Out With The Old, In With The New? Unpacking Member Turnover in Online Production Groups

BOWEN YU, University of Minnesota - Twin Cities
XINYI WANG, University of Minnesota - Twin Cities
ALLEN YILUN LIN *, Northwestern University
YUQING REN, University of Minnesota - Twin Cities
LOREN TERVEEN, University of Minnesota - Twin Cities
HAIYI ZHU, University of Minnesota - Twin Cities

Nearly any group is subject to *turnover*: some people leave, while others join. Turnover can be especially high in online groups, since participation typically is strictly voluntary. We investigated the effects of member turnover in online groups, specifically in Wikipedia's WikiProjects. We based our studies on theories from organizational science, which suggest that it is not just the amount of turnover, but the *characteristics* of those leaving and those joining that matter. We characterized leavers and newcomers by their prior productivity, tenure (in the group or community), and participation in other groups within the larger community. Furthermore, we considered the moderating effect of group size on turnover. We analyzed data from 88,427 editors who participated in 1,054 WikiProjects, finding that (1) the positive effects of newcomers to a group were larger than the negative effects of leavers, (2) prior productivity, tenure, and participation in other groups all played significant roles, and (3) the effects of leavers and newcomers were amplified in larger groups.

CCS Concepts: • **Human-centered computing → Computer supported cooperative work**;

Additional Key Words and Phrases: Online Communities; Peer Production; Online Groups; Wikipedia; WikiProjects; Membership Turnover; Leavers; Newcomers; Group Size; Group Performance

ACM Reference format:
Bowen Yu, Xinyi Wang, Allen Yilun Lin *, Yuqing Ren, Loren Terveen, and Haiyi Zhu. 2017. Out With The Old, In With The New? Unpacking Member Turnover in Online Production Groups. *Proc. ACM Hum.-Comput. Interact.* 1, 2, Article 117 (November 2017), 19 pages.
https://doi.org/10.1145/3134752

1 INTRODUCTION

Peer production is now a well-established and successful method of knowledge production, with Wikipedia and open source software systems such as Linux and Apache as prime examples. Like all online communities, participation in such communities is strictly voluntary, with people entering

* Allen Yilun Lin contributed to the project when he was pursuing his Master Degree at the University of Minnesota.
University of Minnesota researchers were supported by the College of Science and Engineering, University of Minnesota.
Authors' addresses: Bowen Yu, Xinyi Wang, Loren Terveen, Haiyi Zhu: College of Science and Engineering, University of Minnesota - Twin Cities, 200 Union St SE, Minneapolis, MN, 55414, US; Allen Yilun Lin: Northwestern University, 2240 Campus Road, Evanston, IL 60208, US; Yuqing Ren: Carlson School, University of Minnesota, 321 19th Ave S, Minneapolis, MN, 55414, US.

2573-0142/2017/11-ART117
https://doi.org/10.1145/3134752

and leaving continuously. For example, in an average month of the English Wikipedia, 38,628 users register new accounts and 38,138 users leave the site[1]. In the MovieLens film recommendation community, 60% of new users never returned after one week [27]. The standard term for this change of membership in a group over time is *membership turnover* [4].

Membership turnover is ubiquitous in nearly all groups, for example, social groups, work teams, non-profit organizations, as well as online communities. Therefore, its antecedents and consequences have been studied extensively in the organizational science literature [7, 11]. A common theme in the literature is the negative consequences of turnover on groups' productivity due to the loss of experienced and often productive members [45] who have had central roles in the organization's social networks [14], and the cost of training new and inexperienced members [13]. Some studies have also shown the positive effects of turnover, such as helping screen out underperforming members [29].

Another perspective, as represented in Hausknecht and Holwerda [21], suggests that the level of turnover may not be as important as the specific details of the turnover: *who leaves, who stays, and who enters to replace the leavers*. Thus, the same level of turnover could have different consequences depending on these details. For example, turnover would be more damaging when losing a critical manager in the organization compared to losing a regular employee.

We adopt the last perspective to study the effects of membership turnover in the context of *WikiProjects*. WikiProjects are self-organized groups of editors within Wikipedia that are devoted to improve the quality of articles on specific topics. As a widely-studied platform, Wikipedia provides rich data archives that allows us to track editors' editing records and to further investigate the dynamics between editors and WikiProjects. We analyzed 1,054 WikiProjects in which 88,427 editors contributed a total of 44,135,006 edits over 14 years. We studied both leavers and newcomers, and characterized both in terms of their *productivity*, *tenure*, and *additional commitments* (i.e., membership in 'other' WikiProjects). We analyzed the effects of leavers and newcomers on project performance, including *project productivity* (i.e., the amount of editing on the project's topic) and *project planning* (i.e., the amount of intra-group planning activities). We also examined how group size moderates the effects of leavers and newcomers on project performance.

As expected, we found that the ratio of leavers was negatively associated with project performance and that the ratio of newcomers was positively associated with project performance. More importantly, we also found that (1) overall, the positive effects of newcomers dominated the negative effects of leavers; (2) the negative effects of leavers were greater when the leavers were more productive, had shorter project tenure, and had fewer additional commitments; (3) the positive effects of newcomers were greater when the newcomers were less productive within Wikipedia prior to joining the projects, had longer Wikipedia tenure, and had fewer additional commitments, and (4) group size was a significant moderator: large groups' performance suffered more from the same ratio of leavers and benefited more from the same ratio of newcomers. Overall, our findings advance our knowledge of the impact of membership turnover in online peer production communities and provide practical implications for online community design.

The rest of the paper is organized as follows: (1) we review organizational science and online community literature on member turnover and use this literature to ground our hypotheses, (2) we describe the research setting of Wikipedia and WikiProjects along with our data and analyses, and (3) we present our results and discuss their implications for effective management of online peer production communities.

[1]On average, 30,012 users leave within their first month, and 8,316 users had been active more than one month. These numbers are for the 175 months from January 2001 to June 2015 (the range of our dataset).

2 RELATED WORK

The effects of employee turnover on the performance of groups and organizations have received much attention [21]. In recent years, scholars have also begun to study the effects of turnover in online settings due to the rapid growth and ubiquity of online communities [42]. Overall, scholars have drawn mixed conclusions about the effects of turnover with some studies showing it to be harmful [25, 46] and others showing it to be beneficial [18, 29].

The more dominant view argues that turnover is detrimental to organizational performance for a couple reasons. When people leave, organizational resources are needed to recruit and train new employees to replace those who leave [13]. Departing employees can not only take away their human capital (e.g., knowledge, skills, and experience) [6] but may also cause social capital losses (e.g., disrupt the social networks or work environment for the remaining people) [14]. Meanwhile, the training and socialization of newcomers impose additional costs and risks. If not appropriately socialized, newcomers may experience reality shock [33] and cause distraction in the organizations. The online community literature generally emphasizes the importance of sustaining a steady group of members to the success of online communities [31]. Members who stick with the communities help with community activities and sustain the group through difficult times. They are those who are most likely to create content that others value, such as answers to people's questions in technical communities [3], informational and emotional support in online health support groups [50], software code in open source projects [34], and edits and feedback in Wikipedia [28, 55]. Similarly, when members leave the community, it also depletes the collective-level human capital and social capital. Therefore, turnover diminishes the resources available to the community and may threaten its sustainability.

The second view on turnover highlights the potential benefit of turnover when it serves as a screening function to filter out employees who are unproductive or dissatisfied with the current organization [29]. Prior studies have also identified turnover to be a valuable mobility process that vitalizes the organization and enables it to adapt to internal demands and environmental pressure [19]. In the online environment, Kane and Alavi suggested that certain technology-enabled features, such as preserving the previous contributions and meta-data in a searchable fashion, can retain the knowledge in case of member leaving [26]. When the entire collaboration history of an online production collaboration is preserved, subsequently joined members can review the history to get a good sense of the norms and past decision processes, which may help mitigate the negative effects of turnover.

