Newcomer Retention and Productivity in Online Peer-Production Communities

A Dissertation

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Raghav Pavan Srivatsav Karumur

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Joseph A. Konstan

July, 2018

Dedication

THE SOURCE OF EVERYTHING

Acknowledgments

A number of people have supported my journey as a doctoral student. I want to begin by thanking my adviser Joseph Konstan. He was patient with me when things did not work, gave me useful directions and pointers when I got stuck, encouraged me in both successes and failures, guided me patiently in my career decisions, taught me so many things about being a good researcher, and even being a person in life, and was always available to respond to my emails, to schedule meetings, and mentored me even multiple times when preparing for a talk.

I want to thank the National Science Foundation for funding my research through the grant IIS-1319382, and many anonymous reviewers who criticized my work and made it better. My gratitude goes to Daniel, Francois, Melissa, and Julie for being able to provide me letters of recommendation to get into the Ph.D. program at the nick of the moment, to Georganne Tolaas and Rui Kuang for their assistance in bringing me into this program, to Morten Warncke-wang, Daniel Kluver, Abhisek Mudgal, Max Harper, and Tien Nguyen for their mentoring during the initial years of my Ph.D., to Haiyi Zhu for advising me on some of my projects and supporting me and encouraging me in my job search, to Rui Kuang and Adam Rothman for serving on my committee, to Qian Zhao for all the interesting research conversations I had with him, and to other GroupLensers who have been good friends and supporters, to my mentor Andrew Frenkiel and manager Evelyn Duesterwald for giving me an opportunity to intern at IBM Research, to Ashish Karn, and Sri Vasudevacharyulu garu and Smt. Srivani garu for their emotional support and encouragement during tough times, and to Gokul Hariharan and Sri Sateesh Burgadda garu for their physical support and care.

My special thanks to my mother for believing in me, supporting me, and praying for my well-being all along, for all the hard work she had done to bring me up and make sure

I have the best of education, and to my father for everything he has done to me. My thanks also to all my other family members, especially my uncles and aunts who supported me in various ways during this intense time. My thanks to my late grandfather, who had dreamed of getting me into the Indian Administrative Services but transgressing whose wish I chose to pursue a Ph.D. I hope he will be happy with my achievement and the overall direction I have taken in my life. My thanks also to my late grandmother for the effort and love she invested in raising me despite all her ill-health.

My very special thanks to my spiritual preceptors and mentors Bhagavad Ramanuja, Dr. P.V. Krishnan, Swami Bhaktivedanta, Chinna Jeeyar Swamiji, Gopalacharya Swami, and Swami Sivananda - whose life and teachings have inspired me and whose blessings have shaped me spiritually and given me the strength to do my duties sincerely in both stormy and sunny weathers.

My most special thanks to the Lord for the time He has provided, and for being the source of strength, support, inspiration, intelligence, wisdom, memory, courage, determination, enthusiasm, resources, and for everything else.

Abstract

Online communities are online interaction spaces for people that break the barriers of time, space, and scale and provide opportunities for companionship and social support, information exchange, retail and entertainment. Among them are online peer-production communities that have a fantastic business model where volunteers come together to produce content and drive traffic to these sites. Although as a class these communities are successful, the success of individual communities greatly varies. To become and remain successful, these communities must meet a number of challenges related to starting communities, retention of members, encouraging commitment, and contribution from their members, regulating behavior of members and so on.

This dissertation focuses on the specific challenge of newcomer retention and productivity in the context of online peer-production communities. Exploring three different communities with entirely different structures and compositions – MovieLens, GitHub, and Wikipedia and building upon prior work in this space, this dissertation offers a number of important predictors of retention and productivity of newcomers. First, this dissertation explores the value of early activity diversity in the presence of amount of early activity as a predictor of newcomer retention. Second, this dissertation digs into more fundamental psychological traits of newcomers such as personality and presents findings on relationships between personality and newcomer retention, preferences, and productivity. Third, this dissertation explores and presents results on relationship between community interactions (apart from norms, policies and rigid structures) and newcomer retention. Fourth, this dissertation studies and presents the effects of various kinds of prior experience of newcomers on retention and productivity in a new group they join. This dissertation concludes by offering a number of directions for future research.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Online Communities

Online Communities are among the most popular destinations on the Internet. They are spaces where people come together in a website to converse, exchange knowledge, experiences, information, media, or digital goods, learn, play, or just be with each other and sometimes use public or semi-public user profiles to form personal relationships online through communication and interaction [E+07, KRK+12]. A 2012 Forrester Research Study shows that 84% of U.S. Adults use the web daily by 2012 [1] and among them are users who use online communities such as Facebook actively, and make purchases online.

