Achieving Success in Community Crowdsourcing: Lessons from the Field

Triparna de Vreede, Gert-Jan de Vreede, Naif Alawi
University of South Florida

tdevreede@usf.edu, gdevreede@usf.edu, naifalawi@usf.edu

Abstract

Community crowdsourcing is a relatively new phenomenon where local institutions, such as cities and neighborhoods, invite citizens to engage in a public discussion and solve problems that directly affect them. While community crowdsourcing has been around for over a decade, relatively little is known about what drives the success of these initiatives. In this exploratory study, we analyze field data from over 1,000 community crowdsourcing projects that were hosted on a professional community crowdsourcing platform. Our exploration reveals interesting insights into characteristics of community crowdsourcing projects that are associated with higher levels of user engagement. These insights allow us to speculate on guidelines to organize and execute community crowdsourcing initiatives.

1. Introduction

In a tale of two cities where one was a community engagement success story and the other was not, first is the story of the city of San Francisco. This city sought to involve their citizens in several community projects, ranging from discussions about parks and improving public transportation to a new design for the public library membership pass. Rather than setting up traditional face-to-face townhalls, they created online sites where citizens could share their ideas and opinions and vote on proposals. The initiative was a great success with thousands of people participating and several initiatives being implemented by the city. However, when Kansas City attempted something similar and replaced physical townhalls with online netizen discussions, it was a complete failure. After some initial enthusiasm from their netizens, these online sites lost their buzz and participation dropped to near zero levels. Why was this the case? What made one initiative so successful while the other resulted in a great disappointment? What was different about San Francisco?

The above example refers to a new problem-solving model that cities and other forms of local government have embraced to involve citizens in local issues: community crowdsourcing. Crowdsourcing is “the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the

form of an open call” [1]. There are different crowdsourcing models, such as labor markets, open innovation, citizen science, crowdfunding, and community crowdsourcing [2, 3]. Community crowdsourcing, also referred to as digital participatory platforms, is a special form of crowdsourcing in which online participants collaborate to produce a final outcome [3, 4]. For example, parents can discuss traffic safety improvements in their children’s school district or a neighborhood can discuss how to improve recreational areas like parks or playgrounds. Community crowdsourcing is typically used to solve problems organically by a community through complementing and improving upon each other’s ideas as they attempt to find a solution. It focuses on building connected and contributing communities by promoting community dialog online, ultimately leading to community action offline [5]. Common applications include city planning, neighborhood development, school district discussions, and sentiment sampling in a community. In this sense, it can be considered a type of e-government service [6].

While many cities and local organizations have turned to community crowdsourcing, setting up a successful crowdsourcing initiative is no trivial task as evidenced by the introductory example. Organizers face major challenges related to the organization and management of the crowd’s collaboration process in terms of idea generation, convergence, evaluation, and selection [7-10]. There are many considerations to be taken into account, including but not limited to the potential characteristics of the crowd members that are required for the initiative, the rewards and other approaches to motivate crowd members to make high quality contributions, the process that the crowdsourcing project has to follow to reach its goals and produce the desired deliverables, and the moderation of the crowd and the process so that there is sustained and purposeful engagement throughout the project [9]. In short, a community crowdsourcing project requires a dedicated design effort before it can be initiated followed by dedicated supervision to ensure that it continues to make acceptable progress towards a high-quality outcome [11].

Unfortunately, the field of community crowdsourcing research is still relatively young and does not offer much in the way of guidance on how to design and manage crowdsourcing efforts so that the purported benefits may be achieved [12, 13]. Given the expenses

involved in running a community crowdsourcing project, both in terms of financial costs and time investment, it is useful to get a better understanding why some community crowdsourcing projects are successful while others are not. This is the objective of the research reported in this paper: Through an in-depth analysis of field data, we aim to uncover what factors are associated with successful and engaging community crowdsourcing efforts. We summarize the insights gained from these analyses through a series of lessons learned from field experiences with community crowdsourcing.

The remainder of this paper is structured as follows. In the next section, we discuss crowdsourcing, community crowdsourcing, and engagement in more detail. Then we describe the method that we employed, followed by an overview of the results of our exploration. We conclude the paper with a discussion of our key findings, their implications, limitations to our study, and directions for future research.

