Association for Information Systems

AIS Electronic Library (AISeL)

ICIS 2019 Proceedings

Crowds, Social Media and Digital Collaborations

Team Formation and Performance in Online Crowdsourcing Competitions: The Role of Homophily and Diversity in Solver Characteristics

Indika Dissanayake
The University Of North Carolina At Greensboro, i_dissan@uncg.edu

Sridhar Nerur
University of Texas at Arlington, snerur@exchange.uta.edu

Jie Zhang
University of Texas Arlington, jiezhang@uta.edu

Team Formation and Performance in Online Crowdsourcing Competitions: The Role of Homophily and Diversity in Solver Characteristics

Completed Research Paper

Indika Dissanayake
University of North Carolina at Greensboro
Greensboro, NC 27402, USA
i_dissan@uncg.edu

Sridhar Nerur
University of Texas at Arlington
Arlington, TX 76019, USA
snerur@uta.edu

Jie Zhang
University of Texas at Arlington
Arlington, TX 76019, USA
jiezhang@uta.edu

Abstract

Rapid advances in Information Technology (IT) have enabled organizations to venture beyond their workforce to seek solutions to vital business problems through online crowdsourcing platforms. Such platforms are characterized by geographically dispersed self-organizing teams that compete with one another to evolve the best solutions to challenging issues that confront organizations. Despite the growing popularity of crowdsourcing, there is a paucity of empirical research on: a) how participants on these platforms form teams; and b) how the composition of these teams affects their performance. In this paper, we investigated solvers' teaming preferences and their impact on performance in an online crowdsourcing competition platform. Specifically, we explored demographics and acquired characteristics as potential predictors of the choice of a teammate. The findings of this study provide insights on the role of homophily and diversity of solver characteristics on team formation and performance in crowdsourcing competitions.

Keywords: Crowdsourcing, team formation, team performance, virtual teams, homophily

Introduction

Social media technologies have made it feasible and cheaper for organizations to draw on expertise from around the globe. Rapid growth of intermediary networks, such as Kaggle, Innocentive and TopCoders, has made it possible for companies to invite large communities to solve their problems via open calls. This practice – commonly referred to as crowdsourcing – is defined as "the process of taking tasks that have traditionally been performed by employees or contractors and outsourcing them through an open call to an undefined group" (García et al. 2012, p.373; Howe 2008). Crowdsourcing technologies and platforms allow individuals who are typically unaffiliated with an organization to organize through digital platforms and find solutions to vital business problems. Many organizations successfully use this web-based business model to find creative solutions to their problems. A celebrated example of this is Netflix's competition to use the "wisdom of the crowd" to gain insight into better movie recommendation algorithms.

Crowdsourcing has been used in a diverse range of disciplines, including marketing, new product development, healthcare, software development, and customer service (Dissanayake et al. 2015). A survey conducted in 2014 indicated that 85% of the best Global brands used crowdsourcing to find solutions for their business problems (Olenski 2015). Studies have discussed multiple benefits of crowdsourcing, including lower cost (Huang et al. 2011), higher quality solutions (Girotra et al. 2010), lower risk (Boudreau et al. 2011), diverse perspectives that transcend boundaries/borders, and potentially more novel discoveries (Wazny 2017). Performance improvement of crowdsourcing model has become an important research topic for both researchers and practitioners. With this growing popularity, crowdsourcing has emerged as a mainstream research area in the information system (IS) domain (see Dissanayake et al. 2015; Lee et al. 2018; Leimeister et al. 2009).

Prior studies in crowdsourcing have mainly investigated how contest design characteristics such as complexity, duration, and reward structure influence performance outcomes such as quality of the winning solution (Boudreau et al. 2011), number of solvers participating in the contest, and contest completion rate (Yang et al. 2009). In addition to contest design characteristics, solver characteristics (demographic or acquired) also contribute to crowdsourcing performance. In many crowdsourcing competition platforms, solvers organize into teams and participate in contests. Crowdsourcing teams are different from traditional organizational teams, for they are not only global and virtual but also self-organizing. Moreover, these teams compete against each other for a reward. Thus, while teaming in crowdsourcing can enhance the odds of winning, it also means that members now incur a cost of sharing the reward. Therefore, teaming up with solvers who possess the right mix of characteristics is important to enhance the chances of winning without incurring a huge cost. Despite the popularity of crowdsourcing, not many studies have investigated how solvers organize into teams in these competitions and how team composition impacts the quality of solutions. Our study aims to fill this void. Specifically, we empirically explore answers to the following research questions:

1) What factors do solvers consider when selecting their teammates in crowdsourcing competitions?

2) How do the differences and similarities among team members influence performance of the team? In particular, we investigated how language similarity, skill disparity, tenure disparity, past collaborations, and geo distance play a role in teammate selection and team performance.

