Pay-per-Question: Towards Targeted Q&A with Payments

Steve T.K. Jan, Chun Wang
Computer Science
Virginia Tech
Blacksburg, VA, USA, 24060
{tekang, wchun}@vt.edu

Qing Zhang
School of Education
Virginia Tech
Blacksburg, VA, USA, 24060
qingz@vt.edu

Gang Wang
Computer Science
Virginia Tech
Blacksburg, VA, USA, 24060
gangwang@vt.edu

ABSTRACT
Online question and answer (Q&A) services are facing key challenges to motivate domain experts to provide quick and high-quality answers. Recent systems seek to engage real-world experts by allowing them to set a price on their answers. This leads to a "targeted" Q&A model where users to ask questions to a target expert by paying the price. In this paper, we perform a case study on two emerging targeted Q&A systems Fenda (China) and Whale (US) to understand how monetary incentives affect user behavior. By analyzing a large dataset of 220K questions (worth 1 million USD), we find that payments indeed enable quick answers from experts, but also drive certain users to game the system for profits. In addition, this model requires users (experts) to proactively adjust their price to make profits. People who are unwilling to lower their prices are likely to hurt their income and engagement over time.

ACM Classification Keywords
H.5.0 Information Interfaces and Presentation (e.g. HCI): Miscellaneous; J.4 Computer Applications: Social and Behavioral Sciences

Author Keywords
Online Q&A Service; Crowdsourcing; Payments

INTRODUCTION
The success of online question and answer (Q&A) services depends on the active participation of users, particularly domain experts. With highly engaging experts, services like Quora and StackOverflow attract hundreds of millions of visitors worldwide [47]. However, for most Q&A systems, domain experts are answering questions *voluntarily* for free. As the question volume going up, it becomes difficult to draw experts' attention to a particular question, let alone getting answers on-demand [34].

To motivate domain experts, one possible direction is to introduce monetary incentives [10]. Recently, a payment-based Q&A service called *Fenda* [4] is rising quickly in China. Fenda is a social network app that connects users to well-known domain experts and celebrities to ask questions with payments. Launched in May 2016, Fenda quickly gained 10 million registered users, 500K paid questions, and 2 million US dollar revenue in the first 2 months [44]. The success of Fenda has created a new wave of payment-based Q&A services in China (Zhihu, DeDao, Weibo QA) and the U.S. (Whale, Campfire.fm, Yam).

Fenda focuses on verified, real-world domain experts, which is different from earlier payment-based Q&A services driven by an anonymous crowd (*e.g.*, Google Answers, ChaCha [2, 11, 19]). More specifically, Fenda uses a *targeted model* where users ask questions to a target expert by paying the question fee set by the expert. This model seeks to better engage and motivate experts. In addition, Fenda is the first system that explicitly rewards people for asking good questions. After a question is answered, other users in the network need to pay a small amount ($0.14) to access to the answer. This "listening fee" will be split evenly between the question asker and the answerer (Figure 1). A good question may attract enough listeners to compensate the initial question fee.

In this paper, we seek to understand the effectiveness of the targeted Q&A model and the impact of monetary incentives to the Q&A system. By performing a case study on Fenda and a U.S.-based system Whale [42], we explore the answers to a list of key questions: How does the question price affect the answering speed? What is the potential problematic user behavior caused by the monetary incentives? Whether and how could the pricing behavior predict user income and engagement level? These questions are critical for payment-based Q&A design, and Fenda and Whale provide a unique opportunity to study them.

For our analysis, we collected a large dataset of 88,540 users and 212,082 answers from Fenda (2 months in 2016), and 1,419 users and 9,199 answers from Whale (6 months in 2016–2017), involving over 1 million dollar transactions. Our analysis makes three key findings:

- *First*, using the new incentive model, both Fenda and Whale successfully attract a small group of high-profile experts who make significant contributions to the community. Fenda experts count for 0.5% of the user population, but have contributed a quarter of all answers and nearly half of the revenue.
- *Second*, the incentive model has a mixed impact on user behavior. Monetary incentive enables quick answers (av-

GROUP 2018, January 7–10, 2018, Sanibel Island, FL, USA

ACM ISBN 978-1-4503-5562-9/18/01 ...$15.00.
http://dx.doi.org/10.1145/3148330.3148332

erage delay 10–23 hours) and motivates users to ask good questions. However, we find a small number of manipulative users who either aggressively ask questions to make money from listeners, or collude/collaborate to improve their perceived popularity.

- *Third*, we find that different pricing strategies of users (question answerers) can affect their own engagement level. Users who proactively adjust their price are more likely to increase income and engagement level. However, certain celebrities are unwilling to lower their price, which in turn hurts their income and social engagement.

To the best of our knowledge, this is the first empirical study on payment-based, targeted Q&A services. Our study provides practical guidelines for other arising payment-based Q&A services (Zhihu, DeDao, Campfire.fm, Yam) and reveals key implications for future online Q&A system design. We believe this is a first step towards understanding the economy of community-based knowledge sharing.

RELATED WORK

Online Question Answering. In recent years, researchers have studied online Q&A services from various aspects [33]. Early studies have focused on identifying domain experts [26, 8] and routing user questions to the right experts [20, 27, 39]. Other works focused on assessing the quality of existing questions and answers [36, 29, 45, 30, 37, 1, 9, 35] and detecting low quality (or even abusive) content [16]. Finally, researchers also studied Q&A activities in online social networks [25, 7]. As the sizes of Q&A systems rapidly grow, it becomes challenging to engage with experts for timely and high-quality answers [34].

Crowdsourcing vs. Targeted Q&A. As shown in Table 1, most Q&A systems rely on crowdsourcing where any users in the community can answer the question. Fenda and Whale adopt a targeted Q&A model where users can ask questions to a target expert with payments. In this targeted Q&A model, it is the answerer (*e.g.*, the expert) who has the upper hand to set the price for their answers. This differs Fenda and Whale from the earlier crowdsourcing Q&A services (*e.g.*, Google Answers [2], and Mahalo [11]), and the broader crowdsourcing marketplace (*e.g.*, Mechanical Turk) [17]. In those crowdsourcing marketplaces, monetary incentive could affect the work quality and/or the response time [15, 23, 14, 46].

User Motivations in Q&A Services. Prior works have summarized three main user motivations to answer questions online: "intrinsic", "social" and "extrinsic" [13]. Intrinsic motivation refers to the psychic reward (*e.g.*, enjoyment) that users gain through helping others [48, 24]. Social factors refer to the benefits of social interactions, *e.g.*, gaining respect and enhancing reputation. Intrinsic and social factors are critical incentives for non-payment based Q&A services [13]. Extrinsic factors refer to money and virtual rewards (*e.g.*, badges and credit points) [24, 6].

