An Analysis of Design Process and Performance in Distributed Data Science Teams

Torsten Maier, Joanna F. DeFranco (jfd104@psu.edu), and Christopher McComb
The Pennsylvania State University

Abstract

Purpose – Often, it is assumed that teams are better at solving problems than individuals working independently. However, recent work in engineering, design, and psychology contradicts this assumption. This work examines the behavior of teams engaged in data science competitions. Crowdsourced competitions have seen increased used for software development and data science, and platforms often encourage teamwork between participants.

Design/methodology/approach –We specifically examine teams participating in data science competitions hosted by Kaggle. We analyze data provided by Kaggle to compare the effect of team size and interaction frequency on team performance. We also contextualize these results through a semantic analysis.

Findings – This work demonstrates that groups of individuals working independently may outperform interacting teams on average, but that small, interacting teams are more likely to win competitions. The semantic analysis revealed differences in forum participation, verb usage, and pronoun usage when comparing top- and bottom-performing teams.

Research limitations/implications- These results reveal a perplexing tension that must be explored further: true teams may experience better performance with higher cohesion, but nominal teams may perform even better on average with essentially no cohesion. A limitation of this research includes not factoring in team member experience level and reliance on extant data.

Originality/Value – These results are potentially of use to designers of crowdsourced data science competitions as well as managers and contributors to distributed software development projects.

Keywords software engineering teams, teamwork, technical teams, Kaggle dataset, data science, global teamwork, distributed teams

Paper type Research Paper

1. Introduction

Large and small businesses, sports clubs, and social groups rely on teams to complete complex shared goals. In this way, team efficiency and performance are often crucial to task completion. Teams are particularly common in engineering and business (Paulus, Dzindolet, & Kohn, 2011; Reiter-Palmon, Herman, & Yammarino, 2008). This work specifically investigates teams engaged in crowdsourced data science competitions. Many platforms for crowdsourcing compose and organize many different individuals into teams with the objective of solving a single problem. The crowdsourcing of engineering and design problems has seen increased interest in recent years (Ball & Lewis, 2018; Ren, Bayrak, & Papalambros, 2016). For these reasons, a team-based crowdsourcing platform is a useful laboratory to facilitate the study of team processes.

Despite the prevalence of teams and the important role that they play in increasingly popular crowdsourced platforms, there is a fundamental lack of knowledge in terms of how to structure and manage teams to maximize performance. For instance, there are a number of studies that indicate the benefits of small teams (Amason & Sapienza, 1997; Gooding, 2018; Karau & Williams, 1993; Lam, 2015; McComb, Goucher-Lambert, & Cagan, 2017; Mullen, Symons, Hu, & Salas, 1989; North, Linley, & Hargreaves, 2000) but there are also studies that indicate benefits of large teams (Majuika & Baldwin, 1991; Steffens, Terjesen, & Davidsson, 2012; Stewart, 2006). Similarly, there are studies that indicate teams achieve better performance when interating frequently (He, Butler, & King, 2007; Pentland, 2012; Pinto, 1990) and studies showing that teams that do not communicate are superior (Hinsz, Tindale, & Vollrath, 1997; McComb, Cagan, & Kotovsky, 2017a; Salomon & Globerson, 1989). There is a need for greater research on these concepts in an attempt to bring clarity to the literature. The current work investigates the effect of team size and interaction style on several measures of performance, and also performs an exploratory analysis of semantic indicators in team communication. A large existing dataset of team performance is employed.

We specifically examine teams participating in data science competitions hosted by Kaggle, a company that is dedicated to challenges of this nature (Carpenter, 2011). Data science shares a number of similarities with design (Kazakci, 2015), including large solution spaces and similarities in procedural approaches to finding solutions. Data science problems also have the potential for formalization through C-K theory (concept-knowledge theory) a theory presenting design as a series of logic expansion processes (Hatchuel & Weil, 2009). These characteristics make the data collected from the Kaggle platform useful for drawing potentially generalizable insights. The Meta Kaggle data set is public data provided by Kaggle that describes their competitions, submissions, and scores. These competitions range from underwater image analysis that assesses ocean health to analytics in cardiology that assesses heart functionality (Kaggle, 2017). The wide variety of team strategies employed by Kaggle competitors makes this extant data set a strong basis on which to build comparative analysis of different teaming patterns. This data is analyzed here to compare the effect of different team sizes and team interaction frequencies on team performance. Specifically, this study attempts to answer the following research questions:

RQ1: Do teams perform better than individuals in Kaggle competitions?

RQ2: How do team size and interaction type affect overall solution quality?

RQ3: How do team size and interaction type affect the probability of winning and magnitude of expected winnings?

RQ4: *Are there identifiable semantic indicators that align with team performance?*

The rest of this paper is organized as follows. Section 2 provides supporting background literature and uses this to introduce hypotheses aligning with each of the above research questions. Section 3 summarizes the Meta Kaggle dataset and details the variables used in this study. Section 4 provides an explanation of the methodology used in this work and outlines the numerical experiments that were conducted. Section 5 details the results of the numerical experiments and semantic analysis. Section 6 provides a discussion of limitations inherent in this study. Finally, Section 7 concludes the paper with a summary and a discussion of future directions.

2. Background

Even though teams are common in industry (Paulus et al., 2011; Reiter-Palmon et al., 2008), there is little agreement between definitions of the word *team* that have been proposed in the literature (for examples see (Cannon-Bowers, Salas, & Converse, 1993; Dyer, 1984; Morgan, Glickman, Woodard, Blaiwes, & Salas, 1986; Orasanu & Salas, 1993)). However, most definitions share two concepts: multi-agency (the composition of a team as two or more individuals) and communication (the ability of those individuals to exchange information) (McComb, Cagan, et al., 2017a). These two

concepts are fundamental to understanding team-based solutions. This work aims to provide a deeper knowledge base of both components within the context of software development for distributed data science challenges.

Interaction frequency is one commonly-cited factor that affects team performance and the sharing of information (Mello & Ruckes, 2006) and is a common metric used when modeling or investigating team dynamics (He et al., 2007; Pinto, 1990). Studies have found a positive relationship between communication frequency and team cognition and cooperation (He et al., 2007; Pinto, 1990). Some studies show team communication patterns to be the strongest indicator of team success and found that revising the break schedules at call centers to increase teammate socialization lead to increased performance (Pentland, 2012). E-mail and face-to-face interaction frequency was found to have a curvilinear relationship with team performance (Patrashkova-Volzdoska, McComb, Green, & Compton, 2003), with optimal performance located between high and low interaction frequency. In contrast to these findings, "best-member strategy" has shown to yield similar results to conventional teams (Yetton & Bottger, 1982). When employing a "best-member strategy", team members adhere to the decisions of the team member that they identify as "best." The weak interactive ties in such an approach facilitate a diverse perspective amongst team members, beneficial in the problem-solving process (Granovetter & Granovetter, 1983; McComb, Cagan, & Kotovsky, 2015). In the extreme, a team employing a "best-member strategy" becomes a *nominal team.* Nominal teams are groups of individuals who do not interact. Rather, they work independently towards a full solution, with the best individually-generated solution being provided as the solution of the team. A growing body of literature demonstrates that nominal teams may outperform interacting teams in tasks such as concept generation (Hinsz et al., 1997; Salomon & Globerson, 1989) and configuration design (McComb, Cagan, et al., 2017a).

RQ1 addresses the comparative performance of individuals, standard interacting teams (referred to here as "true teams"), and nominal teams. We hypothesize the following:

H1: On average, individuals will perform worse than true teams, but nominal teams will perform better than true teams.

True teams will outperform individuals simply due to the ability to perform more work. The hypothesis that nominal teams will outperform true teams is informed by a growing body of literature (Hinsz et al., 1997; McComb, Cagan, et al., 2017a; Salomon & Globerson, 1989).

Team size is another factor that affects team performance, but findings related to team size are mixed. Several studies have identified a positive relationship between team size and factors such as dissatisfaction and cognitive conflict (Amason & Sapienza, 1997; Mullen et al., 1989). In addition, computational work has demonstrated that smaller teams are capable of making decisions that have better axiomatic characteristics (McComb, Goucher-Lambert, et al., 2017). A meta-analysis of 31 published field studies found negative results for increased team sizes, specifically citing low efficiency as a leading cause (Gooding, 2018). In support of these findings, higher rates of social loafing (defined by Karau and Williams as the "reduction in motivation and effort when individuals work collectively compared with when they work individually or coactively") have also been linked to increasing group size (Karau & Williams, 1993; Lam, 2015; North et al., 2000). In contrast, other studies have shown that larger teams may experience increased effectiveness as the heterogeneity of tasks increases (Majuika & Baldwin, 1991). This would indicate that larger teams are more effective as scope increases, most likely, due to their enlarged set of experiences and backgrounds that can be brought to bear in solving a problem. Indeed, the degree of heterogeneity in new venture teams has been shown to have an effect on persistence and performance over time (Steffens et al., 2012). A quantitative review of 93 studies found a slight increase in performance of project and management teams when they include more members (Stewart, 2006). However, further analysis determined that task autonomy led to increased performance for specific tasks (Stewart, 2006). This may indicate that optimal team size is task-dependent, a premise supported by computational work (McComb, Cagan, et al., 2017a).

