Crowdsourcing Contests: Understanding the Effect of Environment and Organization Specific Factors on Sustained Participation

Hanieh Javadi Khasraghi  
University of Delaware  
ahanieh@udel.edu

Xuan Wang  
University of Texas Rio Grande Valley  
xuan.wang@utrgv.edu

Yaojie Li  
Columbus State University  
Li_yaojie@columbusstate.edu

Abstract

Crowdsourcing has increasingly become a recognized problem-solving mechanism for organizations by outsourcing the problem to an undefined crowd of people. The success of crowdsourcing depends on the sustained participation and quality-submissions of the individuals. Yet, little is known about the environment-specific and organization-specific factors that influence individuals’ continued participation in these contests. We address this research gap by conducting an empirical study using data from an online crowdsourcing contest platform, Kaggle, which delivers data science and machine learning solutions to its clients. The findings show the statistically significant effects of structural capital, familiarity with organization, and experience with the organization on individuals’ sustained participation in crowdsourcing contests. This research contributes to the literature by identifying the environment-specific and organization-specific factors that influence individuals’ sustained participation in crowdsourcing contests. Moreover, this study offers guidance to organizations that host a crowdsourcing platform to design, implement, and operate successful crowdsourcing contest platforms.

1. Introduction

Various business fields have been using crowdsourcing for their problem solving and idea innovation since “crowds can solve problems faster, better, and cheaper than companies are able to in house” [1]. A crowdsourcing request may involve anything that the company needs to be done, ranging from simple tasks (e.g. Amazon Mechanical Turk HITs) to much more complex problems including new idea generation and R&D problems, which create new knowledge for the company [2]. Many crowdsourcing activities are organized as contests and these contests can be hosted on crowdsourcing platforms (e.g. TopCoder and Kaggle) [3]. Crowdsourcing contests have emerged as an innovative way for companies to solve business problems and have enabled them to have access to the knowledge of the crowd external to the firm [4]. This type of crowdsourcing is the same as winner-take-all or selective crowdsourcing and only one or a few best solutions are accepted and rewarded (Zhao and Zhu 2014). Within the past decade, crowdsourcing contests have become established in many business fields and can offer various benefits to the companies. First, By using crowdsourcing in the form of contest companies can obtain diverse and innovative ideas and solutions from a large volume of heterogeneous people [1], [6]. Second, Companies also can have access to their customers’ ideas, innovations, preferences, and suggestions for product development and improvement [7], [8]. Third, Through crowdsourcing contests companies can lower the costs of getting innovative ideas and solutions compared to when they use internal resources [9]–[11]. For example, in crowdsourcing contest platforms (e.g. Kaggle) while companies receive hundreds or thousands of solutions from the crowd, they only pay the individuals or the teams that have submitted the best solutions. Companies also can save time by inviting a large number of people to participate in completing various tasks they want to be done [9], [12], [13].

