Activity Lifespan: An Analysis of User Survival Patterns in Online Knowledge Sharing Communities

Jiang Yang[1], Xiao Wei[1], Mark S. Ackerman[1, 2], Lada A. Adamic[1, 3]

[1]School of Information, [2] Electrical Engineering and Computer Science, [3]Complex Systems
University of Michigan, Ann Arbor
Ann Arbor, MI, USA, 48109
{yangjian, xiaowei, ackerm, ladamic}@umich.edu

Abstract

Retaining participation is crucial for information services, online knowledge sharing services among them. We present the first comprehensive analysis of users' activity lifespan across three predominant online knowledge-sharing communities. Extending previous work focusing on initial interactions of new users, we use survival analysis to quantify participation patterns that can be used to predict individual lifespan over the long term. We discuss how cross-site differences in user participation and the underlying factors can be related to differences in system design and culture. We conduct a longitudinal comparison of the communities' evolvement between two distinct stages, the initial days just after the site launch and one year later. We also observe that sub-communities corresponding to different topics differ in their ability to sustain users. All results reveal the complexity and diversity in users' engagement to a site and design implications are discussed.

Introduction

Across the globe, peer-based Question-Answer (Q&A) communities have been rapidly accumulating knowledge and expertise to serve as vast knowledge repositories. Examples include Yahoo! Answers in English, Naver Knowledge iN in Korean, and Baidu Knows in Chinese. All these sites provide both demand and supply for people's knowledge and expertise, accommodate a large number of users of diverse interests and expertise (Nam, et al., 2009; Yang & Wei, 2009), and form thriving online communities of tremendous scale.

We would like to know more about what keeps these kinds of communities going. Due to the sheer size of their populations, Internet-scale Q&A communities might suffer more from sparse social interactions and thus low levels of commitment. However, relatively little is known about motivating people over the long run in online communities. A few studies have focused on what makes help sharing systems sustainable over time (e.g., Ackerman & Palen, 1996), and some work has investigated this problem in Internet-scale communities. Joyce and Kraut (2006) investigated newcomers' retention across six newsgroups and found interrelations among newcomers' initial post properties, reply properties, and the probability of posting again. Arguello et al. (2006) conducted a similar study with eight Usenet newsgroups, comparing the interaction patterns between newcomers and old timers, and found that they differ in their ability to get replies and in the ways they write messages. In an exploration of the online forum Slashdot, Lampe and Johnston (2005) found that how a newcomer's post is rated and moderated affects her probability of returning.

These studies, however, are limited. They largely consider initial participation, and commitment is measured only to the second post. In this paper, on the other hand, we examine users' participation lifespans to assess how systems might sustain users for the long term. As we will discuss, survival analysis shows that participation patterns and performance factors can account for considerable variance in predicting participation lifespan.

As well, this is a comparison study across three major Q&A sites (the three sites mentioned above). Thus, this is not only the first user retention study on Internet-scaled Q&A sites, but also the first comparison study among these three large communities.

The paper proceeds as follows: First, we describe our data and the methods we used. Next, we describe the factors we found to be important in users' retention over time. We also investigate differences in Q&A categories and discuss the significant differences. Finally we present a discussion of these results.

Data and Methods

Data Description

As mentioned, we investigated three Internet-scale Q&A sites across languages and cultures: Yahoo! Answers (YA) in English, Baidu Knows (BK) in Chinese, and Naver Knowledge-IN (NK) in Korean. NK, started in 2002, was the earliest, while YA and BK were launched a short time apart in 2005. All three experienced a boom in user population and traffic starting in 2006. The set of sites is well suited to a comparative study, since they are similar in scale, purpose, and basic functionality. They, however, vary in cultural context, incentive structures, and site design, and we will discuss the implications in the discussion section.

All of these sites presented challenges in collecting data. Below we briefly describe each site and its data collection. We were limited to two years of use at each site for reasons we will explain below.

Yahoo! Answers (YA) is an online knowledge sharing community website launched by Yahoo! in Dec 2005. YA sites exist in various languages, but we limit our analysis to the English site, which is by far the largest.

YA, like NK and BK, allows any user to ask or answer a question and provides a virtual point system to realize its knowledge market. Users pay a flat fee in virtual points to ask a question and can recover some of those points by selecting the best answer among those received in response to the question. If the asker does not select a best answer, it is selected by votes from other users. On the answerer side, a small number of points are awarded for contributing an answer, and a bigger, flat number of points for being selected as best. The total number of points earned, as well as the percentage of a user's answers that were selected as best are displayed in a user's profile.