2. Background

Community crowdsourcing initiatives exist on the highest levels of social media strategies according to the framework by Mergel [14, 15]. Mergel distinguishes between three strategies: At the lowest level, the *representation of an agency* level, the government entity merely has a presence on the same social media platforms as its citizen. There is little actual interaction at this level; the focus is on the government pushing out information [16]. The next level, *engagement of citizens*, involves more interaction: citizens are invited to contribute and engage in a dialogue with their government entity. The government entity can use this dialogue to inform its decision-making. The highest level, *networking with the public*, involves actual co-creation between the government entity and its citizens, where the government entity takes a more facilitative role and allows the citizens to develop solutions to actual issues. Thus, depending on the extent to which a government entity wishes to give ownership to its citizens, it operates at the *engagement* or at the *networking* level. These strategies represent the degree to which the problem owner focuses on actual engagement of citizens in community discourse.

The genesis of community crowdsourcing can be traced back to the notion of ‘cognitive surplus’. Coined by Shirky [17], cognitive surplus refers to the fact that there are billions of hours of free time that individuals worldwide have available to contribute to activities of their choice. Cognitive surplus is one of the main reasons behind the increasing popularity of other crowdsourcing models like citizen science [18]. Since community crowdsourcing typically focuses on initiatives that concern someone’s own neighborhood or

city, there can be significant motivation to use some ‘cognitive surplus’ to contribute to the online discourse [19].

However, initial motivation to contribute to a community crowdsourcing initiative is insufficient for the ultimate success of such initiatives [5]. If citizens make initial contributions, but never return to further participate, the initiative is likely to die out. Therefore, a critical success criterion for community crowdsourcing is *engagement*. An engaging experience is one that draws a participant in and holds their attention [20]. Without engagement, there is no community involvement and the crowdsourcing initiative is doomed. Since individuals with ‘cognitive surplus’ have different options to spend their free time, a successful community crowdsourcing project will entice them to keep contributing or, at least, monitoring and following the ideas and opinions that others contribute.

In the context of our study, we follow de Vreede and colleagues [20], who define engagements a three-part phenomenon, which manifests in one or more of the following forms:

1. *Affective/Emotional engagement: the extent to which individuals experience a positive psychological reaction or attachment towards a specific activity or situation.*
2. *Behavioral engagement: the extent to which the individuals can be observed to exert effort and show persistence to remain involved in an activity or situation.*
3. *Cognitive engagement: the extent to which individuals are cognitively absorbed in a task or activity resulting in a reduced awareness of their surroundings.*

For the analyses reported in this paper, we only focus on the behavioral component of engagement. As crowdsourcing is Internet-based, participants’ online activities (i.e. evidence of behavioral engagement) are fully recorded and can be analyzed straightforwardly. Since we are using historical field data, it is impossible to measure affective and cognitive engagement. Furthermore, even if participants were to report being highly engaged emotionally and cognitively, but do not convert this psychological engagement into actions, it will be of less interest to the community crowdsourcing organizers.

3. Method

3.1. Case organization

For this study, we partnered with a community crowdsourcing provider, headquartered in the Mid-West of the United States. Founded in 2011, the Company

originated from the domain of urban planning. An important part of their services involved organizing townhalls for interested citizens for urban planning projects. When they noticed that these townhalls were typically poorly attended, they developed an online environment that their clients could use to engage with citizens in their geographical target area. To date, the company has run hundreds of community crowdsourcing projects across North America for different types of clients, including but not limited to cities, counties, schools, school districts, neighborhoods, and municipal organizations (e.g. libraries, transportation).

The Company made part of their production database available for analyses. This database includes a record of all activities that citizens perform for each of their projects up to 2017. Typically, the database is organized into distinct projects representing different clients. Examples of such projects are the ones initiated by City of San Francisco and Kansas City respectively. Each of these projects consisted of a number of topics where citizens shared ideas and comments. Also, a project could involve voting activities to sample which option a group of citizens would prefer. Finally, for many projects, citizens could share links to the topic discussions on their own social media.

3.2. Data preparation

Real-world data typically is fraught with inconsistencies, incompleteness, and noise among other discrepancies that require preparation to derive useful insight from data mining [21]. The Company's dataset we used in this study also contained some of these common issues as other real-world datasets. The Company's dataset consists of 2.8 million activities performed by 741,136 users on their platform from May 2011 to October 2017. When retrieving the overall architecture of the dataset, we found that the dataset schema contains 177 entities that store data for a hierarchical structure of 1208 projects, 23,612 topics, and 207,987 ideas.

