Incorporating Participant Reputation in Community-driven Question Answering Systems

Liangjie Hong, Zaihan Yang and Brian D. Davison
Department of Computer Science and Engineering
Lehigh University, Bethlehem, PA 18015 USA
Email: {lih307, zay206, davison}@cse.lehigh.edu

Abstract—Community-driven Question Answering services are gaining increasing attention with tens of millions of users and hundreds of millions of posts in recent years. Due to its size, there is a need for users to be able to search these large question answer archives and retrieve high quality content. Research work shows that user reputation modeling makes a contribution when incorporated with relevance models. However, the effectiveness of different link analysis approaches and how to embed topical information—as a user may have different expertise in various areas—are still open questions. In this work, we address these two research questions by first reviewing different link analysis schemes—especially discussing the use of PageRank-based methods since they are less commonly utilized in user reputation modeling. We also introduce Topical Page-Rank analysis for modeling user reputation on different topics. Comparative experimental results on data from Yahoo! Answers show that PageRank-based approaches are more effective than HITS-like schemes and other heuristics, and that topical link analysis can improve performance.

I. Introduction

Community-driven Question Answering (CQA) has existed for decades as part of bulletin board systems and Usenet, but has recently been popularized within web portals in which users answer questions posed by other users. CQA has proven to be more effective since users can post natural language questions rather than issuing several word queries to search engines. One typical example of a CQA system is Yahoo! Answers, which already attracts tens of millions of users and stores hundreds of millions of questions and answers [1]. Unfortunately, users may post similar or identical questions multiple times and the quality of answers varies drastically. Recent work [2] shows that a large portion of content in CQA is not useful for users. On one hand, it is not appropriate for users to re-post existing questions. On the other hand, users may find it difficult to browse within the large question-answer archive. Therefore, there is of increasing interest to build retrieval mechanisms to automatically search in a question-answer archive and provide high quality content to users.

Not surprisingly, much research work (e.g., [3], [4]) has shown that reputation of users are good indicators of the quality and reliability of the content. Many ranking schemes which take advantage of user reputation (e.g., [5], [6], [3]) have been proposed to provide search results to users. The assumption behind these methods is that highly authoritative users may provide high quality content. Since the naturally bipartite structure of CQA where users who post questions

and users who provide answers can be seen as two sub-communities within CQA, several ranking approaches (e.g., [5], [6]) derived from the HITS algorithm [7] have been shown to improve retrieval performance. However, there is no evidence to show whether this is the most effective way to model users' expertise. In addition, PageRank-like ranking schemes are less used to model reputation in a CQA context due to several possible reasons. One of them is that it is relatively difficult to see whether CQA has the "hierarchical ranking structure" that PageRank provides where the reputation of users depend not only on the number of questions and answers a participant produces but also on who the user interacts with, compared to naturally bipartite structure of HITS.

In this paper, we discuss how to use PageRank to model user reputation in CQA. We view the link between users as reflecting the likelihood of one user providing an answer to the other. In addition, we introduce topical link analysis [8], which has shown success in modeling web page authority into CQA and show how to incorporate topical information. Our specific contributions include: 1) The use and justification of a PageRank-based method for user reputation modeling in CQA; 2) The introduction of topical link analysis for user reputation modeling in CQA. The method does not use any site-specific features and can be easily applied to other social media; 3) Showing how probabilistic Latent Semantic Analysis (pLSA) can be embedded into user reputation modeling; 4) A comparative study for several popular user reputation modeling schemes in terms of retrieving best answers in Community Question Answering services.

In Section II, we review related work in terms of some general CQA studies, user reputation modeling and link analysis. In Section III, we review several existing user reputation modeling schemes and discuss how to use PageRank in CQA. In Section IV, we discuss topical link analysis in CQA and its challenges. Section V describes experimental results showing the effectiveness of different ranking schemes. Section VI provides discussions and future work.