Using YA APIs, we were able to crawl all questions in each category and the corresponding users within the first year after the launch of the site. However, we could not reach questions posted in the second year using this method. Instead, we used a random sample of 150K users (from Adamic, et al., 2008) over a period of three months starting a month into YA's second year.

Baidu Knows (BK) founded in June 2005, is the largest Chinese Q&A online community. As successful as its co-named search engine in China, BK has garnered a huge user base and traffic. To date, more than 100 million questions have been asked with more than half of them successfully solved (i.e., the best answer was chosen by the asker or voted on by other users).

Table 1. General Site Info. Comparison

	YA	BK	NK[1]
Founded in	2005/12	2005/6	2002
Incentive for answering	**Earn** flat point rate	**Earn** flat point rate+ flexible points	
Incentive for asking	**Pay** flat rate of points	**Earn** flat rate of points, but optionally offer extra flexible points	

Table 2. Data Description & General Characteristics

Year 1	YA	BK	NK
#sampled users	6841	14,683	6,460
Ave # ques per asker	14.3	4.9	4.64
Ave # ans per answerer	217.1	26.08	7.01
Ave # ans per question	12.1	6	3
%asker	54.1%	54%	21.4%
%answerer	7.1%	10.4%	43.5%
%doBoth	38.8%	35.6%	35.1%
Year 2	YA	BK	NK
#sampled users	72,099	18,871	61,177
Ave # ques per asker	5.32	4.17	1.78
Ave # ans per answerer	51.33	16.61	5.44
Ave # ans per question	12.71	5.1	1.72
%asker	59.3%	58.1%	60.5%
%answerer	6.8%	12.4%	24.7%
%doBoth	33.9%	29.5%	14.8%

As YA's Chinese peer, BK shares many features with YA such as using virtual point system as well as some deviations listed in Table 1. Our BK data includes a full history of the first two years of the site, including all undeleted questions and answers (57 million posts in total), with corresponding users.

Naver Knowledge-iN (NK) is the largest online Q&A community in South Korea. The site has over 43 million questions[2]. Since Naver's API access to NK data is restrictive, we manually crawled 2.6 million questions and their answers from 15 categories between 2002 and 2007. The data collection technique we used is described in Nam et al. (2009).

The site differences are summarized in Table 1. Table 2 presents the data description of the sampled users who joined each site during the first three month of each year.

Survival Analysis

Survival analysis (Cox & Oaks, 1984) is the main method in this work to measure the lifespan of users' participation. The technique has been widely used in biological and medical science, engineering, and sociology. It involves modeling of a lifetime against a specific event. For example, two uses include how many days a cancer patient will survive (against death) and how long a marriage will last (against divorce).

[1] To our knowledge, the flexible rate award was not widely used during the observation time for this study

[2] http://kin.naver.com/, retrieved on Dec, 23, 2009

Defining Lifespan. In our context, users "survive" on the site if they keep participating. Defining lifespan with respect to participation is tricky because, unlike actual death or divorce, from which few recover, users can just be inactive for a while and return again. As long as the account is still valid and the site is still running we cannot be sure that a user will not return. We chose to define "death" to be a period of inactivity exceeding 100 days, since a user inactive this long has reached the end of her lifetime that is useful to the site. As Figure 1 shows, the cumulative distribution of the maximum intervals between any two sequential actions for the users is heavy tailed. More than 70% users had no more than 100 days between posts. We therefore calculate the users' lifespan as the duration of active participation from when a user first posts to the forum to her last post, with no gap greater than 100 days. We performed a sensitivity analysis using alternate cutoffs of 50 and 150 days and obtained no major differences in the statistical analyses presented below.

Characterizing User Lifespan Across Sites

To study the sites' ability to retain users, we split the data into two stages: the initial year after the site was launched and a following year, representing a more mature period for each site. In addition, in order to be able to observe the lifespan of a user for a sufficiently long time period following her debut, we further restrict our sample to include only users whose first post occurs in the first three months of either year (as described in Table 2). Therefore, we have a range of 9 to 12 months to observe how long a user continues to participate. At the end of the observation period we then can assign each individual's life status as either dead or censored. A censored user is one who has not exceeded the cutoff interval of inactivity at the conclusion of the observation period and can thus still be considered alive. Survival analysis allows us to properly account for censored data.

Identifying Participation Patterns

We define the following variables which were used in the statistical tests to predict users' lifespans.

