Exploring the Relationship between Membership Turnover and Productivity in Online Communities

Xiangju Qin[♯], Michael Salter-Townshend[♢], Pádraig Cunningham[♯]
[♯]School of Computer Science & Informatics, University College Dublin
[♢]Department of Statistics, University of Oxford
xiangju.qin@ucdconnect.ie, salter@stats.ox.ac.uk, padraig.cunningham@ucd.ie

Abstract

One of the more disruptive reforms associated with the modern Internet is the emergence of online communities working together on knowledge artefacts such as Wikipedia and OpenStreetMap. Recently it has become clear that these initiatives are vulnerable because of problems with membership turnover. This study presents a longitudinal analysis of 891 WikiProjects where we model the impact of member turnover and social capital losses on project productivity. By examining social capital losses we attempt to provide a more nuanced analysis of member turnover. In this context social capital is modelled from a social network perspective where the loss of more central members has more impact. We find that only a small proportion of WikiProjects are in a relatively healthy state with low levels of membership turnover and social capital losses. The results show that the relationship between social capital losses and project performance is U-shaped, and that member withdrawal has significant negative effect on project outcomes. The results also support the mediation of turnover rate and network density on the curvilinear relationship.

Introduction

With the popularity of Web 2.0, recent years have witnessed a growing population of online communities which rely on contributions from online volunteers to build knowledge and software artifacts. Despite the success of a few communities, such as Linux, Apache, Wikipedia and OpenStreetMap, many of them fail to generate desired outcomes. Many of these failures are due to the intrinsic characterstics of open collaboration in online communities, namely, high levels of member withdrawal (Faraj, Jarvenpaa, and Majchrzak 2011; Ransbotham and Kane 2011). Even in successful online communities, such as the English version of Wikipedia, high churn in membership is still the norm, with 60% of registered users staying only a day (Panciera, Halfaker, and Terveen 2009; Dabbish et al. 2012). Figure 1 presents the statistics of yearly active registered editors in English Wikipedia from its inception. Figure 1 shows that each year the number of new editors dominates the amount of active editors, suggesting that Wikipedia experiences high level of membership turnover.

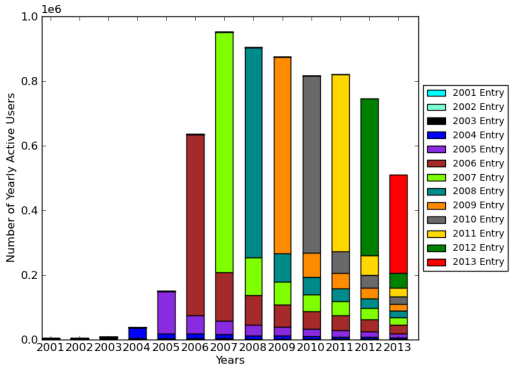

Figure 1: Statistics of Yearly Active Editors in English Wikipedia from 2001 to 2013. The data for 2013 was only till August 5, 2013.

In traditional organizations, the cost of leaving a job can be extremely high in terms of sacrificing material or psychological benefits, such as giving up colleagues, interesting projects, corporate stock and pension benefits (Mitchell et al. 2001). By contrast, in online communities, without committing to any tasks, projects, or conversations, the barriers to entry and exit are very low, each participant is free to join and leave (Ransbotham and Kane 2011). Joining an online group can be as easy as filling in a form or writing a profile (Dabbish et al. 2012), leaving an online group is even easier by stopping contributions or showing no activities in the community. For volunteers, leaving an online community generally incurs little cost (except losing opportunity to learn from peers). While for the online community, member withdrawal can have a significant influence on its functionality: membership turnover may improve group performance by bringing in novel skills and expertise, but it may also threaten the cognitive structures and processes that are so useful for expertise coordination (Lewis, Belliveau, and Herndon 2007; Faraj, Jarvenpaa, and Majchrzak 2011), and inevitably disrupt routines and established social ties and further reduce its social capital (Shaw et al. 2005).

High membership turnover in online communities has attracted much attention from researchers to study the influence of turnover on community health. The most dominant view suggests that membership turnover is disadvantageous to effective collaboration, and the ability to engage and retain members represents a key aspect of success (Arguello et al. 2006). Those who leave the community, take with them not only unique knowledge and insight, but also the experience and social relationships they have gained and established through participation. On the other hand, some research suggests that the unique capabilities of Web 2.0 techniques could help mitigate the negative effects of turnover. For example, Web 2.0 platforms are typically equipped with functions to preserve all previous contributions by past members in an organized and searchable manner, so as to facilitate the utilization of past contributions for future community members as needed (Kane and Fichman 2009; Wagner and Majchrzak 2006). While these studies shed light on the relationship between turnover and performance in the online scenario, a major limitation is that they fail to consider social capital loss through member turnover and its impact on the functionality of communities.

Despite its success as a means for knowledge sharing and collaboration, little is known about the mechanism behind Web 2.0 platforms from the perspective of social network analysis. The social network approach provides a unique perspective to study the evolution of online communities over time, and to study the resilience of social networks against node removal (i.e., member drop-out) in these platforms. In this study, we take a social network approach to investigate the relationship between member turnover, social capital losses and group performance in WikiProjects. We are interested in the following questions: Does the drop-out of members in different network positions have the same effect on group performance? How does social capital loss due to departing members influence group performance? Is there consistency across the evolution of online groups? Overall, the eventual goal of this study is the provision of multi-dimensional indicators related to the health and performance of online communities, which provides insight into the management and evaluation of these communities.

The rest of this paper is organized as follows. The next section provides a brief review of related work. In Section 3, we explain network measures and project characteristics related variables, and develop hypotheses. Next, we discuss data collection and model specification. Section 5 presents the results, followed by discussion and conclusion.