Projects in the dataset refers to a set of topics that belong to a problem owner in which they seek ideas and feedback pertaining a topic or several topics from their community. Topics are subsets of projects and they refer to a specific theme that has single or multiple interconnected and related issues that require ideas and feedback from the community. Ideas are subsets of topics and could be in the form of a suggestion, opinion, or thought that is related to the topic of interest. Users typically participate in the platform by either providing ideas or commenting on other ideas proposed by other users.

The current architecture of the dataset along with issues pertaining inconsistencies, incompleteness and noise, would make it difficult to derive useful lessons and insights from its initial state. Accordingly, we proceeded with a systematic approach in preparing the data and making it suitable for data mining.

Following the data preparation approaches suggested by [21, 22], we designed a target data structure (Figure 1) required for analysis to derive important lessons from the data mining process of the dataset. Accordingly, we performed a three-phase data preparation process consisting of data cleaning, data transformation, and data integration.

Figure 1. Target dataset structure.

Data cleaning. Based on our target data structure, we defined a set of rules on how to approach the data cleaning process. The rules that we set include:

1. The data should be clean of demonstration and testing instances.
2. The data should be clean of nonsensical contributions such as “asdfjkl;”.
3. The data should be related to the ideation and engagement interaction between users.

Following these rules, we applied a set of cleaning filters at four different levels based on the hierarchical structure of the dataset: projects, topics, ideas, and users' activities. Starting with the project level, we filtered projects that were created for the purpose of testing and demonstrations. This process led to reducing the number of projects to 895 projects.

Next, we filtered the topics in the dataset through two stages. The first stage of filtering topics was related to the type of topics available in the dataset. Our inspection of the dataset revealed that there are five different types of topics which are: Standard, Instant Poll, Participatory Budget Allocation, Participatory Budget Fund, Survey, and Photo share. Only one of those five types – Standard – was relevant to ideation and community engagement as defined in our rules as the other four topic types did not have interactive capabilities in terms of sharing ideas and thoughts via the platform. The second stage of filtering topics

focused on removing testing and demo topics that were created within real projects. As a result of those two filters, we ended up having 9,440 topics. Similar to filtering topics, we filtered ideas in two stages. The first stage consisted of removing testing and demonstration ideas while the second stage consisted of removing nonsensical contributions such as “asdfjkl;”. As a result, the final tally of ideas included in our study is 150,960 ideas.

The dataset further consisted of 51 different types of activities. Upon further inspection of these activity types, we found that most were either system administrator exclusives or nondeployed features. Accordingly, we found only three activities that were performed in relation to the ideation and community engagement interaction between users in the platform: create idea, vote idea, and create comment. Based on our findings, we filtered the dataset to include only the activities that are one of the three identified user activities. As a result, the number of activities included for data mining after filtering is 915,468.

Data transformation. After cleaning the data, we performed data transformation processes to extract important semantic features from the dataset. To do so, we used the 2015 version of the Linguistic Inquiry and Word Count tool (LIWC 2015) [23] to extract the semantic features from users’ textual contributions. LIWC is a validated tool that extracts several semantic, psychological, and sociological features from text. It has been used in different research domains including psychology and digital humanity [24]. By exporting textual content from the dataset pertaining to topics and idea descriptions as well as users’ feedback to .txt files, we analyzed these files using LIWC and generated a semantic profile for all extracted content.

Data integration. Following data cleaning and transformation, we finalized the process of data preparation by integrating the clean data and the semantic results extracted using LIWC. This process resulted in creating a dataset with 9376 observations. To account for outliers in our analysis, we use Mahalanobis distance to detect and eliminate outliers [25]. The resulting final dataset consisted of 8814 observations.