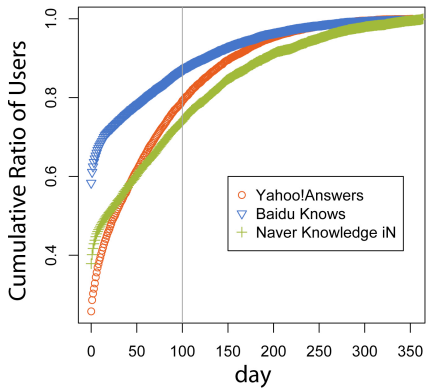

Figure 1. Cumulative distribution of maximum inactivity intervals for users

Table 3. Variables of Participation Pattern

Asking Variables	Description
# Questions	# questions user asked during a period of time, indicating activity level of asking
# Answers/Q	Ave. # answers obtained per question, indicating ability to get answers
Len_Ans	Ave. length of answers obtained
%Answered (BK)	Whether the question has >= 1 answer
%Solved (BK)	Whether a best answer was selected (either by the asker or by a vote)
# Points	Ave. # points user offered per question
%chosenBest (NK)	Whether best answer is chosen by asker
%userChosenBest(NK)	Whether best answer is chosen by others

Answering Variables	Description
# Questions	# questions answered during a period of time, indicating activity level of answering
# Answers	Ave. # answers for each question, indicating level of competition
Len_Ans	Ave. length of answers
# win	# times that user's answer was selected as best
winRate	# win/#question
Guru	winRate incorporating question competitiveness
Points earned	Ave. # points earned per question answered
Points expected	(Ave. points offered for each question the user answered)/(# questions)
%comment	% of answers that were commented on by the asker

Two other variables describe users' asking and answering activity. One is the ***ask/reply ratio*** (***A/R ratio***) representing users' preference between asking and replying to questions (as defined below). The second, ***netPoints***, is defined as the net point balance: points earned minus points expended in asking and answering activity during a period of time.

Analysis and Results

Comparing User Lifespans Across Sites.
We first compared the general survival curves across the three sites as shown in Figure 2.

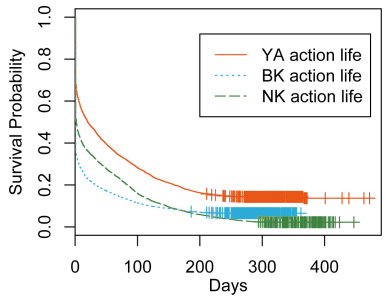

Figure 2. Active user lifespan survival curves of year 1

All sites had a stark initial drop-off, with 30%~70% of users leaving after posting just once. The observation echoes previous studies (Joyce & Kraut, 2006), which found that the first interaction is critical for sustaining a large number of potential users. Subsequently, the curves for all three sites flatten, suggesting that the longer a user remains active, the more likely they are to remain even longer. Overall, YA users are significantly more likely to remain active than users on the other two sites. NK is better able to retain users during the first 100 days than BK. However, the 10% of BK users who remain active past the first 200 days are likely to stay the full year, while NK users' survival continues to drop to as little as 3%.

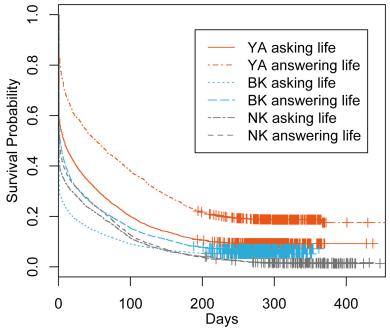

Figure 3. Users' asking lifespan versus answering lifespan (Best seen in color)

So far, our analysis has looked at combined asking and answering activity in aggregate. However, a user need not do both or may participate in one type of activity longer than the other. As shown in Figure 3, answering activity is more likely to persist than asking activity, but this difference is pronounced for just YA and BK. However, even when broken down by the type of activity, the relative difference in survival likelihood remains between sites. Interestingly, answering activity in BK has a similar lifespan to that of NK users, but it is the asking lifespan in BK that is significantly shorter; and the decay rates for both activities on YA are similar though having different initial drop-off rates.

