

Social Tagging Behaviour in Community-driven Question Answering

Eduarda Mendes Rodrigues
Microsoft Research
7 JJ Thomson Avenue
Cambridge, CB3 0FB, UK
eduarda@microsoft.com

Natasia Milic-Frayling
Microsoft Research
7 JJ Thomson Avenue
Cambridge, CB3 0FB, UK
natasamf@microsoft.com

Blaz Fortuna
Dept. of Knowledge Technologies
Institute Jožef Stefan, Jamova 39
1000 Ljubljana, Slovenia
blaz.fortuna@ijs.si

Abstract

On-line community services such as Live QnA and Yahoo! Answers enable their members to ask questions and have them answered by the community. The questions are labelled by the users to facilitate search, navigation, and recommendations. In this paper we provide an in-depth analysis of the question labelling practices by contrasting the use of community generated tags in the Live QnA service with the use of topic categories from a fixed taxonomy in the Yahoo! Answers service. We found that community tagging is related to higher levels of social interactions amongst users. Analysis of the most frequently used community tags reveals that active users may establish strong social ties around specific tags. Furthermore, the discriminative value of individual community tags can be low since the corresponding questions may cover a variety of topics. Thus, appropriate care needs to be taken when designing search, browsing, and recommender features for question discovery.

1. Introduction

Community question answering (Q&A) services have gained popularity over the recent years, building on the model first introduced by the top Korean search portal *Naver* [14]. The main purpose of these services is to provide support for users with specific information needs to obtain prompt responses to their questions from other community users. Often, the questions are requests for advice or opinion, which are unlikely to be provided through standard Web search. In fact, the content generated within question answering services represents a rich knowledge base and a valuable resource for other Web users to search and explore. Even though the answers can be submitted by users of all levels of expertise, the quality of answers can compare, or even surpass, the quality of answers given by expert networks and library reference services [8].

Recent research on community Q&A services has focussed on the Yahoo! Answers service, investigating ways to detect quality answers [1, 2, 8] and characterize users authority [7, 10]. In our work we are particularly interested in the practice of labelling questions by the users in order to ensure that their questions can be found and answered by the rest of the community. Within the Yahoo! Answers service, the user content is organized into fixed topic categories and can be accessed through navigation and filtering. The Live QnA service, on the other hand, has adopted an organization of content based on community tags that are created by the users to describe their questions. This is in line with the trend set by the services such as *Flickr.com* and *Del.icio.us*, which popularized the concept of tagging online resources with users' own keywords.

In this paper, we study community labelling practices using community tags and pre-defined topic categories and investigate the implications they have on the community and the design of the service. Our approach involves several techniques. We analyze tag co-occurrence graphs to identify topicality of tag clusters and study the *answer-to* social network to assess the strength of social ties around specific tags. We discovered that the very mechanism of community tagging is conducive to a higher level of social interaction and, consequently, the use of non-topical tags, i.e., social tags. This, in turn, has an effect on the dynamics of the social network. We show that the users who ask and answer questions associated with social tags form sub-groups with strong ties with each other.

Our paper is organized as follows. We first discuss the related research and describe the datasets from the two services that we aim to explore. In Section 4 we present an in-depth analysis of the tagging practices and in Section 5 we discuss the implications for the community dynamics. We end with the concluding remarks in Section 6.

2. Background and Related Work

Community Q&A services include basic features for submitting questions and answers, and several mechanisms for self-regulation of the content quality, such as providing comments on answers, voting for best answers, reporting abuse, and assigning reputation points to community members.

The Live QnA and Yahoo! Answers services are very similar in nature. Figure 1 represents the main entities present in these Q&A services and the types of interactions that result in implicit social networks of users who: 1) *answer to* other users' questions, 2) *comment to* other users answers, and 3) cast *votes on* other users' answers. After a question is submitted to the service, there is an answering period during which the community can respond. The question may receive several answers from multiple users during that period. The community may vote on the answers during a voting phase, with the aim of electing the best answer.

On Yahoo! Answers community members can cast positive and negative votes (thumbs-up and thumbs-down). Comments can be given to 'resolved' questions for which a best answer has already been elected. In contrast, within Live QnA a user's vote represents the selection of the best answer, while comments can be given to any answer at any time during the question lifecycle. This flexibility opens up opportunities for richer user interactions during the answering process and results in a different structure for the Q&A threads.

