

The Structure of Consensus Building in Online Design Discussions

Roshanak Zilouchian Moghaddam

University of Illinois at Urbana
201 N Goodwin Ave.
Urbana, IL, 61822
rzilouc2@illinois.edu

Brian P. Bailey

University of Illinois at Urbana
201 N Goodwin Ave.
Urbana, IL, 61822
bpbailey@illinois.edu

Wai-Tat Fu

University of Illinois at Urbana
201 N Goodwin Ave.
Urbana, IL, 61822
wfu@illinois.edu

ABSTRACT

While consensus building is inherently complex, the remote and asynchronous nature of online discussions, coupled with the large number of participants further intensifies the complexity. To better understand the nature of online consensus building, we first interviewed designers and developers contributing to design discussions in open source projects to understand the key factors in online consensus building. We then collected a large data set from an online design Web site and used regression analysis to further test how these factors affected consensus building. Our analysis showed that participants who were more experienced in the community facilitated consensus building. In addition, participants who had previous social interactions with each other would more likely reach a consensus in future discussions. We also found that even though the number of participants in each discussion was often large, small-group personal interactions did emerge, which significantly predicted the likelihood of reaching a consensus. Implication to design of interfaces that facilitate online design discussions is discussed.

Author Keywords

Online design discussions, consensus building, CSCW

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Consensus building is a key part of the any collaborative design process [1]. For instance, in open source software projects, user interface design issues are usually resolved through consensus building as they need to balance between

expertise from both user interface designers and developers. A successful consensus building process can encourage and support this form of *commons-based peer productions* [2] in two ways. First, it supports effective allocation and utilization of available human and technical resources in resolving product issues that are prioritized collectively by the online communities. Second, successful consensus building can be perceived as a positive signal that motivates the participants by reassuring them that their contribution has resulted in an improvement in the product, which is important for fostering stronger attachment among community members.

However, consensus building is a very complex process. Consensus building is difficult even in face-to-face settings. Thus, the remote and asynchronous nature of online discussions, coupled with the large number of participants, may further intensify its complexity. Researchers have recognized the importance yet complexity of consensus building and have found different factors affecting consensus building. Most notably they found that in face-to-face settings more social interactions between participants in the discussion [3] and less number of participants can foster consensus building [4]. Studies of decision making in computer-mediated discussions suggest that groups are more likely to come to consensus engaging in structured synchronous discussions [5]. When discussions are conducted online in open source projects, however, social interactions are often forced to be asynchronous, and they tend to involve a large number of participants. More research is needed to understand how previous research on consensus building in face-to-face and synchronous computer mediated settings can transfer to online design discussions.

However, only a few studies have studied consensus building in online design discussions. Ko et al. studied design discussions in the bug reports from the Bugzilla repositories of three open source projects with the goal of understanding interface design discussions and decision making process[6]. Unlike their study, our work focuses on the consensus building process when interface designers are present in the discussion. Zilouchian Moghaddam et al. analyzed interface design discussions in two open source

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2012, May 5-10, 2012, Austin, TX, USA.

Copyright 2012 ACM xxx-x-xxxx-xxxx-x/xx/xx...\$10.00.

projects [7]. They recommended a few improvements to the consensus building process such as enhancing the discussion interface by adding voting mechanisms. However, their early evaluations showed that participants mostly care about design arguments and are not in favor of a voting system.

To better understand the online consensus building process, we first surveyed previous research on consensus building and carried out semi-structured interviews with seventeen designers (N=8) and developers (N=9) participating in online design discussions. We identified three categories of potential factors affecting consensus building and we operationalized them into 23 quantitative metrics. We then calculated these metrics on a large data set crawled from an online software development community called Drupal, and tested how each category of factors might increase the chance of consensus building by casting these metrics as predictors in a binary (consensus or not) logistic regression analysis .

RELATED WORK

We describe prior work for consensus building in different domains and how our study builds upon and extends that work by exploring the nature of consensus building in online design discussions. We also describe the studies on collaboration in online communities and how we extend their work by studying collaborative decision making in collaborative online design.

Consensus Building

Consensus building is a common method for making decisions in resolving complex problems where no one person has all the required expertise or resources to solve the problem [8]. The consensus building process for solving complex problems has been studied in myriad domains. For instance Innes et al. studied how consensus building is useful for coordinating growth management in urban planning and environmental management [9]. One of their findings was that all interested stakeholders should be included in the process to ensure implementation. In another study Grünbacher et al. studied requirement negotiation and consensus building in eXtreme Programming (XP) [10]. They proposed a negotiation approach called EasyWinWin that defines a set of activities to guide the stakeholders in the negotiation process, such as brainstorm stakeholders' interests, prioritize win conditions, and reveal issues and constraints. We have extended these studies by exploring the consensus building process in the domain of online collaborative interface design. The complexity of design process and the difficulty of evaluating design alternatives can increase the complexity of consensus building in this domain.