The Role You Play: The Life You Have

Activity preference between asking & replying. As asking and replying lifespan patterns of individual users are different, we wanted to measure how users' preference for either of these two roles was related to their continued participation. We define:

$$A/R\_ratio = \frac{\#questionsAsked}{\#questionsAsked + \#questionsAnswered}$$

Activity preference and survival. We then used the A/R ratio to characterize individuals' activity lifespans. Table 4 provides the regression result. Note that the exponentiated coefficient indicates the direction of the effect: when larger than 1, it presents a negative relationship between the variable and estimate of lifespan. It presents a positive relationship when smaller than 1. For example the exp(coef) for YA is 2.497, meaning that when the A/R ratio is higher then the lifespan is shorter (i.e., users who primarily ask questions tend to leave earlier). Pr(>|z|) indicates the statistical significance; R^2 is the strength of correlation.

Table 4. Correlating survival and A/R ratio

	exp(coef)	Pr(>|z|)	R^2
YA	2.497	<0.001 ***	0.125
BK	3.297	<0.001 ***	0.147
NK	1.591	<0.001 ***	0.040

The results are consistent across all three sites: users who stay longer prefer answering to asking. Also consistent with Figure 3, this difference is more pronounced in YA and BK than NK. This is a corroboration of what one might intuitively assume: Answerers demonstrate much greater commitment to Q&A communities, by contributing more and staying significantly longer.

How First Time Experience Matters

As mentioned, previous work measured the effect of the first interaction experience on the probability of a user's returning to the online forum. Here we extend the analysis past the probability of returning once to quantifying the degree to which the initial interaction correlates with the length of the entire lifespan of a user's participation.

Table 5. First time action preference in year 1 and year 2

1st action	YA1	YA2	BK1	BK2	NK1	NK2
= asking	63.7%	73.8%	73.4%	74%	36.5%	68.4%
= answering	36.3%	26.2%	26.6%	26%	63.5%	31.6%

Since users can take two different initial actions, asking or answering, we examine them separately and predict users' participation lifespan using variables corresponding to asking and answering. Table 5 presents the ratio of users who initially ask to those who initially answer for the three sites. Except for year 1 of NK, all sites present a significantly larger preference for asking as the initial action. Interestingly, BK is rather stable between the two years, and YA and NK gained a larger portion of users who joined by asking first. Understanding this change will require additional research. It may reflect changes in how users initially discover the sites (e.g. by being routed from a search engine while performing a query) or new community-oriented services and designs. It may even reflect initial social instability in new sites.

Predicting activity lifespan by the first question asked. Table 6 presents the results of Cox proportional-hazards regressions on user lifespans using individual variables and the overall multiple regression using all predictor variables relating to the question. The results were statistically significant for both YA and BK, but not for NK.

In BK, whether the best answer was chosen, either by the asker or by voting, is the most significant factor that results in longer lifespans for the asker. In addition, obtaining more answers (in YA and BK) and longer replies (in BK) for the initial question also encourages askers to

stay longer. As indicators of the level of investment on the part of the asker, writing longer questions (on YA and BK) and offering more points (on BK), can not only attract more answers (Yang & Wei, 2009) but are also positively associated with longevity.

Table 6. Predicting lifespan by first asking activity

YA				
Individual predictor variables				
	exp(coef)	z	Pr(>|z|)	R^2
#answers	0.9817	-4.979	< .001 ***	0.006
Len_Ans	1	-1.558	0.119	0.001
Len_Ques	0.9994	-6.188	< .001 ***	0.01
Multiple R^2= 0.015, p < .001				
BK				
#answers	0.941	-10.84	< .001 ***	0.013
Len_Ans	0.9999	-5.445	< .001 ***	0.004
Len_Ques	0.9956	-9.817	< .001 ***	0.011
?answered	0.6573	-16.31	< .001 ***	0.026
?solved	0.5648	-20.33	< .001 ***	0.043
# Points	0.9982	-1.546	0.122	0
Multiple R^2= 0.049 , p=0				
NK				
#answers	1.010	0.729	0.466	0
?chosenBest	1.397	0.747	0.455	0
?userChosenBest	1.273	2.266	0.024 *	0.002
Multiple R^2= 0.003, p=0.122				

Table 7. Predicting lifespan by first answering activity

YA				
Individual predictor variables				
	exp(coef)	z	Pr(>|z|)	R^2
#Ans	1.01	0.827	0.408	0
?Win	0.8015	-4.07	< .001 ***	0.007
Multiple R^2 = 0.007, p < .001 ***				
BK				
#Ans	1	0.65	0.516	0
?Win	0.9217	-3.097	0.002 **	0.001
earnedPoints	0.9985	-1.824	0.068	0
Len_Ans	1	0.032	0.974	0
?best_Commented	0.9042	-3.377	< .001 ***	0.001
Multiple R^2 = 0.001, p=0.0238				
NK				
#Ans	1.014	1.762	0.078	0.001
%Win	0.98	-0.622	0.534	0
Multiple R^2= 0.001, p=0.214				

Predicting activity lifespan by the first answer. Similarly, we predicted lifespan for those users who started by answering questions (shown in Table 7). The results show very limited prediction power with small R^2 for YA and BK; the variables remain non-significant for NK.