Monetary incentive is an extrinsic factor implemented in payment-based Q&A services such as Google Answers, Mahalo, ChaCha and Jisiklog [2, 11, 19, 18]. These systems (most are defunct) are driven by an anonymous crowd instead of a social network that engages real-world experts. Users are primarily driven by financial incentives without a strong sense of community [19, 10]. This is concerning since research shows monetary incentive plays an important role in getting users started, but it is the social factors that contribute to the persistent participation [28].

Service	Q&A Model	Fee?	Mobile?	Content
Fenda	Targeted	Y	Y	Text/Audio
Whale	Targeted	Y	Y	Text/Video
Jisiklog	Crowdsource	Y	Y	Text
ChaCha	Crowdsource	Y	Y	Text
Google Answer	Crowdsource	Y	N	Text
Mahalo Answer	Crowdsource	Y	N	Text
Naver Q&A	Crowdsource	N	N	Text
Quora	Crowdsource	N	N	Text
Yahoo Answer	Crowdsource	N	N	Text
StackOverflow	Crowdsource	N	N	Text

Table 1. Fenda/Whale vs. other Q&A services.

Researchers have studied the impact of monetary incentives but the conclusions vary. Some researchers find that monetary incentives improve the answer quality [9] and the response rate [49]. Others suggest that payments merely reduce the response delay but have no significant impact on the answer quality [2, 12, 11]. Studies also show that payment-based Q&A can reduce low-quality questions since users are more selective regarding what to ask [10, 11].

Mobile Q&A. Mobile Q&A services leverage the ubiquitous mobile devices to enable user-friendly Q&A experience [18, 19]. Systems like ChaCha and Jisiklog allow users to interact with an online crowd via text messages. Fenda and Whale are also mobile-only Q&A services (their web interfaces are read-only). The questions in Fenda and Whale are still written in text, but the answers are recorded vide/audio messages.

RESEARCH QUESTIONS AND METHOD

Systems like Fenda and Whale are leading the way to socially engage with real-world experts for question answering. The introduction of monetary incentives makes user interactions even more complex. If not carefully designed, monetary incentives can lead the systems down to the wrong path with users chasing financial profits and losing engagement in the long run. In this paper, we use Fenda as the primary platform to investigate how monetary incentives impact the user behavior and engagement. We include Whale (a younger and smaller system) for comparison and validation purposes.

We choose Fenda and Whale for two main reasons. First, Fenda and Whale represent the first targeted Q&A model with a unique incentive model to motivate both question askers and respondents. Second, the system (Fenda in particular) has received an initial success with a significant volume of data and revenue flow. We aim to understand the reasons behind their success and potential problems moving forward, which will benefit future Q&A system design.

Background of Fenda. Fenda is a payment-based Q&A app in China, which connects users in a Twitter-like social network. Launched in May 2016, Fenda quickly gained 10

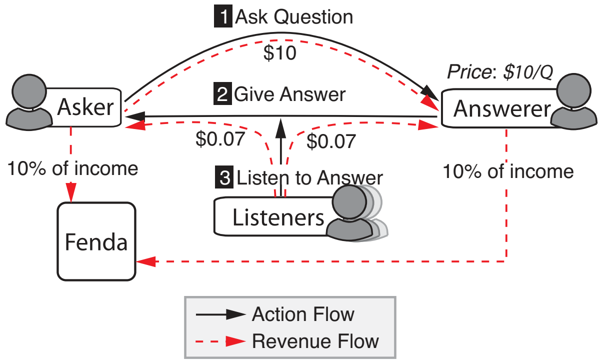


Figure 1. Fenda's Q&A workflow and revenue flow: a user can ask another user a question by making the payment. Any other users who want to listen to the answer need to pay a small amount ($0.14) which will be split evenly between the asker and the answerer. Fenda takes 10% commission fee.

million registered users and over 2 million US dollars' worth of questions answers in the first two months [44].

As shown in Figure 1, Fenda has a unique monetary incentive model to reward both question askers and answerers. A user (asker) can ask another user (answerer) a question by paying the price set by the answerer. The answerer then responds over the phone by recording a 1-minute audio message. If the answerer doesn't respond within 48 hours, the payment will be refunded. Any other user on Fenda can listen to the answer by paying a fixed amount of 1 Chinese Yuan ($0.14), and it will be split evenly between the asker and answerer. A good question may attract enough listeners to compensate the initial cost for the asker. Users set the price for their answers and can change the price anytime. Fenda charges 10% of the money made by a user.

There are two types of users on Fenda: verified real-world experts (*e.g.*, doctors, entrepreneurs, movie stars) and normal users. There is an *expert list* that contains all the experts that have been verified and categorized by the Fenda administrators. Users can browse questions from the social news feed or from the public stream of popular answers (a small sample). To promote user engagement, Fenda selects 2-4 answers daily on the public stream for free-listening for a limited time.

Background of Whale. Whale is a highly similar system launched in the US in September 2016. By analyzing the Whale's app (as of June 2017), we noticed a few differences: First, Whale users record video (instead of audio) as their answers. Second, Whale has free questions and paid questions. For paid questions, Whale takes a higher cut (20%) from the question fee. Third, listeners use the virtual currency "whale coins" to watch the paid answers. Users can receive a few *free coins* from the platform by logging-in each day, or purchase *paid coins* in bulks ($0.29 – $0.32 per coin). Only when a listener uses *paid* coins to unlock a question will the asker and answerer receive the extra payment ($0.099 each).

Our Questions. In the following, we use Fenda and Whale as the platform to analyze how monetary incentives impact user behavior and their engagement-level. We seek to answer the following key questions.

- First, as an expert-driven Q&A system, to what extent does the system rely on experts to generate content and particularly revenue?
- Second, how does the monetary incentive affect the question answering process? Does money truly enable on-demand answers from experts? Do monetary incentives encourage users to game the system for profits?
- Third, in this targeted Q&A model, how do users set and dynamically adjust the price of their answers? How does the pricing strategy affect their income and engagement-level over time?

Service	#Questions	#Users	#Askers	#Answerers
Fenda	212,082	88,540	85,510	15,529
Whale	9,199	1,419	1,371	656

Table 2. Summary of Fenda and Whale dataset.

DATA COLLECTION

We start by collecting a large dataset from Fenda and Whale through their mobile APIs. Our data collection focused on user profiles, which contained a full list of historical questions answered by the user. Data collect has a few challenges. First, there is no centralized list to crawl all registered users. Second, a user's follower list is not public (only the total number is visible). To these ends, we started our crawling from the expert list. For each expert, we collected their answered questions and the askers of those questions. Then we collected the askers' profiles to get their answered question list and extract new askers. We repeated this process until no new users appeared. In this way, we collected a large set of active users who asked or answered at least one question[1].