II. Related Work

Community-based Question Answering (CQA) has become an active research area. Much of the work has focused on Yahoo! Answers due to its popularity. Song et al. [9] propose a metric "question utility" for studying usefulness of questions and showed how question utility can be integrated

into question search as static ranking. Various retrieval models or systems have been proposed (e.g., [10], [11]) to extract answers from QA or FAQ services. Jeon et al. [12] present a framework for using non-textual features like click counts to predict the quality of answers, incorporated with a language modeling-based retrieval model. Agichtein et al. [3] present a supervised approach to mining user interaction and content-based lexical features to identify high quality content in CQA. Bian et al. [13] develop a ranking system to retrieve relevant and high-quality answers. Most models above do not explicitly integrate content quality and user reputation information into the ranking process. Hong and Davison [4] show that user authority is a good indicator for retrieving answers from discussion boards. Zhang et al. [5] applied both ExpertiseRank and HITS to identify users with high expertise. Jurczyk and Agichtein [6] show an application of the HITS algorithm to a CQA portal, especially the user interactions graph, and show a positive correlation between authorities calculated with the HITS algorithm and answer quality. Zhou et al. [14] propose a method for co-ranking authors and their publications using their networks. Most of the work discussed above do not provide a comparative study how their ranking scheme outperforms others. At the same time, most ranking schemes are based on the HITS algorithm.

Two of the most prominent link analysis algorithms, Page-Rank [15] and HITS [7], have been shown to be successful in the context of evaluating quality of Web pages. Nie et al. [8] proposed Topical PageRank and Topical HITS which embed topical information when propagating authority scores. They showed that topical PageRank and topical HITS outperform PageRank and HITS respectively. As far as we know, no research work has shown that whether these ranking schemes can be applied to user reputation modeling especially in the context of CQA.

III. USER REPUTATION MODEL REVIEW

In this section, we will review several user reputation models based on link analysis and simple heuristics.

A. HITS-like scheme

Kleinberg [7] identifies two important properties for a web page: hubness and authority, and proposes a mechanism to calculate them effectively. The basic idea behind HITS is that pages functioning as good hubs will have hyperlinks pointing to good authority pages, and good authorities are pages to which many good hubs point. Authority and hub scores of a web page can be computed via mutual reinforcement, which can be described as follows:

$$A(i) = \sum_{j:j \to i} H(j) \qquad (1)$$

$$H(i) = \sum_{i:i \to j} A(j) \qquad (2)$$

If we treat each user as a node and if user i answers a question posted by user j, there will be a link pointing from i to j. Therefore, a user who often posts good questions which

receive informative answers will have many in-links and a user who often answers questions from others will have many out-links, indicating that the first type of users displays the hubness property and the second type shows authority. Using equation (1) and (2), we can calculate hub and authority scores for each user in CQA. Sometimes, however, we need a single score to represent the rank of a user. For example, when combining other models (e.g., relevance models), a single user rank score may help us simplify our overall model. One easy way is to combine them in a linear fashion, which is used in our experiments:

$$UserRank(i) = \gamma * A(i) + (1 - \gamma) * H(i) \qquad (3)$$

B. PageRank-like scheme

Page et al.'s PageRank [15] is a static ranking of web pages based on the measure of prestige in social networks. PageRank can be seen as a random surfer model in which a surfer on a given page i can choose with probability $(1 - d)$ to select uniformly one of its outlinks and with probability d to jump to a random page from the web. The PageRank score of node i is defined as the stationary probability of finding the random surfer at node i. One formulation of PageRank is:

$$PR(i) = (1 - d) \sum_{j:j \to i} \frac{PR(j)}{O(j)} + d\frac{1}{N} \qquad (4)$$

PageRank is not a popular choice to model user reputation in the context of CQA. One possible reason is that there is no obvious evidence implying that "hierarchical ranking" is better than the bipartite structure used in HITS (or even as effective). In addition, equation (4) indicates that a node would share its PageRank score by uniformly distributing the value to each out-going link. However, if we treat each user as a node and there would be a link from user i to j if j answers a question posted by i, it does not make much sense that user i would share its importance to user j since user j should have higher expertise because of answering questions. Furthermore, PageRank needs to randomly "jump" to any page on the Web even there is no hyperlink between them. This lacks an intuitive explanation since it is difficult to think about a user who can share authority with other users with whom the user never interacts. In this paper, we think about the links between user nodes as the ***possibility*** that interactions could happen between users. If no interactions ever happen between two users, they still might invoke interactions in the future with a certain low probability, captured by the "random jump" part of PageRank If they already have interactions, the probability of their future interactions would be higher than random and indicated by the number of existing interactions, which is captured by "out-going links" part of PageRank. Therefore, the PageRank score of a user measures the activeness of this user.