Crowdsourcing contest is considered an important opportunity for businesses to tap into the creative potential, distributed work patterns, and expansive knowledge of large online crowds [1] for a variety of activities such as carrying out tedious work, collecting product ideas, and promoting brand awareness [13]–[15] at a lower cost [16], [17]. Because of these potential benefits, companies are increasingly opening up their boundaries in order to utilize the knowledge, experience, skills, and expertise of external resources for innovation activities and solving problems [1]. Evidence of this trend can be seen in the most prestigious companies - such as SAP, Dell, Google, General Electric, Fiat, LEGO, and Procter & Gamble - which have started their own crowdsourcing platforms [1]. Despite the widespread adoption of and
the many success stories of crowdsourcing in organizations, not all organizations master their crowdsourcing challenges successfully (e.g. Villa Enterprises, General Motors, Kraft Heinz Company, Henkel, and McDonald’s) [1]. Organizations may encounter various challenges in obtaining benefits from crowdsourcing activities. One of the frequent challenges is motivating individuals to participate [18] and to continue their participation [19]. Sustaining individual’s participation in the crowdsourcing contests and increasing their engagement with the platform are important for the success of the platform [39], [40]. Although there is considerable research which examined what factors affect individuals’ sustained participation in crowdsourcing platforms and contests, these previous studies typically were conducted to identify task-specific and individual-specific factors that affect individuals’ sustained participation in crowdsourcing contests. Individuals typically do not work in isolation in crowdsourcing platforms. Some crowdsourcing platforms are designed as contests where the individuals compete with each other and the best submission will be awarded [20]. Some crowdsourcing platforms provide an environment in which individuals can collaborate with other members [21]. Therefore, it is important to pay particular attention to the effect of environment-specific factors on individuals’ behavior in crowdsourcing platforms. Previous studies have examined some environment-specific factors such as competitors’ rating, number of competitors, number of super-star competitors, number of non-super-star competitors, collaboration, and communication on individuals’ participation behavior in crowdsourcing platform [21], [24], [35], [36]. But, none of these studies have examined how the structural capital (the number of times team members have teamed up) affect individuals’ sustained participation. Moreover, the crowdsourcing platforms host contests from different organizations which can have effect on individuals’ participation behavior in these platforms [3]. There are few studies that have explored the effect of environment-specific factors on individuals’ sustained participation, and fewer still have examined how organization-specific factors affect individuals’ sustained participation in crowdsourcing contests. For example, previous studies examined the effect of organization-specific factors such as brand-strength and marketplace maturity on individuals’ participation behavior in crowdsourcing platforms [37]. But, to our knowledge none of the previous research examined how the familiarity and experience of individuals with the crowdsourcing organization affect their sustained participation. We believe this area has been largely under-investigated and in need of attention from researchers. Our research attempts to help fill this research gap by using a rich dataset obtained from Kaggle.com, a Web-based platform which delivers data science solutions and models to its clients through problem solving contests. Kaggle is an intermediary platform which hosts data science competitions from different organizations and individuals can team up and compete against the other teams in the competitions.

The objective of this research is to investigate the effect of environment-specific factors (structural capital - the number of times team members have teamed up) and organization-specific factors (the familiarity and experience of individuals with the crowdsourcing organization) on individuals’ sustained participation. In short, the study addresses the question of how to sustain individuals’ participation in crowdsourcing contests.

The remainder of this work is structured as follows: First, we review the existing research on crowdsourcing. Second, we present the theoretical background of our study and develop a set of hypotheses. Third, we explain the methodology for our study by describing the underlying dataset, the variables, and the model that is used to test the hypotheses. Finally, we discuss our findings, outline the implications, and present the directions for the future research.

2. Literature

Previous studies extensively investigated individuals’ behavior in crowdsourcing platforms including micro-task platforms[22], idea-ovation contests [23], and problem-solving contests [24]. The extant literature has focused on four general categories that may affect individuals’ behavior and their performance in crowdsourcing contests. These four general categories are: (1) task/contest-specific factors (reward, task type, task complexity, and contest duration for task, etc.) [25]–[32]; (2) individual-specific factors (extrinsic motivations, intrinsic motivations, individuals’ strategy, and individuals’ experience) [25], [28], [29], [33], [34]; (3) environment-specific factors (competitors’ rating, number of competitors, number of super-star competitors, number of non-super-star competitors,
collaboration, and communication) [21], [24], [35], [36]; and (4) organization-specific factors (brand-strength and marketplace maturity) [37]. The effect of these factors have been investigated on different elements including individuals’ motivation to join a platform, individuals’ sustained participation, and individuals’ high quality solutions and their performance [38]. Sustaining the participation of an appropriate community of individuals (problem solvers and innovation providers) is essential to the success of crowdsourcing [39], [40]. Most of the previous studies concentrated on the factors affecting individuals’ initial participation in crowdsourcing. However, Sun et al. in a field survey with 205 subjects in TaskCn tried to understand individuals’ sustained participation in crowdsourcing contests. They found that extrinsic and intrinsic motivations significantly influence sustained participation. They also found that task complexity negatively moderates the relationship between extrinsic motivation and sustained participation and self-efficacy positively moderates the relationship between intrinsic motivation and sustained participation [19]. Boudreau et al. found a significant relationship between cash incentive and continuous level of effort of individuals, but they could not find a significant relationship between cash incentive and the individuals’ choice to participate or the relationship of cash incentive and collaboration across team members [41]. Studies have argued that individuals felt rewarded for their participation when they received feedback from the requesters (or crowdsourcing platform) regardless of whether their solution was selected or not [39]. This kind of reward (non-financial- knowledge acquisition, enhancing skills, having fun, and sense of accomplishment) gives the impression that future success is possible and strongly affects the chance of future participation by individuals [39]. Boons et al. in their field study found that feeling of pride drive ongoing member activity in crowdsourcing platforms. Platform management by engaging members in communication practices can increase members’ feelings of pride and respect [39], [40]. Feller also found that the periodic success or a belief that future success is possible strongly influences the chance of future participation by individual innovators [39]. Nguyen et al. findings illustrate that the interaction among individuals and requesters results in individuals’ higher engagement [42]. The current study on Kaggle platform shows statistically significant effects of amount of prize, tenure, number of competitions, previous team performance, and competition duration on individuals sustained participation in crowdsourcing contests [43]. None of the previous research studied how the team members’ structural capital, or the familiarity and experience of an individual with the crowdsourcing organization affect their sustained participation. In this paper we address this research gap and answer the following research question: how to sustain individuals’ participation in crowdsourcing contests? To answer this research question, we investigate the effect of environment-specific factors (structural capital-team members’ previous ties and collaboration with each other) and organization-specific factors (the familiarity and experience of an individual with the crowdsourcing organization) on individuals’ sustained participation.