The third view on turnover suggests that moderate levels of turnover are the most optimal for organizational performance. The rationale is when organizations hire new people to replace those who have left, the new employees will bring new skills and knowledge to update the existing skills and knowledge that might have been obsolete or stagnant [44]. Similarly, recent studies of turnover in online production communities have also suggested that certain levels of turnover can help online communities thrive. For example, Ransbotham and Kane [42] and Qin et al [41] showed that moderate levels of membership turnover would benefit content creation, content retention, and group productivity in Wikipedia collaboration. The reason is that moderate levels of turnover create a balance between incorporating new information and skills and preserving those the community already possesses.

To reconcile the mixed findings around the effects of turnover, Hausknecht and Holwerda proposed a new conceptualization of turnover [21]. Specifically, they argue that the traditional aggregated measures of turnover as "turnover rates", although valuable, "conceal variation in key causal factors that ultimately determine how turnover shapes performance". They recommend that, to understand the consequence of turnover, it is crucial to unpack the fundamental structure and

properties of turnover, particularly who leaves, who stays, and who enters to replace the leavers [22].

In this paper, we followed Hausknecht and Holwerda's suggestion to examine the effects of turnover in subgroups embedded in the large online production community of Wikipedia. Our hypotheses focus on not only the ratio of leavers and newcomers, but more importantly the attributes of the leavers and newcomers, such as *productivity* (i.e., the amount of work the member has done in the past), *tenure* (i.e., the amount of time the member has been participating in the community or subgroup), and *additional commitments* (i.e., the number of other subgroups the member participated in simultaneously), and how they might affect the group performance.

Prior studies have explored different factors that can affect group performance. For example, Chen et al [8] and Vasilescu et al [48] found that group diversity – in terms of member experience, tenure, etc. – is a strong predictor of group productivity and member retention in peer production communities. Lam and colleague [32] showed that group composition can influence group decision quality. Our work, however, studies how the change of group composition (i.e., specific types of people joining and leaving) affects group productivity. In the following sections, we explain the theoretical reasoning behind our hypotheses.

3 HYPOTHESES

3.1 Effects of Different Types of Leavers

We first consider the effects of leaver attributes: past productivity and tenure in the subgroup, and additional commitments at the time they left the subgroup.

Productivity. We hypothesize that the departure of members who are more productive in the group should be more detrimental than the departure of less productive members for two reasons. First, productive members take on more responsibilities on creating and maintaining the content [39]. When productive members leave, the group loses their valuable human capital to do the work and need to recruit and socialize new members to replace them. Second, productive members often are more socially embedded and communicate more with other group members [38]. The departure of well-connected central members might disrupt the social networks and the important social capital to facilitate group collaboration [30]. In contrast, the departure of unproductive members are less impactful on group performance because these members make relatively fewer contributions and communicate with other members less frequently. Therefore, we hypothesize.

Hypothesis 1a: *The departure of more productive members has a larger negative impact on group performance than the departure of less productive members.*

Tenure. We speculate that the departure of members who have been with the group for a longer period of time harms the group performance more than those who have been with the group a shorter period of time. First, long-term engaged members (old-timers) gain knowledge and experience through participation and observation. Their knowledge and experience is invaluable to the group's productivity and sustainability. Those members often play an important role in socializing new members, by answering questions, sharing relevant resources, and mentoring newcomers. In contrast, we expect the departure of members with a short tenure with the group has a less negative consequence on group performance because they in general play a marginal role and have not made any substantive contributions to group performance. Moreover, their early leaving might also be a signal of lack of fit with the group [40].

Hypothesis 1b: *The departure of members who have been with the group for a longer period of time has a larger negative impact on group performance than the departure of members who have been with the group for a shorter period of time.*

Additional commitments. Lastly, we hypothesize the departure of members without additional commitments will impact the group more than those with additional commitments. Many online communities are consist of subgroups, and it is common for a member to belong to multiple subgroups. For example, a Wikipedia editor may belong to more than one WikiProject. We consider involvement in other subgroups, in addition to the focal subgroup, as additional commitments. Individual members have limited time and resources and when a member belongs to multiple subgroups, these subgroups need to compete with each other for members' time and effort [49]. Additional commitments or involvement in other groups distract members and limit the time and attention they can spare to help the focal group [23]. Therefore, the departure of members who do not have additional commitments has a great impact on group performance; whereas members who have additional commitments are already allocating their time and effort among multiple groups and their leaving should have less of an impact on group performance.

Hypothesis 1c: *The departure of members who have no additional commitments has a larger negative impact on group performance than the departure of members who have additional commitments.*

3.2 Effects of Different Types of Newcomers

Our next set of hypotheses focuses on the impact of newcomer attributes, such as community-level productivity, tenure before joining, and additional commitments at the time they joined the group, on group performance.

Productivity. The impact of newcomer joining depends on the quality of newcomers. Although the newcomers are new additions to the subgroup, they may not be new to the large community. In the case of Wikipedia, an editor may have been with Wikipedia for a while before joining any WikiProjects. Therefore, we get to observe their levels of productivity prior to joining the subgroups. We hypothesize that newcomers who were productive before joining the subgroup are more likely to continue to be productive and take on more crucial tasks in the group [40]. Therefore, the joining of productive newcomers should benefit group performance more than nonproductive newcomers.

Hypothesis 2a: *The joining of newcomers who were more productive before joining the subgroup has a larger positive impact on group performance than newcomers who were less productive before joining.*

Tenure. Newcomers often differ not only in their prior productivity, but also in their community tenure and experience. We posit that newcomers with a longer tenure, and hence more experience with the large community, will benefit group performance more than newcomers with a shorter tenure or less experience. In both online and offline groups, considerable training and socialization are required to help new members to adjust to a new environment [1, 9]. Prior experience with the large community should help newcomers learn group-specific knowledge more quickly and lead to a smoother transition. Therefore, we predict that the joining of experienced newcomers is more beneficial to online groups than inexperienced newcomers.

Hypothesis 2b: *The joining of newcomers who have more community-level experience has a larger positive impact on group performance than newcomers with less community-level experience.*

Additional commitments. In terms of the effects of the additional commitments, we believe the same logic for leavers can be applied to newcomers. When newcomers do not belong to any other subgroups, they are more focused and contribute more, compared to newcomers who belong to multiple subgroups simultaneously. Therefore, newcomers without additional commitments should benefit group performance more than newcomers with additional commitments.

Hypothesis 2c: *The joining of newcomers without additional commitments has a larger positive impact on group performance than newcomers with additional commitments.*

3.3 Moderating Effects of Group Size on Turnover

One relevant factor that influences the effects of group input on group output is group size [12, 17]. Previous research has shown that group size can moderate the effects of turnover [7].

Compared to small groups, large groups can be more resilient in coping with the departure of a larger proportion of members [37]. As large groups tend to have more resources and more members, the departure of some members can be compensated by the remaining members with similar skills or knowledge.

Hypothesis 3a: *The negative effect of a larger proportion of member leaving on group performance is smaller in larger groups than in smaller groups.*

We also speculate that large groups can benefit more from a larger proportion of newcomers than small groups. Large groups possess more available resources to socialize and train newcomers, which is critical when newcomers enter unfamiliar environments [33]. Therefore, large groups can be more effective at incorporating a large proportion of newcomers than small groups.

Hypothesis 3b: *The positive effect of a larger proportion of newcomer joining on group performance is larger in larger groups than in smaller groups.*

4 METHODS

4.1 Study Platform: Wikipedia and WikiProjects

We used WikiProjects within Wikipedia as our study platform. As a prominent example of online production communities, Wikipedia has attracted millions of editors to contribute to the largest online encyclopedia in the world. In Wikipedia, WikiProjects provide places for Wikipedia editors with similar topic interests to collaborate and coordinate their work[2]. Each WikiProject has a project page and a project talk page. Project pages are used to maintain the overall status of the project, including the content coverage of the articles claimed within the scope of the project, member guidelines, membership lists, task lists, etc. Project talk pages are used for project members to communicate with each other and discuss project plans.

