Competition-Based Networks for Expert Finding

Çiğdem Aslay*
Web Research Group, UPF
Barcelona, Spain
aslayci@acm.org

Neil O'Hare
Yahoo! Research
Barcelona, Spain
nohare@yahoo-inc.com

Luca Maria Aiello
Yahoo! Research
Barcelona, Spain
alucca@yahoo-inc.com

Alejandro Jaimes
Yahoo! Research
Barcelona, Spain
ajaimes@yahoo-inc.com

ABSTRACT

Finding experts in question answering platforms has important applications, such as question routing or identification of best answers. Addressing the problem of ranking users with respect to their expertise, we propose Competition-Based Expertise Networks (CBEN), a novel community expertise network structure based on the principle of competition among the answerers of a question. We evaluate our approach on a very large dataset from Yahoo! Answers using a variety of centrality measures. We show that it outperforms state-of-the-art network structures and, unlike previous methods, is able to consistly outperform simple metrics like best answer count. We also analyse question answering forums in Yahoo! Answers, and show that they can be characterised by factual or subjective information seeking behavior, social discussions and the conducting of polls or surveys. We find that the ability to identify experts greatly depends on the type of forum, which is directly reflected in the structural properties of the expertise networks.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

Keywords

expert finding, community question answering, competition-based expertise network, knowledge sharing

1. INTRODUCTION

Search engines are not fully capable of answering complex information needs that require deep semantic understanding or high coverage of human knowledge and experience. *Community Question Answering* (CQA) portals (e.g., Yahoo! Answers, Stack Overflow), in which people can use natural language instead of keywords to ask questions, have emerged in response to these limitations. Although such portals provide a number of feedback mechanisms, these mechanisms are open to abuse and do not provide automatic ways to identify experts or high-quality answers. For this reason, ranking users with respect to their expertise is needed for applications such as question routing [17], best answer prediction, and for improving reward mechanisms.

Expert Finding, identifying knowledgeable people on a given topic, has been an active research area even before it was introduced as a task in Text Retrieval Conference (TREC) in 2005 [10]. In addition to Information Retrieval approaches, graph-based methods have been used to find experts by identifying the most central actors in *expertise networks*, under the assumption that graph centrality is correlated with expertise [16]. Usually, the networks are built from the links between askers and answerers of a question, or between askers and best answerers. These networks ignore the information encoded in the inherent *competition* among the answers of a question to be selected as the best answer. We propose the *Competition-Based Expertise Network* (CBEN), a novel structure that builds expertise networks by creating ties between the best answerer and the other answerers they have "beaten". Combined with graph centrality metrics, CBEN provides a natural way to identify experts.

We evaluate our approach on a large dataset from Yahoo! Answers, using a variety of centrality measures, and show not only that it outperforms state-of-the-art graph-based techniques but, unlike previous approaches [5, 16], it consistently outperforms answer count and best answer count. Additionally, we show that in certain cases our approach outperforms personal best answer ratio, which has been considered an upper bound on expert prediction accuracy. We also analyse different types of Q&A forums and show that the ability to identify experts greatly depends on the type of forum, which is directly reflected in the structural properties of the expertise networks.

2. RELATED WORK

*This work was done while the first author was visiting Yahoo! Research Barcelona, under the Yahoo! internship program for EM-DMKM.

SIGIR'13, July 28–August 1, 2013, Dublin, Ireland.

Early work in expert finding used Information Retrieval methods to model the relevance of candidate users to a given topic. In *Profile-Based Methods*, each candidate is represented by a textual profile, which are ranked with respect to an expertise query [9], while in *Document-Based Methods* relevant documents for the query are first retrieved, and candidates are then ranked based on the co-occurrence of topic and candidate mentions in the retrieved documents [3].

In forums with a narrow topical focus, such as Yahoo! Answers leaf-level categories, all users can be assumed *relevant* to the topic, and rather than considering IR approaches, interactions between askers and answerers can be used to find experts. Interaction networks involving question-answering activity, or expertise networks [1], have been studied in the past [15, 13], and several of their structural properties, such as the disassortativity of the expertise level, have been highlighted [1, 7]. Under the assumption that the most central actors correspond to the most expert users, standard centrality measures, such as PageRank and HITS [16], and custom centrality metrics, like *ExpertiseRank* [16, 11], have been employed to detect expert users. None of these approaches performed better than simpler metrics such as indegree or best answers count [5] and, to date, the best answer ratio of the replier has been found to be the best predictor of best answerer, and it is considered as an upper bound on the prediction accuracy [8].