We collected data from Fenda in July 2016. The dataset contains 88,540 user profiles and 212,082 question-answer pairs ranging from May 12 to July 27, 2016. Each question is characterized by the asker's userID, question text, a timestamp, question price, and the number of listeners. Each answer is characterized by the answerer's userID, a length of the audio and a timestamp. UserIDs in our dataset have been fully anonymized. We briefly estimated the coverage of the Fenda dataset. Fenda announced that they had 500,000 answers as of June 27, 2016 [44]. Up to the same date, our dataset covers 155,716 answers (about 31%). For Whale, we collected 1,419 user profiles and 9,199 question-answer pairs (1114 paid questions and 8085 free questions) from September 7, 2016 to March 8, 2017. It is difficult to estimate the coverage of the Whale dataset since there is no public statistics about Whale's user base. Table 2 shows a summary of our data.

ENGAGING WITH DOMAIN EXPERTS

We first explore the roles and impact of domain experts in the system. More specifically, we examine the contributions of domain experts to the community in terms of generating content and driving financial revenue.

[1]Our study has received IRB approval: protocol # 16-1143.

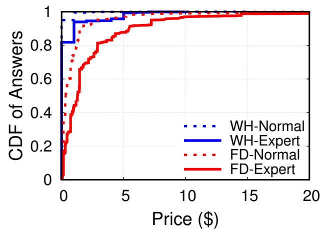


Figure 2. Price of each answer.

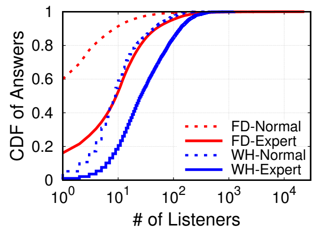


Figure 3. # of Listeners per answer.

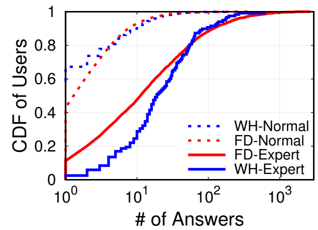


Figure 4. # of Answers per answerer.

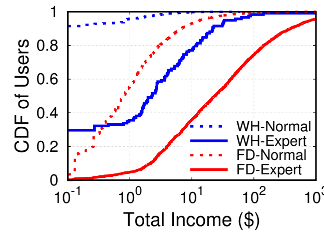


Figure 5. Income per answerer.

Fenda			Whale		
Category	Income	Experts	Category	Income	Experts
Health	\$123K	204	Startups	\$1.9K	63
Career	\$81K	222	Tech	\$1.8K	61
Business	\$81K	108	Entertain.	\$877	2
Relation.	\$73K	90	Snapchat	\$869	1
Movies	\$52K	84	Motorcycle	\$869	1
Entertain.	\$52K	51	Marketing	\$471	20
Academia	\$49K	64	Design	\$383	15
Media	\$45K	138	Travel	\$203	18
Real Estate	\$43K	28	Fitness	\$191	19
Education	\$39K	174	Finance	\$141	8

Table 3. Top 10 expert categories based on total income.

Fenda maintains a list of verified experts and celebrities. As of the time of data collection, there were 4370 verified experts classified into 44 categories by Fenda administrators. We refer these 4,370 users as *experts* and the rest 84,170 users as *normal users*. Whale has a similar expert list (118 experts), and we refer the rest 1301 Whale users as normal users.

Money. Experts play an important role in driving revenue. In total, the questions in the Fenda dataset were worth \$1,169,994[2] Experts' answers generated \$1,106,561, counting for a dominating 95% of total revenue in our dataset. To gauge experts' contribution in the context of the entire network, we again performed an estimation: Fenda reached 2 million revenues as of June 27 in 2016 [44]. Up to this same date (June 27), expert answers in our dataset have attracted \$909,876, counting for a significant 45% of the 2 million revenue. Figure 2 and Figure 3 show that, on average, experts charge higher (\$2.9 vs. \$1.0) and draw more listeners (27 vs. 5) than normal users. Individually, experts also make more money than normal users as shown in Figure 5.

On Fenda, a small group of experts (5%) made more than \$1000. The highest earning is \$33,130 by Sicong Wang, a businessman and the son of a Chinese billionaire. He answered 31 questions related to gossip and investment. He charged \$500 for each of his answers, which drew 9484 listeners (\$664 extra earning) per answer on average.

On Whale, experts are also the major contributors to the revenue flow. The total collected questions on Whale worth \$2,309 and experts contributed to \$2,028 (89%). Compared with Fenda (FD), Whale (WH) users earned significantly less money (Figure 5). A possible reason, as shown in Figure 2, is that most users (more than 80%) provide answers for free.

[2]We convert Chinese Yuan to US dollar based on \$1 = 6.9 Yuan.

Experts of different categories have distinct earning patterns. Table 3 shows the top 10 categories ranked by the total earnings per category. In Fenda, the most popular experts are related to professional consulting. The top category is `health`, followed by `career`, `business`, and `relationship`. In the `health` category, many experts are real-world physicians and pediatricians. They give Fenda users medical advice on various (non-life-threatening) issues such as headache and flu with the expense of several dollars. Other popular categories such as `movies` contain questions to celebrities about gossip. Whale, on the other hand, has fewer experts. The highest earning experts are related to `startups` and `technology`.

Question Answering. The small group of experts have contributed to a significant portion of the answers. Out of the 212K answers in the Fenda dataset, 171K (81%) are from experts. Using this dataset, we can briefly estimate the experts' contribution in the context of the entire network. On June 27 of 2016, Fenda officially announced total 500K answers and 10 million users [44]. Up to the same date, our dataset shows the 4,370 experts (0.44% of the population) have contributed 122K answers (24.4% of total answers). As shown in Figure 4, experts have answered significantly more questions than normal users. Whale (WH) has a similar situation where 118 experts (8% of users) have contributed 4,967 answers (54% of answers).

Engagement. Finally, we quickly examine whether users are more engaged on Fenda and Whale, compared to non-payment based services (*e.g.*, StackOverflow). We use the mean value of the number of answers per day per user as a proxy for engagement (e). On Fenda, the value is 0.51 for experts and 0.006 for normal users. One Whale, the value is 0.23 for experts and 0.02 for normal users. As a comparison, StackOverflow's e value is 0.01 [22]. This indicates that experts are more engaged on Fenda and Whale.

