Individual and Collective User Behavior in Crowdsourcing Services

Dominic DiPalantino
Stanford University
380 Panama Street
Stanford, CA 94305
domdip@stanford.edu

Thomas Karagiannis
Microsoft Research
7 JJ Thomson Avenue
CB3 0FB Cambridge, UK
thomkar@microsoft.com

Milan Vojnovic
Microsoft Research
7 JJ Thomson Avenue
CB3 0FB Cambridge, UK
milanv@microsoft.com

June 2011
Technical Report
MSR-TR-2010-59

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
http://www.research.microsoft.com

Abstract – We study profiles of user behavior in selecting tasks and the resulting profits in online crowdsourcing services. Specifically, we focus on (1) understanding the individual user behavior as well as the underlying collective behavior, (2) understanding the effects of competition among users on the resulting profits and (3) the evolution of user behavior with experience. Our analysis is based on data from a popular crowdsourcing service covering thousands of workers and jobs posted over a period of more than a year.

We found two distinct characteristics when looking at individual worker behavior versus the collective behavior of the community. On the one hand, we show evidence of a market segmentation where individual workers tend to direct their effort to tasks from specific ranges of rewards, reflecting their level of skill. On the other hand, the contribution of the community as a whole spreads more evenly across the entire range of rewards and this contribution exhibits a diminishing increase with the value of the reward. Furthermore, we found significant correlations between different measures of competition among users and the resulting profit and characterized how the user's performance improves with experience.

Our results would provide valuable insights to the designers of crowdsourcing services and may inform the design of novel features such as task recommendation based on user skill.

1. INTRODUCTION

Soliciting solutions to tasks via open calls to large-scale online communities have proliferated since the advent of the Internet; the term *crowdsourcing* was recently coined to refer to these approaches. When communities are offered the chance to be compensated for the work involved in their solutions, a crowdsourcing site takes on the role of an intermediary in the labor market. Paid crowdsourcing has become a large industry, both with respect to revenue and number of workers. Guru, for instance, reports that it has over a million registered freelance workers. Elance, founded in 1999, states that its workers have been paid over $245 million since its founding, and there are at least four other sites that each claim over $100 million in gross payments [1]. Companies of all scales, from small businesses to major corporations, have used crowdsourcing sites to find freelance workers.

Early crowdsourcing sites such as Guru were designed primarily to match workers with those who require their assistance (with the actual labor to be performed after the matching). Since 2005, several sites such as Topcoder and Zhubajie began to use a mechanism more akin to a contest. A task would be posted, several people would work on the task and respond with submissions, and the winner would receive a monetary reward. Typically, the set of workers is mostly disjoint from the set of those who require work (the latter herein referred to as *task owners*). Thus, crowdsourcing sites induce a bipartite network structure with links between workers and task owners.

It is important to understand how workers select tasks to work on in presence of competition in order to inform the design of crowdsourcing services. We study this by using data retrieved by crawling one of the most popular crowdsourcing sites, namely Taskcn. In particular, we focus on the following three main questions:

Q1 The Market Segmentation:

> What is the individual user strategic behavior and incentives when participating in labor markets; how is this behavior reflected collectively at the community level?

Q2 The Role of the Competition:

> How does the competition affect user earnings and success?

Q3 The Importance of Experience:

> How does user strategic behavior evolve over time?

With regard to the first question, our findings indicate an intricate *segmentation* of the labor market based on the skill level of the workers; individuals appear to concentrate their efforts in specific reward ranges that are distinct across users depending on their skill. This implies that the reward sought per submission reflects a user's skill level, and can be used by a crowdsourcing service to profile workers in order to achieve a better matching between task owners and workers. Yet, this segmentation disappears at the community level, where the effort is spread more evenly across the various reward values. At the community level, higher rewards attract a larger number of submissions and this increase appears diminishing with the value of the rewards sought. These observations provide an empirical evidence for the market segmentation observed in previous work based on a game theoretic model [2].

By addressing the second question, we try to understand how this market segmentation relates to worker earnings. To this end, we introduce the concept of the *competition network* (Section 4) to capture the effects of the competition. The competition network is formed by edges between workers weighted by the number of times they competed for the same reward, therefore, inducing an underlying social network; users explicitly compete against each other when they submit work for the same task. To study the competition network effects, and how a worker's revenue is affected by her direct or indirect competition in the network, we examine several centrality measures such as the degree, eigenvector, closeness and betweenness centralities. While some of these measures of centrality relate to the local competition of a worker (direct competitors), other more globally describe the position of a worker within the competition network. Our analysis suggests that both the intensity of competition and the frequency of repeated competitions between same workers are important factors for predicting earnings of a worker.

Our third question examines the importance of experience on worker's success in performing tasks and resulting profit (Section 5). The analysis suggests that both the reward sought and the probability of winning exhibit a diminishing increase with the number of submissions by a worker. On the other hand, we show that the expected revenue of a worker tends to increase with the number of submissions until it settles around a constant. This implies that workers do improve over time; yet, this improvement appears to occur after their first few submissions.

Structure of the paper.

In Section 2 we summarize some basic information about workers and tasks observed in our data which provide insights for addressing our main questions. The three main questions of this paper are addressed in Section 3, 4 and 5, respectively. Section 6 discusses related work. Finally, we conclude in Section 7.

2. BASIC PROPERTIES

In this section we summarize some basic properties observed about workers and tasks as observed in our data. The results presented in this work are based on data collected from crawling Taskcn.com, one of the largest crowdsourcing websites in China. Note that we implicitly consider crowdsourcing sites employing contests, and, henceforth, we use the term "crowdsourcing" to refer to such services.

Participants on Taskcn are split into two disjoint groups: task *owners*, and *workers*. Task owners post tasks for which they specify an objective, time period, and reward. Workers directly compete with one another and submit solutions to the posted tasks within the specified time period. Unlike Q&A websites such as Yahoo! Answers, the reward for the contributors of the winning submissions is monetary in nature; the recipients are chosen by the owner of the task. Taskcn charges 20% of the offered reward as commission. For the majority of tasks, a single winner is selected after the deadline of the task; the number of winners is announced when the task is posted. Given the winner-take-all structure and multiplicity of submissions, this model implies that most workers attempting a given task invest effort and time without a corresponding payoff.

We collected two types of data from Taskcn. First, we obtained basic task-related statistics for about 17 thousand tasks for which there were roughly 1.7 million submissions over a four year period. This information includes, for each task, the number of views, submissions, and registrations (similar in function to 'watching' an item on eBay, and a prerequisite for submission), the user ID of the task owner, the category in which it was posted, the reward offered, and the starting and closing times of the task. Table 1 presents summary statistics of the task data for each category present in Taskcn (discussed in the following section).