We believe WikiProjects provide an appropriate setting for our study. First, WikiProjects is a well-established example of online production subgroups within a larger community. Given the similar voluntary nature and structure among those communities, we believe the findings can be applicable and generalized to other groups, such as projects in GitHub, projects in OpenStreetMap, etc. Second, the availability of high-quality data archives enables us to derive rich measures of member attributes and group performance. For example, we can measure editor activities not only inside and outside the groups, but also before and after joining the groups. This enables us to unpack the blackbox of turnover and examine the specific attributes of member leaving and joining.

4.2 Data Collection

Our study used the English Wikipedia data dump of June 2, 2015 provided by the Wikimedia Foundation[3], which includes the complete editing history of all articles in the English Wikipedia. We used an open source Python package to parse the dump files[4]. To construct our sample of WikiProjects, we used the project templates on articles' talk pages, which provided information about which WikiProjects an article belongs to (i.e., some articles belong to multiple WikiProjects). We excluded projects that never grew to more than three members in any of their active time periods, which is the minimal size of a group (details about the time period construction will be explained in the next section). We further excluded projects that are not topical but rather

[2]https://en.wikipedia.org/wiki/WikiProject
[3]https://dumps.wikimedia.org/
[4]https://mwparserfromhell.readthedocs.io/en/latest/

serve a collaboration purpose, e.g., WikiProject: Manual of Style. Our final sample included 1,054 WikiProjects with an average of 6,368 articles per project.

4.3 Operationalizations

4.3.1 Longitudinal Data Generation. Turnover is a process that happens continuously over time. Therefore, we needed a longitudinal dataset to investigate its effects. We divided our dataset into 3-month time periods from the inception of the first WikiProject in January 2001 to the end of our dataset in June 2015, resulting in 62 time periods. We followed previous work [41] and chose three months as the time window. At the same time, each WikiProject has its own life cycles that span some time periods between 2001 and 2015. The average lifespan of a project is 25.7 time periods, or about six and a half years.

We computed all of our independent variables, dependent variables, and control variables for each project in all time periods during which the project was active.

4.3.2 Types of Editors. We defined three types of editors by examining data for the current 3-month analytic period and the immediately preceding 3-month period.

Leavers were active in the previous time period, but not in the current period. We defined an editor as active in a time period if the editor made any edits on the project page or the project talk page, or made at least 5 edits on the articles belonging to the project. Any of these types of activities indicate project activity, either directly improving relevant articles or else communicating about the project generally.

Remaining members were active in both the previous and current time periods (using the same notion of an active editor).

Newcomers were active in the current time period, but not the previous period. Here we took a slightly more restrictive approach, following prior research to operationalize the notion of *becoming* active or *joining* a project [35, 53]. Specifically, we defined an editor joining a project in a time period only if the editor made his or her first edit to the project page or project talk page (having more than five edits on the relevant article pages is no longer a criteria), because we considered joining a project to be more intentional and explicit that required a clear "sign-up" action compared to leaving a project.

4.3.3 Independent Variables. We built four sets of models to explore the effects of different turnover factors on project productivity and planning. The independent variables for the models represented leavers from and newcomers to projects. For both of these types of editors we operationalized the three factors we described earlier: their *productivity*, *tenure*, and *additional commitments*. Moreover, we used the *ratio* of each type of editors over the total number of project members in a project during a time period to normalize for project size; intuitively for example, when 5 out of 10 editors leave a project, the effect is likely to be significantly different than when 5 out of 100 leave. Therefore, *all* our independent variables represent *ratios*, so each variable descriptions should be understood to end as follows: "... divided by the total number of project members at the given time period".

Basic variables

- **Leavers**: the number of leavers.
- **Newcomers**: the number of newcomers.

Variables for Leavers

- **Productive leavers** were those whose accumulated activity level in the project was *above* the median activity level of remaining project members (Recall that we included edits on the project page, project talk page, and articles belonging to the project in computing activity).

As a reminder, for our analysis we used the ratio of this number to the number of project members.

- **Unproductive leavers** were those whose accumulated activity level in the project was *below* the median activity level of remaining project members.
- **Short-term leavers** were editors who stayed in a project for only one time period; in other words, they joined the project in the previous time period and left (were not active) in the current time period.
- **Long-term leavers** were editors who left the project in the current time period and had been active for more than one previous periods.
- **Leavers with additional commitments** were leavers who were also members of at least one other WikiProjects.
- **Leavers without additional commitments** were leavers who were not members of any other WikiProjects.

Variables for Newcomers

- **Productive newcomers** were those whose cumulative activity level in Wikipedia was *above* the median activity level of remaining project members.
- **Unproductive newcomers**: were those whose cumulative activity level in Wikipedia was *below* the median activity level of remaining project members.
- **Experienced newcomers** were newcomers who had been active in Wikipedia for more than one time period before joining the project.
- **Inexperienced newcomers** were newcomers who were also *new to Wikipedia*; that is, they joined Wikipedia and the project in the same time period.
- **Newcomers with additional commitments** were newcomers who were members of at least one other WikiProject.
- **Newcomers without additional commitments** were newcomers who were not members of any other WikiProjects.

4.3.4 *Control Variables.*

- **Wikipedia time index**: the number of quarters (3-month periods) between the creation of Wikipedia and the current time period.
- **Project size**: the number of all active members (i.e., remaining members and newcomers) in the project in the current time period.
- **Project age**: the number of quarters (3-month periods) between the creation of the project and the current time period.
- **Project prior productivity**: the number of edits in the previous time period to the articles within the project scope; this variable is used for models with project productivity as the dependent variable.
- **Project prior planning**: the number of edits in the previous time period to the project page and the project talk page; this variable is used for models with project planning as the dependent variable.

4.3.5 Dependent Variables. We defined two measures of group performance: (1) *project productivity*, the number of edits project members made to articles within the project scope – these represent the actual *work* of the project; and (2) *project planning*, the number of edits to the project page and project talk page – these represent *talking about* the project work. We operationalized as variables representing the percentage change from the previous period to the current period to normalize for varying amounts of activity. Therefore, a positive effect refers to an increase

Variables (Member Proportion)	Median	Mean	Std. Dev.
Leavers	16.7%	17.7%	13.1%
Productive Leavers	4.0%	5.9%	7.5%
Unproductive Leavers	10.9%	11.8%	10.0%
Short-Term Leavers	2.8%	5.3%	7.8%
Long-Term Leavers	11.1%	12.4%	11.2%
Leavers with Commitment	14.3%	15.6%	12.3%
Leavers without Commitment	0.0%	2.1%	4.7%
Newcomers	18.2%	21.9%	14.6%
Productive Newcomers	5.9%	8.9%	10.8%
Unproductive Newcomers	10.2%	13.0%	12.0%
Experienced Newcomers	16.7%	19.6%	13.8%
Inexperienced Newcomers	0.0%	2.3%	5.2%
Newcomers with Commitment	11.1%	13.9%	11.8%
Newcomers without Commitment	4.8%	8.0%	10.4%

Table 1. The descriptive statistics show the median, mean, and standard deviation of the proportion of each type of members for all the projects over all the time periods. For example, the median of the ratio of leavers for a project in a time period was 16.7% and the mean was 17.7%. Note that the range of these median and mean values is from 0.0% to 100.0%.

in productivity or planning in the current time period compared to the previous period, while a negative effect refers to a decrease.