3.3. Measures

To measure the level of engagement that topics elicited from participants, we focused mainly on participant engagement per topic as the main construct of interest in this study. Drawing from [5], we measured participant engagement using the Participant Engagement Index (PEI), an operational measure that accounts for engagement magnitude. Since the interest of our research was at the topic level, we utilized PEI with the objective of analyzing topics engagement

performance in the dataset. Accordingly, we first calculated the *PEI* score for each participant pertaining to each topic they participated in. The *PEI* score for each participant was calculated based on the three types of activities related to topics: Idea creation (*I*), Feedback (comment) (*F*), and Vote (*V*). It is to be noted that these activities were different in being reactive (*F*, *V*) or proactive (*I*) as well as the level of efforts required to perform these different activities. Given that these activities represented different levels of engagement, the activities were assigned different weights when calculating *PEI* as follows: Idea creation ($W_I = 5$), Feedback ($W_F = 3$), Vote ($W_V = 1$). These values were based on the perceived relative effort of each activity. Accordingly, the weighted *PEI* score for participants was measured based on the formula where *U* is a participant and *X* is the range of all users:

$$PEI_{U_X} = W_I \sum I_{U_X} + W_F \sum F_{U_X} + W_V \sum V_{U_X}$$

We computed the Topic *PIE* by averaging the sum of *PEI* ($\sum PEI_U$) of all participants for each topic based on the number of weeks of each topic as in the formula:

$$Topic\ PEI = \frac{\sum PEI_{U_X}}{Topic\ Number\ of\ Weeks}$$

Based on *Topic PEI*, we arranged the topics in a descending order from highest engagement topics to lowest. Then, we categorized the topics into three separate groups: high engagement topics (top 33 percentile), medium engagement topics (middle 34 percentile), and low engagement topics (bottom 33 percentile). The computed users’ individual *PEI* and *Topic Weekly PEI* as well as the three topics groups were used to discover insights in the results sections.

3.4. Analysis approach

Given the exploratory nature of this study, we adopted an agile data analytics approach to derive important lessons from the dataset. The agile data analytics approach was drawn from Larson and Chang [26] who proposed a Fast Analytics/Data Science framework consisting of seven phases which are: scope, data acquisition, analyze, model, validate, and deploy. Hence, we went through a cyclical process through discussion between the Company and the researchers to define the scope for each lesson, acquire and prepare the data, analyze the data, develop and test analytical proposed models, statistically validate the results, and synthesized the findings accordingly. Based on a

discussion of findings, additional or new analyses were defined in close consultation with the Company.

4. Results

4.1. Lesson 1: Avoid positive wording in topic descriptions

Impact of sentiment has been commonly analyzed in prior research to understand how sentiment affects content virality [27], firm equity value [28], and political outcomes [29]. With regard to content virality, Berger and Milkman [27] found that both positive and negative sentiment have a positive association with content virality; with positive sentiment specifically having a highly likelihood of making content becoming viral. Drawing from the concept of content virality and sentiment, we explore whether topics with high average PEI scores differ from topics with lower average PEI scores in terms of using positive or negative sentiment in the topic description. Accordingly, we propose that:

H1: Topics with high average PEI scores differ from topics with low average PEI scores in terms of positive sentiment wording in the topic description.

H2: Topics with high average PEI scores differ from topics with low average PEI scores in terms of negative sentiment wording in the topic description.

To test these hypotheses, we used multivariate analysis of variance (MANOVA) to compare the topics with high average PEI scores to the topics with lower average PEI scores since it is the most suitable test given the characteristics of the variables. Testing for assumptions, we found that our data does not meet the normality distribution and homogeneity of variance assumptions (Table 1). However, MANOVA was found to be robust for violations of these assumptions when having equal, or somewhat equal, large sample sizes for groups [30, 31].

Table 1: Normality distribution and homogeneity of variance tests for positive and negative sentiments

	Anderson - Darling's test	Levene's Test
Positive sentiment	$p = 0.000***$	$p = 0.000***$
Negative sentiment	$p = 0.000***$	$p = 0.000***$

*** $P \leq 0.001$

The MANOVA test results and the post-hoc Tukey's test show support for H1 but not H2 (Table 2).

Interestingly, the tests indicate that high average PEI scores have a lower use of positive sentiment wording in their description compared to the mid-level and low-level average PEI score topics.

Table 2: MANOVA and post-hoc Tukey's test results for positive and negative sentiment in topics' description

	DF	F	p value
Positive sentiment	2	7.720	0.000***
Negative sentiment	2	0.757	0.469

Positive Sentiment Comparison

	Diff	p adj	
Mid - Low	0.144	0.368	
High - Low	- 0.271	0.032*	Support H1
High - Mid	- 0.416	0.000***	Support H1

* $P \leq 0.05$. *** $P \leq 0.001$

4.2. Lesson 2: Focus on inclusive language and avoid exclusive language

In addition to sentiment, we examined the differences of using inclusive and exclusive language between the topics with high average PEI scores and the topics with lower and average PEI scores. In LIWC, the inclusive words category includes words such as “and”, “with”, “include” while the exclusive words category consists of words such as “but”, “without”, and “exclude”. In policy making, research suggests that the use of inclusive language enhances public engagement [32]. Also, Mahmud et al. [33] found that people who use inclusive language are more likely to be more engaged in social media by retweeting. We thus explore whether:

H3: Topics with more inclusive words will have higher average PEI scores than topics with fewer inclusive words.

