Understanding the "Few that Matter" in Online Social Production Communities: The Case of Wikipedia

Mihai Grigore	Bernadetta Tarigan	Juliana Sutanto	Chrysanthos Dellarocas
Swiss Federal Institute of Technology Zurich	Swiss Federal Institute of Technology Zurich	Swiss Federal Institute of Technology Zurich	Boston University
mgrigore@ethz.ch	btarigan@ethz.ch	jsutanto@ethz.ch	dell@bu.edu

ABSTRACT

Several past studies have commented on the uneven distribution of contributions in online social production communities while at the same time highlighting the successful end products of many such communities. These two seemingly paradoxical situations are made possible through smaller groups of highly devoted volunteers who act as catalysts in organizing and maintaining community outputs. These volunteers have been referred to as "knowledge janitors." There is currently limited understanding of how the group composition and interaction patterns of knowledge janitors affect social production quality outcomes. This study provides answers to these questions in the context of Wikipedia. By analyzing 11,359 changes in Wikipedia article quality, we found that cohesiveness, diversity, and equal distribution of communication turn-taking of an article's janitors increase the likelihood of that article's quality improvement. These main findings are further refined by considering how the main effects differ at different development stages of an article. The study's contributions to research and implications to practice are discussed.

Keywords: online social production, Wikipedia, Wikipedia janitors, group composition, group interaction, article quality

1. INTRODUCTION

Online social production communities have become an increasingly viable and popular way to create information products that are often of relatively high quality (Giles, 2005; Tapscott & Williams, 2006). Open source software such as Linux and R allow people to install and use the software with no cost as well as to contribute back to the improvement of the software. Peer recommendations such as Amazon's book reviews and YouTube's ratings rely on community submissions and social monitoring to produce reliable product reviews. Wikipedia, a peer content production, has become one of the world's most popular sources of information; the quality of Wikipedia entries has repeatedly been found to be on par to other traditionally organized processes, carried out by professional editors over several years, such as the Encyclopedia Britannica (Giles, 2005; Tapscott & Williams, 2006).

While it is common knowledge that contributions in online social production communities follow a long tail distribution (Collier & Kraut, 2012; Zhu et al., 2012), the ways in which the most highly devoted volunteers act as catalysts in the development of high quality output are still not well understood. This study will address this research gap in the context of Wikipedia. A WikiProject is a group of contributors who want to work together as a team to improve Wikipedia. These groups often focus on a specific topic area (for example, women's history) or a specific kind of task (for example, checking newly created pages). The English Wikipedia currently has about 2,000 WikiProjects.[1] Members of Wikiprojects are highly devoted volunteers who act as catalysts in the development of high quality articles, i.e., they are able to organize and keep Wikipedia articles stable. Following (Sundin, 2011), we refer to those individuals as *Wikipedia janitors*, indicating their essential role in the development and organization of content in Wikipedia (Chen et al., 2010; Choi et al., 2010; Sundin, 2011). Focusing on Wikipedia janitors, this study aims to provide answers to the following

[1]. See http://en.wikipedia.org/wiki/Wikipedia:WikiProject

questions: What are the relationships between the *composition* of Wikipedia janitor groups (diversity, cohesiveness) and their *interactions* (communication turn-taking) on the article quality improvement?

The contributions of this study are twofold. First, this study complements the extant literature on online social production in general and Wikipedia in particular, by focusing on Wikipedia janitors rather than following the common approach of looking at the entire Wikipedia community (Kittur & Kraut, 2008; Ransbotham & Kane, 2011; Zhu et al., 2012). We unveil how the composition of Wikipedia janitor groups in a particular article and the way they interact with one another affect the likelihood of that article's quality improvement. Second, as the virtual, self-organizing workplace constantly evolves towards more spontaneous and decentralized forms of collaboration, with the growing tendency of organizations to outsource complex tasks to large masses of workers via distributed labor networks using limited or no monetary incentives (Downs et al., 2010; Aniket Kittur et al., 2008; Ross et al., 2010), this study offers insights for organizations on how to achieve high quality outcomes in such online distributed labor networks.

2. THEORETCIAL BACKGROUND AND RESEARCH HYPOTHESES

There are numerous research articles dealing with Wikipedia, as it is an excellent resource to examine online social production. There are four specific reasons for this. First, the underlying wiki technology records the full editing activity and all editors' social interactions for each article. Second, Wikipedia's articles can only be edited using the Wikipedia platform, allowing researchers to have a complete editing and social interaction history of each article. Third, any Internet user can contribute content to the articles, allowing researchers to examine group interactions in an uncontrolled setting. Fourth, there are formal guidelines and mechanisms for assessing and evaluating quality ratings of Wikipedia's

articles, allowing researchers to have a somewhat objective measurement of group performance outcomes.

In Wikipedia, the contributions and involvement of editors may drastically vary over time (Ransbotham & Kane, 2011), with some editors moving from peripheral participation to full involvement (Bryant et al., 2005), while others abandoning the article (Brandes & Lerner, 2009; Halfaker et al., 2011). In contrast to the plethora of studies on the distribution of editor contributions, a recent sociotechnical investigation of Wikipedia proposes a change of research focus towards analyzing the importance of the "practices of fixing, erasing, voting, changing, proof-reading, and, […] inserting references to external sources" for the construction of articles in Wikipedia (Sundin, 2011) – which are the main activities of Wikiproject members. Joining a Wikiproject can be considered as a proxy of an editor's interest for or familiarity with a specific sub-domain, as well as of the willingness to contribute to the Wikipedia content in that specific sub-domain (Chen et al., 2010; Choi et al., 2010). In this article we refer to Wikipedia editors who joined at least one Wikiproject as *Wikipedia janitors*. We analyze how these Wikipedia janitors, through their essential roles in developing and organizing content in Wikipedia, improve the quality of articles.

Previous research claimed that the success of social production in Wikipedia depends on the ways in which peers self-organize and coordinate their work, e.g. by directly communicating with each other and by concentrating most of the editing among a subset of the editors (Arazy & Nov, 2010). Besides interaction patterns, the characteristics of those who participate in the group work appear to be important in increasing the quality of group outcomes (Butler et al., 2007; Liu & Ram, 2011). Because of the importance of Wikipedia janitors to the development of article content, we focus our analysis on the effect of group composition and interaction patterns of Wikipedia janitors on the change in article quality over time.