We relied on the theoretical underpinnings of homophily, social identity theories and diversity literature to build our hypotheses and tested them using a large dataset from Kaggle.com, a popular online crowdsourcing platform. Despite the fact that crowdsourcing helps to add diversity to the innovation process, our results showed that homophily plays a key role in solvers' team formation decisions. In other words, solvers prefer to team up with similar others who speak the same language, live in the same or neighboring countries, have comparable skills, are similar in terms of tenure, and with whom they have had some prior collaboration experience. On the other hand, our results show that team diversity is a determinant of team performance. Diversity enhances creativity and is important for solving these intellectually demanding tasks. Specifically, team members' skill disparity, tenure disparity, and geo distance are positively correlated with team performance.

Our study makes several theoretical and practical contributions. First, diversity is an important factor in online collaboration environments in general, and crowdsourcing environments in particular. Surprisingly, there is a paucity of empirical research on the relationship between diversity and both team formation and team performance in a crowdsourcing context. Our study helps to fill this gap in the literature. Second, our study extends the extant literature on diversity beyond traditional co-located and virtual teams in organizational settings. Third, our study reveals opposing roles of diversity in team formation and performance in crowdsourcing teams. Finally, findings of this study provide practical implications for crowdsourcing solvers, seekers or sponsors, and platform providers.

The remainder of the paper is organized as follows. The next section presents the theoretical backdrop for our hypotheses. Subsequently, we describe our data and measures and then present our findings. Finally, we conclude our paper with a discussion of the study's theoretical and managerial implications, followed by limitations and directions for future research.

Theoretical Framework and Hypotheses

Crowdsourcing teams

Crowdsourcing is a problem-solving model where clients (seekers) seek solutions to their business problem with the aid of digital platforms (platform-providers). In these online platforms, geographically distributed individuals (solvers) work in self-organized teams whose members collaborate and compete with other teams to find better solutions to seekers' problems (Dissanayake et al. 2015). The competitions are typically knowledge intensive and intellectually demanding, and often give a reward to the winning team. (Ren et al. 2016) noted three main differences between online self-organizing teams and co-located teams in organizational or laboratory settings. These include: "face-to-face versus computer-mediated communications, organizational hierarchy versus meritocracy, closed and static versus open and fluid group memberships (Matins et al.2014)" (Ren et al. 2016, p.1671). In addition to these differences, as mentioned earlier, it is common for crowdsourcing teams to compete with one another for rewards.

There are some notable differences between crowdsourcing teams and traditional organizations teams. First, unlike in face-to-face interactions, individual differences are not readily available in computer-mediated interactions (Ren et al. 2016). In such environments, online profiles are commonly used to learn information about potential teammates. For example, in Kaggle, user profiles display solvers' information, such as their country, tenure, and skill score based on their prior performance. Second, in the absence of organizational hierarchies a participant's status mainly originates from their performance and recognition within the community (Ren et al. 2016). Furthermore, participants have more freedom when teaming up because there is no fear of loss of job or promotion opportunities. Third, group memberships are more fluid in the absence of strong entry and exit barriers (Ren et al. 2016). Fourth, teams compete for rewards; therefore, adding a member could enhance the chances of winning but it also means that there is a cost incurred because the reward has to be shared among more individuals. All these factors make crowdsourcing teams unique, thereby providing a strong motivation for us to explore the effect of similarities and differences in solver characteristics on team formation and performance in this context.

Homophily in team formation

Homophily refers to "the desire to associate with those similar to you" (Gompers et al. 2017, p.3) or the notion that "birds of a feather flock together" (e.g., Muller et al. 2014, p.779). Prior studies in various social networks such as marriage (Fiore and Donath 2005), friendship, support, and professional networks, provide evidence for the existence of homophily (McPherson et al. 2001). Mascia et al. (2011) argued that people are more likely to connect with similar others because "similarity of personal characteristics implies common interests and worldviews and best explains the formation of expressive ties based on interpersonal attraction" (Ibarra 1992, p.423). There are two main sources of homophily that we consider in our study - homophily arising from demographic similarities (e.g., same country, language, gender) and homophily that occurs because of acquired characteristics (e.g., tenure/ experience, skill, education) (Gompers et al. 2017). Social identity theory also supports the view that people would like to engage in activities with others who share common identity attributes (e.g., geographical location) (Muller et al. 2014).

Gompers et al. (2017) conducted an experiment with MBA students and found that ethnicity is one of the strongest homophilic forces. Specifically, the study claimed that South Asian, East Asian, Latin American, Middle Eastern and European students like to pair up with students from the same region (Gompers et al. 2017). In a different study, Ruef et al. (2003) investigated team composition of founding entrepreneurial teams and claimed that the probability of teaming up with a person with the same ethnic background is higher compared to a random matching process. They also noted that the "the importance of geographic proximity in group formation has long been recognized in both the microsociological (Goffman 1963) and macrosociological (Hawley 1950) literature" (Ruef et al. 2003, p.203). Furthermore, language and country have been identified as proxies for culture, and people prefer to team up with others coming from the same cultural backgrounds (Daniel et al. 2013). As per social identity theory, geographical location often serves as an indicator of similarity (Muller et al. 2014). Barriers to communication and coordination among geographically dispersed individuals have been significantly reduced by rapid advancements in digital technology. Despite this, people still prefer to team up with others who are similar to them, or someone they already know. There can be multiple reasons for such behavior. For example, speaking the same language not only makes it easier for team members to get to know one another, but also makes it easier for them to communicate and coordinate activities. This is particularly helpful in the team formation stage. Thus, we argue that solvers like to team up with others who have similar demographic characteristics.

