

The Effects of Diversity on Group Productivity and Member Withdrawal in Online Volunteer Groups

Jilin Chen*, Yuqing Ren**, John Riedl*

*Dept of Computer Science and Engineering
University of Minnesota
200 Union Street SE, Minneapolis, MN 55455
{jilin, riedl}@cs.umn.edu

**Carlson School of Management
University of Minnesota
321 19th Ave S, Minneapolis, MN 55455
chingren@umn.edu

ABSTRACT

The “wisdom of crowds” argument emphasizes the importance of diversity in online collaborations, such as open source projects and Wikipedia. However, decades of research on diversity in offline work groups have painted an inconclusive picture. On the one hand, the broader range of insights from a diverse group can lead to improved outcomes. On the other hand, individual differences can lead to conflict and diminished performance. In this paper, we examine the effects of group diversity on the amount of work accomplished and on member withdrawal behaviors in the context of WikiProjects. We find that increased diversity in experience with Wikipedia increases group productivity and decreases member withdrawal – up to a point. Beyond that point, group productivity remains high, but members are more likely to withdraw. Strikingly, no such diminishing returns were observed for differences in member interest, which increases productivity and decreases member withdrawal in a linear fashion. Our results suggest that the low visibility of individual differences in online groups may allow them to harvest more of the benefits of diversity while bearing less of the cost. We discuss how our findings can inform further research of online collaboration.

Author Keywords

Online volunteer group, diversity, performance, Wikipedia.

ACM Classification Keywords

H.5.3: Group and Organization Interfaces, K.4.3: Organizational Impacts

General Terms

Experimentation, Theory

INTRODUCTION

The Internet has evolved into a powerful platform to support online collaboration, ranging from writing software and editing Wikipedia articles, to designing fashion T-

shirts. Online collaboration harvests the “wisdom of crowds” by involving a group of people with diverse backgrounds, information, and perspectives. Appropriately organized, these diverse independently-deciding individuals can make better decisions than an expert [25]. Diversity, nonetheless, has been shown to work as a double-edge sword in offline work groups in organizations. On the one hand, consistent with the “wisdom of crowds” argument, groups that integrate a diverse set of information and perspectives in their information-gathering or decision-making processes can be more productive and make better decisions. On the other hand, different perspectives and opinions can trigger dysfunctional group processes (e.g., newcomers versus them old-timers) that may cause conflict, delays in action, and member dissatisfaction [26, 28].

Wikipedia serves as a good example to show that similar processes may occur in the context of online collaboration. As an online encyclopedia solely created by hundreds of thousands of volunteers, the English version of Wikipedia alone includes more than 3 million articles, more than 10 million registered user accounts, and has been consistently ranked among the top ten visited websites on the Internet by Alexa.com. Studies have shown its quality to be comparable to traditional encyclopedias [11], and despite widespread concern, most incidents of vandalism are repaired within minutes [24]. At the same time, since the birth of Wikipedia, editors and administrators have been dealing with the challenge of resolving differences effectively and preventing conflict from escalating. Out-of-control conflict threatens to ruin articles, or destroy the collaborative spirit of Wikipedia. In fact, one recurring problem on Wikipedia is “edit wars”, in which two editors fight back and forth over the content of an article. Wikipedia even includes a page with a long list of the “lamest edit wars”, including debates over Obama's eligibility as the Commander-in-Chief, Jimmy Wales' status as the founder (or co-founder) of Wikipedia, and the question of whether the tiger is the most powerful living cat, just to name a few.

We know little about how the effects of diversity manifest in online collaboration where social cues are less visible and members never or rarely meet each other in person. Most diversity studies have examined work groups

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 2010, April 10–15, 2010, Atlanta, Georgia, USA.

Copyright 2010 ACM 978-1-60558-929-9/10/04...\$10.00.

in organizations or ad hoc groups in laboratory experiments with clear membership boundaries, and whose members interact face-to-face rather than with the assistance of computer technologies. The present paper seeks to extend the results of these studies to online volunteer groups like Wikipedia or open source projects.

By *online volunteer groups*, we refer to a group of people who meet and work online as volunteers to create artifacts of lasting value to a broad community. Both Wikipedia and open source projects like Linux or Apache are high-profile examples of online volunteer groups. Understanding the impact of diversity on the performance of online volunteer groups is important and valuable. These groups are increasingly creating artifacts that are being relied on for a wide variety of activities, including commercial activities such as running major Web sites, and non-commercial activities, such as education and research from grade school through graduate school. Our long-term goal is to understand the conditions under which the group members thrive and perform most effectively.

In this study we investigate the effects group diversity – on collaboration outcomes in the context of Wikipedia Projects (<http://en.wikipedia.org/wiki/WikiProject>, referred to as WikiProjects hereafter). A WikiProject is an entity created within Wikipedia to help coordinate and organize the writing and editing of a collection of pages devoted to a specific topic or family of topics. Using data from the January 2008 full dump of the English Wikipedia, we measure the extent to which members of a WikiProject differ in their level of experience as Wikipedia editors (tenure diversity) and the extent to which members differ in their domains of interest (interest diversity), and examine how the two types of diversity affect the amount of work that a project accomplishes and members' willingness to stay and contribute to the project.

Our results suggest that both tenure and interest diversity significantly influence group outcomes of WikiProjects. Increased tenure diversity increases group productivity but with diminishing returns, until very high tenure diversity actually decreases group productivity. Moderate tenure diversity is associated with low member withdrawal while extremely low or high tenure diversity drives members away. In contrast, interest diversity increases productivity and decreases member withdrawal in a linear fashion. We believe that our findings are not only meaningful in extending diversity theories from offline to online, but also useful in promoting effective online collaboration.