Membership Turnover and Performance

There has been substantial work studying the effect of turnover on performance in traditional organizations, but perspectives on this relationship vary widely. The most dominant view suggests that turnover has a negative impact on performance, because the turnover of members can incur replacement costs, disrupt the social network, or weaken the knowledge resources of the organization (Huselid 1995). A second argument indicates that in certain situations turnover may benefit organizations since departing employees are often those most dissatisfied with the current organization, so those who remain behind enjoy better working conditions and performance (Krackhardt and Porter 1985). A third perspective suggests that moderate levels of turnover lead to the best organizational performance (Abelson and Baysinger 1984) in terms of creating opportunities for organizations to access new skills and knowledge through the influx of new employees (Madsen, Mosakowski, and Zaheer 2003). In their empirical study on 38 locations of a restaurant chain, Shaw et al. found that employee turnover rate moderates the curvilinear relationship between social capital losses and performance. Despite the fact that online communities (e.g., Wikipedia) generally experience considerable membership turnover, the effects of turnover on collaborative outcomes in these communities are not well understood.

Because online communities rely on the contributions of volunteers to survive and succeed, member turnover can have significant effects on collaborative activities in terms of resource availability and the health of a community. Faraj, Jarvenpaa, and Majchrzak suggested that the fluidity of members can bring opportunities for knowledge collaboration when managed appropriately. By experiments on 2,065 featured articles from Wikipedia, Ransbotham and Kane found a curvilinear relationship between the turnover of Wikipedia editors and the quality of an article. Specifically, they discovered that a Wikipedia article edited by experienced editors is more likely to be of higher quality up to an optimal level of experience, after which the average experience of editors decreases the quality of the article (by demoting it from FA status to lower quality status). Wang and Lantzy provided a framework to systematically examine the relationship between member turnover and the health of a community, where community health is measured as the content volume, traffic, responsiveness, interactivity of the community. However, the authors haven't provided empirical results to evaluate their hypothesis. Zanetti et al. studied the evolution of social structual in open source software communities along multiple dimensions, such as the cohesiveness of the communities, the distribution of responsibilities and the resilience against membership turnover in the community. However, this study is limited in that it did not take into account community performance indicators. Different from other studies, Kivran-Swaine, Govindan, and Naaman explored the breaking of ties using social network theories and found that network structure can significantly explain the breaking and persistence of ties in online services like Twitter.

The aforementioned works provide interesting insights into the impact of member turnover on the online communities in terms of the influx of novel knowledge, the community health, the evolution and breaking of network structures. However, none of them explore the impact of membership turnover on the collaborative outcomes of voluntary groups. Given that Web 2.0 platforms generally experience high levels of member withdrawal, it is of great significence to understand the influence of member turnover on the functionality of these platforms. In this work, we build on the existing liturature and take a social network approach to investigate the relationship between social capital losses and group productivity in online communities (in particular Wikipedia).

Measures

To investigate the research questions, we explore the relationship between the performance of online groups, membership turnover, and social capital loss related measures. The following introduces the dependent, independent and control measures used in this study.

Dependent Variable

Previous work on Wikipedia generally measures group performance[1] as the amount of work done by members (i.e., the total number of edits to Wikipedia article pages by members). Admittedly, this is a very raw proxy of contributions by members to Wikipedia, and it generally neglects the quality of the contributions. We further calculate the edit longevity of each edit using WikiTrust software by (Adler and Alfaro 2007), which computes the value of each contribution by combining its quality and quantity.

Independent Variables

Social Capital Losses (SC Losses). Social capital theory has been developed to explain the value inherent in social relationships. Social capital is created when relationships facilitate reciprocal action among people (Coleman 1988). Although there is no universal and precise definition for network social capital (e.g., network closure by (Coleman 1988) vs structural holes by (Burt 1992) as social capital), social capital can be broadly defined as the benefits that network members secure from their memberships in social networks or other social structures (Portes 1998). These studies further contend that organizational social capital can improve performance by enhancing commitment, increasing flexibility, and promoting knowledge sharing and information diffusion that fosters intellectual capital (Coleman 1988; Burt 1992; Shaw et al. 2005).

If social capital at the collective level is generated when connections facilitate instrumental action among people (Coleman 1988), it is also lost when the connections among people are dissolved (Burt 1992; Shaw et al. 2005). These connections, which are generally referred to as structural holes by (Burt 1992), play an essential role in facilitating resources exchange and knowledge management for network members. Researchers suggest that turnover among core members who occupy key network positions should be more damaging to organizational performance than withdrawal among members who occupy less central network positions (Shaw et al. 2005). We expect that in online communities, social capital losses due to member turnover should have similar impact on group performance.

As with other researchers, we used "betweenness centrality" by (Freeman 1979) to measure the extent to which members occupied structurally advantageous positions (i.e., structural hole bridges) in the collaboration network of a WikiProject. Following Shaw et al., we calculate social capital losses as a ratio: the average betweenness scores of the leavers divided by the average betweenness scores of all project members.

Turnover rate. Measured as the percentage of members who were active in previous quarter and not active in current quarter. We consider a member being active in a quarter if and only if the member made at least one edit to articles within project scope.

Network Density. This variable measures the proportion of existing connections in the network to the number of possible pairwise combinations of members, and takes value from zero to one, with larger values indicating increasing density. We include this measure because previous work found that in case of social capital losses due to member turnover, a dense network with experienced members and the abundance of redundant social ties would be more tolerant to network disruptions or gaps created by turnovers (Shaw et al. 2005).

Control Variables

Quarter. This variable indexes the quarters in which project-related measures are collected: starting with quarter 0 from the time a project is created, till the last quarter of study periods. The coefficient of this variable indicates the dynamics of project performance over time.

Project creation time (A proxy of project age). Measured as the number of quarters (90-day periods) from the baseline timestamp (the timestamp that the first WikiProject was created) to the creation timestamp of a project. Chen, Ren, and Riedl found that the creation time of a project has negative influence on its performance. They explained that due to the growth in the number of projects and participants over time, projects created later may face a more different environment than those created earlier. In this work, we also include this variable to control for the effect.

Project scope. Measured as the accumulated number of articles that are tagged by a project template up to the current quarter. Most changes reflect new articles being added to the project's scope.

Project size. Measured as the number of project members and active participants during a quarter.