RQ2 addresses the effect of team size on performance, for both nominal teams and true teams. For RQ2 ,the following hypothesis is introduced:

H2: For true teams, team size and solution quality will have a curvilinear relationship with a well-defined maximum. For nominal teams, there will be a positive correlation between team size and solution quality.

The hypothesis for true teams is informed by previous work demonstrating the benefits of both small teams (Amason & Sapienza, 1997; Gooding, 2018; Karau & Williams, 1993; Lam, 2015; McComb, Goucher-Lambert, et al., 2017; Mullen et al., 1989; North et al., 2000) and large teams (Majuika & Baldwin, 1991; Steffens et al., 2012; Stewart, 2006). The hypothesis for nominal teams, in contrast, is informed by statistics. Assuming that every individual on a nominal team produces a solution with quality determined by a normal distribution, then the quality of the best solution on the team will follow an extreme value distribution (Gumbel, 1935). Therefore, the value should increase with larger team sizes. A similar hypothesis with similar underlying reasoning is posed for RQ3, which addresses the relationship of team size and solution quality with probability of winning and expected magnitude of winnings:

H3: For true teams, team size and probability of winning (and magnitude of winnings) will have a curvilinear relationship with a well-defined maximum. For nominal teams, there will be a positive correlation between team size and probability of winning (and magnitude of winnings).

The current work also relates to software engineering. A recent mapping study on software engineering teams research demonstrated that teamwork/collaboration, process/design, and coordination are prevalent research themes (DeFranco & Laplante, 2018). A research gap identified in that mapping study was team communication. With the evolution of software developing modality, it makes sense that the communication between developers and stakeholders should also evolve. Poor team communication is often cited as a root cause of failure as it affects requirement elicitation, estimations, and schedules. In fact, in global software engineering projects, when communication and coordination obviously become a larger challenge, the quality of the product is affected (Jiménez, Piattini, & Vizcaíno, 2009). In another literature review on the communication challenges in managing global software development projects, out of 30 challenges cited, communication ranked first (da Silva, Costa, Franca, & Prikladinicki, 2010).

RQ4 investigates the relationship of communication to effective team performance. Aligning with previous work, we hypothesize that:

H4: Semantic indicators will emerge in forum posts made by teams that help to differentiate high- and low-performing teams.

This hypothesis is founded on prior work that has leveraged semantic analytics to track team performance (Agogino, Song, & Hey, 2006; Zhang, Gloor, & Grippa, 2013).

3. Data

This work specifically utilizes the 2016 Meta Kaggle dataset (Kaggle Inc., 2016) which contains information on 316 different competitions run through Kaggle's platform. This data contains information for 594,892 different users and 7,949 distinct teams. The Meta Kaggle dataset includes 24 different tables of information for each of those competitions. Of those tables, only five were used for this study: users, teams, submissions, competitions, and team memberships. The Users dataset contains data related to individual Kaggle user's profiles. The Teams dataset contains data related to unique teams (of one or more users). The Submissions dataset corresponds to the data recorded when a competition entry is submitted. The Competitions dataset contains information pertinent to completed Kaggle competitions. The Team Memberships dataset is a key used to link users with their teams. All users are also permitted to post to open forums to share solutions, help other competitors, and discuss approaches. The Forum Posts data set contains the full text of every post as well as an identifier for the user who posted the comment. A full list of the variables used, and their descriptions are provided in Table I.

Table I. Dataset Variables and Descriptions

Dataset	Variable Name	Description
Users	ID	Unique identifier for the user
	Display Name	User's display name
Teams	ID	Unique identifier for the team
	Team Name	The team's name
	Competition ID	Unique identifier for the competition that the team participated in
	Score	The team's final score
	Ranking	The team's final ranking
Submissions	ID	Unique identifier for the submission
	Submitting User ID	Unique identifier for user who made submission
	Team ID	Unique identifier for the team corresponding to the user who made the submission
Forum Posts	ID	Unique identifier for the forum post
	Posting User ID	Unique ID for the user who posted to the forum
	Text	Full text of the post
Competitions	ID	Unique identifier for the competition
	Competition Name	URL slug for the competition
	Title	title for the competition
Team Memberships	ID	unique identifier for the team membership
	Team ID	Unique identifier for the team that the user belongs
	User ID	Unique identifier for the user

4. Methodology

This section introduces the analytical methodology employed to investigate the three research questions. First, the primary competitive styles observed in Kaggle competitions are introduced and defined (individual, team, and *nominal* teams simulated here). Second, the metrics used to evaluate and compare these competitive styles are introduced.

4.1 Competitive Styles

Competitors in Kaggle's competitions may compete either as individuals or band together with other individuals as part of a team. We analyze both of those conditions in this work. It should be noted that individual competitors are not completely isolated from the ideas of others. Kaggle implements an extensive forum on which players may share approaches, code, or advice. In addition, we analyze the performance of *nominal teams.* Nominal teams are groups of individuals who do not interact. Rather, they work independently towards a full solution, with the best individually-generated solution being provided as the solution of the team. Although there is no direct evidence that such teams exist in the Kaggle environment, a growing body of literature demonstrates that nominal teams may outperform interacting teams in tasks such as concept generation (Hinsz et al., 1997; Salomon & Globerson, 1989). In the current work, nominal teams are simulated via a bootstrapping approach in which several individuals from the Meta Kaggle dataset are randomly selected and the best solution found by any individual in that group is taken as the team solution. This methodology has been used in several other studies (McComb, Cagan, et al., 2017a; McComb, Cagan, & Kotovsky, 2017b; Wright, 2007). Although other work on concept generation tends to sum or average the outputs of individuals in nominal teams, the approach used here of taking the best solution depicts a reasonable approach to using nominal teams for solving tasks with well-defined and quantitative objectives. The primary comparison featured in this work is between nominal teams are compared against the teams directly available in the Meta Kaggle dataset, which will be referred to as *true teams.* Comparisons are also made against single individual competitors as a baseline.

For the analyses in this paper, only competitions with at least 10 individual competitors and at least 10 team competitors were used. Figure 1 shows the distribution of team sizes for the 127 competitions that met these criteria, and Figure 2 shows the distribution of team sizes for the winning team from each competition. These figures provide additional motivation for the research questions driving this work. Is the high rate of individual winners driven by the prevalence of individual competitors or by some differentiation in performance? This question will be further elucidated in Section 4.

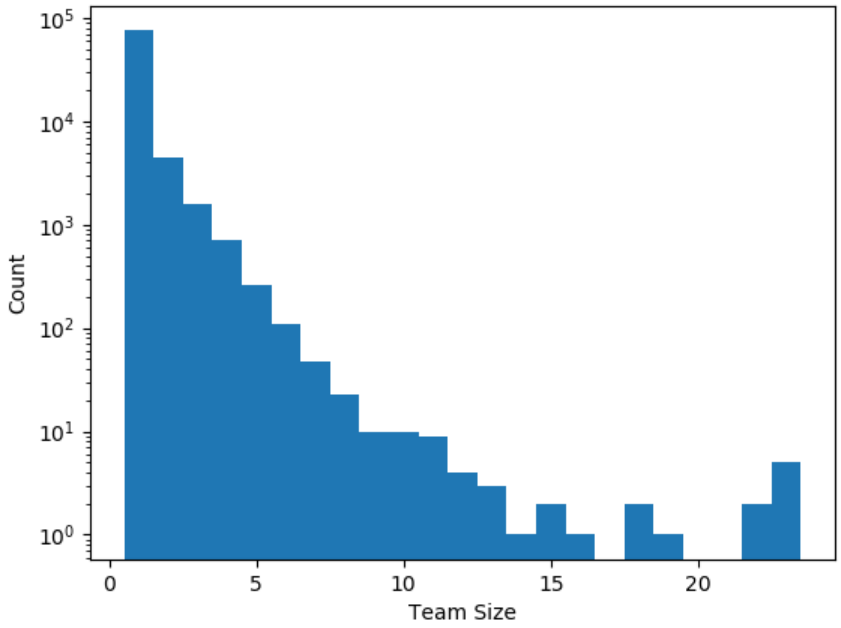

Figure 1. Team size histogram for all teams.

