Navigational efficiency of broad vs. narrow folksonomies

Denis Helic
Knowledge Management Institute
Graz University of Technology
Graz, Austria
dhelic@iicm.org

Christian Körner
Knowledge Management Institute
Graz University of Technology
Graz, Austria
christian.koerner@tugraz.at

Michael Granitzer
Chair of Media Informatics
University of Passau
Passau, Germany
Michael.Granitzer@uni-passau.de

Markus Strohmaier
Knowledge Management Institute and Know-Center
Graz University of Technology
Graz, Austria
markus.strohmaier@tugraz.at

Christoph Trattner
Institute for Information Systems and Computer Media
Graz University of Technology
Graz, Austria
ctrattner@tugraz.at

ABSTRACT

Although many social tagging systems share a common tripartite graph structure, the collaborative processes that are generating these structures can differ significantly. For example, while resources on Delicious are usually tagged by all users who bookmark the web page `cnn.com`, photos on Flickr are usually tagged just by a single user who uploads the photo. In the literature, this distinction has been described as a distinction between *broad vs. narrow folksonomies*. This paper sets out to explore navigational differences between broad and narrow folksonomies in social hypertextual systems. We study both kinds of folksonomies on a dataset provided by Mendeley - a collaborative platform where users can annotate and organize scientific articles with tags. Our experiments suggest that broad folksonomies are more useful for navigation, and that the collaborative processes that are generating folksonomies matter qualitatively. Our findings are relevant for system designers and engineers aiming to improve the navigability of social tagging systems.

Categories and Subject Descriptors

H.5.4 [**Information Interfaces and Presentation**]: Hypertext/ Hypermedia—Navigation; H.5.3 [**Information Interfaces and Presentation**]: [Group and Organization Interfaces—Collaborative computing]

General Terms

Experimentation, Measurement, Algorithms

Keywords

Navigation, Folksonomy, Keywords, Tags

1. INTRODUCTION

In social tagging systems, users organize information using so-called *tags* – a set of freely chosen words or concepts – to annotate various resources such as web pages on Delicious, photos on Flickr, or scientific articles on BibSonomy. In addition to using tagging systems for personal organization of information, users can also socially share their annotations with each other. The information structure that emerges through such processes has been typically described as "folksonomies[1]" (**fol**k-generated ta**xonomies**). Usually, such folksonomies are represented as tripartite graphs with hyper edges. These structures contain three finite, disjoint sets which are 1) a set of users $u \in U$, 2) a set of resources $r \in R$ and 3) a set of tags $t \in T$ annotating resources R. A folksonomy as a whole is defined as the annotations $F \subseteq U \times T \times R$ (cf. [26]). A *bookmark* or *post* refers to a single resource r and all corresponding tags t of a user u.

Although this tripartite structure of folksonomies can be mapped onto a broad range of different systems in heterogeneous domains (such as Delicious, Flickr, Mendeley and others), the *collaborative processes that are generating these structures can differ significantly*. For example: While resources on Delicious are usually tagged by a larger group of users (e.g. by everybody who has bookmarked the web page `cnn.com`), photos on Flickr are usually tagged just by a single user (e.g. just by the user who has uploaded the photo). In past discussions, this distinction has been described as a distinction between *broad vs. narrow folksonomies*[2].

Thus, while broad folksonomies are structures that have been generated as a result of aggregating data from *many people tagging the same resource*, narrow folksonomies are structures that have been generated as a result of aggregating data from *single users tagging their own resources*. Although both kinds of folksonomies can be mapped onto the tripartite structure of folksonomies, it is reasonable to expect that they differ with regard to their overall network characteristics and topology, form and function. In this paper we will argue that without thorough investigations of the different characteristics of different kinds of folksonomies (e.g. broad vs. narrow), our understanding of the poten-

HT'12, June 25–28, 2012, Milwaukee, Wisconsin, USA.
ACM 978-1-4503-1335-3/12/06.

[1] `http://www.vanderwal.net/folksonomy.html`
[2] `http://personalinfocloud.com/2005/02/explaining_and_.html`

tials and limitations of social tagging systems will be limited. Therefore, understanding the usefulness and utility of different kinds of folksonomies for different tasks - such as navigation, emergent semantics or information retrieval - represents a problem of both theoretical and practical importance.

Similar classifications of metadata have been analyzed in other application areas such as learning objects metadata. In their analysis in [29] the authors distinguish between "authoritative" metadata that is provided by official data descriptors, e.g. learning object authors and "non-authoritative" metadata which emerges through the usage of learning objects in different contexts, e.g. it is created by a user community. In our terminology "authoritative" metadata corresponds to narrow folksonomies and "non-authoritative" metadata to broad folksonomies. The authors argued in their study that there are significant differences in the utility of different types of metadata. For example, they demonstrated that the "non-authoritative" metadata is crucial for effective discovery and reuse of learning objects in different contexts.

In this paper, we aim to systematically compare differences between broad and narrow folksonomies on a large tagging system (Mendeley). Mendeley is a collaborative platform for scientists where users can annotate and organize scientific articles with tags. Because Mendeley not only captures data about the set of tags assigned by users, but also about the set of keywords assigned by the authors of articles (extracted from library and metadata information), *we can generate both broad and narrow folksonomies for the same set of resources* (i.e. scientific articles) at the same time. This means that we can generate broad folksonomies based on the tags users assigned to scientific articles, *and* we can generate narrow folksonomies for the same set of resources based on the keywords that authors assigned to their papers.

In this work, we will compare the usefulness of broad vs. narrow folksonomies for a given *task*: navigation. We start by applying hierarchical clustering algorithms (such as the algorithm by [2] and others) to create hierarchies of tags and keywords as navigational structures between resources. We then use an existing framework for simulating navigation in social tagging systems [13] based on Kleinberg's decentralized search [17] to simulate a hypothetical user navigating the resource space using information provided by keywords vs. tags. In particular, we are going to model a navigational task where the user starts at an arbitrary keyword/tag and navigates to another keyword/tag to reach the list of articles with that keyword/tag. In our simulations, we adopt a greedy routing strategy based on Kleinberg's decentralized search. As a result, we use keyword/tag hierarchies as background knowledge that guides the simulation towards a particular destination by providing information on distances between keywords/tags in the resource network. To reflect the limitations of a real-world user interface, we then repeat the simulations by introducing constraints related to different user interface elements inspired by previous work [12]. The overall outcome of our investigations allows us to shed light on the differences between broad vs. narrow folksonomies in theoretical but also in practical navigation settings (by considering UI constraints). For our simulations we use a dataset that currently includes about 150 million scientific articles and has a community of about 1,5 million of users who tag articles in an unconstrained manner.

