Predicting Member Productivity and Withdrawal from Pre-Joining Attachments in Online Production Groups

Bowen Yu*, Yuqing Ren**, Loren Terveen*, and Haiyi Zhu*

*GroupLens Research, University of Minnesota
**Carlson School of Management, University of Minnesota
bowen@cs.umn.edu, chingren@umn.edu, {terveen, haiyi}@cs.umn.edu

ABSTRACT

Productive and dedicated members are critical to the success of online production communities like Wikipedia. Many communities organize in subgroups where members voluntarily work on projects of shared interest. In this paper, we investigate how members' pre-joining connections with the subgroup predict their productivity and withdrawal after joining. Drawing insights from attachment theories in social psychology, we examine two types of pre-joining connections: *identity-based* attachment (how much members' interests were aligned with the subgroup's topics) and *bonds-based* attachment (how much members had interacted with other members of the subgroup). Analyses of 79,704 editors in 1,341 WikiProjects show that 1) both identity-based and bonds-based attachment increased editors' post-joining productivity and reduced their likelihood of withdrawal; 2) identity-based attachment had a stronger effect on boosting direct contributions to articles while bonds-based attachment had a stronger effect on increasing article and project coordination, and reducing member withdrawal.

CCS Concepts

•Human-centered computing → Computer supported cooperative work;

Keywords

Online Communities; Peer Production; Online Groups; Wikipedia; Wikiproject; Bonds-based Attachment; Identity-based Attachment; Pre-joining Attachment; Productivity; Withdrawal

Categories and Subject Descriptors

H.5.3: [**Group and Organization Interfaces**]
; K.4.3: [**Organizational Impacts**]

1. INTRODUCTION

Peer production has become a popular and influential means of knowledge production, with prominent examples including Wikipedia and open source software projects such as Linux and Apache [3, 29]. These communities organize and accomplish a remarkable amount of work by recruiting volunteer participants and supporting their efforts to produce complex and valuable artifacts.

When an online production community grows large, a common way to manage effort is to allow subgroups dedicated to specific topics or interests to self organize. Examples include: (1) Wikiprojects in Wikipedia, where groups of editors create encyclopedia articles about specific topics; (2) projects in SourceForge and Github, where groups of programmers come together to write software; and 3) projects in OpenStreetMap that seek to map geographic features in particular regions. Prior research has shown that subgroups play a major role in directing contributors towards specific group goals [15], encouraging citizenship behaviors [15], providing opportunities to find collaborators, socialize and network [9], and protecting the joint work [9].

Despite the critical role they play in maintaining the health of the online communities, many subgroups struggle or fail because they lack a steady supply of productive and dedicated participants [4]. For example, on Wikia, a platform hosting collaborative editing groups on topics of interest such as food, games, and travel, 22% of the groups were abandoned within one month of being created due to lack of contributions [35].

Much existing research has focused on understanding various tactics that online production groups can deploy to promote member productivity and reduce withdrawal. Some of the commonly studied tactics are socialization [6], group identificiation [15, 37], and goal setting [37]. Most existing research on subgroups has focused on what happens after a member joins a group. However, Panciera et al pointed out in their "Wikipedian are born, not made" paper [24] that members' long-term productivity can be largely predicted by their activities on the first day. In other words, an editor's likelihood of becoming an active Wikipedian was largely determined before his or her joining. In this paper, we build on this result to study member behaviors in subgroups within the large community of Wikipedia. We speculate that the connections that members have forged before joining a subgroup can predict their subsequent contribution to the subgroup and their likelihood of staying with or leaving the subgroup.

We drew insights from social psychology and the online community literature to conceptualize pre-joining connections with subgroups. According to the literature, there are two primary routes through which members can connect to a group: identity-based and bonds-based. With identity-based attachment, members connect with the group as a whole and its shared purpose and interest (e.g., *"I feel connected to the group because I'm interested in the topic of the group and agree with its goal and mission"*). With

CSCW '17, February 25-March 01, 2017, Portland, OR, USA

DOI: http://dx.doi.org/10.1145/2998181.2998227

bonds-based attachment, members connect with individual members of the group (e.g., *"I feel connected to the group because I am friends with its members"*). In the context of online communities, identity-based attachment is likely to develop when members' interests match the goals of the group, e.g., to learn and share information about programming languages or chronic health conditions; bonds-based attachment is likely to develop when members have preexisting friendships or build close relationships through personal disclosure and repeated interaction.

In this paper, we aim to address two research questions. *(1) How do members' pre-joining connections to subgroups predict their productivity and withdrawal after joining the subgroup? (2) Do identity-based and bonds-based attachments have similar or different effects on member productivity and withdrawal in the subgroups?*

We studied these questions in the context of Wikiprojects. A Wikiproject is a subgroup within Wikipedia intended to help organize volunteer effort to improve the quality of articles on a specific topic. We collected our data from the June 2015 full dump of English Wikipedia. We measured pre-joining identity-based attachment by the degree to which members' interests match the topics of the project, and measured pre-joining bonds-based attachment as the amount of direct interactions that members have had with other project members. We measured two outcomes: *productivity*, measured as the number of edits a member contributed after joining a project, and *withdrawal*, measured as the act of leaving or stopping contributing to the project. In addition to direct contributions to a WikiProject, we also studied some types of indirect contributions such as article and project coordination. Our findings advance knowledge about subgroup dynamics and member behaviors in online production communities and inform the designs of cost-effective recruitment and retention strategies.

2. THEORIES AND HYPOTHESES

2.1 Productivity and Withdrawal

Productivity and withdrawal are two primary outcomes in online production communities because they jointly determine how many volunteers are available to contribute and how much work they collectively can accomplish. Productivity measures the quantity of work output. For instance, the total number of edits is a shared measure of productivity in Wikipedia (e.g., [5, 33]). Withdrawal indicates the end of a member's voluntary participation or his or her likelihood of leaving or stopping contributing to the community (e.g., [33]).