3. The Theoretical Background and Hypothesis Development

3.1. Structural capital and continued participation in the contests

Previous research on online social networks have used social capital to explain individuals’ participation behavior [44]. Social capital has been conceptualized as “the sum of the assets or resources embedded in the networks of relationships between individuals, communities, networks, or societies” [45]. Structural capital as one of the dimensions of social capital also has been found to influence individuals’ behavior within a team [46]. According to Robert et al. (2008) structural capital “relates to the ties among actors and reflects the potential resources available to an actor or a team (i.e., “who knows whom”)”. Structural capital refers to the overall pattern of relationships and interactions among team members [44]. Research on virtual communities show that mutual interaction, communication, and long-term relationship are the keys for a sustainable virtual community [45]. The network ties among the team members enable them to exchange information and to integrate knowledge successfully [47]. In summary, higher levels of structural capital and interaction of individuals increase the likelihood that individuals develop shared interest, professional disciplines, common practice, and values [48]. Thus, we propose that individuals tend to sustain their participation in the competitions if they can join teams with high structural capital (structural capital refers to the prior ties and collaborations that team members had with other members within the team). Therefore, we hypothesize that:
Hypothesis 1: Team’s structural capital has positive effect on members’ sustained participation

3.2. Familiarity with the organization and continued participation in the contests

Intermediary crowdsourcing contest platforms host contests from variety of organizations. The list of organizations that have participated in crowdsourcing efforts is long from not-for-profits to Fortune 500 companies [49]. For example, over the past few years, Kaggle has hosted many companies, including General Electric, Allstate, and Facebook in order to solve their business problems [50]. Considering that variety of organizations post competitions in the crowdsourcing contest platforms, we argue that familiarity of individuals with these organizations may have an effect on their participation behavior.

Research on traditional organizations show that an organization’s image (“the way the organization is perceived by individuals” [51]) is a valuable asset which serves as the basis for individuals’ decision-making in their interactions with the organization [52]. The degree to which individuals are familiar with the organizations they are considering for employment is an important element of the employment image formulation process [53]. Organization familiarity is the degree to which an individual is acquainted with the organization [53]. The previous studies shows that individuals are attracted to the organizations that they are familiar with [53]. The more people who have heard about a firm, the more positively the firm is regarded [54]. The organizations that have had more media exposure are less unfamiliar to the individuals than the other organizations since individuals receive more signals regarding those organizations [54]. In online crowdsourcing contests, we also argue that individuals like to continue their participation in the contests from the organizations that have more platform exposure and posted more competitions in the platform. Thus we hypothesize that:

