Knowledge Sharing in Online Discussion Threads: What Predicts the Ratings?

Yuyang Liang
Michigan State University
East Lansing, USA
liangyuy@msu.edu

ABSTRACT

As an important category of user-generated content (UGC) community, Question and Answer (Q&A) community offers internet users opportunities to ask questions and share knowledge with others. In order to understand how the ratings of knowledge contribution quality correlate with the way knowledge is being shared in discussion threads, the study examines user behaviors and profiles in a large knowledge sharing community, `/r/Techsupport`, a discussion based Q&A site in Reddit.com concerning internet and technology problems. Negative binomial regressions and negative binomial mixed models are built to investigate the relationships among thread structure, level of user activity, user profiles and the ratings of threads and comments in the community. Results indicate that in the better rated threads, the structures tend to be more centralized with heterogeneous participants discussing the problem at a deeper level. Meanwhile, contributions with good ratings are more likely to be produced by users who are more engaged in commenting behaviors.

Author Keywords

Knowledge sharing; online community; user generated content; network structure; user profile; threaded discussion.

ACM Classification Keywords

H.4.3 [Information Systems Applications]: Communications Applications—Bulletin; H.5.3 [Information Interfaces and Presentation (e.g. HCI)]: Group and Organization Interfaces—Computer-supported cooperative work

INTRODUCTION

User-generated content (UGC) sites, such as Wikipedia and YouTube, have gained growing popularity and influence on the internet. Among these sites, there is a special type of community which offers platforms for internet users to ask questions as well as provide various supports and is often called Question and Answer (Q&A) community. In some of these communities (e.g. Stack Overflow, Quora), members can evaluate the usefulness of the contribution by voting the content and contributions with more votes often rank higher on the page. In this way, good contributions can reach larger audiences on the platforms.

There is a large body of studies concerning popular Q&A communities such as Google Answers, Yahoo! Answers, and Stack Overflow [1,10,12,13,18,19,23,24]. The quality of the answers is associated with whether the site is free or fee-based [12], the length of the answer, and the track record and the reputation of the answerer [1,23,24]. However, little is known about factors that are associated with the quality of knowledge contributions in threaded discussions.

To understand how the structures and the dynamics of the discussions as well as user profiles are related to the ratings of knowledge contribution quality, this study examines user behaviors and profiles in a large knowledge sharing online community, `/r/Techsupport`, which is a discussion based Q&A site in Reddit.com concerning internet and technology problems, with approximately 82,000 users. This site is suitable for the analysis since it completely consists of threaded discussions, in the form of tree structure; thus various indices (e.g. centralization) can be easily obtained to describe the structure. Also, users are able to upvote or downvote the content based on their perceptions of the quality of the content.

The present study advances the extant research on online knowledge sharing by showing that in Q&A communities, the structure of the better rated thread tends to be more centralized with heterogeneous participants discussing the problem at a deeper level. Meanwhile, contributions with good ratings are more likely to be produced by users who are more engaged in commenting in the community. In addition, from a practical point of view, the study may help users to understand how online knowledge contribution is shaped by the structures and the dynamics of their interactions, hence promoting better knowledge sharing in online communities.

The article starts by surveying the literature on online knowledge sharing, and how the study may extend the scope of the related theories. Next, regression models are built to examine the relationships among the thread structures, the

CSCW '17, February 25-March 01, 2017, Portland, OR, USA

ACM 978-1-4503-4335-0/17/03…$15.00
DOI: http://dx.doi.org/10.1145/2998181.2998217

level of user activity, user profiles and the ratings of contribution. Based on the results, implications are discussed in detail to address the research questions.

LITERATURE REVIEW

Social Capital, Social Network, and Online Knowledge Sharing

Social Capital Theory is tightly related to Social Network Theory, which holds that this type of capital inheres in the social relations an individual possesses [7]. One of the definitions for social capital was given by Nahapiet and Ghoshal [20], which is "the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit" (p. 243). The theory suggests that the network of relationships is an important facilitator of interpersonal knowledge sharing and the intensity of the interaction is related to the depth, breadth, and efficiency of the sharing behaviors [16,20].

Several empirical studies have examined knowledge sharing behaviors within the context of online communities, for the purpose of understanding how the behaviors are influenced by their positions in the network [4,5,26]. Using a multimethod approach, Wasko and Faraj [26] found out that individuals with higher levels of network centrality, or the number of received or posted messages, tended to have more social capital and thus would contribute more responses with better quality.

In another similar study by Chiu, Hsu, and Wang [5], the norm of reciprocity, which indicates a sense of return or acknowledgment when receiving others' supportive responses, was found to be positively related to the quantity. However, this result contradicts the findings in the study of Wasko and Faraj [26], where reciprocity exhibited a negative impact on the volume of contribution.

