents. For our unified graph model, we observe the following relations between different nodes types as edges for our model:

- **Content(Document, Tag, ContentScore):** by annotating a document with a tag as a compact way to describe the document, the user strongly associates the two entities, giving a strong indication about the document's content. We consider a content score associated with a document-tag pair as an indication of how well that tag describes the document and find plenty of concrete approaches in IR literature.
- **Tagging(User, Tag, TagScore):** a tag is naturally associated with the user who associates it with a document, also as a way to describe which topic the user may likely be interested in. The TagScore can also be derived by standard IR measures, e.g., based on how many times the tag has been used by the user or how many times that tag has been used in the whole network.
- **Rating(User, Document, RatingScore):** in many social communities, users can explicitly rate documents, which is captured by a rating score. Another naive instantiation of Rating is authorship of a content item, which (e.g., in the case of bookmarks) can be seen as an endorsement for the document. Alternatively, we can also derive a rating score based on the tag usage of the user.

The result is a graph with different types of nodes, representing the different elements of the network, and links with associated weights, representing the interactions among the nodes. Figure 1 illustrates the network graph. Additional components can be considered as attributes associated with some of the edges. While our model captures all relationships that might occur in a social network, some of the above relationships may be ill-defined for certain social networks. In del.icio.us, for instance, user interactions are mainly through bookmarking and tagging, whereas in Flickr, the vast majority of users has "authored" content, which in this case are are the photos that they have published.

One might argue that tags should be considered as attributes of links between users and documents, instead of nodes on the graph, as in the social content graph suggested in [1]. The decision of considering tags as graph nodes is to represent the tag similarity, which in our view is a particularly intriguing and promising asset for a powerful searching and ranking model.

Given the network model, we can think of different ways to explore the relations in order to efficiently find relevant

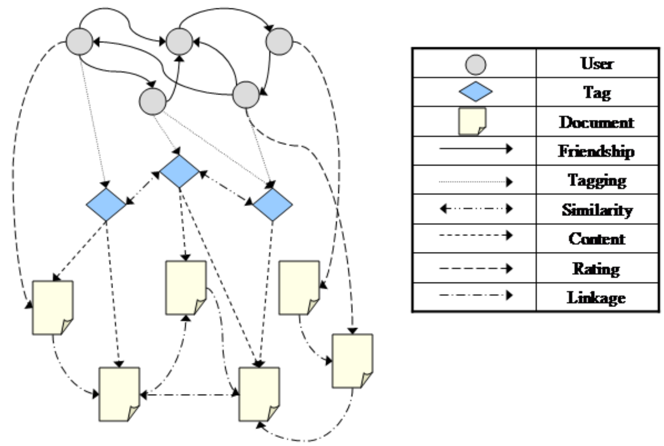

Fig. 1. Network Model

information. The graph structure allows us to model a random surfer, in the spirit of PageRank [5] or HITS [14], for finding the most important nodes on the graph. We can also consider subgraphs, for instance, a *user graph* defined by the Friendship relationship and a *document graph*, on the Linkage relationship.

IV. Scoring Model

In order to introduce our scoring model, we first formalize our notion of a query. In line with the free-text tagging of the social communities, we consider a *query*, issued by a query initiator, as a set of tags (keywords), possibly with weights for each tag (which will be disregarded for the sake of presentation simplicity). Result documents should contain at least one of the query terms and be ranked according to a query-specific document *score*. However, in contrast to the currently existing query models, our document scores also contain a social component: the content-specific score of a document is additionally *user-specific*, i.e., it depends on the context of the query initiator. Even though commercial search engines offer similar personalization approaches, social networks are the natural habitat to further explore and improve this idea.

We have developed a scoring model based on the relations in the network graph that were defined in the previous chapter. Document scores depend on the tags that have been used to annotate the document and on the users that have tagged the document. We perform *semantic expansion* by considering tags that are similar to the keywords that appear in the query, and *social expansion* by preferring documents tagged by close friends.

More formally, given a query Q

$$Q = \{t_1, \dots, t_m\}$$

the *social score* $SOS(Q, d, u)$ of a document d w.r.t. query

Q initiated by user u is defined as

$$SOS(Q,d,u) = \sum_{t_i \in Q} STS(t_i,d,u)$$

where $STS(t_i,d,u)$ is the *single tag score* of the document d w.r.t. the query term t_i and the user u who issued the query. We define STS as follows:

$$\begin{aligned} STS(t_i,d,u) &= \sum_{y \in friends(u)} UserRank(y) \\ &\times Friendship(u,y) \\ &\times \max_{t' \in sim(t_i)} TagSim(t',t_i) \\ &\times Score(y,t',d) \\ &\times DocumentRank(d) \end{aligned}$$

In the equation above, $sim(t_i)$ refers to the list of tags that are *similar* to the query tag t_i (i.e., all tags t where $TagSim(t,t_i) > 0$) and $friends(u)$ refers to the list of friends of user u. These lists are generated using the Similarity and Friends relations and their associated score attributes, respectively. Note that a document does not earn any score from the user who submitted the query; otherwise, results would be dominated by documents tagged (and therefore already known) by the user.