IMPACT OF MONETARY INCENTIVES

So far we show that Fenda and Whale are highly dependent on domain experts' contribution. Then the question is how to motivate experts to deliver timely and high-quality answers. In this section, we perform extensive analysis on the monetary incentive model to understand its impact on user behavior. Noticeably, Fenda and Whale use money to reward both question answerers and askers. Below, we first analyze *answerers* to understand whether payments lead to on-demand responses. Then we focus on *askers* analyzing whether and how users make money by asking the right questions. Finally, we seek

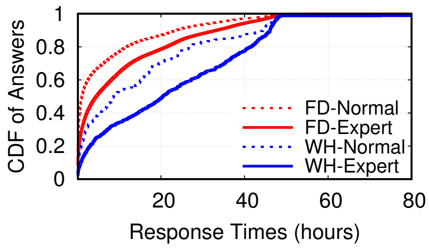


Figure 6. Response time of answers.

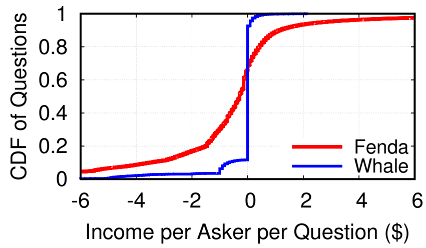


Figure 7. Income of askers per question.

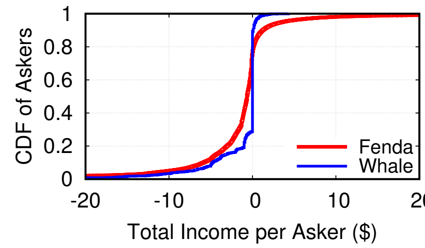


Figure 8. Total income of askers.

Pearson Correlation	Fenda	Whale
# Followers v.s. Answer Price	0.53*	0.30*
Avg. # Listeners v.s. Answer Price	0.65*	0.08*
# Questions Answered v.s. Answer Price	0.04*	0.14*
Avg. Response Time v.s. Answer Price	0.01	-0.07

Table 4. Pearson correlation between a user's answer price and key behavior metrics. * indicates significant correlation with $p < 0.05$.

Service Name	Avg. Resp. Time (hr)	Payment Based?	Crowdsourcing or Targeted?
Yahoo Answers	8.25	N	Crowdsourcing
Fenda	10.4	Y	Targeted
Whale	23.6	Y	Targeted
Google Answers	36.9	Y	Crowdsourcing
Stack Overflow	58.0	N	Crowdsourcing

Table 5. Average response time of the first answer (in hours). We compare Fenda and Whale with different Q&A sites including Yahoo Answers [43], Google Answers [3] and StackOverflow [22].

to identify *abnormal users* who aggressively or strategically game the system for profits.

Answerers

To motivate users (particularly domain experts), both Fenda and Whale allow users to determine the price for their answers. In the following, we investigate how money affects the way users answer questions. Particularly, we examine if monetary incentives truly enable on-demand quick answers.

Setting the Answer Price. To understand how users set a price for their answers, we calculate the Pearson correlation [32] between a user's price and different behavior metrics. In Table 4, we observe that the price has positive and significant correlations with the number of followers, listeners, and answered questions. A possible explanation is that users with many followers and listeners are real-world celebrities who have the confidence to set a higher price. The higher price may also motivate them to answer more questions. Note that these are correlation results, which do not reflect causality.

Surprisingly, there is no significant correlation between price and response time (for both Fenda and Whale). This is different from existing results on crowdsourcing markets, where an asker can use a higher payment to collect answers more quickly [15, 23, 11].

Answering On-demand? We further examine the response time to see if monetary incentives truly enable answering questions on-demand. As shown in Figure 6, answers arrive fast on Fenda: 33% of answers arrived within an hour and 85% arrived within a day. Note that there is a clear cut-off at 48 hours. This is the time when un-answered questions will be refunded, which motivates users to answer questions quickly. After 48 hours, users can still answer those questions for free. We find that only 0.7% of the answers arrived after the deadline, but we cannot estimate how many questions remain unanswered due to the lack of related data. Despite the high price charged by experts, experts respond slower than normal users.

The result for Whale is very similar. Figure 6 shows that for *paid* questions, 50%–70% of answers arrived within a day and normal users respond faster than experts. Comparing to Fenda, Whale has a slightly longer delay possibly because recording a video incurs more overhead than recording a voice message.

We then compare Fenda and Whale with other Q&A systems in Table 5. The response delay in Fenda and Whale is shorter than that of Google Answers and StackOverflow, but longer than that of Yahoo Answers. As payment-based systems, Fenda/Whale beats Google Answers probably because Fenda/Whale only asks for a short audio/video, while Google Answers require lengthy text. Compared to Yahoo Answers, we believe it is the crowdsourcing factor (*i.e.*, a large number of potential answerers) that plays the role. Systems like Yahoo Answers crowdsource questions to a whole community where anyone could deliver the answer. Instead, Fenda/Whale's question is targeted to a single user. The answerer is likely to answer the question within 48 hours in order to get paid, but is not motivated to answer quicker since there is no competition.

Askers

Fenda and Whale implement the first monetary incentive model to reward users for asking good questions. More specifically, once a user's question gets answered, this user (the question asker) can earn a small amount of money from people who want to listen to the answer. This model, if executed as designed, should motivate users to contribute high-quality questions for the community.

Can Askers Make Money? For each question, the question asker's income is half of listeners' payments, with Fenda's commission fee and initial question fee deducted. As shown in Figure 7, out of all questions, 40% have successfully attracted enough listeners to return a positive profit to the asker. For individual askers, Figure 8 shows 40% of them have a positive total income. This demonstrates a good chance of making profits by asking good questions on Fenda. However, for Whale, the vast majority of askers did not earn money. Part of the reason is most people only ask free questions. More

Behavior Metric	Fenda Askers $ >0	Fenda Askers $ ≤0	p	Whale Askers $ >0	Whale Askers $ ≤0	p
Avg. Followers	2155.5	3758.5	*	750.2	790.0	
Avg. Listeners	55.2	16.9	*	28.3	38.1	*
Avg. Price	1.58	4.58	*	0.0	0.3	*
Avg. Questions	3.99	1.86	*	5.4	6.6	

Table 6. Two sample t-test compares the behavior metrics for askers with positive income and those with negative income. * indicates the differences between the two types of askers are significant with $p < 0.05$.

importantly, Whale gives away free coins every day to motivate users to login. If a listener uses free coins (instead of paid coins), the asker will not receive any money.

How Do Askers Make Money? To understand why certain users make money (and others don't), we compare askers who have positive income with those with negative income in Table 6. Specifically, we examine to whom they ask questions (*i.e.*, the number followers and listeners of the answerer), average question price, and total questions asked. A two-sample t-Test [32] shows the significance of the differences between the two groups of askers.