Researchers have examined a wide variety of factors that influence member productivity and withdrawal in online production communities, including but not limited to different types of leadership behaviors [36], newcomer socialization tactics [6], feedback members receive [38], and diversity of subgroups that members belong to [5]. In online health support communities, it has been shown that participants' withdrawal patterns are associated with different types and amounts of social support they were exposed to [34].

Some studies have begun to reveal critical trade-offs between productivity and withdrawal. While online communities strive to encourage productivity and reduce withdrawal, in reality, the two goals may not be accomplished equally well [33]. Lack of member contribution and poor member retention are difficult challenges that many community administrators face [25]. Sometimes one goal may be promoted at the cost of the other. For example, in Wang et al.'s work [33], members who had contributed more edits in the past were more productive moving forward yet were also more likely to stop participating, both in their subgroups and the community as a whole. A possible explanation is a feeling of "mission accomplished" or "burnout".

In this work, we further explore the tension between productivity and withdrawal by investigating how pre-joining connections to subgroups within a large community affect member productivity and withdrawal. Most prior work has focused on environmental factors and their impacts on member productivity and withdrawal. In contrast, we focus on connections that exist before a member joins a subgroup. As Panciera and her colleagues [24] suggested, the amount of work done by Wikipedians (i.e., those made 250 edits over their lifetime) and non-Wikipedians (i.e., those made fewer than 250 edits over lifetime) differed from their very first day. In other words, one's likelihood of becoming an active Wikipedian is at least partially determined by his or her natural tendency prior to joining Wikipedia. We speculate that members' contribution and affiliation with a WikiProject follow a similar pattern. The degree to which a member contributes to and stays with a WikiProject is at least partially determined by his or her pre-joining connections with the project.

2.2 Identity-Based and Bonds-Based Attachment Theories

Social psychology has a long tradition of studying and understanding accounts of group formation and cohesion [21, 31]. According to these theories, there exist two conceptually distinctive ways in which a member can connect with a group: identity-based attachment and bonds-based attachment [26]. The former works through group identity, whereby people feel connected to a group's character or shared purpose [12, 32]. The latter works through interpersonal attraction, whereby people develop relationships with other individual members [8]. The distinction was first made in studying student groups in colleges [26], with identity-based groups like volleyball teams and eating clubs and bonds-based groups like friendship cliques.

In recent years, researchers have used identity-based and bonds-based attachment theories as a lens to study online communities [27]. The two types of attachment apply well to online communities where members join either because they like the community as a whole or they like the people in the community [28]. A study of 75,311 Facebook groups in 2014 [17] showed that the two most common themes were groups associated with an organization, team or club and groups for keeping in touch with family or friends.

In a comprehensive review of the literature, Ren and her colleagues summarized the different antecedents and consequences of the two types of attachment, and discussed how theoretical insights can be applied to inform design decisions to promote either common identity or interpersonal bonds among community members [28]. In laboratory and field experiments, researchers have begun exploring these evidence-based designs to facilitate different types of attachment to online communities. For example, in [27, 7], two sets of features were implemented to increase either identity-based or bonds-based attachment. The identity-based features included: group profiles with information about groups, repeated exposure to the group, and intergroup competition. The bonds-based features included: individual profiles with information about individual users, repeated exposure to interact with individual users, and interpersonal comparison. Both sets of features increased attachment, and identity-based features had stronger effects than bonds-based features.

In this paper, we study identity-based and bonds-based connections with WikiProjects in their naturally-occurring setting. Our

goal is to examine their overall as well as differential effects on member productivity and withdrawal within the subgroup.

2.3 Research Hypotheses

2.3.1 Effects of Identity-Based Attachment

The base for identity-based attachment is group identity, which is defined as "the individual's knowledge that he belongs to certain social groups together with some emotional and value significance to him of the group membership" [30]. Even before a member joins a group, he or she may form some preexisting connection to the group if the person "identifies" with the goal of the group or sees an alignment between the group goal and their personal interests. Group identification has been linked to in-group favoritism and behaviors that are beneficial to the group [2, 12, 20].

In offline groups, identity-based attachment has been associated with outcomes such as increased effort and participation, improved task performance, and ingroup favoritism decisions (see [1] for a review). In online groups, increased identity-based attachment implies that members resonate with the group's mission and share the common interests. As a result, they are more likely to be intrinsically motivated to contribute. Hence, we expect that members with a higher level of pre-joining identity-based attachment will have a higher-level of productivity after joining the group. Furthermore, group identity can trigger positive evaluation towards the group, and lead to continuous participation and increased likelihood of remaining in the group [2, 19]. Hence, we expect that members with a higher-level of pre-joining identity-based attachment will be less likely to leave the group.

Hypothesis 1a: *Members with a higher-level of pre-joining identity-based attachment to a subgroup will be more productive in the subgroup than members with a lower-level of pre-joining identity-based attachment.*

Hypothesis 1b: *Members with a higher-level of pre-joining identity-based attachment to a subgroup will be less likely to withdraw from the subgroup than members with a lower-level of pre-joining identity-based attachment.*

2.3.2 Effects of Bonds-Based Attachment

The base of bonds-based attachment is interpersonal attraction and group cohesiveness. According to [11], group cohesiveness is "the resistance of a group to disruptive forces" (pp. 553-554) and such cohesiveness is associated with the strength of the relational bonds among group members. Bonds-based attachment is related to perceived group cohesiveness and can lead to positive feelings towards the group. Members with strong bonds-based attachment are more motivated to help fellow members of the group [2, 19].

Similar to identity-based attachment, we expect that pre-joining bonds-based attachment will enhance members' evaluation of the group and increase their contribution to the group and likelihood of remaining in the group [2].