- **Project productivity change**. We counted the total number of edits made by project members to articles within the project scope in each time period. Thus, the project productivity change for a time period *T* is the percentage change in number of article edits from period *T-1* to period *T*.
- **Project planning change**. Similarly, we calculated the total number of edits by project members to the project page and project talk page in each time period. Thus, the project planning change for a period *T* is the percentage change in the number of these edits from period *T-1* to period *T*.

4.3.6 Data Preprocessing. We identified outliers by examining the regression residuals. However, whether or not we included outliers had no impact on our results – either statistical significance or effect sizes – so we included all the data in our models. Because many of the data were skewed, we performed base-2 logarithmic transformations on all the variables. Therefore, the coefficients in the regression results are interpreted as the ratio of the change in the dependent variable in percentage when an independent variable increases by one unit (in our case, one unit is one percent due to the log-log transformation).

5 RESULTS

5.1 Descriptive Statistics

Our dataset contains 1,054 projects and 88,427 unique editors. Table 1 shows descriptive statistics for all the independent variables before transformations, i.e., the proportion of each type of members of a project in a time period.

5.2 Statistical Analysis

The level of analysis is WikiProject-period with three months as one period. We ran linear regression models on the panel data. Because the analysis also compared different member ratios of the same project during different time periods, we added in random effects with each WikiProject as a group. We used the package in Stata for our regression models[5].

We included *within* R-squared value (i.e., the variance within the panel units), *between* R-squared value (i.e., the variance between separate panel units), and *overall* R-squared value (i.e., a weighted average of these two) to estimate the explained variation of our models [43]. The R-squared values indicated that our models had good coverage[6] (Table 2, Table 3, and Table 4). In addition, we accessed the significance of the incremental variance explained by each model compared to the baseline by performing the likelihood-ratio test. The results suggest that the models that include independent variables explained more variance and had much better fits than the baseline models that only contain control variables.

From a methodological perspective, it is interesting to note that we used the "entire" population for our study – that is, all WikiProjects – rather than a sample. Thus, one could ask: are p-values meaningful, since they indicate uncertainty whether results for a sample apply to the entire population? However, social scientists still commonly apply statistical methods (and report p-values) even when analyzing a "full" sample, for example, when using the US tax return dataset to study the effects of education on income [36]. The rationale for doing this builds on the notion of a "super population" [15, 16, 20]. In our case, the super population would include all existing and future online production groups (including WikiProjects, OpenStreetMap projects, open source software development projects, etc), and p-values can be interpreted as the likelihood of seeing the same results if we conducted the same analysis on an updated sample of WikiProjects (e.g., through 2017), or on a sample (or the full population) of other online production groups.

5.3 Baseline Models

We first present our baseline models which examine the effects of our control variables on project productivity (Model 1) and project planning (Model 6)[7].

Model 1 shows that project age and project productivity in the previous time period are negatively associated with change in project productivity. Older projects (Coef = -0.23, p <.001) and projects that were more productive in the previous time period (Coef = -0.45, p <.001), tended to have negative changes in productivity in the current time period. Wikipedia time index (i.e., the time since the creation of Wikipedia) is negatively associated with project productivity and project size is positively associated with project productivity, but both are not statistical significant. Model 6 shows that these predictors have similar effects on project planning with an additional significant negative effect from Wikipedia time index (Coef = -0.30, p <.001) and an additional significant positive effect from project size (Coef = 0.14, p <.001).

Next, we add the average leaver and newcomer ratios to examine the overall effects of member turnover (Model 2 and Model 7). An important observation from Model 2 is that the positive impact of newcomers outweighed the negative impact of leavers on project productivity. While both variables had significant effects, the negative effect of leavers (Coef = -0.43, p <.001) was less than the positive effect of newcomers (Coef = 1.92, p <.001). A one-percent increase in the ratio of leavers decreased the change in project productivity by -0.43%, while a one-percent increase in

[5] www.stata.com/manuals14/xtxtreg.pdf

[6] In Social Science studies, R-Squared values around 0.3-0.4 are acceptable.

[7] Note that Model 1 and 6, and Model 2 and 7 in Table 2 are identical to Model 1 and 6, and Model 2 and 7 in Table 3. We include the baseline models in both tables for the purpose of comparison.

	Project Productivity Change										Project Planning Change									
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8		Model 9		Model 10	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Intercept	3.37***	.26	1.16***	.27	1.21***	.26	1.24***	.26	0.48	.30	3.12***	.25	0.00	.24	0.50*	.27	0.55***	.26	-1.65***	.27
Wikipedia Time Index	-0.21	.06	0.20***	.06	0.18**	.06	0.17**	.06	0.09**	.06	-0.30***	.06	-0.05	.06	-0.05	.02	-0.06	.06	0.09	.06
Project Age	-0.23***	.03	-0.15***	.26	-0.13***	.02	-0.14***	.02	-0.12***	.03	-0.12***	.03	-0.07**	.02	-0.02	.02	-0.02***	.02	-0.02	.02
Project Size	0.04	.02	0.18***	.02	0.15***	.02	0.15***	.02	0.26***	.02	0.14***	.02	0.38***	.02	0.29***	.02	0.287***	.02	0.58***	.02
Prior Prod./Planning	-0.45***	.01	-0.48***	.01	-0.45***	.01	-0.45***	.01	-0.47***	.01	-0.49***	.01	-0.59***	.01	-0.52***	.01	-0.52***	.26	-0.61***	.01
Newcomers			1.92***	.10	1.95***	.10	1.96***	.10	2.65***	.13			2.22***	.08	2.08***	.11	2.11***	.11	3.80***	.11
Leavers			-0.43***	.11									-0.85***	.09						
Productive Leavers					-0.69***	.17									-0.90***	.17				
Unproductive Leavers					-0.21*	.13									-0.69***	.13				
Short-Term Leavers							-0.57***	.10									-1.28***	.19		
Long-Term Leavers							-0.30**	.12									-0.65***	.12		
Leavers with Com.									-0.59***	.14									-1.00***	.11
Leavers without Com.									-0.86**	.30									-2.03***	.29
Within R-Sq	0.29		0.32		0.32		0.32		0.31		0.19		0.27		0.29		0.30		0.30	
Between R-Sq	0.59		0.61		0.61		0.61		0.61		0.35		0.41		0.38		0.38		0.38	
Overall R-Sq	0.35		0.39		0.39		0.39		0.38		0.20		0.28		0.30		0.30		0.30	
Likelihood-ratio Test			633.68 ***		634.80 ***		633.99 ***		635.40 ***				1659.26 ***		1658.76 ***		1666.41 ***		1664.80 ***	

Table 2. Results of the Effects of Different Types of Leavers on Project Productivity and Project Planning Changes. $^{*}p$ <.05, $^{**}p$ <.01, $^{***}p$ <0.001

the ratio of newcomers increased the change in project productivity by 1.92%. To put it in context, the results in Table 5 show that a 10% increase in leaver ratio reduces project productivity by 4.3% (equivalent to 50.1 fewer edits on articles), and a 10% increase in newcomer ratio (e.g., from 10% to 20%) increases project productivity by 19.2% (equivalent to 223.7 more article edits). In other words (if all other variables remain the same), when the number of leavers and newcomers are the same, project productivity will increase. Model 7 shows similar effects for project planning (leavers Coef = -0.85, p <.001; newcomers Coef = 2.22, p <.001). Therefore, *turnover on average is beneficial to WikiProjects.*

5.4 Effects of Different Types of Leavers

The models in Table 2 explore the effects of leavers on project performance in more detail, and let us test hypotheses 1a-1c. Models 3 and 8 evaluated the effect of leavers' *prior productivity* (the number of edits the user made in the project), Models 4 and 9 evaluated the effect of leavers' *tenure* (the length of the users' membership in the project), and Models 5 and 10 evaluated the effect of leavers' *additional commitments* (simultaneous participation in other projects).