Other work has explored machine learning approaches [6], including work that is most similar to ours [14], which proposed a competition-based approach to expertise ranking, training an SVM based on the best answer competition between users. Our work differs in that they only use local information, while we exploit global information by building a social network.

3. QUESTION ANSWERING FORUMS

Yahoo! Answers is a general purpose community question answering portal. It includes a hierarchy of categories, with 26 predefined Top-Level Categories (TLC), such as Sports or Family, and a continually growing number of Leaf-Level Categories (LLC): over 1,300 at the time of this study. Yahoo! Answers has a strict question-and-answer format, with questions submitted as short statements, with an optional detailed description, and an obligatory associated category. The asker can select the best answer (the most common feature across all CQA platforms); if no best answer is selected by the asker, the community can vote to determine it.

Next, we analyze the types of Q&A forums on Yahoo! Answers, studying a large data sample from 2011, containing more than *20 million* questions and *87 million* answers.

3.1 Forum Characterisation

User intent in asking questions can vary widely. Basic classifications partition questions into *informational*, if they seek facts or advice, and *conversational*, if they are intended to stimulate a discussion [12]. In practice, every forum category has some mix of requests for factual information, advice seeking and social conversation or discussion. Adamic *et al.* [2] used k-means to cluster Yahoo! Answers leaf level categories, using features such as the average number of replies to a question and the average number of characters in a reply, and identified three different category types: discussion forums, advice-support seeking and factual answer seeking.

Since one of our goals is to test the performance of expert finding across different forum types, we replicated the experiment of Adamic *et al.*, and included additional activity-based features: average character count in questions, average character count in the asker feedback, the proportion of questions with contradictory answer ratings (i.e. with both "thumbs up" and "thumbs down" feedback), and the average magnitude of the rating gap for questions with contradictory answer ratings. We found that the optimal number of clusters is 4 for the extended feature set. For each leaf-level category, we conducted a side by side qualitative comparison of the assigned clusters. K-means clustering with the extended feature set reproduced the good examples of the baseline clustering, and overall it produced a better categorisation with a higher R^2 value. The additional cluster, beyond those reported by Adamic *et al.* [2], contains mostly questions oriented to polls and surveys. We label the Q&A forum types as follows: *factual-information seeking*, *subjective-information seeking*, *social discussion*, and *poll-survey conducting*.

Table 1: Cluster means of forum types.

	Factual	Subjective	Discussion	Poll-Survey
Reply Length	269.30	**351.10**	331.99	218.99
Question Length	209.34	**294.27**	266.33	206.69
Feedback Length	40.09	50.47	**58.34**	42.29
Nr of Replies	2.55	3.14	**6.32**	5.50
Asker/Replier Sim	0.06	0.10	0.35	**0.51**
Contrad. Feedback Ratio	0.02	0.06	**0.35**	0.14
Contrad. Feedback Gap	0.08	0.20	**2.70**	0.59

Table 1 summarizes the cluster means of these forum types with respect to the variables used for clustering. The *factual information seeking* forums (e.g. Computer Networking) have low average values for all the features: questions and answers are short, with little contradictory feedback. *Subjective information seeking* forums (such as Pregnancy & Parenting) have the highest question and answer length. *Social discussion* forums, such as Politics, are characterized by high level of contradictory feedback and high collective participation in answering. Finally, *poll-survey conducting* forums (e.g., Baby Names) have the lowest question and answer length, and the highest asker/replier overlap.

4. COMPETITION-BASED EXPERTISE NETWORK

Community expertise networks [1] are social networks in which nodes represent people and edges represent the flow of expertise and knowledge among them. They can be modeled as weighted graphs, with previous studies using two different structures for the networks: i) *Asker-Replier networks (ARN)*, with directed edges from askers to answerers of questions, weighted by the number of answers [13]; ii) *Asker-Best Answerer Networks (ABAN)*, with directed edges from askers towards best answerers [5, 11], weighted by the number of best answers. ARN treats all answers equally, ignoring "best answer" information, whereas ABAN excludes information about users who are not selected as the best answerer.