Second, each task's page listed the user IDs of workers who submitted tasks, the time at which they uploaded their submission, and the identity of the winner; this was sufficient to provide a profile of each worker's activity. We collected the entire submission histories for workers who made at least one submission to a task of the Design category in 2008 (about 7,000 distinct workers). This category was selected because it contained the largest number of submissions. Table 2 presents the corresponding user statistics.

Tasks on Taskcn vary in kind and difficulty – the worker may be asked to suggest a slogan for a company, or to undertake a complicated graphic design task. Further, the duration for which the task is posted may be brief (a few days) or long (a few months). Each of these factors influences the choices made by workers. Understanding the decisions task owners make, and how workers respond to these decisions, is a crucial step in analyzing a site such as Taskcn. Hence, the rest of this section presents a basic characterization of the posted tasks and the choices made by workers. This characterization will further inform our discussion of user strategic behavior and incentives in Section 3.

Table 2: Worker statistics.

Submissions (mean)	19.3
Submissions (median)	6
Wins (mean)	2.4%
Wins (median)	0%
Mean reward sought (RMB)	542
Median reward sought (RMB)	300

2.1 Task characterization

We begin by characterizing the properties of tasks that are posted to the site. Besides the date on which a task is posted, its principal attributes are the category in which it was listed, the duration of the submission period, and the reward promised to the winning submission, all of which are specified at the time the task is posted.

Figure 1 presents the tasks that were *open* over time, each point corresponding to a week period. Over the four year period, the number of new tasks posted to Taskcn has varied. The site grew rapidly in 2007, and a broader trend of growth continues. The figure presents considerable temporal variation of seasonal nature – for instance, the number of tasks drops dramatically in anticipation of the Chinese New Year, which occurs in late January or February, and activity seems to peak mid-year. There is an unusual spike in the data toward the end of 2007; we are unsure if this reflects an unusual event or a quirk in Taskcn's database.

Categories. Presently, Taskcn organizes tasks into the six categories presented in Table 1, each of which contains several subcategories. We number them as follows, and list some representative subcategories. 1) *Website* (Site design, Flash animation, search engine optimization), 2) *Design* (Logo design, poster design, 2D design), 3) *Programming* (Applications, databases, scripts, mobile apps), 4) *Writing* (Business plans, slogans, translation, creative writing), 5) *Multimedia* (PowerPoint presentations, photography processing, audio processing), 6) *Services* (Sales/marketing, Finance/Accounting, Law).

Heterogeneity between categories is evident in Table 1, both with regard to the choices of task owners and those of workers. Over time, the main categories have changed as presented on the site. However, there appears to have been more continuity with subcategories, and the composition of some categories (such as Design) has changed little over the years. The data is presented according to current categorization – all historical tasks in the Flash subcategory are grouped into the Website category, even though previously this subcategory was organized elsewhere. Changes in categorization may have impacted the visibility and popularity of certain tasks over time.

With the exception of the Services category, the relative popularity of each category has remained somewhat consistent over time. The Design category in particular maintains high popularity, and has seen 8,875 tasks over this time period and over 400,000 submissions; the subcategory logo design contributes significantly to this. By contrast, the Multimedia category has seen only 101 tasks in nearly four years, and fewer than 2000 submissions.

Rewards. Tasks may pay as little as 10 RMB (after

Table 1: Statistics for task across categories.

	Website	Design	Programming	Writing	Multimedia	Services	All
Tasks	1416	8,875	824	3,564	101	2,219	16,999
Submissions (mean)	14.99	49.12	8.26	296.41	18.68	75.36	99.38
Registrations (mean)	32.83	82.91	15.91	387.64	46.99	128.38	145.10
Views (mean)	6,028	7,570	5,859	14,009	10,354	10,215	9,071
Mean reward (RMB)	481.60	390.10	157.12	233.10	299.44	222.70	331.02
Median reward (RMB)	200	240	80	120	80	80	160
Mean duration (days)	22.78	25.89	17.75	26.48	21.81	20.97	24.69
Median duration (days)	17.10	20.07	12.58	21.08	16.07	15.10	19.09

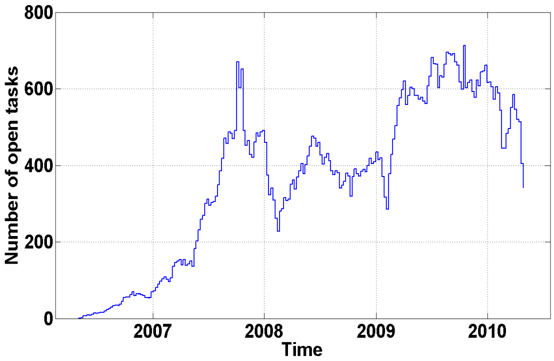

Figure 1: Open tasks over time (weekly bins).

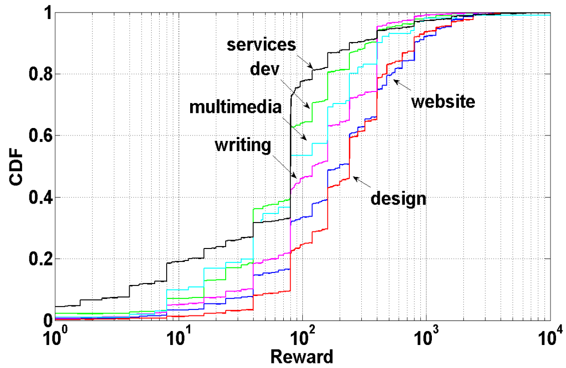

Figure 2: Reward offered per task.

20% commission, about 1.17 USD), or as much as 10,000 RMB (1,171 USD after commission). In Figure 2, we show the Cumulative Distribution Function (CDF) of the rewards chosen by task owners across the various categories. Certain rewards such as 100 RMB and 500 RMB are quite common, whereas others such as 400 RMB are relatively rare. There may be multiple reasons for this – one factor that is likely important is that users may filter tasks according to fixed thresholds using the site's interface, and tasks that fall short of these thresholds may be much less visible to workers. For example, it is easy on the site to view tasks with rewards that are 500 RMB and higher; this increases a task owner's incentive to round up a reward from 400 RMB to 500 RMB. Factors from behavioral economics related to target earning may also come into play [3]. Typically, task owners differ across the various tasks. Overall, we find that more than 75% of task owners submit a single task, and only 10% of task owners submit more than two tasks.