The rest of the paper is organized as follows. The *Diversity in Online Volunteer Groups* section presents an analysis of prior research on diversity that motivates a set of hypotheses about how diversity might operate in online groups. The *WikiProjects as a Study Platform* section outlines the workings of WikiProjects for the reader. The *Data Collection and Methodology* section describes the data collected, and the way Hierarchical Linear Modeling

was used. The *Results* and *Discussion* sections present and discuss our main findings. Finally, the paper ends with a brief section of conclusion, limitation and future directions.

DIVERSITY IN ONLINE VOLUNTEER GROUPS

Group diversity is commonly defined as differences among individuals on any attributes that will lead to the perception that others are different from oneself [26]. The attributes can range from social attributes such as age, gender, race, and nationality [2] to informational attributes such as tenure [21], education [9], and functional areas [8], to deeper-level individual differences such as personality and beliefs.

Most research on diversity focuses on face-to-face work groups in organizations or ad hoc groups in laboratory experiments. In spite of decades of research, the main effects of diversity in work groups remain controversial and inconclusive (see [27, 28] for reviews). Various studies have found positive, negative, or no relationships between diversity and performance. The literature has agreed, however, upon two mechanisms through which diversity affects group performance and member well-being: the informational or decision-making perspective and the social categorization perspective [26, 28].

According to the *informational or decision-making perspective*, heterogeneous groups should outperform homogeneous groups because the former have access to a broader range of knowledge, skills, abilities, and opinions, and are thus able to consider all distinct information related to the task at hand and come up with better decisions.

According to the *social categorization perspective*, homogenous groups should outperform heterogeneous groups because people use differences in social attributes as cues to categorize self and others into social groups and as a result, they favor, pay more attention to, feel more satisfied with, and have more positive evaluations when working with similar others than dissimilar others. This categorization process often results in subgroup dynamics (us versus them) and a high level of interpersonal conflict.

Recent reviews suggest that the overall effects of diversity on group performance may be contingent upon a number of contextual factors such as group type, group size, task complexity, and task interdependency [13, 27]. Meanwhile, factors such as group size, skills and abilities, and task type have been shown to operate very differently in offline and online contexts [19]. As a result, it seems reasonable to expect that the effects of diversity in online groups are different those in the offline groups.

Hypotheses of the Effects of Diversity

We choose to study tenure and interest diversity because both attributes are visible and highly related to tasks in online volunteer groups [27]. Compared to demographic attributes (e.g., age, gender, and ethnicity) or value attributes (e.g., personality, values, and beliefs), information about member tenure and interests is more accessible in online volunteer groups.

We examine two dimensions of group outcomes: group cognitive performance measured as amount of work accomplished, and group affective performance measured through member withdrawal from contributing to group effort. Both measures have been extensively studied in offline work groups, and are meaningful in online volunteer groups. For an online volunteer group to be successful, it needs to fulfill its group goals (i.e., to create artifacts of lasting value) and to fulfill those goals, it needs to keep members satisfied and actively participating in group tasks.

High tenure diversity means high variability among group members in the time they have spent working on group tasks and in the experience they have accumulated. While one review of the diversity literature concludes that the effects of tenure diversity on group cognitive performance are mixed [20], another review reports positive effects of tenure diversity on the quantity of team performance [13].

In offline studies, tenure diversity in project teams has often led to improved team performance, because old-timers provide effective team structure, while newcomers bring in fresh perspectives, skills, and ideas [1]. If these results also apply online, having a mixture of newcomers and old-timers may improve the performance of online volunteer groups. Furthermore, studies of online collaboration have found that members with different tenure are often interested in performing different kinds of tasks [6], so teams with both newcomers and old-timers may also enjoy better task distribution. However, when tenure diversity gets extremely high, old-timers and newcomers may have insufficient common ground, leading to difficulty in communication and coordination. As a result, old-timers may encounter conflict with newcomers, and group productivity may suffer. We thus posit:

Hypothesis 1: High tenure diversity leads to high group productivity, but with diminishing returns. Increasing tenure diversity beyond certain levels will decrease group productivity.

Interest diversity in online volunteer groups is similar to educational or functional diversity in offline work groups. We thus expect positive effects of interest diversity on group productivity because offline research shows that groups with diverse member interests have access to a broad set of knowledge, information, and perspectives. With proper coordination [14], diverse groups benefit from accessing and integrating a broad range of information to accomplish group goals [9]. When interest diversity increases beyond certain levels, however, members are likely to lose common ground that allows in-depth understanding and incorporation of diverse members' input, and group productivity may suffer [27]. For instance, organizational behavior researchers and computer scientists have to work to establish a common vocabulary before they can effectively work together.

Hypothesis 2: High interest diversity leads to high group productivity, but with diminishing returns. Increasing interest diversity beyond certain levels will decrease group productivity.

There is consensus in the diversity literature that tenure diversity reduces social integration and makes people more likely to leave the group when conflict arises [21]. Some studies have also found that even when unsatisfied members do not leave, they are less willing to take on tasks or to contribute effort and ideas to the group [20]. Although a member's tenure within a group is not immediately visible to all members of the group, it can surface or be retrieved with relative ease as members begin interacting with one another. Kriplean et al. [18] described several instances in which Wikipedia newcomers and old-timers fought over the scope of articles, including one case in which the old-timer accused the newcomers as being "like a kid just out of high school" and another case in which one party threatened to stop contributing. We posit:

Hypothesis 3: High tenure diversity leads to high member withdrawal. Groups with high tenure diversity are more likely to experience member withdrawal.

We did not find many studies that have examined the effects of functional diversity on member withdrawal. Some studies found that educational diversity leads to higher withdrawal [20]. We speculate that interest diversity has similar effects as tenure diversity on member withdrawal in online volunteer groups because (1) both tenure and interest are task related attributes, and (2) both types of diversity in principle may elicit social categorization processes and cause negative interactions among members [27]. We posit:

Hypothesis 4: High interest diversity leads to high member withdrawal. Groups with high interest diversity are more likely to experience member withdrawal.