On Fenda, users of positive income are more likely send questions to people who have more listeners and charge less. The counter-intuitive result is the *number of followers*: asking people with more followers is more likely to lose money. Our explanation is the inherent correlation between a user's number of followers and her answer price — famous people would charge higher and the money from listeners cannot cover the initial cost. Askers with a higher income often asked more questions. Again, correlation does not reflect causality: it is possible that the positive income motivates users to ask more questions, or people who asked more questions get more experienced in earning money.

It is hard to interpret the Whale results in Table 6 since only a very small of fraction of askers have a positive income (Figure 8). Noticeably, askers with positive income exclusively ask free questions (average price = 0).

Abnormal Users

Next, we examine suspicious users in the Q&A system who seek to game the system for financial profits.

Bounty Hunters. For certain users, financial gain is the primary reason to participate in payment based Q&A systems as shown in prior works [19, 11]. On Fenda and Whale, users can make a profit not only by answering questions, but also by asking good questions. Below, we analyze askers who aggressively ask questions to gain profits (referred as "bounty hunters").

To identify potential bounty hunters in Fenda, we examine outliers in Figure 9, which is a scatter plot for the number of questions a user asked versus the ratio of questions to experts. We find clear outliers at the right side (*e.g.*, users with >100 questions). They asked way more questions than average, and exclusively interact with experts (ratio of expert questions is close to 1). The most extreme example is a user who asked more than 1300 questions in two months, with 95% of questions to experts. This user earned $194.20, which is much higher than the average income of askers (-$1.95).

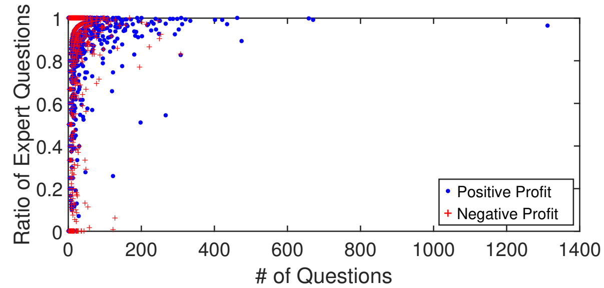
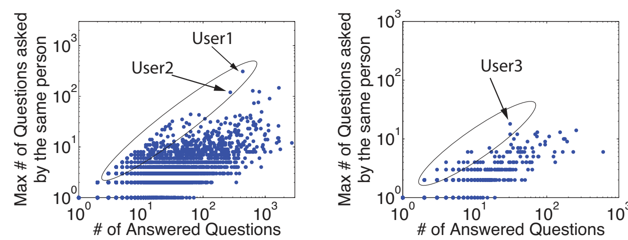


Figure 9. Total # of questions of each asker vs. the ratio of questions to experts in Fenda. Blue dots (red crosses) represent askers with positive (negative) total income.



(a) Fenda (b) Whale

Figure 10. Total # of questions of each answerer vs. Maximum # of questions asked by the same person. Dots in the circled area are likely the collusive users.

To further examine these outliers, we select askers who asked more than 100 questions. This gives us 111 users who count for 0.13% of askers but aggressively asked 11% of the questions. These users carefully target experts who charge a lower price ($0.80 per answer) but still draw significant listeners (15.5 per answer). The rest of the experts on average charge $2.49 and draw 23.0 listeners per answer.

We performed the same analysis on Whale and did not find such outlier users because most askers did not make a positive profit (Figure 8).

Collusive/Collaborative Users. In addition, there are some users who work collaboratively together to make money. For example, an asker can collude with an answerer by asking many questions (with an extremely low price) to create the illusion that the answerer is very popular. Then both the asker and the answerer can make money from the listeners of these questions. This is similar to "Sybil attacks" where multiple fake accounts are controlled by the same attacker to manipulate a social network system [40].

To identify collaborative users, we focus on answerers whose questions are primarily asked by the same user. Figure 10 shows a scatter plot for the number of questions a user answered versus the maximum number of these questions asked by the same person. Users that are close to the diagonal are suspicious. For example, *user1* answered 435 questions and 309 (71%) were asked by the same asker. We notice that this asker did not ask any other users any questions. The questions between these two users charge \$0.16 each, which is much lower than *user1*'s other questions (\$0.25 on average). By using a lower price for collusion, the two users can minimize their loss — the 10% commission fee to Fenda. In this way, *user1* earned \$689.9 in total and this asker also earned \$244 from the listeners. The second example *user2* follows the same pattern.

Figure 10(b) shows the result of Whale. The example user (user3) answered 31 questions, 18 of which were from the same asker. This asker only asked these 18 questions and all 18 questions were free of charge. This is likely an attempt to boost user3's popularity.

Discussion. Our analysis shows that monetary incentives did foster questionable behavior. On the positive side, these users (bounty hunters or collaborative users) are actually working hard to come up with interesting questions in order to earn money from listeners. On the negative side, such behavior has a disruptive impact on the marketplace. For example, bounty hunters are injecting a large volume of questions to experts. The large volume of questions would act as spam to experts, blocking other users' chance to get the experts' attention. The collusive/collaborative behavior creates a fake perception of popularity, which could mislead listeners to making the wrong spending and make it unfair for honest experts.

DYNAMIC PRICING AND USER ENGAGEMENT

Fenda and Whale allow users to set the price for their answers. How users set this price may affect their financial income and their interaction with other users. In this section, we turn to the *dynamic* aspect to analyze how users adjust their answer prices over time and how different pricing strategies affect their engagement level. Understanding this question is critical since keeping users (particularly experts) engaged is the key to building a sustainable Q&A service.

In the following, we first identify common pricing strategies by applying unsupervised clustering on users' traces of price change. Then we analyze the identified clusters to understand what type of users they represent, and how their engagement-level changes over time.

Identifying Distinct Pricing Strategies

To characterize users' dynamic price change, we construct a list of features to group users with similar patterns.

Key Features. For each user, we model their price change as a sequence of events. Given user i, our dataset contains the complete list of her answers and the price for each answer. We use P_i to denote user i's price sequence $P_i = [p_{i,1}, p_{i,2}, \dots, p_{i,N_i}]$ where N_i is the total number of answers of user i. A price change event happens when $p_{i,j-1} \neq p_{i,j}$ for any $j \in [2, N_i]$. We denote the price change sequence as $C_i = [c_{i,1}, c_{i,2}, \dots c_{i,M_i}]$ where M_i is a number of times for price change and $c_{i,j}$ is a price change event (price-up, price-down, or same-price).

