Facilitating Knowledge Exploration in Folksonomies: Expertise Ranking by Link and Semantic Structures

Wai-Tat Fu & Wei Dong
Applied Cognitive Science Lab
University of Illinois at Urbana-Champaign
wfu/wdong@illinois.edu

***Abstract*—We developed user models of knowledge exploration in a social tagging system to test the expertise rankings generated by a link-structure method and a semantic-structure method. The link-structure method assumed a *referential* definition of expertise, in which experts were users who tagged resources that were frequently tagged by other experts; the semantic-structure method assumed a *representational* definition of expertise, in which experts were users who had better knowledge of a particular domain and were better at assigning distinctive tags associated with certain domain-specific resources. Simulations results showed that the two methods of expert identification, although based on different assumptions, were in general consistent but did show significant differences. As expected, the link-structure method was better at facilitating exploration of popular "hot" topics than the semantic-structure method. However, the semantic-structure method was better at guiding users to find less popular "cold" topics than the link-structure method. Resources tagged by domain experts could contain cold topics that were associated with high quality tags, but these resources were less likely highlighted by the link-structure method. We argue that to facilitate knowledge exploration in social tagging systems, it is important to keep a good balance between helping user to follow hot topics and to discover cold topics by including expertise rankings generated by both link and semantic structures.**

Keywords- Knowledge exploration; Exploratory search; User model; Social Tagging; Expert Identification

I. INTRODUCTION

Social tagging systems, such as del.icio.us (http://del.icio.us) and bibsonomy (http://bibsonomy.com), allow users to annotate and share their web resources using short textual labels called tags. The popularity of tagging arises from its benefits for supporting online search and information exploration. When these user-generated sets of tagged documents are aggregated, a bottom-up knowledge structure, often called a folksonomy, is formed. Many argue that folksonomies provide platforms for users in a distributed information space to share not only information, but knowledge among users, as social tags can reveal relationships among structures in the resources that others can follow [4, 8, 9, 15]. In fact, many argue that the effective finding and transfer of meaningful knowledge structures is one of the biggest challenges for future Web technologies.

Many researchers have argued that the openness of social tagging systems may result in a large number of low-quality tags that are not meaningful to other users [11]. Although many data-mining methods have been proposed to distinguish between indices and contents contributed by experts and novices, there is still a lack of systematic evaluation on how the extracted expertise could be utilized to facilitate knowledge exploration by others. The goal of this article is to test how different presentations of expert-generated indices and contents extracted by different methods can facilitate knowledge exploration in a social tagging system.

Research has shown that the definition of expertise can be *referential* or *representational*. In the referential definition, experts are individuals who are recognized and referred to by others [5]. The idea is that the more a person is being referred to, the more likely that others will follow and regard the person as an expert. Many information retrieval methods rely on this definition of expertise. Typically, the referential method identifies experts based on the "hubness" or authoritativeness of the resources by analyzing the aggregate link structures in the system [13]. In contrast, the representational definition of experts is based on performance or sometimes domain knowledge of the person [5]. In terms of information search, this implies that experts are better at describing, locating, and recognizing relevant information and knowledge. Methods based on the representational definition of expertise use formal semantic representations to extract topical structures and indices in resources [2] and to identify users who share similar semantic representations [3, 15]. This method, compared to the referential method, has the advantage of being able to measure how well users can *interpret* and *represent* different topics by tags in a social tagging system, but may not capture as much the social aspect of the definition of expertise. For example, it is possible that domain experts are better at assigning high-quality tags (e.g., more distinctive/representative) to resources, but these properties will not be extracted by the link-structure method.

II. BACKGROUND

A. Expert Identification by Link Structures

Identification of experts is traditionally performed by information retrieval techniques that exploit the structures of linked resources in the system. The general idea is to rank users and/or tags using candidate profiles developed by their associations with certain "high quality" resources. One method that is less sensitive to frequencies of appearances was the SPEAR algorithm by Au Yeung et al [1]. The algorithm identifies experts in social tagging systems by assuming that there is a mutual reinforcement between expertise and quality of the tagged resources. In other words, experts are not only better at identifying quality resources, but the ability to identify quality resources also defines the expert level of the user. In addition, it assumes that experts will be the first ones who "discover" quality resources, while novices tend to be "followers". Au Yeung et al. showed that, compared to the standard HITS and frequency-based algorithms, the SPEAR algorithm was less influenced by spammers, who tend to give many tags randomly to a wide range of resources, and thus would less likely be considered experts by the SPEAR than the traditional HITS algorithm.