Edit activity in project talk pages (Discussion topics). In online communities, members generally rely on discussion pages to communicate with others, ask for help and support, reach consensus on controversial issues and make collective decisions regarding rules and regulations. A large amount of content posted on project-related talk pages is beneficial for the community in terms of enabling participants to arrive at a general understanding or get help for specific questions. Qin, Salter-Townshend, and Cunningham found that the number of discussion topics is positive and significant related to project efficiency in Wikipedia. We calculated the amount of group communication as the total number of discussion topics recorded in the edit history of project-related talk pages in a quarter.

Mean membership tenure (Mean Tenure). The average membership tenure of a WikiProject is calculated as the sum of the membership tenure of its members divided by the number of members in a quarter.

[1]The terms performance, productivity and outcomes are used interchangeably, and refer to the amount of work done by members to project scope articles.

Following Chen, Ren, and Riedl, we also measured membership tenure as the amount of time a member has been active in Wikipedia. Specifically, our measure of membership tenure consists of two parts: membership tenure before and after joining a WikiProject. The previous tenure of a member is calculated as the number of days between the timestamp that the user made the first edit in Wikipedia and the timestamp that this user joined a WikiProject. For a specific quarter, for any month, if a user made at least one edit to any project related pages, we calculated the monthly project tenure of this user as the number of days in that month. Then for each project member, we accumulated his previous tenure and monthly project tenure up to the current quarter to obtain his overall membership tenure in Wikipedia.

Level of controversy. The controversy of project topics may have impact on project performance as it takes time and effort for project members to discuss and reach consensus on controversial issues. Following Chen, Ren, and Riedl, we also include this variable to control for these effects and to separate them from the main effects caused by social capital losses. This variable is measured as the percentage of reverted edits to all articles within project scope during a quarter, normalized by the overall percentage of reverts in Wikipedia in the same time period.

Tenure diversity. In online communities like Wikipedia, it is often the case that members who have been active for a long time tend to be more experienced than new users. These active members play a fundamental role in the community in terms of spreading knowledge, information and experience across the whole community. High tenure diversity indicates high variability among project members in the time they have spent committing to group tasks and in the experience they have accumulated (Chen, Ren, and Riedl 2010). Studies about open collaboration in online communities found that members with different tenure are more likely to perform different kinds of tasks (Bryant, Forte, and Bruckman 2005), so projects with a mix of newcomers and old-timers may enjoy better task distribution.

The coefficient of variation has been widely used in management science literature to measure tenure diversity (Chen, Ren, and Riedl 2010), because the potential benefits and conflict due to tenure are derived from the spread of members' experience and knowledge relative to a project's mean tenure (Harrison and Klein 2007). Following this convention, we measured tenure diversity using the coefficient of variation of tenure of all project members.

Hypotheses

To test the effect of turnover rate, we make the first hypothesis:

Ha *There is a negative relationship between a project's turnover rate and its performance.*

Previous work has suggested curvilinear relationship between social capital losses and organizational performance (Shaw et al. 2005). Motivated by these studies, we develop the following hypothesis regarding the effect of social capital losses:

Hb *Social capital losses of a project due to member turnover relates in a curvilinear fashion to its performance, strongly declining performance with moderate levels of social capital loss but increasing performance at higher levels of social capital loss.*

Shaw et al. found that turnover rate mediates the curvilinear relationship between social capital losses and organizational performance. Following Shaw et al., we develop the following hypothesis:

Hc1 *Social capital losses and turnover rate interact in predicting project performance: the curvilinear relationship between social capital losses and performance is stronger when turnover rate is low.*

Shaw et al. suggested that in case of social capital losses, a dense network with the abundance of redundant social ties would be more tolerant to network disruptions or gaps created by turnovers. However, they found no support for their predictions regarding the mediated effect of density on the relation between social capital losses and organizational performance when density is low. They further suggested that it would be beneficial for future research to explore the conditions under which social capital losses in dense network structures are damaging to organizational performance (Shaw et al. 2005). Following this suggestion, we make the following hypothesis:

Hc2 *Social capital losses and network density interact in predicting project performance: the curvilinear relationship between social capital losses and performance is more pronounced when network density is high.*

Methods

Data Collection

The dataset we used in this study is extracted from the August 2013 dump of English Wikipedia[2], which includes the complete edit history of all pages from the inception of Wikipedia. Project members generally claim an article in its scope by inserting project template in article talk page. We parsed the link to WikiProject in article talk pages and accumulated all articles tagged by a WikiProject up to a specific quarter to estimate its project scope. We parsed the historical edits of a project's member list to identify members and their joining time for each WikiProject. WikiProjects generally provide main pages or subpages to maintain the list of project members, and any editors can join a project by adding one's username to the member list and then leave the project by removing the username from the list.

We calculated the edit longevity of each edit to articles using the WikiTrust software by (Adler and Alfaro 2007). We then obtained the amount of work done by project members in a quarter by accumulating all the edits or the edit longevity of edits contributed by members to project scope articles. We constructed the collaboration network for each WikiProject in a quarter by considering the co-edit activities of members to project scope articles: if two members made edits to an article, and both the edit longevity of their edits were positive, then the two members established a co-authorship connection. By accumulating all the connections among members within a project, we obtained the collaboration network for the project.

[2]http://dumps.wikimedia.org/enwiki/20130805/

We are interested in exploring how member turnover and social capital losses influence project productivity. To make the results more meaningful, we only included those WikiProjects which had at least 50 accumulated tagged articles and at least three members (the minimal size of group) in this study. As a result, we obtained a longitudinal dataset of 891 WikiProjects each with at least 1 to 35 quarterly observations, each observation recording the characteristics, network measures and outcomes of a project in a quarter.