When members of a group coordinate their work through communication, their individual turn-taking dynamics may facilitate the formation of a common ground (Clark, 2005). *Turn-taking* is the set of communication practices by which collaboration is achieved in and through conversational turns (Sacks et al., 1974). Previous studies showed that articles in which a few editors do most of the work are also the ones in which these editors talk to each other on the article talk page (Aniket Kittur & Kraut, 2010). Kittur and Kraut (2008) found that coordination through communication is more efficient when there are few editors. In addition to its coordination role (Butler et al., 2007; Liu & Ram, 2011; Panciera et al., 2009), turn-taking may also be important for knowledge exchange (Woolley et al., 2010). Through laboratory experiments of small size (2-5 members) face-to-face groups, Woolley et al. (2010) showed that *equality in distribution of turn-taking* is positively correlated with group outcome across a variety of tasks. "In other words, groups where a few people dominated the conversation were less collectively intelligent than those with a more equal distribution of conversational turn-taking" (Woolley et al., 2010). The reason being that collective intelligence is a group intelligence that emerges from the knowledge exchange and consensus building of the individual members that is generally superior to simply aggregating the individual members' knowledge. If the discussion is directed by a single individual who imposes a consensus view on the others, then that perspective would not be more powerful than the perspective of the particular individual (Heylighen, 2013). All this implies that an equal distribution of communication turn-taking between the Wikipedia janitors should result in better group performance in terms of higher article quality. Accordingly, we hypothesize:

H1: *There is a positive relationship between the equality of distribution of communication turn-taking between Wikipedia janitors and article quality.*

Besides interaction patterns, the characteristics of the individuals who participate in group work appear to be important in increasing the quality of group outcomes (Butler et al., 2007;

Liu & Ram, 2011). Diversity in group composition has been proposed as a requirement for a group to exhibit 'wisdom of crowds' effects (Surowiecki & Silverman, 2007). The literature concerning group diversity suggests that it may be either beneficial or detrimental in terms of group outcomes. On the one hand, more diversity in knowledge and experience helps group members to avoid biases and overlooking certain aspects; this can lead to improved outcomes. Heterogeneous groups appear to perform well because they have a relatively broad range of information, experiences, and perceptions to draw from. On the other hand, group heterogeneity and differences among individuals may result in conflict and diminished performance (Aral et al., 2008). Social categorization theory suggests that differences between peers may generate tensions and conflicts among them, which may, in turn, negatively influence group outcomes (Van Knippenberg et al., 2004). Recent research on the effect of group diversity on performance has established a curvilinear relationship, in the sense that moderate diversity correlates with higher performance, whereas extremes of too little or too much diversity are detrimental to group performance (Aral et al., 2008). Since group diversity appears to be a double-edged sword, we hypothesize:

H2: *There is an inverted U curvilinear relationship between the diversity of Wikipedia janitors and article quality.*

Previous literature on group processes and outcomes acknowledges that, when group members stay and work together over a period of time, they are able to develop a common ground, unspoken expectations and shared mental models of the task to be accomplished (Aral et al., 2008; Chen et al., 2010). This is especially important for key group members as turnover in the key roles of a work group may negatively influence the way team members interact or coordinate their work (Humphrey et al., 2009; Ransbotham & Kane, 2011). According to the theory of group cohesiveness, the stick-togetherness of group members is positively associated with group performance (Chansler et al., 2003; Festinger et al., 1950).

However, a recent paper argues otherwise. Examining close to 200 groups whose members worked together for a few hours on predefined tasks, Woolley et al. (2010) found that group cohesion was not correlated with group performance. We argue that, in online social production such as Wikipedia, where group size fluctuates dynamically as members come and leave at their own will (Brandes & Lerner, 2009; Bryant et al., 2005; Halfaker et al., 2011; Ransbotham & Kane, 2011), it is important that at least the Wikipedia janitors stick together to be able to improve the article quality over time (Zhu et al., 2012). Hence, we hypothesize:

H3: *There is a positive relationship between the cohesiveness of Wikipedia janitors and article quality.*

3. RESEARCH METHODOLOGY

To test the above hypotheses, we conduct a longitudinal analysis in a similar way to Kittur et al. (2008), i.e., for each quality assessment of an article, we consider the period between the previous and the current change in quality. A graphical representation of the considered time span of analysis is presented in Figure 1. Note that our analysis takes into account only those articles that receive at least two quality assignments during their life span. The instantiations of all the subsequent constructs in our model are computed relative to this time span.

[Insert Figure 1 about here]

3.1 Dependent Variable

Change in Quality. As dependent variable we indicate whether, following a quality assessment, an article was either promoted (to a superior quality class) or demoted (to a lower quality class). The Wikipedia community has developed formal guidelines and mechanisms for assessing and evaluating quality ratings of its articles in an inter-subjective manner, into a large spectrum of *quality ratings*. The ratings vary from a very low quality to the highest quality and are termed (in increasing order): ‘Stub’, ‘Start’, ‘C-class’, ‘B-class’, ‘Good

articles', 'A-class', and 'Featured articles'. These ratings are consistent with Wang and Strong's multidimensional definition of data quality (Liu & Ram, 2011; Wang & Strong, 1996). An article is promoted to a higher rating or demoted to a lower rating only after an official assessment of peer reviewers. In order to maintain objectivity and neutrality, the assessments of good and featured articles are made by Wikipedians who did not participate in heavy or repeated edits of the article proposed for assessments (Liu & Ram, 2011). Previous studies have shown that Wikipedia's internal quality ratings and the ratings from external raters are significantly correlated (Spearman's $\lambda = 0.54$, $p < 0.001$) (Kittur & Kraut, 2008). Hence, even though they may not be completely objective and neutral, Wikipedia's quality ratings constitute a valid *proxy* for *article quality* (Liu & Ram, 2011). This proxy is used to compute the dependent variable in this study.