Hypothesis 2a: *Members with a higher-level of pre-joining bonds-based attachment to a subgroup will be more productive in the subgroup than members with a lower-level of pre-joining bond-based attachment.*

Hypothesis 2b: *Members with a higher-level of pre-joining bonds-based attachment to a subgroup will be will be less likely to withdraw from the subgroup than members with a lower-level of pre-joining bond-based attachment.*

2.3.3 Effects of Identity-Based vs. Bonds-Based Attachment

Although both types of attachments can increase productivity and reduce withdrawal, their relative effects on the two outcomes may differ for several reasons. First, different types of attachments might change the kinds of activities that members engage in after joining a group. Back [2] created identity-based groups (where members had mostly identity-based attachment to the group) and bonds-based groups (where members had mostly bonds-based attachment to to each other) and assigned the same tasks to both groups. He found that members of identity-based groups focused on the task itself, discussed only the matters that were relevant to the tasks, and completed their tasks efficiently; while members of bonds-based groups engaged in longer and often task-irrelevant conversations and completed the tasks less efficiently. In other words, members with strong identity-based attachment tend to focus on achieving the goal of the groups while people with strong bonds-based attachment tend to focus on strengthening their relationships with one another. We expect that identity-based attachment to have a more substantial effect on productivity than bonds-based attachment.

Similarly, we speculate that the two types of attachments affect withdrawal to different degrees and members with the two types of attachments will be sensitive to different changes. Identity-based attachment is more vulnerable to changes in the group's goals and purposes whereas bonds-based attachment is more susceptible to shifts in the membership of the group [28]. In online production communities, the purpose and shared interests are often socially negotiated and constructed by group members [16]. Conflict may arise in the process of negotiation and coordination which threatens members' perception of the alignment of group goals and their interests. We, therefore, speculate that bonds-based attachment has a greater effect in buffering members from withdrawal than identity-based attachment.

Hypothesis 3a: *Pre-joining identity-based attachment has a stronger effect on increasing member productivity than bonds-based attachment.*

Hypothesis 3b: *Pre-joining bonds-based attachment has a stronger effect on reducing member withdrawal than identity-based attachment.*

3. METHODS

3.1 Study Platform

3.1.1 Wikipedia

Wikipedia has been widely studied due to its rich content and complete historical data records. Wikipedia is organized by pages, including pages of the encyclopedic entries which are referred to as article pages, and pages dedicated to distinct purposes such as user pages, project pages, help pages, etc. Any registered user can edit almost any page on Wikipedia by clicking an edit button. Once the edit is done, it is immediately reflected in the article, and saved as a revision which gets recorded into the history data record. Every page also has a talk page, which is used by editors to discuss and coordinate writing the article pages. Article pages are tagged by categories, which are labels editors can use to describe the topics of a page. There are subcategories within high-level categories, and all the categories can be linked together into a large graph structure.

3.1.2 WikiProjects

We specifically study WikiProjects. According to Wikipedia[1], a WikiProject is "the organization of a group of participants in a wiki established in order to achieve specific editing goals, or to

[1] https://en.wikipedia.org/wiki/WikiProject

achieve goals relating to a specific field of knowledge". WikiProjects provide a place for Wikipedia editors with similar topic interests to collaborate and coordinate their work. Each WikiProject has a dedicated page known as the project page. Usually maintained and updated by the creators or senior editors in the project, project pages provide the overall status of the project including the content coverage of the articles associated with the project, membership lists, task lists, and member guidelines. Each project also has a corresponding project talk page where project members can communicate about the project.

WikiProjects provide an appropriate setting for our study for several reasons. First, WikiProjects are well-established subgroups within Wikipedia. Second, WikiProjects have clear goals and a relatively large and complicated structure. Third, the availability of high-quality data archives enables us to measure member behaviors both before and after joining a subgroup.

3.2 Data Collection

The data set used in our study is from the 06/02/2015 version of the English Wikipedia dump provided by the Wikimedia foundation[2]. It includes the complete editing history of all the articles in the English Wikipedia.

To construct our sample of WikiProjects, we used the project templates on articles' talk pages, which provided information about which Wikiprojects articles belong to. Some articles belong to multiple WikiProjects; we included all WikiProjects in these cases. A thorough search of all articles surfaced a total of 1,949 WikiProjects. We then excluded projects that never grew to more than three members, which is the minimal size of a group. We also excluded projects that are not topical but rather serve a collaboration purpose, e.g., *WikiProject: Manual of Style*. Our final sample included 1,341 WikiProjects with an average of 5,036 articles per project.

There are two common ways of determining project membership in the literature: declared membership versus participatory membership. Morgan and his colleague compared the two approaches and found no significant difference [22]. We chose to determine membership and "group joining" using the participatory approach. More specifically, we considered an editor as joining a project when he or she made the first edit to the project page or the project talk page. This yielded a total of 79,704 members of the projects in our sample (excluding the bots).

3.3 Operationalizations

Because we are interested in how editors' pre-joining attachment affects their behaviors after they join the project, we measured the independent and control variables before an editor joined the focal project and the dependent variables after the editor joined the project.

3.3.1 Dependent Variables

Article Productivity: We examined three alternative ways of measuring productivity. Following Kittur et al. [14] and Choi et al. [6], we calculated the number of edits an editor made on the main pages of all articles within the scope of the project after he or she joined the focal project. This variable indicates direct contribution to articles within the project.

According to Kriplean and his colleagues, direct contribution to articles is only one kind of valued work in Wikipedia [18]. Other types of contributions also are important and valuable, such as communication and coordination around articles and projects. Therefore, we created two additional measures of productivity: *article-related communication* and *project coordination*.

Article-related Communication: Measured as the number of edits an editor made on the talk pages of all articles within the scope of the project after he or she joined the focal project. This variable indicates the communication the editor had with other members regarding the articles within the project scope.

Project Coordination: Measured as the number of edits an editor made on the project page or project talk page of the focal project. This variable indicates editors' effort to organize or coordinate activities around the project.

Withdrawal: We regarded an editor as having withdrawn from a WikiProject if he or she had not made any edits for six months to any of the following three pages: the article pages within the scope of the project, the project pages, or the project talk pages.

3.3.2 Independent Variables:

Identity-based Attachment: We operationalized identity-based attachment as the degree to which an editor's editing interests were aligned with the topics of the WikiProject before the editor joined the project. We computed editors' interests and Wikiproject topics using Wikipedia categories.