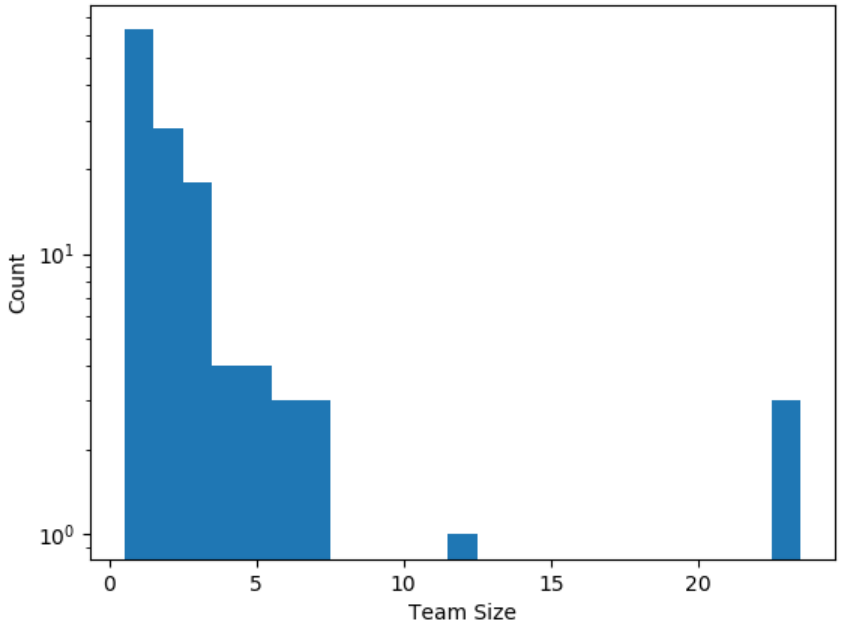

Figure 2. Team size histogram for winning Teams.

4.2 Metrics

Four metrics were utilized in this work to describe team and individual efficacy: solution quality, team effort, probability of winning, and expected payout per team member.

Solution quality is the final score of the algorithm submitted by a team or individual. This is a common metric used to measure both team and individual performance (McComb, Cagan, et al., 2017a; Yu, de Weck, & Yang, 2015). These scores were normalized for those taking part in a given competition using the equation

$$q' = \frac{q - \mu_q^{indiv}}{\sigma_q^{indiv}}, \tag{1}$$

where q' is the normalized solution quality of a team or individual solution, q is the un-normalized solution quality, μ_q^{indiv} is the average solution quality for individuals competing in that competition, and σ_q^{indiv} is the standard deviation for individuals in that competition. Because of this normalization scheme, the distribution of quality for individuals has a mean of 0 and a standard deviation of 1. It should be noted that this normalization allows solution quality to take on negative values. This approach is relatively robust to the presence of outliers in the quality data, whereas an approach that scales the values to fall in a fixed range (e.g., [0, 1]) would be highly sensitive to outliers.

Effort was quantified as the number of submissions made by a team or individual. A submission enables a team to evaluate the approximate score of their solution – it is analogous to an objective function evaluation in optimization. In defining this measure, an implicit assumption is that each submission is the result of an equal amount of cognitive effort. Although this is not precisely true (some competitors might have spent more or less effort between submissions) it should be approximately correct when averaged across many competitors. Effort was normalized for those taking part in a given competition using the equation

$$e' = \frac{e}{\mu_e^{indiv}}, \tag{2}$$

where e' is the normalized effort of a team or individual, e is the un-normalized effort, and μ_e^{indiv} is the average effort for individuals taking part in that competition. This metric will always be positive, as it represents a cardinal quantity. It should be noted that Equations 1 and 2 were both designed to establish the individual condition as a baseline for comparison.

Probability of winning is the likelihood that a team with a given structure and approach will win an average competition. Within the simulation paradigm utilized in this work, this probability is estimated simply as

$$p_{win} = \frac{w}{m}, \tag{3}$$

where p_{win} is the probability of winning, m is the number of simulations conducted, and w is the count of instances in which a team of interest produces a winning solution. Thus, this is a statistical approximation based on a sample drawn through the use of the simulations.

Expected payout per team member quantifies the average fraction of the competition reward that an individual can expect to win when averaged over many competitions (e.g., average winnings over time). This expected fraction is computed by dividing the probability of winning by the number of members on the team as

$$r = \frac{w}{mn}, \tag{4}$$

where r is the expected payout per team member, n is the number of individuals on the team and w and m are as defined previously. This is also a statistical approximation based on the sample drawn through the use of the simulations.

4.3 Linguistic Analysis

In the Kaggle competitions the discussion forums are utilized for individuals to post questions and answers to technical topics related to the competition. Example discussion topics are shown in Figure 3.

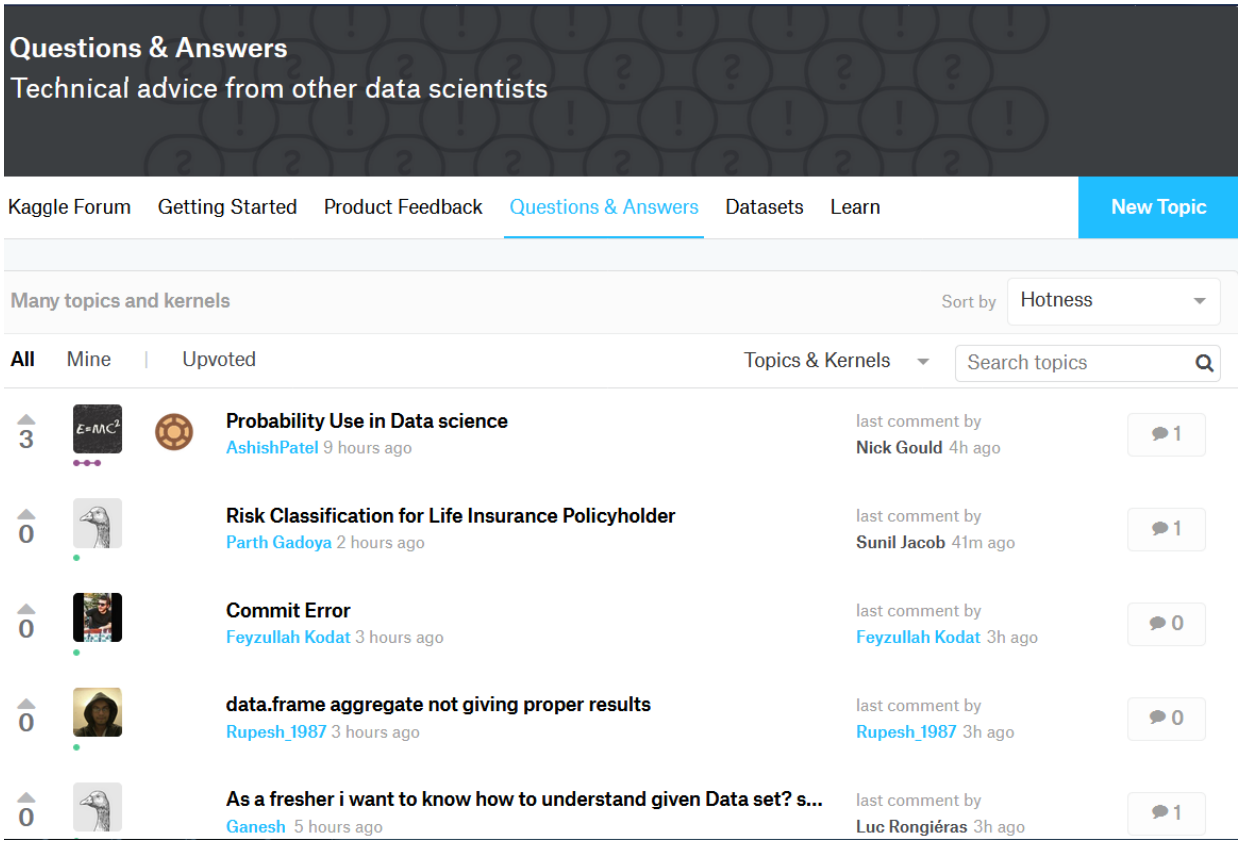

Figure 3. Kaggle forum example.

We analyzed the communication patterns in the discussion forums, comparing the forum posts of the top three ranking teams and the bottom 3 ranking teams. Only competitions for which both echelons of teams made forum posts were used, resulting in a corpus of forum posts from 69 different competitions. This data was pre-processed using a pipeline that integrated readily available tools (Boyd, 2018a, 2018b).

The resulting data was analyzed using the Linguistic Inquiry and Word Count tool, or *LIWC* to perform a linguistic analysis (Tausczik & Pennebaker, 2010). The output from LIWC are percentages of the total text that correspond to the following variables: *Linguistic process* (e.g., word count, words per sentence, dictionary words, words longer than six letters, etc.), *Linguistic categories* (e.g., 1st person, 3rd person, common verbs), *Psychological Processes* (e.g., positive and negative emotion, social, cognitive, and perceptual processes), *relativity* (e.g., time, space), *current concerns* (e.g., work, leisure, money), *spoken categories* (e.g., filler words), and *punctuation* (e.g., commas, periods,

questions). Standard statistical tests enabled comparisons between the LIWC results for top- and bottom-ranking teams.