These studies reveal the connection between social network and online knowledge sharing, which indicate that some aspects of the network structures, including centrality, reciprocity, may impact the way individuals exchange knowledge in online communities. Similarly, in threaded discussions, where networks are formed by posting messages, participants' behaviors and the quality of their contributions are also likely to correlate with the degree of centrality they possess, the norm of reciprocity, and other properties of the network. However, the main focus of the studies mentioned above is the motivation to contribute knowledge in online communities. How the quality of the contributions is affected by the actual interactions among participants and the way online discussions are formed and developed is still a bit of a puzzle.

Structure and Pattern of Online Discussion

One of the major forms of interactions among users in online communities is threaded discussions. Knowledge is exchanged and shared by posting and receiving messages in the communities. In some large online communities (e.g. Reddit, Slashdot), users can produce hundreds or even thousands of new posts during a single day, which greatly increases the complexity of the structure of the discussions. Such intricate structure may imply the underlying patterns of user activity in these communities. For instance, in Slashdot, comments with high scores or from good quality writers are more likely to attract follow-up replies, and the first two levels of a thread contain most of the comments and then the number of comments reduces as the depth increases [11]. Similarly, Choi et al. [6] characterized the online conversation patterns in Reddit, and found that a small proportion of users generate the majority of the comments in a large and viral discussion thread by reciprocally exchanging information to one another. The domination of highly active users can also be found in Q&A sites. These users are more likely to be closely connected with more mutual interactions [1,19], and their online reputation is earned through active participation in answering questions [19].

This line of research reveals extensive details about the configurations of threaded discussions in online communities, though many studies tend to be descriptive. Generally, the lifespan of a threaded discussion is relatively short, and active users may play an important role in the development of the discussion. As online discussions are often transient, it is of great interest to further examine whether a higher level of user activity is more likely to produce better communication results. Especially in Q&A sites, the timeliness of answers can be important to the asker. Also, it is not yet clear whether the quality of the discussion will be limited when the discussion is dominated by a small number of active users. Such examinations will help to advance the understanding of online discussions and improve the system performance.

Cultural Production in Online Fields

Levina & Arriaga [17] introduced Bourdieu's Theory of Cultural Production, a sociological theory, to establish a theoretical framework with respect to the social dynamics in online fields, or more specifically, user-generated content (UGC) communities. Basically, there are two key groups of users in the field, producers and consumers, each of whom possess their own power (resources and capital), and status (distinction and reputation). The dynamics between power and status play a vital role in shaping the community, as participants utilize the resources they have to obtain distinctions and higher status achieved by the users may lead to more power and rewards. Therefore, different producers may take advantage of different strategies to obtain recognitions in the community. Consumers are also affected by their profiles in evaluating the quality of the content. This framework helps to understand how user profiles and characteristics may influence the contribution they make, although it pays more attention to the users while the content is considered as the means to achieve status.

In online communities, the boundary between producer and consumer is often blurred, so anyone can be both producer and consumer at the same time. One common approach to identify salient user profiles is analyzing the log data. Across various platforms (e.g. Usenet, Yahoo Answers), the most common roles are askers (who start new threads and ask questions) and answerers (who largely respond to existing messages) [8,21]. In addition to these two profiles, Adamic et al. found another type of user as the discussion person who is active in both asking and answering questions [1]. However, these studies do not connect the profiles with the quality of the contributions.

To expand present knowledge about how user profiles and behaviors change over time, Furtado et al. [10] categorized contributors on Stack Exchange into ten clusters. Low-activity and occasional users contribute a large portion of the questions while active users are responsible for the majority of the answers and comments to existing questions. In comparison, experts only provide a small amount of content. The study also suggests that highly active users are not the users who consistently provide high-quality contributions, which is particularly interesting.

Hence, following previous research, another important question to ask is what kind of asker/answerer is more likely to make high quality contributions, which can be done by investigating their activity history. In fact, both meaningful questions and answers are critical in sustaining the development of Q&A communities. Identifying good asker and answerer will have benefits for community and information management.

Based on the discussion above, the present study hypothesized that the quality of knowledge contribution in online communities is associated with the following factors: (1) the structure of the threaded discussions; (2) the level of user activity; and (3) user profiles.