Hypothesis 2: Individuals’ familiarity with the organization has positive effect on their sustained participation

3.3. Experience with the organization and continued participation in the contests

In the previous section we argued that familiarity of the individuals with the organizations may have an effect on their participation behavior. Literature on traditional organizations shows that Individuals with previous work experience may have preconceived ideas about the types of organizations they enjoy working for, which could affect their employment decisions [55]. The study on IT outsourcing shows that when two parties have already developed a mutual understanding, the benefits to both parties increase over time when relationships continue to persist [56]. We argue that the individuals’ personal experience of working on organizations’ problems may give them additional sources of information and they get more familiar with the organizations’ problems and contests which led them to continue their participation in the contests of those organizations. Thus we hypothesize:

Hypothesis 3: Individuals’ experience with the organization has positive effect on their sustained participation

Figure 1 summarizes the research model of this study.

4. Research Methodology

4.1. Empirical Context: Kaggle.com

Data for this study comes from the publically available information on Kaggle website. Kaggle is a web-based platform that delivers data science models for its clients through the use of online competitions involving a members’ base of over 536,000 registered users, or ‘Kagglers’, from 194 countries with a variety of backgrounds from computer science to biology.
Established in 2010, Kaggle has served information technology intensive organizations (e.g., General Electric, Allstate, Ford, and Facebook) to identify their data science requirements, which it converts into contests for its member base. Each participating team can submit multiple solutions before the contest deadline. Kaggle competitions are open to all data scientists registered on the site and for competitions with monetary rewards, prize pools vary from $0 to $500,000 depending on the contest. Teams are able to submit multiple times before the competition deadline and they can see their standing in the public leaderboard compared to other competitors. There is also a data set that is concurrently scored but that is not visible until the end of the competition (Dissanayake et al. 2015).

4.2. Data collection and analysis

For this study, we used data from Kaggle’s contests starting from the launch of the platform in April 2010 through April 2019, which is the most updated data that is publicly available. Kaggle is an appropriate case for this study because it hosts contests from variety of companies and it allows the individuals to participate as a team and compete against other teams in the competitions. Additionally, Kaggle is the most popular crowdsourcing platform with the most significant number of participants compared to the other crowdsourcing platforms in the area of data science, and many technology giant companies such as Google, Apple, and Microsoft have posted competitions through the Kaggle platform. By using data from Kaggle platform we can investigate how environment-specific factors and organization-specific factors affect individuals’ participation behavior. Our sample includes 6680 observations for 78 competitions and 3502 users. Each observation is measured for each user that has participated the corresponding competition, and all the users in the sample have participated at least two times in the competitions. For example, if a user has participated in two competitions after his/her registration in the kaggle platform, there is only one observation in the sample for this user.

4.3. Variable measurement

To test our hypothesis, we measured the dependent variable and independent variables as summarized and described in Table 1.

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sustained participation</td>
<td>The number of days between two competitions for each user</td>
</tr>
<tr>
<td>Structural capital</td>
<td>The number of times team-members have teamed up before the current competition</td>
</tr>
<tr>
<td>Familiarity with organization</td>
<td>number of the competitions that the organization posted before the current competition</td>
</tr>
<tr>
<td>Experience with organization</td>
<td>the number of times the user has participated in the competitions that are posted by the same organization before the current competition</td>
</tr>
</tbody>
</table>

Six other variables are included as control variables for model adjustment. We controlled for current competition’s “prize. The literature shows monetary reward as one of the important factors affecting individuals’ participation in crowdsourcing contests [57]. We controlled for the “number of competitors”. This variable shows the number of teams competing with each other in a certain competition. The literature shows that the number of competitors negatively affects individuals participation behavior [24]. We controlled for “previous team performance” which indicates the rank that the individual’s team received in the previous competition. The literature shows that the individual’s performance can affect her/his self-efficacy. Previous research indicates that self-efficacy moderates the relationship between motivation and sustained participation [19]. We controlled for “Tenure” which indicates the number of days that have passed from the individual’s registration date in the platform until the current competition [43]. Previous research shows that individuals with longer tenure take longer time to come back to the platform and continue their participation in the other contests [43]. We controlled for the competition “duration” because duration is one of the factors that have received considerable attention in crowdsourcing research that affects individuals’ participation in crowdsourcing contests [58]. We also controlled for the “Number of previous competitions” which measures the number of times the individual has participated in the competitions before the current competition. Number of previous competitions have been identified as one of the important factors.
affecting individuals’ participation behavior in crowdsourcing platforms [59]. Table 2 provides a descriptive statistics table for all the independent variables and dependent variable.