Studies on consensus building also explored different factors that can affect consensus. For instance studies of consensus building in face-to-face settings suggest that social interaction among group members can induce

divergent viewpoints to converge over time and lead to consensus [3] and members tend to experience less satisfaction and less consensus when participating in large groups [4]. Studies of decision making in synchronous computer-mediated discussions suggest these discussions are more focused than face-to-face discussions[11], but they will take more time and have more difficulty in reaching consensus[11, 12]. Studies have also found that groups are more likely to come to consensus engaging in structured synchronous discussions [5]. We extend these works by studying the factors that affect consensus building in online communities where social interactions are asynchronous, social ties may be weaker, and the group size is often larger.

Online Discussions in Distributed Software Teams

Researchers have started to study online discussions in distributed software teams. Ko et al. studied bug report discussions in Firefox and found that only a fraction of these reports lead to changes in the software [13]. Barcellini et al. studied software discussions occurring in mailing lists of Python with the goal of extracting software and architectural design relevant information [14].

Other studies have been focused on usability related bugs or user interface design discussions. Twidale et al. analyzed discussions in the usability bugs from Bugzilla [15]. They recommended several improvements such as enhancing the classification of usability bugs and adding explicit representations of design arguments to the reports. Zilouchian Moghaddam et al. also examined usability discussions in two open source projects. They found that tracking design alternatives and having a mechanism to share opinions regarding design alternatives needs to be explicitly supported in the discussion interface [7]. Similarly Ko et al. analyzed design discussions in the bug reports from the Bugzilla repositories of Firefox, Linux, and Facebook API [6]. They observed common activities in design discussions such as establishing scope, proposing ideas, identifying design dimensions, defining claims with rationale, moderating process, and making decision.

We also studied online interface design discussions but we mainly focused on the consensus building process and the factors affecting this process.

Collaboration in Online Communities

Several studies have been conducted on different aspects of collaboration in online communities such as motivation of collaborators, the collaboration process, and the success factors. For instance, Nov explored the motivation of Wikipedia contributors [16]. Cranshaw et al. studied online collaboration in the Polymath Project, a group of mathematicians who collaborate online to solve open math problems [17]. Luther et al. examined the underlying reasons for the success of online creative collaboration [18]. Costa et al. explored the scale, range, and volatility of coordination requirements in large-scale distributed projects

and discussed the implications of their findings for design of collaborative tools [19]. We extend the prior work by shedding more light on the collaborative decision making and consensus building process in online design discussions. In addition, we propose a set of recommendations for supporting consensus building in online settings that can be utilized by designers of collaborative tools.

METHODOLOGY

To better understand the consensus building process and the factors affecting this process in online user interface design discussions, a mixed methods study was conducted. Our study consisted of analysis of a large corpus of interaction data, a set of 17 semi-structured interviews, and manually reviewing thirty discussion threads.

Collection of Interaction Data

The interaction data was pulled from the discussion threads (issues) in the issue management system of Drupal, an open source content management system initiated in 2001. The project has received a lot of attention over the past ten years. The project has managed to extend both its usage and number of contributors. Currently more than 490000 websites are using Drupal core to run their site and more than 440000 users have registered in drupal.org to participate in discussions and contribute to the project.

There are four types of discussion threads in Drupal issue management system: bug reports, feature requests, task, and support requests. According to drupal.org, bug reports are aimed to resolve functionality and usability problems. Feature requests are for adding completely new functionality. Tasks are not functional bugs, but things that ‘need to be done’, and support requests are asking for support. We only looked at the bug reports and feature request as they had almost all types of UI design discussions we would like to study. There are a total of 285008 bug reports and feature requests in Drupal issue management system. From these threads, we selected all discussion threads tagged with “Usability” or “d7ux” (usability in Drupal 7) for a total of 537 threads. The selected threads have been created between March 2004 and September 2011. The usability issues range from significant redesigns to minute details. For instance an issue called “Initial D7UX admin overlay” was aimed to completely change the interaction mechanism of admin pages in Drupal by adding an overlay. Another issue called “Shortcut “Edit shortcuts” link has insufficient color contrast ratio” which only suggest changing the color of a link. Figure 1 shows a sample usability discussion thread from Drupal website.

Any of the community members can initiate a discussion thread by describing a problem or ask for a new feature. Then other community members can participate in the discussion by proposing alternatives, examining the proposed alternatives, implementing an alternative (writing

a patch), reviewing a patch written by another member, or clarifying the scope of the problem. The participants can also indicate the progress of the discussion by updating the discussion’s status value (active, needs work, needs review, reviewed and tested by the community, fixed, and closed).