5. Results

This section details the results of three numerical experiments and a semantic analysis conducted to better understand the role of team performance in the Kaggle competition environment. A full implementation of these analyses is available in the Python language under an MIT License.[1]

5.1 Do teams perform better than individuals?

The results of the comparison between the three competitive styles (individual, true team, and nominal team) are shown in Figure 4. The data for each competition was normalized separately and only aggregated after normalization within the competition. The data points plotted for the true teams and individuals are based directly on data from the Meta Kaggle dataset. Nominal teams were simulated by selecting several individuals and selecting the best individual solution as the team solution. By construction, the distribution of team sizes for the simulated nominal teams matches the distribution of team sizes for the true teams (see Figure 1). Note that the error bars for standard error of the individual data point are too small to appear in the figure because of the large sample size.

True teams put in more average effort (quantified by the number of submission, shown normalized on the x-axis) than individuals, and consequently are capable of producing solutions with higher quality (shown normalized on the y-axis). Nominal teams put in only slightly more effort than true teams and achieve substantially higher solution quality. Since nominal teams and true teams have the same distribution of team size, the difference in the number of submissions (our approximation of effort) is unexpected. For nominal teams, the mean normalized effort is equal to the average team size of 2.61 individuals. However, the true teams only put in 2.42 individual worth of effort on average. In other words, members of true teams put in 92.7% of the effort that would be expected if they competed individually. This difference in expended effort between true and nominal teams is potentially evidence of social loafing.

[1] https://github.com/THREDgroup/KaggleTeams

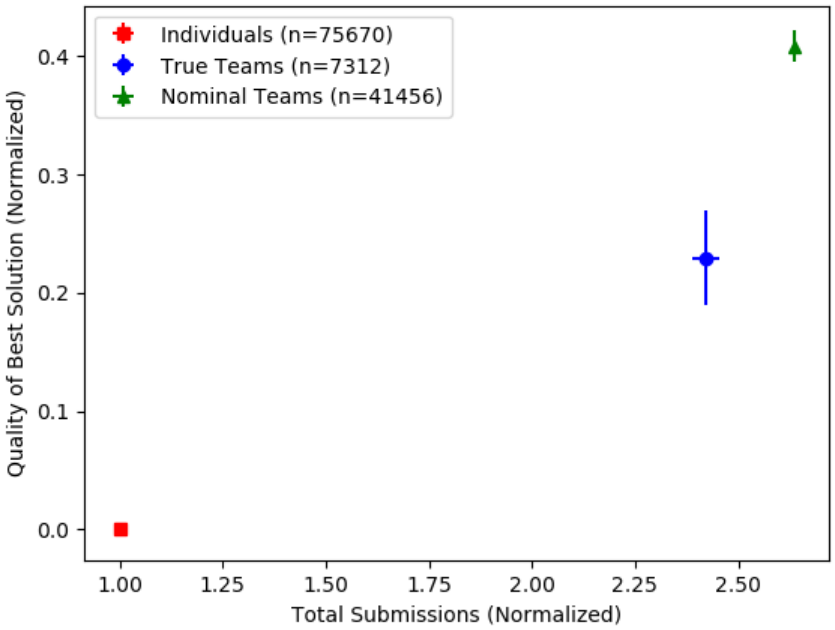

Figure 4. Comparison of individuals, interacting teams, and nominal teams. Error bars show ±1 Standard Error.

5.2 How do team size and interaction type affect overall solution quality?

The analyses in the remainder of this paper were limited to team sizes for which more than 10 instances were observed in the Kaggle dataset (see histogram in Figure 1). Thus, the maximum team size explored here is 8, as larger team sizes show substantial variance due to the low number of instances.

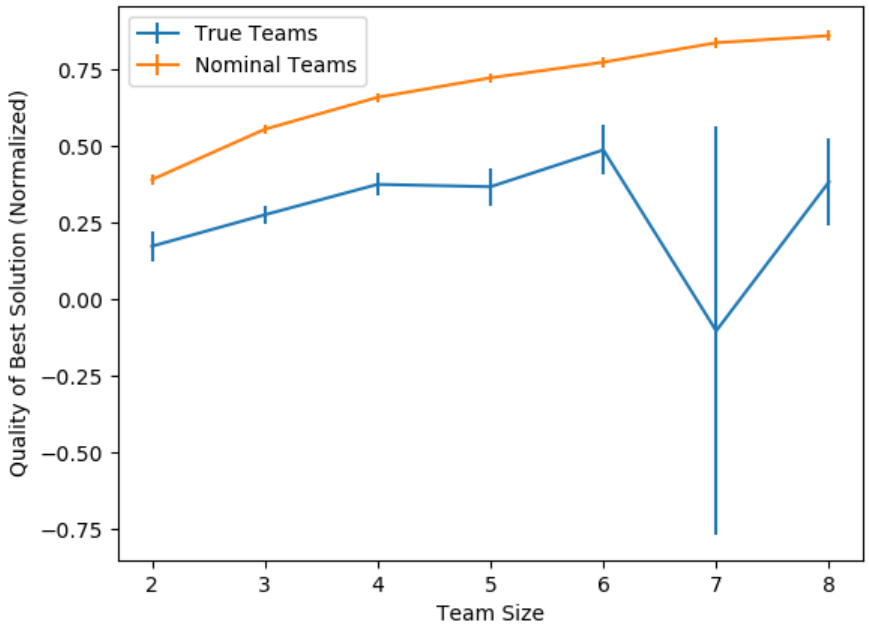

Figure 5. Solution quality versus team size for true and nominal teams. Error bars show ±1 Standard Error.

The results of the analysis of team performance as a function of team size are provided in Figure 5. Exactly 1000 nominal teams were generated for each team size shown in the figure. The number of true teams for each team size is provided in Figure 1. This analysis confirms that the effect observed in Figure 4 is robust across team sizes. A Kruskal-Wallis test was used to compare the performance of true and nominal teams for each team size (Kruskal-Wallis was chosen over ANOVA because homoscedasticity was violated). Every comparison was statistically significant ($p < 0.001$). On average, nominal teams consistently outperform true teams by approximately 0.25 standard deviations. Of particular note is the fact that 2-person nominal teams offer performance that is on par with 4-person true teams. In addition, it should be recognized that the average solution quality for 7-person true teams is substantially lower than adjacent team sizes, and the standard error is higher. This difference is driven by a small number of teams with very low solution quality.

Further, the relationship between cumulative team effort and team size is provided in Figure 6. A Kruskal-Wallis test was again used to compare the effort of true and nominal teams for each team size. Every comparison was statistically significant ($p < 0.001$). The nominal team effort follows a diagonal line with a one-to-one increase in effort for every individual added. This is expected, as nominal teams are simulated coalitions of individuals. The difference in effort displayed between the nominal and true team conditions grows at a constant rate for team sizes of up to 6, reaching a maximum of almost 50%. This aligns with other work on social loafing that indicates an higher predisposition towards the behavior in larger teams (Karau & Williams, 1993; Lam, 2015; North et al., 2000). However, for team sizes of 7-8, the difference in effort shrinks and standard error increases. This indicates that some teams were prone to social loafing while others were not. One possible explanation is that some of these larger teams employed an approach that utilized sub-teams or a clear decomposition of work into sub-tasks, effectively negating negative social effects.

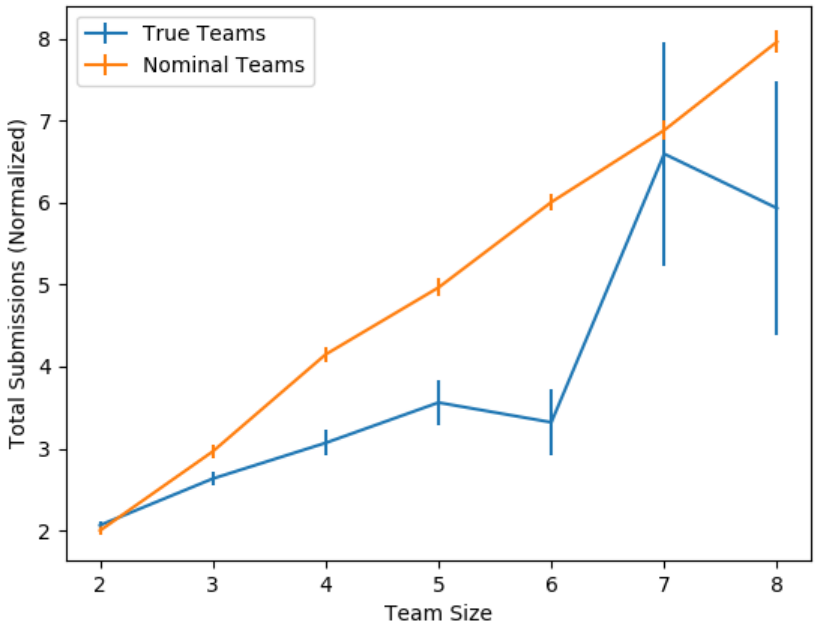

Figure 6. Total Submissions versus team size for true and nominal teams. Error bars show ±1 Standard Error.