3.2 Dataset

The dataset we use in this study is extracted from the June 2011 dump of English Wikipedia. The dataset includes full texts of Wikipedia pages and their complete edit history from the beginning of Wikipedia. Since our intention is to examine the influential factors of Wikipedia article quality, we selected an equal proportion of articles of different Wikipedia-designated quality levels as follows. We randomly sampled 60% of the articles (about 2.15 million of 3.6 million articles). We chose random sampling against purpose sampling (e.g. snowball sampling), because, for example, selecting articles that are more popular and attract more editors would not create a representative subset population. Random sampling ensures that each article has identical probability to be chosen independently of other articles it is connected to. Similar to Liu and Ram (2011), out of the selected articles, we extracted all the articles labeled as featured (fa), good (ga), B-class (b), and C-class (c), totalling a number of 84,915 articles. Note that we investigated featured, good, B-class, and C-class articles because these four represent well-distinguished, non-transitional quality classes (Nemoto et

al., 2011). We ensure unbiased statistics by chosing Wikipedia articles with equal likelihood to represent one representative quality class. Thus, by using random sampling, the insights are more likely to be generalizable to the entire Wikipedia articles as well. Finally, we identified all the changes in article quality during an article's life span, ranging from 3-level demotion (fa→c) to 3-level promotion (c→fa). The resulting dataset contains in total 11,359 changes in article quality, on which we base our analysis. Table 1a and Table 1b give an overview of the counts and proportions of each type of change in article quality.

[Insert Table 1a and Table 1b about here]

3.3 Independent Variables

Distribution of Communication Turn-Taking. We operationalize the equality of distribution of communication turn-taking in the same manner as Woolley et al. (2010, p. p. 688). For each article, we counted the number of times each Wikipedia janitor contributed a discussion point on the talk page of that article. The distribution of these counts represents the distribution of *communication turn-taking* among the Wikipedia janitors who worked on the article (TT_i, $1 \leq i \leq N$ with N is the number of janitors). We then computed the coefficient of variation of communication turn-taking (CV_{TT}) as the ratio between the standard deviation and the mean of the turn-taking scores (μ_{TT}): $\mathrm{CV_{TT}} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{TT_i} - \mu_{TT})^2}}{\mu_{TT}}$. To interpret this measure of equality, $CV_{TT}=0$ if there are for instance 16 communication turns in a talk page of 4 janitors who each communicate 4 times. That is, $CV_{TT}=0$ means an equality in the communication turn-taking of the janitors on the talk page of an article. The smaller the value of CV_{TT} suggests the more equality in the distribution of communication turn-taking of the janitors on the talk page of an article.

Group Diversity. WikiProjects provide a valuable setting to quantify the amount of diversity among Wikipedia janitors according to their editing activity and interests (Chen et al., 2010,

p. p. 18). In this sense, we use the Blau index (Agresti, 2002) to measure the probability that two entities (e.g. Wikipedia janitors), taken at random from the dataset of interest (editorial team of an article), represent different types, i.e. do not share any Wikiproject in common (Blau, 1977). The formal definition of the Blau index is $\mathrm{BI} = 1 - \sum_{k=1}^{K} {p_k}^2$, where K is the number all possible categories (i.e., of Wikiprojects), and p_k is the proportion of the Wikipedia janitors who share (i.e. are all members of) project k. A value of the Blau index equal to zero indicates that all Wikipedia janitors who are part of the editorial team of an article share at least one Wikiproject. If all Wikipedia janitors share no Wikiproject among them, the corresponding Blau index equals 1.

Group Cohesiveness. Group cohesiveness is achieved when the members of the group are committed to work together to accomplish a collective task (Guzzo et al., 1995; Hsu et al., 2011; Kidwell et al., 1997). Group cohesiveness refers to "the resultant of all the forces acting on members to remain in the group" (Festinger, 1950). In other words, group cohesiveness is the 'stick togetherness' of the group, the bonds of unity that hold group members together to achieve their common goals (Ren et al., 2012; Schultz, 1988). Members of strongly cohesive groups are more willing to participate readily and to stay with the group (Dyaram & Kamalanabhan, 2005). In line with these definitions, we instantiate group cohesiveness relative to the development process of articles in Wikipedia as the proportion of Wikipedia janitors who previously contributed to the article development, and stayed in the team during the most recent interval of quality change (Chansler et al., 2003; Dyaram & Kamalanabhan, 2005).

3.4 Control Variables

Editing Effort Distribution. This is an important control variable, as previous studies pointed to the long tail distribution of edits on Wikipedia articles (Kittur & Kraut, 2008; Wilkinson, 2008). Here we expect to see that such distributions exist among Wikipedia

janitors as well, and so we decide to control for it in our analysis. One of the most commonly used metrics to describe the inequality of a distribution (in this case, equality of contributions) is the Gini coefficient of homogeneity, with higher values indicating greater inequality (Dorfman, 1979). In the context of Wikipedia, a low Gini coefficient of the edit distribution means that the work is equally distributed among Wikipedia janitors; conversely, a large value of the Gini coefficient reflects situations where a few Wikipedia janitors are performing most of the work (Allison, 1978).

To account for exogenous variables, we also consider the following control variables related to Wikipedia articles that may influence the article quality: **article age, article popularity, group size, and initial article quality**. A detailed description of these variables and a summary of all the variables used in this study are presented in Table 2.

[Insert Table 2 about here]

4. DATA ANALYSIS

Since the aim of our analysis is to identify influential predictors and gain insights into the relationship between the predictors and the outcome, we employ a regression model (Harrell, 2001, p. 97). A longitudinal analysis is conducted as follows: for each quality assessment of an article, we considered the period between the previous and the current change in quality. Similar to Kittur and Kraut (2008), the metrics computed in this time span are used in order to explain the change in article quality, as depicted in Figure 1.

As the scales/units of the main predictors are not easy to interpret (they are not natural metrics), we employ a *standardized* version of the binary logistic regression model. As explained in (Schielzeth, 2010), standardization facilitates not only a common scale of all the input variables so that the coefficients are better understood, especially in the presence of interactions, but also a comparison of the relative importance of the individual main effects: *marginal effect size estimates*. Table 3a shows the summaries of unstandardized (original)

and standardized datasets that are used for modeling. We note here that the original dataset is a subset of the main dataset, obtained by keeping only non-missing observations of the eight covariates (the main predictors and the controls variables). In this way, the meaning of promotion and demotion are not violated since the aforementioned articles cannot be promoted to any higher quality level than "fa" and cannot be demoted to any lower quality level than "c".