Duration. The task durations can vary from a few days to several weeks; the median is roughly at 20 days, while the 10^{th} and 90^{th} percentiles at 7 and 47 days respectively (Table 1). Across categories, some variability is also evident. For example, programming tasks have the shortest duration (median at 13 days, mean at 18 days), while Writing tasks have the longest (median at 21 days, mean at 26 days).

We observe a positive correlation between longer durations and higher rewards; yet, the relationship is not robust statistically. One possible cause is that owners of tasks may face a tradeoff between the speed with which they receive a solution and the quality of the solution, speculating that longer durations are associated with greater quality, other factors held constant. They may employ a higher reward to shorten the duration, particularly if they face time constraints. Thus, while we might expect more difficult tasks both to command higher rewards and to require more time, constraints on available time could break this relationship.

2.2 Worker characterization

The high number of submissions per task evident in Table 1 implies that much of workers' efforts is unrewarded and unused. This is an essential but unfortunate feature of contests. A mitigating factor is that the high number of submissions per task in the Writing category (with a mean of nearly 300) is heavily influenced by slogan writing tasks; this type of task is not conspicuously taxing for workers and a contest structure is perhaps intuitively sensible. Indeed, we found that conditioned on the subcategory, the mean number of submissions per task amounts to nearly 640 for Slogans while it is in the order of tens otherwise; as expected, the exceptions are the categories requiring specialized skills like Translation and Business Plan Design that each received roughly 20 submissions per task on the average.

Submissions per worker. The distribution of submissions per users exhibits a power-law decay, with the me-

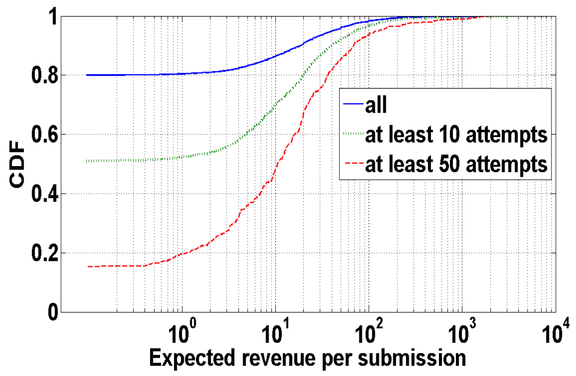

Figure 3: Mean revenue per worker per submission.

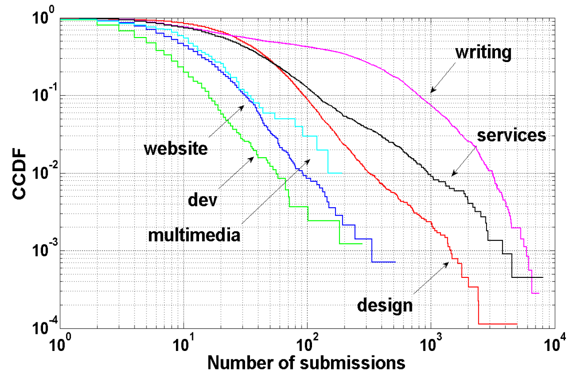

Figure 4: Number of submissions per task, CCDF conditioned on category.

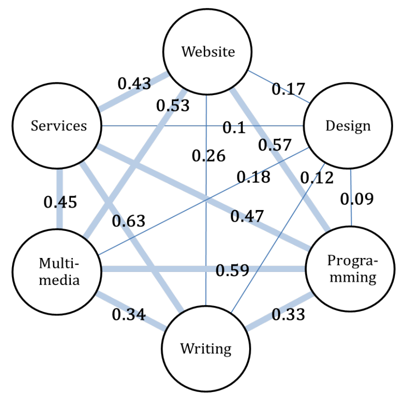

Figure 5: Correlation of workers' submissions over task categories.

dian number of submissions per worker being 6, and the mean 19.3. Workers are not restricted to one submission for each task. Examining submissions over time, the median is roughly one submission per week per worker. Workers participate in a "bursty" manner with a median of less than a day between subsequent submissions (mean of four days).

Revenue. A worker's earnings are plainly one of the most important statistics for a labor intermediary. We emphasize, however, that a worker's earnings are not identical to a worker's welfare, particularly since we do not observe the quantity of time or exertion of effort workers put into their submissions; this costly effort is even more notable since it is spent even when the worker does not receive the reward. Figure 3 shows the CDF of workers' average revenue per submission. Three groups are presented – all workers, those who have submitted to tasks at least 10 times (above the median of 6 but below the mean of 19), and those who have submitted to tasks at least 50 times. The relationship between the three groups is clear: conditioning on more submissions increases the expected revenue.

Analyzing the workers' wage rates, we observe that the median revenue after commission is about 100 RMB per month for workers that submitted at least 50 times, and that 10% of workers earn more than 800 RMB per month. Note that these are substantial amounts in comparison to the average monthly income in China in 2008 of 2,436 RMB (6,218 RMB for a worker in the software industry, or 1,352 RMB for a textile worker) [4]. While for many Taskcn is perhaps only a slightly profitable hobby, for others it is capable of providing a meaningful wage.

Submissions per category. In Table 1 we presented the number of submissions per task for different categories. We now examine this in more detail; further, we inspect cross-correlations of participation between categories.

Figure 4 shows the CCDF of the number of submissions per task, conditioned on each category. We observe that these distributions are mostly consistent with a power-law decay over a wide range of values. The degrees of the distributions are roughly comparable, though the Services category is notably more heavy-tailed and the Writing category is not very well-behaved. Though the categories are similarly heavy-tailed, the number of submissions can differ in magnitude.

In order to understand the joint distribution of workers' submissions across categories, we examine the pairwise correlation coefficients (see Fig. 5). We observe that in all cases these correlation coefficients are positive. While there are some intuitive correlations between selections of task categories, e.g., Website design tasks selected alongside with Programming tasks, others may not be expected a priori, e.g., Writing tasks selected alongside with Programming. Finally, we observe that the Design category (the largest category by number of tasks) appears disconnected from the other categories, indicating existence of an isolated community of workers.

3. MARKET SEGMENTATION

The previous section discussed how the task rewards offered are distributed across tasks, and showed that for a fraction of workers, providing solutions to posted tasks can be quite profitable. We now examine in more detail how workers select tasks of offered rewards and provide evidence of a labor market segmentation where individual worker behavior is different from collective worker behavior.