C. Other Heuristics

In [5], Zhang et al. propose a heuristic ranking scheme called *Z-Score* that is based on the number of questions and

answers one user generates. Z-Score is calculated as follows:

$$Z = \frac{a - n/2}{\sqrt{n}/2} = \frac{a - q}{\sqrt{a + q}} \qquad (5)$$

where a is the number of answers provided by one user, q is the number of questions produced by one user, and n is the sum of a and q. Zhang et al.'s rationale for the heuristic is to measure the difference in behavior from a "random" user who posts answers with probability $p = 0.5$ and posts new questions with probability $1 - p = 0.5$. If the user equally asks or answers questions, the z-score will be close to 0. A positive z-score captures a user who asks answers more than asks. Another simple heuristic is derived from [4] in which the authors found that the number of posts a user generates and the number of replies a user provides are two good indicators for the user reputation in forums. Here, we use the linear combination of the number of questions and answers a user generates as the model of reputation:

$$SimpleRank = \theta * a + (1 - \theta) * q \qquad (6)$$

where a is the number of answers one user provides and q is the number of questions that user produces. The parameter θ is used to control whether we emphasize the capability to post new questions or to answer questions for a user. In our experiment, we use $\theta = 0.8$ to focus on the capability to answer questions.

IV. TOPICAL LINK ANALYSIS FOR USER REPUTATION

So far, all user reputation models we reviewed are trying to give a "global" user reputation score, which means that the score represents the user's authority across all topics. However, one may argue that an expert in Computer & Internet may not give good suggestions in Gardening. Obviously, it is better to give authority scores according to different topics and rank user reputation differently. That is why some ranking schemes are designed to take topical information into account, such as Topic-Sensitive PageRank [16]. Here, we want to review Topical PageRank [8], one successful topical ranking scheme, and discuss how to adapt it into the context of user reputation modeling in CQA.

A. Topical PageRank

The main motivation of Topical PageRank is that the authority score of a web page should be modeled with respect to different topics. The basic idea of Topical PageRank is to incorporate topic distribution into the representation of each web page as well as the importance score of each page. Therefore, two vectors are associated with each page: the content vector and the authority vector. The content vector $C_u : [C(u_1), C(u_2), ..., C(u_T)]$ is a probability distribution representing the content of u, solely determined by content itself. The authority vector $A_u : [A(u_1), A(u_2), ..., A(u_T)]$ is used to measure the importance of the page where $A(u_K)$ is the importance score on topic K.

Topical PageRank is also a random surfer model. On each page, the surfer may either follow the outgoing links of the

page with probability $1 - d$ or jump to a random page with probability d. When following links, the surfer may either stay on the same topic to maintain topic continuity with probability α or jump to any topic i on the target page with probability $1 - \alpha$. The probability of jumping to topic i is determined by $C(u_i)$. When jumping to a random page, the surfer is always assumed to jump to a random topic i. Therefore, the authority score $A(i)$ on page u is calculated as follows:

$$A(u_i) = (1 - d) \sum_{v:v \to u} \frac{\alpha A(v_i) + (1 - \alpha)C(v_i)A(v)}{O(v)} \\ + \frac{d}{N}C(u_i) \qquad (7)$$

where $A(v) = \sum A(v_i)$. Note that authors in [8] also proposed a topical version of the HITS algorithm, which may be interesting to adapt into CQA in future work.