We used the status value to categorize the discussion threads into consensus, non-consensus, and ongoing threads (threads that may or may not reach to consensus in new future). We considered the threads with *closed* status value as consensus threads. A challenge was to differentiate between non-consensus and on-going threads. We calculated the number of weeks between the time that the last comment was posted in a thread to the time of the study (idle duration). We considered a thread to be non-consensus if its idle duration is more than 90% of the idle durations between comments in consensus threads. We regarded the rest of the threads as ongoing threads. In sum, we had 278 consensus threads, 233 non-consensus threads, and we discarded 51 ongoing threads because we were unable to categorize them.

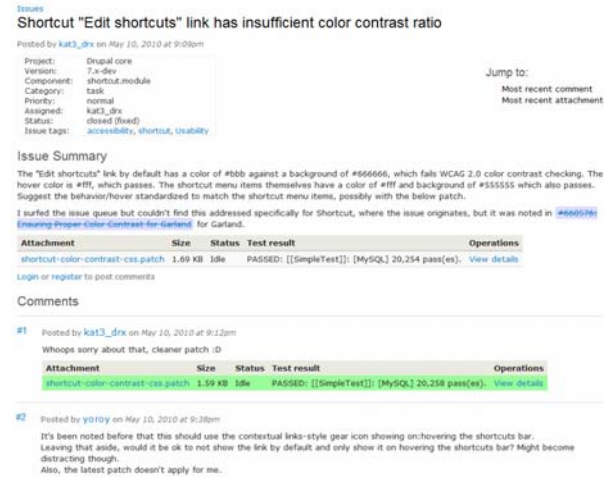


Figure 1. A sample user interface design discussion from Drupal issue management system.

In Table 1, we report summary statistics for the consensus and non-consensus categories.

We filtered our collected data set to only include threads that had a non-trivial amount of discussion. By reading a large fraction of collected data and experimenting with different appropriate thresholds, we found that the threshold of six comments filtered out almost all trivial discussions. After the filtering we had 196 consensus threads and 137 non-consensus threads.

Characteristic	Type	Mean	Stdev
Duration of a thread (weeks)	C	41.73	50.99
	N	98.17	75.39
Number of comments	C	48.38	61.61
	N	32.38	29.63

Number of authors	C	12.85	10.52
	N	11.63	6.89

Table 1. Summary statistics for the consensus and non-consensus categories (C indicates consensus and N indicates non-consensus threads).

Qualitative Data

The analysis of interaction data was complemented by conducting a set of 17 semi-structured interviews with designers and developers participating in two open source projects, Drupal and Ubuntu as well as reviewing thirty discussion threads from Drupal issue management system.

We interviewed eight active designers, five from Drupal and three from Ubuntu with an average of 4.5 years of experience within the community ($sd=2.6$). We also interviewed nine developers: six from Drupal and three from Ubuntu with an average of 5 years of experience within the community ($sd = 2.58$). Each interview lasted about an hour and was conducted via phone and instant messaging. The subjects were compensated with an Amazon gift card of at least \$25. We will refer to Drupal designers and Ubuntu designers as DD# and DU# and Drupal developers and Ubuntu developers as DevD# and DevU# accordingly.

We first asked the subjects to point out to one or two recent discussion threads they participated in. Then in the context of those threads, we asked them to describe the consensus building process, what is hard about this process, how they promote consensus around a specific solution, what factors affect the consensus building process, and what are some consequences of not reaching consensus.

The interviews helped steer the features we used in our data analysis, explain our results, and gain insight on the consensus building process. Twelve of the interviews preceded the data analysis and five interviews were conducted after where we tailored a few of our questions based upon the results. All the interviews were coded using a grounded theory approach.

In addition to the interviews, we selected thirty discussion threads from the list of the threads we chose for our data

Category	Metrics	Explanation\Justification
Content	(1) Average # of words in the comments of a thread	(1) - (3): Content qualities of the comments is one of the commonly used sources of data in research studies on online communities [23, 24] [21, 22]. From the content qualities we extracted the length of the comments and the length of the description of the issue. Usually, a longer text contains more information. However, long comments may be ignored or partially read by participants due to the lack of time. (4) - (8): In the interviews participants mentioned performing usability testing, writing issue summaries, providing screenshots, offering links to outside resources, and providing code reviews as factors that can accelerate the consensus building process.
	(2) Total # of words in a thread	
	(3) # of words used to describe the issue	
	(4) # of times usability testing was mentioned	
	(5) # of times summary was mentioned	

analysis. We sorted the threads based on three of the factors that we found to be important for consensus building in our interaction data analysis and we selected ten threads from each sorted list for a total of thirty threads. We coded the discussion threads using a grounded theory approach.

Research Framework

Based on preliminary interview results, consensus building literature [20], and prior research on content analysis in online communities [21, 22], we identified three categories of factors that can affect consensus building to use for an in-depth statistical analysis of interaction data: Content, Process, and User Relationships.