[Insert Table 3a and Table 3b about here]

We first consider a baseline model in which the covariates are the control variables. We refer to this as Model 1. Afterwards, similar to Kittur and Kraut (2008), we add Edits as the main predictor and all the interactions with the variables in Model 1. The resulting model is referred to as Model 2. Finally, in Model 3, which is the main contribution of this study, we add Communication, Diversity, and Cohesiveness as main predictors, along with the corresponding interaction effects. We computed the *AIC* value corresponding to each model together with two summary measures for the goodness of fit: the Nagelkerke's R^2 (also known as Cragg & Uhler's R^2) and the chi-square test for goodness of fit (*chi-square GOF*) (Hosmer and Lemeshow, 1980; Hosmer et al., 1997). Values of *chi-square GOF* larger than 5% indicate that the considered model is fit. A summary of the results of the logistic regression is presented in Table 4.

[Insert Table 4 about here]

5. DISCUSSIONS OF THE FINDINGS

The results of our data analysis confirm that contribution inequality (Edits) among Wikipedia janitors has a positive effect on article promotion in Wikipedia. More precisely, a one standard deviation increase in the inequality of editing effort distribution (Edits) among Wikipedia janitors corresponds to an increase of 169% (i.e., exp(0.99)-1) in the likelihood for

an article to be promoted (vs. demoted), when all the other variables are kept at their mean level (see Table 4, Model 3).

With respect to our first main variable of interest, i.e., the distribution of communication turn-taking (Communication), our results show that unequal distribution of communication turn-taking is not favorable for promoting an article to a superior quality class. In other words, higher uniformity in the distribution of communication turns correlates with higher article promotion likelihood. A one standard deviation increase in the equality of communication turn-taking corresponds to an increase of 15% (i.e., exp(0.14)-1) in an article's promotion likelihood (see Table 4, Model 3). This result confirms H1 and the finding of Woolley et al. (2010). It is worth noticing the consistency of the two results, given the obvious difference in the two study contexts: we study real-life groups working on fuzzy tasks for an extended period of time whereas Woolley et al.'s (2010) study laboratory-controlled face-to-face groups working on predefined tasks for a few hours. This remarkable consistency of results suggests that, regardless of the mode (offline or online), or the degree of certainty of the task, or the work duration, the more the communication turns are uniformly distributed among the group members, the higher the quality of group outcomes.

We found that the diversity of Wikipedia janitors has a significantly positive impact on the likelihood for an article to be promoted. A one standard deviation increase in diversity corresponds to a 12% (i.e., exp(0.12)-1) increase in the promotion likelihood of an article (see Table 4, Model 3). Following H2, we further tested for a quadratic effect of diversity: although we found a negative coefficient of the quadratic term, the coefficient was not significant. This does not offer sufficient support for H2. Looking at the interaction effect of the diversity variable with the control variables (see Table 4), one notable interaction is the interaction with group size where the positive effect of diversity became much stronger when the number of Wikipedia janitors increased. To dig deeper into this interaction effect, we run

additional analysis (see Figures 2a, b, c). Collectively these figures show that the effect of Wikipedia janitors' diversity on an article's promotion probability is positive when the article's initial quality is "b" (indicating a less developed article) and, further, that this effect becomes stronger with the increasing size of Wikipedia janitors. However when the article's initial quality is "ga" (indicating a more developed article), the size of Wikipedia janitors appears to undermine the effect of diversity, i.e., the effect of diversity on the probability for an article to be promoted becomes negative. Altogether these results imply having many diverse Wikipedia janitors is beneficial when the article is at a less developed stage, while having few and less diverse Wikipedia janitors is beneficial when the article is at a more developed stage.

[Insert Figures 2a, b, c]

Group cohesiveness is found to be the most important main effect relative to an article's likelihood of promotion. A one standard deviation increase in group cohesiveness corresponds to a 68% (i.e., exp(0.61)-1) increase in an article's promotion likelihood (see Table 4, Model 3). This result confirms H3. In contrast with Woolley et al. (2010), but in line with the theory of group cohesiveness (Festinger et al., 1950), this finding shows the importance of the stick-togetherness of the Wikipedia janitors for producing high quality outcomes. As with the diversity variable, the most notable interaction effect of this variable is its interaction with the size of the Wikipedia janitors, i.e., the higher the number of Wikipedia janitors working on an article, the more positive the effect of their cohesiveness on the article's promotion likelihood. However when we dig deeper into this interaction, we found a marginal reverse effect when the article is at a less developed stage (see Figure 2b). This is peculiar because one would expect a tightly bonded group to be important in improving the quality of a less developed article, especially when the number of individuals working on the article is quite large. One possible explanation of this finding could be that people that stick

together typically think alike (homophily); having too many like-minded people at the early stage of article development may not be beneficial on the likelihood of the article being promoted to a higher quality level. In this case, the stick-togetherness of the Wikipedia janitors editing the article may be undermined by biases resulting from the fact that they think alike. Indeed we found a negative correlation between diversity and cohesiveness (corr = -0.42).

6. CONTRIBUTIONS AND IMPLICATIONS

Findings of this study have to be viewed in light of its limitations. With regard to generalizability and endogeneity, we acknowledge that several areas dealing with the dynamics of social interaction in online collaboration that were not examined in this study, such as the issues of social power or culture (Baym, 2006; Jiang et al., 2011). Pragmatically, there may be several perspectives for examining peer interactions in online social production communities. Conditions other than group interactions and group composition characteristics – such as users' capabilities and goals, their interpretations of technology, and institutional contexts, power, or culture – may play key roles in causal explanations. Due to the possibility of omitted variable bias, simultaneous causality bias, and errors-in-variable bias, future research should examine our identified relationships using more controlled settings or methods such as instrumental variables regression. Nonetheless the findings of this study contribute to the extant literature and have important implications for practitioners.