Both services provide search and browsing features for users to explore recently posted questions as well as answered ones. Furthermore, the services allow users to assign labels to their questions, for making them easier to find. On Yahoo! Answers the question is assigned to a single category from the Yahoo! topic hierarchy. On Live QnA the service suggests tags that have been created by the community. The user can select the recommended tags and/or specify new tags to describe the question. Thus, in contrast to the fixed taxonomy, Live QnA provides a non-hierarchical community generated categorization scheme.

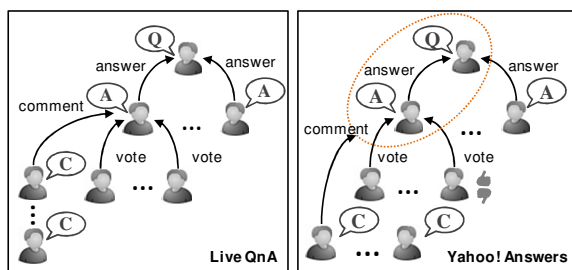


Figure 1. Types of entities and user interactions present in the Live QnA and Yahoo! Answers services (Q: question, A: answer, C: comment and V: vote).

Throughout this paper we refer to the labels assigned to questions within the Yahoo! Answers service as *topic categories* and to those generated within the Live QnA service as *community tags*.

2.1 Related work

Community Q&A services have grown in popularity over the last couple of years, greatly due to the success of Yahoo! Answers. The research community has also gained interest in investigating various aspects of this service, leading to a number of studies reported over the past year [1,2,7,8,10]. There has been a great emphasis on identifying and predicting quality answers [1,2,8], and modelling users authority [7,10].

Adamic et al. [1] analysed Yahoo! Answers social network interactions on three topic categories, identifying users with similar behaviour to the 'answer-person' role verified in newsgroup communities [14]. Agichtein et al. [2] provided a classification model for estimating answer quality based on features derived from the content and also authority measures from the social network. Gyongyi et al. [7] and Jurczyk et al. [10] applied variants of Kleinberg's HITS algorithm to the Q&A social network graph to model user reputation and level of expertise.

There have been numerous studies on tagging systems that have analysed collaborative tagging in online services. Golder et al. [6] observed convergence in tag usage patterns over time for the *Del.icio.us* collaborative bookmarking tags. Marlow et al. [11] proposed a framework to characterize collaborative tagging systems into several dimensions, such as tagging support or social connectivity, among others, with the aim to inform the design of new services.

3. Datasets

The analyses we present in this paper are based on two datasets. The first one was obtained from the Live QnA service and spans the first year of its beta release (Sep. 2006 until Sep. 2007). It consists of 488,760 questions, 1,330,819 answers and 901,752 comments. The questions were submitted by 241,616 unique users, while the answers and comments were contributed by 42,941 and 34,068 unique users, respectively. On average, 45.5% ($\pm 16.2\%$) of active users on each day answered questions. Figure 2 shows the ratios of new questions, answers and comments per day with respect to the overall daily content over a 1-year period. These plots enable us to analyze the relative growth rate of each type of content contribution. Increased growth rates for all three types of content were observed starting around Dec. 2006, which coincided with a new release of the Live QnA service [11].

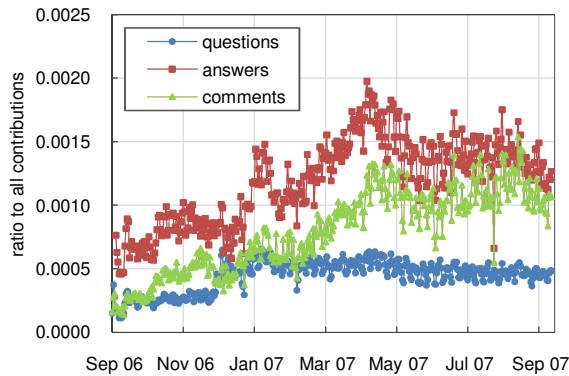


Figure 2. Ratio of questions, answers and comments to the overall content contributions per day, on Live QnA.