Being a framework, our scoring model allows different instantiations for most of its components. We now present the instantiations used in our experimental evaluation.

- $UserRank(y)$**:** The UserRank is the PageRank score defined by a a random walk on the users graph, with a random jump probability of 0.15.
- $Friendship(u,y)$**:** The Friendship measure favors users that are closer to the query initiator on the user graph, since users are likely to prefer results from their friends. We define it as

  $$Friendship(u,y) = \frac{1}{dist(u,y)^2}$$

  where $dist(u,y)$ is the distance between u and y in the user graph, i.e., the number of edges in a shortest path connecting the two users.
- $TagSim(t',t_i)$**:** The similarity between two tags is determined by the co-occurrence of the tags in the document collection. We compute the *Dice coefficient* defined as

  $$TagSim(t',t_i) = \frac{2 \times df_{t',t_i}}{df_{t'} + df_{t_i}}$$

  where df_{t',t_i} is the number of documents that contain both terms t' and t_i; $df_{t'}$ and df_{t_i} are the number of documents that have been tagged with t' and t_i, respectively.
- $Score(u,t,d)$**:** We define the score of a document d that was tagged with tag t by a user u as a user-specific BM25 score

  $$score(u,t,d) = \frac{k_1 + tf_u(t,d)}{K + tf_u(t,d)} \cdot \log \frac{N_u - ef_u(t) + 0.2}{ef_u(t) + 0.5}$$

  where

  $$K = \left((1-b) + b\frac{length_u(d)}{avg_{d' tagged by u}\{length_u(d')\}}\right)$$

  $tf_u(t,d)$ is the number of times u tagged d with t, $ef_u(t)$ is the number of times u tagged any document with t, N_u is the number of documents tagged by u, $length_u(d)$ is the number of tags given to d by u, and k_1 and b are constants, set to $k_1 = 1.2$ and $b = 0.5$.
- $DocumentRank(d)$**:** Analogous to UserRank, DocumentRank is the PageRank score of a document in the document graph.

V. Experiments

A. Data Collections and Benchmark Queries

In July 2007, we performed automated Web crawls on the social community platforms del.icio.us and Flickr and harvested the following two real-world data sets.

- **del.icio.us:** We have crawled parts of del.icio.us with a total of $13,515$ users, $4,582,773$ bookmarks, and $152,306$ friendship connections. In addition, we have crawled and indexed the actual target pages of the bookmarks.
- **Flickr:** We have crawled parts of Flickr with a total of $2,274$ users, $1,357,424$ images and $72,703$ friendship connections. A user might associate a text description to some images of his collection. This text information is not present in all image descriptions implying that we only consider tags for indexing.

We identified the 62 (for Flickr) and 74 (for del.icio.us) most frequent tag pairs that were used to describe the documents as our benchmark query workload. Typical example queries are 'landscape','nature' or 'insect','bug' for Flickr and 'cooking','recipes' or 'firefox','extension' for del.ici.ous.

B. Relevance Assessments

While many previous works have been focusing on recall (i.e., the fraction of relevant documents that is retrieved from the global set of documents), we do not consider this a useful measure for social communities. Building the ground truth, i.e., the global set of relevant documents for a query, is virtually impossible in a social network though, as the notion of relevance is highly subjective and dependent from the query initiator and her personal context. To capture this, we have conducted a manual relevance assessment user study based on the following observations that are peculiar to social search.

- The query results obtained from the system using a social search strategy depend on the query initiator, i.e., different users will obtain different results for the same query. This is obvious as the scoring model explores friendship links and tag similarities, which are user-specific.
- The relevance of a result document w.r.t. a query may be even harder to determine than in standard information retrieval. For example, a picture of a person may only be relevant if she is known to the query initiator. However, when we execute our queries in the context of a user, we

obviously do not have that user available to assess the subjective relevance of a result item.