When an answer is selected as the "best answer" to a particular question, it is done so in comparison with to the other answers of the same question. There is, therefore, an inherent *competition* between the answers of a question, and we can assume that the relative expertise of the best answerer is higher than that of the other answerers of the question.

To leverage this information, we propose the *Competition-Based Expertise Network (CBEN)*, a novel structure for com-

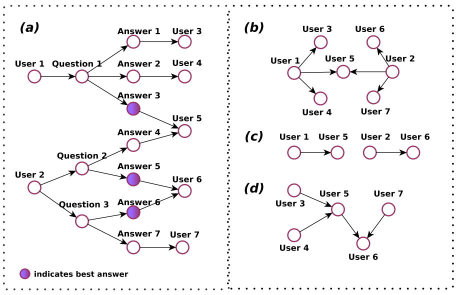

Figure 1: (a) Relations between users, questions, answers and best answers. On the right, the corresponding (b) asker-replier network (ARN), (c) asker-best answerer network (ABAN), and (d) competition-based expertise network (CBEN).

munity expertise networks, with directed edges from the non-best answerers towards the best answerer of a question. This keeps track not only of how many best answers a user has contributed, but also of who they have "beaten" in the competition to be selected as best answer. In general purpose CQA communities, unlike specialized technical forums, asking a question is not necessarily related to a lack of expertise [1, 16], so we do not include the relation between the askers and answerers of a question in CBEN. Examples of the three expertise networks are shown in Figure 1.

5. EVALUATION

We consider each leaf-level category of Yahoo! Answers separately, assuming that each category represents a topic and the users within it are topically relevant. Unlike previous work which relies on human-created ground truth [1, 16, 8, 14], we adopt a prediction-oriented approach for evaluation, and evaluate expert finding methods according to their ability to predict the best answerer of a question, based on the knowledge of past interactions. For each leaf-level category, we extract the CBEN, ARN and ABAN networks, and compute graph centrality measures for them. For each question, we rank all answerers by their centrality score, and the user with the highest score is predicted as the best answerer. We evaluate using *prediction accuracy*, the ratio of questions for which the best answerer is predicted correctly.

We study centrality measures belonging to three main families: *degree*-based centrality (*Degree* and *InDegree*), *distance*-based centrality (*Harmonic* centrality [4]: $c_h(x) = \sum_{y \neq x} 1/dist(y,x)$), and *spectral* centrality (weighted *PageRank* and *HITS*). We do not consider *path*-based centrality measures, such as *betweenness*, since their computational complexity is prohibitive for large datasets.

We partition the dataset by time, extracting networks using data from Jan to Oct 2011 (training period), and based on that we predict the best answerers for questions submitted between Oct 2011 to Jan 2012 (test period).

5.1 Prediction results

In Table 2 we present the average prediction accuracy for graph-based methods and a number of baselines (random selection, number of answers, number and ratio of best answers), grouped by forum type. Factual-information seeking forums have the top performance, followed by subjective-

Table 2: Forum Type Average Prediction Accuracy.

		Discussion	Factual	Poll	Subjective
ARN	PageRank	0.352	0.497	0.369	0.447
	HITS	0.342	0.484	0.363	0.436
	Harmonic	0.308	0.477	0.322	0.419
	Indegree	0.347	0.495	0.364	0.445
	Degree	0.346	0.492	0.359	0.441
ABAN	PageRank	0.401	0.526	0.405	0.485
	HITS	0.405	0.527	0.402	0.482
	Harmonic	0.351	0.516	0.364	0.467
	Indegree	0.400	0.534	**0.412**	0.490
	Degree	0.395	0.531	0.401	0.487
CBEN	PageRank	0.404	0.536	0.411	0.500
	HITS	**0.415**	**0.545**	0.411	**0.506**
	Harmonic	0.361	0.524	0.371	0.481
	Indegree	0.402	0.538	**0.412**	0.498
	Degree	0.377	0.506	0.387	0.466
	Random	0.326	0.352	0.340	0.323
	Total Answer Count	0.349	0.494	0.366	0.445
	Best Answer Count	0.403	0.535	0.413	0.491
	Best Answer Ratio	0.465	0.610	0.468	0.582

Table 3: Performance Comparison Matrix.