We operationalized these categories into a set of 23 metrics. The metrics in the content category can be used to determine whether the content qualities of the discussion correlate with reaching consensus. For instance, whether a lengthier discussion is more likely to reach consensus or whether a more detailed description of the issue promotes consensus. The process related metrics can reveal whether task activities as well as work and collaboration process affect consensus. For instance, whether use of synchronized communication media promotes consensus or whether submitting more patches to a thread increases the chance of reaching consensus. The metrics identified in the user relationships group can verify whether the relationships between users and content as well as the relationships between users affect consensus. For example, whether previous interactions between participants in a discussion promotes consensus or whether presence of more experienced participants supports consensus building.

The second column in Table 2 describes the metrics in each category. To calculate these metrics we have incorporated information from three different sources: the textual properties of the discussion threads (e.g. length of comments), metadata of the discussion threads (e.g. duration of the thread), and information from thread participants' Drupal profile. These metrics were computed by running algorithms such as pattern matching and page rank on the gathered information.

	(6) # of screenshots	
	(7) # of times code review was mentioned	
	(8) # of links to resources outside Drupal website	
Process	(9) # of times participants mentioned IRC	(9): From the interviews we found that having a conversation in IRC can promote consensus when the result of the conversation is reported back to the discussion thread. We used the number of times they mentioned “IRC” in the thread to capture the use of IRC.
	(10) # of questions marks	(10): Promoting mutual understanding and common ground fosters collaboration and consensus building [5]. The number of question marks can be used as a rough estimation for the number of times people tried to create mutual understanding.
	(11) Duration of the thread	
	(12) Average duration between comments	(11) – (12): Giving more time to a discussion may lead to finding the best solution. However, engaging in a discussion that spans over time is usually complex.
	(13) # of comments	
	(14) # of patches	(13) – (15): The thread with higher amount of activity may have a better chance of reaching consensus.
	(15) # of authors/participants	
(16) # of comments posted by thread creator	(16): A motivated thread creator usually tries to lead the discussion towards consensus.	
User Relationships	(17) # of triads in the social graph	(17)– (18): The relationship between users and content as well as the relationship between users are another common source for extracting metrics for further analysis on online communities [10, 11], [14]. We can view the data set of users, content (threads), and their relationship as a social graph. In such a graph, the relationships are represented by edges in a graph, with threads and users as nodes. There is an edge between a user and a thread when a user posts a comment in a thread. Also, there is an edge between two users when one replies to the other. The edges can be weighted based on the number of words in the comments. This graph can help in determining the whether the participants who had previous social interaction promote consensus. It can also help finding whether active participants (participants with higher page rank score [25]) can facilitate consensus building.
	(18) Average page rank score of participants	
	(19) Average # of participation weeks for participants	
	(20) Average # of previous comments posted by participants in other threads	
	(21) # of participation weeks of the creator	
	(22) # of comments the creator of a thread posted in other threads	
	(23) # of times participants reply to each other’s comments	
		(19) – (22): Our interview results showed that having experienced people participate in a discussion promotes consensus.
		(23): Posting direct replies to others comments can indicate a more focused discussion.

Table 2. Categories of metrics affecting consensus in online design discussions.

Analysis

To investigate the impact of the above metrics on consensus, we used them as variables to run a binary logistic regression analysis on the consensus and non-consensus discussion threads we pulled from Drupal issue management system. Binary logistic regression is a type of regression that is used to describe the relationship between one or more independent variables and a binary response variable (in this reaching consensus). We first inspected the correlations between the variables we extracted from the discussion threads. To avoid problems with collinearity in

the regression analysis, we left out twelve of the variables that had strong correlations ($r > 0.4$).

We used the binary logistic regression analysis as implemented in SPSS, and used step-down regression to identify our partial model, i.e., we first entered all variables at once, and took out each variable that did not show significance and repeated until we reached a set of variables that were all significant.

RESULTS

In order for consensus to be reached in online UI design discussions, participants need to agree on design as well as

the implementation of a proposal. This process usually consists of iterations of discussions on different design ideas, patch development, and code reviews.

In the following sections we first compare consensus building in usability and non-usability discussion threads. Then, we elaborate on the consequences of not reaching consensus. Finally, we turn our attention to our regression analysis results and we explain how the factors we identified in our regression analysis can affect the iterative process of design, development, and code review, and promote consensus. In our explanation we use our interview results, findings from the consensus building literature, and the insights we gained from analyzing thirty design discussion threads in Drupal.

Consensus Building in Usability vs. Non-Usability Discussion Threads

We found that consensus building in interface design issues is different from other technical issues. More participants are involved in design discussions both because these discussions are subjective and the outcome is visible to everyone. For instance, a total of 1404 unique participants posted comments on discussion threads tagged with “Usability”, however only 703 unique participants posted comments on discussion threads tagged with “Performance”. On average 9.60 (sd=9.57) unique participants contributed to each usability discussion and on average 7.70 (sd=7.73) unique participants posted comments on each performance discussion.