Table 7 list our 9 features: the overall frequency of price change (*i.e.*, $\frac{M_i}{N_i}$), a frequency for price-up and price-down, and the frequency difference between price-up and down. In addition, we consider the average price change magnitude

id	Feature Name	Feature Description
1	Price Change Freq.	# of price change / # answers
2	Price Up Freq.	# price up / # answers
3	Price Down Freq.	# price down / # answers
4	Price Up - Down	(# price up - # price down) / # answers
5	Price Up Magnitude	Average percentage of price increase
6	Price Down Magnitude	Average percentage of price decrease
7	Consecut. Same Price	Max # consecutive same price / # answers
8	Consecut, Price Up	Max # consecutive price up / # answers
9	Consecut. Price Down	Max # consecutive price down / # answers

Table 7. A list of features for price change dynamics.

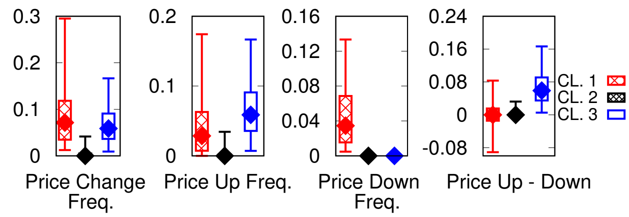


Figure 11. The distribution of top 4 features for the 3 clusters in Fenda. We depict each distribution with box plot quantiles (5%, 25%, 50%, 75%, 95%).

for price-up and price-down events. Finally, we consider the maximum number of consecutive events of same-price, price-up and price-down in the sequence.

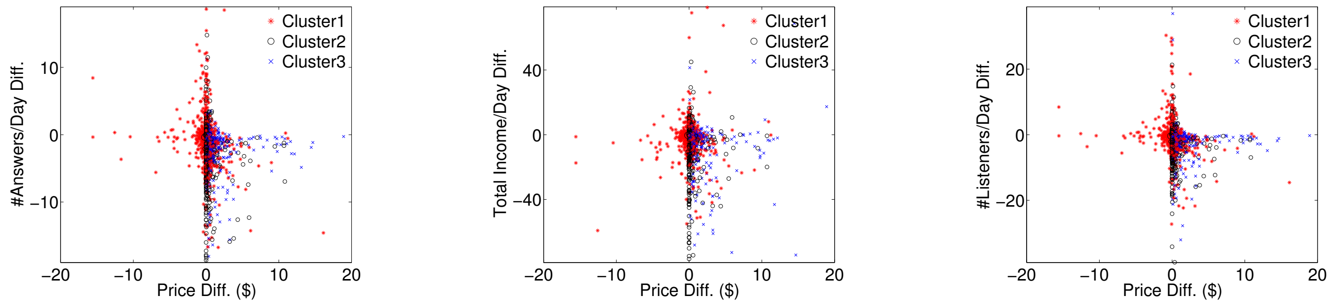
User Clustering. Based on these features, we then cluster similar users into groups. First, we compute the pair-wise Euclidean distance between users based on their feature vectors. This produces a fully connected similarity graph [41] where each node is a user and edges are weighted by distance. Then, we apply hierarchical clustering algorithm [5] to detect groups of users with similar price change patterns. We choose hierarchical clustering for two reasons: 1) It does not pre-define the number of clusters. 2) It is deterministic and the clustering result does not depend on the initial seeding.

To determine the number of clusters, we use *modularity*, a well-known metric to measure clustering quality [5]. High modularity means users are more densely connected within each cluster than to the rest of the users. We choose the number of clusters that yields the highest modularity.

Data. For this analysis, we only consider users who have answered enough questions. Otherwise, discussing their dynamic price change would be less meaningful. We heuristically set the threshold as 10 (we have tested 5 questions and the conclusion is consistent). On Fenda, this filtering produces 2094 users who have answered 171,322 questions (85% of all questions). On Whale, however, only 68 users meet the criteria. The following clustering analysis will focus on Fenda. The results of Whale are omitted due to the small number of qualified users.

Clustering Results

Our method produces 3 clusters for Fenda (modularity 0.59). To understand the pricing strategy of each cluster, we plot their feature value distributions in Figure 11. Due to space limitation, we plot 4 (out of 9) most distinguishing features that have the largest variance among the 3 clusters selected by Chi-Squared statistic [32].



(a) Engagement. (b) Income. (c) Listeners.

Figure 12. Impact of pricing strategy on user engagement, income, and listeners in Fenda. We divide a user's lifespan into two even parts, and compute the difference between the later half and the first half. A positive value indicates an upward trend.

Metrics	Cluster-1	Cluster-2	Cluster-3
Avg. #Followers	627.6	749.5	951.4
Avg. #Listeners	16.6	27.0	25.9
Avg. Price ($)	1.7	2.4	2.6
Avg. #Questions	106.5	68.8	71.4

Table 8. User statistics of the identified clusters.

- **Cluster-1 (33%):** *Frequent price up and down.* 687 users (76% are experts) who have a high price change frequency. Price up and down are almost equally frequent.
- **Cluster-2 (43%):** *Rarely changing price.* 908 users (76% are experts) who rarely change their price.
- **Cluster-3 (24%):** *Frequent price up.* 499 users (74% are experts) who increase price frequently but rarely lower their price.

We find that the 3 types of pricing patterns on Fenda correspond to users of different popularity. As shown in Table 8, cluster 1 represents the least popular answerers, who have the least followers and listeners but answered more questions. These users constantly adjust their price (primarily dropping the price), possibly to test the market. Cluster 3 represents the most popular experts and celebrities. They charge higher than others and keep increasing the price. Cluster 2 stands between cluster 1 and 3 in terms of popularity, and its users rarely change the price. The results indicate that popular users on Fenda have the luxury to keep increasing the price while less popular users need to carefully adjust the price to survive in the market.

Impact on User Engagement

Next, we analyze how price adjustments affect a user's engagement level over time. Price is a key parameter within users' control, and adjusting price is a way to test their answers' value in the market.

Figure 12(a) shows the interplay between price change and engagement level over time for 3 identified clusters on Fenda. We quantify engagement-level using number of answers per day. To measure changes over time, we divide a user's lifespan (time between her first and last answer in our dataset) into two even parts. Then we compute the differences for average price and engagement-level between the later half and first half. In a similar way, we also measure the changes in income (Figure 12(b)) and listeners (Figure 12(c)), which represent the strength of monetary and social incentives

We observe different patterns: for cluster 2 and 3, more users are located in the lower right corner than upper right, indicating a decrease of engagement, income and number of listeners. A possible explanation is that there is a mismatch between the answer's price and its value, but users did not make the right adjustments. In contrast, we find a significant number of users in cluster 1 located in the upper left corner. By lowering their price, these users get to answer more questions, and receive more money and listeners over time. We validate the statistical significance of the results by calculating the Pearson correlation [32] between the price change (x) and behavior metrics (y) for all three clusters in Figure 12. We find 8/9 of the correlations are significant ($p < 0.05$) except for cluster1's income/day metric.