H4: Topics with fewer exclusive words will have higher average PEI scores than topics with more exclusive words.

As with the previous lesson, we utilized MANOVA and Tukey's post-hoc comparison to test the hypotheses. In terms of measures, we used LIWC inclusive and exclusive scores for topic descriptions and compare between topics with high engagement and lower engagement in terms of using inclusive and exclusive language. The results of the test support H3 partially as we found significance only between high PEI topics and low PEI topics where high PEI topics have more inclusive language compared to low PEI topics. Additionally, the test results support H4 as we found

that high PEI topics use less exclusive language compared to mid and low PEI topics (Table 3).

Table 3: MANOVA and Post-hoc Tukey's test results for inclusive and exclusive language in topics' description

	DF	F	p value
Inclusive	2	4.052	0.017*
Exclusive	2	14.215	0.000***

Inclusive Language Comparison			
	Diff	p adj	
Mid - Low	0.168	0.171	
High - Low	0.265	0.013*	Support H3
High - Mid	- 0.416	0.552	

Exclusive Language Comparison			
	Diff	p adj	
Mid - Low	- 0.096	0.205	
High - Low	- 0.298	0.000***	Support H4
High - Mid	- 0.202	0.001**	Support H4

* $P \leq 0.05$. ** $P \leq 0.01$. *** $P \leq 0.001$

4.3. Lesson 3. Listen and acknowledge

In community crowdsourcing, projects where problem owners provide feedback to users are expected to have a higher level of user engagement compared to projects where there is little to no feedback [5]. In order to explore the impact of feedback on user engagement from the dataset, we investigated the relation between users' engagement and the level of feedback engagement they received for their ideas. To this end, we compared individual users' PEI. As demonstrated earlier, PEI can be calculated for a single topic. Similarly, PEI can also be calculated for individual participants by performing a weighted addition of all activities that a specific user executed. We postulated that:

H5: There is a positive correlation between a user's PEI score and the PEI score of the feedback they receive.

To test this hypothesis, we used the computed users' PEI scores for users who contributed ideas. In addition, we aggregated the PEI scores of the reactions (feedback and votes) that users received for their ideas. Then, we utilized non-parametric ranking measures such as Kendall τ and Spearman ρ to test the hypothesis given the non-normality distribution characteristics of the dataset [34]. The tests results show strong support for H5 (Table 4).

Table 4: Correlation test between user's PEI and the PEI of the reaction they receive

	Kendall τ	Spearman ρ	p value
User PEI & Feedback PEI	0.463	0.569	0.000***

*** $P \leq 0.001$

4.4. Lesson 4: Know your crowd

Community crowdsourcing platforms such as this Company have a large user base that represents different populations and sub-populations. Understanding the characteristics of the user base helps in identifying important observations and building analytical models [35]. Such observations and models can help guide crowdsourcing platforms and problem owners to enhance the engagement of the userbase. In this analysis, we investigated the characteristics of the user base based on their demographic and geographic information.

A summary analysis of the demographic information indicated that participants belong to different age groups between 21 years old and over 65 years old. While there is a fairly even spread across age groups, the majority of the users (82.8%) are between 26 and 65 years old (Figure 2). In terms of gender, the userbase shows a higher proportion (55% vs. 45%) of female users (Figure 3).

Figure 2: User age distribution.

Figure 3: User gender distribution.

Beyond the demographic information, the dataset has also revealed a wide distribution of users geographically. The heat map generated from the data

set using users zip codes from the United States revealed that the user base expands across most states (Figure 4).

Figure 4: Distribution of userbase over the US.

4.5. Lesson 5: Engage power users

Identifying power users by their level of engagement is an integral component in community crowdsourcing platforms. Power users are important as they can greatly influence engagement by diffusing positions and perspectives in crowdsourcing platforms [36]. To investigate if power user engagement was indeed important, we plotted user's PEI as a percentage of the overall engagement score for the dataset (Figure 5). The results show that 29.5% of users are responsible for 80% of the activities in the platform.