H1a: Geo distance among team members is negatively related to team formation.

H2a: Language similarity among team members is positively related to team formation.

In addition to demographic characteristics, we contend that solvers would also like to collaborate with others possessing similar acquired characteristics, such as skill, tenure, and past collaboration experiences. For example, Gompers et al. (2017) argued that even though functional diversity could lead to performance improvements, people still would like to collaborate with others who have similar experience. Consistent with this reasoning, we argue that solvers feel comfortable to team up with those who have similar level of skills and tenure. High tenure disparity means high variability among team members in the time they have worked on the platform. The viewpoints and perspectives that team members bring to bear on the problem may be different if tenure diversity among them is high. We argue that solvers would have fewer conflicts and be more comfortable to team up with others who have similar viewpoints.

In an open source context, Hahn et al. (2008) showed that cohesion cues are important when forming online teams. Cohesion cues address individuals' preference for repeat collaborations when they have had prior associations with existing members of the team. Specifically, past collaboration plays an important role in virtual teaming environments where there is lack of opportunity for face-to-face communications. Prior collaborative experience facilitates accurate estimation of the risk and uncertainty involved in collaborative activities (Hahn et al. 2008), increases the availability of information related to skills and capabilities of the members (Hahn et al. 2008), fosters efficient and effective communication, develops human capital and transactive memory systems (Huckman et al. 2009), and enhances psychological safety (Huckman et al. 2009).Thus, we argue that solvers are likely to team up with others who have similar acquired characteristics.

H3a: Skill disparity among team members is negatively related to team formation.

H4a: Tenure disparity among team members is negatively related to team formation.

H5a: Past collaboration experience among team members is positively related to team formation.

Diversity in team performance

Prior literature on the relationship between diversity and team performance have been largely inconclusive. On the one hand, diversity improves team performance through integration of a variety of information. On the other, it hinders team performance because of inherent conflicts and coordination inefficiencies (Ren et al. 2016). Prior literature has identified several dimensions of diversity, ranging from social demographic attributes (e.g., age, gender, race, and nationality) to acquired/ informational attributes (e.g., tenure, education, and expertise) to much deeper personal attributes such as individual personalities and values (Ren et al. 2016). In traditional organizational settings, scholars have provided evidence for the association between group diversity and performance (Horwitz and Horwitz 2007; Williams and O'Reilly 1998). Specifically, in an extensive review of 80 empirical papers on diversity, Williams and O'Reilly (1998) found that heterogeneity in age, tenure, race, and sex hinder group processes and performance. Even though diversity has been widely investigated in off-line teams, it has received scant or no attention in the literature on online self-organizing teams, the notable exception being (Ren et al. 2016). We explored team diversity and performance relationship along the same dimensions discussed in the team formation section above.

Diversity in the crowd is presumably an essential element for the success that participants are able to achieve in a crowdsourcing task. When people from diverse backgrounds examine a problem from different perspectives and arrive at a consensus, there is a high likelihood that the solution will be closer to the optimal solution. This represents the "wisdom of the crowd" phenomenon that has lately received a lot of attention (Bernstein 2017). On the other hand, a homogeneous crowd could apply the same lens or perspective to a problem and evolve a solution quite far from optimal. This scenario is a characteristic of the "madness of the crowd" phenomenon (Bernstein 2017). In the open-source software development context, Daniel et al. (2013) found that culture-based separation diversity positively influences market success. They used country and language as proxies for culture-based separation. In fact, the result was contrary to what they had hypothesized in the paper. This shows that the general belief that team members coming from the same country, or speaking the same language, experience increased performance due to fewer group conflicts and enhanced communication efficiencies may not hold true in all contexts. In a crowdsourcing context, "the higher performance of heterogeneous groups significantly overweighs the risks of potential conflicts" (Ivanov 2017). Moreover, the literature on psychology suggests that culture influences the cognitive thinking process of individuals. For example, Nisbett et al. (2001) notes that East Asians are more holistic thinkers while westerners are more analytical thinkers. Since it is important to look at these cognitively demanding tasks from multiple perspectives, we argue that teaming up with people from diverse geographical backgrounds will increase performance. Thus, we hypothesize:

H1b: Geo distance is positively related to team performance.

H2b: Same language (of team members) is negatively related to team performance.