B. *Expert Identification by Semantic Structures*

In addition to link structures, researchers have used resource contents to identify experts. For example, Wu et al. [16] used a probabilistic topic model to extract semantic structures from del.icio.us and applied it to discover new Web resources. The model used the hierarchical Bayesian Latent Dirichlet Allocation (LDA) model to extract topical distributions of words in the Web documents [2]. Pirolli [15] used the same model to extract topics from the community web site related to the TV series *Lost* to keep track of the growth of popular topics on the Web site. Pirolli found that topics extracted by the model matched well with the actual events happening in the TV series, providing support of the method to extract semantics from resource contents that reflect events in the real world.

III. THE SIMULATIONS

A. *The Database*

We used the database provided by Bibsonomy (January 1st of 2009), which contains 3859 users, 201,189 tags, 543,43 resources, and are connected by 1,483,767 tag assignments. We selected the most recent 6 months of tag assignments in our simulations, which contained data from 537 users, 18,278 tags, 52,098 resources, and connected by 101, 428 tag assignments.

B. *Expert identification by Link Structures*

We chose the SPEAR algorithm, which used mutual reinforcement to generate the lists of experts and quality resources. Following Au Yeung et al., the lists were represented as two vectors: E represented the vector of expertise scores of users, i.e., $E = (e_1; e_2; \ldots; e_M)$, and Q represented the vector of quality resources, i.e., $Q = (q_1; q_2; \ldots; q_N)$, where M and N were the total number of users and resources in the set respectively. Mutual reinforcement was implemented by preparing an adjacency matrix A of size M x N, where $A_{ij} = 1 + k$ if user i had assigned a tag to document j, and k users had assigned tags to document j after user i, and $A_{ij} = 0$ otherwise. Thus, if user i was the first to tag resource j, A_{ij} would be set to the total number of users who tag resource j; but if user i was the last one, then A_{ij} would be set to 1. This effect of this was to create a bias to those users who discovered quality resources. Following Noll et al., in order to balance the impact of the discovery and hubness effect, the value of A_{ij} was adjusted by the square root function, such that $A_{ij}' = \sqrt{A_{ij}}$. Based on this adjacency matrix, the calculations of expertise and quality scores followed an iterative process. Specifically, E and Q were updated as:

$$E' = Q \times AT \text{ and } Q' = E \times A \qquad (1)$$

The final lists of E and Q would represent the expertise and quality scores of the users and resources, which could be sorted to identify the top experts and resources in the system.

C. *Identifying semantic structures*

To study the differences in the semantics structures of the resources tagged by experts and non-experts, we extracted topics from the resources using the LDA model [2]. However, because topic extraction is computationally extensive and many resources were tagged by very few users (a typical power-law distribution), we selected samples of the resources to focus on the relations between link and semantic structures. We first selected the top 5,000 quality resources identified by the SPEAR algorithm, and then randomly sampled another 5,000 resources from the rest of the resources. We called these the *high-quality* and *low-quality* sets of resources. Because the SPEAR algorithm calculates quality based on how often the resources are tagged, high quality resources in our set represented mostly "popular" resources, while the low quality resources represented the "regular" resources. Topic distributions in these two sets of resources will therefore reveal how likely one can find certain kinds of topics by following popular (and unpopular) resources.

We also identified the first 50 experts from the SPEAR algorithm, and then randomly sampled another 50 users in the rest of the dataset. We extracted the resources tagged by these experts and non-experts to form the *expert* and *non-expert* sets of resources. We then processed the HTML files based on the URLs of the resources in the database. We filtered out any non-English pages and pages that contained fewer than 50 words, and eventually obtained 5000 usable resources from each of the four sets. We performed the topic extraction algorithm derived from the standard LDA model on each set of resources.