METHOD

Overview of Reddit and `/r/Techsupport`

Reddit is one of the largest UGC communities where users are able to create their own content by: (1) submitting a link to an external site; (2) writing a text post; (3) commenting on other users' posts or comments. Reddit also consists of thousands of subreddits, or sub-communities, which can be created by any user who is interested in certain topics. `/r/Techsupport` is one of the subreddits with the longest history in Reddit which is dedicated to providing supports for general computer and internet problems. Community members can either raise related questions or provide solutions. Once the asker thinks the issue is solved, the thread will be marked as "Solved" in the title. Also, each user can upvote or downvote the whole thread or a single comment regarding the value of the contribution. Some of the members carry a "Trusted" badge with their username to indicate that they have solid knowledge in the IT field and their skills are proven by the moderators of the community.

Figure 1 shows an example of a discussion thread in the community.

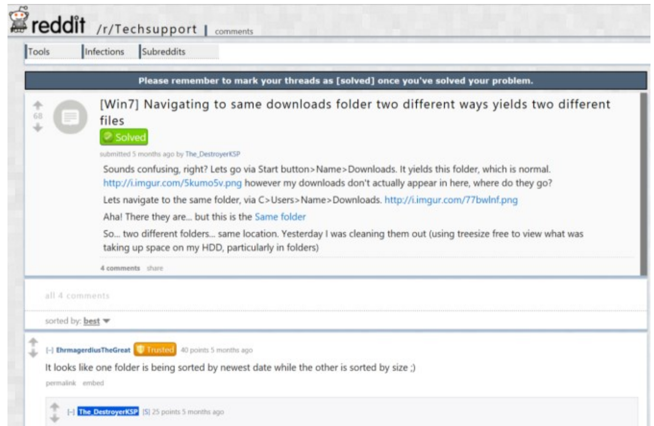

Figure 1. An example of a discussion thread in `/r/Techsupport`.

Data and Variables

The dataset used in the current study contains one year of user activities in the community, ranging from March 16, 2015 to March 15, 2016. Threads with no comments are eliminated in order to perform the multilevel analysis. In total, 286 threads and 16,512 comments are collected from the API (application program interface) offered by Reddit. The number of comments within each thread ranges from 7 to 640, with 44 as the median.

To operationalize the notion of the quality of contributions, the ratings of the threads (TR) and comments (CR) are used. The rating, or score, according to Reddit, is the difference between the number of upvotes and the number of downvotes, which can be negative. Therefore, a large number indicates that there are more users who like the contribution than users who dislike it, and these ratings can be a reasonable *proxy* of how helpful and valuable the content is based on the users' perceptions.

According to the nature of threaded discussions, there are two levels of variables. The first is the thread level variables, which is associated with each thread. In particular, degree centralization (CEN), reciprocity (REC) and maximum depth (MD) are calculated to describe the structure of a thread.

Centralization

The calculation of centralization is based on converting the entire thread into an ego network, where each poster is a node in the network and each comment is treated as a directed edge (tie). According to the definition given by Freeman [9], degree centralization is calculated as:

$$\frac{\sum_i [\max C(p_i) - C(p_i)]}{\max \sum_i [\max C(p_i) - C(p_i)]}.$$

The numerator denotes the sum of the differences between the node with the largest degree centrality and all other nodes, while the denominator gives the theoretical maximum for a network with the same number of nodes; thus the index is bounded between 0 and 1 and it measures how heterogeneous the degree centralities are. Higher centralization suggests that the most central node in the network has much higher degree centrality than all other nodes. In other words, the asker receives much more comments than all other answerers. Thus, the distribution of the comments (edges) is less even among the posters (nodes) than a thread with a lower centralization index.

Reciprocity

Reciprocity is given by the proportion of reciprocated ties in the ego network. If two users respond to each other by posting messages, they are considered as a pair of reciprocated ties. In adjacency matrix notation, reciprocity is calculated as $\sum_{ij}(A \cdot A^T)_{ij} / \sum_{ij} A_{ij}$, where $A \cdot A^T$ is the element-wise product of the adjacency matrix A and its transpose. Therefore, the numerator calculates the number of reciprocated ties while the denominator is the total number of ties in the network. The index also ranges between 0 and 1. Thus, greater reciprocity means the participants in the discussion frequently reply to one another.

	Min.	Max.	Mean	SD
TR	32	577	68.18	57.79
ALK	1	32,2185	4072	21,723
ACK	0	32,7949	8261	24,430
AT	2,654,804	319,899,719	83,691,134	51,914,806
REC	0	.92	.40	.19
CEN	.08	.73	.34	.14
MD	2	10	6.83	2.30
SOL	0	1	.31	-

Table 1. Descriptive statistics for the thread level variables.

Maximum Depth

Besides an ego network, a threaded discussion can also be represented as a tree structure, where the top-level post is the root of the tree. Direct comments to this post are the first level nodes and subsequent comments continue to increase the nesting levels. The maximum depth of a thread is the deepest level of nested comment in the comment tree structure, which may reflect the deepness of the discussion or the level of controversy.