In particular, our analysis provides support for the following hypotheses:

H1 *Individual worker behavior*: A typical worker tends to focus submissions to a specific range of rewards. Specifically, a typical worker tends to submit most frequently solutions to tasks from a narrow range of rewards and

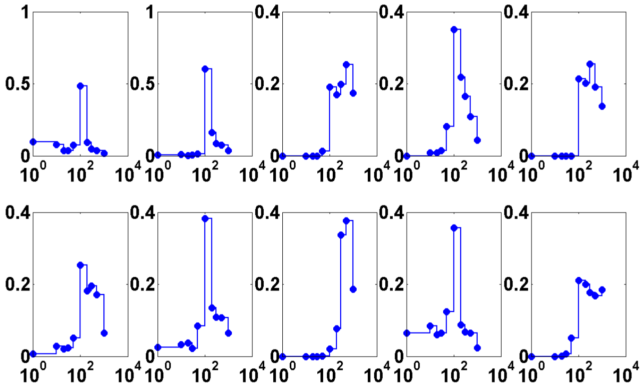

Figure 6: Histograms of the number of submissions across reward ranges by individual workers for the top ten workers with respect to the number of submissions.

attempts higher-reward tasks with diminishing frequency with the value of the reward.

H2 *Collective worker behavior*: When workers are viewed as a community, however, higher rewards tend to attract larger number of submissions and this increase is diminishing with the value of the reward.

This result (especially the first item) implies that one may be able to understand a worker's skillfulness by examining her history of submissions. In particular, *under the premise that the reward is granted to the worker with the highest quality submission (and that submission quality is increasing in a worker's skill), our findings suggest that the revenue per submission reflects a worker's skill.*

3.1 Individual behavior

We first note that if a typical worker would direct most of her submissions to tasks of a unique category and within this category the rewards would highly concentrate to a narrow set of rewards, then as a result, hypothesis H1 would trivially hold. However, we show that this is not the case. While in Figure 2 we observe that for every category, some specific values of rewards are selected by workers more often than others, the values of selected rewards span a wide range across more than two orders of magnitude. Thus, any observed concentration of values of rewards of selected tasks by a worker *is a result of user choice* over a set of tasks of different values of rewards.

We examine the distribution of rewards sought by individual workers by considering the number of submissions made by a worker to tasks across the following ranges of rewards: $[0, 10)$, $[10, 20)$, $[20, 30)$, $[30, 50)$, $[50, 100)$, $[100, 200)$, $[200, 300)$, $[300, 500)$, $[500, 1000)$, $[1000, \infty)$. We consider this partition of rewards as it is reasonably fine grained, it is well suggested by the distributions of rewards showed in Figure 2, and is well aligned with the grouping of tasks with respect to the offered rewards used in the user interface of the service.

We first demonstrate the concentration of rewards sought by individual workers; to this end, we focus on the set of workers with the highest number of submissions. These are

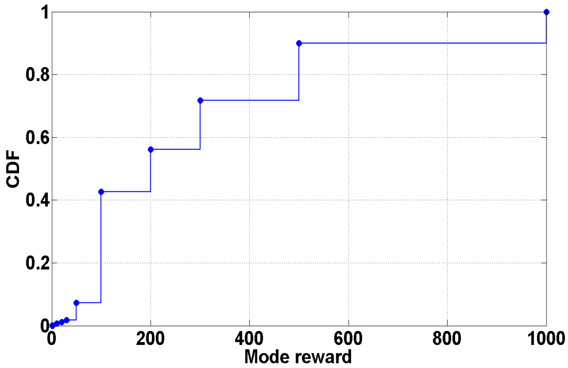

Figure 7: The distribution of the mode reward range per worker.

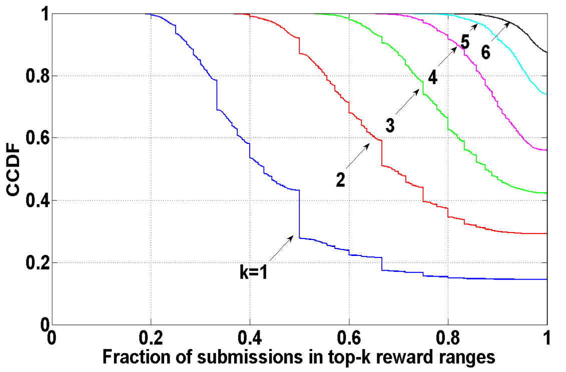

Figure 8: Submissions over reward ranges.

workers that are most experienced as measured by the observed number of submissions made per worker. In Figure 6, we observe that the histogram of rewards of selected tasks by a worker (1) exhibits a **unique mode**, (2) much of the mass of the histogram is contained in the mode, (3) the histograms concentrate around the mode, and (4) rewards higher than the mode tend to be sought with diminishing frequency. In the following, we support the latter observations by examining various statistics over the entire set of workers.

The mode of the reward sought is unique for every worker in our dataset. Furthermore, we found that *the location of this mode reward varies across workers.* The distribution of the location of the mode, showed in Figure 7, spans a large range of values from a few tens to a thousand and more. The most common mode is 100 RMB followed by higher rewards of 200, 300, 500, and 1000+ RMB which occur for approximately equal portions of workers.

For almost half of the users, user effort is concentrated in their top-2 most frequent reward ranges. This finding is supported by examining the portion of submissions made by a worker to the set of her k most frequently selected ranges of reward. Therefore, $k = 1$ corresponds to considering the portion of submissions made by a worker to her most frequently selected range of rewards, $k = 2$ corresponds to considering her two most frequently selected ranges of rewards, and so forth. In Figure 8, we show the complementary dis-

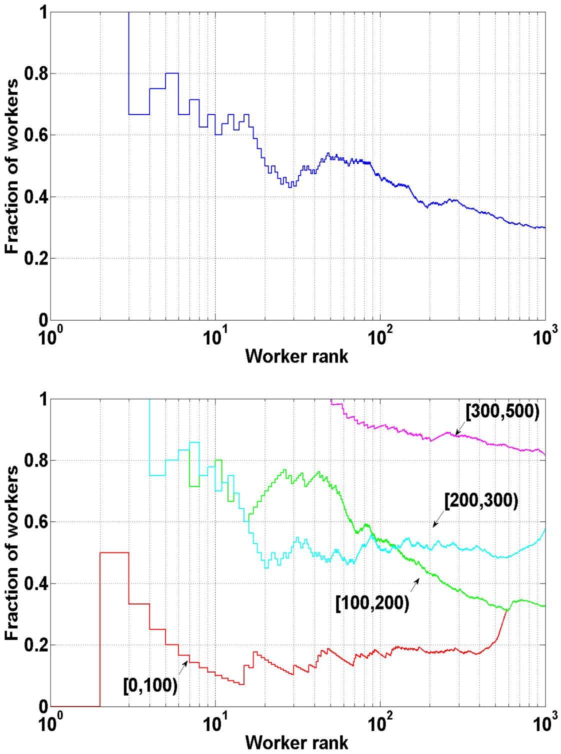

Figure 9: (Left) Fraction of top workers who select rewards higher than the mode with non-increasing frequencies. (Right) Conditioned on the location of the most frequently selected reward range.

tribution of this quantity for various values of parameter k. We observe that more than 10% of workers *always* submit to their most frequently selected reward range, and half of workers direct more than 40% of their submissions to the their most frequently selected reward range. While we observe that a substantial portion of workers direct most of their submissions to one or two or their most frequently selected ranges of rewards, an appreciable portion of workers submit across a larger number of reward ranges. Concretely, for half of workers, the three most frequent ranges account for 90% of their submissions while for 90% of workers the five most frequent ranges account for 90% of their submissions. The histogram of rewards sought by a worker is typically skewed to larger rewards, i.e., reward ranges larger than the mode, are selected more often than those smaller than the mode. Specifically, we observe a median frequency of 10% for selecting a reward smaller than the mode, 40% for selecting the mode, and 30% for selecting a reward larger than the mode.