Table 2. Descriptive Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sustained Participation</td>
<td>0</td>
<td>1800</td>
<td>132.87</td>
<td>176.47</td>
</tr>
<tr>
<td>Structural capital</td>
<td>0</td>
<td>124</td>
<td>4.66</td>
<td>11.18</td>
</tr>
<tr>
<td>Familiarity with Organization</td>
<td>0</td>
<td>32</td>
<td>1.68</td>
<td>5.89</td>
</tr>
<tr>
<td>Experience with Organization</td>
<td>1</td>
<td>4</td>
<td>1.05</td>
<td>0.24</td>
</tr>
<tr>
<td>Prize</td>
<td>0.05</td>
<td>150</td>
<td>7.7636</td>
<td>19.36</td>
</tr>
<tr>
<td>Number of Competitors</td>
<td>109</td>
<td>7198</td>
<td>2555.82</td>
<td>1881.51</td>
</tr>
<tr>
<td>Previous team performance</td>
<td>0.00</td>
<td>1.00</td>
<td>0.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Tenure</td>
<td>2</td>
<td>3129</td>
<td>749.31</td>
<td>574.30</td>
</tr>
<tr>
<td>Competition Duration</td>
<td>31</td>
<td>231</td>
<td>85.98</td>
<td>29.33</td>
</tr>
<tr>
<td>Number previous competitions</td>
<td>0</td>
<td>34</td>
<td>2.45</td>
<td>3.03</td>
</tr>
<tr>
<td>Number of Observations: 6680</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.4. Research Model and Analysis

Since the sustained participation is measured by the number of days between two competitions for each user, the variable type for the dependent variable is discrete, which is treated as count data. The negative binomial regression model is employed in this model because the variance have been way larger than the mean of the response variable continued participation. Therefore, the Possion-Gamma mixture model is the best fit to overcome the issue of dispersion in the Poisson regression models. Moreover, the possession model has been widely used in the setting of information systems study [58]-[59]. Additionally, the individual effects have been fixed in the model because the differences among users also need to be controlled. The model in this study can be presented as:

\[
Sustained\ \text{Participation} = \exp(C + \beta_1 \cdot \text{Structural capital} + \beta_2 \cdot \text{Familiarity with Org.} + \beta_3 \cdot \text{Experience with Org.} + \beta_4 \cdot \text{prize} + \beta_5 \cdot \text{number of competitors} + \beta_6 \cdot \text{previous performance} + \beta_7 \cdot \text{tenure} + \beta_8 \cdot \text{competition duration} + \beta_9 \cdot \text{num. of prev. competitions} + \varepsilon)
\]

4.5. Results

The results of the Negative binomial model are listed in Table 3. As shown in Table 3 Hypothesis 1 is supported at p<0.05 significance level which indicates that the number of times the team members have teamed up with each other in the previous competitions have significant effect on individuals decision to come back to the platform and participate in another competition in a shorter period of time. Hypothesis 2 is supported at p<0.01 significance level which indicates that the number of times the company posted competitions on the platform before the current competition have significant effect on individuals to come back to the platform in a shorter period of time and participate in another competition posted by the same company. For Hypothesis 3 although we found the significant effect of the number of times the user has participated in the competitions that are posted by the same organization and sustained participation but the coefficient in positive (negative effect on sustained participation). It means that as the number of times the users have participated in the same company’s competitions increase, it takes longer for them to come back to the platform and participate in that company’s competition again. The findings also show the significant effect of prize, number of competitors, previous team performance, tenure, competition duration, and number of previous competitions that the user has participated in, on the continued participation.