Also, conducting usability testing to assess the effectiveness of a design needs a lot of time and effort, while testing and gathering evidence in other areas such as performance and security is usually easier. For instance, participants on average spent 65.05 (sd=68.16) weeks on usability discussion threads versus 47.6 (sd=53.05) weeks on performance discussion threads. Having more people participate in a subjective discussion where it is hard to get evidence makes consensus building in design discussions harder than other technical areas. As DevD4 said:

“...I think the usability testing is a little harder to do often. It takes a little more time, if you do an informal one it's not so bad, but you certainly can't do with just one person, [...] you have to get several, and it takes time to get evidence. In some other areas in Drupal it's easier to get evidence. You know, what percent [it makes the Drupal website] faster. Usability also, is an area usually more people can get involved, some of the highly technical stuff not that many people can get involved in the issue without background. In usability it would be easier, anyone can suggest a design: I think we should do it like that. It tends to get a lot more people involved; it's also a lot more visible [...]. Often more people are participating, more people means it's harder to [build] consensus.”

Consequences of Not Reaching Consensus

Of the threads we analyzed, 41% did not reach consensus. A total of 1605 participants contributed to the threads that

did not reach consensus by posting 4469 comments and developing 413 patches. A lot of these comments and patches were not even considered or reviewed by other participants. Ignoring people’s contribution and failure to reach consensus can result in an unimproved product, build resentment among team, and demotivate people and make them leave other discussions temporarily or even leave the community permanently. As DD5 said:

“[Consequences of not reaching consensus are] stupid interfaces surviving yet another version in Drupal, known issues not being fixed, frustrated contributors. Consequences can be that people disappear for a couple of weeks or entirely because they get burned out on a too long discussion that didn't reach consensus...”

On the other hand when consensus is reached, community members experience stronger connections with each other and they can make a better Drupal. As DD4 said:

“...when we reach consensus we are taking our strengths [to] make the world together, we have something that is at least as good as what the two of us could bring to separately, and probably is better because our strengths tend to reinforce each other.”

This emphasizes the importance of studying the nature of consensus building and designing tools to support the consensus building process.

Factors affecting consensus building in online design discussions

Running the binary logistic regression analysis, three of the 23 metrics showed significance (p<0.05): Average # of participation weeks, # of triads in social graph, and IRC. Table 3 shows these variables. To assess the goodness of fit of our model we also performed the Hosmer-Lemeshow test (Chi-square=10.557, p=0.228) which confirmed the validity of our model. In this test the model is valid if the p-value is greater than 0.05.

	B	df	Sig.	Exp(B)
Average # of participation weeks	.006	1	.007	1.006
# of triads in social graph	.136	1	.014	1.145
IRC	.265	1	.050	1.304
Constant	-1.02	1	.019	.361

Table 3. Regression analysis results. The Hosmer-Lemeshow test confirmed the overall validity of our model (Chi-square=10.557, p=0.228).

The average number of participation weeks represents the experience of the thread participants in the Drupal community, number of triads in social graph shows the number of three participants who had previous interactions with each other, and IRC shows the number of times participants used synchronized communication media in addition to the issue management system. In the following sections we use the result of our interviews as well as the

insight we gained from reading thirty discussion threads in Drupal to explain how each of the factors we found in our regression analysis promote consensus.

Experience with Drupal

Our regression analysis showed that having people in a discussion who have participated in Drupal longer can promote consensus. Research studies also confirm the positive influence of the experienced people on group performance [26].

Our interview results and the outcome of analyzing discussion threads allowed us to shed more light on different ways experienced people can facilitate the consensus building process. First, we learned that people who have been in the community for a long time are familiar with the community's communication standards and participation process. They can facilitate the consensus building process by helping other participants and especially new contributors to understand the process of participation. As DD4 said:

"...what's important for reaching consensus is having common ground rules or communication and process and you know working those out and to that extend more experienced in Drupal community might help people be better at reaching consensus, because they'll understand that you don't say things this way or you do, or these are the options for contacting somebody if you have a problem or that kind of thing..."

Second, we found that experienced people can unblock a discussion. Comments and opinions posted by experienced members are valued more than those posted by other participants. As DD1 said:

"...there are those people in the community that are recognized – people who have been in the community a long time, or who are respected because they have written a lot of code, or they have written a lot of patches, or they are the maintainer of a certain bit of code – and when those people chime in, it tends to hold a little bit more weight when someone unknown chimes in."

As a result, experienced members can direct the discussion toward a specific solution or help to unblock a discussion. As DDev6 who was involved in the community for more than 6 years said:

"...in a situation where discussion is sort of going around circles and no consensus is forming or a situation where a heated thread is starting to get really ugly and people are getting personal and things like that then often times we will chime in and [...] suggest a next step..."

Third, our qualitative analysis showed that experienced people could also promote consensus by proposing solutions that satisfies all opposite parties. Sometimes consensus may be delayed because of two or more competing alternatives. Proposing a new alternative that

accommodate the taste of opposite parties can resolve the conflict. As DevU1 said:

"If there is a way to produce an outcome which solves both problems then I prefer to take that option...four different options that you have can be combined together to produce a new idea that might actually solve more problems and be more beneficial"

Conforming our result, seeking a new alternative is a suggested strategy in prior work for resolving conflicts and building consensus [20].