Consistently between YA and BK, having one's answer selected to be the best is a promising sign for a longer lifespan. On BK, earning points also had positive effect, and importantly, getting feedback about the answers from the asker (?best_Commented) also was correlated with users staying longer.

Participation Patterns That Predict Lifespan

Next we look past the first interaction to see how users continued participation patterns can be used to predict total lifespans on the site. These patterns can be only observed and identified through a period of time. For example, users' performance can be measured as the average points earned per question answered.

Thus, we selected users who had stayed for more than 30 days and used the variables obtained during this period of time to predict how long those users would continue to participate. The results show that users' aggregate participation patterns can yield considerably more predictive power than using just a user's initial experience.

Predicting asking lifespan by the first 30 days. We first predict users' asking lifespans, based on their asking patterns in the first 30 days of participation (shown in Table 8). First, on all three sites, those who asked more questions remained longer after the 30 days, and this accounts for a significant portion of total explained variance. On NK, this difference in activity level accounts for the majority of prediction. This may imply that NK is less capable of sustaining low-use users for the long term, while the other two sites might be able to accommodate users at a variety of activity levels.

Some aspects of the general experience are important. Askers who continuously put in more effort (as measured by the average questions length) will also stay longer. As well, across the three sites, getting more answers each time

Table 8. Predicting asking lifespan by first 30 days

YA				
Individual predictor variables				
	exp(coef)	z	Pr(>|z|)	R^2
# Question	0.972	-10.49	< .001 ***	0.056
Ave# Answer	0.982	-2.857	0.004 **	0.003
Len_Ans	1.000	-0.935	0.35	0
Len_Ques	1.000	-2.455	0.014 *	0.002
A/R ratio	0.598	-10.08	< .001 ***	0.036
Multiple R^2= 0.09, p=0				
BK				
# Question	0.800	-38.12	< .001 ***	0.194
Ave# Answer	0.880	-23.86	< .001 ***	0.058
Len_Ans	1.000	-13.79	< .001 ***	0.022
Len_Ques	0.991	-21.96	< .001 ***	0.052
%answered	0.417	-32.61	< .001 ***	0.089
%solved	0.347	-36.99	< .001 ***	0.116
Ave_offerPoint	1.000	0.202	0.84	0
A/R ratio	3.442	38.03	< .001 ***	0.121
Net_Points	1.080	67.33	< .001 ***	0.292
Multiple R^2= 0.392, p=0				
NK				
# Question	0.7688	-27.34	< .001 ***	0.372
Ave# Answer	0.9681	-2.326	0.0200 *	0.002
%chosenBest	1.277	0.686	0.493	0
%userChosenBest	1.577	3.672	< .001 ***	0.004
A/R ratio	1.432	6.339	< .001 ***	0.012
Net_Points	0.9886	-2.75	< .001 ***	0.002
Multiple R^2= 0.382, p=0				

is correlated with the asker's continued participation. On BK, getting longer answers and the proportion of questions being answered (demonstrating a greater effort on the part of the community) encouraged askers to keep asking. Considering that almost 45% of questions have never been solved on BK, we can imagine many askers' being discouraged by obtaining no answer. However on NK, where both the asker and other users can select one of the answers as best, whether the asker made a selection is not statistically significant, while other users selecting the best answer, for reasons unclear to us, actually has a negative effect on lifespan.

Preference for role was mixed. On BK and NK, a user who prefers asking tended to continue asking. However, on YA, she would be less likely to continue asking (Table 8, rows for A/R ratio). This might imply the more marked tendency of YA users to switch roles between asking and answering; there is larger portion of users on YA who have both asked and answered (Table 2).

The incentive structure also had a mixed effect. Net point balance had a different effect size and direction between BK and NK. (We were unable to collect best answer selections, and therefore point balances, for YA.) For BK the factor yields a high predictive power. This might imply that on BK, net point surplus could also yield a negative effect in sustaining users.