5.3 How does the probability of winning vary as a function of team size and interaction type?

Previous sections analyzed the average performance of different competitive styles and assessed them with respect to different team sizes. However, average performance of a group is not necessarily a strong indication of how that group will perform in an authentic environment. In this section, we simulate the actual competitiveness of both nominal and true teams of varying size in an average competition. Specifically, an average competition in the Meta Kaggle dataset has 653 competitive entities (either individuals or true teams). In order to simulate competitions, we employed a statistical bootstrapping approach that leveraged the available Kaggle data. To generate one competition, we first randomly sampled 653 competitive entities from the Meta Kaggle dataset in such a way that they followed the size distribution shown in Figure 1. Ten thousand such competitions were randomly generated. For a given team size and style, ten thousand instances of the team were sampled from the Meta Kaggle dataset. Then, every combination of the set of average competitions and the set of teams was assessed by injecting the simulated team into the simulated competition. This resulted in 10 million different simulated competitions. This made it possible to estimate the relatively low probabilities of winning with high accuracy. Figure 7 shows the probability of winning an average competition (i.e., having a solution with the highest score), and Figure 8 shows the expected individual payout per team member.

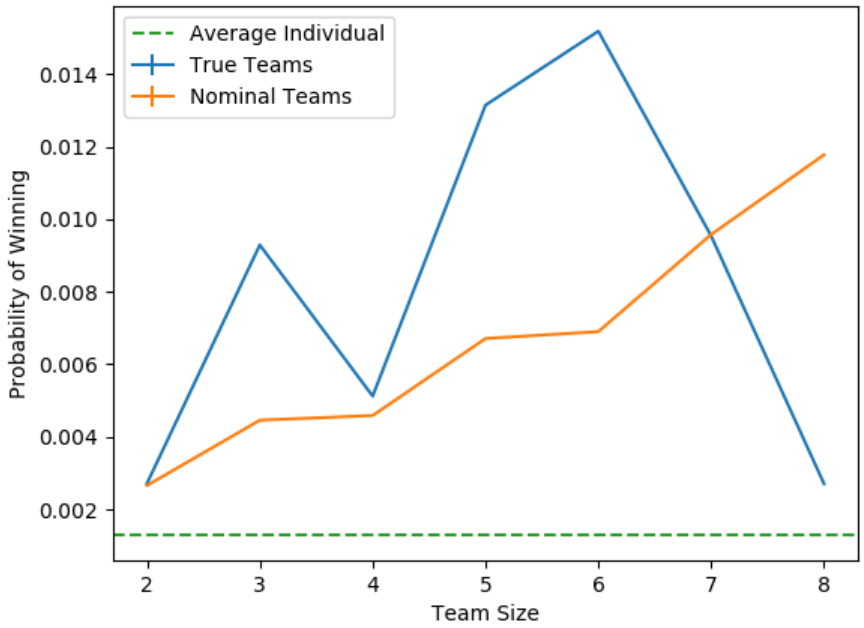

Figure 7. Probability of winning an average competition as a function of team size for both true teams and nominal teams. bars for standard error are too small to appear on this figure because of the large sample size.

In Figure 7, the probability of winning for true teams following a roughly parabolic shape, with a peak at a team is of 6. For that size, the probability of winning is approximately 1.5%. For a competition with nearly 700 competitive entities, this is significantly above what would be expected by chance. The probability of winning for the nominal teams, in contrast, increases almost linearly for the team sizes shown. The probability of winning with any team-based competitive style (either true, interacting teams or nominal teams) is substantially higher than the probability of winning as an individual.

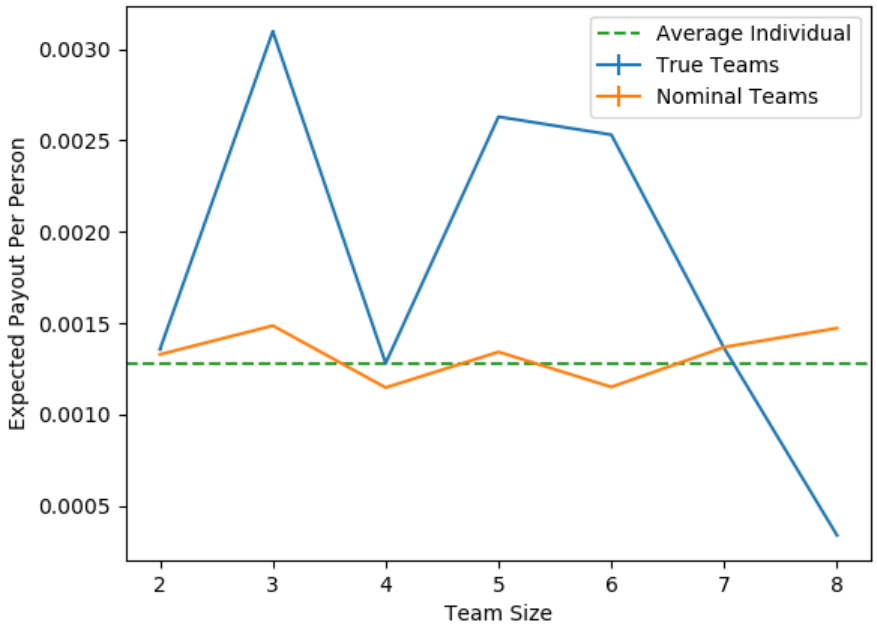

Figure 8. Expected individual payout in an average competition as a fraction of total prize amount. bars for standard error are too small to appear on this figure because of the large sample size.

Figure 8 shows the average expected individual payout to each team member as a fraction of total competition reward (probability of winning divided by team size). This represents the average payout that an individual could expect if they consistently employ a single competitive style over time. The highest observed value for true teams occurs at a size of 3, although team sizes of 5 or 6 offer comparable payouts. Interestingly, the expected individual payout for members of nominal teams is essentially equivalent to the expected payout for an individual competing alone, regardless of team size. In other words, an individual has no personal monetary incentive to join a nominal team; they can expect to reap the same return regardless of nominal team size. However, joining true teams with sizes of 3, 5, or 6 could yield significantly higher returns.

The results for both probability of winning and individual payout seem counter-intuitive, especially in comparison to Figure 5 that clearly demonstrates that nominal teams display better mean performance than true teams. However, in competitions such as these, mean performance of a group directly indicate a winning team. By their very nature, winning teams typically fall in the extreme tail of the distribution they belong to. Therefore, the parameter of interest in predicting a win probability is not strictly mean performance, but mean performance in combination with standard deviation. To illustrate this fact in greater detail, the 95th percentile of solution quality as a function of team size is provided in Figure 9. This demonstrates that the highest performing true teams deliver solutions with higher quality than the highest performing nominal teams, helping to explain the results in Figures 6 and 7. This result provides a new lens through which to view the growing literature on the superiority of nominal teams relative to interacting teams. Although nominal teams may provide better mean performance, that performance has smaller standard deviation. The higher standard deviation of interacting teams makes higher performance accessible with a higher probability, increasing the likelihood that they will succeed in competitive environments. This may help to explain the prevalence of teams in practice for both engineering and business (Paulus et al., 2011; Reiter-Palmon et al., 2008).

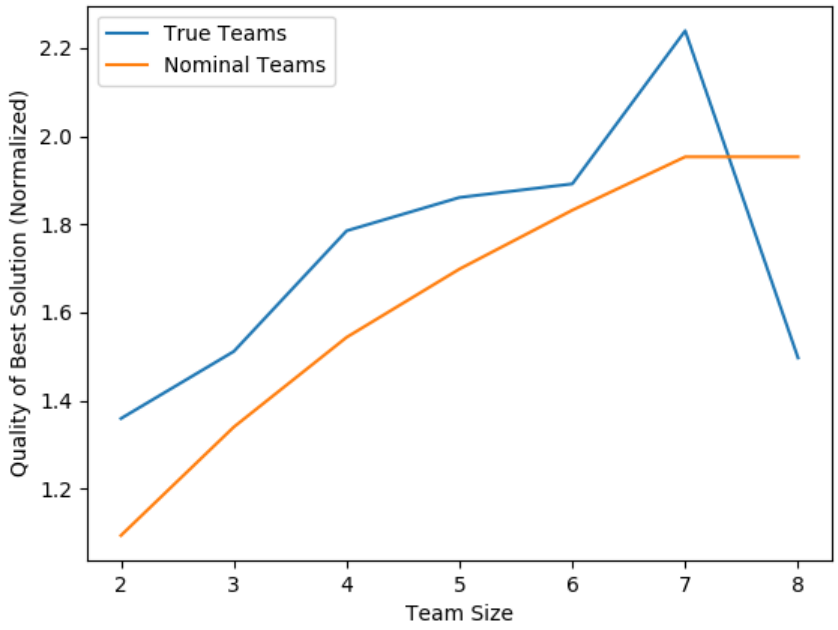

Figure 9. 95th percentile of solution quality as a function of team size for true and nominal teams.