WIKIPROJECTS AS A STUDY PLATFORM

Understanding some basic concepts about Wikipedia is important for understanding this research. Wikipedia is organized in *pages*, including pages of encyclopedic entries, which we refer as *articles*, as well as pages dedicated to specific editors, guidelines and projects. Any Web user can become an *editor* of Wikipedia, by visiting nearly any page and clicking the *edit this page* link. After editing the page, the editor can save it as a *new revision*. We call this process an *edit*. Every regular page also has an associated *talk* page, which is used for discussion related to the page. Talk pages are edited in the same way as other Wikipedia pages. Pages within Wikipedia are loosely organized into *categories*. Categories are labels that can be applied to any page by an editor. Some categories are declared to be subcategories of other categories, and most categories are linked together via these relationships into a large graph structure. Categories are a useful way to classify the pages of Wikipedia into topics.

One way that Wikipedia editors organize their efforts are through WikiProjects. A WikiProject is "a collaboration area and open group of editors dedicated to improving Wikipedia's coverage of a particular topic, or to organizing some internal Wikipedia process". As shown in Figure 1, since 2002 more than 20,000 Wikipedia editors have joined more than a thousand projects. In this study, we focus on

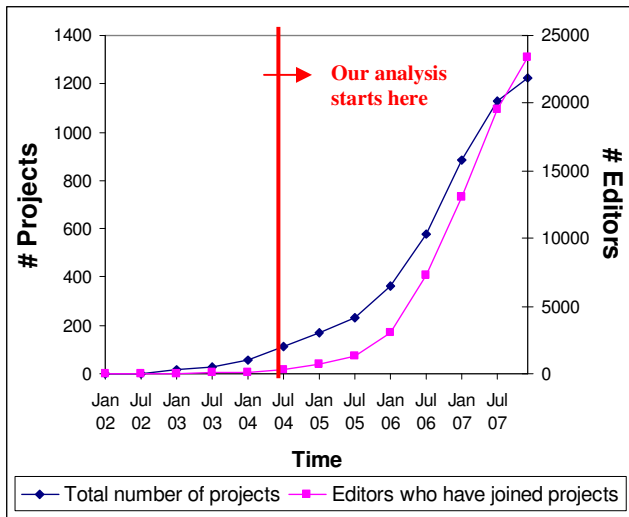


Figure 1: Timeline of WikiProjects in Wikipedia.

The x axis is time from the beginning of WikiProjects to the time when the dump used in this research ends. The left y axis is the number of projects in Wikipedia. The right y axis is the number of editors who have joined WikiProjects.

“topical WikiProjects”, which are created to improve articles within a certain topic area (e.g., Africa or Computer Science) by expanding article content, aligning articles to the same style of writing, and peer review of article quality in a topic area. Figure 2 shows the main page of *WikiProject Computer Science* as an example. It states the scope of the project, a list of the members of the project, a list of tasks to be done, and various guidelines for members.

WikiProjects provide a good setting to study the effects of diversity in online volunteer groups for several reasons. First, compared to Wikipedia articles with open boundaries, most WikiProjects manage their membership using a member list on which members can sign or remove their names. Second, WikiProjects have clear goals, organized activities to meet those goals, and rich historical data so that we can measure performance and membership changes over time. Third, WikiProjects are representative of other online volunteer groups, such as open source projects; therefore, we believe our findings can be generalized.

DATA COLLECTION AND METHODOLOGY

The dataset we use in this study is extracted from the January 2008 dump of English Wikipedia, which includes full text of all pages and their complete edit history from the beginning of Wikipedia. To gather our sample of projects, we traversed from the main directory page of WikiProjects, and excluded projects that are not topical (e.g., WikiProject Citation Cleanup). We also excluded projects that never grew to have at least three members (the minimal size of a group), projects that do not have a member list to track membership, and projects whose scope cannot be estimated using categories. In the first two years of Wikipedia the level of activity in WikiProjects is very different from more recent years due to the small number of projects and editors involved (see Figure 1). We therefore

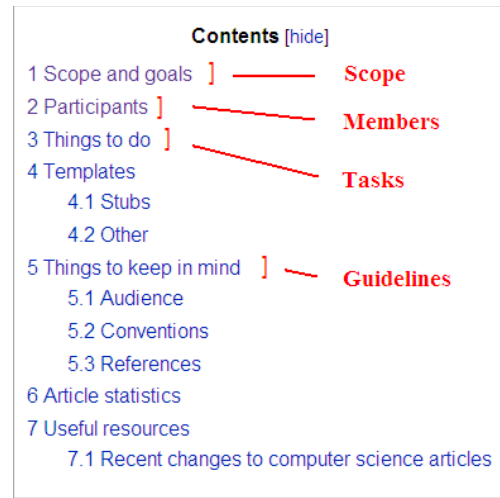


Figure 2: Screenshot of Main Page of WikiProject on Computer Science.

excluded projects created before June 2004. The resulting dataset has 683 WikiProjects.

We estimated the scope of each WikiProject (i.e. what articles fall under the project) by finding the Wikipedia category that matches its title (like category *Computer science* for *WikiProject Computer science*), and finding all articles that fall under that category. We traversed through all subcategories of the matched category down to the 4th level, and considered all articles in those categories to be within the scope of the WikiProject¹. We also considered an alternative measure, *claimed scope*, by including articles whose talk pages link to the WikiProject. Creating such links from talk pages is the established way for projects to claim articles into their scope. However, because claiming articles is a manual process, the claimed scope grows inconsistently over time for different projects. Overall this measure appears to be a serious underestimate of scope. For instance, WikiProject C++, created in August 2004, had only claimed three articles as of Dec 2007. Further, projects frequently work on articles that they do not bother to claim. As a result, in the end we estimated the scope of projects using categories, which seemed a more reliable measure to compare breadth across projects.