Skill disparity considers both differences in the level of skill as well as the direction. To clarify, consider two three-member groups, one consisting of one high-skill member and two low-skill members (group A) and the other consisting of two high-skill members and one low-skill member (group B). If we consider only the separation or difference in level of skill (i.e., std dev), it is the same for both teams. However, high-skill members have more influence on performance. If we consider the disparity (i.e., coefficient of variation), it takes into account both the direction and the separation. Thus, skill score diversity has a greater impact on performance when a team consists of more low-skill score members (group A) than a team with more high-score members (group B). Thus, we hypothesize:

H3b: Skill disparity is positively related to team performance.

Prior studies are inconclusive with regard to the impact of tenure diversity on team performance. Daniel et al. (2006) indicated that tenure diversity increases performance in open source software development teams due to synergetic effects arising from interactions between old-timers and new-comers. Old timers have experience within the platform and are more likely to have worked on similar projects, while newcomers bring new skills and perspectives. Furthermore, based on data from traditional organizational teams and online self-organizing teams such as in Wikipedia, scholars have shown that although old timers have necessary skills and experience, their motivations could diminish with time (Ren et al. 2016). Thus, teams will benefit if they can combine the experience of the old with the novelty of perspectives that newcomers bring to bear on a task. However, very high levels of tenure disparity could engender conflicts because of opposing viewpoints that members cannot easily resolve. For example, Williams and O'Reilly (1998) showed evidence for negative effects of high tenure disparity, such as reduced communication, lack of social integration, and increased conflicts. Therefore, we expect tenure disparity to be beneficial up to a point, after which it is detrimental to performance. Thus, we hypothesize:

H4b: Tenure disparity has an inverted U-shaped relationship with team performance.

If team members always collaborate with the same set of individuals, they may start looking at problems from the same perspective. While this would help to reduce conflicts and influence team cognitions (He et al. 2007), it will be detrimental to innovativeness. New interactions can bring new knowledge and fresh perspectives that may enhance creativity. Thus, we hypothesize:

H5b: Past-collaboration experience is negatively related to team performance.

Figure 1 summarizes the research model.

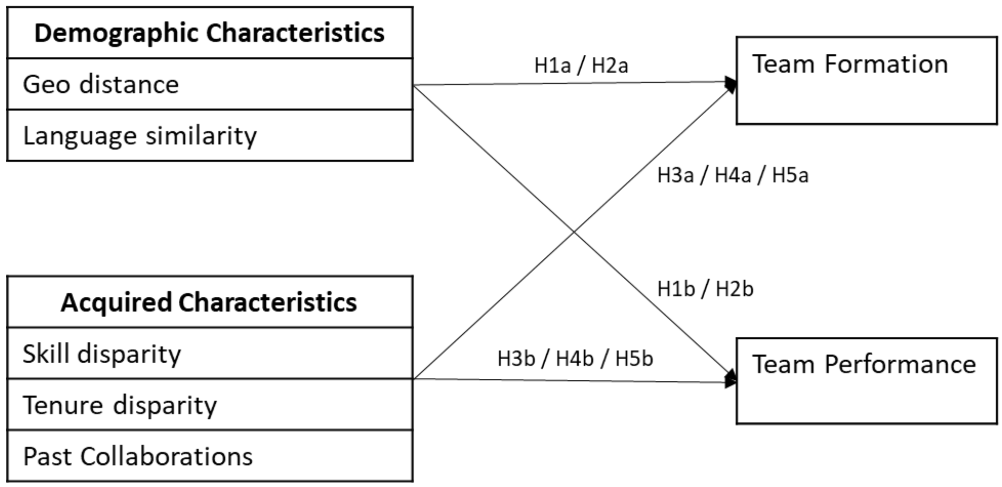

Figure 1. Research Model

Research Context

We examined solvers' teaming decisions on Kaggle, a crowdsourcing platform for predictive modeling competitions. Companies, government, and researchers provide datasets and problems for statistical and analytics outsourcing, while participants compete by producing the best solutions as determined by certain accuracy measures. The competition host pays rewards in exchange for the intellectual property behind the winning model. Geographically distributed individuals are self-organized into teams to compete in these competitions. Each participant or participating team can submit multiple solutions. The platform evaluates all submissions in real-time and provides participants instant feedback through a leaderboard, which gives participants information on the predictive accuracy of their models and their relative positions (i.e., rank) in the contest. The platform also provides an online profile for each solver that includes personal information such as geographical location, solver's performance history, date they joined the platform, and overall performance score based on rankings in contests in which he/she participated. Solvers use these online profiles to learn about their potential teammates. In addition, the platform facilitates interactions among solvers through discussion forums.

Analysis and Results

Team formation

Dependent variable

Teaming (solver-solver) matrix: We selected solvers who participated in teams in the year 2016 and created a dichotomous (i.e., binary) matrix of solver-solver pairs based on whether or not they were on the same team. After eliminating solvers with missing country information, our final sample consisted of a matrix of 1378 x 1378 solver-solver pairs.