Our results suggest that both broad (tag-based) and narrow (keyword-based) folksonomies support efficient navigation in theory. However, taking some practical limitations of typical user interfaces into account, we find that broad folksonomies outperform narrow folksonomies significantly on our dataset.

In summary, this paper reports on the following findings based on our dataset:

- Narrow folksonomies create less effective navigational structures than broad folksonomies when real-world user interface constraints are considered.
- Our analysis suggests that navigational effectiveness of tags comes from the different viewpoints of readers provided through tagging resources.
- Broad folksonomies provide substantially higher quality of navigational structures than narrow ones. We speculate that with growing numbers of tags in broad folksonomies, their navigational advantage becomes even greater. More research on this question is warranted though.

The remainder of this paper is organized as follows. In Section 2, we discuss related work. In Section 3 we shortly present our simulation model for user navigation. In Section 4, we outline our experimental setup and in Section 5 we present our experimental results. In Section 6 we discuss the results and provide a possible explanation for the observed difference in navigational efficiency.

2. RELATED WORK

Related work in this field of research can be split up into two different parts: *folksonomies*, and *navigation and hierarchies in networks.*

Folksonomies: In the past, folksonomies have been studied from at least two different perspectives – from an ontological and an information retrieval perspective. From the ontological perspective, our community analyzed emergent semantic structures. For example [2, 14, 24] propose algorithms for constructing semantically sound tag hierarchies from social tagging data. A detailed analysis of approaches to semantic relatedness of tags in social tagging systems can be found in e.g. [6]. In our own previous work [20, 21], we investigated the extent to which tag semantics are influenced by user motivation and usage practices. In [31] we investigated the quality of semantic relations in automatically constructed tag hierarchies. By measuring Taxonomic Recall and Precision [9] against a huge number of existing human created concept hierarchies we have shown that algorithms such as e.g. [2] outperform other popular tag hierarchy induction approaches such as Affinity Propagation [11] or Hierarchical K-Means [10].

From the information retrieval perspective, Chi at al. [7] investigated the ability of tags to efficiently encode resources for later retrieval and found out that this ability decreases over time. In [15] and [1] the authors proposed and evaluated search ranking algorithms such as FolkRank and SocialSimilarity Rank. In our own previous work[13], we evaluated the suitability of different tag hierarchies to support navigation in social tagging systems on a theoretical level – not taking user interface constraints into account. There we showed that tag hierarchies created with algorithms such as [2, 14] are able to, at least in theory, provide an efficient support for *navigation* in tagging systems. In subsequent work, we also modeled typical limitations of a standard user interface such as e.g. *directories*, and were able to deduce a new algorithm that produces tag hierarchies that are still able to support efficient navigation even when restricted by a real-world user interface [12]. These hierarchies were evaluated by simulations with the same decentralized approach as it is also used in this paper.

Navigation and hierarchies in networks: Research on navigation in complex networks was initiated by the famous small-world experiment conducted by Milgram [27]. In that experiment randomly selected persons were required to pass a letter to a target person through their social networks. The striking result of the experiment was that the average chain length length was only six. Apart from the findings that humans in that social network are connected by short paths, another conclusion was that humans can efficiently navigate social networks although they have only *local knowledge* of that network – humans can efficiently perform *decentralized search*. Kleinberg concluded that humans possess *background knowledge* of the network structure and that this knowledge allows humans to efficiently find short paths [16, 18, 19]. Kleinberg represented such background knowledge as a hierarchy of nodes, where more similar nodes are situated closer to each other in the hierarchy.

In [30] the authors extend the notion of background knowledge to the notion of *hidden metric spaces*. In such hidden metric spaces nodes are identified by their co-ordinates – distance between nodes is their geometric distance in a particular metric space. Navigation strategies in complex networks are then based on the distances between nodes – an agent always navigates to the node with the smallest distance to a particular destination node. An interesting research question is the structure of such hidden metric spaces that underlie observable networks. In [4], the authors introduce a model with the circle as a hidden metric space and show its effects on routing in the global airport network. In [22] the authors discuss hyperbolic geometry as a hidden metric space (which can be approximated by a node hierarchy) whereas in [5] the authors apply hyperbolic geometry as a model of the hidden metric space of the Internet and design a novel greedy Internet routing algorithm. In [23] the authors describe a novel decentralized search model for efficient navigation in social networks. The model is based on the users interest. By simulating navigation on the co-author network of DBLP[3] they evaluate the model and show the importance of one step lookahead in decentralized search algorithms for social networks.

Hierarchies that are extracted from networks play an important role in many of these network navigation models. Apart from the tag hierarchy induction algorithms based on bipartite networks such as e.g. [14, 2, 12], researchers also proposed hierarchy extraction algorithms for general networks. In [28] the authors discuss an algorithm for hierarchy construction in Wikipedia networks based on metrics for estimating hierarchy level of single nodes. Also, Clauset et al. [8] present a hierarchy induction algorithm based on prediction of hierarchical links. Links prediction problem (in general settings) has been also studied by Liben-Nowell and Kleinberg [25]: They studied the extent to which interactions among members of a social network are likely to occur in the near future.

West and Leskovec [32] performed a study of user navigation behavior. The authors analyzed a collection of click paths of users playing a navigation game in a network of links between the concepts of Wikipedia. In their work they found out that user navigation behavior differs from shortest paths. For example, users typically navigate through high-degree hubs in the early phase and then apply content similarity as a criteria for reaching the destination concept.

3. METHODOLOGY

Our methodology for comparing the usefulness of broad vs. narrow folksonomies for navigation is simulation. We simulate a user who visits a digital library in search for a set of scientific articles and applies thereby a set of standard information seeking strategies. A recent study that investigated user behavior in Web search [33] showed that not many users satisfy their information need with their first search query. Instead, users visit one of the first search results, follow links on that result page, backtrack, follow some other links, then in many cases refine their search, and so on.