We conducted our user study as follows. First, the participants were shown a query. Next, one particular user that has previously used the query tags was randomly selected from the database as the (fictitious) query initiator. The participants are displayed exactly those documents (i.e., pictures or bookmarks) from that user that contain at least one of the query tags, in order to understand the personal context of the query initiator. This way, we try to overcome the aforementioned problem of subjectively assessing result qualities with the eyes of the query initiator. Next, the participants are displayed a 6-column result page that illustrates the top-10 results from each of our processing strategies under evaluation. The columns are not labeled and presented in random order. For each result item, the participant has to mark as to whether the item is relevant to the query in the context of the query initiator or not. From these assessments, we finally compute the *Precision@10* as a measure of user satisfaction:

$$precision = \frac{\#\, of\, relevant\, docs\, retrieved}{total\, \#\, of\, retrieved\, docs}$$

C. Search Strategies

From the scoring framework that we have developed in Section IV, we have identified a number of interesting instances for evaluation.

Both datasets have an explicit user-to-user relationship that we leverage to create our Friendship relation, namely *Contacts* in Flickr and *your network* in del.icio.us. We obtain the user graph according to the connections between users. For del.icio.us, *your network* typically contains a list of people a user considers as the owner of good bookmarks, which can be seen as a form of endorsement for these users. On Flickr, on the other hand, the links among users mainly represent some friendship, typically disregarding the contents published by the users. Although relevant, this information might not suffice for computing authority scores for users. Therefore, we additionally enhance the user graph using the comments for user photos. More precisely, if user u_1 has commented on at least one picture from user u_2, we create an additional edge from u_1 to u_2 in the user graph. As the documents we harvested by following the bookmarks from del.icio.us did not exhibit a reasonable amount of hyperlinks among them (and as the Flickr data by nature has no such thing as inter-document links) we do not consider DocumentRank in our experiments.

1) Semantic Search: The query initiator u sends the query Q to all users SU holding similar content to the query, i.e., users who used at least one of the query tags $\{t_1, ..., t_m\}$ in describing their content. For each t_i, a retrieved result d is ranked using the BM25 score presented in section 3.2:

$$STS(t_i, d, u) = \sum_{y \in SU} Score(y, t_i, d)$$

2) Social Search: The query initiator u sends the query Q to all his friends $F(u)$ who used at least one of the query tags $\{t_1, ..., t_m\}$ in describing their content. For each t_i, a retrieved result d is ranked using BM25 and Friendship scores:

$$STS(t_i, d, u) = \sum_{y \in F(u)} Friendship(u, y) \times Score(y, t_i, d)$$

3) Expanded Semantic Search: This strategy uses a semantic search with query expansion. k possible expansion tags with the largest similarity to the original tags are added to the query Q to enrich its results. For each query tag t_i, the query initiator u, ranks a result d using BM25 and tag similarity scores:

$$STS(t_i, d, u) = \sum_{y \in SU} \max_{t' \in sim(t_i)} TagSim(t', t_i) \times Score(y, t', d)$$

4) Expanded Social Search: This strategy uses a social search where k possible expansion tags are added to the original query. For each query tag t_i, the query initiator u, ranks a result d using BM25, Friendship and tag similarity scores:

$$\begin{aligned} STS(t_i, d, u) &= \sum_{y \in F(u)} Friendship(u, y) \\ &\times \max_{t' \in sim(t_i)} TagSim(t', t_i) \times Score(y, t', d) \end{aligned}$$

5) Social Search with User Rank: The query initiator u sends the query Q to all his friends $F(u)$ who used at least one of the query tags $\{t_1, ..., t_m\}$ in describing their content. For each t_i, a retrieved result d is ranked using BM25 and Friendship scores, and User Rank:

$$\begin{aligned} STS(t_i, d, u) = \sum_{y \in F(u)} & Friendship(u, y) \times Score(y, t_i, d) \times \\ & UserRank(y) \end{aligned} \tag{1}$$

6) Expanded Social Search with User Rank: This strategy uses a social search where k possible expansion tags are added to the original query. For each query tag t_i, the query initiator u, ranks a result d using BM25, Friendship and tag similarity scores, and User Rank:

$$\begin{aligned} STS(t_i, d, u) &= \sum_{y \in F(u)} UserRank(y) \\ &\times Friendship(u, y) \\ &\times \max_{t' \in sim(t_i)} TagSim(t', t_i) \\ &\times Score(y, t', d) \\ &\times DocumentRank(d) \end{aligned}$$

D. Retrieval Effectiveness

We have used the results from our user study to compare the retrieval effectiveness for all the above strategies and for both data sets. The results are summarized in Table I. First, we note that the precision results on the Flickr dataset are constantly higher than for the del.icio.us data set. This is partly due to the fact that tags on Flickr are used much more descriptive than for del.icio.us, as one can clearly see from the queries (that, in turn, were derived from popular tag pairs). But mainly this is due to the more sophisticated context for the del.icio.us users, as presented to the participants of the user study, which made precision drop due to the better context information.