Independent variables

We used performance and profile data from 2010 to 2015 for the 1378 solvers previously identified to derive the following matrices: geo distance matrix, skill disparity matrix, tenure disparity matrix, and language similarity matrix. It must be noted that each matrix is of size 1378 x 1378 and represents an independent dyadic variable.

Geo distance (solver-solver) matrix: We used latitude and longitude data of each country to calculate the nautical distance between two countries. The nautical distance between countries of each solver pair was calculated for the entire sample of solver pairs.

Language similarity (solver-solver) matrix: For each pair, language similarity was considered to be 1 if their countries' primary language was the same and a 0 otherwise.

Tenure disparity (solver-solver) matrix: This is the coefficient of variation of tenure of each pair of solvers. A tenure of a solver was calculated as the difference in days between the date they joined the platform and December 31, 2015.

Past collaboration (solver-solver) matrix: Solver-solver value for this matrix was a 1 if the pair had collaborated as members of the same team during the period 2010-2015 and a 0 otherwise.

Skill disparity (solver-solver) matrix: This is the coefficient of variation of skill score of each pair of solvers before December 31, 2015. Kaggle uses a formula to derive skill scores based on their performance on the platform.

Team formation analysis

Our data were relational (dyadic) in nature and solver-solver pairs formed the primary unit of analysis. Thus, we chose Double Dekker Semi-Partialling Multiple Regression Quadratic Assignment Procedure (MRQAP) method to run our analysis (Dekker et al. 2007). This method is commonly used in social network analysis models when the dependent variable is a social relational matrix and explanatory variables are also relational matrices derived from attributes (e.g., gender, tenure, age) of the actors (Mascia et al. 2011). Tenure disparity, skill disparity, and geo distance are relational matrices based on continuous variables. Basically, these variables represent the distance between values for each pair of solvers. A small difference implies that they are similar with respect to the given attributes. A negative sign for these variables indicates that homophily of solvers positively predicts their being on the same team. Language similarity and past collaborations are binary matrices. The positive sign for these variables indicates that solvers are likely to team up with similar others. We conducted the analysis using UCINET 6.623, a popular software package for analyzing social networks (Borgatti et al. 2002). Table 1 summarizes the results.

MRQAP Regression Results	
	Teaming
Skill disparity	-0.0114 ***
Tenure disparity	-0.0057 ***
Geo distance	-0.0217 ***
Language similarity	0.0230 ***
Past collaboration	0.2398 ***
*** p < 0.01, ** p < 0.05, * p < 0.1	

Table 1. MRQAP Regression Results

The coefficients of language similarity and past collaboration are positive and significant while geo distance, skill disparity, and tenure disparity are negative and significant. Thus, hypotheses 1a, 2a, 3a, 4a, and 5a are all supported. Our results suggest that homophily in general has a positive impact on team formation.

Team performance

Our final sample for the analysis of team performance consists of 4571 teams that participated in 213 projects from April 2010 to June 2016. We eliminated all single member teams and their associated projects from our sample.

Dependent variable

Team performance: Our main dependent variable is team performance. We used the final ranking of the team (relative performance) as a measure of team performance.

Independent variables

Skill disparity: This is the coefficient of variation of team members' skill scores at the beginning of the contest. A solver skill score is dynamically updated based on their team performance in previous contests. Kaggle derives it using variables such as team rank, team size, number of teams, and the timing of each contest in which they have participated.

Tenure disparity: Following prior studies, we used the coefficient of variation of tenure of team members as a measure of tenure disparity (Harrison and Klein 2007; Ren et al. 2016). The tenure of a member is the difference in days between the date they joined the platform and a project start date. If a solver joins the platform after a project is launched, the value is considered to be 0.

Geo distance: As mentioned earlier, nautical distance between countries was calculated for each pair of members in a team. The average of geo distance for all the pairs in a team was used as the team's geo distance.

Language similarity: The average of language similarity for all the pairs in a team was used as the measure of team's language similarity.

Past collaboration: This is the number of times a team member had collaborated with other team members in different projects prior to the formation of the current team. We calculate the average of all team members past collaboration ties as the team's past collaboration value.

Team skill: This is the average skill score of all the members in a team. Prior studies in crowdsourcing have identified average skill as a predictor of team performance. Thus, we add it as a covariate to the model to control for the effect (Dissanayake et al. 2015).

Team size: This is the number of members in a team. Prior studies have shown that team size has an impact on team performance (Girotra et al. 2010). It was therefore used as a control variable.