To illustrate asker profiles, the asker link karma (ALK), comment karma (ACK) as well as tenure (AT) are obtained. Karma is simply the score users received in their previous contributions and shows the quality and popularity of their previous contributions. As the name suggests, link karma is the score received by submitting links while comment karma concerns posting comments. For example, when a user submitted a link (or posted a comment) to Reddit and received 320 upvotes and 20 downvotes, the user will gain 300 link (comment) karma. It is worth noting that when writing a text-only post (e.g. describing a problem needs to be solved), the user will not receive any link karma even though the post can still be rated by other users, which is the case in this particular community. In addition, the user tenure indicates the age of the account (in second). Moreover, whether the problem is solved (SOL) is also included as a control variable. Table 1 presents the descriptive statistics for the thread level variables.

	Min	Max	Mean	SD
CR	-60	374	3.85	10.81
CLK	1	742,873	2,928	12,741
CCK	-100	611,949	15,582	33,673
CT	925,504	334,540,800	103,864,895	60,734,104
TRU	0	1	.045	-
NNC	0	105	.67	1.30
RT	5	15,500,000	177,500	1,173,614

Table 2. Descriptive statistics for the comment level variables.

The second set of variables is the comment level ones. Likewise, commenter profiles include the commenter link karma (CLK), comment karma (CCK), and tenure (CT, in second). Besides, an indicator of whether the user is a trusted expert or not (TRU) is also included in the profiles.

To illustrate the level of user activity in the threads, in terms of the intensity and frequency of interactions, two variables are obtained: the number of nested comments (NNC) and response time (RT, in second). Specifically, the number of nested comments for a particular comment only includes comments that are directly connected to it, thus excluding the replies to the nested comments. Response time is the time difference between a certain comment and its follow-up reply. Table 2 summarizes the descriptive statistics for the comment level variables.

In order to better perform the analysis and obtain reliable results, some of the variables are transformed. First, since ALK, ACK, AT, CLK, CCK, CT, and RT are highly skewed, with their variances much larger than their means, log transformations are applied to these variables (represented as TALK, TACK, TAT, TCLK, TCCK, TCT, TRT). Specifically, in order to apply the log transformation, ACK and CCK are shifted by subtracting their own minimums (0 and -100, respectively) and adding one to the values to avoid zero or negative numbers. Meanwhile, CR is also shifted by subtracting the minimum (-60) to avoid negative values (represented as TCR).

The Pearson correlation coefficients are calculated after the transformation is finished, in order to examine the potential collinearity problem. Table 3 and Table 4 present the coefficients for thread level variables and comment level variables, respectively. Although some of the variables exhibit relatively high pairwise linear correlations (e.g. 0.69 between TALK and TACK), the results of multicollinearity test indicate that there is no significant multicollinearity issue within the variables, as the largest variance inflation factor (VIF) is less than 2.5 for the thread level and comment level variables.

	TR	TALK	TACK	TAT	CEN	REC
TALK	-.14					
TACK	-.05	.69				
TAT	-.15	.48	.50			
CEN	.03	.07	.15	.02		
REC	-.07	.22	.28	.16	.53	
MD	.18	.07	.09	.03	-.01	.42

Table 3. Pearson correlation coefficient for the thread level variables.

	TCR	NNC	TCLK	TCCK	TCT
NNC	.36				
TCLK	.02	.02			
TCCK	.04	.02	.62		
TCT	.02	-.003	.49	.51	
TRT	-.10	-.15	-.08	-.13	-.05

Table 4. Pearson correlation coefficient for the comment level variables.

Table 5 is a summary of the variables used in the regression models.

Analysis

For the purpose of examining how the thread structure, the level of user activity, and user profiles predict the ratings of the threads and comments, negative binomial regressions are applied on the thread level while negative binomial mixed models are built to analyze the comment level data. In fact, negative binomial distributions are often used to model count data, especially for dependent variables with over-dispersion. In this case, both thread ratings and comment ratings can be treated as counting the number of net votes (the number of upvotes minus the number of downvotes). Also, both ratings are excessively dispersed, whose means are much smaller than the variances. Hence, negative binomial models are suitable for fitting these particular data. For each thread j, the rating is modeled as:

$$\log\big(E(TR_j)\big) = \beta_0 + \beta_1 TALK_j + \beta_2 TACK_j + \beta_3 TAT_j + \beta_4 SOL_j + \beta_5 CEN_j + \beta_6 REC_j + \beta_7 MD_j \,.$$

Variables	Description	
TR	Thread ratings	Quality of contribution
TCR	Transformed Comment ratings	
TALK	Transformed asker link karma	Asker profile
TACK	Transformed asker comment karma	
TAT	Transformed asker tenure (in second)	
TCLK	Transformed commenter link karma	Commenter profile
TCCK	Transformed commenter comment karma	
TCT	Transformed commenter tenure (in second)	
TRU	Trusted user	
CEN	Centralization	Thread structure
REC	Reciprocity	
MD	Maximum depth	
NNC	Number of nested comments	Level of user activity
TRT	Transformed response time (in second)	
SOL	Solved problem	Control

Table 5. Summary of variables used in the regression models.