This study contributes to the extant literature of online social production communities in general and Wikipedia in specific by investigating how the properties and interactions of the people who are most engaged in the production of the common goods, i.e., the Wikipedia janitors, affect quality outcomes. Linus' law that "many eyes make all bugs shallow" have been repeatedly challenged by findings that show that, in reality, only few volunteers significantly contribute to common goods. However no study has ever investigated how the

few Wikipedia janitors contribute to the success of Wikipedia. Ours is a pioneering study that investigates the composition of the Wikipedia janitors, how they interact, and how their composition and interaction affect output quality. Through a longitudinal analysis of the composition of Wikipedia janitors, their communication turn-taking, and the quality changes (promotion or demotion) of articles they worked on, we found that the most important ingredient to an article's likelihood of promotion is the cohesiveness of Wikipedia janitors. At the same time, however, cohesiveness implies lower diversity and this, in turn, may not be beneficial for an article at its early development stage. An interesting question for future research would be to explore further the tension between cohesiveness and diversity in other online social production communities, e.g., examine whether this dilemma persists in OSS projects (which produce technically more challenging output than Wikipedia) at different software development stages.

While comparisons between offline and online social production are not within the scope of the study, it is interesting to note here the similarities and differences between our findings and the findings of Woolley et al. (2010), which is perhaps the most comprehensive study of face-to-face group interactions to date. Our context is real-life groups working on fuzzy tasks for an extended period of time whereas Woolley et al. (2010) study laboratory-controlled face-to-face groups working on predefined tasks for a few hours. Despite the obvious contextual differences, we similarly found the positive effect of an equal distribution of communication turn-taking on output quality. However, while Woolley et al. (2010) found that group cohesiveness has no effect on output quality, we found it to be the most important antecedent. This difference can be possibly related to the different lengths of work duration of the groups. Future research could pursue a 2 x 2 x 2 research design that manipulates the interaction mode (offline or online), the degree of task certainty (low or high), and the work duration (short or long) for a more direct comparison of the different settings.

From a managerial perspective, organizations increasingly consider the outsourcing of knowledge tasks to large masses of workers via distributed labor networks using limited or no monetary incentives; this is possible, in part, due to the fact that the virtual, self-organizing workplace constantly evolves towards more spontaneous and decentralized forms of collaboration. While open collaborative innovation can potentially displace producer innovation at many parts of the economy (Baldwin and Hippel 2011), the fluid generativity of distributed innovation suggests that knowledge resources will be increasingly heterogeneous and often only temporarily integrated (Yoo et al. 2012). Reflecting from the Wikipedia case, having janitors (who could be individuals employed by the organization) tasked with constantly organizing and refining the collectively produced knowledge resources is important for organizations who would like to outsource knowledge tasks to large masses of online distributed workers. This study provides insights on the most effective composition of janitor groups, as well as on how they should interact with one another. While the focal organization may not be able to control the way janitors interact with one another, it should be able to define and alter their composition (diversity and cohesiveness) over time. Findings in this study can guide the organization in doing so.

REFERENCES

Aaltonen, A., & Lanzara, G. (2011). *Governing Complex Social Production in the Internet: The Emergence of a Collective Capability in Wikipedia.* Paper presented at the Decade in Internet Time symposium.

Agresti, A. (2002). *Categorical data analysis* (Vol. 359): Wiley-interscience.

Allison, P. D. (1978). Measures of Inequality. *American Sociological Review, 43*(6), 865-880. doi: 10.2307/2094626

Aral, S., Brynjolfsson, E., & Van Alstyne, M. (2008). Antecedents and Consequences of Mutual Knowledge in Teams. *Social Science Research Network (SSRN)*.

Arazy, O., & Nov, O. (2010). *Determinants of wikipedia quality: the roles of global and local contribution inequality.* Paper presented at the Proceedings of the 2010 ACM conference on Computer supported cooperative work.

Arazy, O., Nov, O., Patterson, R., & Yeo, L. (2011). Information quality in Wikipedia: The effects of group composition and task conflict. *Journal of management information systems, 27*(4), 71-98.

Baym, N. K. (2006). Interpersonal life online. *Handbook of new media: social shaping and social consequences of ICTs*, 35-54.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*: Yale University Press.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*: Free Press New York.

Brandes, U., & Lerner, J. (2009). *Is Editing More Rewarding Than Discussion? A Statistical Framework to Estimate Causes of Dropout from Wikipedia*. Paper presented at the Proceedings of the WWW 2009.

Bryant, S. L., Forte, A., & Bruckman, A. (2005). *Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia*. Paper presented at the Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work, Sanibel Island, Florida, USA.

Butler, B., Sproull, L., Kiesler, S., & Kraut, R. (2007). Community effort in online groups: Who does the work and why. *Leadership at a Distance, 11*, 171-194.

Chansler, P. A., Swamidass, P. M., & Cammann, C. (2003). Self-Managing Work Teams: An Empirical Study of Group Cohesiveness in "Natural Work Groups" at a Harley-Davidson Motor Company Plant. *Small Group Research, 34*(1), 101-120. doi: 10.1177/1046496402239579

Chen, J., Ren, Y., & Riedl, J. (2010). *The effects of diversity on group productivity and member withdrawal in online volunteer groups*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA.

Choi, B., Alexander, K., Kraut, R. E., & Levine, J. M. (2010). *Socialization tactics in wikipedia and their effects.* Paper presented at the Proceedings of the 2010 ACM conference on Computer supported cooperative work.

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse Studies, 7*(4-5), 507-525.

Collier, B., & Kraut, R. (2012). Leading the Collective: Social Capital and the Development of Leaders in Core-Periphery Organizations. *arXiv preprint arXiv:1204.3682*.

Conaldi, G., & Lomi, A. (2013). The dual network structure of organizational problem solving: A case study on Open Source Software development. *Social Networks*.

Crowston, K., Li, Q., Wei, K., Eseryel, U. Y., & Howison, J. (2007). Self-organization of teams for free/libre open source software development. *Information and software technology, 49*(6), 564-575.

Dorfman, R. (1979). A Formula for the Gini Coefficient. *The Review of Economics and Statistics, 61*(1), 146-149. doi: 10.2307/1924845

Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). *Are your participants gaming the system?: screening mechanical turk workers.* Paper presented at the 28th International Conference on Human Factors in Computing Systems, Atlanta, GA, USA.

Dyaram, L., & Kamalanabhan, T. (2005). Unearthed: The other side of group cohesiveness. *Journal of Social Science, 10*(3), 185-190.