Team performance analysis

As mentioned earlier, our main dependent variable was team rank. Therefore, we used rank ordered logistic regression model with reverse preference order (lower the rank value better the performance). Model was tested on Stata 14. We also controlled for contest specific characteristics using contest fixed-effects.

$$\begin{aligned} Performance_{ij} = {} & \alpha_0 + \alpha_1\, TenureDisparity_{ij} + \alpha_2\, TenureDisparity_{ij}{}^2 + \alpha_3\, GeoDistance_{ij} \\ & + \alpha_4\, LanguageSimilarity_{ij} + \alpha_5\, SkillDisparity_{ij} + \alpha_6\, PastCollaboration_{ij} \\ & + \alpha_7 TeamSize_{ij} + \alpha_8 TeamSkill_{ij} + \delta_j + \varepsilon_{ij} \qquad (1) \end{aligned}$$

where α_k (*k*= 0…8) represents the coefficients of the variables, δ_j is the coefficient

for the contest fixed effect, subscript *i* is for the team, and subscript *j* is for the contest.

Rank-ordered Logistic Regression Results		
Reversed team rank	Performance	
	Model 1	Model 2
Skill disparity	0.1414 ***	0.1319 ***
Tenure disparity	17.2680 ***	25.3254 ***
Tenure disparity sq		-759.5181*
Geo distance	0.0024 ***	0.0023 ***
Language similarity	-0.0539	-0.0578
Past collaboration	-0.0074	-0.0082
Team skill	0.0000 ***	0.0000 ***
Team size	0.0443 ***	0.0431 ***
Log likelihood	-12529.71	-12527.73
Sample size	4,571	4,571
*** p < 0.01, ** p < 0.05, * p < 0.1		

Table 2. Rank-ordered Logistic Regression Results

The coefficients of geo distance, skill, and tenure are positive and significant. Thus, hypotheses 1b, 3b, and 4b are supported. Language similarity and past collaboration do not have any significant impact on performance. Thus, hypotheses 2b and 5b are not supported. All control variables have significant effect on team performance. In general, results suggest that diversity has a positive impact on team performance. The variance inflation factor (VIF) <4 suggests there is no evidence of multicollinearity in our model.

Robustness tests

We conducted an additional analysis to investigate the impact of extremely skillful solvers on team performance. Kaggle assigns a performance tier to solvers based on their overall performance in the platform. In our sample, performance tier of solvers varies from 2 to 10, where 10 represents the best performance. Based on the tier of the highest tier group member, we categorized teams into three tier groups. Low tier group (highest tier value <4), medium tier group (4 ≤ highest tier value <10), and high tier group (highest tier value = 10). Percentages of low, medium, and high tier groups were 10%, 75%, and 15%, respectively. We considered the lowest tier group as the base group and added dummy variables to represent other groups and re-ran the analysis. As shown in Table3, having a member belonging to the high tier group has a significant positive impact on performance. However, the results still support our main argument that diversity is positively related to performance.

Rank-ordered Logistic Regression Results	
Reversed team rank	Performance
Skill disparity	0.0617 **
Tenure disparity	10.9511 **
Tenure disparity sq	-484.0606
Geo distance	0.0017 **
Language similarity	-0.0465
Past collaboration	-0.0475 ***
Medium tier dummy	0.1756 *
High tier dummy	1.4321 ***
Team skill	0.0000 ***
Team size	0.0283 ***
Log likelihood	-12284.08
Sample size	4,571
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$	

Table 3. Rank-ordered Logistic Regression Results with Tier Groups

Discussion

Key findings

Our results showed that solvers are likely to team up with others who have similar geographical or acquired characteristics. Specifically, solvers like to team up with others who speak the same language, come from the same geographical region, have similar level of skill and tenure, and with whom they have had prior collaboration experience. Interestingly, our results suggest that diversity has a significant impact on team performance in crowdsourcing contests. Heterogeneous teams showed better performance than homogeneous teams. Specifically, skill disparity, tenure disparity, and geo distance showed positive impact on team performance.

Theoretical implication

This study makes several contributions to the IS literature. It elucidates the importance of understanding the dynamics of group formation and performance in online self-organizing teams. Prior studies on online collaborations, virtual teams, or diversity have focused mainly on diversity in traditional organizations and virtual teams, with Ren et al. (2016) being a notable exception. Our study complements these studies by investigating diversity in a new context. Furthermore, it is a valuable contribution to the diversity literature, which has largely been inconclusive with regard to the relationship between diversity and performance. This study provides additional insight into the diversity-performance relationship, particularly in the context of crowdsourcing teams. Companies or organizations use crowdsourcing as a way to engage diverse pools of solvers to evolve creative solutions for their business problems. Our results confirm our hypothesis that diversity has a positive influence on performance. In crowdsourcing, positive effects of diversity appear to outweigh some of the reported negative effects of diversity, such as communication and coordination inefficiencies.

Even though diversity in group performance has been investigated extensively in other contexts, diversity as a basis for group formation has rarely been explored. Unlike in face-to-face communication settings, solvers interact with each other through digital media. (Carte and Chidambaram 2004) indicated that visual anonymity reduces the salience of individual diversity in computer-mediated teams. However, in contemporary work environments, digital technologies play a significant role in online collaboration and team members have numerous ways to learn about each other. For example, in our context, solvers can use public profiles of potential teammates to learn about their demographic and acquired characteristics. The study supports Ren et al.'s suggestion that the notion of visibility has to be revisited in online teams (Ren et al. 2016).