Table 3. Results from Negative Binomial Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Prob &gt; Chi Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural capital</td>
<td>-0.0026</td>
<td>0.0243 **</td>
</tr>
<tr>
<td>Familiarity with Organization</td>
<td>-0.0061</td>
<td>0.0098 ***</td>
</tr>
<tr>
<td>Experience with Organization</td>
<td>0.2478</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Prize</td>
<td>0.0040</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Number of Competitors</td>
<td>0.0217</td>
<td>0.0029 ***</td>
</tr>
<tr>
<td>Previous team performance</td>
<td>0.2745</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Tenure</td>
<td>0.0009</td>
<td>&lt;.0001 ***</td>
</tr>
<tr>
<td>Competition Duration</td>
<td>-0.0010</td>
<td>0.0827 *</td>
</tr>
<tr>
<td>Number of previous competitions</td>
<td>-0.1246</td>
<td>&lt;.0001 ***</td>
</tr>
</tbody>
</table>

Negative sign for coefficient means positive effect on sustained participation since the sustained participation is measured by the number of days between previous participation and current participation.

* Level of significance: p < 0.1
** Level of significance: p < 0.05
*** Level of significance: p < 0.01
5. Conclusions

5.1. Discussion and Implications

The results of our study yield several important findings. First, the results indicate that the number of times team members have teamed up in the previous competitions has significant positive effect on sustained participation. The explanation for this finding is that the prior collaboration and ties indicates the likelihood that the team members can work well with one another and win the competition [50]. Second, the results indicate that the number of times the crowdsourcing organization have posted contests in the platform have significant positive effect on sustained participation. It means that when individuals are more familiar with the organization, they are more inclined to come back to the platform in the shorter period of time. The explanation for this finding is that organizations that post more competitions in the platform provide more information about the themselves and individuals can get familiar with the organizations [53] and consider participating in the competitions from those organizations. Third, the results show that the number of times individuals have worked on the competitions from the same organization has negative effect on sustained participation. It means that when individuals have experience of working on the problems of the same organization, it takes longer time to come back and work on the problems of that organization again. This finding is in contrast with what we hypothesized. One possible explanation for this finding is that since individuals have experience of working for that organization and are aware of the expectations, therefore they take longer time to prepare themselves for the competition and submit the solutions. Moreover, in the meanwhile the individuals probably are active in the kernels and discussion forums of that competition to gather as much information as they can. Finally, our results support the findings in the literature. We found significant effect of prize, number of competitor teams, previous team performance, tenure, competition duration, and the number of competitions the individual has participated before the current competition on the sustained participation. Our findings supports the findings in the literature. But, In contrast to previous work that has demonstrated the positive effect of prize on individuals’ sustained participation [19], the present study shows the negative effect of the prize. One possible explanation is that the competitions that offer higher prices are usually more complex and competitive that makes the likelihood of winning the competition less. Therefore, individuals are less inclined to come back to the platform in the shorter period of time.

Our research results highlight the importance of environment-specific and organization-specific factors on individuals’ sustained participation. Our research suggests that crowdsourcing platform sponsors should focus on encouraging individuals to collaborate and improve communications with each other. Moreover, providing more information about the crowdsourcing organizations and creating positive image of them will result in sustained participation of individuals in the crowdsourcing contest platforms. Our study provides guidance for the research into existing and new practices to study and improve the sustained participation of crowds in crowdsourcing contest platforms.

5.2. Limitation and Future Research

This study contains several limitations that should lead to future research. First, we conducted this research entirely within the Kaggle environment. To generalize the findings, future research can investigate the individuals’ sustained participation in other crowdsourcing contest platforms (e.g. TopCoder). Second, for the environment-specific and organization-specific factors we focused on individuals’ previous interactions with the team members and organizations. Future research can focus on other environment-specific and organization specific factors including the team members’ knowledge and skills, the organizations’ reputations and brand-strength and etc. Third, we conducted empirical study on publically available data from Kaggle platform and we did not capture the individuals’ perception on sustained participation. To further our findings and to get an in-depth understanding of individuals’ sustained participation in crowdsourcing contests, future studies can conduct a qualitative study to complement this study.

6. References