Thus, we model a user who starts the inquiry by issuing a search query either at an external search engine or using the integrated search function provided by the digital library. Upon selecting one of the search results the user lands at a particular page in the digital library and explores the links from that page in order to satisfy her information need. We model this first step by randomly selecting words from broad (tags) vs. narrow (keywords) folksonomies from the library. We represent the user information need as another randomly selected destination keyword together with the list of articles for which this destination keyword was assigned. We then simulate the navigation from the starting keyword to the destination keyword. In our previous work we simulated the navigation in tagging systems by simulating a user traversing links between tags from tag clouds [13] or links in a hypothetical directory-like user interface for tags [12]. The former was an assessment of the navigability of tags in an unconstrained settings whereas the latter represents a more realistic settings of a user interface that has limitations in the number of items that are presented to the user. Please note that an important advantage of simulation as an evaluation strategy is the possibility to experiment with various configurations and parameters and in this way cover a wide range of different settings – something that would not be possible in more traditional user studies. Thus, we apply the same methodology in this paper and evaluate different settings in which keywords might be used to support navigation, such as unconstrained navigation, or different variations of navigation limited by constraints of a typical user interface.

In [12, 13] we introduced a simple user navigation model – in this paper we just shortly explain its basic principles. Essentially, user navigation in information networks (such as networks of tags, or networks of keywords and scientific articles) is a kind of so-called decentralized search, or search with local knowledge of the network [16, 17, 18, 19]. At each step of navigation towards a specific destination node the user is aware only of links emanating from the current node. The user does not possess the global knowledge of the network and is therefore required to adopt a navigation

[3]`http://dblp.uni-trier.de/`

strategy that will guide her as fast as possible to the destination node. In his research on the search in social networks inspired by the famous small-world experiment by Milgram [27] Kleinberg introduced a simple greedy strategy [16, 17]. The prerequisite for this strategy is the existence of an external background knowledge on the network that defines the notion of distance or similarity between network nodes. An agent applying the greedy strategy selects from currently available links the link that leads to the most similar, i.e. to the node closest to the destination node. Kleinberg was able to show that such a greedy strategy is a very efficient one and that an agent applying that strategy always finds the destination node in a small number of steps that is bounded poly-logarithmically in the number of nodes.

Thus, we simulate user navigation by applying such a greedy strategy in search from the start to the destination node. In [12, 13] we represented the background knowledge as various tag hierarchies. Clearly, the structure of this hierarchy influences navigational capability. We assessed navigational efficiency provided by those hierarchies by measuring how often the search for the destination node is successful and if successful how fast is it. We were able to show in those papers that tag hierarchies can indeed support efficient navigation. We also designed a new algorithm that induces tag hierarchies that are efficiently navigable even under the restrictions of a realistic user interface. In this paper we apply those same algorithms on collections of keywords and scientific articles, measure the navigability of keywords and compare those results with the results that we obtained for tags on the same set of resources.

Moreover, in this paper we extend our navigation model to account for a situation where the user looks for a specific scientific article. Thus, we are not only interested in how quickly we can find keywords – we also want to know how easy it is to find a particular article once when we reach one of its keywords.

4. EXPERIMENTAL SETUP

4.1 Simulation and Evaluation Metrics

We divide our evaluation into two parts: We compare the usefulness of broad vs. narrow folksonomies by comparing their (i) encoding efficiency and (ii) navigational efficiency.

Encoding efficiency. First, we evaluate how good different folksonomical data is at encoding articles for later retrieval. This evaluation provides an insight in the intermediate exploration steps of the navigation process – the user has already reached a potentially interesting keyword or tag and the system presents a list of articles associated with that keyword or tag. We want to estimate how easy is it to find a specific article in this list. This is typically measured in terms of conditional entropy [7]. Entropy is a measure of uncertainty in a random variable. In information theory entropy is expressed in the number of bits that are needed to encode a random variable. Entropy reaches the maximal value when the random variable is distributed uniformly (uncertainty in the value of that random variable is maximal) and is minimal, i.e. it is equal to zero if the random variable always takes on a single value. Entropy of a single random variable (e.g. tags or keywords) is calculated by:

$$H(X) = -\sum_{x \in X} p(x) log(p(x)) \qquad (1)$$

In turn, conditional entropy quantifies uncertainty in one random variable (articles) once we know a specific value of another random variable (keywords or tags). Thus, conditional entropy of articles measures how difficult is to find a specific article within the presented list. Higher values of conditional entropy mean that there is more uncertainty and it is therefore more difficult to reach a particular article. On the contrary, lower values of conditional entropy mean that the first random variable (keywords or tags) encodes articles more efficiently, decrease uncertainty, and thus it is easier for users to reach a specific article. Conditional entropy of two random variables is given by:

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) log(p(y|x)) \qquad (2)$$

The navigability evaluation consists of four steps:

Network construction. We start with the datasets that include triples of keywords or tags, articles, and authors or users. From those datasets we construct bipartite networks of keywords (tags) and articles and remove the user information as that information is typically not relevant for navigation. Subsequently, we project the bipartite networks onto keyword-to-keyword and tag-to-tag networks as those networks are available for the user for navigation. We assume that article lists are also presented to the user upon selecting a keyword or a tag but only as a means of satisfying the initial information need, whereas keywords or tags are used for exploration, i.e. as a means of making progress towards the final destination.

Hierarchy construction. We induce broad (tag-based) and narrow (keyword-based) folksonomy hierarchies which we will use as the background knowledge to steer navigation. We use two algorithms for constructing hierarchies. In [14], the authors introduce a generic algorithm for producing hierarchies from bipartite networks such as tag-to-resource networks. The algorithm can be applied to arbitrary bipartite structures. The algorithm takes as input two parameters. The first is a ranked list of tags sorted by their centrality in the projected tag-to-tag network. This centrality ranking acts as a proxy to the generality ranking of tags. Benz et al. [3] showed that the centrality provides a viable approximation to the tag generalities. The second input parameter is the tag similarity matrix. The algorithm starts then by a single node hierarchy with the most general tag as the root node and then iterates through the centrality list. At each iteration step, the algorithm adds the current tag to the hierarchy as a child to its most similar tag. The centrality and similarity measure are exchangeable – in [14] the authors use closeness centrality and cosine similarity, whereas in [2] the authors select degree centrality and co-occurrence similarity measure. As both combinations perform similarly in supporting navigation [13], we select the latter combination because of better computational properties. This algorithm produces unbalanced hierarchies that are typically very broad in the top hierarchy levels. As some of the top nodes in real datasets might end up with hundreds or even thousands of children those hierarchies give us the insight in the intrinsic, theoretical, and unconstrained navigational support. We obtain a more realistic assessment

	K	T	OK	OT
Bipartite				
Metadata	1,124,260	399,703	469,952	201,651
Links	28,459,841	12,869,137	3,323,787	1,492,217
Articles	5,172,180	3,649,350	523,488	523,488
$\frac{\#Links}{\#Metadata}$	25.3	32.3	140.8	134.72
Eff.Diam	6.92	7.10	8.25	8.65
Projected Dataset				
Metadata	1,092,655	371,044	455,001	166,957
Links	124,690,988	47,760,792	26,450,686	7,877,564
$\frac{\#Links}{\#Metadata}$	114.18	128.7	58.13	47.5
Eff.Diam	4.06	3.94	4.79	4.68

Table 1: Dataset and network statistics of broad (T, OT) vs. narrow (K, OK) folksonomies. Datasets OT and OK only contain articles for which both tags and keywords are available.

of navigational efficiency by applying a variant of this algorithm. In [12], we extended that algorithm and introduced an algorithm that takes also the branching factor (the maximal number of children) of the final hierarchy as an input parameter. Through re-balancing of the hierarchy and introduction of nested misc categories we were able to produce hierarchies that support efficient navigation even under realistic limitations imposed by a typical user interface.