It is likely that the comments in a high-score thread may generally receive better ratings compared to those in a thread with a lower score. To account for this between-cluster effect, a mixed model with a random intercept is introduced. For each comment i in thread j, the rating is modeled as:

$$\log\big(E(TCR_{ij})\big) = \gamma_{00} + u_{0j} + \gamma_{10} TALK_j + \gamma_{20} TACK_j + \gamma_{30} TAT_j + \gamma_{40} SOL_j + \gamma_{50} CEN_j + \gamma_{60} REC_j + \gamma_{70} MD_j + \gamma_{80} TCLK_{ij} + \gamma_{90} TCCK_{ij} + \gamma_{100} TCT_{ij} + \gamma_{110} TRU_{ij} + \gamma_{120} TRT_{ij} + \gamma_{130} NNC_{ij} \,,$$

where u_{0j} represents the random effect.

RESULTS

Negative Binomial Models on Thread Ratings

Table 6 reports the results for the negative binomial regression models (numbers in the parentheses are the standard errors for the estimates). The first model only includes user profiles as predictors, all of which are significantly associated with the thread ratings (TR). In particular, on average, one unit increase in the transformed asker link karma (TALK) leads to 0.038 decrease in the log of the thread rating, with other variables held constant. On the other hand, the log of the thread rating is expected to increase by 0.034 with one unit increase in the transformed asker comment karma (TACK) and other variables remain unchanged. One unit increase in the transformed asker tenure (TAT) reduces the expected value of the log of the thread rating by 0.095 when other variables remain constant.

The second regression model specifically examines the correlation between the thread structure and the thread ratings. Again, all of the variables show significant impacts. With other variables held constant, if the centralization index (CEN) increases by 0.1, the log of the thread rating is expected to increase by 0.098. However, 0.1 increase in the reciprocity index (REC) will result in 0.119 decrease in the expected value of the log of the thread rating, with other variables held constant. On average, one level increase in the maximum depth (MD) raises the log of the thread rating by 0.104 if other variables remain unchanged.

The third model combines all of the independent variables and the results suggest that all but asker comment karma and tenure remain significant, with the same directions of effect and similar estimates. The chi-square goodness of fit tests for all three models are not significant, with p-values equal to 0.32, 0.35, and 0.30, respectively, indicating that there is no sufficient statistical evidence to reject the null hypothesis that the model fits the data. Also the estimated dispersion parameters $\hat{\theta}$ are 3.99, 4.57, and 4.81, respectively, which are much greater than 1 and suggest that there is indeed an over-dispersion issue with the dependent variable and thus negative binomial models are more proper. Compared to model (1) and model (2), the full model, or model (3), has the lowest AIC (Akaike information criterion) value (2,751 vs. 2,797 and 2,757), suggesting that the full model fits the data better.

Negative Binomial Mixed Model on Comment Ratings

The results for the negative binomial mixed model are summarized in Table 7 (numbers in the parentheses are the standard errors for the estimates). Among all of the thread level predictors, only reciprocity shows a significant negative association with the log of the transformed comment rating (TCR). In other words, as the reciprocity index for a certain thread increases by 0.1, the expected value of the log of the transformed comment rating within that particular thread will decrease by 0.006, with other variables held constant.

	Dependent variable: TR		
Predictors	**Estimates**		
	(1)	(2)	(3)
TALK	-.038**		-.028*
	(.013)		(.012)
TACK	.034*		.027
	(.017)		(.016)
TAT	-.095*		-.061
	(.046)		(.043)
CEN		.979***	.854***
		(.245)	(.241)
REC		-1.194***	-1.063***
		(.209)	(.210)
MD		.104***	.095***
		(.014)	(.014)
SOL			.118
			(.061)
Intercept	5.888***	3.633***	4.702***
	(.796)	(.114)	(.747)
χ^2	292.87	290.59	290.06
	(p = .32)	(p = .35)	(p = .30)
$\hat{\theta}$	3.989	4.569	4.813
AIC	2797	2757	2751
n		286	

*** $p < .001$; ** $p < .01$; * $p < .05$

Table 6. Results for the negative binomial regression models.