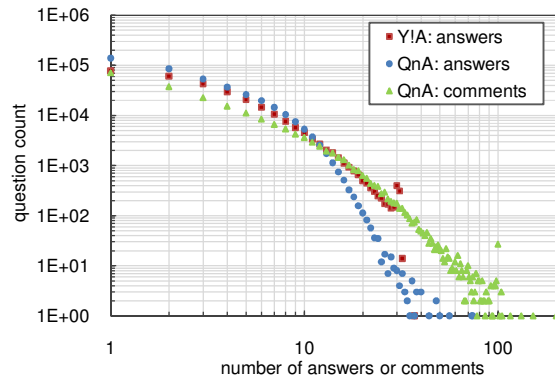


Figure 3. Distributions of the number of answers (on Y!A and QnA) and comments (in QnA) per question.

Questions reached a steady growth after Jan. 2007, while answers and comments sustained an increasing growth rate over a longer period of time.

The second dataset was gathered from the Yahoo! Answers service by seeding a crawler with pages linked to the top level categories that list recent questions with the assigned category and sub-categories. Overall we crawled 309,599 questions, posted by 217,615 distinct users and 1,151,453 answers, given by 202,052 distinct users. On average, 72.5% ($\pm 17.1\%$) of users active in each day answered questions. Over 95% of the content that we crawled was posted during the 3-month period of March-May 2008. Statistics about the evolution of Yahoo! Answers in its early days can be found in [7].

Figure 3 presents the distributions of answers per question and comments per answer, for both services. The responsiveness to individual questions varies widely: in Live QnA 15.4% of questions did not receive any answer and 55.8% of questions received two or more answers within the 1-year observation period. In Yahoo! Answers 5.8% of questions did not receive any answer and 69.2% received two or more

answers within a month after being posted. We expect that the reasons for answering questions, or not, have commonality with behaviour observed in newsgroup communities [3] and include the rhetoric of the content, topic coherence, and linguistic complexity of the question. It is interesting to observe that, on Live QnA, the affordance of interacting with other users through comments at any time during the question lifecycle, led to 36% of the answers getting at least one comment and to 14.6% question threads containing comments on every single answer.

4. Community Tags vs. Topic Categories

Users who are new to the Q&A communities may choose to browse or search for questions on a given topic rather than post their own question. They may be able to satisfy their information need from answers that were given in the past. Live QnA and Yahoo! Answers facilitate browsing questions based on tags or topic categories that were assigned to questions by the users who asked them. However, the assignment process differs significantly. On Yahoo! Answers, users select a topic from the predefined taxonomy and assign only one category to the question. The taxonomy comprises 26 top level categories, each branching into $11(\pm 6.9)$ second level categories on average. On Live QnA, users can assign one or more tags, without constraints on the vocabulary. Thus, the set of tags increases over time and individual questions can be accessed via multiple tags by browsing the evolving tag vocabulary.

In the following sections we analyze the usage of community tags and topic categories in Live QnA and Yahoo! Answers, respectively. We study the relations among community tags in Live QnA and characterize their topic coverage by classifying tagged questions into the Yahoo! Answers taxonomy. This provides further insights in the nature of community tags.

4.1 Analysis of Live QnA tags usage

In the Live QnA dataset questions were labelled with $2(\pm 2.3)$ tags on average. Overall, the community applied 188,468 distinct tags. Some of these tags were used very frequently, possibly due to the automatic recommendation of tags that is provided by the service. Among the 10 most frequently used tags, which are listed in Table 1, the technology-related ones ('Internet', 'Technology', 'Computers', 'Windows', and 'Microsoft') received 2.4 answers per question on average, whereas the remaining ones ('Fun', 'life', 'people', and 'Family') received 5.1 answers per question on average. This indicates that the community members responded more actively to questions on certain topics.

Table 1. Ten most frequently used Live QnA tags.

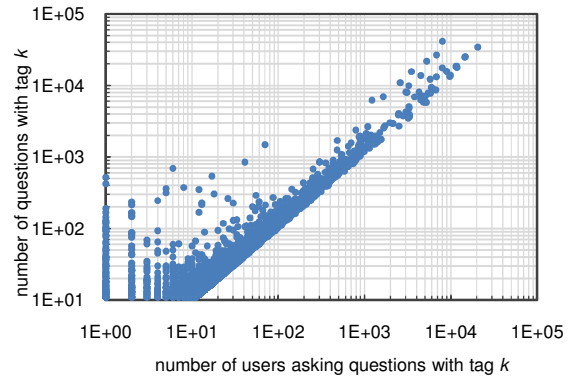
Tag Name	Number of Questions Tagged	Number of Answers Received
Fun	41,259	212,874
Internet	34,005	83,822
People	26,583	138,409
Technology	25,116	61,872
Computers	24,633	59,867
Life	21,739	124,064
Windows	18,499	42,078
Microsoft	18,343	42,557
Windows Live	17,644	40,045
Family	17,498	74,355