Table 9. Predicting answering lifespan by first 30 days

	exp(coef)	z	Pr(>|z|)	R^2
YA				
# Question	0.998	-6.452	< .001 ***	0.029
Ave#Ans	0.999	-0.388	0.698	0
#win	0.994	-4.85	< .001 ***	0.016
winRate	0.840	-1.697	0.090	0.002
guru	0.894	-2.472	0.013 *	0.003
A/R ratio	8.181	13.66	< .001 ***	0.075
Multiple R^2= 0.091, p=0				
BK				
# Question	0.977	-19.49	< .001 ***	0.1
Ave# Answer	1.001	5.745	< .001 ***	0.005
#win	0.946	-13.56	< .001 ***	0.053
winRate	0.802	-3.868	< .001 ***	0.003
guru	0.986	-1.684	0.092	0.001
Ave_earned_Point	0.997	-1.549	0.121	0
Ave_expectedPoint	1.002	4.822	< .001 ***	0.004
Len_Ans	1.000	-2.175	0.030 *	0.001
%best_Commented	0.840	-2.351	0.019 *	0.001
A/R ratio	1.519	7.937	< .001 ***	0.01
Net_Points	1.023	47.77	< .001 ***	0.14
Multiple R^2= 0.488, p=0				
NK				
# Question	0.9073	-30.11	< .001 ***	0.335
Ave# Answer	0.9263	-8.385	< .001 ***	0.015
winRate	0.9128	-2.059	0.040 *	0.001
guru	0.965	-2.308	0.021 *	0.001
A/R ratio	0.9384	-0.986	0.324	0
Net_Points	0.9856	-6.737	< .001 ***	0.009
Multiple R^2= 0.371, p=0				

Predicting answering lifespan by the first 30 days.

Similarly, activity level in answering (#questions in the 30 days) is also an important factor for staying longer (see Table 9). Furthermore, users' performance in answering (as measured by #win, winRate, or guru score) is often positively, but weakly correlated with continued answering. Interestingly for BK, users who self-select for participating in higher reward questions (Ave_expectedPoint) die earlier, which is consistent with the finding in Yang & Wei (2009) and Yang et al. (2008) that experienced users tend to adopt a strategy of choosing less well-rewarded and therefore less competitive questions. The willingness to put in more effort in the form of longer answers is also weakly correlated with a higher survival rate. While unpredictable based on their first post, NK users lifespan becomes much more predictable once one accounts for the first 30 days of activity.

Community Evolvement

Next we investigated how users' participation patterns change over time. We conducted a longitudinal comparison between two different periods: the first year after the launch of site and the subsequent year, with the YA sample falling 1.5 month behind the beginning of the second year.

Figure 4 displays the user distributions over two dimensions: #posts and A/R ratio during a year. BK had a rather consistent distribution over A/R ratio and #posts between the years. On the other hand, YA and NK both had a similar shift: users who made a few posts, mainly asking, took over the largest portion of user population in the subsequent year. The shift for NK was more dramatic, consistent with Table 2.

We then compared users' activity lifespans between the two years (Figure 5). Interestingly, all sites presented a decline in survival rate from year 1 to year 2, especially for YA. In the second year, user retention in YA dropped to a similar level as BK, which maintained around 5% users after 250 days. NK suffered more difficulty in sustaining users in the second year as almost no users were left after 250 days. If new users become less committed one year after launch, this might suggest the difficulty of sustaining

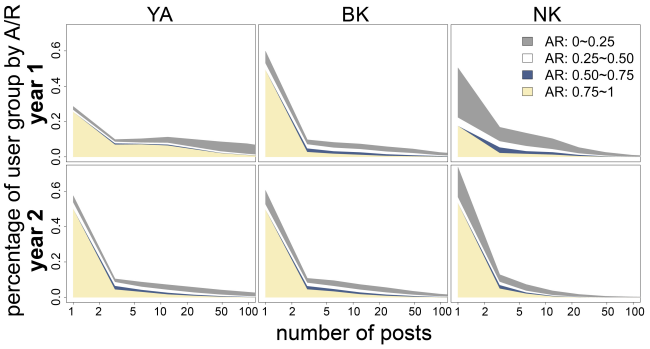

Figure 4. User distribution over #posts and A/R ratio: Users are grouped via A/R ratio and presented in different color area; for each level of #post, the length on the Y-axis presents the portion of all users who fall in the combination of #post and A/R ratio

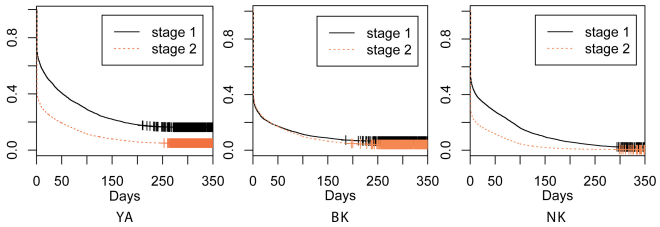

Figure 5. Comparison in user retention between early and established periods for Q&A sites.

a quickly expanding population. It may also reflect a difference in enthusiasm between early and later adopters. We do note, however, that YA users were still more active (asking or answering) in both years (Table 2).