With respect to the comment level predictors, the effects of the transformed commenter link karma (TCLK) and comment karma (TCCK) are similar to the ones for the asker. For one unit increase in the transformed commenter link karma, the log of the transformed comment rating is expected to decrease by 0.0011, while one unit increase in the transformed commenter comment karma leads to 0.0017 increase in the expected value of the log of the transformed comment rating, when other variables stay the same. Moreover, if the commenter is a trusted user (TRU), the log of the transformed comment rating is expected to be 0.03 greater than the one from a commenter who is not a verified expert.

Both the number of nested comments (NNC) and response time (TRT), representing the level of user activity, significantly predict the log of the transformed comment rating, with opposite directions. Specifically, on average, one additional nested comment increases the log of the transformed comment rating by 0.021 with other variables being constant. One the other hand, one unit increase in the transformed response time reduces the expected value of the log of the transformed comment rating by 0.007 when other variables remain unchanged.

The estimated standard deviation for the random effect ($\hat{\sigma}^2_{u_0}$) is 0.22, which denotes the variation in the log of the transformed comment ratings between the threads. Compared to the effects of the predictors, the random effect is relatively large and therefore the between-thread variations can be substantial. The estimated dispersion parameter $\hat{\theta}$ for the negative binomial mixed model is 1.21, suggesting that there exists over-dispersion in the dependent variable.

Dependent variable: TCR			
Thread level		Comment level	
Predictors	Estimates	Predictors	Estimates
TALK	-.0002	TCLK	-.0011*
	(.0008)		(.0005)
TACK	-.0004	TCCK	.0017*
	(.0010)		(.0008)
TQT	-.0046	TCT	.0034
	(.0027)		(.0018)
CEN	0.0132	TRU	.0304***
	(.0156)		(.0052)
REC	-.0557**	NNC	.0207***
	(.0144)		(.0004)
MD	-.0007	TRT	-.0070***
	(.0010)		(.0006)
SOL	.0062	Intercept	4.2558***
	(.0039)		(.0545)
$\hat{\sigma}^2_{u_0}$	.0005 (s.d. = .0218)		
$\hat{\theta}$	1.211		
AIC	116180		
n_t	286		
n_c	16512		

*** $p < .001$; ** $p < .01$; * $p < .05$

Table 7. Results for the negative binomial mixed model.

DISCUSSION

The current study examines how the thread structure, the level of user activity and user profiles predict the ratings of the contributions in a knowledge sharing online community. As the results show, all of the three groups of factors exhibit significant associations with the ratings of the contributions.

In this study, the structure of a thread is operationalized as the reciprocity index, the centralization index and the maximum depth. First of all, the proportion of reciprocated messages in the thread is found to have a negative association with the ratings of the whole thread and the comments within the thread. In fact, research on Social Capital Theory in the context of knowledge sharing suggests that, the norm of reciprocity increases the quantity and quality of knowledge sharing since participants' time and effort spent in online community are being recognized and justified if they receive replies from others [3,5]. Joyce and Kraut [14] also found in public newsgroups, the probability of posting a second or more messages was higher for the newcomers who received a reply. However, in this case, a highly reciprocated thread tends to receive fewer upvotes, or be regarded as less useful by Reddit members. One possible explanation is that the number of participants in highly reciprocated discussions is limited. Though the mutual replies may encourage these participants to make further contributions, in fact, the value of these contributions is only pertinent to a small number of users. In contrast, according to Critical Mass Theory, the level of heterogeneity in a group is positively associated with the group's long-term success [22]. Similarly, in less reciprocated knowledge sharing threads, the participants and contributions are more diversified, which is likely to produce more collective good; hence more users are able to obtain benefits from the discussions. By attracting a larger number of participants, the threads are more likely to receive more upvotes from the users.

There is a positive connection between the rating of a Q&A thread and the degree centralization. Usually, in this type of discussion thread, centralization reflects how central the asker is in relation to how central all other participants are. Hence, greater centralization suggests that the asker receives more replies and posts more follow-up comments compared to other participants. In those threads with top ratings, the flow of information tends to focus on the asker. A discursive, or decentralized, thread structure can reduce the value of the discussion as the replies do not directly answer the main question. Hence, the moderators of the community may need to provide proper guidance to reduce the number of irrelevant topics in the threads.

The maximum depth of a thread is positively related to its rating, which means that as the discussion develops into a deeper level, it tends to receive more upvotes. In an online Q&A community, like `/r/Techsupport`, maximum depth may indicate that based on previous comments, there are follow-up questions, or users are actively contributing additional knowledge, which is likely to promote the quality of knowledge sharing.