Finally, we are the first to investigate two opposing influences of diversity on team formation and performance in the same setting. This is especially important for self-organizing teams where there is no control over the team formation process. This opens up a new window for diversity studies.

Practical implication

The findings of this study provide several practical implications for seekers, solvers and platform providers of crowdsourcing contests. All three parties will benefit from improving solution quality through enhancing team performance. Teams can enhance the chances of winning rewards by forming a group with the right combination of solvers. Platform providers can use these insights to provide guidance to solvers during the team formation process, which, in turn, can lead to better performance. By doing so, they can attract more seekers and solvers to their platforms, which, in turn, will increase revenues through platforms charges from seekers. Seekers, on the other hand, will be able to find better solutions for their problems.

The key insight of our study is that the diversity in acquired characteristics (e.g., skill sets, tenure) among team members facilitates creative problem-solving in a crowdsourcing environment characterized by knowledge intensive tasks. In contrast, team homophily in terms of language similarity, geographical proximity, and familiarity because of past collaboration may facilitate efficient communication and coordination. These findings could be generalized to other crowdsourcing settings that focus on cognitively demanding knowledge work, such as software development and research and development (R&D) projects.

In addition to these main contributions, our study also generates some insights into the effect of visibility of solver attributes presented in their online profiles. Future research can be done in this area to identify the advantages and disadvantages of disclosing different solver attributes in their online profiles. The platform providers could perhaps use these findings to design online profiles to better assist the team formation process.

Limitation and future research

The findings of this study should be interpreted in the context of its limitations. Understanding these limitations also creates opportunities for future studies. First, our results are based on a limited number of geographical and acquired diversity dimensions. More insights could be generated by adding a more comprehensive set of diversity measures. Second, the role of diversity could depend on the nature of the problem at hand as well as on the characteristics of the crowd. Kaggle focuses only on machine learning and predictive analytics problems. Therefore, generalizability of the results to other types of crowdsourcing platforms could be limited. Future studies can study other crowdsourcing platforms to affirm our findings and/or to generate deeper insights about these relationships. Third, our data is limited to the Kaggle platform and does not take into account the information that solvers could have possibly obtained about other solvers from other social media sites. Fourth, although the platform facilitates online team formation, it is conceivable that some of the teams are formed offline. However, what still matters is the preference of an individual to team-up with someone in the same geographical location.

Conclusion

Crowdsourcing teams are increasingly used by organizations to find creative solutions for their business problems. Forming a team with the right people is the key to finding a better solution to these problems. This paper attempts to investigate how solvers select their team members in crowdsourcing competitions and how these decisions impact their performance. Understanding factors that influence solvers teaming decisions and its consequential impact on performance would not only be helpful to seekers and solvers, but would also provide insights to platform providers. Platform providers can use these insights to design their sites and competitions in such a way as to enable solvers to form better teams, thus enhancing their performance and increasing their likelihood of finding the optimal winning solution.

References

Bernstein, W. 2017. "Best Practices for Diversity in Crowdsourcing," IdeaScale, , January 19. (https://ideascale.com/diversity-in-crowdsourcing/, accessed September 23, 2018).

Borgatti, S. P., Everett, M. G., and Freeman, L. C. 2002. "UCINET 6 for Windows: Software for Social Network Analysis," Harvard, Ma: Analytic Technologies.

Boudreau, K. J., Lacetera, N., and Lakhani, K. R. 2011. "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis," Management Science (57:5), pp. 843–863.

Carte, T., and Chidambaram, L. 2004. "A Capabilities-Based Theory of Technology Deployment in Diverse Teams: Leapfrogging the Pitfalls of Diversity and Leveraging Its Potential with Collaborative Technology," Journal of the Association for Information Systems (5:11), p. 448-471.

Daniel, S., Agarwal, R., and Stewart, K. 2006. "An Absorptive Capacity Perspective of Open Source Software Development Group Performance," ICIS 2006 Proceedings, (Vol. 59) pp. 927–936.

Daniel, S., Agarwal, R., and Stewart, K. J. 2013. "The Effects of Diversity in Global, Distributed Collectives: A Study of Open Source Project Success," Information Systems Research (24:2), pp. 312–333.

Dekker, D., Krackhardt, D., and Snijders, T. A. B. 2007. "Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions," Psychometrika (72:4), pp. 563–581.

Dissanayake, I., Zhang, J., and Gu, B. 2015. "Task Division for Team Success in Crowdsourcing Contests: Resource Allocation and Alignment Effects.," Journal of Management Information Systems (32:2), pp. 8–39.