Lifespan differences by category

All three sites have a similar category structure, encompassing topics ranging from science to relationship advice to entertainment. Previous studies found that users have different interaction patterns within categories, according to whether the category is of more informational or conversational nature (Harper, et al., 2009; Adamic, et al., 2008; Nam et al. 2009). While conversational categories usually involve substantial mutual interactions, informational categories consist of more one-directional information seeking and giving (Adamic et al. 2008). Here, we examine whether categories also differ in users' survival patterns, reflecting different commitment to the sub-communities defined by categories.

We found that there is significant and consistent difference in survival patterns among categories on YA and BK, as presented in Figure 6. On both sites, categories like "entertainment" and "trouble/advice", which would be considered as conversational, have higher survival rates over time than informational categories such as "games" on YA where users ask how to access and maneuver different computer games, or "medicine" on BK where users seek medical information.

Note, however, that the category of "computer/internet" on BK is an exception. It is often considered informational but has a significantly higher survival rate than all other categories. We suspect that there is an important factor in this: culturally-based information use.

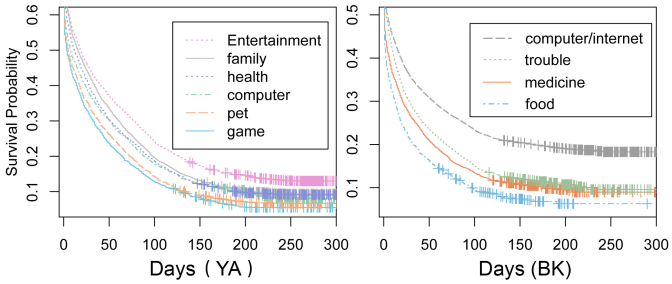

Figure 6. Survival curves in sample categories on YA & BK (Best seen in color)

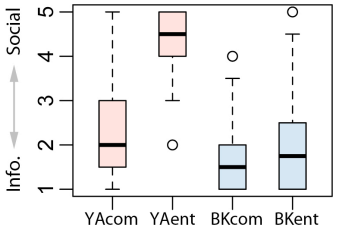

Figure 7. Rating in sample categories on YA & BK

To examine this further, we conducted human coding to evaluate a random sample of BK and YA questions. To control the variance across categories, we sampled 80 questions from each of two meta-categories of YA and BK: "entertainment" and "computer/Internet", which should represent conversational and informational topics respectively. Raters rated each question with all of its answers on a 5-point Likert scale from 1 (information seeking and providing objective information) to 5 (social discussion and conversation with subjective opinions and attitude). The Spearman inter-rater correlation was 0.83.

We found that both categories in YA have higher average scores than those of BK, as shown in Figure 7. The matched differences between the sites are statistically significant. This confirms our suspicion that there are more social conversations going on in YA, consistent with the observation that YA has significantly more answers per question on average. Both the properties of questions being asked and the patterns of answering questions contribute to the Q&A sites' dynamics. BK users ask more questions seeking objective information, prominently questions regarding online resources and computer assistance (e.g., Where can I download XXX?); YA users like to raise discussion topics to garner others' opinions or simply for fun (e.g., What is your favorite website besides this one? Or, have you lived an enchanted life?). On the answering side, we observe that compared to BK users who merely provide answers, YA users tend to add more humor, offer personal opinions, and express sociable statements.

Therefore, we tentatively interpret this significant difference across sites as a consequence of the complicated interactions among: (1) information needs and (2) cultural differences. BK is a valuable source of information about online resources and computer assistance, because there are not many other Chinese information resources for this. As well, it may be that Westerners tend to be more willing to express their opinions and feelings (Russel & Yik, 1966; Song, 1985). More research, however, will be required to understand this.

Discussion

In this paper we employed survival analyses to compare users' participation lifespans across three major Q&A sites. First, we examined participation roles, finding consistently across sites that users who preferred answering tend to have a longer, more active life within the site.