In terms of user profiles, the study shows that users with higher comment karma tend to produce questions and comments with higher ratings, so good askers and answerers are probably those users who are more engaged in commenting behaviors. Actually, they can be considered as "content creators", who more often produce their own original content in the community. Specifically, in `/r/Techsupport`, most of the posts are self-created (other than submitting links to other sites), therefore those "content creators" are more likely to make better quality contributions. On the other hand, link karma is negatively associated with the ratings of the threads as well as the comments. Users with higher link karma tend to be "content transporters", who more often engage in transporting content from other sources to Reddit by submitting links. These users might be less often engaged or interested in commenting others' posts or creating original content, and thus their replies are more likely to be considered as less helpful when compared to those created by the content creators.

The difference between these two types of users in this particular community partially reflects Bourdieu's notion of habitus, which suggests that agents (i.e., users) have the motivation and ability to maintain and advance the cultural distinction achieved in their prior history [2]. More specifically, content creators have accumulated a large comment karma by posting good quality replies in the community and therefore they may continue to do so in the future. Nonetheless, since Reddit is a content community with low self-presentation based on the framework proposed by Kaplan and Haenlein [15], users on this platform may not be particularly searching for distinction. Still, their contributions are indeed related to their previous activities.

Furthermore, response time and the number of nested comments are used to describe the level of user activity in terms of the frequency and intensity of interactions. In particular, comments with shorter response interval tend to rank higher. Online discussions are asynchronous and time sensitive [6,11], so faster replies are more likely to be noticed and evaluated. It also implies that timeliness may be a critical characteristic of an online knowledge sharing field, which is included as one of the properties of UGC community that appeal to consumers' taste [17]. Besides, it serves as a non-verbal social cue, which indicates users' level of engagement and attentiveness and thus facilitates the conversation [25]. Although fast replies are not always equal to good replies, moderators can encourage members to actively participate in the discussion to increase the potential of receiving timely and useful answers.

The study also finds a positive relationship between the number of nested comments and comment ratings, which suggests that good comments are often the ones with more follow-up replies.

CONCLUSION

By empirically examining the threaded discussions in an online knowledge sharing community, the current study suggests that several factors are positively related to the ratings of contributions in this type of community: (1) the heterogeneity of the participants in the thread; (2) the high centrality of the asker; (3) the depth of the thread; (4) users who are more engaged in commenting and who are experts; (5) shorter response time; and (6) more nested comments. Further, the study also extends Bourdieu's cultural production theory by focusing on how the content in UGC communities is shaped by the power of the users, especially when there are no remarkable distinctions between producers and consumers, and when users do not specifically search for status in the community.

As the study suggests, expanding user base, encouraging more users (especially those experienced ones) to actively share their knowledge while limiting the number of off-topic comments may have a significant effect on improving the quality of the discussions and the performance of the community. These results may be generalized to other similar discussion forums where users are able to discuss their problems on specific topics (e.g., WebMD, WordReference).

Using ratings as the proxy for the quality of contribution may induce some potential problems. It is likely that a post receives many upvotes from other community members because it is amusing rather than being truly helpful. Besides, the scores users observed can influence their subsequent voting behaviors: they may upvote the high ranking posts and downvote the low ranking ones regardless of their quality. Hence, future studies can contextualize the messages using text mining or content analysis techniques to further operationalize the quality of the shared knowledge as well as investigate additional factors that may be associated with the quality. Also, previous interactions between users can be included as a measure of tie strength since the knowledge sharing behaviors may be more frequent and active between strong ties.

REFERENCES

1. Lada A. Adamic, Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. 2008. Knowledge sharing and yahoo answers: everyone knows something. In *Proceedings of the 17th international conference on World Wide Web*, ACM Press, 665. http://doi.org/10.1145/1367497.1367587
2. Pierre Bourdieu. 1990. *In other words: essays towards a reflexive sociology*. Stanford University Press, Stanford, Calif.
3. Hsin Hsin Chang and Shuang-Shii Chuang. 2011. Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. *Information & Management* 48, 1: 9–18. http://doi.org/10.1016/j.im.2010.11.001
4. Chih-Jou Chen and Shiu-Wan Hung. 2010. To give or to receive? Factors influencing members' knowledge sharing and community promotion in professional virtual communities. *Information & Management* 47, 4: 226–236. http://doi.org/10.1016/j.im.2010.03.001
5. Chao-Min Chiu, Meng-Hsiang Hsu, and Eric T.G. Wang. 2006. Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems* 42, 3: 1872–1888. http://doi.org/10.1016/j.dss.2006.04.001