Model 3 shows that *productive* a member leaving is more detrimental to the project. More specifically, while losing both productive (Coef = -0.69, p <.001) and unproductive members (Coef = -0.21, p <.05) was associated with a negative change in project productivity, the effect of losing productive members was more than twice that of losing unproductive members. Again, to put it in context, a 10% increase in the ratio of productive leavers decreased project productivity by 80.4 edits compared to the previous time period, estimated by the mean of prior project productivity (Table 5), and that number reduced to 24.5 in the case of unproductive members under the same condition. Model 8 shows similar effects on project planning; both types of leavers have negative affects on the change in project planning, with productive leavers (Coef = -0.90, p <.001) having a larger affect than unproductive leavers (Coef = -0.69, p <.001). Thus, *Hypothesis 1a is supported.*

Models 4 and 9 show that short-term leavers hurt project performance more than long-term leavers. Both affected project productivity negatively, but the effect size of short-term leavers was twice that of long-term leavers for both project productivity (Coef = -0.57, p <.001 *V.S.* Coef = -0.30, p <.01) and project planning (Coef = -1.28, p <.001 *V.S.* Coef = -0.65, p <.001). Thus, *Hypothesis 1b is not supported.* We speculate about reasons for this unexpected result in the Discussion section.

	Project Productivity Change										Project Planning Change									
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8		Model 9		Model 10	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Intercept	3.37***	.26	1.16***	.27	1.37***	.26	1.80***	.44	1.31	.30	3.12***	.25	0.00	.24	0.33	.23	0.40	.40	0.35**	.27
Wikipedia Time Index	-0.21	.06	0.20***	.06	0.16***	.06	0.21**	.10	0.25***	.06	-0.30***	.06	0.05	.06	-0.03	.05	0.08	.08	0.14*	.06
Project Age	-0.23***	.03	-1.52***	.26	-0.10***	.02	-0.08*	.04	-0.12***	.03	-0.12***	.03	-0.07**	.02	0.03*	.02	0.12**	.04	-0.01	.02
Project Size	0.04	.02	0.18***	.02	0.13***	.02	0.25***	.04	0.25***	.02	0.14***	.02	0.38***	.02	0.28***	.02	0.52***	.04	0.56***	.02
Prior Prod./Planning	-0.45***	.01	-0.48***	.01	-0.45***	.01	-0.55***	.02	-0.47***	.01	-0.49***	.01	-0.59***	.01	-0.52***	.01	-0.59***	.02	-0.61***	.11
Leavers			-0.43***	.11	-0.33***	.10	-1.00***	.23	-0.66***	.13			-0.85***	.09	-0.73***	.08	-1.30***	.21	-1.14***	.02
Newcomers			1.92***	.10									2.22***	.08						
Prod. Newcomers					0.90***	.07									1.08***	.06				
Unprod. Newcomers					1.23***	.08									1.39***	.07				
Exp. Newcomers							3.24***	.21									4.74***	.20		
Inexp. Newcomers							0.13***	.04									0.30***	.03		
Newcomers w/ Com.									2.16***	.14									3.32***	.12
Newcomers w/o Com.									2.86***	.16									3.78***	.14
Within R-Sq	0.29		0.32		0.32		0.32		0.32		0.19		0.27		0.29		0.30		0.29	
Between R-Sq	0.59		0.61		0.61		0.61		0.61		0.35		0.41		0.39		0.38		0.38	
Overall R-Sq	0.35		0.39		0.39		0.39		0.39		0.20		0.28		0.30		0.30		0.30	
Likelihood-ratio Test			633.38 ***		626.77 ***		638.59 ***		635.52 ***				1659.26 ***		1613.66 ***		1624.22 ***		1591.06 ***	

Table 3. Results of the Effects of Different Types of Newcomers on Project Productivity and Project Planning Changes. $^{*}p$ <.05, $^{**}p$ <.01, $^{***}p$ <0.001

Models 5 and 10 show that losing members without additional commitments hurts project performance more than losing members with additional commitments. While losing either type of member had significant negative effects on project productivity, the effect size of losing members without additional commitments was much larger than losing members with additional commitments for both project productivity (Coef = -0.86, p <.01 *V.S.* Coef = -0.59, p <.001) and project planning (Coef = -2.03, p <.001 *V.S.* Coef = -1.00, p <.001). Thus, *Hypothesis 1c is supported.*

5.5 Effects of Different Types of Newcomers

The models in Table 3 explore the effects of newcomers on project performance in more detail, and let us test hypotheses 2a-2c. Models 3 and 8 evaluated the effect of newcomers' prior *productivity* (the number of edits the user made in the *community*, i.e., Wikipedia), Models 4 and 9 evaluated the effect of newcomers' *tenure* (the length of the user's participation in the community), and Models 5 and 10 evaluated the effect of leavers' *additional commitments* (simultaneous participation in other projects).

Models 3 and 8 show that adding newcomers who were previously *less productive* (in the broader community, i.e., Wikipedia) is more beneficial to a project. Specifically, while both productive (Coef = 0.90, p <.001) and less productive (Coef = 1.23, p <.001) newcomers were associated with a positive change in project productivity, the effect size of less productive newcomers was larger. Similarly, less productive newcomers (Coef = 1.39, p <.001) were associated with a larger increase in project planning than more (externally) productive newcomers (Coef = 1.08, p <.001). Thus, *Hypothesis 2a is not supported.* We again will suggest reasons for this in the Discussion section.

Models 4 and 9 show that adding newcomers with previous experience in the broader community (Wikipedia) benefits a project more than adding newcomers who were also new to the community. Table 3 shows that both experienced and inexperienced newcomers were associated with positive changes in project productivity, but the effect size of experienced newcomers (Coef = 3.24, p <.001) was much larger than inexperienced newcomers (Coef = 0.13, p <.001). Similar results can be observed in the effects on project planning (experienced Coef = 4.74, p <.001 *V.S.* inexperienced Coef = 0.30, p <.001). Thus, *Hypothesis 2b is supported.*

Models 5 and 10 show that newcomers without additional commitments are more valuable for a project. While both types of newcomers were associated with a positive change in project

	Productivity Change		Planning Change	
	Coef.	S.E.	Coef.	S.E.
Intercept	-1.23***	.26	-1.100	***
Wikipeida Time Index	0.25**	.01	0.08*	.05
Project Age	-0.13***	.03	0.01	.02
Project Size	0.14***	.01	0.29***	.02
Prior Productivity/Planning	-0.46***	.01	-0.54***	.01
Leavers	-0.63***	.17	-1.49***	.10
Newcomers	2.14***	.13	2.85***	.08
Group Size X Leavers	-0.32**	.11	-0.63***	.07
Group Size X Newcomers	0.63***	.08	1.12***	.05
Within R-Sq	0.31		0.28	
Between R-Sq	0.61		0.41	
Overall R-Sq	0.38		0.28	
Likelihood-ratio Test	637.46 ***		1740.23 ***	

Table 4. Results of Moderating Effects of Group Size on Ratios of Leavers and Newcomers on Project Productivity and Project Activity Changes. $^{*}p$ <.05, $^{**}p$ <.01, $^{***}p$ <0.001

productivity, the effect of newcomers without additional commitments (Coef = 2.86, p <.001) was larger than newcomers with additional commitments (Coef = 2.16, p <.001). Further, adding newcomers without additional commitments (Coef = 3.78, p <.001) also increased project planning more than adding newcomers with additional commitments (Coef = 3.32, p <.001). Therefore, *Hypothesis 2c is supported.*

5.6 Moderating Effects of Group Size

The models in Table 4 explore the moderating effect of project size – how it affects the relationship between leavers/newcomers and project performance – and it lets us test hypotheses 3a and 3c. To compute the interaction effects, we centered independent variables by mean to reduce collinearity among independent variables and interaction terms.