Our result suggests that users need to set their price carefully to match their market value. This requires proactive price adjustments and lowering their price when necessary. Right now, highly popular users on Fenda (*e.g.*, cluster 3) are less motivated or unwilling to lower their price, which in turn hurts their income and engagement level over time.

DISCUSSION

Next, we discuss the key implications of our results to future Q&A system design.

Answering On Demand. Fenda and Whale adopt a targeted Q&A model where experts set a price for their answer. This model is suitable for targeted questions (users know who to ask), but can have a longer delay compared to crowdsourcing (where anyone can be a potential answerer). Fenda and Whale achieve faster responses than most Q&A services, but are still not as fast as the crowdsourcing based Yahoo Answers. Recently, Fenda added a new crowdsourcing channel for "medical" and "legal" questions. This channel is customer-driven: users post their questions with a cash reward, and any experts can give their answers to compete for the reward. We did a quick crawling on the crowdsourcing channel and obtained 1344 questions. We find their average response time is 4.38 hours, which is even faster than the 8.25 hours of Yahoo Answers.

Rewarding Good Questions. Fenda and Whale are the first systems that reward users financially for asking good questions. This leads to a mixed effect. On the positive side, users are motivated to ask good questions that attract broad interests. 40% of the questions on Fenda received enough listeners to cover the asker's cost. On the negative side, this model motivates a small number of users to game the system for profits. We find "bounty hunters" who aggressively ask questions to low-priced experts, and collaborative/collusive users who work together to manipulate their perceived popularity. Manipulators mainly introduce unfairness, but they still need to come up with good questions to attract listeners.

Bootstrapping for New Users. In this targeted Q&A model, a well-known expert has the key advantage to receive questions. As a result, the top 5% answerers get about 90% of the total profits in Fenda. For new comers or less known users, they receive much fewer questions. To help users to bootstrap popularity, Fenda recently introduced a system update, which allows users to set their answers "free-for-listening" for 30 minutes after posting. Whale also recently (on June 26, 2017) opened up all the questions and answers for free to encourage user participation.

Q&A Communication Channels. Fenda and Whale allow users to directly record their answers in audio/video, to avoid the inconvenience of typing text on the phone. In the context of education and communication, audio and video are also more effective than text to enhance the social bounding between communicators [21, 31], which seem to be the natural choices for mobile Q&A systems. To make online Q&A even more interactive, another possible direction is to use live streaming channel such as Periscope and Facebook Live [38]. The common challenge for audio and video communication is that answerers need to react to the questions on the fly, which makes it difficult for them to give longer and more in-depth answers. Future research can examine the proper communication channels (text, audio, video) for different Q&A contexts.

Fenda vs. Whale. Finally, we want to briefly discuss the differences between Fenda and Whale. Although these two sites are similar by design, Fenda's has been more successful so far with a significantly larger user base and more content. Other than cultural differences (China vs. U.S.), one possible factor is Fenda has been taking advantage of China's largest social network WeChat. First, users can directly signup in Fenda using their WeChat accounts, which helps users to directly locate their friends. Second, Fenda's payment is made by "WeChat pay", a mobile payment service already integrated with WeChat account. The social network effect may have helped Fenda to quickly gain a wide adoption. A similar effect has been observed in Periscope, which has successfully bootstrapped through Twitter [38].

LIMITATIONS

Our study has a few limitations. First, our study only focuses on two services: Fenda and Whale. A broader comparison with other payment-based Q&A services can help to further generalize our results. Second, our dataset is not perfect. The crawler produces a dataset with a complete list of experts but an incomplete list of normal users. We argue that most of the missing users are likely lurkers (or inactive users) who are less influential in the community. We also used Fenda's official numbers to justify parts of our results. Third, much of our analysis is based on correlation analysis, which is a simple and powerful tool to examine the interplay of different factors in a given system. However, correlation analysis has limitations to capture more complex system dynamics (*e.g.*, revealing causality). Future work will consider using tools such as time series analysis to study the causal relationship.

CONCLUSION

In this paper, we discuss lessons learned from the first targeted, payment-based Q&A systems. By analyzing a large empirical dataset, we reveal the benefits of applying monetary incentives to Q&A systems (fast response, high-quality questions) as well as potential concerns (bounty hunters and over time engagement). As more payment-based Q&A systems arise (Campfire.fm, DeDao, Zhihu Live), our research results can help system designers to make more informed design choices.

ACKNOWLEDGMENTS

The authors want to thank the anonymous reviewers for their helpful comments. This work is supported by NSF grant CNS-1717028. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any funding agencies.

REFERENCES

1. Lada A Adamic, Jun Zhang, Eytan Bakshy, and Mark S Ackerman. 2008. Knowledge sharing and yahoo answers: everyone knows something. In *Proc. of WWW*.
2. Yan Chen, Tech-Hua Ho, and Yong-Mi Kim. 2010. Knowledge market design: A field experiment at Google Answers. *Journal of Public Economic Theory* 12, 4 (2010), 641–664.
3. Benjamin Edelman. 2011. Earnings And Ratings At Google Answers. *Economic Inquiry* 50, 2 (2011), 309–320.
4. Fenda. 2017. (2017). `http://fd.zaih.com/fenda`.
5. Santo Fortunato. 2010. Community detection in graphs. *Physics Reports* 486 (2010), 75 – 174.
6. Scott Grant and Buddy Betts. 2013. Encouraging user behaviour with achievements: an empirical study. In *Proc. of MSR*.
7. Rebecca Gray, Nicole B. Ellison, Jessica Vitak, and Cliff Lampe. 2013. Who Wants to Know?: Question-asking and Answering Practices Among Facebook Users. In *Proc. of CSCW*.
8. Benjamin V. Hanrahan, Gregorio Convertino, and Les Nelson. 2012. Modeling Problem Difficulty and Expertise in Stackoverflow. In *Proc. of CSCW*.
9. F Maxwell Harper, Daphne Raban, Sheizaf Rafaeli, and Joseph A Konstan. 2008. Predictors of answer quality in online Q&A sites. In *Proc. of CHI*.