According to Preece et al. [Pre00], an online community consists of (i) a computer system to support social interaction and facilitate a sense of togetherness with (ii) a shared purpose, such as an interest, a need, or a service that provides a reason for the community to exist where (iii) policies, in the form of tacit assumptions, rituals, protocols, rules, and laws guide people's interactions where (iv) people interact socially as they strive to satisfy their own needs. But the term applies to any social configuration from small close-knit groups to sites with millions of participants. The area of Computer Science called *social computing* deals with the success of social behavior of users in online communities.

[1] https://techcrunch.com/2012/12/19/forrester-84-of-u-s-adults-now-use-the-web-daily-50-own-smartphones-tablet-ownership-doubled-to-19-in-2012/

1.2 The Promise of Online Communities

Online communities are not new. Long before social networking sites like Facebook emerged, early forms of online communities existed. Back in the 1960s and 70s, first computer networks were established to facilitate the connection between geographically dispersed people [WSD+96]. Over the years, text-based communication systems such as the Usenet newsgroups, bulletin boards, mailing lists or chats have been used to exchange information and connect people electronically [HH98, Rhe93]. Today, these communities have grown to what we call Recommender Systems such as MovieLens[2] where users rate and review products and services, e-commerce sites such as Amazon where users purchase products too, online groups such as those in Wikipedia or Wikia where users contribute scholarly articles, Q&A sites such as TheStackOverflow, Open Source Software code repositories such as those in GitHub, weblogs for expressing one's views, microblogs such as Twitter to express current feelings and opinions, and other Internet forums for other kinds of information exchange.

Online communities break the barriers of time, space, and scale that limit offline interactions but essentially serve the same range of purposes that offline groups, networks, and communities serve: They provide their members with opportunities for information exchange, companionship and social support, and entertainment. Some of them also provide non-members with product information, code-bases, reviews, or content. People with addiction, in recovery or with unusual medical conditions can get social support from others who share their condition or who can assist but live far away, and they can do so whenever they need it rather than only at a weekly or a monthly scheduled meeting. On Ravelry[3], a hobby community for people who knit and crochet, knitters can share patterns with thou-

[2] http://movielens.org
[3] http://www.ravelry.com

sands more people than they could stitch with in person. On Q&A sites, users get answers from experts living across the globe.

1.3 The Challenges of Online Communities

Online communities have a fantastic business model where volunteers come together to produce content and drive traffic to these sites. However, although as a class these communities are successful, the success of individual communities greatly varies. Some struggle to succeed, whereas most fail. To become and remain successful, online communities must meet a number of challenges that are common to many groups and organizations in general, offline as well as online. The challenges can be broadly classified under the following categories:

- Starting a New Online Community
- Encouraging Contribution
- Encouraging Commitment
- Regulating Behavior
- Dealing with Newcomers

A brief overview of each challenge is presented below:

1.3.1 Starting a New Online Community

When creating a community from scratch, community organizers are faced with three major challenges. The first is to carve out a useful niche in terms of (i) the purpose of the community, the scope of the community (i.e., the breadth of topics to cover), and the kinds of people

to try to attract as members, (ii) the extent of compatibility and integration with other sites, including the borrowing of features and user interface elements, the sharing of user identifiers, and the import and export of content and people, and (iii) the internal organization of the people, content, and activities within the community. The second is to defend that niche in the ecology of competing online communities and alternative ways that potential members can spend their time. The third is that the new site doesn't have enough content to attract users and there are too few users to create the content that might attract others. This is the reason why a vast majority of online communities never really get off the ground. For example, hundreds of proposals for new Q&A sites are made on StackExchange, but most get closed because of lack of sufficient members supporting the ideas.

1.3.2 Encouraging Contribution

A number of online communities work on the model that their volunteers build their content. Therefore, in order to be successful, online communities need volunteers who make contributions based on their need. There are two kinds of challenges here. The first challenge is of volunteers who contribute little in general. For instance, on Gnutella, a popular peer-to-peer music sharing service, two-thirds of users share no music files and 10% of them provide 87% of all the music [AH00]. And, in a majority of active mailing lists, it was found that fewer than 50% of subscribers posted even a single message in a four-month period [But99].