B. Adapting Topical PageRank to CQA

One question for Topical PageRank is how to obtain the content vector C_u. In [8], a text classifier trained on the pages selected from the twelve top categories (e.g., Arts, Computers, Games) of the dmoz Open Directory Project (ODP) was used. For CQA, a fine-grained topic distribution like Software and Hardware is needed, which is usually hard to obtain. In order to adapt Topical PageRank for CQA, we propose to use an unsupervised learning algorithm to obtain content vector. In this work, we use probabilistic Latent Semantic Analysis (pLSA) [17], a simple unsupervised topic modeling method. pLSA is a generative model in which documents are not "hard" classified to topics but characterized by a mixture of topics with weights. The model is to maximize the log-likelihood function

$$L = \sum_d \sum_w n(d, w) log P(d) \sum_z P(w|z)P(z|d) \qquad (8)$$

where $n(d, w)$ denotes the number of times w occurred in d. The standard computation procedure for maximum likelihood estimation in latent variable models is the Expectation Maximization (EM) algorithm. We do not include details of EM in our paper.

After knowing the topic distribution of each document (here, in CQA, each question and each answer can be seen as one document), we want to know the topic distribution of each user if we treat users as nodes. One simple way is to add the topic distribution of each document one user U_i generates together. Therefore,

$$P(z|U_i) = \sum_{d \in Q(U_i)} P(z|d) \qquad (9)$$

where $Q(U_i)$ represents all the documents user i produces.

Another approach is to introduce a new variable u into pLSA model to represent users. Therefore, the log-likelihood function is :

$$L = \sum_u \sum_d \sum_w n(d, w, u) log P(d) \sum_z P(u|z)P(w|z)P(d|z) \qquad (10)$$

The advantage of this approach is that we can directly obtain the topic distribution for each user through EM algorithm. However, this would require more computation especially for large-scale data. In this work, we do not introduce the new variable and focus instead on the simpler method.

V. EXPERIMENTAL RESULTS

In this section, we will compare several ranking schemes including what we introduced in previous sections.

A. Experimental Setup

We crawled 59,936 posts from Yahoo! Answers through its API, where 14,122 are questions and 45,814 are answers. Among all answers, 12,731 are selected as "best answers" by users to particular questions. 37,940 unique users are extracted from the dataset. Since we do not have real queries that users issued in Yahoo! Answers for searching the question answer archive, we treat each question in our dataset as a query and all answers as potential answers to the query. Therefore, we have 12,731 questions as queries and their corresponding best answers as relevant results. (We do not consider those questions that have no best answers.) We want to measure ranking schemes in two ways. First, we want to see whether a ranking scheme can return the best answer selected by user high in the return-list. Second, we want to see whether the ranking scheme can return more best-answers higher than other answers. Specifically, we are looking at these metrics:

- Precision@1(Strict): If the question has the answer selected as the best answer, this best answer should return at the first place if we use the whole question as a query.
- Mean Reciprocal Rank (MRR): Same as the metric above, we want to see the position where the best answer ranked if the question has the best answer chosen by the user.

The above two metrics are strict metrics since each question (or query) only has one answer (or relevant result). In order to evaluate ranking schemes for our second goal, we relax the constraint of returned answers by treating all best answer as relevance results.

- Precision@1(Relaxed): We only want to see whether the top result is a best answer regardless whether it is the best answer selected for the query or not.
- Precision@10: We want to see how many results in top 10 positions are best answers.
- Mean Average Precision (MAP): We sum up the precision score whenever a new best answer retrieved and average all scores across all queries (questions).

Since we need to calculate the relevance score for each answer, user reputation model itself is not enough. We combine user reputation model with the Okapi BM25 weighting function. For each answer, we calculate two rankings, one from BM25 and the other one from a user reputation model and combine them as follows:

$$\lambda * rank_{BM25}(a) + (1 - \lambda) * rank_{USER}(a) \qquad (11)$$

Unsurprisingly, the parameter λ affects the final results. Thus,

TABLE I
RESULTS OF BM25

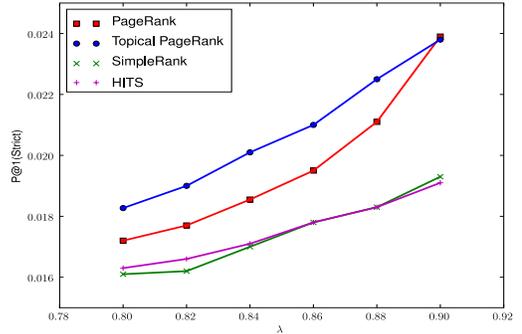| P@1(S)	MRR	P@1(R)	P@10	MAP
0.0857	0.1414	0.3410	0.3170	0.3081