We were interested in the quality of tags assigned by experts and non-experts and those on low and high quality resources in terms of their ability to serve as good navigational cues to the users. To measure this, we assumed that users would adopt a tag-based topic inference process [9, 10] during exploration, which allowed them to predict whether the tagged resource would contain topics that they were interested in. This value could be calculated by the posterior probability $p(c_j|tags)$, where c_j is the topic of interest. For the current purpose, it was useful to compare the predictive distribution of tags in each set of resources to compare the usefulness of the tags. This empirical distribution could be derived from the LDA model (see [12]). Another useful measure for understanding the semantic structures in the different sets of resources is to compare the predictive distributions of *topics* in the different set of resources, which again can be estimated from the LDA model. Comparing the predictive distributions of topics between the set of resources would therefore show how the distributions of popular or "hot" topics would correlate with the experts and quality resources identified by the SPEAR algorithm.

D. *Simulating exploratory search in social tagging systems*

In simple fact retrievals, search efficiency can be measured by how easily one could reach a specific target resource; but in exploratory search, one is more likely navigating in the system to browse and explore for topics of interest rather than to simply reach a specific resource [14]. Previous research has shown that exploratory search is often guided by the semantic interpretation of tags to infer the possible topics in the resources [10]. Therefore, we decided to select a topic of interest and simulate exploratory search by model-searchers that interpreted social tags during navigation, and measured performance by counting the number of resources that were "picked" up by the model-searcher as relevant resources.

To understand how different presentations of expertise rankings would impact exploratory search, we created search environments that utilized different metrics to present the lists of tags, users, and resources and tested how they would influence exploratory search performance. In an actual search, this assumption of the search processes and environments might be over-simplistic (see e.g., [7]), However, given that the current goal is to understand the general computational characteristics of the systems in terms of their efficiencies in presenting knowledge in the folksonomies to facilitate exploratory search, we decided to abstract away certain interface design details to highlight the effects of general presentation methods in a social tagging system on exploratory search performance.

To simulate exploratory search, we first randomly pick a topic (represented as a topic-word distribution) and a random tag, and calculate average performance of the model-searcher in different search environments through repeated simulations. The model-searcher would navigate in the folksonomies and collect resources that were relevant to the topic of interest. Topical relevance of a resource was calculated by Kullback–Leibler (KL) divergence between the desired and the best matching topic-word distribution in the resource. If the KL divergence reached the threshold, the resource was considered relevant. To measure exploratory search performance, we limited each model-searcher to perform 1000 steps of "clicking" to count the number of relevant resources it could find. In other words, we assumed that the average number of relevant resources found within 1000 transitions in the hypergraph formed by the folksonomy reflected how well the environment could support exploratory search conducted by the model-searcher. We also randomly selected the topic of interest based on the predictive probabilities of topics, such that half of the simulations were looking for popular topics, the other half were looking for less popular topics. We could then compare how different model-searchers would perform differently when searching for popular or less popular topics.

1) The model-searchers

We simulated three types of model-searchers: The *position-searcher*, *topic-satisficer,* and *perfect-searcher*. All model-searchers would start with a tag, and select a resource to evaluate whether it was relevant. It would then select another tag assigned to the resource (or user), and so on; however, the model-searcher would never repeat the same click (i.e., cannot pick the same hyperedge and resource). Depending on the type of the model-searcher, either the presentation order and/or the topical relevance of the nodes would influence exploratory search performance. We will elaborate on this next.

Position-searcher: For the position-searcher, selection of the next node was determined based on the presentation order of the connected nodes. In other words, the position-searcher would always pick the first unselected node on the list. Performance of the position-searcher would therefore indicate how presentation orders of nodes in different search environments would influence exploratory search performance.