Table 4 shows that larger projects suffered more from losing members and benefited more from adding newcomers. First, project productivity was negatively associated with leaver ratio (Coef = -0.63, p <.001) and with the interaction term between project size and leaver ratio (Coef = -0.32, p <.01). In other words, if leaver ratio stays constant, increasing project size by 1% increased the negative effect of the leavers on project productivity by -0.321%. Second, project productivity was positively associated with newcomer ratio (Coef = 2.14, p <.001) and with the interaction term between project size and newcomer ratio (Coef = 0.63, p <.001). Project planning was affected in a similar direction and comparable magnitude (interaction with leavers Coef = -0.63, p <.001 *V.S.* interaction with newcomers Coef = 1.12, p <.001). Therefore, *Hypothesis 3a is not supported, but Hypothesis 3b is supported.* We reflect on possible reasons for the unexpected result in the Discussion section.

6 SUMMARY AND DISCUSSION OF RESULTS

Table 5 summarizes our main results, i.e., the effects of different types of leavers and newcomers on project performance. We indicate whether there is a statistically significant difference between each pair of coefficients, and further interpret the percentage changes in project productivity and project planning with the numbers of edits increased or decreased on the project.

Hypothesis	Increasing Certain Type of Members by 10% Points[8]	Productivity Change in %	Productivity Change in #	Coef. Diff. χ^2	Planning Change in %	Planning Change in #	Coef. Diff. χ^2
	Leavers	-4.3%	-50.1		-6.9%	-5.2	
1a (Supported)	Productive Leavers	-6.9%	-80.4	4.73 ***	-9.0%	-6.8	1.52
	Unproductive Leavers	-2.1%	-24.5		-6.9%	-5.2	
1b (Not Supported)	Short-Term Leavers	-5.7%	-66.4	1.86	-12.8%	-9.6	13.27 ***
	Long-Term Leavers	-3.0%	-35.0		-6.5%	-4.9	
1c (Supported)	Leavers with Commitment	-5.9%	-68.7	0.67	-10.0%	-7.5	11.15 ***
	Leavers without Commitment	-8.6%	-100.2		-20.3%	-15.3	
	Newcomers	19.2%	223.7		22.2%	16.7	
2a (Not Supported)	Productive Newcomers	9.0%	104.9	9.82 ***	10.8%	8.1	7.14 ***
	Unproductive Newcomers	12.3%	143.3		13.9%	10.5	
2b (Supported)	Experienced Newcomers	32.4%	377.5	203.14 ***	14.7%	11.1	493.22 ***
	Inexperienced Newcomers	13.2%	153.8		15.9%	12.0	
2c (Supported)	Newcomers with Commitment	21.6%	251.7	14.11 ***	33.2%	25.0	9.74 **
	Newcomers without Commitment	28.6%	333.2		37.8%	28.5	

Table 5. Summary of the main effects of different types of leavers and newcomers on project productivity and project planning changes. We computed the statistical significance of each pair of coefficients, and further showed the impacts of changes in percentage with the numbers of edits increased or decreased on the project. $^{*}p$ <.05, $^{**}p$ <.01, $^{***}p$ <0.001

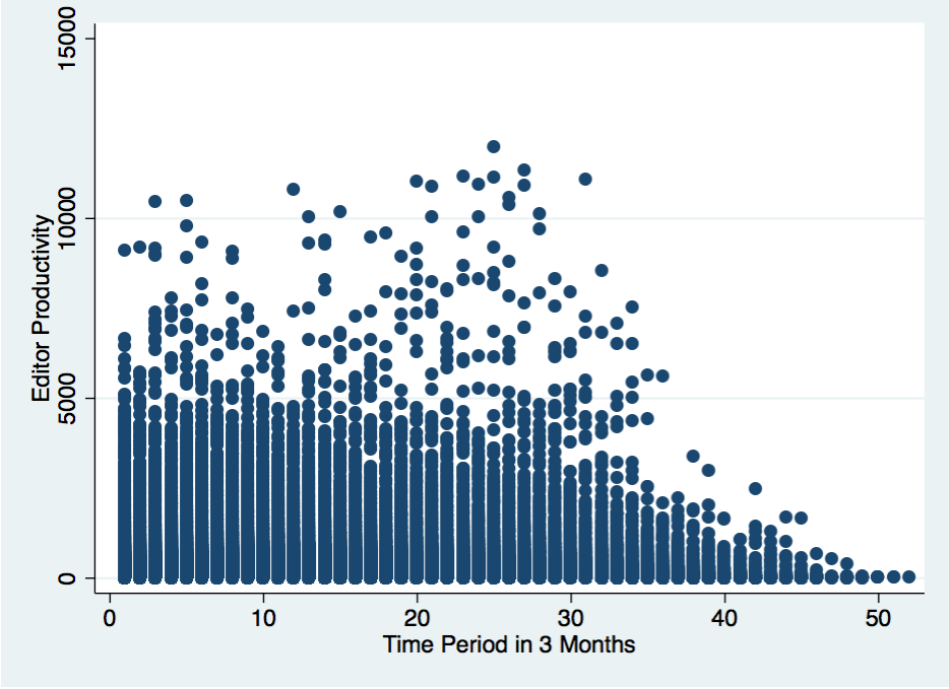

Fig. 1. Trend of editors' productivity in the project after joining for all the editors in all the projects. Time Period 0 is the time point they joined the project. Clearly, editors contributed less over time after joining the project.

Regarding the effects of different characteristics of leavers, the results show support for Hypothesis 1a and 1c, but not 1b. The departure of productive members without additional commitments

[8]Increasing 10% points means increasing by addition, e.g., increasing the newcomer ratio from 10% to 20%.

indeed is more harmful to a group than the departure of unproductive members with additional commitments (Hypothesis 1a and 1c). However, in contrast to our hypothesis 1b, the departure of members who stayed in the group for a short period of time hurts group performance more than the departure of those who had been in the group for a long time. One possible reason why short-term members are important to the group is that their lack of experience is outweighed by greater motivation and effort. Short-term members may be more energetic and willing to take on more tasks. In contrast, longtime members might gradually reduce their interest in and commitment to the group, contribute less, and slowly "retire" [49]. The trend in Figure 1 shows clearly that member productivity within a project decreases over time. Therefore, if a project loses a short-term member, that member will in general be in a more productive era of his or her life cycle than a long-term member. Observe that the results are consistent with the overall trend seen in the baseline models: the positive impact of newcomers outweighs the negative impact of leavers.

Regarding the effects of newcomers, Hypotheses 2b and 2c are supported, but Hypothesis 2a is not. Newcomers with more experience in Wikipedia and without additional commitments are more beneficial to a group than newcomers with less experience and with additional commitments (Hypotheses 2b and 2c). Interestingly, newcomers who had been less productive (in Wikipedia) increased group performance more than more productive newcomers. A possible explanation is that project newcomers who had been productive in the large community were at greater risk of "burnout" [10]. Indeed, prior research found that burnout is quite common among the most productive members in Wikipedia [49] and other voluntary activities [52]. Therefore, a takeaway is that projects may benefit most from *experienced but not exploited* newcomers.