Model Specification

A preliminary analysis of the data revealed that some of the independent variables were positively skewed. Following (Gelman and Hill 2007), we performed a logarithmic transformation on these independent variables to make the coefficients more comparable. To explore the possible non-linear relationship between social capital losses and project performance, we included the squares of the social capital losses in the model. To validate the possible mediated effects that member turnover or network density has on the relationship between social capital losses and performance, we added the interaction of turnover rate (or density) and social capital losses in the model. There is a non-linear dependence of productivity on social capital losses. We fit a second order polynomial (using the poly function in R) and discuss the curvilinear relationship we find. We chose a second order polynomial (quadratic) over a smoothing spline to facilitate comparison with existing work (Shaw et al. 2005). The preliminary analysis also suggested that the collinearity of the model is very high (with kappa value larger than 30), which indicates potentially harmful collinearity among variables (Baayen 2008). We used z-score to scale the Mean Tenure variable, which reduces the collinearity of the model significantly. To ease the interpretation, we represent the log-transformed and z-score normalization variables with"Ln" and "Sc" prefix, respectively.

Our data has a nested structure – quarterly observations nested within projects – which suggest Hierarchical Linear Model (Bryk and Raudenbush 1992) for our analysis. HLM is an advanced form of linear model which takes into account potential autocorrelations among time periods that are nested with the same project, while allowing us to examine the main effects of independent variables on dependent variables. Meanwhile, to account for project-level unobserved differences, we include project level random effects in the model. The model specification is as follows:

$$\begin{aligned}\text{LnDV}_{jt} &= \beta_0 + \beta_1 \text{Quarter}_{jt} + \beta_2 \text{Project creation}_{jt} + \beta_3 \text{LnProject scope}_{jt} \\ &+ \beta_4 \text{LnProject size}_{jt} + \beta_5 \text{LnDiscussion topics}_{jt} \\ &+ \beta_6 \text{ScMean tenure}_{jt} + \beta_7 \text{Level of controversy}_{jt} + \beta_8 \text{Tenure diversity}_{jt} \\ &+ \beta_9 \text{Turnover rate}_{jt} + \beta_{10} \text{Density}_{jt} \\ &+ \beta_{11} \text{SC Losses}_{jt} + \beta_{12} \text{SC Losses}_{jt}^2 \\ &+ \beta_{13} \text{SC Losses}_{jt} \times \text{Turnover rate}_{jt} \\ &+ \beta_{14} \text{SC Losses}_{jt}^2 \times \text{Turnover rate}_{jt} + \upsilon_j \end{aligned} \quad (1)$$

where variables are indexed across project (j) and quarter (t); LnDV_{jt} represents the dependent variable; υ_j is the random effect at the project level. We obtained our estimates using the lmer function in lme4 package (Bates, Maechler, and Bolker 2012) for **R** software (version 2.15.3). We check for violations of the assumption of our regression analyses and find no substantive violations.

Table 1: Descriptive Statistics

Variables	Mean	Std Dev	Min	Max
Edit Longevity	12727.3	39887.8	0	808926.6
Edit Count	741.5	2388.4	0	49665.0
Quarter	12.83	7.66	1	35
Project creation	18.61	6.1	0	45.6
Project scope	1852	11158.2	1	439598.0
Project size	24.55	4.73	1	735
Discussion topics	11.84	3.17	0	589
Mean Tenure	1081.4	452.9	6	3379
Level of controversy	0.17	0.11	0	2.93
Tenure diversity	0.51	2.01	0	1.6
Turnover rate	0.4	0.44	0	7
Density	0.1	0.08	0	0.83
SC Losses	0.29	0.5	0	8

Results

The descriptive statistics and correlation matrix of the study variables are presented in Table 1 and Table 2, respectively. It is obvious from Table 1 that, the amount of work (i.e., Edit Longevity and Edit Count), discussion topics, project scope and size are of reasonable variation and have a heavily right skewed distribution. By including quarter and random effects in the model, we control for the changes in overall trend of activity within WikiProjects and Wikipedia over time. We observe from Table 2 that the correlation between the two dependent variables (i.e., Log2(Edit Longevity) and Log2(Edit Count)) is positive and significant, suggesting that it is reasonable to use the edit count as a measure of the amoung of work done by project members.

Table 3 presents the results of our empirical analysis in an incremental manner, with the upper and lower panel including the results when the edit longevity and edit count is used as the dependent variable, respectively. Model 1 is the baseline model including all control variables, Model 2 adds the linear terms for the independent variables into Model 1, Model 3 adds the quadratic term of SC Losses into Model 2, Model 4a and Model 4b add the interaction terms into Model 3. Comparison of χ^2 reveals that models with interaction terms (Model 4a and Model 4b) fit the data better than the simpler models.

Our results in Model 1 about the significance of control variables are consistent with the findings by (Chen, Ren, and Riedl 2010): Five control variables have significant effects on the amount of work (edit longevity or edit count) done by project members (p<.001). Specifically:

- A project would be more productive if it is of larger scope (0.241 and 0.299) and size (1.386 and 1.132), or if it deals with a less controversial topic (-0.566 and -1.161).
- The negative and significant coefficient for quarter variable (-0.045 and -0.025) suggests that when holding other

Table 2: Correlation Matrix for Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Log2(Edit Longevity)	-													
2. Log2(Edit Count)	0.92**	-												
3. Quarter	0.04**	0.07**	-											
4. Project creation	-0.29**	-0.29**	-0.21**	-										
5. LnProject Scope	0.49**	0.59**	0.49**	-0.22**	-									
6. LnProject size	0.71**	0.75**	-0.00	-0.42**	0.50**	-								
7. LnDiscussion topics	0.60**	0.64**	-0.07**	-0.46**	0.39**	0.77**	-							
8. ScMean Tenure	-0.01	0.01	0.57**	0.08**	0.31**	-0.16**	-0.18**	-						
9. Level of controversy	-0.05**	-0.12**	0.09**	-0.09**	-0.07**	-0.05**	-0.00	0.03**	-					
10. Tenure Diversity	0.26**	0.24**	-0.15**	-0.11**	0.03**	0.39**	0.22**	-0.43**	0.00	-				
11. Turnover rate	-0.15**	-0.17**	0.03**	0.02^{t}	-0.05**	-0.13**	-0.10**	-0.01	0.04**	0.02^{t}	-			
12. Density	-0.06**	-0.07**	-0.22**	0.24**	-0.19**	-0.21**	-0.18**	-0.03**	-0.09**	0.00	0.22**	-		
13. SC Losses	-0.08**	-0.11**	-0.02**	0.00	-0.07**	-0.03**	-0.05**	-0.07**	0.04**	0.04**	0.23**	0.16**	-	
14. SC Losses^2	-0.05**	-0.05**	-0.00	0.07**	-0.03**	-0.10**	-0.06**	0.03**	-0.03**	-0.05**	-0.17**	-0.02**	0.0	-

Significance level: ** p<0.001, * p<0.01, ^{t} p<0.05.

variables constant, projects generally become less productive over time.