The varied number of answers may be due to various factors including the nature of the question or the community interest in a particular topic area. Indeed, for some topics, users may be predominantly asking for information, e.g., where to find a download site for a computer game or when to plant a particular type of flowers. Such questions can be addressed by a single or a couple of answers. This is in contrast with topics such as family issues where one may seek for advice and receive many answers covering multiple opinions and advice. It may also be that the community attracts people who are less technology savvy and thus do not provide answers to technical questions.

Overall, the Live QnA tags vary widely in the number of questions they were assigned to and the number of distinct users who applied them. From Figure 4 we observe that the common tags were used not just frequently but also by the majority of the user population. Most tags with question frequency above 100 were used by more than 100 distinct users. But, there are several outliers corresponding to tags that were applied often by very few users. For example, the tag ‘oreeeeeeeeeellly?’ was applied to 420 questions by a single user and the tag ‘dwayne’ was applied to 190 questions, also by a single user. While other commonly used tags tend to represent generic topics, similar to the Yahoo! Answers categories, these outliers illustrate a very personal and social use of the tags.

4.1.1 Topic relevance of community tags

The assignment of community tags may potentially be misused, e.g., due to a lack of understanding of their purpose, and thus lead to tags that do not describe the topic of the question. Given that tags are often used for indexing of the content and are recommended for question labelling based on their usage, it is important to understand how topical the community tags are. Thus, we decided to analyze samples of questions from the Live QnA and the Yahoo! Answers datasets. We asked assessors to inspect the tags, or topic category, of each question, and mark them as ‘on-topic’, ‘off-topic’ or otherwise of ‘unknown’ purpose.

**Figure 4. Number of questions classified with tag k vs. number of users tagging questions with tag k .**

In total, the assessors inspected 2,748 tags from the Live QnA sample, consisting of 1,060 questions, and 942 topic categories from the Yahoo! Answers sample, consisting of 471 questions. This revealed that 75.5% of the Live QnA tags were related to the actual topic of the questions and 23.1% were found to be off-topic, in contrast with 93.8% on-topic and 4.2% off-topic Yahoo! Answers categories. In both datasets, the fraction of undecided judgements was small, 1.4% and 1.9%, respectively. While this does not represent a significant statistical analysis of the two data sets, it does indicate that community tagging leads to practices that need to be carefully considered when designing search and tag recommendation features.

4.2 Inferring topics from tag co-occurrences

The guidelines provided by the Live QnA service advise the users to assign multiple tags to a question. They recommend using at least three tags per question to include general and specific terms and synonyms. They also suggest avoiding spelling mistakes, since that may hinder finding the question, and words that are already included in the question text. Essentially, tags are expected to be descriptive and discoverable.

Figure 5 shows the distribution of questions with a given number of tags. Despite the service guidelines, around 50% of questions were tagged with a single tag and just over 10% of questions are assigned three or more tags. Figure 6 shows, for every community tag, the average number of distinct tags that were co-assigned to the same question. We see that the most popular tags co-occur with 3 to 4 additional tags, on average. Looking at the tag co-occurrence in more detail we expect to gain further insights into the topicality of questions and emerging topics.

To that effect, we model tag associations by a directed graph, where vertices represent individual tags and edges represent co-occurrence.

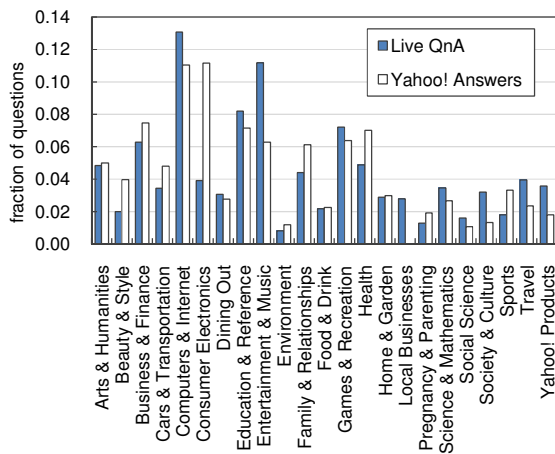


Figure 8. Fraction of questions per Yahoo! Answers topic category, for both datasets.