We used historical edits of a project’s member list (see Participants in Figure 2) to identify members of each WikiProject. The member list is usually on the project’s main page or one of its subpages. In Wikipedia, any editor can join a project by adding his or her username to the member list, and later leave the project by removing the username from the list.

¹ We only traverse to the 4th level because the Wikipedia subcategory structure is not a hierarchy, but a more general graph structure. Through experimentation we discovered that traversing four levels covers most meaningful subcategories without reaching many problematic links.

We created a longitudinal dataset in which each observation records the characteristics, composition, and outcomes of each project for each quarter in its life span, where each quarter includes 90 days. Take WikiProject on Computer Science as an example, the first record measures the first 90-day period following its date of creation, the second measures the second 90-day period, and so on, until the end of the dump at the end of year 2007. There are altogether 3899 project quarters from 683 projects in our dataset.

Using a longitudinal approach enables us to temporally separate our independent and dependent variables and thus better understand their causal relationships. We have proposed that diversity levels affect member withdrawal. Withdrawal in turn changes the member composition of the project, which affects diversity levels. As a result, a cross-sectional analysis that does not contain temporal information will not be able to distinguish whether it is diversity changes that cause membership changes or the reverse. To resolve this problem, in our analysis for each quarter we measure diversity at the beginning of the quarter and member withdrawal over that quarter, so the change in composition of the project only affects the diversity measure for the next quarter but not the current one.

Independent Variables

Tenure Disparity: We measured an editor’s tenure by how long the editor has been a member of Wikipedia, that is, the number of days elapsed from a member’s first edit in Wikipedia to the end of a quarter (*Wikipedia Tenure*). We explored two alternative measures of tenure: how long an editor has been a member of a specific WikiProject (*WikiProject Tenure*) and how many edits the editor has performed (*Total Edits*).

We preferred *Wikipedia Tenure* over *WikiProject Tenure* because experience in Wikipedia as a whole transfers readily to projects. For instance, editors who, over time, learn how Wikipedia policies work can apply those policies in discussions within any WikiProject. We preferred *Wikipedia Tenure* over *Total Edits* because it does a better job of capturing the experience that lurkers gain by observing the interactions of other editors [23]. Our results remain qualitatively the same using either the *WikiProject Tenure* or *Total Edits* measures. We report our findings using *Wikipedia Tenure* as our tenure measure.

We measured tenure diversity using the coefficient of variation of *Wikipedia Tenure* (simply called “tenure” from here forward) of all project members. We chose coefficient of variation as an appropriate measure for tenure because it has been widely used in past research to measure tenure diversity [3], and because the likely sources of benefit and conflict from tenure are from the spread of members’ experience relative to the project’s mean tenure [12]. If we denote each member’s tenure as T_i and the mean tenure over n members as T_{mean} , their coefficient of variation can be calculated using the following formula [12]:

$$[\sum (T_i - T_{mean})^2 / n]^{1/2} / T_{mean}$$

Interest Variety: We considered the edits that an editor contributed to different topic areas as a proxy to measure their interests. For example, an editor primarily editing in Philosophy is assumed to be more interested in Philosophy than average editors. The challenge is to differentiate editing behaviors driven by member interest from incidental edits – such as fixing spelling errors that an editor accidentally encountered.

We first compose eight primary interest areas from the categorical index portal of Wikipedia: *Arts, Geography, Health, History, Science, People, Philosophy* and *Religion*. Then following a procedure similar to [16], we assign a Wikipedia article to an interest area if the article is closest to the top level category of the interest area in the subcategory structure. For example, article *Computer science* is in category *Computer science*, which is a 3rd level subcategory of *Science* (through *Science, Scientific disciplines, Applied sciences* to *Computer science*), closer than all other top level categories, and is thus assigned to be in the interest area of *Science*. An article can be assigned to more than one interest area if it is equally close to several top level categories.

We classified a Wikipedia editor as being interested in an area if he has done more than 10 edits on articles in that area and those edits comprise at least 40% of all edits he has made in Wikipedia. Of all editors in our dataset, we were able to assign 24% of them into no interest area, 45% into one interest area, 22% into two interest areas, and only 9% into three or more areas. Our results remain qualitatively the same using 20 edits or 20% as cutoffs.

Once we classified all editors into the eight areas, we measured interest diversity of a project using the Blau index because interest is a categorical variable [12]. For a particular project quarter, we calculated interest variety by counting the number of project members in each interest area. If we denote the percentage of project members in each area as P_i , the Blau index can be calculated using the following formula [12]:

$$1 - \sum P_i^2$$

Dependent Variables

Amount of Work: We measured amount of work by the number of edits done by members of a WikiProject on articles within the scope of the project. We repeated our analyses with an alternative measure: the number of words added by members into articles within the scope, and our main results remained unaffected.