Topic-satisficer: For the topic-satisficer, selection of the next node was determined by both the presentation order and the topical relevance of the neighboring nodes. Following previous computational cognitive models of information search [6, 7], the selection was based on a satisficing mechanism that aimed at matching actual human navigational behavior. We assumed that there were separate lists of resources, tags, and users, and the topic-satisficer would first random pick one of the lists and sequentially evaluate the node in the list. The topic-satisficer would stochastically choose one of the three possible actions: (1) select a node, (2) evaluate the next node, and (3) leave the list. Specifically, if S_k represented the topical relevance of k^{th} node in the list, the utilities of the actions at the n^{th} cycle were updated by the following equations:

$$U_{select}(n) = \frac{U_{select}(n-1) + \max(S_1, S_2, ..., S_k)}{1 + C_1 + k}$$
$$U_{evaluate}(n) = \frac{U_{evaluate}(n-1) + S_k}{1 + k}$$
$$U_{leave}(n) = C_2 - (\sum_{j=1}^{k} S_j) / k \qquad (2)$$

In (6), C_1 and C_2 were constant scaling parameters and were set to 5 and 1 respectively (same as in previous models). The probability for selection one of the three actions U_i was calculated by the softmax equation.

As (6) shows, when the topic-satisficer first evaluated a list, it would be biased to evaluate (because of C_1) the next node on the list. However, as k (and n) increased, the best node evaluated so far would become more likely to be selected (i.e., it satisficed on a node). Therefore, in general, nodes that had higher topical relevance would more likely be selected. However, it was possible that the topic-satisficer would select a node located at an earlier part of the list, even when the node had a low topical relevance than another node at the later part of the list. In addition, as k increased, the probability that the topic-satisficer leaving the current list would increase. Thus, in cases where nodes were estimated to have low topical relevance, the topic-satisficer would more likely switch to another list. This satisficing process was found to describe the human sequential evaluation process well, as it was shown that human searchers were sensitive to both presentation orders and information relevance when navigating in information spaces. Performance of the topic-satisficer therefore reflected how presentation orders would interact with topical relevance of nodes during exploratory search, and it also provided a metric for predicting how human users would perform in the environments.

Perfect-searcher: The perfect-searcher had "perfect" knowledge in predicting which neighboring node would have the highest chance of leading to the resource that contained the topic of interest. Specifically, the perfect-searcher would select the next node that had the highest topical relevance among the neighboring nodes (therefore it was not really perfect because it could not look ahead). The calculation of topical relevance depends on whether the node was a resource, a tag, or a user. When evaluating a resource node, the same KL divergence similarity measure between topic-word distributions of the target and the resource would be used to estimate its topical relevance. When evaluating a tag node w_i, it would calculate the probability $p(z|w_i)$ that the tag would lead to the topic of interest z, which could be estimated by summing over the topic-word distributions over all resources. Finally, when evaluating a user node, the probability $p(z|u_i)$ that the user u_i would lead to resources that contained the target topic would be calculated. This probability could be calculated by summing over the topic-word distributions over all resources tagged by u_i. Performance of the topic-searcher would not be influenced by the presentation orders of nodes in different search environments, but would have "perfect" knowledge to pick the best node among the neighboring nodes.

2) The search environments

We simulated exploratory search performance in different search environments to compare how changes in the interface representations may impact efficiencies of exploratory search. Each of these interface representations can be implemented in multiple ways, but the current purpose is to reveal the computational characteristics underlying the folksonomies in terms of general exploratory search efficiencies. The following search environments were simulated:

- In the *no-ranking* environment, tags, users, and resources were randomly presented in each cycle of search. In other words, the probabilities of visiting any of the neighboring nodes in the hypergraph were equal. This environment would serve as the base-line performance of exploratory search in an "unstructured" interface.
- In the *resource-ranking* and *expert-ranking* environments, after the model-searcher clicked on a node (tag, user, or resource), the list of resources or experts, respectively, associated with the node would be ranked based on their quality scores from the SPEAR algorithm.
- In the *tag-topic-ranking* environment, the list of tags associated with the current node would be ranked. The ranking was implemented by calculating the conditional probabilities of the tag given that the model-searcher was in a particular node. If the model-searcher was in a tag node (that it just clicked on a tag in the last search cycle), the ranking of tags would be determined by $p(w_i|w')$, where w' is the current tag, and w_i is the i^{th} tag node that connects to w'. This conditional probability could be calculated by summing over all topics *z* that the tags were associated with. When the model-searcher is in a user node, the summation of topics would be all topics associated with the resources that the user tagged; when in the resource node, the summation of topics would be over all topics in the resource. The probabilities would then be used to rank the list of nodes and presented to the model-searcher during the simulation. *Note that in this environment, ranking was based on the tag quality derived from the semantic structures extracted from the resources, while in the resource- and expert-ranking environments, rankings were derived from link structures in the system derived using the referential expert identification method*.
- In the *tag-expert*, *tag-resource*, and *resource-expert* environments, the lists of tags and experts, tags and resources, and resources and experts were ranked based on their corresponding ranking scores as described above, but different pairs of lists were combined and presented in each of the environments. In the *all-ranking* environment, the lists of tags, resources, and users would all be ranked according to the corresponding calculations discussed earlier.