% Factual LLCs	ARN PR	CBEN HITS	ABAN IND	# ANS	# BA	BA %
ARN PR	-	5.1	4.3	59.4	3.4	2.6
CBEN HITS	**94.9**	-	**63.2**	**94.9**	**63.2**	7.8
ABAN IND	**95.7**	36.8	-	**93.2**	17.9	6.7
# ANS	40.6	5.1	4.3	-	3.4	2.6
# BA	**96.6**	36.8	**26.5**	**94.9**	-	7.7
BA %	**97.4**	**92.2**	**93.3**	**97.4**	**91.5**	-

information seeking forums: these are both more suitable for expert ranking than social discussion and poll-survey forums. The best answer ratio has the highest average prediction accuracy for all forum types. This is expected, as numerous studies show the high correlation of this metric with expertise [13, 8, 11]. Even best answer ratio performs poorly, however, in social discussion and poll-survey forums.

To test if the differences among the performance of the metrics are statistically significant, we performed paired t-tests ($p < 0.01$) on the prediction accuracies obtained for each LLC. First, we performed the tests among five centrality measures within each forum type and network structure separately, finding that PageRank is the best centrality measure of ARN, while HITS is best for CBEN, and InDegree, as shown previously [5], performs best on ABAN.

We repeated the t-tests between pairs of such best network / centrality-measure combinations. CBEN HITS significantly outperforms both ARN PageRank and ABAN InDegree in the factual and subjective information seeking categories; it is also the only network-based metric that significantly outperforms both the number of answers and best answers for those categories. This is a strong result compared to previous work [5, 16], which suggested that graph-based algorithms cannot outperform these simple metrics. A summary of the paired t-tests for the factual information seeking forum type is given in Table 3, with statistically significant results in bold text. The matrix entries represent the percentage of leaf level categories where the metric reported in the row has higher prediction accuracy that in the column.

Since ARN does not use best answer information, caution is needed when comparing it with other approaches on a best answer prediction task. For this reason we complement the evaluation by computing the *rank correlation* of the network-based metrics with the best answer ratio, as in previous studies [13, 11]. We selected the 10 users with highest centrality (from training data) and computed the Spearman correlation with their personal best answer ratio (from test data). CBEN HITS is also the best approach for this evaluation, with the highest correlation in 40% of

Table 4: Forum Type Network Statistics.

		Discussion	Factual	Poll-Survey	Subjective
ARN	$\#wcc$	69	2220	529	951
	$gwcc_s$	0.99	0.89	0.98	0.94
	τ	14.70	0.19	12.00	0.87
CBEN	$\#wcc$	33	1021	240	475
	$gwcc_s$	0.99	0.90	0.98	0.94
	τ	23.02	2.22	18.08	4.76
ABAN	$\#wcc$	1127	7824	3647	4916
	$gwcc_s$	0.76	0.55	0.77	0.65
	τ	1.70	0.03	2.15	0.11

factual-information seeking forums and 43% of subjective-information seeking forums.

5.2 Dependency with Network Structure/Type

In the previous section, we saw that prediction accuracy strongly depends on the forum type. In this subsection, we look at the network structure to ask whether such qualitative aspects can be captured quantitatively.

We studied a number of network characteristics, but due to space limitations, we report only the metrics that showed interesting correlations with expert prediction: 1) *Number of Weakly Connected Components* ($\#wcc$) 2) *Relative Greatest Weakly Connected Component size* ($gwcc_s$), and 3) *Triangulation rate* (τ), the proportion of edges that belong to an undirected triangle. Table 4 shows their average values for each forum type. For ARN and ABAN τ is determined by how many users become both askers and answerers of the questions. For CBEN it indicates that users show highly overlapping participation in answering questions. Discussion and poll-survey forums show high τ and $gwcc_s$ while, as expected, $\#wcc$ is considerably lower. These results provide a strong evidence that there is no clear distinction of expert and novice roles in non-expertise based categories: users show highly overlapping participation in question answering and there is no real specialization in the choice of questions to answer. Within every LLC, the prediction accuracies have a negative linear correlation with and τ and $gwcc_s$, and a positive correlation with $\#wcc$. These results demonstrate that expertise should not be sought in forums where the dominant user intent is socialising rather than quality information seeking.