We represented each question as a vector comprising term frequencies of individual terms and word n -grams, i.e., 2- and 3-word terms co-occurring frequently in the Yahoo! Answers dataset. Prior to gathering statistics, we stemmed the text using the Porter’s algorithm and eliminated standard stop words for English language.

To assess the quality of the classifiers, we performed a 5-fold cross-validation on the set of questions from Yahoo! Answers. We applied one-vs-all approach for multi-class classification using SVM as the binary classifier. In order to account for imbalance between the positive and negative class in our data sample, we modified the SVM cost parameter to increase the penalty for misclassifying positive documents. For each topic category we ranked the classified questions based on the SVM score and calculated the break-even-point (BEP) for the ranked list – the BEP value indicates the rank at which the classification precision and recall are equal. The average BEP obtained across all the Yahoo! Answers categories was 69.8% ($\pm 19.6\%$). This result gave us confidence that we can reasonably predict a topic category for each question, despite the fact that many questions typically offer short snippets of text.

Assuming that the users of both services come from the same population of online users who ask questions, we applied the classifiers to the Live QnA questions and use the resulting assignment of topic categories as the basis of the further analysis. Figure 8 shows the distribution of questions, from Yahoo! Answers and from Live QnA, over the top level topic categories of the Yahoo! Answers taxonomy.

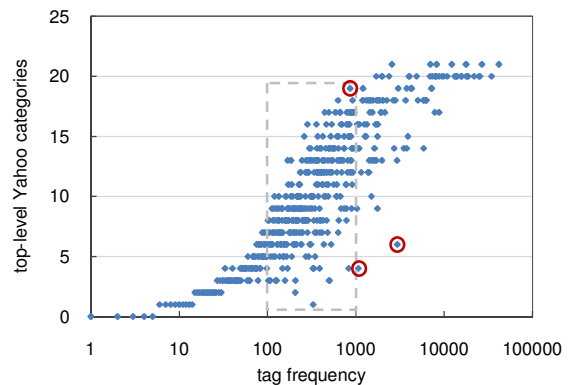


Figure 9. Tag frequency (log scale) vs. number of related top-level Yahoo! Answers categories. The circled data points correspond to the ‘curious’ tag (top), ‘ipod’ tag (bottom left) and ‘windows vista’ tag (bottom right).

4.3.2 Topic analysis

For each community tag used in Live QnA we can now track the distribution of its questions across the Yahoo! Answers topic categories. From Figure 9 we observe that questions labelled with frequently used community tags (frequency > 1,000) are generally classified into many distinct Yahoo! categories – typically 10 or more topics. Those questions that are assigned tags with overall frequency between 100 and 1000 can vary significantly in the number of topic categories they cover. Thus, if we consider the spread across Yahoo! Answers topic categories as an indicator of a tag’s specificity, or generality, the community tags vary significantly in that respect. That could be attributed to the recommended practice by the Live QnA service to use multiple tags and ensure they cover general and specific terms. Nevertheless, as we noted in Section 4.2, only 10% of Live QnA questions have more than 2 tags assigned. Thus, we expect that other factors play a role.

Here are some examples of the variability in the tag specificity. The Live QnA questions tagged with the tag ‘ipod’, 1,066 questions in total, were classified onto 29 second-level Yahoo categories that spanned 4 top-level categories. Similarly, 2,940 questions labelled with ‘windows vista’ were assigned to 44 second-level categories, under 6 top-level categories. At the same time, questions tagged with ‘curious’, 860 questions in total, were classified onto 114 second-level categories, spanning 19 top-level categories. Thus, one could argue that ‘ipod’ and ‘windows vista’ are generally better descriptors of the question topic than the ‘curious’ tag which, in fact, could be applied to questions on numerous topics.

5. Social interactions in Q&A services

While designed primarily to facilitate answering questions, the Q&A services are based on the premise that their communities are formed, active, and self sustainable. Thus it is not surprising that some members seek to communicate and connect with others asking questions, such as, ‘How are you?’ or ‘I’m eating a slice of home-made pie. Anyone wants some?’. This behaviour does not comply with the intended use of the service but aims to engage with and perhaps entertain the community. We now explore how that is reflected in the individuals’ tagging practice and users interaction patterns.