The categorical index portal of Wikipedia [3] shows 12 top categories or primary topical areas: *Arts, Geography, Health, History, Mathematics, People, Philosophy, Religion, Science, Society, and Technology*, and *Not Found* (not under any of the other categories). Following a procedure similar to [14] and [5], we assigned each Wikipedia article to a topic area by first checking the subcategories to which the article belongs and then finding the closest top level categories in the category graph structure. A subcategory can be assigned to multiple top categories if they have the same path length in the category graph. Articles can also have multiple subcategories. As a result, an article can belong to multiple topic categories.

Let's consider the *Computer Science* article as an example. The article is tagged by four subcategories: Electrical engineering, Electronic engineering, Computer engineering, and Computer science. Following the category graph in Wikipedia for a subcategory tag, *Computer Science*, we obtained the shortest path from *Computer Science* to one of the top categories, *Science*, through two subcategories, *Applied sciences* and *Scientific disciplines*. By searching the top categories for all the four subcategory tags, we identified the top categories of the article as *Science* and *Technology*. We constructed topic coverage vector for each Wikiproject by counting the number of articles that are within the project's scope and fall into each topic category.

To measure an editor's interests, we calculated the number of articles the editor had edited in each category or topic area prior to joining the project. The overall interests of an editor can be represented by a vector of twelve elements, with each element representing the number of articles the editor had edited in that topic area. We then normalized the vector so that the sum of the twelve added up to 1.

Finally, we calculated identity-based attachment as the cosine similarity between an editor's editing interests vector and a project's topical coverage vector. The variable indicates the degree to which the editor's interests were aligned with the Wikiproject's topics (i.e., how much the editor might be interested in the topics of the Wikiproject). Because our independent variable was *pre-joining* identity-based attachment, both vectors and the variable were based on data before the joining of the Wikiproject.

Table 1 shows an example between the *WikiProject: Human Computer Interaction* and an editor named *Diego Moya*. The top row shows *Diego Moya*'s interest vector and the bottom row shows

[2] https://dumps.wikimedia.org/

[3] https://en.wikipedia.org/wiki/Category:Portals

Category	Arts	Geography	Health	Mathematics	History	Science	People	Philosophy	Religion	Society	Technology	NF
Editor:*Diego Moya*	0.130	0.034	0.006	0.048	0.038	0.052	0.106	0.094	0.028	0.242	0.200	0.024
Wikiproject: *HCI*	0.034	0.023	0.008	0.008	0.013	0.086	0.039	0.077	0.017	0.302	0.375	0.018
											Cosine Similarity:	0.180

Table 1: An Example of Identity-based Attachment Calculation Between An Editor and A WikiProject

Variable	1	2	3	4	5	6	7	8	VIF
1. Project Age	1.00								2.48
2. Project Member	0.74	1.00							2.96
3. Project Article	0.27	0.43	1.00						1.25
4. Editor Tenure	0.17	0.02	-0.04	1.00					2.33
5. Editor Prior Edits	0.06	-0.09	-0.07	0.74	1.00				4.56
6. No. of Projects Involved	0.10	-0.11	-0.12	0.53	0.76	1.00			2.68
7. Identity-based	-0.05	-0.08	0.03	0.14	0.16	-0.03	1.00		1.10
8. Bonds-based	0.27	0.23	0.07	0.40	0.57	0.48	0.04	1.00	1.89
Mean VIF									2.41

Table 2: Correlations and Collinearity Diagnostics on Independent and Control Variables

the project's topic coverage vector. Overall the editor's interests and the project's topics were pretty well aligned, especially in the *Society* and *Technology* categories. The cosine similarity of their vectors is 0.180[4]. The match between *Diego Moya* and *WikiProject: Human Computer Interaction* was among the top 25% of all editor-Wikiproject pairs.

Bonds-based Attachment: Prior research has shown that the frequency and amount of personal communications are strong predictors of relationship strength between people [10]. Based on this rationale, we operationalized pre-joining bonds-based attachment as the amount of communication an editor had with project members of a Wikiproject on their user talk pages before he or she joined the project.

We explored three measures of pre-joining bonds-based attachment: 1) the total number of edits the editor made to other project members's user talk pages before joining the project; this represents the total volume of personal communication with other members; 2) the total number of project members whose user talk pages the editor posted to before joining the project; this represents the number of ties the editor had developed; and 3) the maximum number of edits the editor made to a single project member's user talk page before joining the project; this represents the closest bond the editor had developed. Moreover, as communication is a bi-directional process, we also measured how other project members made edits on the user page of the focal editor. We therefore had six possible measures. We found that all six were highly correlated and led to similar results, so we selected the total number of user talk page edits the focal editor made to other project members to use in our main analysis.

3.3.3 Control Variables

Project Age: The number of months from the creation of the project until the focal editor joined. We used this variable to control for the project maturity, which may affect the ease with which new members could connect and contribute to the project.

Project Members: The total number of members who had participated in the project before the focal editor joined.

Project Articles: The total number of articles in the project scope.

[4] The 25% percentile, 50% percentile, and 75% percentile of the cosine similarity in the data set are respectively 0.111, 0.141, and 0.167.

Editor Tenure: The number of months the focal editor had been a Wikipedia editor before joining the project.

Editor Prior Edits: The total number of edits the focal editor had made on all Wikipedia pages before joining the project.

Number of Projects Involved: The number of *other* projects the editor had joined before the focal project.

3.3.4 Data transformation

Many of the independent and control variables were skewed so we performed base-2 logarithmic transformations of all the independent and control variables except for identity-based attachment. We then standardized all variables for ease of interpretation. We did not apply any transformation on the dependent variables because the negative binomial models we used could account for skewness of the outcome variables.

4. ANALYSIS AND RESULTS

Table 2 presents correlations and collinearity diagnostics of the independent and control variables after the logarithmic transformation and grand mean centering. The two types of attachments were not correlated. *Variance Inflation Factor* or the VIF values are frequently used to check collinearity issues [23]. As a rule of thumb, a variable whose VIF value is greater than 10 may merit further investigation. All of our VIF values were well below 10 indicating that multicollinearity was not a concern.