Search pairs selection. We randomly select one million of so-called search pairs consisting of a start node and a destination node. Both start and destination nodes are low degree nodes as searching for high degree nodes is a trivial task. For those pairs we calculate the global shortest path that we will use as our metric to assess the navigation efficiency.

Navigation simulation. We run simulation with greedy navigation on those search pairs and measure the success rate s and stretch τ which is the ratio of the number of simulator steps and the global shortest path. We calculate the global averages of both metrics (s_g and τ_g), as well as distribution of both values over the global shortest path. Also we calculate average of the global shortest path ($\bar{l}$), as well as average number of simulator hops ($\bar{h}$), i.e. average number of clicks of the simulated user.

4.2 Datasets

Mendeley[4] claims to be the largest research database, with 150 million papers and 1,5 million users. For our experiments, we used tagging data (dataset **T**) from the system gathered in September 2011 as well as a snapshot from the Mendeley system which includes papers as well as the corresponding keywords provided by the authors (dataset **K**). For dataset **T** we lowercased the tags and removed typos and personal bookmarks, i.e. tags that were used only once by a single user. Lowercasing of the keywords was also performed for dataset **K**. Furthermore we constructed an "overlapped" dataset - a dataset which includes only articles for which both keywords and tags are available. These datasets are called **OT** – overlapped tags and **OK** – overlapped keywords respectively.

Projection of the Dataset: After this preprocessing step, we construct the bipartite networks of keywords and articles and tags and articles. From those bipartite networks we extract the largest connected component (which typically contains around 99% of the network nodes). Finally, we project the largest connected component onto keyword-to-keyword and tag-to-tag networks and obtain the final networks on which we perform our analysis. The dataset and network statistics are shown in Table 1.

[4] `http://www.mendeley.com`

	K	T	OK	OT
Entropy	15.09	14.23	12.74	12.39
Cond. Entropy	6.40	5.92	4.18	3.81

Table 2: Entropy and Conditional Entropy for broad (T, OT) vs. narrow (K, OK) folksonomies. Datasets OT and OK only contain articles for which both tags and keywords are available.

The first important property here to note is that the quantitative ratio of the number of links and the number of metadata items (i.e. nodes) is comparable between the data set. The second property – the effective diameter (which is the longest shortest path that connects 90% of all network nodes) – is also comparable in all datasets. Thus, this basic *quantitative network-theoretic properties* indicate that all networks possess similar navigational properties. Hence, any differences in navigational efficiency have to be accounted for *qualitative differences in the network topology.*

5. RESULTS

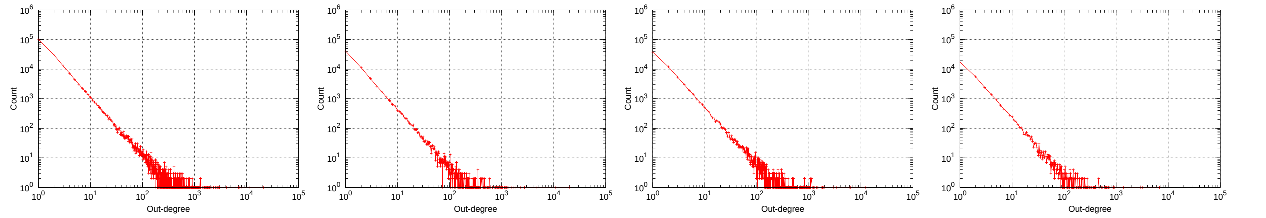
5.1 Tag and Keyword Entropy

Table 2 shows the entropy of articles conditional on keywords and tags in all four datasets. Although it is difficult to interpret absolute values obtained for the conditional entropy, a comparison of entropy values obtained for different datasets provides insight in the relative encoding efficiencies of broad vs. narrow folksonomies. From this analysis we can observe that in our dataset, broad folksonomies (T, OT) encode articles more efficiently than narrow folksonomies (K, OK). In other words, on average we know more about articles annotated by a particular tag than about articles annotated by a particular keyword. This is important when considering that users navigate for resources, not for tags. Our simulation currently does not take into account that users would have to investigate all resources attached to a particular keyword. Hence, the more uncertainty there is on the articles captured by a node, the more time users have to invest for searching the list of articles.

5.2 Unconstrained Navigation

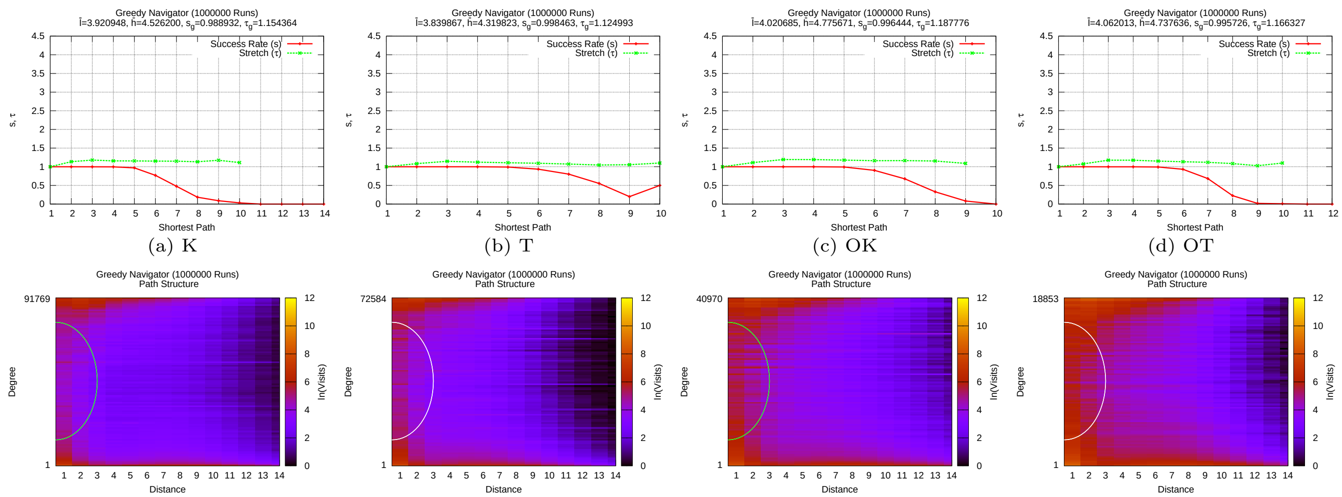
We start our navigational analysis with an estimation of the theoretical navigability of keyword and tag hierarchies. Thus, we construct hierarchies by using Heymann's algorithm [14] which does not consider any user interface constraints. The algorithm produces broad and flat hierarchies in which the nodes from the top hierarchies have hundreds or even thousands of children nodes. Figure 1 shows the degree distributions of the hierarchies depicting the existence of hub nodes.