Fig. 1. P@1 (Strict)

for each method and metric, we show how this parameter influences the ranking results. The results of only using BM25 are shown in Table I, where P@1(S) indicates Precision@1(Strict) and P@1(R) indicates Precision@1(Relaxed). We do not include the results of Z-Score in the following discussions since it performs worst in our experiments and the values for each evaluation metric is low. Figure 1 and 2 show the result of using "Strict Metrics" as λ is varied from 0.8 to 0.9. Two obvious observations can be quickly obtained. First, all the results are worse than using BM25 ranking result alone. The reason may be that sometimes the best answer could be produced by users that might not be the first authoritative user (e.g., may be second or third). Since "Strict Metrics" only measure whether **the** best answer can be retrieved or not, we argue that the results may not reflect what would happen in real world where users often issue short queries that are less likely to match a whole question. However, this result does give us hints about how different ranking schemes perform in terms of "Strict Metrics", which leads to the second observation that PageRank-like approaches perform better than
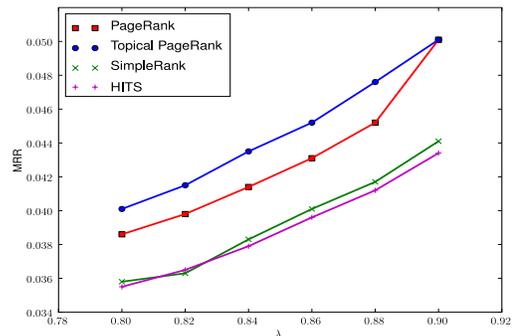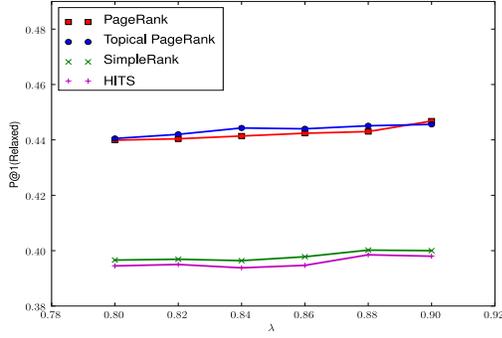
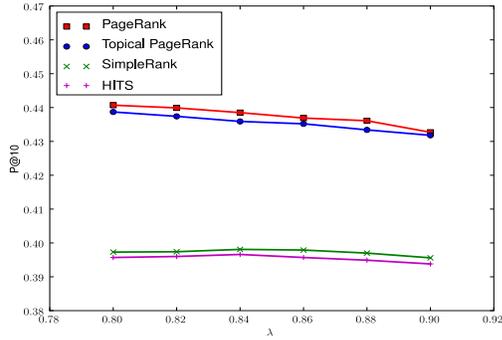

Fig. 2. MRR

Fig. 3.   P@1(Relaxed)

Fig. 4.   P@10(Relaxed)

HITS-like schemes and other heuristics. HITS-like schemes cannot capture the notion of "hierarchical authorities", which means that the user who can answer a question posted by an authority should have higher authority score. PageRank-like approaches naturally model this notion and give better approximation than HITS.

B. Relaxed Metrics

If we use "Relaxed Metrics", Figures 3, 4 and 5 show that all ranking methods combined with BM25 can improve
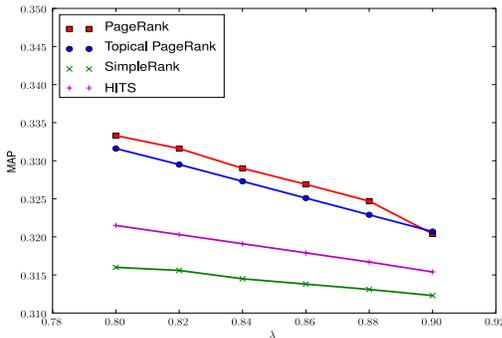