4.1 Article Productivity

4.1.1 Statistical Model

We had a nested data set with editors embedded in projects. Every editor in a project was a data point. Because our dependent variable, productivity, is a count variable with non-normal distributions truncated at zero, we used negative binomial regression models. Because the analysis compared the editing differences of multiple editors in the same project, we used the random effects negative binomial regression model, with each WikiProject as a group [13].

4.1.2 Results

Table 3 shows the results of our analysis of article productivity. Because we performed grand mean centering, our baseline was the mean values of all variables. The effects were assessed as *Incidence Rate Ratios* (IRR), the ratio of change of the dependent variable when an independent variable increases by one unit (in our case, by one standard deviation).

Table 3 presents three models predicting member productivity. Model 1 includes only the control variables, Model 2 adds the independent variables, and Model 3 adds the interaction terms between the independent variables and the number of projects an editor was involved in.

According to Model 1, project age, project members, editor tenure, and the number of projects are negatively associated with member productivity. The longer the project had existedIRR = 0.888, p <.001), the more members the project had (IRR = 0.959, p <.001), the longer tenure the editor had with Wikipedia (IRR = 0.873, p <.001), and the more other projects the editor was involved in (IRR = 0.589, p <.001), the less productive the editor was after joining

	Article Productivity						Article-related Communication						Project Coordination					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
	IRR	P	IRR	P	IRR	P	IRR	P	IRR	P	IRR	P	HR	P	HR	P	HR	P
Intercept	0.165	***	0.165	***	0.164	***	0.118	***	0.118	***	0.120	***	0.500	***	0.502	***	0.505	***
Project Age	0.888	***	0.878	***	0.877	***	0.836	***	0.826	***	0.828	***	1.037	***	1.034	***	1.034	***
Project Member	0.959	***	0.947	***	0.949	***	0.984	***	0.942	***	0.940	***	0.774	***	0.765	***	0.765	***
Project Article	1.392	***	1.385	***	1.384	***	1.387	***	1.386	***	1.375	***	1.016	***	1.014	***	1.014	***
Editor Tenure	0.873	***	0.877	***	0.875	***	0.869	***	0.878	***	0.878	***	0.963	***	0.965	***	0.966	***
Editor Prior Edits	2.349	***	2.189	***	2.209	***	2.168	***	1.924	***	1.904	***	1.124	***	1.082	***	1.077	***
Number of Projects	0.589	***	0.600	***	0.595	***	0.680	***	0.676	***	0.680	***	0.924	***	0.927	***	0.931	***
Identity-based			1.170	***	1.178	***			1.140	***	1.168	***			1.039	***	1.048	***
Bonds-based			1.085	***	1.073	***			1.178	***	1.195	***			1.051	***	1.059	***
Identity X Bonds					0.988						0.975	***					1.010	***
Identity X No. of Projects					1.018	***					1.047	***					1.011	***
Bonds X No. of Projects					1.017	***					0.982	***					0.987	***
Likelihood-ratio Test			5191.20	***	73.80	***			4537.58	***	156.71	***			729.12	***	108.68	***

Table 3: Results of Predicting Three Productivity Measurements – Article Productivity, Article-related Communication, and Project Coordination

a project. On the other hand, the larger the project scope (IRR = 1.392, p <.001), and the more edits editors made before joining (IRR = 2.349, p <.001), the more productive they were.

Despite the great strength of control variables such as prior edits (Model 1), adding identity-based attachment and bond-based attachment (Model 2) significantly increases the ability to predict participation. We assessed the significance of the incremental variance explained by performing the likelihood-ratio test, similar to assessing changes in R-square in traditional regression analysis. Likelihood-ratio test statistics suggests that Model 2 explained more variance and has a much better fit than Model 1.

Model 2 in Table 3 shows that both pre-joining identity-based and bonds-based attachments had significant positive effects on editor productivity. The effect size of identity-based attachment was larger than the size of bonds-based attachment. One standard deviation increase in identity-based attachment increased editor's productivity by 17.0%, while one standard deviation increase in bonds-based attachment increased editor's productivity by 8.5%. The coefficient test in Table 5 suggested the difference between the two coefficients was significant.

According to Model 3, the number of projects the editor joined has a significant negative effect (IRR = 0.600, p <.001). One standard deviation increase in the editor's number of projects decreased productivity by 40%. Although the number of projects the editor was involved in had a negative effect on productivity, its interactions with the two attachments were significant and positive (IRR = 1.018, p <.001 with identity-based; IRR = 1.017, p <.001 with bonds-based), which indicated that editors with strong pre-joining attachments to the project were less vulnerable to the pressure of being involved in multiple projects. The interaction between the two types of attachments, however, was not significant at the 0.05 level.

Overall, our results supported Hypotheses 1a, 2a and 3a. Both pre-joining identity-based attachment and bonds-based attachment increased article productivity, and the effect of identity-based attachment on productivity was stronger than bonds-based attachment.

4.2 Article-related Communication and Project Coordination

We analyzed two alternative measures of productivity, the degree to which editors contributed to article talk page communication and the degree to which editors participated in project talk page communication. We refer to these variables as article-related communication and project coordination. Similar to article edits, these were also count data so we ran negative binomial regression models with random effects.

4.2.1 Article-related Communication

Table 3 presents three models predicting article related communication.

Model 2 in Table 3 shows that both pre-joining identity-based and bonds-based attachments had significant positive effects on editor's article-related communication. The effect size of bonds-based attachment (IRR = 1.178, p <.001) was stronger than identity-based attachment (IRR = 1.140, p <.001), which means people who had strong bonds-based attachment to the project communicated more on article talk pages than people with identity-based attachment.

Model 3 shows the effects of interaction terms. Being involved in multiple projects had a negative effect on article related communication (IRR = 0.680, p <.001). However, its interaction with identity-based attachment was positive and significant (IRR = 1.047, p <.001), suggesting that identity-based attachment provided a buffer for the negative effect of being involved in multiple projects. The interaction with bonds-based attachment was negative (IRR = 0.982, p <.001). Interestingly, when people had both strong identity-based and bonds-based attachments, their article-related communication decreased (IRR = 0.975, p <.001).