These analyses also have deeper significance for interpreting the motivations of the participants in Kaggle's competitions. If these competitors were motivated purely by the prestige associated with winning, it is expected that they would be competing in larger teams of five or six. If they were motivated purely by monetary rewards they would likely compete in teams of three to maximize individual payout. However, the dominant competitive style in Kaggle is individual. This hints that these extrinsic motivations (prestige and compensation, respectively) may pale in comparison to intrinsic motivation offered by the Kaggle platform and community. This competition between intrinsic and extrinsic motivation in crowdsourcing has been noted by other authors (Paolacci & Chandler, 2014; Rogstadius et al., 2011; Zheng, Li, & Hou, 2014).

5.4 Are there semantic indications that align with team performance?

Across the 69 competitions analyzed, individuals on the top ranked teams collectively contributed 131.29 times on the discussion forum on average versus 20.13 times for individuals on the bottom ranked teams. A full analysis using the LIWC software was performed to further assess the differences between the semantics of these posts. Figure 10 provides a summary of the effect sizes for the LIWC comparisons. More detailed information for statistically significant comparisons is provided in Table .

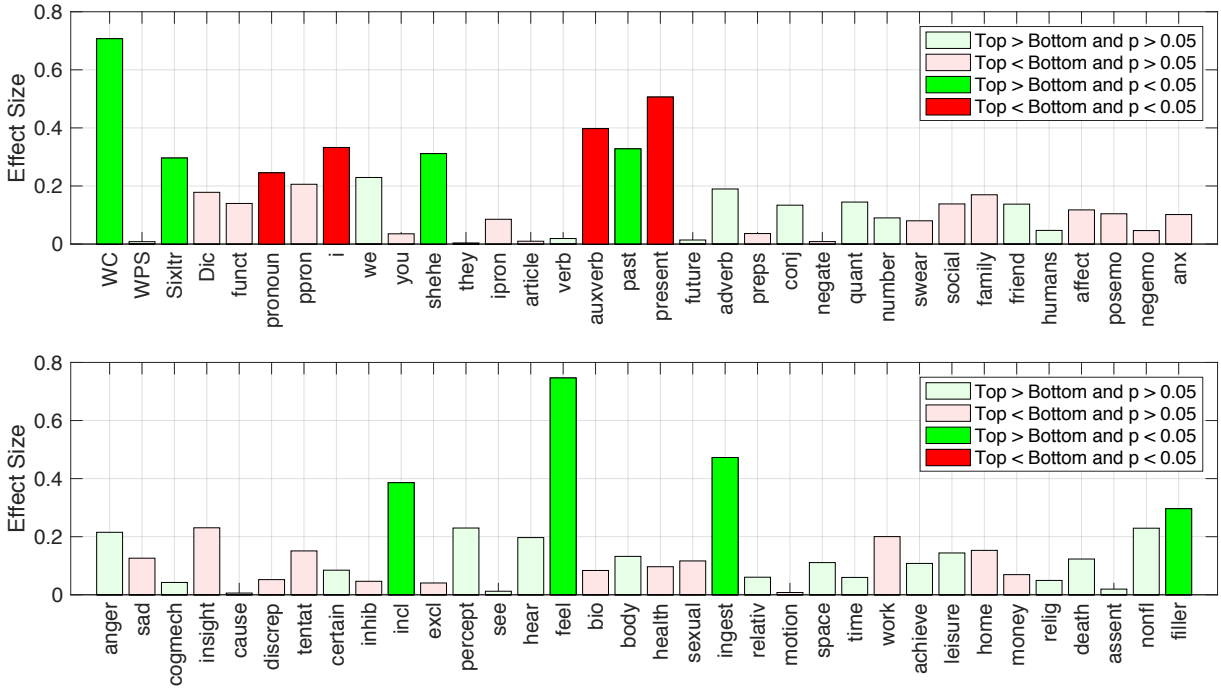

Figure 10. Summary of effect sizes for LIWC comparisons.

Table II. Significant LIWC results.

Parameter	Description	Examples	Effect size	Top mean	Bottom mean
WC	Word count	Word count	0.707	30851	2809
Sixltr	Words>6 letters	dictionary, engineering	0.296	20.00	18.64
pronoun	Total pronoun usage	I, them, itself	0.245	10.17	11.05
i	First person singular	I, me, mine	0.333	2.93	3.824
shehe	Third person singular	She, her, him	0.311	0.08	0.04
auxverb	Auxiliary verbs	Am, will, have	0.397	7.84	8.89
past	Past tense verbs	walked, were, had	0.328	2.28	1.78

present	Present tense verbs	walk, is, have	0.506	7.37	8.82
incl	Inclusive language	with, and, include	0.386	3.65	3.11
feel	Sensory feeling	touch, hold, felt ("I felt", "they felt")	0.747	0.32	0.12
ingest	Ingesting	East, swallow, taste	0.473	0.08	0.03
filler	Filler words	You know, I mean	0.296	0.16	0.09

In addition to contributing more posts to the Kaggle forum, this analysis indicates that the forum posts from the top teams also had more words (and presumably more content). Regarding verb usage, bottom team used present tense more often than the top teams. This could be an indication that these teams post about the current state of their solution or progress. Top teams, in contrast, tended to use past tense more than bottom teams, which could indicate sharing lessons learned and discussing successful past solutions. This may even be indicative of a greater concentration of experience in these top teams (manifesting here as a prevalence for drawing on past experiences).

Differences in pronoun usage in the forum posts are also significant across several variables. These include greater use of first person singular pronouns by bottom teams and greater use of third person singular pronouns by top teams. In addition, top teams tended to use first person plural (e.g., we) more often, although this difference was not significant. Taken together, these observations may indicate that bottom teams experience less cohesion (indicated by more first person singular pronouns in posts) whereas top teams were more cohesive (evidenced by first person plural and third person pronouns in posts). This hypothesis is supported by the more prevalent usage of inclusive language by top teams. In addition, the greater use of filler words (e.g., words that provide little or no information) by top teams may be indicative of greater team cohesion; *cheap talk*, the use of non-binding language that is often non-technical, has been associated with better interpersonal cooperation (Balliet, 2010; Sally, 1995).

6. Limitations

This work is subject to several implicit limitations derived from the design of the study and the nature of the database that was used for analysis. These include the lack of information on communication patterns of true teams, limitations given the definition of measures used here, and the lack of psychometric information on the Kaggle participants.

The first limitation is the lack of information on communication patterns in the true teams. The assumption made in the analysis is that true teams engaged in a non-zero amount of communication, and this assumption is based on observations of teams engaged in technical tasks who preferred to communicate with high frequencies (McComb, Cagan, et al., 2017b). However, this is not necessarily the case. For instance, it is possible that teams engaged in the Kaggle design challenges were not co-located, and thus engaged in more sparse patterns of communication than would be expected for co-located teams.

The second limitation is the definition of the effort measure used in this work. Here, effort is defined using the number of submissions (for each submission, participants received feedback on the performance of their code). The implicit assumption is that a similar amount of cognitive effort was devoted to producing each submission, thus correlating the number of submissions to total effort. It

is likely that they are correlated to some degree, but the strength of the correlation must be assessed in future work.

The third limitation is the lack of psychometric information for the participants. The current data provides rough behavioral information for the users (submission timing and quality) and some information public communications (on the forums). However, this information doesn't provide a robust way to estimate characteristics of the participants that might lead to interpersonal issues or synergistic team combinations. Specific psychometrics that are impactful in teams and may be of use for building a more nuanced model of team performance here include: personality, using a five-factor model for instance (Salleh, Mendes, Grundy, & Burch, 2009, 2010); cognitive style, a measure of an individual's preferred approach in solving problems (Buffinton, Jablokow, & Martin, 2002; Jablokow, Teerlink, Yilmaz, Daly, & Silk, 2015); and cognitive level, commonly referred to as expertise or knowledge (Ehrlich & Chang, 2006; Faraj & Sproull, 2000).