Finally, we found that experienced people can facilitate consensus building by providing good arguments for or against proposed alternatives. The comments that contain profound design arguments advocate shared understanding and help the participants in a discussion to come to an agreement. As DU3 said:

"... I think that the general solution is to talk about the alternatives you considered, and talk about what benefits and what disadvantages were. And, if you do that, then people are much more likely to, to come to an agreement, and say that was a good decision, or you're gonna say disagree, to disagree with the decision, at least I can see why [you disagree]."

This also conforms with research studies that suggest promoting mutual understanding and resolving differences of mental model as strategies for building consensus [20].

Previous Social Interaction with Other Participants

Our regression analysis showed that having more triadic closures in the discussion threads increase the chance of coming to consensus. Triadic closures represent three people who had strong social interaction with each other. Triadic closures produce closed social structures that can promote greater trust among individuals [27].

Interviews and studying discussion threads helped us confirm trust as an important factor in consensus building. First, we found that participants are more likely to read, learn from, and evaluate the comments posted by individuals whom they trust. This exchange of knowledge can create mutual understanding and consequently promotes consensus. As DevD4 said:

"I think I'm less likely to dismiss something if it's from somebody I know and I respect. It's a little more likely to read carefully what they say and believe that they have something meaningful to say"

This finding reflects findings in other research studies that indicate trust in dense parts of the social network facilitates exchange of complex knowledge [28].

Second, our qualitative analysis showed that strong social interaction and trust among participants in a discussion thread can also promote consensus by provoking agreement among individuals. People tend to agree with the individuals whom they trust. As DevD6 said:

“... it's sort of like a trust matrix type of thing, because if I don't know you and you are suggesting this thing that sounds like a bad idea to me, I probably fight against it, but if you are proposing something and I don't know you but three other people that I do know are saying yeah actually that's a great Idea and this is why, then I'll be far more likely to be like alright let's go with it then.”

Finally, we found that trusting other participants' technical abilities can save time in the process. For instance, knowing that the person who wrote a patch usually conforms to coding standard can accelerate code review. As DevD6 said:

“... if people know who somebody else is, it saves a hack of a lot of time, at all levels, like, for example, if I know the person who wrote the patch and I know that traditionally they write pretty good patches that conform to coding standard and stuff like that and then I see the person that reviewed it is the person I associate with being the smart person about that thing and the person who marked it as reviewed and tested by community [...], and that person was also someone I recognize as if they say something is RTBC it's actually good to go. Then it saves all kinds of time.”

Interaction through Synchronized Communication Media

Our regression analysis showed that threads that contain more occurrence of the word “IRC” are more likely to reach consensus. The interview results and analysis of discussion threads allowed us to elaborate on the nature of using IRC in Drupal discussion threads.

We found that as part of their work process, participants in online design discussions use synchronized communication media such as IRC to ask questions, request for development help, or have more detailed discussion on a controversial topic. As DevD3 said:

“... the other main venue is on IRC when people ask me to look at stuff or just discuss something...”

A group of two to five people are usually participating in IRC discussions. Reporting the result of such discussions back to the issue queue can foster consensus building among the larger group of participants.

From both the interview study and reading discussion threads, we identified three major ways that IRC can help the consensus building process. First, we found that having discussions in IRC can accelerate agreement between two opposite parties. As DD4 said:

“... it [IRC] can help if there is one or two people who are disagreeing about something, if those people go to IRC they can chat it out much faster than the issue queue.”

Second, in IRC people can come up with an initial proposal for solving the problem that may not be possible in the issue queue. As DevD6 said:

“IRC is great for say a small group of people going off and coming up with an initial proposal that they all agree on and then proposing that to the community.”

Reporting this proposal back to the issue queue advances the consensus building process, because with this proposal people will have a position to argue about and discuss, as opposed to trying to come to a position. As DevD6 said:

“...then it becomes let's argue against this position as opposed to try to come to a position to argue against...”

Finally, we learned that participants in the discussion threads form small groups in IRC to engage in synchronized collaborative design review, programming, or debugging sessions. These sessions can help participants to quicken the design review, development, and code review activities. Therefore accelerates the whole process of coming to consensus. For instance in an issue where participants try to design and implement an overlay for Drupal interface, one of the developers (Y) asks another developer (X) to join him in an IRC chat for a collaborative debugging session. Y says:

“X, some of your files are being cut off, such as overlay-parent.css(?) Please come onto IRC so we can help you debug?”