4.2.2 Project Coordination

Table 3 presents three models predicting project coordination of editors after they join a project.

According to Model 2, both types of attachments had a positive, significant effect on project coordination, and the effect size of bonds-based attachment (IRR = 1.051, p <.001 with bonds-based) was stronger than that of identity-based attachment (IRR = 1.039, p <.001). In other words, editors with strong bonds-based attachment contributed more on project pages for project planning and development than editors with identity-based attachment. Model 3 shows the interaction terms. Again, identity-based attachment served as a buffer for the negative effects of editors being involved in multiple projects (IRR = 1.011, p <.001), but the interaction between number of projects and bonds-based attachment was not significant (IRR = 0.987, p <.001).

To summarize, the results of two additional productivity measures (i.e., article-related communication and project coordination) provided mixed support for our hypotheses. Both identity-based and bonds-based attachment increased article-related communication and project coordination, consistent with Hypothesis 1a and

2a. However, contra Hypothesis 3a, identity-based attachment did not have a stronger effect than bonds-based attachment. Instead, the effect of bonds-based attachment on article-related communication and project coordination was slightly stronger than identity-based attachment, and the difference was statistically significant. We will discuss the implication of these results later.

4.3 Withdrawal

4.3.1 Statistical Model

We ran survival analysis to predict withdrawal since the complete survival time for some editors cannot be censored due to truncation of the data set.

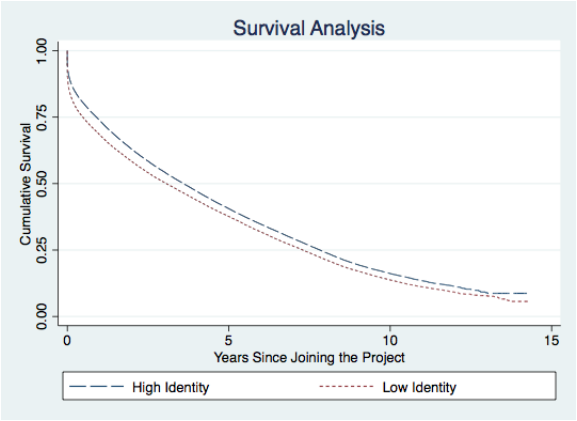

Figure 1: Survival Curves for Members with Different Levels of Identity-based Attachment

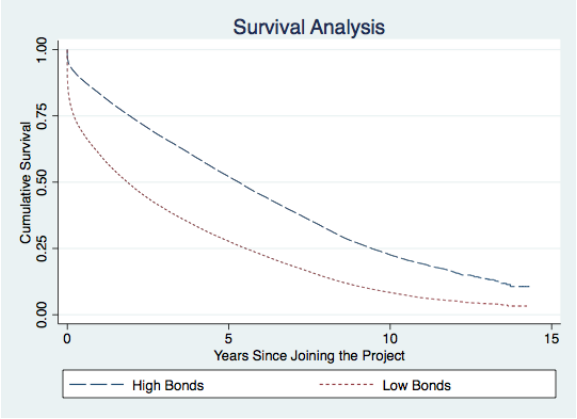

Figure 2: Survival Curves for Members with Different Levels of Bonds-based Attachment

4.3.2 Results

Survival analysis is used for analyzing the expected duration of time until one or more events happen. In our case, the event is the withdrawal of project members. It solves the problem of the proportion of the population which will survive past a certain time, reported by *Hazard Ratio* (HR). HR can be interpreted as the ratio change of the hazard (i.e., the likelihood of withdrawal in our case) when increasing an independent variable by one unit. More specifically, if HR is greater than 1, then withdrawal increases as the independent variable increases; if HR is smaller than 1, then withdrawal decreases as the independent variable decreases.

	Withdrawal					
	Model 1		Model 2		Model 3	
	HR	P	HR	P	HR	P
Intercept	0.278	***	0.502	***	0.024	***
Project Age	1.051	***	1.061	***	1.063	***
Project Member	1.133	***	1.148	***	1.142	***
Project Article	0.620	***	0.622	***	0.623	***
Editor Tenure	0.665	***	0.661	***	0.663	***
Editor Prior Edits	0.515	***	0.553	***	0.534	***
Editor Workloads	1.289	***	1.275	***	1.294	***
Identity-based			0.931	***	0.914	***
Bonds-based			0.912	***	0.958	***
Identity X Bonds					0.982	***
Identity X Workloads					0.978	***
Bonds X Workloads					0.933	***
Likelihood-ratio Test			1530.74	***	599.50	***

Table 4: Results of Predicting Withdrawal

Table 4 shows the results of the survival analysis. According to Model 1, the longer a project had existed (HR = 1.051, p <.001), the more members the project had (HR = 1.133, p <.001), and the more projects editors were involved in (HR = 1.289, p <.001), the more likely editors were to withdraw. On the other hand, the larger the project scope (HR = 0.622, p <.001), the more senior the editor (HR = 0.665, p <.001), and the more prior edits the editorhad made (HR = 0.515, p <.001), the less likely the editor was to withdraw.

Model 2 showed that both identity-based (HR = 0.931, p <.001) and bonds-based attachments (HR = 0.912, p <.001) had significant effects on reducing withdrawal. Bonds-based attachment had a larger effect size, meaning bonds-based attachment had a stronger effect on member withdrawal than identity-based attachment. Table 5 showed that the difference between the coefficients of the two attachments was significant at 0.001 level. According to Model 3, the number of projects an editor was involved in increased withdrawal (HR = 1.294, p <.001). The interactions between the number of projects and the two attachments were negative (HR = 0.978, p <.001 with identity-based; HR = 0.933, p <.001 with bonds-based), meaning that both attachments served as buffers for the negative effects of being involved in many projects and reduced the chance of editors withdrawing from projects. Bonds-based attachment had a stronger effect than identity-based attachment.