Espinosa, J. A., Slaughter, S. A., Kraut, R. E., & Herbsleb, J. D. (2007). Familiarity, Complexity, and Team Performance in Geographically Distributed Software Development. *Organization science, 18*(4), 613-630. doi: 10.1287/orsc.1070.0297

Feller, J., Finnegan, P., Hayes, J., & O'Reilly, P. (2008). Governance Structures for Open Innovation: A Preliminary Framework *Open IT-Based Innovation: Moving Towards Cooperative IT Transfer and Knowledge Diffusion* (pp. 511-525): Springer.

Festinger, L. (1950). Informal social communication. *Psychological review, 57*(5), 271.

Festinger, L., Schachter, S., & Back, K. (1950). The spatial ecology of group formation. *Social pressure in informal groups*, 33-60.

Ghosh, R. A. (2005). *CODE: Collaborative ownership and the digital economy*: The MIT Press.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature, 438*(7070), 900-901.

Guzzo, R. A., Salas, E., Goldstein, I. L., & Guzzo. (1995). *Team effectiveness and decision making in organizations*: Jossey-Bass San Francisco.

Hahn, J., Moon, J. Y., & Zhang, C. (2008). Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties. *Information Systems Research, 19*(3), 369-391.

Halfaker, A., Kittur, A., & Riedl, J. (2011). *Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work*. Paper presented at the Proceedings of the 7th International Symposium on Wikis and Open Collaboration, Mountain View, California.

Harrell, F. E. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*: Springer.

Heylighen, F. (2013). Self-organization in Communicating Groups: the emergence of coordination, shared references and collective intelligence *Complexity Perspectives on Language, Communication and Society* (pp. 117-149): Springer.

Hosmer, D. W., Hosmer, T., Le Cessie, S., & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in medicine, 16*(9), 965-980.

Hosmer, D.W. and Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods*, **A9**(10), p. 1043-1069.

Hsu, J. S. C., Chang, J. Y. T., Klein, G., & Jiang, J. J. (2011). Exploring the impact of team mental models on information utilization and project performance in system development. *International Journal of Project Management, 29*(1), 1-12.

Humphrey, S. E., Morgeson, F. P., & Mannor, M. J. (2009). Developing a theory of the strategic core of teams: a role composition model of team performance. *Journal of Applied Psychology, 94*(1), 48.

Jiang, L., Bazarova, N. N., & Hancock, J. T. (2011). From Perception to Behavior: Disclosure Reciprocity and the Intensification of Intimacy in Computer-Mediated Communication. *Communication Research*.

Jullien, N. (2012). What we know about Wikipedia: A review of the literature analyzing the project (s). *SSRN, 2053597*.

Kidwell, R. E., Mossholder, K. W., & Bennett, N. (1997). Cohesiveness and Organizational Citizenship Behavior: A Multilevel Analysis Using Work Groups and Individuals. *Journal of Management, 23*(6), 775-793. doi: 10.1177/014920639702300605

Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web, 1*(2), 19.

Kittur, A., Chi, E. H., & Suh, B. (2008). *Crowdsourcing user studies with Mechanical Turk.* Paper presented at the 26th Annual SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy.

Kittur, A., & Kraut, R. E. (2008). *Harnessing the wisdom of crowds in wikipedia: quality through coordination*. Paper presented at the Proceedings of the 2008 ACM conference on Computer supported cooperative work, San Diego, CA, USA.

Kittur, A., & Kraut, R. E. (2010). *Beyond Wikipedia: coordination and conflict in online production groups*. Paper presented at the Proceedings of the 2010 ACM conference on Computer supported cooperative work, Savannah, Georgia, USA.

Kittur, A., Lee, B., & Kraut, R. (2009). *Coordination in collective intelligence: the role of team structure and task interdependence.* Paper presented at the Proceedings of the 27th international conference on Human factors in computing systems.

Laniado, D., & Tasso, R. (2011). *Co-authorship 2.0: Patterns of collaboration in Wikipedia.* Paper presented at the Proceedings of the 22nd ACM conference on Hypertext and hypermedia.

Lanubile, F., Ebert, C., Prikladnicki, R., & Vizcaíno, A. (2010). Collaboration tools for global software engineering. *Software, IEEE, 27*(2), 52-55.

Lerner, J., & Tirole, J. (2002). Some simple economics of open source. *The journal of industrial economics, 50*(2), 197-234.

Liu, J., & Ram, S. (2011). Who does what: Collaboration patterns in the wikipedia and their impact on article quality. *ACM Trans. Manage. Inf. Syst., 2*(2), 1-23. doi: 10.1145/1985347.1985352

Mehra, S. (2012). PLATFORMS, TEAMWORK AND CREATIVITY: MEDIATING HIERARCHS IN THE NEW ECONOMY. *JL Econ. & Pol'y, 9*, 15-519.

Menard, S. (2011). Standards for standardized logistic regression coefficients. *Social Forces, 89*(4), 1409-1428.

Nemoto, K., Gloor, P., & Laubacher, R. (2011). *Social capital increases efficiency of collaboration among Wikipedia editors*. Paper presented at the Proceedings of the 22nd ACM conference on Hypertext and hypermedia, Eindhoven, The Netherlands.

Panciera, K., Halfaker, A., & Terveen, L. (2009). *Wikipedians are born, not made: a study of power editors on Wikipedia.* Paper presented at the Proceedings of the ACM 2009 international conference on Supporting group work.

Priedhorsky, R., Chen, J., Lam, S., Panciera, K., Terveen, L., & Riedl, J. (2007). *Creating, destroying, and restoring value in Wikipedia.* Paper presented at the Proceedings of the 2007 international ACM conference on Supporting group work.

Ransbotham, S., & Kane, G. C. (2011). Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly-Management Information Systems, 35*(3), 613.

Ren, Y., Harper, F. M., Drenner, S., Terveen, L. G., Kiesler, S. B., Riedl, J., & Kraut, R. E. (2012). Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds. *MIS Quarterly, 36*(3), 841-864.

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). *Who are the crowdworkers?: shifting demographics in mechanical turk.* Paper presented at the 28th of the International Conference on Human Factors in Computing Systems, Atlanta, GA, USA.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696--735.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution, 1*(2), 103-113.

Schultz, B. G. (1988). *Communicating in the small group: Theory and practice*: Harper & Row.

Shah, S. K. (2006). Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science, 52*(7), 1000-1014.

Shirky, C. (2010). *Cognitive surplus: Creativity and generosity in a connected age*: ePenguin.