IV. Results

A. Expertise rankings as navigational cues

1) Topic distributions

Figure 2 shows the mean predictive probabilities of topics (top) and tags (bottom) against the ranked list of resources in the expert and non-expert sets of resources (left) and against the ranked list of experts in the high- and low-quality sets of resources (right). These probability distributions show that the semantic structures in each set of the resources identified by the referential method had very different properties. In terms of the predictive probabilities of topics, low-ranking resources found by experts tended to contain more "popular" topics than those found by non-experts, but this difference seemed to diminish quickly as the resource quality rank increased (top-left of Figure 2). On the other hand, comparing the sets of high and low quality resources identified by the referential method, there seemed to be consistently more popular topics in the high

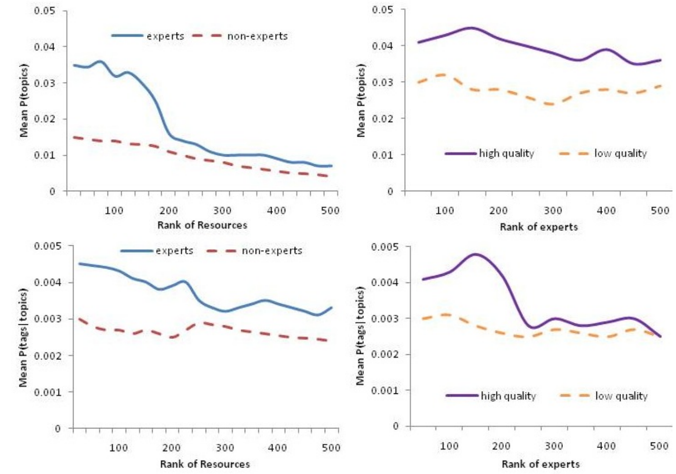

Figure 2. Predictive distributions of topics (top) and tags (bottom) plotted against ranks of resources and experts (lower rank is better) in the sets of resources tagged by experts and non-experts (left) as identified by the algorithm, and the sets of resources identified as high and low quality by the algorithm (right).

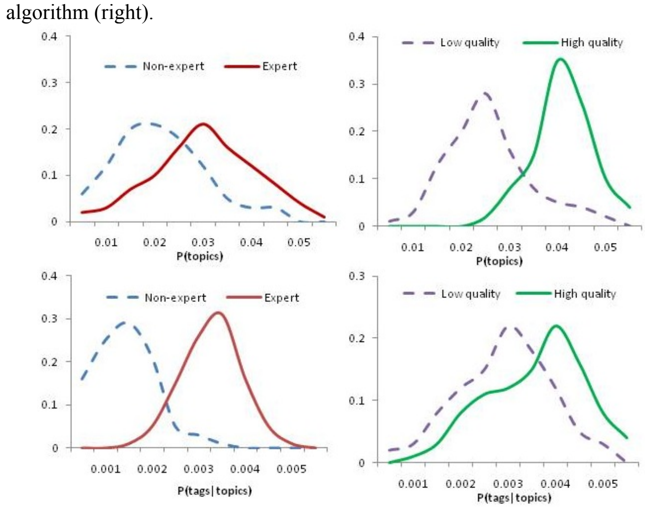

Figure 2. The empirical PDF for the predictive probabilities of topics and tags in each of the four sets of resources.

quality resources than low equality resources, and this difference seemed to be relatively insensitive to the rank of experts within the sets (top-right of Figure 2).