- Project creation time has very weak effect on edit longevity (β=-0.045, p>0.1) or edit count (β=0.014, p<0.01) if we observe quarterly productivity for projects over a longer period of time, which is the case in this study (observation over 8 years).
- The average member tenure and tenure diversity are positively and significantly associated with the amount of work, suggesting that projects with high level of tenure diversity enjoy better group outcomes.

In Model 1, the positive and significant coefficient for the discussion topics variable implies that when controlling for other factors, having more discussion topics in project talk pages generally makes a project more productive. Consistent with the results by (Qin, Salter-Townshend, and Cunningham 2013), this finding suggests that the amount of information and knowledge resources exchanging among participants via communication in project talk pages has a positive impact on project productivity and efficiency.

In Model 2, all three independent variables have significant effects on group productivity (p<.001). The coefficient for network density is positive and significant (2.596 and 2.146), suggesting that when controling for other factors, increasing the connectedness of the collaboration network for a project generally improves its productivity. The negative and significant coefficient for turnover rate (-0.478 and -0.345) implies that controlling for other factors, projects with high levels of turnover in general accomplish less work in a quarter. This provides support for hypothesis **Ha**.

Model 3 shows that both the linear and squared term of SC Losses are significant in the expected direction. Specifically, the linear term of social capital losses has a negative main effect (β=-11.722, p<0.001 in the edit longevity model and β=-13.801, p<0.001 in the edit count model) and its quadratic term has a positive effect (β=5.38, p<0.01 for edit longevity and β=3.939, p<0.001 for edit count). As shown in Figure 2, as social capital losses increases from a very low level to a moderate level, the amount of work done by project members declines; after the moderate level, with any further increased in social capital losses from turnover, the amount of work done by project members increases. This finding is consistent with the results by (Shaw et al. 2005). One explanation for this negative curvilinear relationship between social capital losses and group productivity can be found in (Shaw et al. 2005): initially, when social capital losses are low, the loss of structural hole spanners will create the first communication gaps in a network (can also be knowledge gaps or other gaps), which will have a negative effect on group performance; when social capital losses are high, key network relationships have not been well established and maintained, which should be less damaging to group performance. The results provide full support for hypothesis **Hb**.

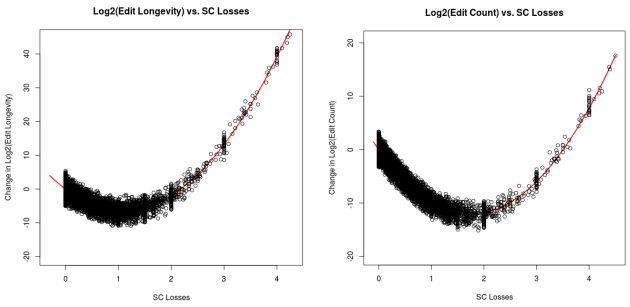

Figure 2: A U-shaped relationship for social capital losses and productivity of WikiProjects. The red lines show the model fit for the relationship of SC Losses to Log2(Edit Longevity) on the left and Log2(Edit Count) on the right. The y-values of the dots correspond to the observed productivity minus the productivity predicted by all other variables. Thus the scatter plots depicts the covariance of productivity (Log2(Edit Longevity) and Log2(Edit Count)) with SC Losses conditional on all other variables.

Hypothesis **Hc1** states that the negative curvilinear relationship between social capital losses and project productivity is more pronounced when turnover rate is low. In Model 4a, the interaction of turnover rate and the linear term of SC Losses is significant in the expected direction, but the interaction with the squared term is insignificant. We plotted

Table 3: HLM results of Predicting Amount of Work Done by Project Members (Using Maximum Likelihood estimates)

Variables	Model 1		Model 2		Model 3		Model 4a		Model 4b	
	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.
Log2(Edit longevity) as dependent variable (#Obs: 16048, #WikiProjects: 890, #Quarters:1–35)										
Intercept	3.493***	0.259	4.115***	0.259	4.121***	0.26	4.119***	0.259	4.122***	0.259
Quarter	-0.045***	0.004	-0.043***	0.004	-0.043***	0.004	-0.043***	0.004	-0.043***	0.004
Project creation	-0.012	0.01	-0.03**	0.01	-0.031**	0.01	-0.03**	0.01	-0.03**	0.01
LnProject scope	0.241***	0.014	0.253***	0.014	0.253***	0.014	0.251***	0.014	0.253***	0.014
LnProject size	1.386***	0.031	1.249***	0.032	1.248***	0.032	1.246***	0.032	1.247***	0.032
LnDiscussion topics	0.16***	0.014	0.159***	0.014	0.159***	0.014	0.161***	0.014	0.161***	0.014
ScMean tenure	0.255***	0.033	0.246***	0.032	0.245***	0.032	0.239***	0.032	0.246***	0.032
Level of controversy	-0.566**	0.164	-0.518**	0.162	-0.516**	0.162	-0.508**	0.162	-0.508**	0.162
Tenure diversity	0.697***	0.101	0.74***	0.1	0.74***	0.1	0.724***	0.1	0.74***	0.1
Turnover rate			-0.478***	0.032	-0.458***	0.033	-0.397***	0.036	-0.461***	0.033
Density†			2.596***	0.19	2.616***	0.19	2.495***	0.193	2.511***	0.194
SC Losses†			-11.341***	1.622	-11.722***	1.626	-5.838**	2.224	-7.476**	2.851
SC Losses^{2}†					5.38**	1.596	3.834^{t}	2.151	0.215	2.827
SC Losses† × Turnover rate							-12.263***	3.288		
SC Losses^{2}† × Turnover rate							2.493	3.699		
SC Losses† × Density†									-34.47*	17.283
SC Losses^{2}† × Density†									40.64*	17.281
Fit										
AIC	61235.0		60862.0		60852.0		60839.0		60848.0	
BIC	61320.0		60969.0		60968.0		60970.0		60979.0	
logLik	-30607.0		-30417.0		-30411.0		-30403.0		-30407.0	
χ^2 / df vs. previous nested model			379.53***		11.36***		17.19***		8.10*	