Figure 5: Comparing the number of users relative to engagement.

Lesson 6: Articulate your expectations

In community crowdsourcing platforms, topic descriptions provide key information about the problem in terms of background and context that users will base

their ideas on. Accordingly, understanding the impact of the topic's description articulation is key for understanding underlying factors that enhance community engagement with topics. According to Fan & Yan [37], poor wording is a key contributing factor for reduction in the number of participants in online polls. Drawing from that ideas, we posit that:

H6: There is a positive relationship between topics engagement score and topics description articulation.

We tested this hypothesis using the topic's PEI score as a measure for topic engagement and the word count of the topic description as a measure for topic articulation. Given the non-normality characteristics of our dataset, we used Theil-Sen nonparametric linear regression for its robustness to test our hypothesis [38]. The test results show significant support for H6 providing evidence that there is a positive relationship between topic engagement and topic articulation (Table 5).

Table 5: Nonparametric regression of the relation between topic description word count and topic PEI score

	Estimate	MAD	<i>p</i> value
Description Word count	0.022	0.52	0.032*

* $P \leq 0.05$

5. Discussion & Conclusions

The exploration of community crowdsourcing field data results in a number of interesting and, at times, surprising insights. Table 6 summarizes the lessons we learned from our exploration.

Table 6: Lessons learned

Lesson 1: Avoid positive wording in topic descriptions
Lesson 2: Use inclusive language and avoid exclusive language
Lesson 3: Listen and acknowledge
Lesson 4: Know your crowd
Lesson 5: Engage power users
Lesson 6: Articulate your expectations

The first interesting observation is that positive wording in topic description is associated with lower engagement. This is contrary to our expectations. A possible explanation may be that a positive articulation of a topic may only appeal to a subset of potential users and 'turn off' some others. It will be worthwhile to explore whether more neutral wordings are associated with higher levels of engagement as such topics may appeal to any user, regardless of what their initial opinion of a topic is.

The next two insights were less surprising. First, our analyses suggest that it may be useful to focus on inclusive language in topic description and avoid exclusive language. Inclusive language can be seen as more welcoming to users, drawing them in and stimulating them to contribute. Second, our analyses clearly show that it is important for sustained engagement that users experience a sense that someone is listening to them. This is most directly evident from receiving feedback on their contributions. Going back to the opening example, a deeper investigation of the Kansas City initiative showed that many users repeatedly asked the city for feedback on their ideas and received no response. This may not only have disappointed the users, but it may also have discouraged new users from making additional contributions.

Achieving a better understanding of the nature of contributors to a community crowdsourcing effort was the focus of lessons 4 and 5. It was interesting to see that age was not correlated with engagement. We witnessed fairly evenly spread engagement for all age groups from 30 years and up. This may be explained by the fact that around that age, people get more settled into their neighborhood and thus may be more motivated to contribute to public discourse about their environment. Additionally, it is interesting to note that there is a relatively high proportion of users (almost 30%) that is responsible for 80% of the online activity. This means there is potentially a broad user base for community crowdsourcing discussions which will make it easier for organizers to facilitate ongoing interaction.

Our last lesson suggests additional guidance in terms of articulating a topic. Results show that longer topic descriptions are associated with higher levels of engagement. Articulation appears to be an important driver for users to better understand the issue at hand and contribute. Combining this lesson with lessons two (inclusive vs. exclusive language) and three (feedback), appears to offer actionable guidance for community crowdsourcing organizers.

Our study makes several contributions to research and practice. Community crowdsourcing is still an understudied phenomenon. Most publications focus on the conceptualization of the phenomenon (e.g. [39, 40]) or describe small-scale case studies (e.g. [16]). To the best of our knowledge, our study is the first large scale, empirical field analysis of community crowdsourcing that has been undertaken with the goal of extracting practical guidance to increase community user engagement. Using real data from hundreds of projects allowed us to derive rich insights about factors that are associated with different levels of engagement. Further, our findings offer initial suggestions for managers of community crowdsourcing projects on how to set up and manage the projects. Our findings may also inspire a set

of evaluation criteria to assess the health and success of a community crowdsourcing project in terms of sustained engagement.