5.1 Analysis of question types

Through manual inspection of the random samples from Live QnA and Yahoo! Answers (see 4.1.1) we identified several types of questions, including (a) *information seeking* – requesting information about a fact or a resource that can satisfy the user information need, (b) *opinion seeking* – requesting an opinion about a topic, possibly of personal nature, and (c) *chit-chat* – question instigating community reaction for the purpose of socializing. We observed that information seeking questions were predominant in both datasets: 62.3% on the Live QnA sample and 78.1% on the Yahoo! Answers sample. The percentages of opinion seeking questions were also comparable: 19.0% on Live QnA and 15.3% on Yahoo! Answers. However, we found a substantial difference in the proportion of chit-chat questions: 14.2% on Live QnA and 3.6% on Yahoo! Answers.

Considering the Live QnA questions that are associated with the 10 most frequent community tags we observed the frequency of question types across tags. Figure 10 shows that tags referring to technology and computer-related topics were predominantly associated with questions of the information seeking type. In contrast, tags like ‘Fun’, ‘People’ and ‘life’ were mostly associated with chit-chat questions. The ‘Fun’ tag, in particular, is highly correlated with this type of question.

5.2 Community tags & social network activity

With new insights about the question types and community tags, we investigate the properties of the social network that emerges from answering questions with specific tags. We performed an analysis of the *answer-to* social network derived from the Live QnA data set (see Section 3). In such network the nodes correspond to active users and the directed edges indicate that, for example, a user A has responded to a question of the user B.

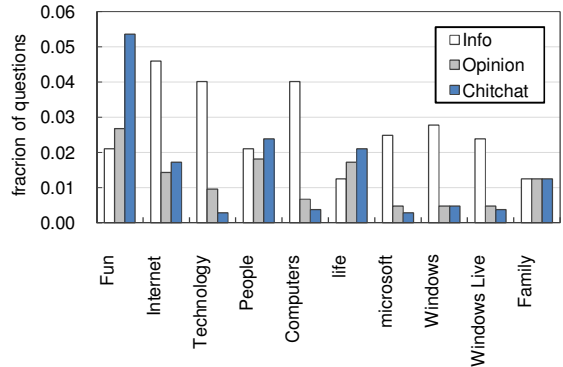


Figure 10. Distribution of labelled questions across the top 10 most frequent Live QnA tags.

Looking at the in- and out-degree distributions we observe the typical power-law with few users having many social ties, i.e. answering to many users, and many users having very few social ties, i.e., answering to few users or none. For each community tag we consider the corresponding sub-graph of the answer-to network and calculate the *density of the sub-graph* to assess the strength of the social ties among the involved users. The graph density measures how close a subset of vertices is to forming a clique (i.e., to include the maximal number of edges):

Definition. For a directed graph with $|E|$ edges and $|V|$ vertices, the graph density is defined as $D = \frac{|E|}{|V| \cdot (|V| - 1)}$.

We examined sub-graphs consisting of the 100 most active ‘answers’ and 100 most active ‘questioners’ for each of the top ten Live QnA tags. In Table 2 we show for each tag sub-graph the overlap between the top questioners and top answerer ($V_Q \cap V_A$) and the density of the sub-graphs associated with answerers (D_A) and questioners (D_Q).

Table 2. Density of the social network resulting from *answer-to* interactions between top answerers (D_A) and questioners (D_Q), on the specified tag.

Tags	Questions	H_t	D_A	D_Q	$V_Q \cap V_A$
Fun	41,259	2.981	0.588	0.613	52%
Internet	34,005	2.709	0.243	0.255	31%
People	26,583	2.646	0.450	0.459	42%
Technology	25,116	2.190	0.092	0.089	17%
Computers	24,633	2.063	0.092	0.088	21%
Life	21,739	2.595	0.365	0.357	38%
Windows	18,499	1.942	0.067	0.066	19%
Microsoft	18,343	1.919	0.066	0.069	16%
Windows Live	17,644	1.987	0.107	0.120	27%
Family	17,498	2.602	0.307	0.326	40%

We note that the community exhibits different behaviour across tags. For tags like ‘Fun’, ‘People’ and ‘Family’, a high percentage of users who post questions also engage very actively in giving answers to other users, indicating that there are sub-communities of active users formed around such tags. Furthermore, the density of the sub-graphs hints that a particular type of questions may be pre-dominant for a given tag.