Variables	Model 1		Model 2		Model 3		Model 4a		Model 4b	
	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.
Log2(Edit count) as dependent variable (#Obs: 16171, #WikiProjects: 891, #Quarters: 1–35)										
Intercept	0.016	0.18	0.474**	0.178	0.48**	0.178	0.475**	0.178	0.484**	0.178
Quarter	-0.025***	0.003	-0.023***	0.003	-0.023***	0.003	-0.023***	0.003	-0.023***	0.003
Project creation	0.014*	0.007	-0.0001	0.007	-0.001	0.007	-0.0003	0.007	-0.0003	0.007
LnProject scope	0.299***	0.01	0.307***	0.01	0.307***	0.01	0.307***	0.01	0.307***	0.01
LnProject size	1.132***	0.022	1.031***	0.022	1.03***	0.022	1.031***	0.022	1.03***	0.022
LnDiscussion topics	0.157***	0.01	0.155***	0.01	0.155***	0.01	0.155***	0.01	0.155***	0.01
ScMean tenure	0.068**	0.022	0.06**	0.022	0.059**	0.022	0.058**	0.022	0.059**	0.022
Level of controversy	-1.161***	0.114	-1.131***	0.112	-1.131***	0.112	-1.129***	0.112	-1.127***	0.112
Tenure diversity	0.14*	0.069	0.174*	0.068	0.175*	0.068	0.168*	0.068	0.172*	0.068
Turnover rate			-0.345***	0.022	-0.332***	0.022	-0.315***	0.025	-0.333***	0.022
Density†			2.146***	0.13	2.164***	0.13	2.132***	0.132	2.103***	0.133
SC Losses†			-13.541***	1.126	-13.801***	1.127	-11.143***	1.547	-13.549***	1.961
SC Losses^{2}†					3.939***	1.096	4.843**	1.671	-0.109	1.97
SC Losses† × Turnover rate							-5.709*	2.266		
SC Losses^{2}† × Turnover rate							-2.404	2.823		
SC Losses† × Density†									-3.978	11.692
SC Losses^{2}† × Density†									32.119*	12.992
Fit										
AIC	50019.0		49473.0		49462.0		49459.0		49460.0	
BIC	50103.0		49580.0		49577.0		49590.0		49550.0	
logLik	-24998.0		-24722.0		-24716		-24713.0		-24713	
χ^2 / df vs. previous nested model			552.03***		12.90***		6.35*		6.11*	

Note. † one quarter lag for network measures.
Significance level: *** p<0.001, ** p<0.01, * p<0.05, ^{t} p<0.1.

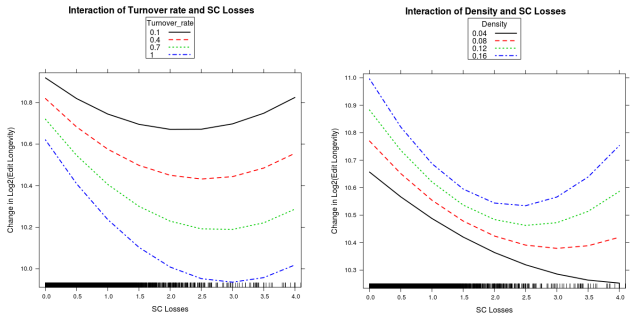

Figure 3: Visualization of the mediated effects of Turnover rate and Network Density on the relationship between social capital losses and project productivity when the edit longevity is the dependent variable. The curves are generated when holding the values of turnover rate (or network density) at a specific level.

the coefficients to examine the level of support for **Hc1**. The left figure in Figure 3 shows this relationship when the edit longevity is the dependent variable[3]. Consistent with (Shaw et al. 2005), we also observe that project performance is the most pronounced when social capital losses and turnover rate are at their lowest levels. The figure shows that, with low turnover rate, the curvilinear relationship is negative at low levels of social capital losses, but this effect becomes attenuated as these losses increase beyond moderate levels. One explanation for this can be explained by (Shaw et al. 2005): Members in long-standing groups generally establish routine, trust and norms of reciprocity that reduce opportunistic behaviour. Thus, the effects of the loss of structural hole bridgers should be more pronounced when turnover is low – a situation in which reciprocity norms and routines are well-established – than when turnover is high, so key network relationships have not been well maintained. The results provide substantial support for **Hc1**.

Hypothesis **Hc2** states that the negative curvilinear relationship between social capital losses and project productivity is more pronounced when network density is high. In Model 4b, the interaction of density and SC Losses is significant ($p<0.05$) in the expected direction when the edit longevity is the dependent variable, but only the interaction with the squared term is significant ($p<0.05$) when the edit count is the dependent variable. The right figure in Figure 3 visualizes this relationship. Different from the left figure in Figure 3, we observe in the right figure that project performance is the highest when social capital losses are at the lowest levels and network density is at higher level. One explanation for this may be about the nature of the collaboration network: a dense collaboration network implies more collaboration among members to work on group tasks, thus promoting trust, reciprocity norms and routines in the group, which has potentially positive effect on group performance. We also observe that as social capital losses increase from low to moderate levels, the slope of the line becomes steeper when density is high than that when density is low. One explanation may be: Dense networks indicate more collaborations among members, so dense networks are more susceptible to social capital losses. Moreover, the overall network density in this study is not very high as we see from Table 1, there is less redundancy of social ties that would make the networks tolerant to disruptions. The figure shows that the negative curvilinear relationship is more pronounced when network density is high. Thus, **Hc2** is partially supported.

[3]We observe a similar trend when the edit count is the dependent variable. The figures are omitted due to space limitation.