The results of the simulation with Mendeley tags are comparable with our previous experiments with tagging datasets from Flickr, Delicious, LastFM, BibSonomy, and CiteULike [12, 13]. In such theoretical settings Mendeley tags are efficiently navigable. Keyword networks show similar behavior – in theory, keywords support efficient navigation. The complete results of the experiments are shown in Figure 2.



(a) K (b) T (c) OK (d) OT

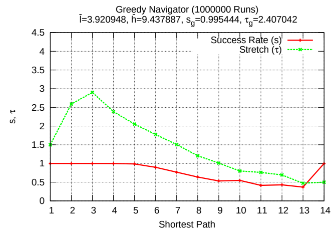
Figure 1: Out degree distribution of unconstrained hierarchies. The top hierarchy levels are populated by high-degree hubs – nodes that have hundreds or even thousands of children nodes. The hierarchies are very broad and flat.



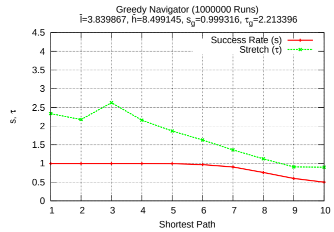
(a) K (b) T (c) OK (d) OT

(e) K (f) T (g) OK (h) OT

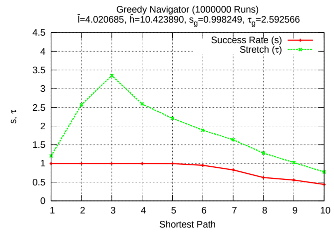
Figure 2: Results of the simulation with unconstrained user interface. ***Top:*** **Average shortest path $\bar{l}$, average hop count $\bar{h}$, greedy navigator success rate s and stretch τ – global average values (s_g and τ_g) and distribution over shortest paths. Theoretical evaluation of Mendeley tag hierarchies produces results comparable to other tagging datasets. In theory, tag hierarchies support efficient navigation – both success rate and stretch are close to 1. Similarly, keyword hierarchies aid efficient navigation – success rate and stretch are excellent.** ***Bottom:*** **Navigator path structure without user interface constraints. The density maps visualize visit frequency to nodes of a given degree at a given distance to the destination node – the color is logarithm of the visit frequency (black and violet indicating less visits; orange and yellow indicating more visits). As already observed by [4] in e.g. airport network or the Internet, the navigation path structure follows the zoom-out/zoom-in phase pattern. In the zoom-out phase, navigation starts at a low degree node in the network periphery and continues from there by visiting the nodes of increasing degrees into the network core to one the network hubs there. In the second, zoom-in phase, navigation continues over decreasing node degrees to its low-degree destination node in the periphery. Over all datasets, the top nodes are the most visited nodes – these are the nodes from the network core where the phase transition in the navigation process occurs. A specific property of navigation paths in tagging networks are so-called shortcuts between related mid-degree nodes occurring at the smaller distances to the destination node (see e.g. white marked region of a large orange-colored area in 2h). Those shortcuts are taken between sibling tags of a high-degree parent in the cases where the destination node is situated in the sub-hierarchy of one of the siblings. The density maps reveal a slightly different path structures between keyword and tag navigation. The green marked regions of shortcut areas in the keyword navigation (2e and 2g) show that shortcuts between related mid-degree and siblings are taken less frequently in the case of keyword navigation – high-degree hubs are more frequently visited in keyword than in tag networks. Since the global success rate and stretch in both networks are comparable to each other this phenomenon indicates that there exist structural differences between keyword and tag hierarchies – a possible explanation would be that tag hierarchies are somewhat richer in structure, i.e. keyword hierarchies more broad and flat. Nevertheless, in this theoretical navigational settings without any user interface constraints this does not impede the keyword navigation.**



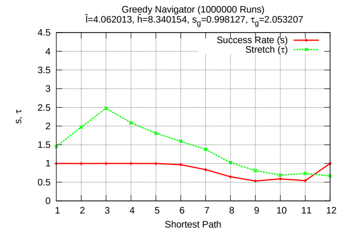
(a) K, $m = 20$



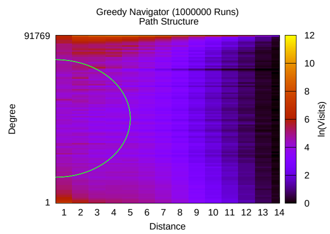
(b) T, $m = 20$



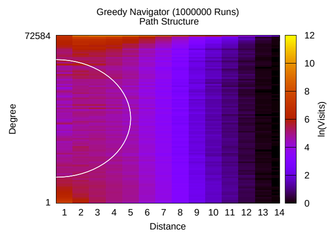
(c) OK, $m = 20$



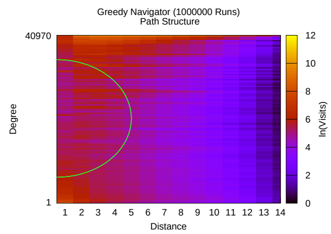
(d) OT, $m = 20$



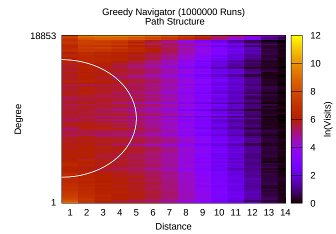
(e) K, $m = 20$



(f) T, $m = 20$



(g) OK, $m = 20$



(h) OT, $m = 20$

Figure 3: Results of the simulation with constrained user interface. The number of siblings is limited to $m = 20$. *Top:* Average shortest path $\bar{l}$, average hop count $\bar{h}$, greedy navigator success rate s and stretch τ – global average values (s_g and τ_g) and distribution over shortest paths. Although the success rates remain excellent over all datasets, stretch increases slightly in keyword datasets. This results in path lengths that are on average longer by 1 or 2 in keyword networks. *Bottom:* Path structure with user interface constraints. The green marked regions of shortcut areas in keyword networks (3e and 3g) demonstrate less frequent shortcuts than in tag networks (white regions in 3f and 3h) explaining the increased stretch values in keyword networks.