Semantic search already yields strong results, in particular for Flickr. Social expansion based on the Friendship graphs can further improve the precision remarkably, which makes a strong case for social search strategies.

On the other hand, tag expansion did not yield the desired results in our experiments, but had an negative impact. Taking a closer look at the individual numbers indicates that the Dice coefficient does not do a good job of grouping tags. In effect, tag expansion often makes the query execution drift away from the original topic.

Finally, UserRank does not seem to have any remarkable influence on the precision results. Again, taking a closer look at the individual numbers indicate that the nature of PageRank-style authority scores assigning global (i.e., non query-specific) authority scores to users makes the effect on precision vanish for particular queries. In other words, a high UserRank by design can not give an indication as to whether the authority has been accumulated w.r.t. the particular information need. This makes a strong case for considering personalized authority-style analyses that capture authority from the viewpoint of the query initiator and her personal interest profile [9], [12].

Approach	**P@10 - Flickr**	**P@10 - Delicious**
Semantic	72%	29%
Social	76%	37%
Expanded Semantic	72%	28%
Expanded Social	66%	36%
Social + UserRank	77%	33%
Expanded Social + UserRank	67%	31%

TABLE I
PRECISION@10

VI. CONCLUSIONS

Social search is a promising direction to increase user-perceived query result quality. This paper has developed a framework that represents a social community by means of a network graph model that represents the three major entities of a social network: users, documents, and tags. These entities are connected via manifold relationships that can be attributed with individual edge weights. While we have presented straightforward approaches for each of these weights, future work will have to identify the most powerful approaches to quantify the respective relations. A comprehensive scoring model combines the scores to a social score for a document, capturing user- and tag-specific contexts. Our experimental evaluation on real-world data has shown the potential of retrieval effectiveness of social search.

REFERENCES

[1] S. Amer-Yahia, M. Benedikt, and P. Bohannon. Challenges in searching online communities. *IEEE Data Eng. Bull.*, 30(2):23–31, 2007.

[2] S. Bao, G.-R. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *WWW*, pages 501–510, 2007.

[3] M. Bender, T. Crecelius, M. Kacimi, S. Michel, J. X. Parreira, and G. Weikum. Peer-to-peer information search: Semantic, social, or spiritual? *IEEE Data Eng. Bull.*, 30(2):51–60, 2007.

[4] N. Borch. Social peer-to-peer for social people. In *The International Conference on Internet Technologies and Applications*, 2005.

[5] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[6] A. Damian, W. Nejdl, and R. Paiu. Peer-sensitive objectrank - valuing contextual information in social networks. In *WISE*, pages 512–519, 2005.

[7] P. A. Dmitriev, N. Eiron, M. Fontoura, and E. J. Shekita. Using annotations in enterprise search. In *WWW*, pages 811–817, 2006.

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

[9] T. H. Haveliwala. Topic-sensitive pagerank. In *Proceedings of the Eleventh International World Wide Web Conference*, Honolulu, Hawaii, May 2002.

[10] P. Heymann and H. Garcia-Molina. Collaborative creation of communal hierarchical taxonomies in social tagging systems. Technical Report 2006-10, Stanford University, April 2006.

[11] A. Hotho, R. Jschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In Y. Sure and J. Domingue, editors, *The Semantic Web: Research and Applications*, volume 4011 of *LNAI*, pages 411–426, Heidelberg, June 2006. Springer.

[12] G. Jeh and J. Widom. Scaling personalized web search. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 271–279, New York, NY, USA, 2003. ACM Press.

[13] M. Khambatti, K. D. Ryu, and P. Dasgupta. Structuring peer-to-peer networks using interest-based communities. In *DBISP2P*, pages 48–63, 2003.

[14] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[15] A. Mathes. Folksonomies – cooperative classification and communication through shared metadata. http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html, December 2004.

[16] J. Meattle. Top-20 websites: Where do we spend our time online? http://blog.compete.com/2007/01/25/top-20-websites-ranked-by-time-spent/, January 2007.

[17] A. Mislove, K. P. Gummadi, and P. Druschel. Exploiting social networks for internet search. In *Proceedings of the 5th Workshop on Hot Topics in Networks (HotNets'06)*, 2006.