The second challenge is the challenge of imbalance in contributions. Often, there are volunteers who are experts in certain areas and not others. As a result, popular topics of expertise get more contributions whereas the less popular ones suffer from under-contribution. Therefore, it is not uncommon to find even highly popular communities like Wikipedia suffering from the problems of under-contribution in some areas. For instance, roughly two-

thirds of the articles in the English version of Wikipedia have been classified as articles with only a few sentences of content that are too short to provide encyclopedic coverage of a topic [Sta17]. About half of social, hobby, and work mailing lists had no traffic over a four-month period [But99]. More than 20% of the movies listed in MovieLens do not have sufficient ratings for recommender algorithms to be able to make accurate predictions about whether users will like them [BLW+04].

Although not everyone needs to contribute for an online community to be successful [NP00], online communities with large proportions of under-contributors or under-contributions in some areas have difficulty providing needed services. As we see above, in Wikipedia, certain articles don't end up having sufficient content. In OSS projects, bugs remain unfixed and enhancements are not delivered, and less popular movies may not be evaluated. Therefore, this is a challenge online communities have to deal with in order to be successful.

1.3.3 Encouraging Commitment

Creating commitment is more difficult to accomplish in an online community than an offline community because the forces keeping the volunteers online are weak and non-binding. Conventional organizations have employment contracts and offer monetary benefits in most cases. In contrast, most developers in OSS projects participate voluntarily, with neither employment contracts nor monetary benefits encouraging them to stay and contribute. The physical location of the conventional organization also places constraints on members' willingness to go elsewhere. If someone wants to leave a job, church, or club, for example, only a relatively small number of alternatives are nearby and convenient to join. In contrast, if someone wants to leave a particular online community, they could join any other comparable community online with no constraints imposed by geographic proximity. Although

most online communities offer virtual rewards like leaderboards, statuses, badges, or barnstars, it is easy for volunteers (at any stage of volunteering) to leave when faced even with minor adversities. Therefore, encouraging commitment is a challenge.

1.3.4 Regulating Behavior

The volunteers who come to online communities often have different and sometimes competing interests. Most large online discussion groups (especially those that deal with controversial topics) attract trolls, people who post controversial, inflammatory, irrelevant, or off-topic messages to provoke other users into an emotional response. Some have manipulators who try to make the community produce particular outcomes (e.g., to pump up the rating of a restaurant). Commercial spammers often try to drive traffic to their external websites. In more mundane conflicts of interest, some participants in a hobby site may prefer that the discussion stay focused on the hobby, but others may want to engage in more personal conversation with other members they have become friends with. Most sites in the StackExchange [4] forum discourage opinion-type questions although users sometimes post them. When there are conflicting interests in a group, there must be mechanisms to help participants regulate behavior. The challenges here are to deter inappropriate behavior by group members, prevent trolls and other outside attackers, and limit the damage that is caused when inappropriate behavior occurs.

Although these challenges confront almost all groups and organizations, online communities may have more difficulty overcoming them than conventional groups and organizations because of three characteristics that are typical of online communities but unusual in conventional groups and organizations. The first is that users are allowed to post/contribute anonymously. Disallowing this would make the newcomers feel the pressure for social ac-

[4] https://stackexchange.com

countability which they may not be ready for yet. The second is ease of entry and exit: spammers and trolls can enter easily, and when strict regulations are imposed that cause inconvenience a normal user may leave the site. The third is the use of text and emojis for communication which may be prone to misinterpretation because a) they lack some of the fluidity and nonverbal cues of face-to-face interaction and b) emojis are prone to both miscommunication as well as misinterpretation due to platform incompatibility [MTSC+16].

1.3.5 Dealing with Newcomers

Newcomers in any community serve as sources of new energy, activity, fresh perspectives, skills, innovation, and work procedures [AC92, KRK+12], and maintain the critical mass [SW14]. A higher number of newcomers to a site is an indication that the community is still valued among its competitors. A higher number of newcomers also attracts more business opportunities and ads leading to profits for the online community. Therefore, newcomers are important to an online community and even established online communities must attract a stream of newcomers to replace others who leave. When dealing with newcomers, online communities must solve six basic problems [KRK+12]:

- **Newcomer Recruitment** Recruiting newcomers is important to replenish periodic member loss and for the growth of the community.