5.3 Constrained Navigation

In our next experiments we configure the simulator to reflect typical limitations of a standard user interface, e.g. a directory-like interface, such as Yahoo directory[5]. Thus, we model constraints such as limited number of children nodes that are shown (e.g. 20 children), limited number or related items (e.g. 20 siblings), or combination of both restrictions. As we have shown in [12], such restrictions seriously impede the navigation properties of tag hierarchies and we obtain similar results for both keyword and tag hierarchies. The most interesting observation that we make with those experiments is the difference in stretch values for the limitation of the number of related items that are presented to the user. In our experiments, we observe increased stretch values for keyword navigation resulting in one or two more clicks that are needed on average to reach the destination node. This result is consistent over all datasets and it might reflect an intrinsic property of keyword networks and keyword hierarchies. Our explanation for this phenomenon is that within a group of co-occurring keywords there exist a single keyword which "dominates" the group, i.e. other keywords co-occur more frequently with that "dominating" keyword and less frequently with other keywords from the group. The "dominator" becomes a parent node in the hierarchy and all other nodes are attached as children to that node (see also 6). Thus, the limitation of the number of siblings that are presented to the user causes that a longer path over the parent node is taken and increases the path length by 1 or 2 (see Figure 3).

5.4 Realistic Constrained Navigation

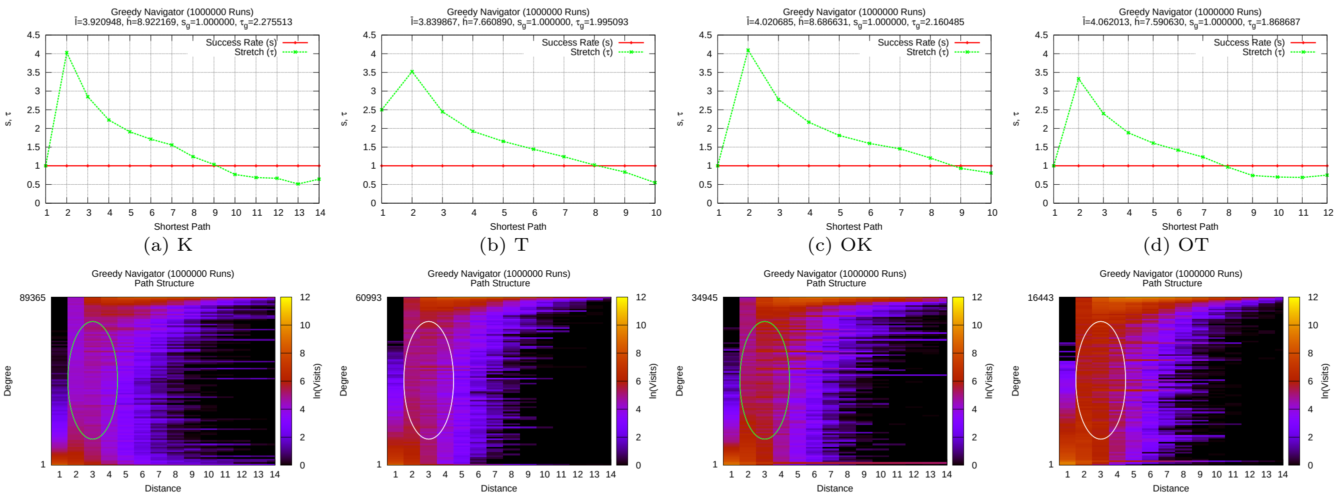
Finally, we want to perform experiments using an alternative algorithm for hierarchy induction to better reflect the realities of user interfaces. We apply the algorithm presented in [12] that produces balanced hierarchies with a maximal number of children (we set e.g. 20 children to reflect a typical user interface limitation). The algorithm produces a nested sub-hierarchy of so-called misc categories in which it inserts nodes with the smallest similarities to their parent node. In a typical case, low-degree nodes from the long tail are inserted into such nested misc categories. In our experiments, we obtain similar results as in experiments limiting the number of siblings. Consistently and over all datasets, keywords perform slightly worse exhibiting increased stretch and an increase of the average number of clicks by 1 (see Figure 4).

Finally, we remove misc categories completely to reflect another situation – a case where users might not navigate within misc categories. In those experiments we obtain smaller success rates that are comparable to each other over all datasets. As before, we observe an increased stretch in keyword networks resulting in the average number of clicks to increase by 1 in those networks (see Figure 5).

6. DISCUSSION

Our results show that in realistic navigational settings - when we take into account user interface limitations - tag navigation is slightly more efficient than keyword navigation. Moreover, tag encoding efficiency is also higher than keyword encoding efficiency. The density maps reveal the reason for this finding – there are more shortcuts taken between mid-degree and high-degree siblings in tag hierarchies than between such keywords in keyword hierarchies. A possible cause for that is a lower average overlap between sibling keywords compared to sibling tags. We will explain this situation with the following simple example. Suppose