10. Gary Hsieh and Scott Counts. 2009. mimir: A market-based real-time question and answer service. In *Proc. of CHI*.
11. Gary Hsieh, Robert E Kraut, and Scott E Hudson. 2010. Why pay?: exploring how financial incentives are used for question & answer. In *Proc. of CHI*.
12. Grace YoungJoo Jeon, Yong-Mi Kim, and Yan Chen. 2010. Re-examining price as a predictor of answer quality in an online Q&A site. In *Proc. of CHI*.
13. Xiao-Ling Jin, Zhongyun Zhou, Matthew KO Lee, and Christy MK Cheung. 2013. Why users keep answering questions in online question answering communities: A theoretical and empirical investigation. *IJIM* 33, 1 (2013), 93–104.
14. Manas Joglekar, Hector Garcia-Molina, and Aditya G. Parameswaran. 2013. Evaluating the Crowd with Confidence. In *Proc. of SIGKDD*.
15. Aikaterini Katmada, Anna Satsiou, and Ioannis Kompatsiaris. 2016. Incentive mechanisms for crowdsourcing platforms. In *Proc. of INSCI*.
16. Imrul Kayes, Nicolas Kourtellis, Daniele Quercia, Adriana Iamnitchi, and Francesco Bonchi. 2015. The Social World of Content Abusers in Community Question Answering. In *Proc. of WWW*.
17. Aniket Kittur, H. Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proc. of CHI*.
18. Uichin Lee, Hyanghong Kang, Eunhee Yi, Mun Yi, and Jussi Kantola. 2012. Understanding mobile Q&A usage: an exploratory study. In *Proc. of CHI*.
19. Uichin Lee, Jihyoung Kim, Eunhee Yi, Juyup Sung, and Mario Gerla. 2013. Analyzing crowd workers in mobile pay-for-answer q&a. In *Proc. of CHI*.
20. Baichuan Li and Irwin King. 2010. Routing questions to appropriate answerers in community question answering services. In *Proc. of CIKM*.
21. Tom Lunt and John Curran. 2010. Are you listening please? The advantages of electronic audio feedback compared to written feedback. *Assessment & Evaluation in Higher Education* 35, 7 (2010), 759–769.
22. Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. Design lessons from the fastest Q&A site in the west. In *Proc. of CHI*.
23. Winter Mason and Duncan J Watts. 2010. Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter* 11, 2 (2010), 100–108.
24. Kevin Kyung Nam, Mark S Ackerman, and Lada A Adamic. 2009. Questions in, knowledge in?: a study of naver's question answering community. In *Proc. of CHI*.
25. Jeffrey Nichols, Michelle Zhou, Huahai Yang, Jeon-Hyung Kang, and Xiao Hua Sun. 2013. Analyzing the Quality of Information Solicited from Targeted Strangers on Social Media. In *Proc. of CSCW*.
26. Aditya Pal, Shuo Chang, and Joseph A. Konstan. 2012. Evolution of Experts in Question Answering Communities. In *Proc. of ICWSM*.
27. Aditya Pal, Fei Wang, Michelle X. Zhou, Jeffrey Nichols, and Barton A. Smith. 2013. Question routing to user communities. In *Proc. of CIKM*.
28. Daphne Ruth Raban. 2008. The incentive structure in an online information market. *Journal of the American Society for Information Science and Technology* 59, 14 (2008), 2284–2295.
29. Sujith Ravi, Bo Pang, Vibhor Rastogi, and Ravi Kumar. 2014. Great Question! Question Quality in Community Q&A. In *Proc. of ICWSM*.
30. Chirag Shah and Jefferey Pomerantz. 2010. Evaluating and predicting answer quality in community QA. In *Proc. of SIGIR*.
31. Lauren E. Sherman, Minas Michikyan, and Patricia M. Greenfield. 2013. The effects of text, audio, video, and in-person communication on bonding between friends. *Cyberpsychology* 7, 2 (2013).
32. David J. Sheskin. 2007. *Handbook of Parametric and Nonparametric Statistical Procedures*.
33. Ivan Srba and Maria Bielikova. 2016a. A Comprehensive Survey and Classification of Approaches for Community Question Answering. *ACM TWeb* 10, 3 (2016), 18:1–18:63.
34. I. Srba and M. Bielikova. 2016b. Why is Stack Overflow Failing? Preserving Sustainability in Community Question Answering. *IEEE Software* 33, 4 (2016), 80–89.
35. Qi Su, Dmitry Pavlov, Jyh-Herng Chow, and Wendell C. Baker. 2007. Internet-scale Collection of Human-reviewed Data. In *Proc. of WWW*.
36. Yla Tausczik, Ping Wang, and Joohee Choi. 2017. Which Size Matters? Effects of Crowd Size on Solution Quality in Big Data Q&A Communities. In *Proc. of ICWSM*.
37. Qiongjie Tian, Peng Zhang, and Baoxin Li. 2013. Towards Predicting the Best Answers in Community-based Question-Answering Services. In *Proc. of ICWSM*.
38. Bolun Wang, Xinyi Zhang, Gang Wang, Haitao Zheng, and Ben Y. Zhao. 2016b. Anatomy of a Personalized Livestreaming System. In *Proc. of IMC*.
39. Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y. Zhao. 2013a. Wisdom in the Social Crowd: an Analysis of Quora. In *Proc. of WWW*.
40. Gang Wang, Tristan Konolige, Christo Wilson, Xiao Wang, Haitao Zheng, and Ben Y. Zhao. 2013b. You are How You Click: Clickstream Analysis for Sybil Detection. In *Proc. of USENIX Security*.

41. Gang Wang, Xinyi Zhang, Shiliang Tang, Haitao Zheng, and Ben Y. Zhao. 2016a. Unsupervised Clickstream Clustering For User Behavior Analysis. In *Proc. of CHI*.
42. Whale. 2016. (2016). `https://techcrunch.com/2016/10/31/justin-kan-launches-video-qa-app-whale/`.
43. Dan Wu and Daqing He. 2014. Comparing IPL2 and Yahoo! Answers: A Case Study of Digital Reference and Community Based Question Answering. In *Proc. of IConf.*
44. Li Xuanmin. 2016. Putting a price on knowledge. `http://www.globaltimes.cn/content/997510.shtml`. (August 2016).
45. Yuan Yao, Hanghang Tong, Feng Xu, and Jian Lu. 2014. Predicting Long-term Impact of CQA Posts: A Comprehensive Viewpoint. In *Proc. of KDD*.
46. Teng Ye, Sangseok You, and Lionel P. Robert. 2017. When Does More Money Work? Examining the Role of Perceived Fairness in Pay on the Performance Quality of Crowdworkers. In *Proc. of ICWSM*.
47. Ken Yeung. 2016. Quora now has 100 million monthly visitors, up from 80 million in January. VentureBeat. (March 2016).
48. Jie Yu, Zhenhui Jiang, and Hock Chuan Chan. 2007. Knowledge Contribution in Problem Solving Virtual Communities: The Mediating Role of Individual Motivations. In *Proc. of SIGMIS CPR*.
49. Haiyi Zhu, Sauvik Das, Yiqun Cao, Shuang Yu, Aniket Kittur, and Robert Kraut. 2016. A Market in Your Social Network: The Effects of Extrinsic Rewards on Friendsourcing and Relationships. In *Proc. of CHI*.