Fig. 5.   MAP(Relaxed)

retrieval performance significantly, which validate the conclusions from other related work that user reputation models can help retrieval tasks in CQA. PageRank-like approaches still outperform simple heuristics and the HITS-like scheme. The results also indicate that SimpleRank performs similarly to the HITS-like scheme. This is more evidence that HITS cannot model "hierarchical authorities" as discussed above. One interesting observation is that as λ increases, P@1(Strict), P@1(Relaxed) and MRR also increase but P@10 and MAP decrease. Because the first three metrics only focus on one answer (or relevance result) per question (or query), as we discussed before, user reputation modeling may not help much and P@1(Strict) and MRR are actually worse than only using BM25. On the other hand, if we care about returning more relevant results, P@10 and MAP show the value of user reputation modeling and indicate a significant improvement.

C. Topical PageRank

For Topical PageRank, we use pLSA as our topic model and specify 20 latent variables (topics). Since our dataset is from the Computer & Internet category of Yahoo! Answers, which has 7 sub-categories, we arguably think the number of latent variables (topics) would necessarily cover major topics. However, this number can be given by the number of positive singular values of the word-chunk matrix, a popular technique used in text mining [18].

In Figures 1-5, PageRank and Topical PageRank perform similarly and we want to see whether there is a significant difference or not. We perform a t-test on each evaluation metric, showing that Topical PageRank does significantly better than PageRank on P@1(Strict), MRR and P@1(Relaxed) (p-value=0.05) while PageRank does significantly better than Topical PageRank on P@10 and MAP (p-value=0.05). The possible reason that Topical PageRank performs better on those metrics that only consider one result per query is that Topical PageRank can capture the notion that certain users only have expertise on some topics. So a user may not be an overall computer expert but can provide several good answers just for hardware repair. In other words, PageRank would average the authority score for all topics and therefore give better approximation for more macro-level evaluation metrics such as P@10 and MAP. Another reason that the results of PageRank and Topical PageRank are close is that our dataset only consists of questions and answers in one main category, Computer & Internet. Compared to [8] where they used topics of the top level of the ODP hierarchy, the difference between topics in our dataset is relatively small. You can imagine that a good expert in Computer may not be an authority in Sports but we probably need to agree that a good expert in Computers may also be an expert in Computer Software. In this case, Topical PageRank shows similar performance as PageRank itself. However, we postulate that if a more topic-diverse dataset is used, Topical PageRank would provide more benefit than PageRank because this less diverse dataset already shows the improvement of Topical PageRank.

TABLE II
RESULTS OF DIFFERENT α

α	P@1(S)	MRR	P@1(R)	P@10	MAP
0.70	0.0225	0.0476	0.4450	0.4334	0.3229
0.75	0.0223	0.0475	0.4447	0.4335	0.3229
0.80	0.0223	0.0475	0.4458	0.4336	0.3229
0.85	0.0224	0.0475	0.4460	0.4336	0.3230
0.90	0.0221	0.0473	0.4456	0.4337	0.3230

Since Topical PageRank has a parameter α to indicate the probability of whether to stay on the topic or jump to other topics, we choose several different α values to see how this parameter influences our experimental results. In this set of experiments, we also combine the ranking with BM25 where λ is chosen as 0.88, the value at which the best performance of Topical PageRank is achieved. Table II details the results. Obviously, different α values do not influence final ranking results so much. We think it is due to the lower diversity of our dataset since "jumping" to other topics would still remain in Computer & Internet.

VI. CONCLUSION & FUTURE WORK

In this paper, we reviewed two popular ranking schemes, HITS and PageRank, and their applications in user reputation modeling. Due to the naturally bipartite structure of CQA, the HITS scheme and its variations attract more attention in related work and they do improve retrieval performance according to our experiments. On the other hand, we discussed the possibility of using a PageRank-like scheme and introduced topical link analysis into user reputation models. We showed how to incorporate unsupervised topic analysis, in our case pLSA, into topical link analysis. The performance of PageRank and Topical PageRank is much better than HITS and other heuristics in our experiments, which indicates a "hierarchical property" of user reputation. In addition, Topical PageRank is slightly better than PageRank in P@1(Strict) and MRR. We also found that while in general user reputation can help retrieval performance when incorporated with BM25, the performance of returning the exact best answer to a particular question (P@1(Strict) and MRR) decreases. We view this problem as a difficulty that most user reputation models give a "global" reputation for a user which may not reflect the authority of a user in certain topics. Additionally, BM25 plus user reputation models may not be a good approach to Question Answering task but are good indicators for returning high quality content (which is different from Question Answering!).