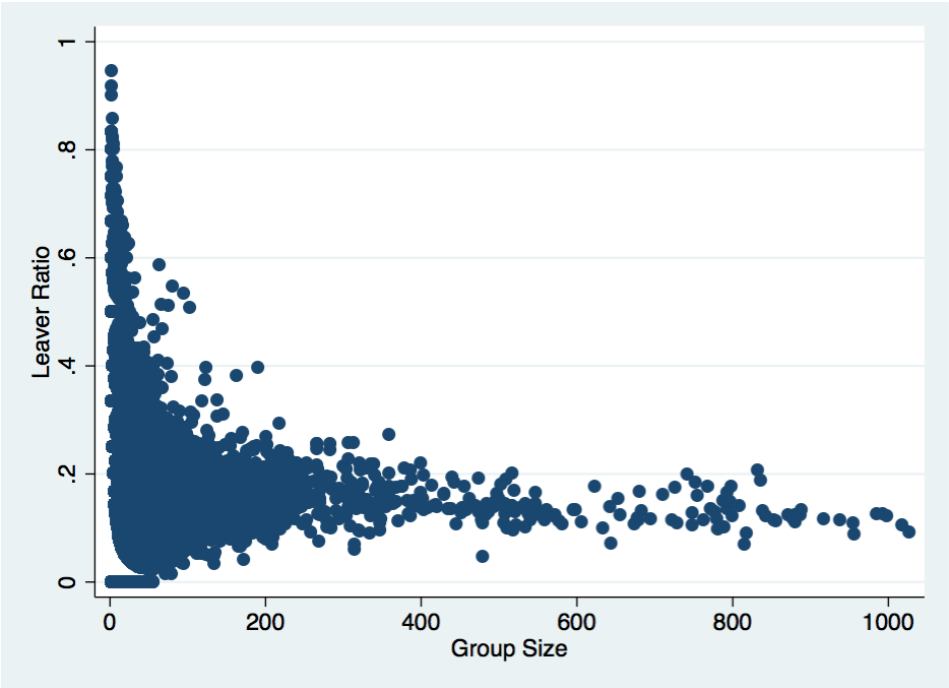

Fig. 2. Scatter plot of the relationship between group size and leaver ratio over all the time periods. Apparently, large groups have a steady rate of leavers, while small groups face a various range of leaver ratios.

Further, we found that the effects of leavers and newcomers were both amplified for larger groups, which is consistent with Hypothesis 3b but not with Hypothesis 3a. It is interesting to see that larger groups were harmed more than smaller groups by the same proportion of member leaving. A closer look at the data revealed that larger groups generally have lower leaver rates than smaller groups. Figure 2 clearly shows that leaver ratios of larger groups are scattered across a relatively low and stable range, while smaller groups are more likely to suffer higher proportions of member leaving as indicated by a higher mean and larger variance. In other words, it is more common for a project of 10 members to lose 5 than a project of 1000 members to lose 500. Because smaller groups experience large member departures more frequently, it is possible that they might develop protection mechanisms against member departures. The results suggest directions for future research to investigate why and how small online groups can operate productively in the face of frequent member departures.

In summary, our findings advance knowledge of the effects of membership turnover in online production communities by investigating exactly how the properties of turnover affect group performance. We bridge the gap between turnover theories in organizational science literature and user behavior studies in online production communities. Our results clearly demonstrate the significant effects of membership turnover in online production communities.

6.1 Design Implications

Our findings suggest practical design implications for community organizers and designers. First, our work can inform the design of visualization and analytic tools to help predict the effects of membership turnover on project performance. These tools can automatically calculate the numbers and types of leavers and newcomers for a given project and estimate their effects on project performance. Project leaders can use the data to better manage and organize project activities accordingly.

Our findings also can inform the design of more effective and targeted retention and recruitment strategies in online production groups. For example, to minimize the costs of member leaving and maximize the benefits of member joining, online groups should focus on retaining relatively new and productive existing members who do not have additional commitments, and recruiting experienced but not exploited new members who do not have additional commitments. Project leaders also could "look ahead" to mitigate harmful effects of member departures: if some current members appear to be reaching the end of their participation, the project could begin recruiting an appropriate number of new members to make up for their departures.

6.2 Limitations and Future Work

We also note several limitations of our study. First, we used the quantity of contributions not their quality to measure the performance of WikiProjects. However, contribution quality is also a key metric for online production groups, and previous research on Wikipedia has offered metrics for computing contribution quality [51]. Future research could extend our study by using these metrics to compute changes in project quality outcomes in relation to the turnover factors we defined.

Second, our study is correlational, and thus does not support causal inferences. Further, as a quantitative study of behavioral data, it does not explain *why* certain types of leavers and newcomers have (unpredicted) influence on group performance. In future research, we plan to conduct natural experiments (i.e., examining the effects of external events that cause members to leave and join projects) and qualitative analysis such as interviews to complement the current findings and investigate causality. Social computing research topics are best studied through multiple studies that employ different methods. A good example is the line of work on the lifecycles of participants in peer production communities that began with an interview study of Wikipedia editors [5],

continued with a quantitative data analysis of Wikipedia editors [40], and eventually led to multiple studies in additional peer production systems using a variety of methods [2, 24].

Third, we followed prior work in defining an editor *joining* a project as the editor making their first edit to the project page or project talk page (i.e., participatory membership) [47, 54]. However, other operationalizations of project membership are reasonable, e.g., only counting as project "members" those editors whose names appear on explicit lists of project members [49]. We do note that Morgan and colleague compared these two membership identification approaches and found no significant differences in members' content production and planning activities [35]; despite this, we think it reasonable future work to experiment with different operationalizations of membership and see whether they affect our results.

7 CONCLUSION

To summarize, we investigated the effects of turnover on subgroups in peer production communities. In contrast to prior research that focused on the *amount* of turnover, we studied how the *characteristics of leavers and newcomers* (their productivity, tenure, and additional commitments) impacted group performance – both actual work and work planning – and the moderating effect of group size. We found that these characteristics had significant predictive powers. In general, the positive effects of newcomers outweighed the negative effects of leavers. Specifically, the departure of members who were productive, who were in the group for a short time, and who did not have additional commitments hurt a project more, and the joining of newcomers who previously had not been productive in the broader community, but who had some prior experience in the community and did not have other specific additional commitments helped a project more. Further, group size amplified the effects: larger groups were hurt more by departures and benefited more from newcomers. Our work extends the organizational science literature by carefully exploring the effects of membership turnover in an online community. Our findings suggest design implications for online community analytics and recruitment strategies which can contribute to building effective online production communities.

8 ACKNOWLEDGMENT

We thank the members of the GroupLens Research at the University of Minnesota for the helpful feedback, with special thanks to Aaron Halfaker. We acknowledge the College of Science and Engineering, University of Minnesota for funding this project.

REFERENCES

[1] Mara B Adelman and Lawrence R Frey. 1997. *The fragile community: Living together with AIDS.* Lawrence Erlbaum Associates.

[2] Ofer Arazy, Hila Lifshitz-Assaf, Oded Nov, Johannes Daxenberger, Martina Balestra, and Coye Cheshire. 2017. On the" How" and" Why" of Emergent Role Behaviors in Wikipedia.. In *CSCW*. 2039–2051.

[3] Anita L Blanchard and M Lynne Markus. 2004. The experienced sense of a virtual community: Characteristics and processes. *ACM Sigmis Database* 35, 1 (2004), 64–79.

[4] Allen C Bludedorn. 1978. A taxonomy of turnover. *Academy of Management Review* 3, 3 (1978), 647–651.

[5] Susan L Bryant, Andrea Forte, and Amy Bruckman. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. ACM, 1–10.

[6] Kathleen Carley. 1992. Organizational learning and personnel turnover. *Organization Science* 3, 1 (1992), 20–46.

[7] Gaylen N Chandler, Benson Honig, and Johan Wiklund. 2005. Antecedents, moderators, and performance consequences of membership change in new venture teams. *Journal of Business Venturing* 20, 5 (2005), 705–725.

[8] Jilin Chen, Yuqing Ren, and John Riedl. 2010. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 821–830.

[9] Boreum Choi, Kira Alexander, Robert E Kraut, and John M Levine. 2010. Socialization tactics in wikipedia and their effects. In *Proceedings of the ACM conference on Computer supported cooperative work*. ACM, 107–116.

[10] Cynthia L Cordes and Thomas W Dougherty. 1993. A review and an integration of research on job burnout. *Academy of management review* 18, 4 (1993), 621–656.