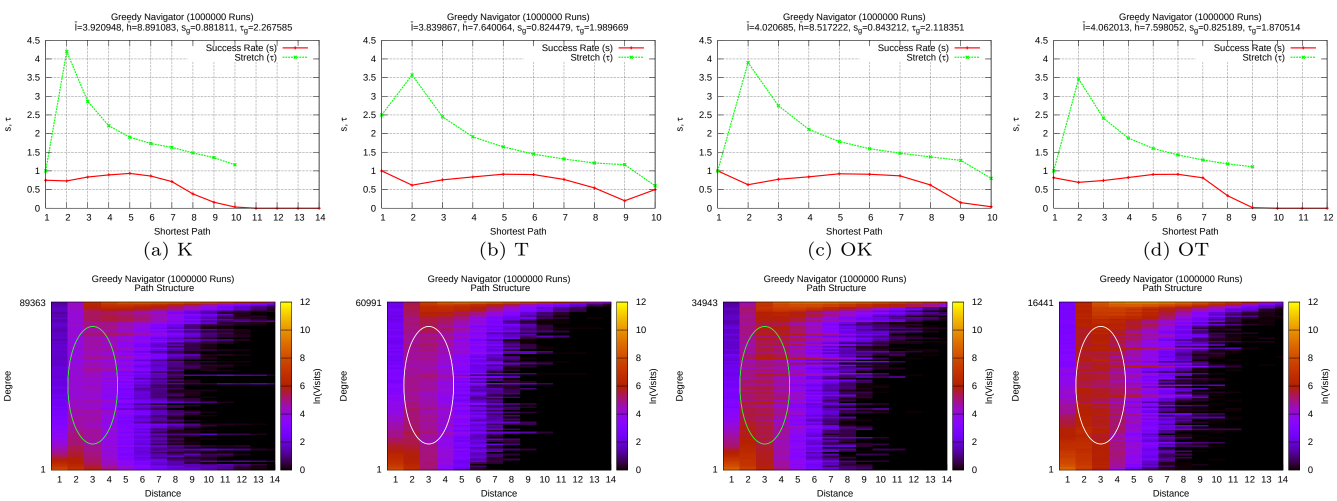
[5] http://dir.yahoo.com/



(a) K (b) T (c) OK (d) OT

(e) K (f) T (g) OK (h) OT

Figure 4: Results of the simulation with balanced hierarchies. The number of children and siblings is set to 20. *Top:* Average shortest path $\bar{l}$, average hop count $\bar{h}$, greedy navigator success rate s and stretch τ – global average values (s_g and τ_g) and distribution over shortest paths. As previously observed the success rates remain stable and excellent over all datasets, whereas stretch increases slightly in keyword datasets. This results in path lengths that are on average longer by 1 in keyword networks. *Bottom:* Navigator path structure with balanced hierarchies. Again, the green marked regions of shortcut areas in keyword navigation (4e and 4g) indicate smaller shortcut frequencies than in tag navigation (white ellipses in 4f and 4h).



(a) K (b) T (c) OK (d) OT

(e) K (f) T (g) OK (h) OT

Figure 5: Results of the simulation with balanced hierarchies without misc categories. The number of children and siblings is set to 20. *Top:* Average shortest path $\bar{l}$, average hop count $\bar{h}$, greedy navigator success rate s and stretch τ – global average values (s_g and τ_g) and distribution over shortest paths. The success rates is smaller than before over all datasets. Again, stretch increases slightly in keyword datasets. *Bottom:* Navigator path structure with balanced hierarchies. The green marked regions of shortcut areas in the keyword navigation (5e and 5g) and white marked regions in tag datasets (5f and 5h) show differences in the number of shortcuts.

we have an article dealing with navigation in tagging systems. The authors define the following keywords for this article: "`folksonomy`", "`tagging`", "`navigation`" (see $r1$ in Figure 6). Suppose also that the authors calculate entropy in that article, but do not include "`entropy`" as a keyword in their article. Thus, the authors define their single viewpoint that defines a narrow navigation structure in the proximity of that article and its keywords. Now, suppose that multiple users annotate that article with tags. For example, the first user annotates it with "`folksonomy`" and "`tagging`". The second user annotates it with "`navigation`", and the third user with "`entropy`" (because that is the most interesting

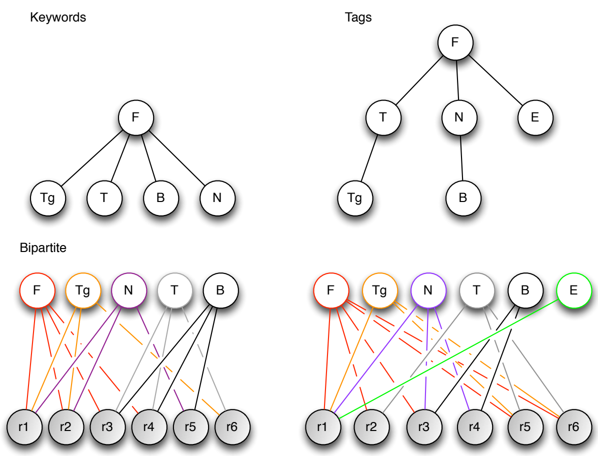


Figure 6: Two simple examples showing the emergence of hierarchies in keyword networks (left) and tag networks (right) with metadata "`folksonomy`" (F), "`tagging`" (Tg), "`tags`" (T), "`navigation`" (N), "`brwosing`" (B), and "`entropy`" (E). In keyword (narrow) folksonomies keywords are applied for grouping of articles. On contrary, in tag (broad) folksonomies tags are assigned by many users with multiple and possible alternative viewpoints. This results in tag distributions that impose richer overlap between similar tags. As a consequence the hierarchies that are based on tag generality and their mutual similarities are richer in structure than keyword hierarchies. Our experiments show that those structurally richer hierarchies are more stable and robust to the negative effects of the user interface constraints.

part of the article for that user). Now, there are multiple viewpoints on the same article and there are multiple navigational structures that are broader and *overlap* with each other. Suppose now that a user is interested in an article about entropy. Now a user may reach that article in a number of alternative ways – one of these paths leads also over our sample article as its "**`entropy`**" tag represents an entrance to a completely different cluster in the network. Thus, the user might come from a cluster related to e.g. social tagging and then upon arriving on the sample article take a *shortcut* over "**`entropy`**" tag and enter the entropy cluster. Thus, tags provide different, alternative, and more heterogeneous access paths to articles. In other words, tag folksonomies result in tag distributions whereas keyword folksonomies result in simple almost independent groups of keywords.

Moreover, such multiple viewpoints from many users tagging the same resource collection result in richer hierarchical structures – at least under the algorithms that we applied in our paper. Figure 6 depicts an example with a group of similar articles dealing with e.g. social tagging systems – the constructed hierarchies differ in their structures. Richer structures that emerge in tag hierarchies are more robust to the restrictions imposed by a user interface – less tags are affected by e.g. limiting the number of related tags as compared to more keywords that are removed when we limit the number of related keywords presented to the user. We can provide a remedy for this problem of keyword networks by e.g. enriching the keywords with additional metadata such as categorizations, or subject descriptors to turn *narrow* keyword folksonomy into a *broad* folksonomy similar to the tag folksonomy.