The predictive probabilities of tags also showed interesting difference between the sets of resources. In general, resources tagged by experts contained more predictive tags compared to those by non-experts, and this difference seemed to be relatively stable across the set of resources (bottom-left of Figure 2). Similarly, the predictive probabilities of tags between the high and low quality sets of resources also showed differences, but this difference diminished quickly as the expert rank increased (bottom-right of Figure 2).

The patterns of results shown in Figure 2 suggest that, as expected, the experts and quality resources extracted based on the link structures in the folksonomies were useful in revealing popular topics among the resources. However, the referential method did show some interesting patterns between the set of resources tagged by experts and those ranked as high quality by the algorithm. As shown in the top two graphs in Figure 2, while high quality resources tend to be associated with popular topics, experts (as identified by the referential method) did not always

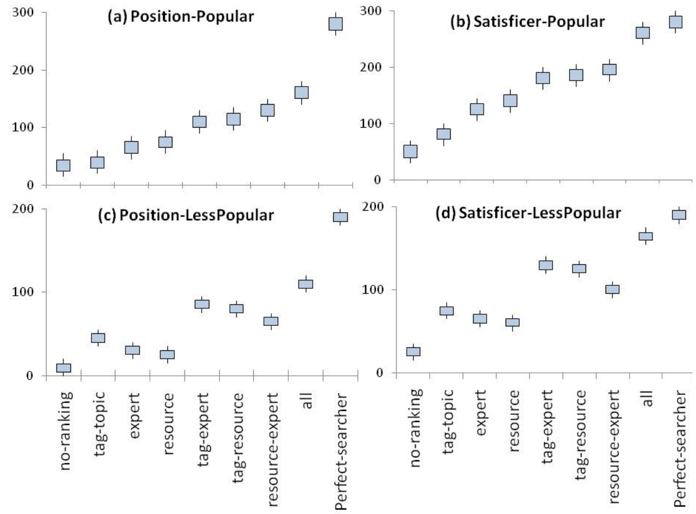

Figure 3. Exploratory search performance of the position-searcher (left) and the topic-satisficer (right) when the topics were popular (top) and not. (bottom).

tagged resources that contained popular topics. In terms of exploratory search, this difference implied that while following the list of high quality resources would allow users to discover mostly "hot" topics, following the list of experts would allow users to sometimes discover "cold" topics (but could be useful for a subgroup of users) in the folksonomies.

2) Tag Distributions

Figure **2** (bottom) also shows that tags created by experts and those assigned to quality resources were in general predictive of the topics in the resources, suggesting that they were potentially useful as navigational cues to help users to infer what topics could be found in the tagged resources (more results in the next subsection). However, while tags created by experts were in general predictive, not all tags in the high quality resources were predictive. The results suggest that while tags created by experts were in general useful for inferring topics in the resources, tags in high quality resources were not necessarily useful as navigational cues during exploratory search of knowledge.

To confirm these differences, we compared the empirical probability distribution functions (PDF) of the predictive probabilities of topics and tags in each sets of resources (see Figure **2**). One could see that the topic distributions between experts and non-experts were indeed less distinguishable than those between low and high quality resources (top-left of Figure **2**), but the reverse was true for the tag distributions (bottom-left of Figure **2**). Indeed, only the distributions between high and low quality resources in P(topics) and those between experts and non-experts in P(tag|Topics) were significantly different ($p<0.05$), while the other two pairs of distributions were not significantly different. This suggested that quality of resources were in general better at predicting "hot" topics; but high quality resources did not necessarily contain fewer "hot" topics. Rather, expert-generated tags tended to be more predictive of "cold" topics than resource quality. For example, resources tagged by a focused group of domain experts could contain cold topics that were associated with high quality tags, but these resources were less likely picked up by the link-structure method.