Member Withdrawal: We measured member withdrawal by the number of people who were active members in the previous quarter but removed their names from the member list, or stopped contributing by the end of the current quarter. We considered a member to be active for a quarter

Descriptive Statistics									
	Mean	Std Dev	1% Percentile	25% Percentile	50% Percentile	75% Percentile	99% Percentile		
1. Amount of work	1492	2559	0	107	556	1756	13174		
2. Member withdrawal	3.972	5.535	0	0	1	3	40		
3. Quarter	3.186	2.805	0	1	3	5	11		
4. Project size	23.31	31.64	0.5	5.0	12.0	27.5	163.4		
5. Project scope	20476	68129	16.5	402.0	2169.0	10086.0	399311.1		
6. Project creation	16.51	3.187	10.3	14.2	17.1	19.0	22.2		
7. Level of controversy	1.109	0.414	0.24	0.88	1.07	1.31	2.28		
8. Mean tenure	503.4	174.5	0	377.0	487.8	601.6	999.2		
9. Tenure disparity	0.556	0.191	0	0.457	0.570	0.675	1.352		
10. Interest variety	0.640	0.188	0	0.592	0.693	0.762	0.851		
Correlations									
Variables	1	2	3	4	5	6	7	8	9
1. Amount of work									
2. Member withdrawal	0.58 ***								
3. Quarter	0.22 ***	0.24 ***							
4. Project size	0.64 ***	0.87 ***	0.41 ***						
5. Project scope	0.20 ***	-0.02	0.13 ***	0.01					
6. Project creation	-0.12 ***	-0.03 +	-0.57 ***	-0.14 ***	-0.09 ***				
7. Level of controversy	0.01	0.02	0.03 +	0.04 *	0.05 **	-0.08 ***			
8. Mean tenure	-0.03 *	-0.11 ***	0.49 ***	0.04 *	0.05 **	-0.12 ***	0.00		
9. Tenure disparity	0.20 ***	0.26 ***	-0.08 ***	0.16 ***	-0.02	0.11 ***	-0.01	-0.29 ***	

Table 1: Descriptive Statistics and Correlations of Variables

We use the following notation in tables to represent p-values: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$

if the person had at least one edit to: an article within the project scope, the talk page of such an article, any of the project organization pages, or the user pages or user talk page of any other project member during that quarter.

Controlled Variables

Quarter Index: The index of time within the project measured in quarters (90-day periods); starting with quarter 0 from the moment the project is created, until the last full quarter before the end of 2007.

Project Size: Measured as the number of project members during the current quarter.

Project Scope: Measured as the number of articles falling under the project scope during the current quarter. Scope, determined by the structure of the category hierarchy in Wikipedia, changes slowly. Most changes involve new articles being added to the project's scope.

Project Creation Quarter: Measured as the number of quarters (90-day periods) from Jan 2002 to the date the project is created. Larger number means the project was created later. Due to the sheer increase in the number of projects and participating editors over time, we suspect that projects created later may face a different environment than projects created earlier. This variable controls for that.

Level of Controversy: Measured as the percentage of reverts in all edits on articles inside the project scope during the current quarter, normalized by the overall percentage of reverts in Wikipedia over the same period of time. Reverting other editors' edits is one of the common expressions of conflict and controversy within Wikipedia [15]. We use this measure to control for effects caused by controversy that is inherent in project subject matters and separate them from effects caused by diversity among the individuals in the project.

Mean Tenure: Measured as the mean of the tenures of all project members during the current quarter. We measure tenure as how long a member has been a Wikipedia editor.

Analysis Approach

Our data are nested in nature – quarters nested within projects – so we employed Hierarchical Linear Models [7] for our analysis. HLM is an advanced form of linear regression that allows us to examine the effects of independent variables (tenure disparity and interest variety) on dependent variables (amount of work done and member withdrawal), taking into account potential correlations across quarters that are nested within the same project.

Our hypotheses concerning diversity were examined in an incremental approach, where we first specified a null HLM model with only intercept and the quarter index. We then developed a second model with only control variables added, as a baseline against which the later models can be measured. Finally, we developed two models with our independent variables added: one with just the linear version of each variable, and one that also includes the quadratic terms for each, so that we can test whether each set of predictor variables provides significant incremental prediction of the dependent variables [17].

Before estimating the models we log transformed project size and project scope to make them more normal (we used a base 2 logarithm for visualization convenience). Interest variety is negatively skewed so we first used (1 – interest variety) to change it into a positively skewed distribution, and then log transformed it to make it more normal. We then reflected it back so the ordering of its values is consistent with the original variable. We also performed grand mean centering for all predictor variables to reduce multicollinearity between main effects and quadratic terms. We then estimated our HLM models, using maximum

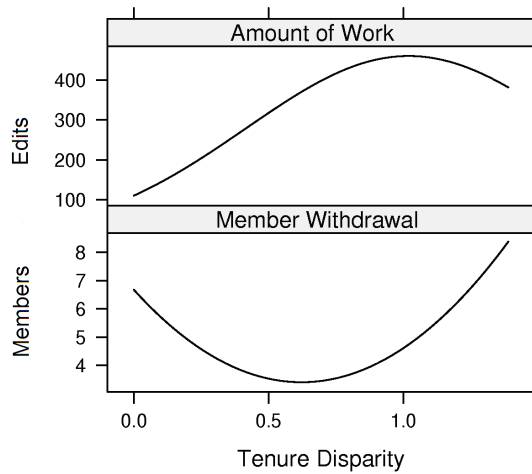


Figure 3: Effects of Tenure Disparity (Hypothesis 1&3).
The log transform for “Amount of Work” was reversed.

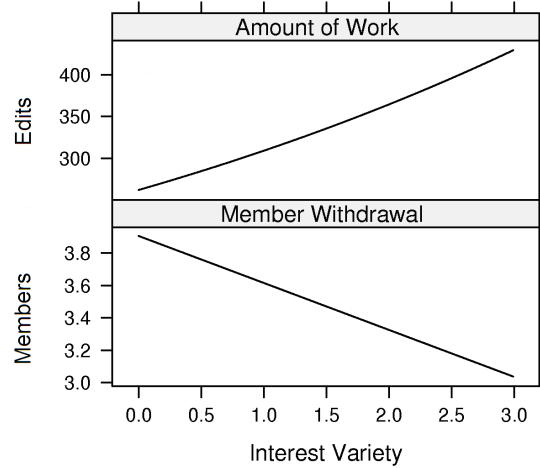


Figure 4: Effects of Interest Variety (Hypothesis 2&4).
The log transform for “Amount of Work” was reversed.

likelihood estimation, random intercepts, and unstructured covariance structure. We conducted a number of diagnosis analyses [4]. After grand mean centering, multicollinearity is low as reflected in condition indices with values of less than 10 for all models. Residual analysis identified two projects as outliers for having more than 300 members or 30,000 edits in a quarter, and as a result we excluded 22 project quarters from our dataset.