More generally, such network properties help identify cases where network-based expertise finding approaches are more effective. We analyzed the network statistics for the 9% of the total of factual and subjective information seeking categories for which the CBEN outperforms best answer ratio, and found that, in particular, τ is much lower for these cases.

6. CONCLUSIONS

In this paper we introduce a novel network structure for modeling expertise in community question answering portals, based on the idea of competition among users to be selected as best answerer. Using a large Yahoo! Answers dataset, we evaluate this network structure for the task of best answer prediction, using a number of centrality measures, and show an improvement over state-of-the-art techniques. We also show that, unlike previously used network-based methods for this task, the proposed network type consistently outperforms answer count and best answer count, and in certain cases outperforms best answer ratio. By clustering Q&A forums by their structural features, we identify 4 forum types, and observe that the prediction task is more difficult for forums dominated by social discussions and surveys, as opposed to factual/subjective information seeking forums. A deeper investigation of the networks revealed that network properties of triangulation and connectivity capture the potential predictability of best answerers. For future work, we will further investigate the network features that affect the performance of the task, and integrate topic modeling to capture better relevance of users to the categories.

Acknowledgments

This research is partially supported by European Community's Seventh Framework Programme FP7/2007-2013 under the ARCOMEM and Social Sensor projects, by the Spanish Centre for the Development of Industrial Technology under the CENIT program, project CEN-20101037 "Social Media", and by Grant TIN2009-14560-C03-01 of the Ministry of Science and Innovation of Spain.

7. REFERENCES

[1] *CommunityNetSimulator: Using Simulations to Study Online Community Networks*, volume 7, Michigan State University, 2007. Springer.

[2] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge sharing and Yahoo Answers: everyone knows something. In *WWW*, pages 665–674. ACM, 2008.

[3] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *SIGIR*, pages 43–50. ACM, 2006.

[4] P. Boldi, M. Rosa, and S. Vigna. Hyperanf: approximating the neighbourhood function of very large graphs on a budget. In *WWW*, pages 625–634. ACM, 2011.

[5] M. Bouguessa, B. Dumoulin, and S. Wang. Identifying authoritative actors in question-answering forums: the case of Yahoo! Answers. In *KDD*, pages 866–874. ACM, 2008.

[6] B.-C. Chen, A. Dasgupta, X. Wang, and J. Yang. Vote calibration in community question-answering systems. In *SIGIR*, pages 781–790. ACM, 2012.

[7] H. Chen, H. Shen, J. Xiong, S. Tan, and X. Cheng. Social network structure behind the mailing lists. In *TREC, Expert Finding Track*, 2006.

[8] L. Chen and R. Nayak. Expertise analysis in a question answer portal for author ranking. In *WI-IAT*, volume 1, pages 134 –140, dec. 2008.

[9] N. Craswell, D. Hawking, A.-M. Vercoustre, and P. Wilkins. Panoptic expert: Searching for experts not just for documents. In *Ausweb Proceedings*, 2001.

[10] N. Craswell, A. P. D. Vries, and I. Soboroff. Overview of the TREC 2005 Enterprise Track. In *TREC*, 2005.

[11] Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina. Questioning Yahoo! Answers. Technical Report 2007-35, Stanford InfoLab, 2007.

[12] F. M. Harper, D. Moy, and J. A. Konstan. Facts or friends?: distinguishing informational and conversational questions in social Q&A sites. In *CHI*, pages 759–768. ACM, 2009.

[13] P. Jurczyk and E. Agichtein. Discovering authorities in question answer communities by using link analysis. In *CIKM*, pages 919–922. ACM, 2007.

[14] J. Liu, Y.-I. Song, and C.-Y. Lin. Competition-based user expertise score estimation. In *SIGIR*, pages 425–434. ACM, 2011.

[15] E. Smirnova and K. Balog. A user-oriented model for expert finding. In *ECIR*, pages 580–592. Springer-Verlag, 2011.

[16] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *WWW*, pages 221–230. ACM, 2007.

[17] T. C. Zhou, M. R. Lyu, and I. King. A classification-based approach to question routing in community question answering. In *WWW*, pages 783–790. ACM, 2012.