B. Exploratory search performance

1) Exploring for "hot" topics

Figure 3 shows the number of relevant resources found by the three model-searchers in the exploratory search simulations. When searching for hot topics, the general uptrend from all four figures suggested that the addition of more rankings of tags, experts, and resources had led to better exploratory performance. Interestingly, for popular topics, rankings of users and resources seemed to lead to slightly better results than the ranking of tags. Consistent with previous results, this could be attributed to the fact that the rankings of users and resources were based on the referential method that were in general better at predicting hot topics. In contrast, ranking of tags was based on their predictability of topics, which depended, to a large extent, on the likelihood that the tags were uniquely associated with the different topics. Given that hot topics tended to be associated with semantically general tags in multiple topics [4], the general predictability of tags for popular topics were therefore lower than the rankings derived from link structures.

Compared to the position-searcher, the topic-satisficer in general found more relevant resources, suggesting that the process of sequential topic evaluation improved exploratory search performance. Similar to the position-searcher, however, tag ranking was slightly less useful for exploring for hot topics compared to expert and resource rankings. On the other hand, the combination of tag and expert or tag and resource rankings did significantly improve performance. As shown in the graph in the top-right of Figure 3, performance of the topic-satisficer in the all-ranking environment was almost the same as the perfect-searcher, which always picked the most predictive nodes in every transition. The results suggested that when exploring for hot topics, ranking of tags, users, and resources were useful in guiding users to explore for popular topics in the folksonomies.

2) Exploring for "cold" topics

When the model-searchers explored for less popular topics, the results were similar but did show differences (bottom of Figure 3) when compared to exploration for hot topics. In particular, while the addition of rankings improved performance, but for both model-searchers, *tag ranking* in general led to better exploratory performance than expert and resource rankings. This could be attributed to the fact that for cold topics, the predictive power of tags was higher than that by expert and resource rankings (see top two graphs of Figure 2). Following the tags therefore led to a higher chance of discovering cold topics than by following expert and resource rankings.

In summary, the patterns of simulation results showed not only that expertise rankings could improve exploratory performance, but it also showed that different expert identification methods could have systematic differences in their influence on exploratory performance. In particular, we found that while expert and resource rankings based on the referential method could facilitate exploration of hot topics, ranking of tags based on the probabilistic topic extraction method could facilitate exploration of cold topics. In our simulations, a combination of the two methods seemed to lead to the best overall result in providing effective navigational cues that facilitate knowledge exploration.

V. Conclusion and General Discussion

Although a significant amount of work has been done to improve methods of expert identification, the link between expertise rankings and knowledge exploration has not yet been systematically studied. The potential for social tagging systems to support knowledge exploration cannot be underestimated. For example, students may utilize a social tagging system to learn collaboratively and to share and structure information to facilitate knowledge discovery and creativity. Scientists may utilize the system to improve knowledge sharing and facilitate knowledge transfer across disciplines, and the interactions may potentially promote idea generation and integration.

The current sets of results provide support to the promising aspect of using social tagging system to facilitate knowledge exploration and sharing among users. Specifically, our results showed that (1) the method based on the referential definition of expertise was more useful for generating rankings that facilitate search for popular topics, while the method based on the representational definition of expertise was more useful for generating rankings that facilitate search for less popular topics, (2) rankings of tags based on their predictive probabilities of topics could facilitate search of less popular topics, and (3) combinations of referential and representational methods of expert identification could facilitate knowledge exploration of both popular and less popular topics.

In the knowledge exploration simulations, we assumed that the topic-satisficer was able to infer topic relevance based on the predictive probabilities. In reality, users could be less sophisticated in inferring topic relevance based on the tags, and could thus perform worse than the topic-satisficer. The position-searcher, on the other hand, relied solely on the rankings and did not select nodes based on topic inference. The exploration search performance of real users would therefore likely fall somewhere between the position-searcher and the topic-satisficer. However, the results were in general consistent with the idea that, users who were better at inferring topics based on the tags (e.g., domain experts, see [9]) were more likely to perform better during knowledge exploration.