7. Conclusions

This work examined the behavior of teams engaged in crowdsourced data science competitions in order to gain insights relevant to engineering design as both researchers and practitioners increasingly embrace crowdsourcing. Specifically, the analysis investigated the Meta Kaggle dataset, which contains data from a large number of data science competitions conducted in the Kaggle crowdsourcing platform. This data was utilized to conduct three numerical experiments.

The first experiment assessed the relative performance of three different competitive styles: individual, true team (those teams observed in the Meta Kaggle dataset) and nominal teams (simulated coalitions of individuals for which the best individual solution becomes the team solution). This analysis showed that true teams perform better than individuals (though with much greater effort). However, nominal teams outperform true teams with only minimal additional effort.

The second experiment explored the robustness of the first experimental result for teams of different sizes. It was observed that nominal teams provide solutions of higher quality than true team for every team size assessed. It was also observed that the effect of social loafing increases with team size until a size of 7-8, at which point teams may naturally develop some degree of internal structure of task decomposition.

The third experiment moved beyond the assessment of average performance and injected nominal and true teams into a series of simulated competition environments. The results from these simulated competitions ran counter to those from the first two experiments: true teams were more likely to win competitions nominal teams. This is caused by the fact that true teams experience higher performance variability than nominal teams (as evidenced by Figure 5). Therefore, although their mean performance is lower, the upper tail of their distribution reaches higher. This helps to explain the prevalence of interacting teams in industry contexts.

The semantic analysis indicated several distinct differences between and top and bottom teams from the Kaggle dataset. First, it was noted that top teams tended to post in the past tense and bottom teams in the present, possibly indicative of a experience asymmetry. Second, a number of factors (pronoun usage, inclusive language, and filler words or "cheap talk") indicated that top teams were more inclusive and perhaps more cohesive. The relationship here of cohesion to high performance is at odds with results indicating that nominal teams (which, by definition, have no cohesion) may offer higher average performance.

One important feature of the data used here is the inherently quantitative evaluation of potential solutions that is afforded by the formulation of typical data science tasks. Often, tasks in engineering and design lack objectives that are easy to quantify. However, the rising prevalence of ratings-based surveys and crowdsourced evaluation approaches (e.g. (Wu, Corney, & Grant, 2015; Yumer, Chaudhuri, Hodgins, & Kara, 2015)) to quantify solution quality in engineering and design may be transferrable to software development as well.

Social loafing was a considerable factor in the true teams analyzed here. However, the nominal teams that were simulated had no mechanism for social loafing. It is possible that simply belonging to a team could lead to social loafing even in the absence of communication with teammates. Therefore, future work should evaluate the role of social loafing in nominal teams. To

this end, validated models of team behavior such as Cognitively Inspired Simulated Annealing Teams framework (McComb et al., 2015) may be useful in addition to behavioral human studies. Other research has shown that expertise plays an important role in determining crowdsourcing outcomes (Burnap et al., 2015), but the current work does not approximate the expertise of Kaggle competitors. This should be remedied in future work to offer greater insight for the motivation that competitors have to either join teams or compete individually. It may be possible to assess expertise by examining performance on past competitions.

In addition, there is significant potential to cross-reference additional open source data to provide greater insight for the analyses performed here. For instance, expertise could be measured by cross referencing the users from the Kaggle competition with a participation analysis of Stack Overflow, an online community for software developers to learn and share their programming knowledge. In addition, many Kaggle competitors host and share their software on GitHub during development. This presents several distinct possibilities, including the use of the contributions that a user has made to non-Kaggle projects to aid estimation of expertise, as well as the use of commits within the Kaggle project codebase to assess team members' contributions and potentially problem decomposition. Further, existing tools can infer gender of a GitHub user based on their username (Vasilescu, Capiluppi, & Serebrenik, 2012), which could relate results to teams' collaborative intelligence (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010).

The primary results from this work may be summarized as the following

(1) nominal teams may provide superior average performance in crowd-sourced environments;
(2) social loafing plays a large role in true teams and should be investigated experimentally in nominal teams as well;
(3) the inherent variability of true teams may enable them to outperform nominal teams in competitive environments across several measures of team performance (solution quality, probability of winning, and magnitude of expected winnings); and
(4) there are a number of semantic indicators that set top teams apart and are potentially indicative of greater team cohesion.

These results are potentially of use to a variety of practitioners in different fields. The most immediate applicability is to designers of crowdsourced data science competitions as well as managers and contributors to distributed software development projects. This work provides insights for how to maximize incentivization as well as performance. However, there are also more general implications for those who manage teams involved in work that entails problem-solving. Namely, this work provides a nuanced comparison between interacting and nominal teams regarding performance and risk. True teams with freedom to interact do not perform well on average, but they experience high variability in solution quality that occasionally results in solution with *very* high quality. In contrast, nominal teams have better average performance with very low variability, ensuring consistently mediocre results. For applications or organizations that are risk-tolerant, true teams may be the best strategy; for risk-averse cases, nominal teams should be preferred.

References

Agogino, A. M., Song, S., & Hey, J. (2006). Triangulation of Indicators of Successful Student Design Teams. *International Journal of Engineering Education*, *22*(3), 617–625.

Amason, A. C., & Sapienza, H. J. (1997). the Effects of Top Management Team Size and Interation Norms on Cognitive and Affective Conflict. *Journal of Management*, *23*(4), 495–516. https://doi.org/10.1177/014920639702300401

Ball, Z., & Lewis, K. (2018). Observing network characteristics in mass collaboration design projects. *Design Science*, *4*, e4. https://doi.org/10.1017/dsj.2017.26

Balliet, D. (2010). Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review. *Journal of Conflict Resolution*. https://doi.org/10.1177/0022002709352443

Boyd, R. (2018a). EZPZTXT. Retrieved from https://ezpztxt.ryanb.cc

Boyd, R. (2018b). TextEmend. Retrieved from https://toolbox.ryanb.cc/

Buffinton, K. W., Jablokow, K. W., & Martin, K. A. (2002). Project Team Dynamics and Cognitive Style. *Engineering Management Journal*, *14*(3), 25–33. https://doi.org/10.1080/10429247.2002.11415170

Burnap, A., Ren, Y., Gerth, R., Papazoglou, G., Gonzalez, R., & Papalambros, P. Y. (2015). When Crowdsourcing Fails: A Study of Expertise on Crowdsourced Design Evaluation. *Journal of Mechanical Design*, *137*(3), 031101. https://doi.org/10.1115/1.4029065

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Individual and Group Decision Making: Current Issues* (pp. 221–246).

Carpenter, J. (2011). May the Best Analyst Win. *Science*, *331*(6018), 698–699. https://doi.org/10.1126/science.331.6018.698

da Silva, F. Q. B., Costa, C., Franca, A. C. C., & Prikladinicki, R. (2010). Challenges and Solutions in Distributed Software Development Project Management: A Systematic Literature Review. *2010 5th IEEE International Conference on Global Software Engineering*, 87–96. https://doi.org/10.1109/ICGSE.2010.18

DeFranco, J. F., & Laplante, P. (2018). A software engineering team research mapping study. *Team Performance Management: An International Journal*, *24*(3/4), 203–248. https://doi.org/10.1108/TPM-08-2017-0040

Dyer, D. J. (1984). Team research and team training: A state-of-the-art review. In F. Muckler (Ed.), *Human Factors Review: 1984* (pp. 285–323). Santa Monica: Human Factors Society.

Ehrlich, K., & Chang, K. (2006). Leveraging expertise in global software teams: Going outside boundaries. *2006 IEEE International Conference on Global Software Engineering (ICGSE'06)*, 149–158. https://doi.org/10.1109/ICGSE.2006.261228

Faraj, S., & Sproull, L. (2000). Coordinating Expertise in Software Development Teams. *Management Science*, *46*(12), 1554–1568. https://doi.org/10.1287/mnsc.46.12.1554.12072

Gooding, R. Z. (2018). *A Meta-Analytic Review of the Relationship between Size and Performance : The Productivity and Efficiency of Organizations and Their Subunits Author (s): Richard Z . Gooding and John A . Wagner III Source : Administrative Science Quarterly , Vol . 30 , . 30*(4), 462–481.

Granovetter, M., & Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, *1*, 201–233. https://doi.org/10.2307/202051

Gumbel, E. J. (1935). Les valeurs extrêmes des distributions statistiques. *Annales de l'Institut Henri Poincaré*, *5*(2), 115–158.