While retaining these more committed users helps sustain the Q&A community, garnering enough questions for them to answer is also important. As might be expected, askers tend to stay longer if they can successfully obtain better, more numerous, and longer responses. It is unclear, however, how much of the askers' lifespan is explained by others' response to them, and how much of it is explained by the askers' intrinsic motivation. Askers who put in more effort, in terms of the number and average length of questions they write, both obtain more answers and tend to stay longer.

For answerers, acknowledgement of one's contribution by having answers selected as best or commented on is tied to longer stays. This implies a potential need to reinforce the dynamics of information seeking and perhaps offering ways to improve both the askers and answerers' experience (for example, a routing system).

In contrast to earlier studies, which focused on users' initial interactions, we find that such interactions are only very weakly significant, and only for two of the sites, in predicting users' participation in the long run. Users whose first action is to ask are a bit more predictable that those who first post an answer. This would suggest that askers are more sensitive to the experience of first time interaction. Thus, in order to incentivize new arrivals, it might be useful to help first time askers by offering help or wizards about how to formulate a question.

There are also some intriguing differences between the sites. The most noticeable cross-site difference we found was that answerers tend to be more active in providing answers on YA, while askers ask a similar number of questions on all three sites. There is a higher rate of answers per question on YA compared to BK and NK.

We believe that the cause may be a subtle one. One possible reason for this difference could be incentive design: Both BK and NK encourage asking by rewarding askers with points while YA deducts points for asking questions and rewards answers. While incentivizing asking activity may have contributed to the incredible growth in question volume on e.g. BK, it might also bring issues such as high drop-off rates, an insufficient supply of answers or even large numbers of unsolved questions. An additional possible factor for the high volume of questions with few answers, and also the low retention of askers on NK and BK, is that many casual askers may come to the sites through search portals provided by Baidu and Naver.

A final reason is suggested by our brief analysis of informational versus conversational content on BK and YA, which found that Q&A interactions on YA tend to be conversational. This hints at potential cultural differences. These cross-site and potential cross-cultural differences will be interesting questions for future study.

Acknowledgements

This work was funded by NSF IIS 0948639 and NSF IIS-0746646. We would like to thank Kevin Nam and Jun Zhang for their help in understanding these sites as well as their data sets.

Ackerman and Adamic are co-equal authors.

References

Ackerman, S. M., & Palen, L. (1996). *The Zephyr Help Instance: promoting ongoing activity in a CSCW system*. Proceedings of the SIGCHI conference on Human factors in computing systems: common ground.

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008). *Everyone knows something: Examining knowledge sharing on Yahoo Answers*. Proceedings of the 17th international conference on World Wide Web, Beijing.

Arguello, J., Butler, B., Elisabeth, J., Kraut, R., Ling, K. S., Rose, C., et al. (2006). *Talk to Me: Foundations for Successful Individual-Group Interactions in Online Communities*. Human-Computer Interactions, Montréal, Québec, Canada.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. London: Chapman & Hall.

Harper, F. M., Moy, D., & Konstan, J. A. (2009). *Facts or Friends? Distinguishing Informational and Conversational Questions in Social Q&A Sites*. CHI, Boston, MA, USA. .

Joyce, E., & Kraut, R. E. (2006). Loyalty: Predicting continued participation in newsgroups. *Journal of Computer Mediated Communication, 11*(3), 723-747.

Lampe, C., & Johnston, E. (2005). *Follow the (slash) dot: Effects of feedback on new members in an online community*. International ACM SIGGroup conference on supporting group work, New York, NY.

Nam, K., Adamic, L. A., & Ackerman, M. S. (2009). *Questions in, Knowledge iN? A Study of Naver's Question Answering Community*. the twenty-seventh annual SIGCHI conference on Human factors in computing systems, Boston, MA.

Russel, J. A., & Yik, S. M. (1966). Emotion among Chinese. In M. H. Bond (Ed.), *The handbook of Chinese psychology*. Hong Kong: Oxford University Press.

Song, W. (1985). A preliminary study of the character traits of the Chinese. In W. S. Tseng & D. Y. H. Wu (Eds.), *Chinese culture and mental health*. Orlando, FL: Academic Press.

Yang, J., Adamic, L. A., & Ackerman, M. S. (2008). *Crowdsourcing and Knowledge Sharing: Strategic User Behavior on Taskcn*. Proceedings of the 8th ACM conference on Electronic Commerce.

Yang, J., & Wei, X. (2009). *Seeking and Offering Expertise across Categories: A Sustainable Mechanism Works for Baidu Knows*. ICWSM.