7. CONCLUSIONS

This paper set out to study differences between broad vs. narrow folksonomies and their usefulness for the task of navigation. Using data from Mendeley, we created both broad (based on tags provided by users) and narrow (based on keywords provided by authors) folksonomies. While our experiments show that broad and narrow folksonomies exhibit comparable quantitive properties, we find interesting qualitative differences with regard to navigation. For example, broad folksonomies create more efficient navigational structures that enable users to find target resources with fewer hops. We find that the reason for better navigational utility of broad folksonomies can be explained by the fact that greater overlap between tags provides better options for users to switch between different parts of the network. Narrow folksonomies are not able to provide this kind of support. While our findings are limited to a single dataset (Mendeley), they warrant future research in this direction. Our results are relevant for designers of social tagging systems and for engineers aiming to improve the navigability of their systems.

8. ACKNOWLEDGMENTS

This work is funded by - BMVIT - the Federal Ministry for Transport, Innovation and Technology (grant no. 829590) to the last author.

9. REFERENCES

[1] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *Proceedings of the 16th international conference on World Wide Web - WWW '07*, page 501, New York, New York, USA, May 2007. ACM Press.

[2] D. Benz, A. Hotho, and G. Stumme. Semantics made by you and me: Self-emerging ontologies can capture the diversity of shared knowledge. In *Proc. of the 2nd Web Science Conference (WebSci10)*, Raleigh, NC, USA, 2010. Web Science Trust.

[3] D. Benz, C. Körner, A. Hotho, G. Stumme, and M. Strohmaier. One tag to bind them all: Measuring term abstractness in social metadata. In *Proceedings of the 8th Extended Semantic Web Conference (ESWC 2011)*, Heraklion, Crete, May 2011.

[4] M. Boguñá, D. Krioukov, and K. C. Claffy. Navigability of complex networks. *Nature Physics*, 5:74–80, Jan. 2009.

[5] M. Boguñá, F. Papadopoulos, and D. Krioukov. Sustaining the Internet with hyperbolic mapping. *Nature Communications*, 1:62, Sept. 2010.

[6] C. Cattuto, D. Benz, A. Hotho, and G. Stumme. Semantic grounding of tag relatedness in social bookmarking systems. In *The Semantic Web – ISWC 2008, Proc. of International Semantic Web Conference 2008*, volume 5318 of *LNAI*, pages 615–631, Heidelberg, 2008. Springer.

[7] E. H. Chi and T. Mytkowicz. Understanding the efficiency of social tagging systems using information theory. In *Proc. of the nineteenth ACM conference on Hypertext and hypermedia, HT '08*, pages 81–88, New York, NY, USA, 2008. ACM.

[8] A. Clauset, C. Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008.

[9] K. Dellschaft and S. Staab. On how to perform a gold standard based evaluation of ontology learning. In *Proceedings of ISWC-2006 International Semantic Web Conference*, Athens, GA, USA, November 2006. Springer.

[10] I. Dhillon, J. Fan, and Y. Guan. Efficient clustering of very large document collections. In *Data Mining for Scientific and Engineering Applications*. Kluwer Academic Publishers, Heidelberg, 2001.

[11] B. J. J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–976, January 2007.

[12] D. Helic and M. Strohmaier. Building directories for social tagging systems. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, pages 525–534, New York, NY, USA, 2011. ACM.

[13] D. Helic, M. Strohmaier, C. Trattner, M. Muhr, and K. Lerman. Pragmatic evaluation of folksonomies. In *Proceedings of the 20th international conference on World wide web*, WWW '11, pages 417–426, New York, NY, USA, 2011. ACM.

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

[15] A. Hotho, R. J, C. Schmitz, and G. Stumme. Information Retrieval in Folksonomies : Search and Ranking. *Data Engineering*, 4011:411–426, 2006.

[16] J. Kleinberg. The small-world phenomenon: an algorithm perspective. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, STOC '00, pages 163–170, New York, NY, USA, 2000. ACM.

[17] J. Kleinberg. Complex networks and decentralized search algorithms. In *International Congress of Mathematicians (ICM)*, pages 1019–1044, Zürich, Switzerland, 2006. European Mathematical Society Publishing House.

[18] J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, August 2000.

[19] J. M. Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems (NIPS) 14*, page 2001, Cambridge, MA, USA, 2001. MIT Press.

[20] C. Körner, D. Benz, M. Strohmaier, A. Hotho, and G. Stumme. Stop thinking, start tagging - tag semantics emerge from collaborative verbosity. In *Proc. of the 19th International World Wide Web Conference (WWW 2010)*, Raleigh, NC, USA, Apr. 2010. ACM.

[21] C. Körner, R. Kern, H. P. Grahsl, and M. Strohmaier. Of categorizers and describers: An evaluation of quantitative measures for tagging motivation. In *21st ACM SIGWEB Conference on Hypertext and Hypermedia (HT 2010), Toronto, Canada, ACM*, Toronto, Canada, June 2010.

[22] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá. Hyperbolic geometry of complex networks. *Phys. Rev. E*, 82(3):036106, Sep 2010.

[23] S. Lattanzi, A. Panconesi, and D. Sivakumar. Milgram-routing in social networks. In *Proceedings of the 20th international conference on World wide web*, WWW '11, pages 725–734, New York, NY, USA, 2011. ACM.

[24] R. Li, S. Bao, Y. Yu, B. Fei, and Z. Su. Towards effective browsing of large scale social annotations. In *Proc. of the 16th international conference on World Wide Web, WWW '07*, page 952, New York, NY, USA, 2007. ACM.

[25] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *Proceedings of the twelfth international conference on Information and knowledge management*, CIKM '03, pages 556–559, New York, NY, USA, 2003. ACM.

[26] P. Mika. Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(1):5–15, March 2007.

[27] S. Milgram. The small world problem. *Psychology Today*, 1:60–67, 1967.

[28] L. Muchnik, R. Itzhack, S. Solomon, and Y. Louzoun. Self-emergence of knowledge trees: Extraction of the wikipedia hierarchies. *Phys. Rev. E*, 76:016106, Jul 2007.

[29] M. Recker and D. Wiley. A non-authoritative educational metadata ontology for filtering and recommending learning objects. *Journal of Interactive Learning Environments*, 9(3):255–271, 2003.

[30] M. A. Serrano, D. Krioukov, and M. Boguñá. Self-similarity of complex networks and hidden metric spaces. *Phys. Rev. Lett.*, 100(7):078701, Feb 2008.

[31] M. Strohmaier, D. Helic, D. Benz, C. Körner, and R. Kern. Evaluation of folksonomy induction algorithms. *ACM Trans. Intell. Syst. Technol.*, 2012.

[32] R. West and J. Leskovec. Human Wayfinding in Information Networks. In *Proceedings of the 21th international conference on World Wide Web - WWW '12*, New York, New York, USA, 2012. ACM Press.

[33] R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 21–30, New York, NY, USA, 2007. ACM.