Hatchuel, A., & Weil, B. (2009). C-K design theory: an advanced formulation. *Research in Engineering Design*, *19*(4), 181–192. https://doi.org/10.1007/s00163-008-0043-4

He, J., Butler, B., & King, W. (2007). Team Cognition: Development and Evolution in Software Project Teams. *Journal of Management Information Systems*, *24*(2), 261–292. https://doi.org/10.2753/MIS0742-1222240210

Hinsz, V. B., Tindale, R. S., & Vollrath, D. a. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin*, *121*(1), 43–64. https://doi.org/10.1037/0033-2909.121.1.43

Jablokow, K. W., Teerlink, W., Yilmaz, S., Daly, S., & Silk, E. (2015). The Impact of Teaming and Cognitive Style on Student Perceptions of Design Ideation Outcomes. *Proc. of the 2015 ASEE Annual Conference on Engineering Education*. Seattle, WA.

Jiménez, M., Piattini, M., & Vizcaíno, A. (2009). Challenges and Improvements in Distributed Software Development: A Systematic Review. *Advances in Software Engineering*, *2009*, 1–14. https://doi.org/10.1155/2009/710971

Kaggle. (2017). Data Science Bowl 2017 | Kaggle.

Kaggle Inc. (2016). Meta Kaggle. Retrieved from https://www.kaggle.com/kaggle/meta-kaggle

Karau, S., & Williams, K. (1993). Social Loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*(4), 681–706. https://doi.org/citeulike-article-id:1260763

Kazakci, A. O. (2015). Data science as a new frontier for design. *International Conference on Engineerring Design*, 1–10.

Lam, C. (2015). The role of communication and cohesion in reducing social loafing in group projects. *Business and Professional Communication Quarterly*, Vol. 78, pp. 454–475. https://doi.org/10.1177/2329490615596417

Majuika, R. J., & Baldwin, T. T. (1991). Team-based employee involvement programs for continuous organizational improvement: Effects of design and administration. *Personnel Psychology*, *44*, 793–812.

McComb, C., Cagan, J., & Kotovsky, K. (2015). Lifting the Veil: Drawing insights about design teams from a cognitively-inspired computational model. *Design Studies*, *40*, 119–142. https://doi.org/10.1016/j.destud.2015.06.005

McComb, C., Cagan, J., & Kotovsky, K. (2017a). Optimizing Design Teams Based on Problem Properties: Computational Team Simulations and an Applied Empirical Test. *Journal of Mechanical Design*, *139*(4), 041101. https://doi.org/10.1115/1.4035793

McComb, C., Cagan, J., & Kotovsky, K. (2017b). Validating a Tool for Predicting Problem-Specific Optimized Team Characteristics. *Volume 7: 29th International Conference on Design Theory and Methodology*, 1–10. https://doi.org/10.1115/DETC2017-67430

McComb, C., Goucher-Lambert, K., & Cagan, J. (2017). Impossible by design? Fairness, strategy, and Arrow's impossibility theorem. *Design Science*, *3*(e2). https://doi.org/10.1017/dsj.2017.1

Mello, A. S., & Ruckes, M. E. (2006). Team Composition. *The Journal of Business*, *79*(3), 1019–1039. https://doi.org/10.1086/500668

Morgan, B., Glickman, A., Woodard, E., Blaiwes, A., & Salas, E. (1986). *Measurement of Team Behaviors in a Navy Environment*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a185237.pdf

Mullen, B., Symons, C., Hu, L.-T., & Salas, E. (1989). Group Size, Leadership Behavior, and Subordinate Satisfaction. *The Journal of General Psychology*, *116*(May), 155–170. https://doi.org/10.1080/00221309.1989.9711120

North, A. C., Linley, P. A., & Hargreaves, D. J. (2000). Social Loafing in a Co-operative Classroom Task. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *20*(4), 454–475.

Orasanu, J. M., & Salas, E. (1993). Team decision making in complex environments. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsambok (Eds.), *Decision making in action: models and methods* (Vol. 327, pp. 327–345). Norwood, NJ: Ablex Publishers.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, *23*(3), 184–188. https://doi.org/10.1177/0963721414531598

Patrashkova-Volzdoska, R. R., McComb, S. A., Green, S. G., & Compton, W. D. (2003). Examining a curvilinear relationship between communication frequency and team performance in cross-functional project teams. *IEEE Transactions on Engineering Management*, *50*(3), 262–269. https://doi.org/10.1109/TEM.2003.817298

Paulus, P. B., Dzindolet, M. T., & Kohn, N. (2011). Collaborative creativity-Group creativity and team innovation. In M. D. Mumford (Ed.), *Handbook of organizational creativity* (pp. 327–357). New York: Elsevier.

Pentland, A. (2012). The new Science of Building Great Teams Analytics for Success The New Science of Building Great Teams : Analytics for Success. *Harvard Business Review*, (April). https://doi.org/citeulike-article-id:10606943

Pinto, M. (1990). Project team communication and cross-functional cooperation in new program development. *Journal of Product Innovation Management*, *7*(3), 200–212. https://doi.org/10.1016/0737-6782(90)90004-X

Reiter-Palmon, R., Herman, A. E., & Yammarino, F. J. (2008). Creativity and cognitive processes: Multi-level linkages between individual and team cognition. In M. D. Mumford, S. T. Hunter, & K. E. Bedell-Avers (Eds.), *Multi-Level Issues in Creativity and Innovation (Research in Multi Level Issues, Volume 7)* (pp. 203–267). https://doi.org/10.1016/S1475-9144(07)00009-4

Ren, Y., Bayrak, A. E., & Papalambros, P. Y. (2016). EcoRacer: Game-Based Optimal Electric Vehicle Design and Driver Control Using Human Players. *Journal of Mechanical Design*, *138*(6), 061407. https://doi.org/10.1115/1.4033426

Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., & Vukovic, M. (2011). An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.

Salleh, N., Mendes, E., Grundy, J., & Burch, G. S. J. (2009). An empirical study of the effects of personality in pair programming using the five-factor model. *2009 3rd International Symposium on Empirical Software Engineering and Measurement*, 214–225. https://doi.org/10.1109/ESEM.2009.5315997

Salleh, N., Mendes, E., Grundy, J., & Burch, G. S. J. (2010). An empirical study of the effects of conscientiousness in pair programming using the five-factor personality model. *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - ICSE '10*, *1*, 577. https://doi.org/10.1145/1806799.1806883

Sally, D. (1995). Conversation and Cooperation in Social Dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*. https://doi.org/10.1177/1043463195007001004

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research*, *13*(1), 89–99. https://doi.org/10.1016/0883-0355(89)90018-9

Steffens, P., Terjesen, S., & Davidsson, P. (2012). Birds of a feather get lost together: New venture team composition and performance. *Small Business Economics*, *39*(3), 727–743. https://doi.org/10.1007/s11187-011-9358-z

Stewart, G. L. (2006). A Meta-Analytic Review of Relationships Between Team Design Features and Team Performance. *Journal of Management*, *32*(1), 29–55. https://doi.org/10.1177/0149206305277792

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Vasilescu, B., Capiluppi, A., & Serebrenik, A. (2012). Gender, Representation and Online Participation: A Quantitative Study of StackOverflow. *2012 International Conference on Social Informatics*, 332–338. https://doi.org/10.1109/SocialInformatics.2012.81

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science*, *330*(6004), 686–688. https://doi.org/10.1126/science.1193147

Wright, D. B. (2007). Calculating nominal group statistics in collaboration studies. *Behavior Research Methods*, *39*(3), 460–470. https://doi.org/10.3758/BF03193015

Wu, H., Corney, J., & Grant, M. (2015). An evaluation methodology for crowdsourced design. *Advanced Engineering Informatics*, *29*(4), 775–786. https://doi.org/10.1016/j.aei.2015.09.005

Yetton, P. W., & Bottger, P. C. (1982). Individual versus group problem solving: An empirical test of a best-member strategy. *Organizational Behavior and Human Performance*, *29*(3), 307–321. https://doi.org/10.1016/0030-5073(82)90248-3

Yu, B. Y., de Weck, O., & Yang, M. C. (2015). Parameter Design Strategies: A Comparison Between Human Designers and the Simulated Annealing Algorithm. *Volume 7: 27th International Conference on Design Theory and Methodology*, V007T06A033. https://doi.org/10.1115/DETC2015-47674

Yumer, M. E., Chaudhuri, S., Hodgins, J. K., & Kara, L. B. (2015). Semantic shape editing using deformation handles. *ACM Transactions on Graphics*, *34*(4), 86:1-86:12. https://doi.org/10.1145/2766908

Zhang, X., Gloor, P. A., & Grippa, F. (2013). Measuring creative performance of teams through dynamic semantic social network analysis. *International Journal of Organisational Design and Engineering*, *3*(2), 165. https://doi.org/10.1504/IJODE.2013.057014

Zheng, H., Li, D., & Hou, W. (2014). Task Design, Motivation, and Participation in Crowdsourcing Contests. *International Journal of Electronic Commerce*, *15*(4), 57–88. https://doi.org/10.2753/JEC1086-4415150402