Discussion

We have found that there is a negative curvilinear relationship between social capital losses and group productivity in WikiProjects – as social capital losses increase from very low level to moderate level, the amount of work done by project members declines; after the moderate level, with any further increase in social capital losses from turnover, the amount of work done by project members increases.

To further understand the characteristics of quarterly observations that locate at different areas (specifically, the far left, middle and top right area) of the curve in Figure 2, we explore the dataset in more details by tracking the evolution of independent variables and average task distribution over quarters. Following other studies, we quantify different tasks as edits to different types of pages within Wikipedia. The file system of Wikipedia is subdivided into "namespaces"[4] which represent general categories of pages based on their function. Average task distribution is calculated as the average number of edits to a specific namespace by members in a quarter. We observe that healthy WikiProjects generally experience lower levels of member turnover and social capital losses, thus enjoy better group outcomes and task distribution, and tend to locate somewhere between the far left and middle areas of the curve in Figure 2; while unhealthy WikiProjects generally experience more fluctuation in member turnover and social capital losses, have extremely unbalanced task distribution and have quarterly observations distributed over the three areas of the curve in Figure 2.

Figure 4 presents the evolution of edit distribution and independent variables for two projects. Because article edits generally account for more that 50% of members' activity, we omitted it from Figure 4 to improve readability of the graph. The upper two figures shows a substantial difference among their average edit distribution: WP Medicine enjoys better and more balanced task distribution than WP Unionism in Ireland, with the members of the former making edits across a range of namespaces each quarter. The lower two figures reveals that WP Unionism in Ireland experiences substantial fluctuation in its member turnover and social capital losses, while WP Medicine has a more stable evolution in its independent variables.

This work has practical implications for future research and management about membership, participation and social capital in online communities:

- Given that membership turnover has a significant negative effect on the outcomes of voluntary groups, it is essential to study how users shift their participation as they spend

[4]http://en.wikipedia.org/wiki/Wikipedia:Namespace

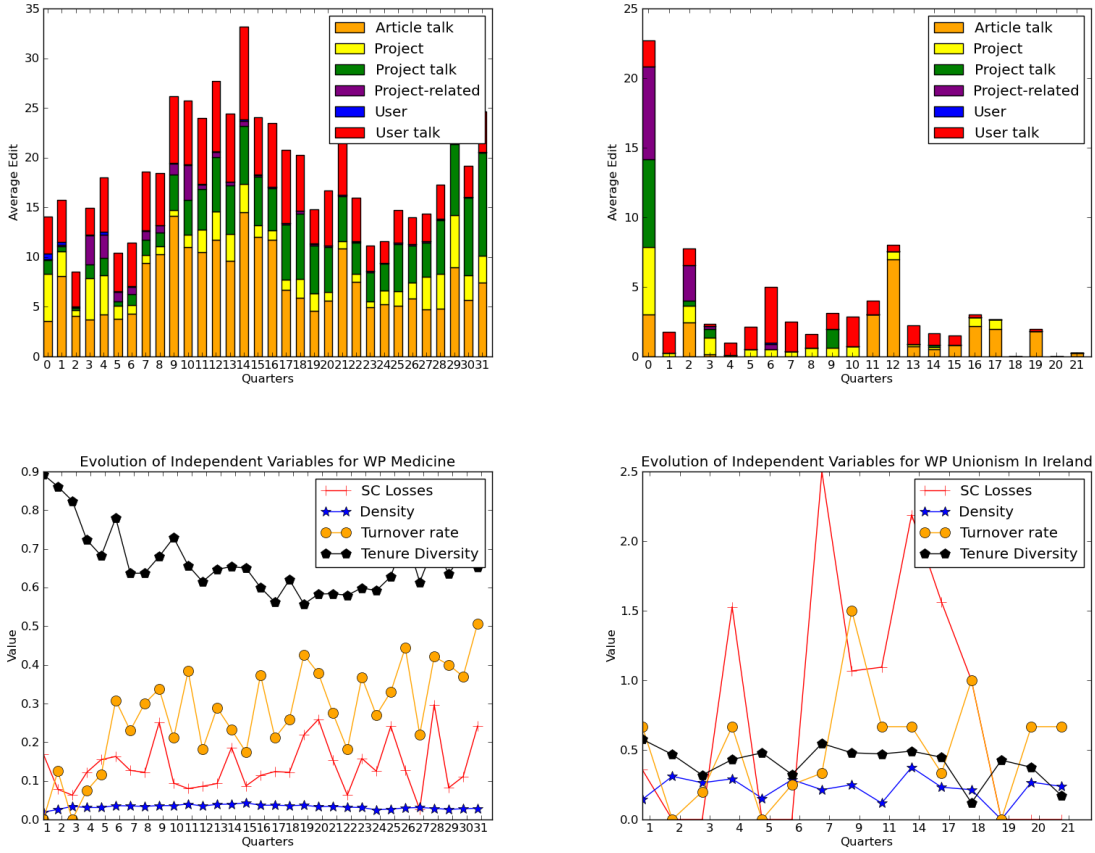

Figure 4: Evolution of Edit Distribution and Independent Variables for a healthy project and a less healthy project.

more time in online platforms and how the communities can fill the gaps created by those turnovers. Modeling user lifecycles could provide insights into issues about the role of each user in the communities and measures taken to improve member retention.

- The findings that the number of discussion topics have a significant positive effect on productivity and that healthy projects generally enjoy better task distribution, suggest that it is beneficial to promote flexible and malleable rules of participation in online platforms.
- Different users generally assume different roles in online platforms, future research can explore the loss of social capital from the perspective of the roles of members and importance of their contributions (e.g., roles in wiki editing defined by (Kostkova and Szomszor 2012) or roles identified by user lifecycle models).