Finally, we examine how typical it is that a worker selects rewards larger than the mode reward with non increasing frequency with the value of the reward (as the Figure 6 suggests). Figure 9 (left) shows that for workers with a large number of submissions, it is rather common that higher rewards are selected with non-increasing frequency (it amounts to more than 40% of submissions for more than 100 of the workers with the most submissions). Whether the latter property holds or not depends significantly on the location of the mode reward. In Figure 9 (right), we observe

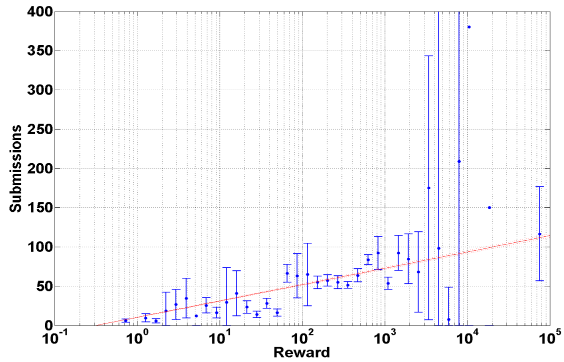

Figure 10: Number of submissions per reward (across all workers) is increasing, but with diminishing returns.

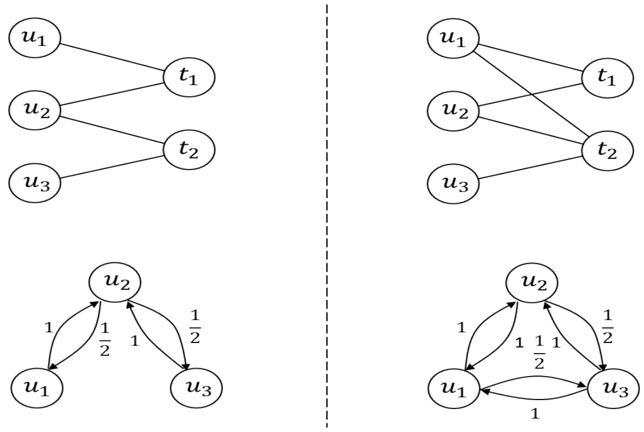

Figure 11: Construction of competition networks for two examples. The top graphs are two matchings of workers and tasks where u_i nodes represent workers and t_i nodes represent tasks. The bottom graphs show the corresponding competition networks.

that for some mode rewards, the property holds for the majority of workers; for instance, for the mode $[300, 500)$, it holds for more than 80% of workers .

3.2 Collective behavior

In the previous section, we observed that the *individual behavior* of a worker is characterized by the existence of modes, where a typical worker invests her effort in submitting solutions to tasks of rewards concentrated around specific values. Typical workers tend to select tasks that offer a reward higher than the mode with diminishing frequency with the value of the reward.

The collective behavior of workers is substantially different. Figure 10 highlights that the number of submissions tends to increase with the offered reward supporting *H2*. Note that this increase is diminishing with the value of the reward.

4. THE ROLE OF THE COMPETITION

Besides the skill level of individual workers that is reflected in user choices, other factors, such as a worker's competition may affect a worker's revenue. In this section, we examine this relationship between the revenue earned by a typical worker and the competition for her selected tasks. To this end, we describe the competition among workers by a *competition network* which enables us to quantify how the revenue earned by a worker depends on her position in this competition network. We examine and provide support for the following two main hypotheses:

H3 The workers that direct their effort to tasks with a larger number of competitors tend to make smaller revenue.

H4 The workers that tend to compete with different competitors tend to make larger revenue.

It is noteworthy that while *H3* is intuitive, *H4* is rather subtle as it implies that workers that tend to repeatedly compete with same competitors tend to earn less.

4.1 The concept of the competition network

The selection of tasks by workers induces a bipartite graph, where W is the set of workers and T is the set of tasks, and an edge $(u, t) \in W \times T$ exists if a worker $u \in W$ selected a task $t \in T$. We describe edges of this bipartite graph by $x_{u,t}$, where $x_{u,t} = 1$, if worker u submitted a solution for task t, and $x_{u,t} = 0$, otherwise. We also denote with n_u the number of submissions made by a worker u, and with m_t the total number of submissions for a task t. This graph represents a tournament of contests where a contest corresponds to the competition of workers that submitted solutions for a given task. From this representation, we construct the *competition network* defined as follows. The competition network is a directional graph $G = (W, E)$ where W is a set of vertices that represent workers, and E is the set of edges that are associated weights defined as follows. A directional edge (u, v) is associated a positive weight only if workers u and v submitted a solution for some task t, and is equal to the fraction of worker u's submissions that are directed to tasks to which worker v also made a submission. Formally, for an edge $(u, v) \in E$ we have the weight

$$w_{u,v} = \frac{\sum_{t \in T} x_{u,t} x_{v,t}}{\sum_{t \in T} x_{u,t}}. \quad (1)$$

We argue that the competition network is a natural characterization of the competition among workers. Note that the weight $w_{u,v}$ reflects the extent by which a worker u observes a worker v as a competitor. These weights are not necessarily symmetric and are normalized to avoid biases due to some workers making more submissions than other. It is noteworthy that the weight of an edge between two workers u and v in the competition network is also related to the cosine similarity that describes similarity of task selections made workers u and v,

$$c_{u,v} = \frac{\sum_{t \in T} x_{u,t} x_{v,t}}{\sqrt{\sum_{t \in T} x_{u,t}^2} \sqrt{\sum_{t \in T} x_{v,t}^2}}.$$

Indeed, the weight $w_{u,v}$ and the cosine similarity $c_{u,v}$ stand in the following relation $w_{u,v} = c_{u,v} \sqrt{\frac{n_v}{n_u}}$. Therefore, if two workers u and v have made exactly the same number of submissions, i.e., $n_u = n_v$, then the weight between these two workers is exactly the cosine similarity. Furthermore, we will see later in this section that some standard measures of node centrality in a network boil down to rather natural measures of competition. In Figure 11, we provide two simple example constructions of competitions networks.