For example, the density values for the ‘Fun’ tag indicate that highly active users interact with a large proportion of other highly active users and thus support our hypothesis that ‘Fun’ tag is associated predominantly with chit-chat questions. We can contrast that with density values for tags like ‘Microsoft’ or ‘Windows’, which are significantly lower. Furthermore, the low overlap between top answerers and top questioners for these tags is more typical of information seeking communities where expert users provide the most answers [14]. We also found that the social network density for the top 10 tags is very strongly correlated with the tag co-occurrence entropy H_t , with a 0.95 correlation value.

6. Conclusions

The Live QnA and Yahoo! Answers services facilitate question answering by online communities of users. They provide different mechanisms for users to label their questions and make them easily discoverable by the community. The Live QnA service allows users to generate their own tags while the Yahoo! Answers requires that users select one of the pre-defined topic categories.

In this paper we analyze the two different models and study the implications on the labelling practices and the interaction patterns within the communities. We found that the tags generated by the Live QnA community reflect both the social interactions among users and the topic of the questions. In fact, we hypothesise that the freedom to create custom tags has led to the possibility of disseminating questions that are of social nature, and vice versa, that the variety of social interactions have influence the evolution of the community taxonomy.

In our study we introduced two methods: the tag co-occurrence graph representation for analyzing the emerging topics of the social interaction and the graph density analysis to describe the strength of the user communities around specific community tags.

We conclude that the strong social aspect in Live QnA question-answering and the low discriminating value of individual community tags require special care when designing navigation or recommendation features. This contrasts with the use of pre-defined

topic categories from the Yahoo! Answers taxonomy that comprises topic categories that are well defined and widely adopted.

In our future work, we shall investigate the design principles of community services, including the mechanisms required to maintain the sense of community and facilitate effective discoverability of questions via tagging mechanisms.

7. References

- [1] Adamic, L. A., Zhang, J., Bakshy, E., and Ackerman, M. S., Knowledge sharing and yahoo answers: everyone knows something. In *Proc. of the 17th Int. Conf. on the World Wide Web*, WWW '08, pp. 665-674.
- [2] Agichtein, E., Castillo, C., Donato, D., Gionis, A., and Mishne, G., Finding high-quality content in social media. In *Proc. of the Int. Conf. on Web Search and Web Data Mining*, WSDM '08, pp.183-193, 2008.
- [3] Arguello, J., Butler, B. S., Joyce, E., Kraut, R., Ling, K. S., Rosé, C., and Wang, X., Talk to me: foundations for successful individual-group interactions in online communities. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI '06, pp. 959-968.
- [4] Cortes, C. and Vapnik, V. Support vector networks. *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [5] *Text Garden*, Grobelnik, M. and Mladenic, D., available at: <http://www.textmining.net/>
- [6] S. A. Golder and B. A. Huberman. The Structure of Collaborative Tagging Systems, In *Journal of Information Science*, 2006.
- [7] Gyongyi, Z., Koutrika, G., Pedersen, J., Garcia-Molina, H., Questioning Yahoo! Answers. *First Workshop on Question Answering on the Web*, held at WWW '08.
- [8] Harper, F. M., Raban, D., Rafaei, S., and Konstan, J. A., Predictors of answer quality in online Q&A sites. In *Proc. of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, CHI '08.
- [9] Joachims, T., “Text Categorization with Support Vector Machines: Learning with Many Relevant Features,” In: *Proc. of ECML'98*, Nédellec, C. and Rouveirol, C. (editors), *Lecture Notes in Computer Science*, vol. 1398, pp. 137-142, Springer-Verlag, 1998.
- [10] Jurczyk, P. and Agichtein, E., Discovering authorities in question answer communities by using link analysis. In *Proc. of the 16th ACM Conf. on Information and Knowledge Management*. CIKM '07, pp. 919-922.
- [11] Live QnA Blog: <http://liveqna.spaces.live.com/blog/>
- [12] C. Marlow, M. Naaman, D. Boyd and M. Davis. Tagging Paper, Taxonomy, Flickr, Academic Article, To Read. In *Proc. of ACM Hypertext 2006 Conference*, HT'06.
- [13] Naver portal: <http://www.naver.com>.
- [14] Welsler, H.T., Gleave, E., Fisher, D., Smith, M. Visualizing the Signatures of Social Roles in Online Discussion Groups. *Journal of Social Structure*, no. 8, vol. 2.