We plotted the survival curves by dividing editors into four groups: editors with high versus low pre-joining identity-based attachment and editors with high versus low pre-joining bonds-based attachment. We used the median values of the two attachments to do the split. As shown in Figure 1, there was a small difference between low and high identity-based attachment groups in editors' survival. In contrast, as shown in Figure 2, there was a large differ-

[5] Estimated by the mean of article productivity, +17.0% increase is equivalent to 43.8 more edits.
[6] Estimated by the mean of article-related communication, +14.0% increase is equivalent to 7.9 more edits.
[7] Estimated by the mean of project coordination, +3.9% increase is equivalent to 0.4 more edits.
[8] Estimated by the mean of article productivity, +8.5% increase is equivalent to 21.9 more edits.
[9] *** means p <.001.

	Article Productivity	Article-related Communication	Project Coordination	Withdrawal
Increasing Identity-based Attachment by 1SD	+17.0%[5]	+14.0%[6]	+3.9%[7]	-6.9%
Increasing Bond-based Attachment by 1SD	+8.5%[8]	+17.8%	+5.1%	-8.8%
Significance of Coefficient Difference χ^2	394.09 ***[9]	56.71 ***	14.04 ***	22.92 ***

Table 5: Summary of the Main Effects of Two Types of Attachments

ence between low and high pre-joining bonds-based attachments. Editors with high bonds-based attachment were less likely to withdraw from the projects compared to those with low bonds-based attachment.

To summarize, our results supported Hypotheses 1b, 2b and 3b. Identity-based attachment and bond-based attachment significantly reduced withdrawal, and the effect was stronger for bonds-based attachment.

5. DISCUSSION

We summarize our main results, i.e., the effects of identity-based and bonds-based attachments on productivity and withdrawal in table 5. Consistent with our hypotheses, pre-joining attachments, both identity-based and bonds-based, increased all three measures of productivity and reduced member withdrawal. The numbers in table 5 also showed that the effects of both types of attachment were practically significant and not trivial. For instance, one standard deviation increase in the identity-based attachment was associated with 17.0% increase in article productivity, which was equivalent to 43.8 more edits. One standard deviation increase in bonds-based attachment increased article productivity by 8.5%.

Our results also suggest that the relative effects of the two types of attachment differed across the outcome measures. More specifically, identity-based attachment had a more substantial effect on increasing article productivity while bonds-based attachment had a stronger effect on the other measures, such as increasing article-related communication and project coordination and reducing member withdrawal.

One way to explain the stronger effects of bonds-based attachment on article-related communication and project coordination may be that members with stronger bonds-based attachment are more interested and have an advantage in participating in activities that require interpersonal exchanges with other members of the project. Article-related communication and project coordination need intense interchange of thoughts, opinions, and information with other members in the group. Members who have interacted and developed bonds with other members are more likely or more willing to take on and succeed in performing these tasks.

Overall, our findings advance knowledge of member productivity and withdrawal in the online production communities by tackling it from a novel perspective, i.e., members' pre-joining connections to the subgroup. We bridge attachment theories in social psychology and user behaviors in the online production communities. The results clearly demonstrate the importance of pre-joining connections to groups in online production communities.

5.1 Implications for Design

Our findings can inform the design of cost-effective and targeted recruitment strategies in online production groups. One effective strategy in recruiting newcomers for WikiProject is for existing members of a WikiProject to send out invitation messages to potential new members. Compared to standardized invitations, these personalized invitations were much more effective in recruiting new members [6]. Meanwhile, personalized invitations were a costly way to recruit new members because old-timers' time and efforts are limited and sending out too many personalized messages may distract them from performing other tasks critical to the project. Our results can help project old-timers to allocate and focus their attentions on members with preexisting connections to the project, which signals their potential to become active contributors and sustaining members. More importantly, depending on the recruiting goals, old-timers could target new members with different types of preexisting connections. For example, if the aim is to boost productivity, they could target those whose interests are more aligned with the project's topical coverage (i.e., people with high identity-based attachment). If the goal is to improve project communication and coordination or to improve member retention, old-timers should target those who have strong personal connections with existing members (i.e., people with high bonds-based attachment).

Because many online communities are very large, with 100K members or more, it may be difficult or impossible for individual members to identify potential members to approach. Our work can inform the design of visualization or analytic tools to help solve this problem. These tools should automatically calculate identity-based attachment and bonds-based attachment of all editors to a focal project, and narrow it down to a manageable list which project leaders can use to identify and target promising new members.

5.2 Generalization

As we mentioned in the introduction, Wikiprojects are a prominent example of online production subgroups within larger communities. We believe our findings can be generalized to other online production groups, such as projects in Github and SourceForge, groups in GoodReads, and projects in OpenStreetMap. Take OpenStreetMap as an example; we can similarly predict project members' productivity and withdrawal based on their pre-joining connections to the project, i.e., how well their interests align with the project goals based on prior map creation activities and communication pattern with existing project members. We encourage replication of our study in other online production communities.

5.3 Limitations

Our study is not without limitations. First, our study applies more to large and developed communities because they are more likely to have established subgroup structures and have historical data to calculate potential contributors' identity-based and bonds-based attachment. Smaller platforms and fledgling communities may not have sufficient data to identify the attachment type of potential members.

Second, we operationalized attachment solely based on behavioral measures such as prior edits for identity-based attachment and interactions with existing members for bonds-based attachment. The strength of this approach is that it allows for automated implementation which is low cost and scalable. At the same time, we acknowledge that attachment is also a psychological state with affective connection [27]. While we believe our behavioral measures serve as a good proxy for both types of attachments, it would be

good to complement the behavioral data with additional measures such as self-report responses in future studies.

6. CONCLUSION

To summarize, we investigated how identity-based and bonds-based attachments to subgroup that existed prior to joining the subgroup affected members' behaviors in that subgroup. We tested our hypotheses in the context of WikiProjects, by examining the relationships between individual's pre-joining attachments and the amount of the work they accomplished, the likelihood of withdrawal after they joined the group, and the interactions they had with other group members. We found that member behaviors after joining a project were significantly determined (and therefore could be predicted) by their pre-joining experiences with the project. Both identity-based and bonds-based attachments increased editor productivity and reduced likelihood of withdrawal. Our work also extended the social psychology literature by demonstrating the differential effects of identity-based and bonds-based attachments on productivity and withdrawal. Our findings provide design implications for recruiting and socializing new members of subgroups in online production communities.