6. Daejin Choi, Jinyoung Han, Taejoong Chung, Yong-Yeol Ahn, Byung-Gon Chun, and Ted Taekyoung Kwon. 2015. Characterizing Conversation Patterns in Reddit: From the Perspectives of Content Properties and User Participation Behaviors. In *Proceedings of the 2015 ACM on Conference on Online Social Networks*, ACM Press, 233–243. http://doi.org/10.1145/2817946.2817959
7. James S. Coleman. 1988. Social Capital in the Creation of Human Capital. *American Journal of Sociology* 94: S95–S120.
8. Howard Welser Danyel Fisher Marc A. Smith. 2006. You Are Who You Talk To: Detecting Roles in Usenet Newsgroups. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, Institute of Electrical and Electronics Engineers, Inc.
9. Linton C. Freeman. 1978. Centrality in social networks conceptual clarification. *Social Networks* 1, 3: 215–239. http://doi.org/10.1016/0378-8733(78)90021-7
10. Adabriand Furtado, Nazareno Andrade, Nigini Oliveira, and Francisco Brasileiro. 2013. Contributor profiles, their dynamics, and their importance in five q&a sites. In *Proceedings of the 2013 conference on Computer Supported Cooperative Work*, ACM Press, 1237. http://doi.org/10.1145/2441776.2441916
11. Vicenç Gómez, Andreas Kaltenbrunner, and Vicente López. 2008. Statistical analysis of the social network and discussion threads in slashdot. In *Proceedings of the 17th international conference on World Wide Web*, ACM Press, 645. http://doi.org/10.1145/1367497.1367585
12. F. Maxwell Harper, Daphne Raban, Sheizaf Rafaeli, and Joseph A. Konstan. 2008. Predictors of answer quality in online Q&A sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 865. http://doi.org/10.1145/1357054.1357191
13. Gary Hsieh, Robert E. Kraut, and Scott E. Hudson. 2010. Why pay?: Exploring how financial incentives are used for question & answer. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 305. http://doi.org/10.1145/1753326.1753373
14. Elisabeth Joyce and Robert E. Kraut. 2006. Predicting Continued Participation in Newsgroups. *Journal of Computer-Mediated Communication* 11, 3: 723–747. http://doi.org/10.1111/j.1083-6101.2006.00033.x
15. Andreas M. Kaplan and Michael Haenlein. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* 53, 1: 59–68. http://doi.org/10.1016/j.bushor.2009.09.003
16. Peter J. Lane and Michael Lubatkin. 1998. Relative Absorptive Capacity and Interorganizational Learning. *Strategic Management Journal* 19, 5: 461–477.
17. Natalia Levina and Manuel Arriaga. 2014. Distinction and Status Production on User-Generated Content Platforms: Using Bourdieu's Theory of Cultural Production to Understand Social Dynamics in Online Fields. *Information Systems Research* 25, 3: 468–488. http://doi.org/10.1287/isre.2014.0535
18. Zhe Liu and Bernard J. Jansen. 2014. Predicting potential responders in social Q&A based on non-QA features. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 2131–2136. http://doi.org/10.1145/2559206.2581366
19. Leydi Viviana Montoya, Athen Ma, and Raúl J. Mondragón. 2013. Social Achievement and Centrality in MathOverflow. In *Complex Networks IV*, Gourab Ghoshal, Julia Poncela-Casasnovas and Robert Tolksdorf (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 27–38.
20. J. Nahapiet and S. Ghoshal. 1998. Social Capital, Intellectual Capital, and the Organizational Advantage. *Academy of Management Review* 23, 2: 242–266. http://doi.org/10.5465/AMR.1998.533225
21. Kevin Kyung Nam, Mark S. Ackerman, and Lada A. Adamic. 2009. Questions in, knowledge in?: a study of naver's question answering community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 779. http://doi.org/10.1145/1518701.1518821
22. Daphne R. Raban, Mihai Moldovan, and Quentin Jones. 2010. An empirical study of critical mass and online community survival. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, ACM, 71–80.
23. Chirag Shah and Jefferey Pomerantz. 2010. Evaluating and predicting answer quality in community QA. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, ACM Press, 411. http://doi.org/10.1145/1835449.1835518
24. Yla R. Tausczik and James W. Pennebaker. 2011. Predicting the perceived quality of online mathematics contributions from users' reputations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 1885. http://doi.org/10.1145/1978942.1979215
25. Joseph B. Walther and Ulla Bunz. 2005. The Rules of Virtual Groups: Trust, Liking, and Performance in Computer-Mediated Communication. *Journal of Communication* 55, 4: 828–846. http://doi.org/10.1111/j.1460-2466.2005.tb03025.x
26. Molly McLure Wasko and Samer Faraj. 2005. Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly* 29, 1: 35–57.