Conclusion

We have empirically examined the impact of membership turnover on productivity in online groups. The results show that membership turnover has significant negative effect on group outcomes in Wikipedia, and that social capital losses through member turnover are associated with group productivity in a negative curvilinear fashion. Our results show that the loss of social capital in collaboration networks can explain significant amount of variation in productivity for WikiProjects. The results contribute to the existing literature that investigates the relationship between social capital losses and group performance (Huselid 1995; Abelson and Baysinger 1984; Shaw et al. 2005; Ransbotham and Kane 2011) in conventional organizations and online communities. The results also support the mediation of turnover rate and network density on the relationship between social capital losses and group productivity.

Previous studies have suggested that editors with different tenure are likely to perform different tasks (Bryant, Forte, and Bruckman 2005), and that groups consisting of members with diversified tenure generally enjoy better task distribution and group performance (Chen, Ren, and Riedl 2010). Similarly, our exploration of the dataset reveals that projects with low levels of social capital losses and turnover rate generally enjoy better task distribution and group outcomes. It will be very interesting to investigate the impact of task distribution among members on group performance and to study the evolution of user behaviours in online communities. Furthermore, this study only considers social capital losses in collaboration networks, it will be worthwhile to

take into account social capital losses in communication networks for WikiProjects and in article talk networks and user talk networks.

Acknowledgements

This work is supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289 (Insight Centre). Xiangju Qin is funded by University College Dublin and China Scholarship Council (UCD-CSC Scholarship Scheme 2011). The authors would like to thank the three anonymous reviewers for their constructive comments to improve the quality of this paper.

References

Abelson, M. A., and Baysinger, B. D. 1984. Optimal and dysfunctional turnover: Toward an organizational level model. *Academy of Management Review* 9:331–341.

Adler, B. T., and Alfaro, L. D. 2007. A Content-Driven Reputation System for the Wikipedia. In *WWW*, 261–270.

Arguello, J.; Butler, B. S.; Joyce, E.; Kraut, R.; Ling, K. S.; Rose, C.; and Wang, X. 2006. Talk to Me: Foundations for Successful IndividualGroup Interactions in Online Communities. In *CHI*, 959–968.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge University Press.

Bates, D.; Maechler, M.; and Bolker, B. 2012. *lme4: Linear mixed-effects models using S4 classes*.

Bryant, S. L.; Forte, A.; and Bruckman, A. 2005. Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. In *GROUP*, 1–10.

Bryk, A. S., and Raudenbush, S. W. 1992. *Hierarchical Linear Models: Applications and Data Analysis Method*. Sage Publications.

Burt, R. S. 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press.

Chen, J.; Ren, Y.; and Riedl, J. 2010. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *CHI*, 821–830.

Coleman, J. S. 1988. Social capital in the creation of human capital. *The American Journal of Sociology* 94:S95–S120.

Dabbish, L.; Farzan, R.; Kraut, R.; and Postmes, T. 2012. Fresh faces in the crowd: turnover, identity, and commitment in online groups. In *CSCW*, 245–248.

Faraj, S.; Jarvenpaa, S.; and Majchrzak, A. 2011. Knowledge Collaboration in Online Communities. *Organization Science* 22(5):1224–1239.

Freeman, L. 1979. Centrality in Social Networks: Conceptual Clarification. *Social Networks* 1(3):215–239.

Gelman, A., and Hill, J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. London: Cambridge University Press.

Harrison, D., and Klein, K. 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *AMR* 32:1199–1228.

Huselid, M. A. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *AMJ* 38(3):635–672.

Kane, G. C., and Fichman, R. G. 2009. The Shoemakers Children: Using Wikis for IS Research, Teaching, and Publication. *MIS Quarterly* 33(1):1–22.

Kivran-Swaine, F.; Govindan, P.; and Naaman, M. 2011. The impact of network structure on breaking ties in online social networks: Unfollowing on twitter. In *CHI*, 1101–1104.

Kostkova, P., and Szomszor, M. 2012. The FEM Wiki Project: A Conversion of a Training Resource for Field Epidemiologists into a Collaborative Web 2.0 Portal. In *E-Health*, volume 69 of *LNICST*. 119–126.

Krackhardt, D., and Porter, L. W. 1985. When friends leave: A structural analysis of the relationship between turnover and stayers attitudes. *ASQ* 30:242–261.

Lewis, K. M.; Belliveau, B.; and Herndon, J. K. 2007. Group cognition, membership change, and performance: Investigating the benefits and detriments of collective knowledge. *Organ Behav Hum Decis Process* 103(2):159–178.

Madsen, T. L.; Mosakowski, E.; and Zaheer, S. 2003. Knowledge retention and personnel mobility: The nondisruptive effects of inflows of experience. *Organization Science* 14(2):173–191.

Mitchell, T. R.; Holtom, B. C.; Lee, T. W.; Sablynski, C. J.; and ErezSource, M. 2001. Why People Stay: Using Job Embeddedness to Predict Voluntary Turnover. *AMJ* 44(6):1102–1121.

Panciera, K.; Halfaker, A.; and Terveen, L. 2009. Wikipedians are born, not made. In *GROUP*, 51–60.

Portes, A. 1998. Social Capital: Its Origins and Applications in Modern Sociology. *Annual Review of Sociology* 24:1–24.

Qin, X.; Salter-Townshend, M.; and Cunningham, P. 2013. The Influence of Network Structures of Wikipedia Discussion Pages on the Efficiency of WikiProjects. *Manuscript under review* 1–19.

Ransbotham, S., and Kane, G. 2011. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in wikipedia. *MIS Quarterly* 35(3):613–627.

Shaw, J.; Duffy, M. K.; Johnson, J. L.; and Lockhart, D. E. 2005. Turnover, social capital losses, and performance. *Academy of Management Journal* 48(4):594–606.

Wagner, C., and Majchrzak, A. 2006. Enabling Customer - Centricity Using Wikis and the Wiki Way. *Journal of Management Information Systems* 23(3):17–43.

Wang, X., and Lantzy, S. 2011. A Systematic Examination of Member Turnover and Online Community Health. In *ICIS*, 25:1–25:10.

Zanetti, M. S.; Sarigöl, E.; Scholtes, I.; Tessone, C. J.; and Schweitzer, F. 2012. A Quantitative Study of Social Organisation in Open Source Software Communities. In *ICCSW*, 116–122.