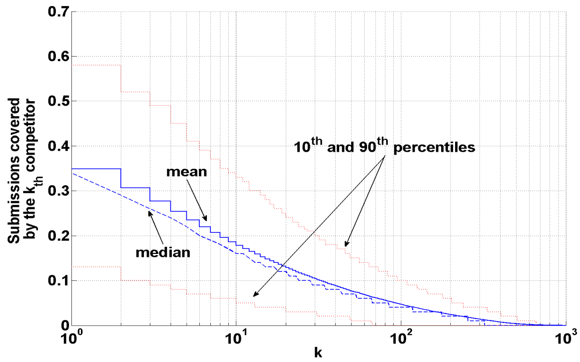

Figure 12: Weights of the competition network in decreasing order.

Before discussing the relation between worker's revenue and position in the competition network, we characterize the weights of the competition network for a typical worker. In Figure 12, we consider the weights for each worker sorted in decreasing order. We observe that for half of the workers, as many as 35% of submissions result in a competition with the most frequent competitor of this worker. We also note that competition with less frequent competitors is appreciable. Figure 13 shows the CDF of the largest weight per worker. We observe that the median of the largest weight per worker is about 40%. Despite the fact that the largest weight per worker is somewhat skewed towards smaller values, Figure 13 shows that the largest weight assumes a wide range of values in $[0, 1]$ for appreciable portion of workers.

We further examine the extent by which a worker's participation overlaps with her frequent competitors. For each worker u, we consider the proportion of her submissions that involve competing with at least one of her k most frequent competitors, and consider how this fraction increases with k. In Figure 14, we find that the latter number of submissions increases with diminishing returns with k. While a typical worker directs a large fraction of submissions to tasks that involve competing with same competitors, still, an appreciable portion of submissions is directed to tasks with new competitors.

4.2 Competition and the distribution of wealth

We examine how the revenue made by a worker relates to the position of this worker in the underlying competition network. To this end, we consider standard measures of node centrality, including the *degree centrality*, *eigenvector centrality*, *closeness centrality*, and *betweenness centrality*. While the centrality measures that we study are typical, it is insightful to discuss their relation to the original bipartite graph of task selections by workers. In the following, we define and discuss the centrality measures that we evaluate using our data and, for some of them, we derive explicit characterizations in terms of the bipartite graph of task selections by workers.

The centrality measures studied. In the following, we study both local and global centrality measures with re-

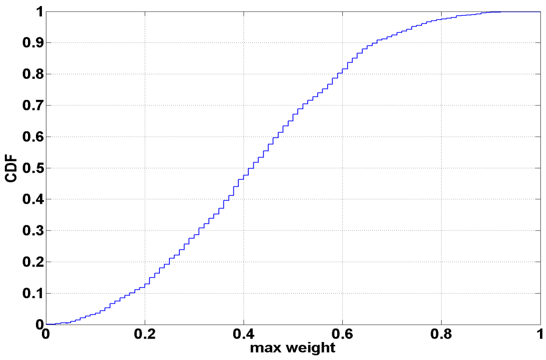

Figure 13: The distribution of the maximum weight per node.

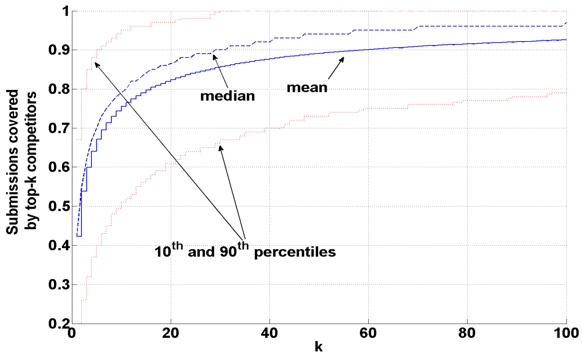

Figure 14: Worker's submissions covered by the top k competitors.

spect to a worker's position in the competition network. Local measures *are functions only of the worker's local neighborhood (up to a multiplicative constant).* Naturally, global measures reflect a worker's global competition.

The *degree centrality* d_u of a worker $u \in W$ is defined as the sum of weights of out-edges of worker u, i.e., $d_u = \sum_{v \neq u} w_{u,v}$. Using (1), we obtain that the degree centrality of a worker u is equal to

$$d_u = \frac{1}{n_u} \sum_{t \in T:\ x_{u,t}=1} (m_t - 1) \qquad (2)$$

where, recall, m_t denotes the total number of submissions to task t. Therefore, the degree centrality d_u of a worker u is exactly the mean number of competitors per submission as observed by worker u.

We define the *eigenvector centrality* e_u of a worker u to be the u-th element of the principal eigenvector $\vec{e}$ of the matrix of the competition network weights $W = (w_{i,j})$. Moreover, we also consider the eigenvector centrality for normalized weights $\mathbf{P} = (p_{u,v})$ where

$$p_{u,v} = \frac{w_{u,v}}{\sum_{v' \neq u} w_{u,v'}}. \qquad (3)$$

Notice that P is a stochastic matrix, i.e., $\sum_{v \in W} p_{u,v} = 1$. From (3) and (1), we obtain

$$p_{u,v} = \frac{\sum_{t \in T} x_{u,t} x_{v,t}}{\sum_{t \in T} x_{u,t}(m_t - 1)}. \qquad (4)$$

Therefore, $p_{u,v}$ corresponds to the number of times the worker u competes with worker v per competitor observed over the submissions by worker u. It turns out that the eigenvector centrality for the graph with weights $\mathbf{P}$ can be represented in a closed form.

PROPOSITION 4.1. *Suppose that the stochastic matrix* $\mathbf{P}$ *is irreducible and aperiodic, then for every worker* $u \in W$*, the eigenvector centrality is*

$$\pi_u = \frac{1}{C} \sum_{t \in T:\ x_{u,t}=1} (m_t - 1) \qquad (5)$$

where $C = \sum_{t \in T} m_t(m_t - 1)$ *is the normalization constant such that* $\sum_{u \in W} \pi_u = 1$.