Singh, P. V. (2010). The small-world effect: The influence of macro-level properties of developer collaboration networks on open-source project success. *ACM Transactions on Software Engineering and Methodology (TOSEM), 20*(2), 6.

Stewart, K. J., & Gosain, S. (2006). The impact of ideology on effectiveness in open source software development teams. *MIS Quarterly*, 291-314.

Sundin, O. (2011). Janitors of knowledge: constructing knowledge in the everyday life of Wikipedia editors. *Journal of Documentation, 67*(5), 840-862.

Surowiecki, J., & Silverman, M. P. (2007). The wisdom of crowds. *American Journal of Physics, 75*, 190.

Tapscott, D., & Williams, A. D. (2006). *Wikinomics: How Mass Collaboration Changes Everything*. New York: Portfolio Hardcover.

Van Knippenberg, D., De Dreu, C. K., & Homan, A. C. (2004). Work group diversity and group performance: an integrative model and research agenda. *Journal of Applied Psychology, 89*(6), 1008.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: what data quality means to data consumers. *J. Manage. Inf. Syst., 12*(4), 5-33.

Weber, S. (2004). *The success of open source* (Vol. 368): Cambridge Univ Press.

Wilkinson, D. M. (2008). *Strong regularities in online peer production.* Paper presented at the Proceedings of the 9th ACM conference on Electronic commerce.

Wilkinson, D. M., & Huberman, B. A. (2007). *Cooperation and quality in wikipedia*. Paper presented at the Proceedings of the 2007 international symposium on Wikis, Montreal, Quebec, Canada.

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

Zhang, X., & Wang, C. (2012). Network positions and contributions to online public goods: The case of Chinese Wikipedia. *Journal of management information systems, 29*(2), 11-40.

Zhu, H., Kraut, R., & Kittur, A. (2012). *Effectiveness of shared leadership in online communities.* Paper presented at the Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work.

Figures and Tables

Figure 1: Analysis window

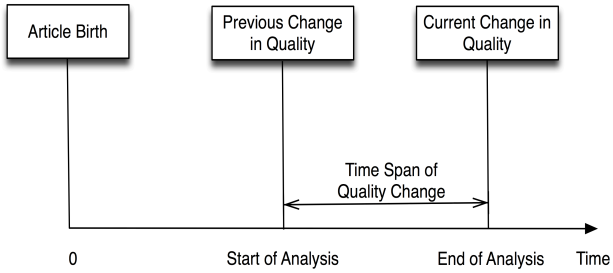

Table 1a: Total counts of quality changes (lines represent the starting quality, whereas columns represent the quality after the change)

From/to quality	c	b	ga	fa
c	0	1285	284	16
b	6239	0	1974	191
ga	146	469	0	501
fa	60	180	14	0

Table 1b: Proportions of quality changes by initial quality and magnitude of changes in quality

Initial quality/magnitude of change	-3	-2	-1	0	1	2	3
InitialQuality=c	NA	NA	NA	**0**	**0.81**	**0.18**	**0.01**
InitialQuality=b	NA	NA	**0.74**	**0**	**0.23**	**0.02**	NA
InitialQuality=ga	NA	**0.13**	**0.42**	**0**	**0.45**	NA	NA
InitialQuality=fa	**0.24**	**0.71**	**0.06**	**0**	NA	NA	NA

Table 2: Operational definitions

Variable	Operational Definition
Main Effects	
Communication Effort Distribution (**Communication**)	Coefficient of Variation of Communication Turn Taking measured on article talk pages (Woolley et al., 2010), e.g. conversations are dominated by single group members.
Group Cohesiveness **(Cohesiveness)**	The proportion of Wikipedia janitors who previously contributed to the article development, and stayed in the team during the most recent interval of quality change (Chansler et al., 2003; Dyaram & Kamalanabhan, 2005).
Group Diversity **(Diversity)**	Editorial team diversity measured using the Blau index (Blau, 1977), i.e. the probability that two entities chosen at random represent different types, with respect to the participation in Wikiprojects.
Controls	
Editing Effort Distribution **(Edits)**	Contribution inequality measured using the Gini Coefficient of the distribution of article editing effort (Allison, 1978; Dorfman, 1979; Espinosa et al., 2007).
Article Age	Duration from the creation of an article until the current change in quality.
Article Popularity	Number of web search hits.
Group Size	Number of Wikipedia janitors (i.e. WikiProject members and administrators) involved in the editing process.
Initial Quality	Quality label of an article at the beginning of the time span of analysis.
Dependent Variable	
Change in Quality	Binary variable measuring whether an article was promoted (+1, i.e. to a superior quality standard) or demoted (-1; i.e. to an inferior quality standard).

Table 3a: Dataset Summary

	Unstandardized						Standardized					
Independent Variables	**min**	**max**	**mean**	**med**	**var**	**sd**	**min**	**max**	**mean**	**med**	**var**	**sd**
Communication	0.00	4.79	0.35	0.00	0.25	0.50	-0.69	8.85	0	-0.69	1	1
Diversity	0.00	0.98	0.53	0.67	0.12	0.35	-1.51	1.26	0	0.38	1	1
Cohesiveness	0.00	1.00	0.28	0.20	0.08	0.28	-0.99	2.57	0	-0.28	1	1
Controls	**min**	**max**	**mean**	**med**	**var**	**sd**	**min**	**max**	**mean**	**med**	**var**	**sd**
Edits	0.00	1.00	0.53	0.54	0.07	0.26	-2.05	1.83	0	0.05	1	1
Article Age	15.00	3739.00	1996.14	2047.00	556050.00	745.69	-2.66	2.34	0	0.07	1	1
Log (Article Popularity)	0.69	22.39	12.00	12.02	12.45	3.53	-3.20	2.94	0	0.01	1	1
Log (Group Size)	0.69	5.06	2.12	2.08	0.89	0.94	-1.51	3.12	0	-0.05	1	1
Initial Quality	**count**						**count**					
c	6445						6445					
b	1934						1934					
ga	2272						2272					
fa	708						708					
Dependent Variable (DV)												
Change in Quality	**count**						**count**					
Demoted (-1)	7108						7108					
Promoted (+1)	4251						4251					
Total nr. of instances	11359						11359					