DISCUSSION

Our analysis showed that the number of times participants mentioned IRC in a discussion thread is predictive of consensus. One may argue that participants can only use IRC for resolving software issues. However it is important to note that participants contribute to Drupal from all over the world and they may not be available in IRC at the same time. Therefore people who commit the codes will only rely on the decisions made on the issue management system. As DD5 said:

“No decisions are made in IRC in the sense that the persons who will commit the files, patched as a code will not rely on IRC discussions; they would want to rely on the actual discussion in the issue...”

IRC is only useful if the result of the discussion would be reported back to the issue queue where all the members of the community have access to. Failure to report these discussions back to the issue queue may cause other participants too loose context and therefore delays consensus. As DevD6 said:

“The danger in IRC becomes when and this happens sometimes when there are huge discussions that go on IRC, big community impacting discussions and only the people who happen to be on IRC at that time, know about them and if those don't make their way back to the issue queue or groups or some other mean of more permanent storage that's really dangerous because a lot of people loose context in these discussions that way”

A number of factors mentioned by our interviewees did not show significance in the regression analysis. For instance, interviewees mentioned performing informal usability studies often positively influence consensus. One reason behind the absence of usability studies from the significant factors may be the overhead of performing these tests. Due to this overhead participants may skip performing any formal or informal usability studies. In our data set, only 0.06% of the threads that reached consensus talked about usability studies. Designers and developer can advocate for more usability testing by setting up a testing platform where users can easily try a new version of the interface without having to worry about deployment. Using such testing platform all non-technical users of Drupal can try a controversial design and provide feedback.

DESIGN IMPLICATIONS

Our results showed that when discussions are struggling to reach consensus, participants in a thread can solicit other participants that have strong social connections with the current participants to promote consensus. Multiple strategies can be employed to find and invite people who have strong social connections with current participants. One strategy might be to implement a searching mechanism that could be used to find participants who had certain amount of previous interaction. Another strategy would be to make available a list of socially close people who could be beneficial for the discussion. Such a list can be compiled based on participants' social network [29] or by processing the history of participants' contributions [30]. According to our results, inviting experienced participants can also promote consensus. Similar strategies can be employed to invite experienced people. Inviting appropriate people to join the discussion will not only help the consensus building process, but also will help participants to find issues that they are interested in.

Our qualitative review of interviews and discussion threads revealed that certain types of comments can help the consensus building process more than the others. For instance comments that include strong arguments for or against design alternatives can advocate agreement between participants, comments that summarize the discussion to date can help participants in making sense of the thread, or comments that include reports of the discussions occurred on IRC can help participants to remain in context. The discussion interface could employ iconic signifiers or color coding to highlight these different types of comments. To classify the type of comments one way would be to ask participants to tag the type of comments, but this may impose a small burden on participants. An alternative strategy would be to automatically infer the comment types. However the results may not be accurate.

Our analysis showed that thread participants value the comments posted by experienced people more than others. Also, participants tend to read the comments and trust the contribution of people who are socially close to them. There

are multiple ways to help people locate these comments: searching for the comments posted by a particular participant, filtering the comments, or creating a summary list of all participants where their experience and social connectedness is displayed.

To further accelerate the design, implementation, and code review in the discussion threads, recent key contributions to the discussion thread can be highlighted. For instance a list of the comments that include key contributions such as proposing the most recent design idea, implementing or reviewing a recent patch, or changing the status of the thread can be provided. It is important to note that not all of the filtering, searching, and highlighting mechanisms discussed above should necessarily be included in the thread at the same time.

Limitations

We did not study all of the potential factors affecting consensus building. We only looked at a range of factors that were easy to compute. Human coding can be used in future studies to include more sophisticated factors such as number of design ideas, number of arguments for or against ideas, or the effective tone of the comments.

We also note that our study only focused on the interface design discussions in Drupal and the results may not be generalizable to other open source projects. Also, the inferences of our qualitative study are based on only 17 participants. We would like to expand our work both by studying the interface design discussions in other open source projects and surveying more participants in Drupal and other open source projects.

CONCLUSION

We studied the nature of consensus building and the factors affecting consensus in online design discussions. Our analysis showed that besides having participants that are experienced with the software and the issue management system, having social ties and connections between participants promote consensus. Further we found that smaller groups are usually formed in design discussions that often utilize synchronous communication media to collaborate and foster consensus among the larger group of participants. This is in contrast to many who believe that commons-based peer production does not involve much bond-based interactions, but our results suggest that small-group personal interactions do emerge and they do play a significant role in consensus building. This may be an interesting phenomenon that deserves further research.

Directions for Future Work

We propose three directions for future work. First, we can expand the study by capturing and analyzing the IRC logs related to each discussion thread. This analysis can better describe the impact of synchronized communication media on consensus. It can also reveal potential strategies that participants use in IRC.

Secondly, our study only examined the interface design discussions in Drupal website. A similar study can be conducted on other types of discussions and in other projects to verify the generalizability of the results.

Finally, we are interested in redesigning the discussion interface based on our findings and evaluate the impact of our findings in actual work environments.