- **Newcomer Selection** As much as recruiting is important, it is also crucial to make sure that the members who come in fit well with the community for the overall health of the community. Mismatch in newcomer goals and community goals will leave the members themselves dissatisfied and unproductive. It may also lead to behaviors which can be harmful to the groups' success and undercut the smooth functioning of the group. Therefore, it is important to identify and encourage potential members

who have the characteristics, skills, and motivation to contribute. For instance, while OSS projects are often looking for potential developers with the right skills to join, many health-support groups often try to screen out inappropriate members and restrict membership to people who have a particular illness or care for someone who does.

- **Newcomer Retention** Despite bringing in members with the right match, both theory and practical experience suggests that attracting and retaining new volunteers in online communities is challenging. The problem of newcomer retention is different from maintaining commitment of other members discussed in section 1.3.3 Newcomers are usually very sensitive to the public image of a community and to their own early experiences in it and may leave or not join when faced with even minor adversity. Also, newcomers, who are potentially choosing from among other similar online communities to join, frequently have insufficient information to make their choices and almost always have less commitment to a community than more established members or old-timers have. Therefore, their connection with the community is even more fragile and it is important to understand what factors lead to retention of newcomers.

- **Newcomer Productivity** Newcomers do not feel the same commitment to the group as felt by old-timers. Therefore, they are less motivated to be helpful to the community or to display good organizational citizenship behavior characteristic of most old-timers [OR95].

- **Newcomer Socialization** Different communities have standards and norms that shape and constrain the behavior of their members. Some of these norms are broad and open to different interpretations whereas others are more narrowly targeted. While some communities explicitly state these norms, they are implicit in others and must

be learned by observation. Because the newcomers have not yet learned the appropriate ways to behave in the community, they may behave in ways that undercut the smooth functioning of the group. When they try to participate, they may offend other members, disrupt the activity of existing members or imperil the work that other community members have already performed. For instance, they may introduce bugs in an OSS project they have joined, cause the (virtual) death of fellow group members in an online role-playing game or ask redundant questions in discussion groups. When participating in Wikipedia, new editors may fail to write in a neutral tone (which is required according to Wikipedia policy guidelines), or may end up adding content to an article that more experienced editors and moderators have determined to belong to a different article. Therefore, teaching them how to climb the ropes is important as part of a socialization process.

- **Protection of the Community from the Newcomers** It is necessary for existing members to socialize with newcomers through friendly initial interactions and explicitly avoid being hostile. However, this does not mean that the newcomers should receive carte blanche access to the community and its resources including people and artifacts produced by them. As mentioned above, there is a chance that newcomers pose real threats to a community they join. It is important to protect code from being committed by a new member to an OSS project like Apache, or protect buyers on Amazon from purchasing expensive items from relatively new sellers. Protection mechanisms should not only prevent damage but also discourage those who are disruptive and encourage those who are a good fit.

These problems vary in degree and nature across the various online communities. However, despite the limited direct control of individual people's actions, online communities

can be designed and managed to achieve the goals that their owners and members desire. This thesis focuses on the challenge of dealing with newcomers to online communities and in particular attempts to address the problems of retention and productivity.This dissertation focuses on the specific subset of peer-production online communities.

1.4 The Importance of Newcomer Retention and Productivity

In general, it is hard to retain newcomers. 80% of downloaded Android apps are no longer used after three days [Che15]. More specifically, research on peer-production online communities shows that an average of 60% of users who register in these communities do not return for a second time[ABJ+06, Duc05, PHT09]. Due to a lack of steady supply of active contributions or less than optimal contributions from existing volunteers, most peer-production communities fail quite early [But99, But01, LH03, MFH02]. And, techniques used in conventional organizations are not effective in retaining new online volunteers in these contexts because of lack of employment contracts, weak interpersonal bonds, and poor communication levels between the smaller pool of organizers and administrators, and the larger pool of other members. Therefore, in the face of inevitable turnover, every online community must make successive generations of newcomers to survive. For this, it is important to understand what factors predict their retention (and eventual productivity) in an online community.