Our future work includes: 1) Compare all ranking schemes in a larger scale and more diverse dataset. We want to see whether the diversity of topics in a dataset impacts the results of user reputation models and retrieval performance. 2) Directly model users in pLSA and see whether it is a reasonable ranking scheme compared to link analysis. And also see performance of the combination of these two approaches. 3) We also want to see the performance of user reputation models on different social media datasets, such as discussion boards, Wikipedia and Twitter. Do they have the same user reputation structure or they are different?

REFERENCES

[1] J. Leibenluft, "A librarian's worst nightmare: Yahoo! answers, where 120 million users can be wrong," *Slate Magazine*, Dec. 2007. [Online]. Available: http://www.slate.com/id/2179393

[2] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha, "Learning to recognize reliable users and content in social media with coupled mutual reinforcement," in *Proc. of the 18th International Conference on World Wide Web (WWW)*, 2009, pp. 51–60.

[3] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, "Finding high-quality content in social media," in *WSDM '08: Proceedings of the International Conference on Web Search and Web Data Mining*. New York, NY, USA: ACM, 2008, pp. 183–194.

[4] L. Hong and B. D. Davison, "A classification-based approach to question answering in discussion boards," in *Proc. of the 32nd Annual Int'l ACM SIGIR Conf. on Research and Dev. in Information Retrieval*, 2009.

[5] J. Zhang, M. S. Ackerman, and L. Adamic, "Expertise networks in online communities: structure and algorithms," in *WWW '07: Proceedings of the 16th International Conference on World Wide Web*. New York, NY, USA: ACM, 2007, pp. 221–230.

[6] P. Jurczyk and E. Agichtein, "Discovering authorities in question answer communities by using link analysis," in *CIKM '07: Proceedings of the 16th ACM Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2007, pp. 919–922.

[7] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[8] L. Nie, B. D. Davison, and X. Qi, "Topical link analysis for web search," in *Proc. of the 29th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*, 2006, pp. 91–98.

[9] Y.-I. Song, C.-Y. Lin, Y. Cao, and H.-C. Rim, "Question utility: A novel static ranking of question search," in *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2008.

[10] M. Surdeanu, M. Ciaramita, and H. Zaragoza, "Learning to rank answers on large online QA collections," in *46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*, 2008.

[11] X. Xue, J. Jeon, and W. B. Croft, "Retrieval models for question and answer archives," in *Proc. of the 31st Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*, 2008, pp. 475–482.

[12] J. Jeon, W. B. Croft, J. H. Lee, and S. Park, "A framework to predict the quality of answers with non-textual features," in *SIGIR '06: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2006, pp. 228–235.

[13] J. Bian, Y. Liu, E. Agichtein, and H. Zha, "Finding the right facts in the crowd: factoid question answering over social media," in *WWW '08: Proc. of the 17th International Conference on World Wide Web*. New York, NY, USA: ACM, 2008, pp. 467–476.

[14] D. Zhou, S. A. Orshanskiy, H. Zha, and C. L. Giles, "Co-ranking authors and documents in a heterogeneous network," in *ICDM '07: Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 739–744.

[15] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Technical Report 1999-66, November 1999, previous number = SIDL-WP-1999-0120.

[16] T. H. Haveliwala, "Topic-sensitive pagerank," in *WWW '02: Proceedings of the 11th International Conference on World Wide Web*. New York, NY, USA: ACM, 2002, pp. 517–526.

[17] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. of the 22nd Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*, 1999, pp. 50–57.

[18] M. D. Choudhury, H. Sundaram, A. John, and D. D. Seligmann, "What makes conversations interesting? Themes, participants and consequences of conversations in online social media," in *Proc. of the 17th Int'l Conf. on World Wide Web (WWW)*, Apr. 2009, pp. 331–340.