REFERENCES

- [1] Everitt, K. M., Klemmer, S. R., Lee, R. and Landay, J. A. Two worlds apart: bridging the gap between physical and virtual media for distributed design collaboration. In *Proceedings of the CHI* (2003).
- [2] Benkler, Y. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press., 2006.
- [3] Fiol, C. M. Consensus, diversity, and learning in organizations. *Organization Science* 5, 3 (1994), 403-420.
- [4] Hare, A. P. A Study of Interaction and Consensus in Different Sized Groups. *American Sociological Review*, 17, 3 (1952), 261-267.
- [5] Farnham, S., Chesley, H. R., McGhee, D. E., Kawal, R. and Landau, J. Structured Online Interactions: Improving the Decision Making of Small Discussion Groups. In *Proceedings of CSCW* (2000).
- [6] Ko, A. J. and Chilana, P. K. Design, discussion, and dissent in open bug reports. *iConference* (2011).
- [7] Zilouchian Moghaddam, R., Bailey, B. and Poon, C. IdeaTracker: An Interactive Visualization Supporting Collaboration and Consensus Building in Online Interface Design Discussions. In *Proceedings of INTERACT* (2011)
- [8] Susskind, L., McKernan, S. and Thomas-Larmer, J. *The Consensus Building Handbook*. City, 1999.
- [9] Innes, J. E., Gruber, J., Thompson, R. and Neuman, M. *Coordinating Growth and Environmental Management through Consensus Building*. University of California, 1994.
- [10] Grünbacher, P. and Hofer, C. Complementing XP with Requirements Negotiation. In *Proceedings of the 3rd International Conference on eXtreme Programming and Agile Processes in Software Engineering* (2002).
- [11] Straus, S. G. and McGrath, J. Does the medium matter? The interaction of task type and technology on group performance and member reactions. *Journal of Applied Psychology*, 79 (1994), 87-97.
- [12] HILTZ, S. R., JOHNSON, K. and TUROFF, M. Experiments in Group Decision Making Communication Process and Outcome in Face-to-Face Versus Computerized Conferences. *Human Communication Research*, 13(1986), 225-252.
- [13] Ko, A. J. and Chilana, P. K. How power users help and hinder open bug reporting. In *Proceedings CHI* (2010).
- [14] Barcellini, F., Détienne, F., Burkhardt, J.-M. and Sack, W. *A Study of Online Discussions in an Open-Source Software Community* (2005).
- [15] Twidale, M. B. and Nichols, D. M. Exploring Usability Discussions in Open Source Development. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences - Volume 07* (2005).
- [16] Nov, O. What motivates Wikipedians? *Commun. ACM*, 50, 11 (2007), 60-64.
- [17] Cranshaw, J. and Kittur, A. The polymath project: lessons from a successful online collaboration in mathematics. In *Proceedings of CHI*(2011).
- [18] Luther, K., Caine, K., Ziegler, K. and Bruckman, A. Why it works (when it works): success factors in online creative collaboration. In *Proceedings of the 16th ACM international conference on Supporting group work* (2010).
- [19] Costa, J. M., Cataldo, M. and Souza, C. R. d. The scale and evolution of coordination needs in large-scale distributed projects: implications for the future generation of collaborative tools. In *Proceedings of CHI* (2011).
- [20] Briggs, R. O., Kolfshoten, G. L. and Vreede, G.-J. d. Toward a Theoretical Model of Consensus Building. In *Proceedings of the AMCIS 2005* (2005).
- [21] Adamic, L. A., Zhang, J., Bakshy, E. and Ackerman, M. S. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *Proceedings of the WWW* (2008).
- [22] Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G. Finding High-Quality Content in Social Media. In *Proceedings of the WSDM* (2008).
- [23] Cook, E., Teasley, S. D. and Ackerman, M. S. Contribution, Commercialization & Audience: Understanding Participation in an Online Creative Community. In *Proceedings of the GROUP* (2009).
- [24] Diakopoulos, N. and Naaman, M. Towards Quality Discourse in Online News Comments. In *Proceedings of the CSCW* (2011).
- [25] Page, L., Brin, S., Motwani, R. and Winograd, T. *The PageRank citation ranking: bringing order to the Web*. Stanford Digital Library Technologies Project, 1998.
- [26] Kerr, N. L. and Tindale, R. S. Group Performance and Decision Making. *Annual Review Psychology*, 55 (2004), 623-655.
- [27] Uzzi, B. Social structure and competition in interfirm networks: the paradox of embeddedness. *Administrative Science Quarterly*, 42, 1 (1997), 35-67.
- [28] Zaheer, A. and Bell, G. G. Benefiting from network position: firm capabilities, structural holes, and performance. *Strategic Management Journal* 26, 9 (2005), 809-825.
- [29] Horowitz, D. and Kamvar, S. D. The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web* (2010).
- [30] Cosley, D., Frankowski, D., Terveen, L. and Riedl, J. SuggestBot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of IUI* (2007).