RESULTS

Table 1 presents the descriptive statistics and correlations of variables used in the analysis before log transformation and grand mean centering. As one can see, the amount of work, member withdrawal, project scope and size are of sufficient variation and are of heavily skewed distributions. By including quarter in the model we control for the changes in overall level of activity within Wikipedia over time.

Predicting Amount of Work Done

Table 2 presents the first four models, predicting the amount of work done by project members each quarter. Model 1 is the null model, Model 2 adds the four controlled variables, Model 3 adds the linear terms for the independent variables, and Model 4 also adds the quadratic terms for the independent variables. We assessed the significance of the

incremental variance explained by examining the difference between the deviance statistics (-2 log likelihood) for each pair of model, similar to assessing changes in *R-square* in traditional regression analysis. Comparison of the deviance statistics suggests that Model 4 explains the most variance and has a better fit than the simpler models, so we interpret its results to test our hypotheses.

All four controlled variables have significant effects on amount of work done in a project ($p < .001$). The positive coefficient 0.708 of project size suggests that when holding other variables constant, increasing the number of members in a project results in more edits being done in a quarter. Similarly, more work would be done if the project is of a larger scope, if the project is created earlier in time, or if the project is dealing with a less controversial topic. The negative coefficient of the quarter variable suggests that when controlling for other factors, projects in general accomplish less work each quarter over time.

Model 4 reveals a positive main effect of tenure disparity ($\beta = 1.837, p < .001$) and a negative effect of its quadratic term ($\beta = -1.978, p < .001$). As shown in the top figure of Figure 3, increased tenure disparity increases the amount of work done by group members in a quarter with diminishing

	Model 1			Model 2			Model 3			Model 4		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P
Intercept	8.325	.1165	***	8.273	.0750	***	8.301	.0759	***	8.419	.0850	***
Quarter	0.085	.0243	***	-0.326	.0311	***	-0.288	.0308	***	-0.291	.0307	***
Project size (log2)				0.911	.0361	***	0.733	.0400	***	0.708	.0406	***
Project scope (log2)				0.409	.0209	***	0.384	.0209	***	0.388	.0209	***
Project creation quarter				-0.060	.0241	***	-0.066	.0238	**	-0.070	.0238	**
Level of controversy				-0.406	.1086	***	-0.394	.1067	***	-0.383	.1065	***
Mean tenure				-0.001	.0003	**	0.000	.0004		0.000	.0004	
Tenure disparity							1.985	.2272	***	1.837	.2294	***
Interest variety							0.338	.0978	***	0.238	.1165	*
Tenure disparity squared										-1.978	.5786	***
Interest variety squared										-0.110	.0936	
Deviance (-2 Log likelihood)	15916.6			14800.0			14573.0			14561.0		
Deviance difference (Δ Dev)				1116.6 ***			227.0 ***			12.0 **		
N	3877			3877			3877			3877		

Table 2: HLM Results of Predicting Amount of Work Done by Project Members (log2)

returns. After tenure disparity goes above 1 (tenure disparity ranges between 0 and 1.5), increase in tenure disparity does not result in more work done. In fact, extremely high tenure disparity decreases the amount of work done in a project. **The results support Hypothesis 1.**

Model 4 reveals a positive main effect of interest variety ($\beta = 0.238, p < .05$) and no effect of its quadratic term. As shown in the top figure of Figure 4, increased interest variety increases the amount of work done by group members in a linear fashion. **The results provide partial support to Hypothesis 2** supporting the main effect but rejecting the diminishing returns part of the hypothesis.

Predicting Member Withdrawal

Table 3 presents the four models that predict the number of members withdrawing from project activities each quarter. The first three models are parallel to the first three models in Table 2. Introducing a quadratic term for interest variety to Model 3 completely changes the result for its linear term, without explaining significantly more variance. Therefore, in Model 4 we include only a quadratic term for tenure disparity, and not for interest variety. Deviance statistics suggests that Model 4 has a better fit than simpler models. As a result, we use the results of Model 4 to test our hypotheses.

As shown in Table 3, on average, projects with more members, a smaller scope, or projects that were created later in Wikipedia history are more likely to have members withdraw from contributing to project effort. Interestingly, level of controversy of the project topic does not have any significant impact on withdrawal.

Model 4 in Table 3 reveals a negative main effect of tenure disparity ($\beta = -1.128, p < .01$) and a positive effect of its quadratic term ($\beta = 8.464, p < .001$) on member withdrawal. As shown in the bottom part of Figure 3, increased tenure disparity decreases member withdrawal in a quarter, with a curvilinear relationship. Projects with a moderate level of tenure diversity seem to have the lowest level of member withdrawal. Either extremely low or extremely high tenure diversity doubles the number of members who withdraw from group activities. **The results support Hypothesis 3.**

Model 4 reveals a negative main effect of interest variety ($\beta = -0.290, p < .05$). As shown in the bottom part of Figure 4, increased interest variety decreases member withdrawal.

The results provide no support to Hypothesis 4. These results are surprising. Involving members with a wide variety of interests in WikiProjects makes members less likely, rather than more likely, to withdraw.

DISCUSSION

Understanding Tenure Diversity Online

We have found generally that increasing tenure diversity led to better group outcomes – but that at very high levels tenure diversity predicted negative group outcomes.