Fiore, A. T., and Donath, J. S. 2005. "Homophily in Online Dating: When Do You like Someone like Yourself?," in CHI '05 Extended Abstracts on Human Factors in Computing Systems - CHI '05, Portland, OR, USA: ACM Press, p. 1371. (https://doi.org/10.1145/1056808.1056919).

García, I., Guillot, M.-N., Zimányi, K., and Langman, J. 2012. "Book Reviews," The Translator (18:2), pp. 373–392.

Girotra, K., Terwiesch, C., and Ulrich, K. T. 2010. "Idea Generation and the Quality of the Best Idea," Management Science (56:4), pp. 591–605.

Gompers, P. A., Huang, K., and Wang, S. Q. 2017. "Homophily in Entrepreneurial Team Formation," National Bureau of Economic Research.

Hahn, J., Moon, J. Y., and Zhang, C. 2008. "Emergence of New Project Teams from Open Source Software Developer Networks: Impact of Prior Collaboration Ties," Information Systems Research (19:3), pp. 369–391.

Harrison, D. A., and Klein, K. J. 2007. "What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations," Academy of Management Review (32:4), pp. 1199–1228.

He, J., Butler, B. S., and King, W. R. 2007. "Team Cognition: Development and Evolution in Software Project Teams," Journal of Management Information Systems (24:2), pp. 261–292.

Horwitz, S. K., Horwitz, I. B. 2007. "The effects of team diversity on team outcomes: A meta-analytic review of team demography," Journal of Management (33:6), pp. 987–1015.

Howe, J. 2008. "Why the Power of the Crowd Is Driving the Future of Business," in Crowdsourcing, New York: Crown Business.

Huang, Y., Singh, P., and Srinivasan, K. 2011. "Crowdsourcing 'Blockbuster' Ideas: A Dynamic Structural Model of Ideation," in Proceedings of the 32nd International Conference on Information Systems, Atlanta, GA, pp. 19–22.

Huckman, R. S., Staats, B. R., and Upton, D. M. 2009. "Team Familiarity, Role Experience, and Performance: Evidence from Indian Software Services," Management Science (55:1), pp. 85–100.

Ibarra, H. 1992. "Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm," Administrative Science Quarterly (37:3), pp. 422–447.

Ivanov, E. 2017. "Crowdsourcing: Adding More Diversity to Your Innovation Process," , February 8. (https://innovationobserver.com/2017/02/08/crowdsourcing-adding-more-diversity-to-your-innovation-process/, accessed September 24, 2018).

Lee, H. C. B., Ba, S., Li, X., and Stallaert, J. 2018. "Salience Bias in Crowdsourcing Contests," Information Systems Research (29:2), pp.401-418.

Leimeister, J. M., Huber, M., Bretschneider, U., and Krcmar, H. 2009. “Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition,” Journal of Management Information Systems (26:1), pp. 197–224.

Mascia, D., Cicchetti, A., Fantini, M. P., Damiani, G., and Ricciardi, W. 2011. “Physicians’ Propensity to Collaborate and Their Attitude towards EBM: A Cross-Sectional Study,” BMC Health Services Research (11:1), p. 172. (https://doi.org/10.1186/1472-6963-11-172).

McPherson, M., Smith-Lovin, L., and Cook, J. M. 2001. “Birds of a Feather: Homophily in Social Networks,” Annual Review of Sociology (27:1), pp. 415–444.

Muller, M., Geyer, W., Soule, T., and Wafer, J. 2014. “Geographical and Organizational Distances in Enterprise Crowdfunding,” in Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, pp. 778–789.

Nisbett, R. E., Peng, K., Choi, I., and Norenzayan, A. 2001. “Culture and Systems of Thought: Holistic versus Analytic Cognition.,” Psychological Review (108:2), p. 291-310.

Olenski, S. 2015. “The State Of Crowdsourcing,” Forbes. (https://www.forbes.com/sites/steveolenski/2015/12/04/the-state-of-crowdsourcing/, accessed October 18, 2017).

Ren, Y., Chen, J., and Riedl, J. 2016. “The Impact and Evolution of Group Diversity in Online Open Collaboration,” Management Science (62:6), pp. 1668–1686.

Ruef, M., Aldrich, H. E., and Carter, N. M. 2003. “The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among US Entrepreneurs,” American Sociological Review, pp. 195–222.

Shiota, M. N., and Levenson, R. W. 2007. “Birds of a Feather Don’t Always Fly Farthest: Similarity in Big Five Personality Predicts More Negative Marital Satisfaction Trajectories in Long-Term Marriages.,” Psychology and Aging (22:4), p. 666.

Wazny, K. 2017. “‘Crowdsourcing’ Ten Years in: A Review,” Journal of Global Health (7:2). (https://doi.org/10.7189/jogh.07.020601).

Williams, K. Y., O’Reilly, C. A. 1998. "Demography and diversity in organizations: A review of 40 years of research," Research in Organizational Behavior. (20), pp. 77–140.