A number of limitations have to be considered in the context of this study. First, we were only able to analyze field data from a single community crowdsourcing provider. As there are dozens of other providers (see e.g. [3]), additional research is required to determine whether our findings hold across different platforms. Second, we were only able to determine certain associations between topic and user characteristics and their engagement levels. It was impossible to uncover underlying causal mechanisms from the historical data that give rise to the effects that we observed.

We foresee the following avenues for future research. First, a number of the lessons learned can be examined in more detail in a laboratory setting. This will not only enable us to replicate field observations in the lab, but also allow us to explore the underlying mechanisms the cause the observed effects. Second, it would be interesting to develop an ongoing assessment instrument to monitor engagement levels in ‘live’ community crowdsourcing projects to show the PEI score for regular intervals. This could illustrate how engagement is developing with respect to a benchmark engagement level in a project (see figure 6 for a conceptualization). It could also allow organizers to see the effects of certain interventions on the users’ engagement with a topic. Finally, an ongoing monitor could be combined with short online contributor surveys to measure affective and cognitive engagement.

Figure 6. Conceptualization of a Project PEI monitor showing the health of a project in terms of engagement compared to a benchmark.

References

- [1] J. Howe. (2006) The rise of crowdsourcing. *Wired magazine*. 1-4.

- [2] O. Oh, C. Nguyen, G.-J. de Vreede, and D. C. Derrick, "Collaboration science in the age of social media: A crowdsourcing view," in *Proceedings of Group Decision and Negotiation*, Recife, Brazil, 2012, vol. 2, p. 69.
- [3] E. Falco and R. Kleinhans, "Digital participatory platforms for co-production in urban development: A systematic review," in *Crowdsourcing: Concepts, Methodologies, Tools, and Applications*: IGI Global, 2019, pp. 663-690.
- [4] K. Boudreau and K. Lakhani, "How to manage outside innovation," *MIT Sloan management review*, vol. 50, no. 4, p. 69, 2009.
- [5] C. Nguyen, N. Tahmasbi, T. De Vreede, G.-J. De Vreede, O. Oh, and R. Reiter-Palmon, "Participant engagement in community crowdsourcing," in *European Conference on Information Systems*, Münster, Germany, 2015.
- [6] M. A. A. Alryalat, N. P. Rana, G. P. Sahu, Y. K. Dwivedi, and M. Tajvidi, "Use of social media in citizen-centric electronic government services: A literature analysis," *International Journal of Electronic Government Research*, vol. 13, no. 3, pp. 55-79, 2017.
- [7] H. Shimazu and S. Koike, "KM2. 0: Business knowledge sharing in the Web 2.0 age," *NEC technical journal*, vol. 2, no. 2, pp. 50-54, 2007.
- [8] I. Seeber, G.-J. De Vreede, R. Maier, and B. Weber, "Beyond brainstorming: Exploring convergence in teams," *Journal of Management Information Systems*, vol. 34, no. 4, pp. 939-969, 2017.
- [9] N. H. Thuan, P. Antunes, and D. Johnstone, "Factors influencing the decision to crowdsource: A systematic literature review," *Information Systems Frontiers*, vol. 18, no. 1, pp. 47-68, 2016.
- [10] T. De Vreede, C. Nguyen, G.-J. De Vreede, I. Boughzala, O. Oh, and R. Reiter-Palmon, "A theoretical model of user engagement in crowdsourcing," in *International conference on collaboration and technology*, Wellington, New Zealand, 2013, pp. 94-109: Springer.
- [11] T. de Vreede, G.-J. de Vreede, and R. Reiter-Palmon, "Antecedents of engagement in community-based crowdsourcing," in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.
- [12] J. I. Criado, R. Sandoval-Almazan, and J. R. Gil-Garcia, "Government innovation through social media," *Government Information Quarterly*, vol. 30, no. 4, pp. 319-326, 2013.
- [13] R. Kleinhans, M. Van Ham, and J. Evans-Cowley, "Using social media and mobile technologies to foster engagement and self-organization in participatory urban planning and neighbourhood governance," *Planning Practice & Research*, vol. 30, no. 3, pp. 237-247, 2015.
- [14] I. Mergel, "A framework for interpreting social media interactions in the public sector," *Government information quarterly*, vol. 30, no. 4, pp. 327-334, 2013.