One reason that increased tenure diversity enhances group performance may be that increased tenure diversity leads to better task distribution. Although active editors do not necessarily do more or better work as their tenure increases [22], editors do develop themselves over time. As a result, editors with different tenure are interested in performing different kinds of tasks, such as coordinating with other members, assessing article quality, writing new content, and proof-reading [6]. In one instance we observed two editors U1 and U2 who both joined WikiProject Croatia in early 2009. U1 was a new editor who began to edit Wikipedia in Feb 2009. Most of U1's efforts were focused on editing and discussing articles on Croatia. U2 has been an editor since 2005, and had prior experience with a number of other projects. Most of U2's efforts were spent in recruiting and coordinating for WikiProject Croatia. The project benefited by having diverse contributions from the two editors.

On the other hand, increased tenure diversity may cause conflict in a WikiProject, reducing performance. Tenure in Wikipedia is sometimes viewed as conferring social status. Old-timers and newcomers may refer disparagingly to each other, resulting in low productivity and high withdrawal [20, 21]. For example, in 2007 on WikiProject Intelligent Design, a new editor U3 complained about bias in the project, and argued that an “inappropriate disclaimer” should be removed from the beginning of many articles in the project's scope. A senior editor U4 then criticized U3 for not understanding several key Wikipedia policies and

	Model 1			Model 2			Model 3			Model 4		
	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P
Intercept	4.507	.2112	***	3.822	.1621	***	3.748	.1583	***	3.436	.1572	***
Quarter	0.729	.0582	***	-0.014	.0660		-0.094	.0670		-0.103	.0659	
Project size (log2)				1.548	.0581	***	1.799	.0724	***	1.960	.0708	***
Project scope (log2)				-0.097	.0277	***	-0.083	.0276	**	-0.092	.0259	***
Project creation quarter				0.091	.0306	**	0.111	.0304	***	0.122	.0286	***
Level of controversy				0.199	.1761		0.229	.1756		0.216	.1676	
Mean tenure				-0.002	.0005	**	-0.002	.0005	***	-0.002	.0005	***
Tenure disparity							-1.474	.3949	***	-1.128	.3852	**
Interest variety							-0.390	.1496	**	-0.290	.1430	*
Tenure disparity squared										8.464	.9785	***
Interest variety squared										-	-	
Deviance (-2 Log likelihood)	16943.2			16326.0			16223.8			16153.3		
Deviance difference (Δ Dev)				617.2 ***			102.2 ***			70.5 ***		
N	3877			3877			3877			3877		

Table 3: HLM Results of Predicting Number of Members Withdrawing from Project Activity

for being disruptive. In the ensuing argument, U4 continually cited policies, which he asserted that U3 did not understand because of his lack of experience, while U3 argued that the policies were not relevant, and that U4 was just listing sets of policies in place of argument. In the end, U3 left the project and added a sentence on his user page – “This user is sick of the bullshit on wikipedia”.

Conflicts like the above can be particularly damaging to online volunteer groups, because online volunteer groups often have low exit barriers, and may be less cohesive with lower member identification compared to other work groups. As a result, when members get frustrated, they are more likely to leave or stop contributing to group effort. For instance, in the case described above it is unlikely that U3 will ever come back and participate in discussions within the project, while in offline work groups people might be more willing to work through conflict because of higher member identification and higher exit barriers. Carrying this line of reasoning one step further, Drenner et al. found that increasing the barrier to entry for an online group led to greater participation and reduced withdrawal among group members who passed through the barrier [10].

High tenure diversity increasing conflict is consistent with prior research on offline groups. However, this consistency may be surprising to some people, because previous theories and empirical results have suggested that status inequalities can be less salient in online groups and thus less likely to cause conflict (see [19] for a review). This may be particularly true in online volunteer groups like Wikipedia where most editors participate on equal footing, without much difference in privilege or rank. However, Kriplean et al. found that more experienced Wikipedia editors use their ability to cite policy as a way to control the outcome of debates [18]. Likewise, our findings suggest that people in online volunteer groups still categorize their peers based on experience and treat them differently.

Understanding Interest Diversity Online

In contrast to the effects of tenure diversity, interest diversity shows only positive effects in our study (see Figure 4). Offline studies suggest that this positive effect is due to the information integration process: project members with different interests are likely to possess knowledge in different subject matter, and thus be able to contribute different types of unique information to the project. For example, in August 2008 on WikiProject China, separate discussion items are created for articles about a historical war in China, a government program associated with the ongoing Olympics, the depiction of Buddhist deities in China, and a Mongolian secessionist movement. Improving those articles required diverse members who could contribute in history, sports, religion and politics.

The lack of negative effects of interest variety is more surprising as it is inconsistent with offline studies. Note that we cannot simply attribute this inconsistency to differences between online and offline groups, because tenure diversity

in our study does exhibit the negative effects predicted by the offline studies. In searching for an explanation, we speculate that the key factor is the differential visibility of tenure and interest on Wikipedia. The visibility of an attribute affects how difficult it is to categorize other people based on that attribute, and how likely conflicts can arise from such categorizations.

Neither tenure nor interest areas of an editor are directly visible in Wikipedia discussions. However, an editor can learn about the tenure of another editor with relatively little effort. Clicking the username of the other editor opens the user page, where some editors, especially more senior ones, disclose when they joined Wikipedia, along with other information about themselves. If an editor does not disclose her tenure on her user page, it is easy to check her edit history to find out when she first edited Wikipedia.

In contrast, learning the interest areas of an editor is more challenging – there are few visually salient indications of user interests that can be picked up on at a glance like in the case of tenure. While the interest of active editors can often be inferred by analyzing their user pages, or by examining the set of pages in their edit history, interpreting cues of interest on user pages such as userboxes and categorizing edited pages into interest areas requires additional effort and knowledge that many editors are unlikely to exert or to possess.