Table 3b: Variable Correlations

	Correlations						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1. Age	1						
2. Log (Article Popularity)	0.34	1					
3. Log (Goup Size)	0.40	0.31	1				
4. Edits	-0.06	-0.06	0.22	1			
5. Communication	0.08	0.26	0.34	0.33	1		
6. Diversity	0.30	0.24	0.79	0.22	0.26	1	
7. Cohesiveness	-0.16	-0.15	-0.44	-0.08	-0.13	-0.42	1

Table 4: Standardized Binary Logistics Models

	Baseline (Model 1)			Edits (Model 2)			All (Model 3)		
	Coef	**Pr(>|Z|)**	**VIF**	**Coef**	**Pr(>|Z|)**	**VIF**	**Coef**	**Pr(>|Z|)**	**VIF**
Intercept	-0.60	***		-0.85	***		-0.84	***	
Article Age	-0.36	***	1.20	-0.27	***	1.20	-0.33	***	1.40
Log (Group Size)	-0.57	***	1.10	-0.83	***	1.40	-0.65	***	4.20
Log (Article Popularity)	-0.08	***	1.10	-0.03	0.14	1.10	0.03	0.26	1.40
Edits				**0.99**	***	**2.60**	**0.99**	***	**2.80**
Article Age * Edits				-0.08	***	1.20	-0.02	0.51	1.50
Log (Group Size) * Edits				0.54	***	2.30	0.47	***	3.50
Log (Popularity) * Edits				0.04	.	1.20	0.01	0.78	1.50
Communication							**-0.14**	***	**1.60**
Article Age * Communication							-0.10	***	1.30
Log (Group Size) * Communication							0.01	0.70	1.80
Log (Article Popularity) * Communication							-0.02	0.40	1.50
Diversity							**0.12**	*	**4.80**
Diversity^2							**-0.06**	**0.28**	**4.30**
Article Age * Diversity							-0.06	*	1.60
Log (Group Size) * Diversity							0.21	***	5.70
Log(Article Popularity) * Diversity							0.17	***	1.60
Cohesiveness							**0.61**	***	**4.10**
Article Age * Cohesiveness							0.05		1.50
Log (Group Size) * Cohesiveness							0.25	***	4.10
Log (Article Popularity) * Cohesiveness							0.05	*	1.40
AIC	13574			12465			12093		
Pseudo-R^2	0.16			0.28			0.31		
Chi-square GOF	0.27			0.33			0.51		

***Significant at < 0.001; **Significant at < 0.01; *Significant at < 0.05

Figure 2a: Slopes of main effects - when taking into account the marginal effect of GroupSize, when all the other factors are kept fixed, using the following formula:

Slope(MainEffect)=Coef(MainEffect)+ Coef(Interaction(MainEffect, GroupSize))*DomainOfDef (GroupSize)

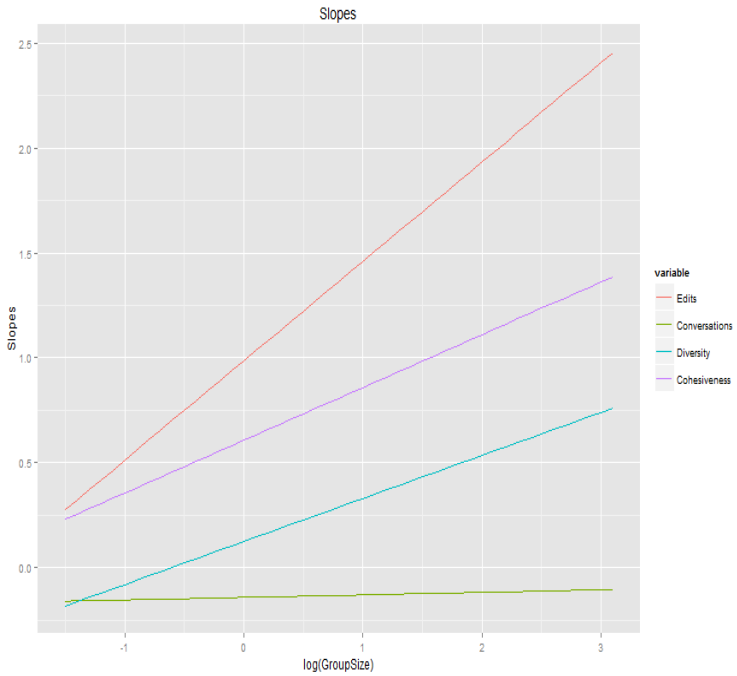

Figure 2b: Slopes of main effects when the initial quality is "b"

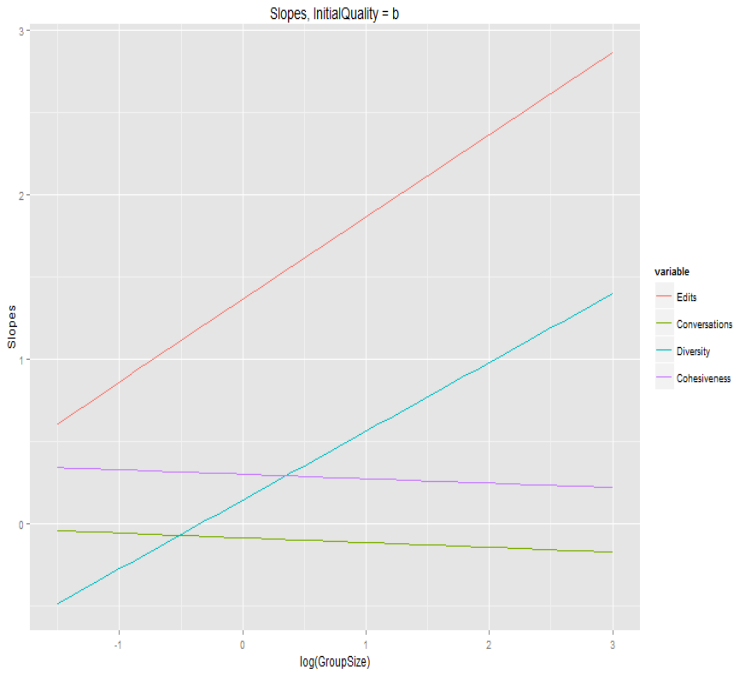

Figure 2c: Slopes of main effects when the initial quality is "ga"