PROOF. The asserted result follows by noting that for the homogeneous Markov chain specified by the transition matrix $\mathbf{P}$, the vector $\vec{\pi}$ satisfies the detailed balance equations:

$$\pi_u p_{u,v} = \pi_v p_{v,u}, \text{ for } (u, v) \in E. \qquad (6)$$

By Corollary 6.1 [5], $\vec{\pi}$ is the stationary distribution of $\mathbf{P}$ (i.e., a principal eigenvector of $\mathbf{P}$). From (3) and (1) we derive (4), which plugged in (6) yields

$$\frac{\pi_u}{\sum_{t \in T} x_{u,t}(m_t - 1)} = \frac{\pi_v}{\sum_{t \in T} x_{v,t}(m_t - 1)}.$$

The assertion of the proposition readily follows from the last identity. □

From (2) and (5), we observe that the degree centrality d_u of a worker u is proportional to the eigenvector centrality π_u weighted with $1/n_u$, where recall n_u is the number of submissions made by worker u. It is noteworthy that for every competition network, both the degree centralities $\vec{d}$ and the eigenvector centralities $\vec{\pi}$ are *local measures*. Instead, the following centrality measures capture *global competition*, as this is manifested in the competition network.

The *closeness centrality* c_u of a worker u is defined as the inverse of the average length of a shortest path that originates from worker u and terminates at a worker v, for every worker v in the competition network. Therefore,

$$c_u = \frac{|W| - 1}{\sum_{v \neq u} \tilde{w}_{u,v}}$$

where $\tilde{w}_{u,v}$ is the sum of the weights along a path from u to v of smallest value.

Finally, the *betweenness centrality* for a node u is defined as the sum of the fractions of shortest paths between every pair of workers that pass through worker u.

Worker's centrality vs. revenue. In Table 3, we present how a worker's centrality correlates with her earnings. All correlation coefficients with the exception of the eigenvector centrality ($\vec{\pi}$) are of statistical significance (p-value at most 0.05). We observe that *local measures, i.e., both the degree and the eigenvector centrality of a worker correlate negatively with the worker's revenue.* This is inline with our previous analysis for the degree centrality, indicating that the metric captures the mean number of competitors a worker observed per submission. Figure 15 depicts in more detail this correlation for the eigenvector centrality.

Table 3: Worker's centrality versus revenue.

Centrality metric	Correlation	p value
Degree	-0.088	0.002
Eigenvector	-0.099	0.001
Eigenvector ($\vec{\pi}$)	0.027	0.340
Closeness	0.081	0.005
Betweenness	0.071	0.013

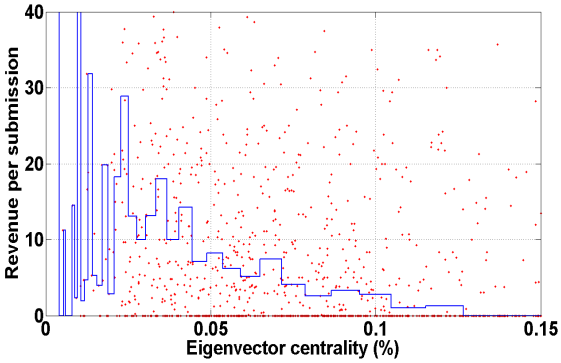

Figure 15: Worker's revenue per submission versus the worker's eigenvector centrality.

Instead, *the global centrality measures (i.e., closeness and betweenness) correlate positively with the worker's revenue.* Intuitively, these centrality measures would tend to be larger for workers who are connected with small weights to other workers. This suggests that higher earners would tend to be those workers who select tasks in a way so that they engage less frequently in *repeated* competitions with the same workers. Finally, it is noteworthy that the eigenvector ($\vec{\pi}$) centrality of a worker positively correlates with the reward in view of the fact that, in (5), we observed that this centrality measure is proportional to the total number of competitors observed over submissions by a given worker. A possible explanation is that the centrality measure tends to be higher for those workers who made more submissions and, thus, earn more as a result of experience (we discuss this in more detail in the next section).

5. THE IMPORTANCE OF EXPERIENCE

Insofar, we have examined worker strategies of picking tasks that are affected by the worker's skill and direct or indirect competition. Yet, we have not directly examined the worker's proficiency in solving tasks and the resulting revenue versus the worker's experience with time. In this section, we investigate this by characterizing the relationship between the worker experience measured by the number of submitted tasks and various measures of success, including the chance of winning a reward, the expected revenue per submission, and typical reward sought. Specifically, we examine and provide support for the following hypotheses:

H5 The probability of winning by typical workers exhibits a diminishing increase with the number of submissions;

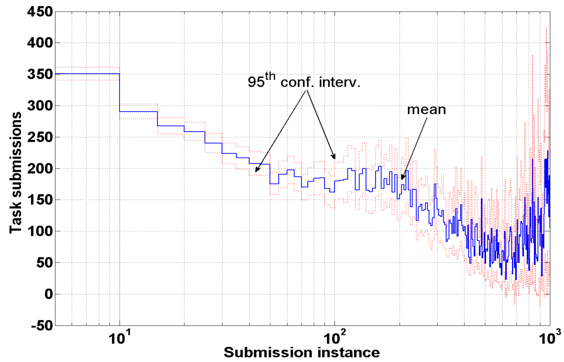

Figure 16: Number of submissions per selected task versus worker's submission instance.

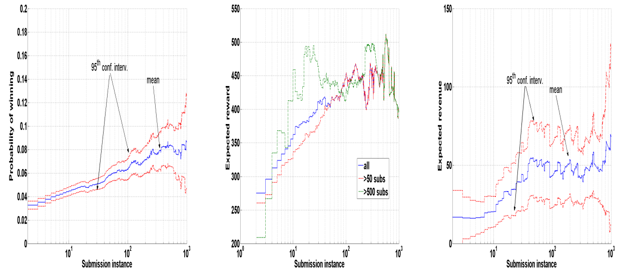

Figure 17: Evolution of the probability of winning, expected reward and expected revenue over time.

H6 The reward of selected tasks by typical workers exhibits a diminishing increase with the number of submissions;

H7 The expected revenue per submission by typical workers tends to increase until it settles around a constant with the number of submissions.

These observations suggest that workers improve with experience and contradict *the paradox of users failing to improve* which was observed in [6] by examining only the first few submissions per user. Indeed, when examining the complete user history and a larger population, experience does appear to play a role in a worker's success.

We first provide evidence that as a typical worker makes more submissions, the worker tends to submit to tasks with fewer competitors (Figure 16). The result suggests an interesting relationship between the number of submissions N for a task selected by a typical worker at submission instance S. From Figure 16, we observe that approximately, over a wide range of submission instances, the number of submissions per task N diminishes logarithmically with the submission instance S. In other words, this suggests that the *virtual probability of winning a task* $1/N$ increases with the submission instance S and over a wide range of submission instances follows the law $1/N = 1/\log\left(\frac{C}{S}\right)$, where $\log(C)$ is the number of submissions per task as observed at the submission instance 1. The latter quantity is dubbed *virtual* as it would correspond to the probability of winning a task if each competitor would have an equal chance of winning the task.