7. ACKNOWLEDGMENT

We thank Aaron Halfaker, Brent Hecht, and the members of the GroupLens Research at the University of Minnesota for helpful feedback.

8. REFERENCES

[1] B. E. Ashforth, S. H. Harrison, and K. G. Corley. Identification in organizations: An examination of four fundamental questions. *Journal of management*, 34(3):325–374, 2008.

[2] K. W. Back. Influence through social communication. *The Journal of Abnormal and Social Psychology*, 46(1):9, 1951.

[3] A. Bruns. *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*, volume 45. Peter Lang, 2008.

[4] B. S. Butler. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information systems research*, 12(4):346–362, 2001.

[5] J. Chen, Y. Ren, and J. Riedl. 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*, pages 821–830. ACM, 2010.

[6] B. Choi, K. Alexander, R. E. Kraut, and J. M. Levine. Socialization tactics in wikipedia and their effects. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 107–116. ACM, 2010.

[7] R. Farzan, L. A. Dabbish, R. E. Kraut, and T. Postmes. Increasing commitment to online communities by designing for social presence. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 321–330. ACM, 2011.

[8] L. Festinger, K. W. Back, and S. Schachter. *Social pressures in informal groups: A study of human factors in housing*. Number 3. Stanford University Press, 1950.

[9] A. Forte, N. Kittur, V. Larco, H. Zhu, A. Bruckman, and R. E. Kraut. Coordination and beyond: social functions of groups in open content production. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 417–426. ACM, 2012.

[10] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 211–220. ACM, 2009.

[11] N. Gross and W. E. Martin. On group cohesiveness. *American Journal of Sociology*, pages 546–564, 1952.

[12] M. A. Hogg and J. C. Turner. Interpersonal attraction, social identification and psychological group formation. *European Journal of Social Psychology*, 15(1):51–66, 1985.

[13] P. Kennedy. *A guide to econometrics*. MIT press, 2003.

[14] A. Kittur, E. H. Chi, and B. Suh. What's in wikipedia?: mapping topics and conflict using socially annotated category structure. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1509–1512. ACM, 2009.

[15] A. Kittur, B. Pendleton, and R. E. Kraut. Herding the cats: the influence of groups in coordinating peer production. In *Proceedings of the 5th international Symposium on Wikis and Open Collaboration*, page 7. ACM, 2009.

[16] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi. He says, she says: conflict and coordination in wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 453–462. ACM, 2007.

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

[18] T. Kriplean, I. Beschastnikh, and D. W. McDonald. Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 47–56. ACM, 2008.

[19] J. M. Levine and R. L. Moreland. *Small groups: key readings*. Psychology Press, 2008.

[20] N. Michinov, E. Michinov, and M.-C. Toczek-Capelle. Social identity, group processes, and performance in synchronous computer-mediated communication. *Group Dynamics: Theory, Research, and Practice*, 8(1):27, 2004.

[21] R. L. Moreland. The formation of small groups. 1987.

[22] J. T. Morgan, M. Gilbert, D. W. McDonald, and M. Zachry. Project talk: Coordination work and group membership in wikiprojects. In *Proceedings of the 9th International Symposium on Open Collaboration*, page 3. ACM, 2013.

[23] R. M. OâĂŹbrien. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5):673–690, 2007.

[24] K. Panciera, A. Halfaker, and L. Terveen. Wikipedians are born, not made: a study of power editors on wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*, pages 51–60. ACM, 2009.

[25] J. Preece and D. Maloney-Krichmar. Online communities: focusing on sociability and usability. *Handbook of human-computer interaction*, pages 596–620, 2003.

[26] D. A. Prentice, D. T. Miller, and J. R. Lightdale. Asymmetries in attachments to groups and to their members: Distinguishing between common-identity and common-bond groups. *Key Readings in Social Psychology*, page 83, 1994.

[27] Y. Ren, F. M. Harper, S. Drenner, L. G. Terveen, S. B. Kiesler, J. Riedl, and R. E. Kraut. Building member attachment in online communities: Applying theories of

group identity and interpersonal bonds. *Mis Quarterly*, 36(3):841–864, 2012.

[28] Y. Ren, R. Kraut, and S. Kiesler. Applying common identity and bond theory to design of online communities. *Organization studies*, 28(3):377–408, 2007.

[29] K. J. Stewart and S. Gosain. The impact of ideology on effectiveness in open source software development teams. *Mis Quarterly*, pages 291–314, 2006.

[30] H. Tajfel, M. G. Billig, R. P. Bundy, and C. Flament. Social categorization and intergroup behaviour. *European journal of social psychology*, 1(2):149–178, 1971.

[31] J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, and M. S. Wetherell. *Rediscovering the social group: A self-categorization theory.* Basil Blackwell, 1987.

[32] J. C. Turner and P. J. Oakes. The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, 25(3):237–252, 1986.

[33] L. S. Wang, J. Chen, Y. Ren, and J. Riedl. Searching for the goldilocks zone: trade-offs in managing online volunteer groups. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 989–998. ACM, 2012.

[34] Y.-C. Wang, R. Kraut, and J. M. Levine. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 833–842. ACM, 2012.

[35] H. Zhu, J. Chen, T. Matthews, A. Pal, H. Badenes, and R. E. Kraut. Selecting an effective niche: an ecological view of the success of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 301–310. ACM, 2014.

[36] H. Zhu, R. Kraut, and A. Kittur. Effectiveness of shared leadership in online communities. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 407–416. ACM, 2012.

[37] H. Zhu, R. Kraut, and A. Kittur. Organizing without formal organization: group identification, goal setting and social modeling in directing online production. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 935–944. ACM, 2012.

[38] H. Zhu, A. Zhang, J. He, R. E. Kraut, and A. Kittur. Effects of peer feedback on contribution: a field experiment in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2253–2262. ACM, 2013.