Our simulations were performed using an existing data set from Bibsonomy, but one should note that, the growth of the folksonomy in a social tagging system is a dynamic process: The exploratory search process and the knowledge structures in the folksonomy will likely influence each other. It is likely that, for example, different interface presentations would influence how people browse and search in the social tagging system, which will influence what resources they will find and tag, and thus collectively influence the development of the knowledge structures in the folksonomies. For example, interface that provides expertise rankings using the referential method may lead to more visibility of popular resources, which may further attract both more users to browse and tag these resources. These “biases” provided by the interface presentations would therefore influence not only exploratory search behavior but also the patterns of tags assigned by users, which will slowly influence the development of inherent structures of the folksonomy. Computational models that capture the dynamics between the use and growth of folksonomies would therefore be critical in predicting the emergent properties of the system.

We have shown how data-mining methods can be combined with mechanistic models of exploratory search to understand how different interface representations and interaction methods could impact overall utilities of the system. Our results highlight the importance of including *user models* to evaluate the functional utilities of structures extracted from different data-mining methods. In general, we believe that it is useful for researchers of social computing to investigate the dynamics between how individual users will actually *utilize* information cues or structures extracted from a social information system, and how they would in turn influence the computational properties of the system itself.

References

[1] C.-M. Au-Yeung, M. G. Noll, N. Gibbins *et al.*, “SPEAR: Spamming-resistant Expertise Analysis and Ranking in Collaborative Tagging Systems,” *International Journal of Computational Intelligence*, 2010.

[2] D. Blei, A. Ng, and M. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research,* vol. 3, pp. 993-1022, 2003.

[3] C. S. Campbell, P. P. Maglio, A. Cozzi *et al.*, "Expertise identification using email communications," *Twelfth international conference on Information and knowledge management*, New Orleans, LA, 2003.

[4] C. Cattuto, V. Loreto, and L. Pietronero, “Semiotic dynamics and collaborative tagging,” *Proceedings of National Academy of Sciences,* vol. 104, pp. 1461-1464, 2007.

[5] P. J. Feltovich, M. J. Prietula, and K. A. Ericsson, "Studies of expertise from psychological perspectives," *Cambridge handbook of expertise and expert performance*, K. A. Ericsson, N. Charness, P. Feltovich *et al.*, eds., pp. 39-68, Cambridge, UK: Cambridge University Press, 2006.

[6] W.-T. Fu, and W. Gray, “Suboptimal tradeoffs in information seeking,” *Cognitive Psychology,* vol. 52, no. 3, pp. 195-242, 2006.

[7] W.-T. Fu, and P. Pirolli, “SNIF-ACT: A cognitive model of user navigation on the World Wide Web,” *Human-Computer Interaction,* vol. 22, pp. 355-412, 2007.

[8] W.-T. Fu, "The microstructures of social tagging: A rational model," *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, pp. 229-238, San Diego, CA, 2008.

[9] W.-T. Fu, T. G. Kannampallil, and R. Kang, "A Semantic Imitation Model of Social Tag Choices. ," *Proceedings of the IEEE conference on Social Computing*, pp. 66-72, Vancouver, BC, 2009.

[10] W.-T. Fu, T. G. Kannampallil, R. Kang *et al.*, “Semantic Imitation in Social Tagging,” *ACM Transactions on Computer-Human Interaction*, in press.

[11] G. W. Furnas, T. K. Landauer, L. M. Gomez *et al.*, “The vocabulary problem in human-system communication,” *Communications of the ACM,* vol. 30, no. 11, pp. 964-971, 1987.

[12] T. Griffiths, and M. Steyvers, “Finding Scientific Topics,” *Proceedings of the National Academy of Sciences,* vol. 101, pp. 5228-5235, 2004.

[13] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” *Journal of the ACM,* vol. 46, no. 5, pp. 604-632, 1999.

[14] G. Marchionini, “Exploratory search: from finding to understanding,” *Commun. ACM,* vol. 49, no. 4, pp. 41-46, 2006.

[15] P. Pirolli, “A Probabilistic Model of Semantics in Social Information Foraging,” in AAAI Spring Symposium, Chicago, 2008, pp. 72-77.

[16] X. Wu, L. Zhang, and Y. Yu, "Exploring social annotations for the semantic web." pp. 417-426.