Further, we estimated the "true" probability of winning a task versus the submission instance of this worker, by identifying the actual winning submissions per user across submission instances. Figure 17-left shows the same trend – the probability of winning a task by a typical worker tends to increase with the number of submission instances by this worker. *The improvement with experience amounts to roughly doubling the probability of winning a task within hundred submissions.*

We next discuss the role of experience as measured by the number of submissions made by a worker and its relationship with the value of reward sought by the worker. In Figure 17-middle, we find that the more experienced the worker is, the larger reward she seeks. This confirms the claim in [6] over a larger range of submission instances. We observe that the value of reward sought by typical worker roughly doubles over the first hundred submission instances.

Combined together, the observations that a typical worker will increase her probability of receiving an award and will compete for tasks that offer higher rewards with the number of submission instances suggest that the expected revenue of the worker would increase with the number of submission instances. Indeed, this is confirmed in Figure 17-right, where we observe that the expected reward, as estimated based on the true winning probability, by a typical worker is roughly constant over a first few submission attempts, and then, tends to increase until settling down around a value, which happens at about a few tens of submissions. *The average earnings by a typical worker more than double after a few tens of submissions.* Our results suggest that over a wider range of submission attempts, workers do tend to earn more.

6. RELATED WORK

Strategic behavior in Taskcn and initial factors that distinguish successful workers were studied in [6]. Besides our large-scale crawl of Taskcn that produced a much larger set of data compared to [6] (including two additional years of data in which the site grew substantially), our analysis focuses on the market segmentation and the competition network which have not been studied before. Our results hence complement [6] by providing insights on how network effects might influence workers' revenue and strategic behavior.

Viewed as a network, crowdsourcing sites share characteristics with online forums and Q&A sites, where users may contribute to many possible topics, or online intermediaries, such as Alibaba and Amazon Marketplace. In such marketplaces and forums, the network may be viewed as a bipartite graph similar to the graph of workers and tasks discussed in this work. Studies have examined participation patterns and influential factors in several online forums and auctions sites (e.g., [7, 8, 9]), as well as the timing of user's submissions on Q&A sites [10, 11]. However, the motives of users in online forums differ substantially from those on crowdsourcing sites – there is no financial incentive, and the presence of many other users may be in itself a positive.

Understanding worker behavior may also illuminate the possibilities of adding monetary incentives to other platforms. Studies contemplate what would happen if market mechanisms were deployed on online Q&A sites [12], or even software development platforms [13]. In both cases, the intention is to use price signals to mitigate inefficiencies. Care must be taken not to displace non-monetary social incentives – these sites often have extensive non-monetary incentive structures in the form of 'points' and 'levels', and users may also derive utility from altruistic motives and the currency of others' attention [14]. Nonetheless, there may be an appropriate place for monetary rewards in such settings.

We note that crowdsourcing sites are not inherently efficient; much of the effort goes into submissions that are unused and unrewarded, and if participation is random some tasks may receive an insufficient amount of attention. The former inefficiency is an inherent property of contests, while to mitigate the latter, solutions such as reputations systems have been proposed [3]. However, the continued growth of these sites and our findings indicate that there remain a large number of new, possibly inexperienced entrants who would potentially benefit from the contest structure.

7. CONCLUSION

We investigated how workers respond to monetary incentives and the effect of the competition as captured by the social network of competitive relationships, as well as the effect of individual worker experience. We found that the scale matters: individual worker behavior differs qualitatively from collective behavior. Our results suggest that both the intensity of the competition as well as the frequency of repeated competitions with specific workers are important factors for the earnings of a typical worker. Finally, we showed that typical workers do improve with experience.

8. REFERENCES

[1] B. Frei. Paid crowdsourcing: Current state & progress toward mainstream business use, 2009. `http://www.smartsheet.com/files/haymaker/Paid%20Crowdsourcing%20Sept%202009%20-%20Release%20Version%20-%20Smartsheet.pdf`.

[2] D. DiPalantino and M. Vojnovic. Crowdsourcing and all-pay auctions. In *Proc. of ACM Electronic Commerce*, pages 119–128, Stanford, CA, USA, 2009.

[3] J. Horton and L. Chilton. The labor economics of paid crowdsourcing. *CoRR*, abs/1001.0627, 2010. informal publication.

[4] National Bureau of Statistics of China. China statistical yearbook, 2009. `http://www.stats.gov.cn/tjsj/ndsj/2009/indexeh.htm`.

[5] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues.* Springer, 1999.

[6] J. Yang, L. A. Adamic, and M. S. Ackerman. Crowdsourcing and knowledge sharing: strategic user behavior on taskcn. In *Proc. of ACM Electronic Commerce*, pages 246–255, Chicago, IL, USA, 2008.

[7] X. Shi, J. Zhu, R. Cai, and L. Zhang. User grouping behavior in online forums. In J. F. Elder IV, F. Fogelman-Soulié, P. A. Flach, and M. J. Zaki, editors, *KDD*, pages 777–786. ACM, 2009.

[8] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Everyone knows something: examining knowledge sharing on Yahoo answers. In *Proc. of WWW*, 2008.

[9] G. Swamynathan, C. Wilson, B. Boe, K. Almeroth, and Ben Y. Zhao. Do social networks improve e-commerce? a study on social marketplaces. In *Proc. of WOSN*, pages 1–6, Seattle, WA, USA, 2008.

[10] S. Jain, Y. Chen, and D. C. Parkes. Designing incentives for online question and answer forums. In

Proc. of ACM Electronic Commerce, pages 129–138, Stanford, CA, USA, 2009.

[11] *Questions in, Knowledge iN? A Study of Naver's Question Answering Community*, Boston, MA, 09/2008 2009.

[12] G. Hsieh and S. Counts. mimir: A market-based real-time question and answer service. In *Proc. of ACM SIGCHI*, pages 769–768, Boston, MA, USA, 2009.

[13] D. F. Bacon, Y. Chen, D. Parkes, and M. Rao. A market-based approach to software evolution. In *Proc. of ACM SIGPLAN conference companion on Object oriented programming systems languages and applications*, pages 973–980, Atlanta, GA, USA, 2009.

[14] B. A. Huberman, D. M. Romero, and F. Wu. Crowdsourcing, attention, and productivity. *Journal of Information Science*, 35(6):758–765, 2009.