Because interest areas are less visible than tenure, it is more difficult for editors to categorize others by interest areas than by tenure. As a result, compared to differences in tenure, differences in interest area are less likely to lead to social categorization and the resulting conflict, and in turn less likely to negatively affect group outcomes.

Similarly, because differences in interest areas are less visible than many types of diversity in offline groups, the above discussion may explain the why our results disagree with prior results from offline studies. We speculate that such phenomenon will appear in other online groups, such as open source projects, as well. Future research should explore whether differences that are more easily visible to group members lead to greater negative effects on outcome.

CONCLUSION

We have examined the effects of group diversity on group productivity and member withdrawal in WikiProjects. We found that increased diversity in experience with Wikipedia increases productivity and decreases withdrawal – up to a point. Beyond that point, productivity remains high, but members are more likely to withdraw. In contrast, no such diminishing returns were observed for differences in member interest in projects, which increases productivity and decreases withdrawal in a linear fashion.

There are rich directions for future research. For instance, instead of analyzing archived data and observing temporal changes in natural environments as we did, future studies can manipulate group diversity and observe the outcomes,

to demonstrate causality more directly. Future studies can also examine the collaboration processes to gain a deeper understanding of why and how diversity has the effects we found. Research in these directions will provide insights to further harvest the “wisdom of crowds”, increasing the long-term productivity of online volunteer groups.

ACKNOWLEDGEMENTS

Thanks to Wikimedia Foundation for providing the data. Thanks to members of GroupLens Research Lab at Dept of Computer Science and OB Research Group at the Carlson School of Management for their feedback and support. This work is supported by National Science Foundation Grant IIS-0808692, IIS 0729344 and IIS 0534939.

REFERENCES

1. Ancona, D.G., & Caldwell, D.F. 1992. Demography and design: predictors of new product team performance. *Organization Science*, 3, 321-41.
2. Bayazit, M. & Mannix, E.A. 2003. Should I stay or should I go? Predicting team members' intent to remain in the team. *Small Group Research* 34, 290-321.
3. Bedeian, A., & Mossholder, K. 2000. On the use of the coefficient of variation as a measure of diversity. *Organizational Research Methods* 3(3), 285-297.
4. Belsley, D., Kuh, E., & Welsch, R.. 1980 *Regression Diagnostics: Identifying Influential Data and Collinearity*. New York: Wiley.
5. Bhappu A.D., Griffith T.L., & Northcraft G.B. 1997. Media effects and communication bias in diverse groups. *Organizational Behavior and Human Decision Processes*, 70(3), 199-205.
6. Bryant, S. L., Forte, A., and Bruckman, A. 2005. Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. *GROUP* 2005.
7. Bryk, A.S., & Raudenbush, S.W. 1992. *Hierarchical linear models for social and behavioural research: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.
8. Bunderson J.S., & Van Der Vegt, G.S. 2005. Learning and performance in multi-functional teams: The importance of collective team identification. *Academy of Management Journal*, 48, 532-547.
9. Dahlin K.B., Weingart L.R., & Hinds P.J. 2005. Team diversity and information use. *Academy of Management Journal*, 48, 1107-1123.
10. Drenner S., Sen S., & Terveen L.G. 2008. Crafting the initial user experience to achieve community goals. *RecSys 2008*, ACM Press, 187-194.
11. Giles, G. 2005. Internet encyclopedias go head to head. *Nature*, 438, 900-901.
12. Harrison, D.A., & Klein, K.J. 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*, 32, 1199-1228.
13. Horwitz, S.K. & Horwitz, I.B. 2007. The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, 33, 987-1015.
14. Kittur, A., & Kraut, R.E. 2008. Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. *CSCW* 2008.
15. Kittur, A., Suh, B., Pendleton, B.A., & Chi., E. 2007. He says, she says: Conflict and coordination in Wikipedia. *CHI* 2007.
16. Kittur, A., Suh, B., & Chi, E. 2009. What's in Wikipedia? Mapping topics and conflict using collaboratively annotated category links. *CHI* 2009.
17. Kreft, I., & de Leeuw, J. 1998. *Introducing multilevel modeling*. Sage, Thousand Oaks, CA.
18. Kriplean, T., Beschastnikh, I., McDonald, D.W., & Golder, S.A. 2007. Community, consensus, coercion, control: CS*W or how policy mediates mass participation. *GROUP* 2007.
19. Martins, L.L., Gilson, L.L, & Maynard M.T. 2004. Virtual teams: What do we know and where do we go from here? *Journal of Management*, 30(6), 805-835.
20. Milliken F.J., Martins L.L. 1996. Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review*, 21, 402-433.
21. O'Reilly, C.A., Caldwell, D.F., & Barnett, W.P. 1989. Work group demography, social integration, and turnover. *Administrative Science Quarterly*, 34, 21-37.
22. Panciera, K., Halfaker, A., and Terveen, L. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. *GROUP* 2009.
23. Preece, J., Nonnecke, B., & Andrews, D. 2004. The top 5 reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior*, 20(2), 201-223.
24. Priedhorsky, R., Chen, J., Lam, S., Panciera, K., Terveen, L., & Riedl, J. 2007. Creating, destroying, and restoring value in Wikipedia. *Proc. GROUP* 2007.
25. Surowiecki, J. 2004. *The wisdom of crowds*. Random House, Inc..
26. Van Knippenberg, D., De Dreu, C.K.W. & Homan, A.C. 2004. Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology* 89, 1008-1022.
27. Van Knippenberg, D., & Schippers, M.C. 2007. Work group diversity. *Annual Review of Psychology*, 58, 515-541.
28. Williams, K.Y. & O'Reilly, C.A. 1998. Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior*, 20, 77-140