- [15] I. Mergel, "Social media adoption and resulting tactics in the US federal government," *Government information quarterly*, vol. 30, no. 2, pp. 123-130, 2013.
- [16] D. Skog, "Push or pull?: Using social media to interact with citizens," in *ICT, Society, and Human Beings*, Lisbon, Portugal, 2017, pp. 231-234: IADIS Press.
- [17] C. Shirky, *Cognitive surplus: Creativity and generosity in a connected age*. Penguin UK, 2010.
- [18] J. Silvertown, "A new dawn for citizen science," *Trends in ecology evolution*, vol. 24, no. 9, pp. 467-471, 2009.
- [19] K. Mossberger, Y. Wu, and J. Crawford, "Connecting citizens and local governments? Social media and interactivity in major US cities," *Government information quarterly*, vol. 30, no. 4, pp. 351-358, 2013.
- [20] T. de Vreede, S. Andel, G.-J. de Vreede, P. Spector, V. Singh, and B. Padmanabhan, "What is Engagement and How Do We Measure It? Toward a Domain Independent Definition and Scale," in *Proceedings of the 52nd Hawaiian International Conference on System Science*, Los Alamitos, 2019: IEEE Computer Society Press.
- [21] S. Zhang, C. Zhang, and Q. J. A. a. i. Yang, "Data preparation for data mining," vol. 17, no. 5-6, pp. 375-381, 2003.
- [22] S. García, J. Luengo, and F. Herrera, *Data preprocessing in data mining*. Cham, Switzerland: Springer, 2015.
- [23] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," 2015.
- [24] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of language social psychology*, vol. 29, no. 1, pp. 24-54, 2010.
- [25] R. De Maesschalck, D. Jouan-Rimbaud, D. L. J. C. Massart, and i. l. systems, "The mahalanobis distance," vol. 50, no. 1, pp. 1-18, 2000.
- [26] D. Larson and V. Chang, "A review and future direction of agile, business intelligence, analytics and data science," *International Journal of Information Management*, vol. 36, no. 5, pp. 700-710, 2016.
- [27] J. Berger and K. L. Milkman, "What makes online content viral?," *Journal of marketing research*, vol. 49, no. 2, pp. 192-205, 2012.
- [28] Y. Yu, W. Duan, and Q. J. D. s. s. Cao, "The impact of social and conventional media on firm equity value: A sentiment analysis approach," vol. 55, no. 4, pp. 919-926, 2013.
- [29] W. Budiharto and M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," *Journal of Big data*, vol. 5, no. 1, p. 51, 2018.
- [30] R. G. O'Brien and M. K. J. P. b. Kaiser, "MANOVA method for analyzing repeated measures designs: an extensive primer," vol. 97, no. 2, p. 316, 1985.
- [31] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. L. Tatham, *Multivariate data analysis* (no. 3). Prentice hall Upper Saddle River, NJ, 1998.
- [32] S. Coleman and J. Götze, *Bowling together: Online public engagement in policy deliberation*. Hansard Society London, 2001.
- [33] J. Mahmud, J. Chen, and J. Nichols, "Why are you more engaged? predicting social engagement from word use," *arXiv preprint arXiv:1402.6690*, 2014.
- [34] J. Hauke and T. J. Q. g. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data," vol. 30, no. 2, pp. 87-93, 2011.
- [35] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. N. Rosenquist, "Understanding the demographics of

Twitter users," in *Fifth international AAAI conference on weblogs and social media*, 2011.

- [36] J. Bright, S. Bermudez, J.-B. Pilet, and T. Soubiran, "Power users in online democracy: their origins and impact," *Information, Communication & Society*, pp. 1-16, 2019.
- [37] W. Fan and Z. Yan, "Factors affecting response rates of the web survey: A systematic review," *Computers in human behavior*, vol. 26, no. 2, pp. 132-139, 2010.
- [38] R. Fernandes and S. G. Leblanc, "Parametric (modified least squares) and non-parametric (Theil-Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors," *Remote Sensing of Environment*, vol. 95, no. 3, pp. 303-316, 2005.
- [39] E. Musa, "Crowdsourcing Social Innovation: Towards a Collaborative Social Capitalism," in *Social Entrepreneurship: Concepts, Methodologies, Tools, and Applications*: IGI Global, 2019, pp. 83-106.
- [40] K. B. Sumra and W. Bing, "Crowdsourcing in local public administration: importance of online platforms," *International Journal of Public Administration in the Digital Age*, vol. 3, no. 4, pp. 28-42, 2016.