[11] John L Cotton and Jeffrey M Tuttle. 1986. Employee turnover: A meta-analysis and review with implications for research. *Academy of management Review* 11, 1 (1986), 55–70.

[12] Jonathon N Cummings, Sara Kiesler, Reza Bosagh Zadeh, and Aruna D Balakrishnan. 2013. Group heterogeneity increases the risks of large group size: a longitudinal study of productivity in research groups. *Psychological science* 24, 6 (2013), 880–890.

[13] Rene Y Darmon. 1990. Identifying sources of turnover costs: A segmental approach. *The Journal of Marketing* (1990), 46–56.

[14] Gregory G Dess and Jason D Shaw. 2001. Voluntary turnover, social capital, and organizational performance. *Academy of management review* 26, 3 (2001), 446–456.

[15] Peng Ding, Xinran Li, and Luke W Miratrix. 2017. Bridging Finite and Super Population Causal Inference. *Journal of Causal Inference* (2017).

[16] Alan H Dorfman and Richard Valliant. 1998. Superpopulation models in survey sampling. *Encyclopedia of Biostatistics* (1998).

[17] Joan Esteban. 2001. Collective action and the group size paradox. In *American Political Science Association*, Vol. 95. Cambridge Univ Press, 663–672.

[18] Terri L Griffith, John E Sawyer, and Margaret A Neale. 2003. Virtualness and knowledge in teams: Managing the love triangle of organizations, individuals, and information technology. *MIS quarterly* (2003), 265–287.

[19] Oscar Grusky. 1960. Administrative succession in formal organizations. *Social Forces* 39, 2 (1960), 105–115.

[20] HO Hartley and Robert L Sielken Jr. 1975. A" super-population viewpoint" for finite population sampling. *Biometrics* (1975), 411–422.

[21] John P Hausknecht and Jacob A Holwerda. 2013. When does employee turnover matter? Dynamic member configurations, productive capacity, and collective performance. *Organization Science* 24, 1 (2013), 210–225.

[22] John P Hausknecht and Charlie O Trevor. 2011. Collective turnover at the group, unit, and organizational levels: Evidence, issues, and implications. *Journal of Management* 37, 1 (2011), 352–388.

[23] Benjamin Mako Hill and A Shaw. 2012. *Is volunteer labor a" fixed and finite" resource? Evidence from peer production.".* Technical Report. Working Paper.

[24] Shih-Wen Huang, Minhyang Mia Suh, Benjamin Mako Hill, and Gary Hsieh. 2015. How activists are both born and made: An analysis of users on change. org. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 211–220.

[25] Mark A Huselid. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of management journal* 38, 3 (1995), 635–672.

[26] Gerald C Kane and Maryam Alavi. 2007. Information technology and organizational learning: An investigation of exploration and exploitation processes. *Organization Science* 18, 5, 796–812.

[27] Raghav Pavan Karumur, Tien T Nguyen, and Joseph A Konstan. 2016. Early Activity Diversity: Assessing Newcomer Retention from First-Session Activity. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 595–608.

[28] Aniket Kittur, Bongwon Suh, Bryan A Pendleton, and Ed H Chi. 2007. He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 453–462.

[29] David Krackhardt and Lyman W Porter. 1985. When friends leave: A structural analysis of the relationship between turnover and stayers' attitudes. *Administrative science quarterly* (1985), 242–261.

[30] Robert E Kraut and Andrew T Fiore. 2014. The role of founders in building online groups. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 722–732.

[31] Robert E Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. 2012. *Building successful online communities: Evidence-based social design.* Mit Press.

[32] Shyong K Lam, Jawed Karim, and John Riedl. 2010. The effects of group composition on decision quality in a social production community. In *Proceedings of the 16th ACM international conference on Supporting group work*. ACM, 55–64.

[33] Meryl Reis Louis. 1980. Surprise and sense making: What newcomers experience in entering unfamiliar organizational settings. *Administrative science quarterly* (1980), 226–251.

[34] Audris Mockus, Roy T Fielding, and James D Herbsleb. 2002. Two case studies of open source software development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 11, 3 (2002), 309–346.

[35] Jonathan T Morgan, Michael Gilbert, David W McDonald, and Mark Zachry. 2013. Project talk: Coordination work and group membership in WikiProjects. In *Proceedings of the 9th International Symposium on Open Collaboration*. ACM, 3.

[36] Gad Nathan. 2011. Superpopulation Models in Survey Sampling. In *International Encyclopedia of Statistical Science*. Springer, 1575–1577.
[37] Pamela E Oliver and Gerald Marwell. 1988. The Paradox of Group Size in Collective Action: A Theory of the Critical Mass. II. *American Sociological Review* (1988), 1–8.
[38] Charles A O'Reilly III, David F Caldwell, and William P Barnett. 1989. Work group demography, social integration, and turnover. *Administrative science quarterly* (1989), 21–37.
[39] Felipe Ortega and Jesus M Gonzalez Barahona. 2007. Quantitative analysis of thewikipedia community of users. In *Proceedings of the international symposium on Wikis*. ACM, 75–86.
[40] Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM international conference on Supporting group work*. ACM, 51–60.
[41] Xiangju Qin, Michael Salter-Townshend, and Pádraig Cunningham. 2014. Exploring the relationship between membership turnover and productivity in online communities. *arXiv preprint arXiv:1401.7890* (2014).
[42] Sam Ransbotham and Gerald C Kane. 2011. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *Mis Quarterly* (2011), 613–627.
[43] Reinhard Schunck et al. 2013. Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *Stata Journal* 13, 1 (2013), 65–76.
[44] Jason D Shaw, Michelle K Duffy, Jonathan L Johnson, and Daniel E Lockhart. 2005. Turnover, social capital losses, and performance. *Academy of Management Journal* 48, 4 (2005), 594–606.
[45] Myra H Strober. 1990. Human capital theory: Implications for HR managers. *Industrial Relations* 29, 2 (1990), 214–239.
[46] Zeynep Ton and Robert S Huckman. 2008. Managing the impact of employee turnover on performance: The role of process conformance. *Organization Science* 19, 1 (2008), 56–68.
[47] Hang Ung and Jean-Michel Dalle. 2010. Project management in the Wikipedia community. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*. ACM, 13.
[48] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark GJ van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. 2015. Gender and tenure diversity in GitHub teams. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3789–3798.
[49] Loxley Sijia Wang, Jilin Chen, Yuqing Ren, and John Riedl. 2012. Searching for the goldilocks zone: trade-offs in managing online volunteer groups. In *Proceedings of the ACM conference on Computer Supported Cooperative Work*. ACM, 989–998.
[50] Yi-Chia Wang, Robert Kraut, and John M Levine. 2012. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM conference on Computer Supported Cooperative Work*. ACM, 833–842.
[51] Morten Warncke-Wang, Dan Cosley, and John Riedl. 2013. Tell me more: an actionable quality model for Wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration*. ACM, 8.
[52] John Wilson. 2000. Volunteering. *Annual review of sociology* 26, 1 (2000), 215–240.
[53] Bowen Yu, Yuqing Ren, Loren Terveen, and Haiyi Zhu. 2017. Predicting Member Productivity and Withdrawal from Pre-Joining Attachments in Online Production Groups. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1775–1784.
[54] Haiyi Zhu, Robert Kraut, and Aniket Kittur. 2012. Organizing without formal organization: group identification, goal setting and social modeling in directing online production. In *Proceedings of the ACM conference on Computer Supported Cooperative Work*. ACM, 935–944.
[55] Haiyi Zhu, Amy Zhang, Jiping He, Robert E Kraut, and Aniket Kittur. 2013. Effects of peer feedback on contribution: a field experiment in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2253–2262.

Received March 2017; revised June 2017; accepted August 2017