Also, in online communities, committed members are those most likely to provide the valuable content such as answers to people's questions in technical and health support groups [BM04, FSW06, RC05], code in open source software projects [MFH02], and edits in Wikis [KCP+07]. They are the ones most likely to exercise voice, demanding change and improvement when dissatisfied [Hir70]. They are the ones that care enough to respond to

and to enforce norms of appropriate behavior [SMO97]. For newcomers to gain benefits from an online community and eventually become committed members who can assume such core responsibilities, it is important to keep them around long enough so that they can learn the ropes, understand what the community offers them and what they can offer back to the community, form relationships with other members in the community, and begin to identify with the community as a whole.

Modeling newcomers and their early user experience can aid in understanding their preferences better, customizing their early user experience by making appropriate recommendations of activities or products, and helping them adapt to the system better. Therefore, this thesis focuses on modeling newcomers and their early user experience in order to understand how various factors at the individual level as well as the community level predict newcomer retention and productivity in online peer-production communities. Specifically, this thesis focuses on the three peer-production communities Movielens, GitHub, and Wikipedia.

1.5 Organization of this thesis

The rest of this thesis is organized as follows:

- In Chapter 2, we present a survey of related work and background literature on the challenge of dealing with newcomer retention and productivity in online as well as offline settings.
- We began our work exploring factors based on a newcomer's early behavior within a site as predictors of newcomer retention. We were specifically interested in understanding how the breadth of exposure to various features in a site early on affects newcomer retention. This study was performed on a movie recommender system called MovieLens. The details of this study are laid out in Chapter 3.

- We then examined the effect of more fundamental traits such as a user's personality in predicting a newcomer's retention, activity preferences, usage of various features, intensity of user engagement, and distribution of user activity. This study was also performed on MovieLens. This study is described in Chapter 4.

- We then decided to look at how factors at the community level play a role in determining a newcomer's retention in a site. This study is presented using the case of GitHub in Chapter 5.

- We then wanted to understand how what online volunteers carry from their prior work experience affects their retention and performance in the context of a new online community. To answer this question, we reviewed prior literature and identified three types of prior experience and explored their effects on newcomer retention and productivity in the specific context of WikiProjects. A description of this study forms the content of Chapter 6.

- We summarize our findings and map out some future work in Chapter 7.

Chapter 2

Related Work

2.1 Introduction

Looking at newcomer churn is an example of the challenge of user churn in online communities, and customer churn, more broadly in online as well as offline businesses. We are not aware of any work that has studied churn of *new* customers in businesses. However, customer churn, more generally, has been studied in several settings such as banking, commerce, wireless telecommunication and subscription services. Below, we present a review of relevant work in each of these domains.

2.1.1 Banking

Dudyala et al. studied the churn of credit card customers in the banking domain [AKR08]. Mutanen et al. showed success in identifying customers who churn from those who don't using conventional statistical methods such as logistic regression using data from a consumer retailer banking company [MAN06].

2.1.2 Commerce

For e-commerce companies, very high turn rate has a devastating effect on their customer base. Therefore, early detection of potential churning customers enables them to target the churning customers using specific retention actions to increase their profits. Buckinx

et al. tried to predict the partial defection of loyal customers in the commerce domain using techniques such as Logistic Regression, Automatic Relevance Determination (ARD) Neural Networks, and Random Forests, and showed that future partial defection can be successfully predicted even exceeding the benchmark hurdle of the null model [BVdP05]. Using the real-life data of a European pay-TV company, Burez and Van den Poel built different churn-prediction models and showed that profits could be doubled using their best churn-prediction model [BVdP07].

2.1.3 Wireless telecommunication and Subscription services

Coussement and Van del Poel built churn prediction models using Support Vector Machines (SVM), Logistic Regression, and Random Forests in a newspaper subscription context and provided an overview of the most important churn drivers [CVdP08]. In mobile telecom networks, Dasgupta et al. explored the relationship between the likelihood of a subscriber churning out of a service provider's network and the number of social ties (friends) that have already churned [DSV+08]. Using second order social metrics, Richter et al. exploited the structure of customer interactions to predict which groups of subscribers are most prone to churn, before even a single member of the group has churned [RYTS10]. Datta et al. used learning methods such as decision trees and genetic algorithms for selecting features and a cascade neural network for predicting which customers will discontinue a cellular phone service for over a hundred cellular phone markets [DMML00]. Others used regression, regression trees, and neural networks to predict churn from complaints data [HTRR06, MWG+00]. Pendharkar used genetic-algorithm (GA) based neural network (NN) models to predict customer churn in subscription of wireless services [Pen09]. Verbeke et al. used advanced rule induction techniques to build customer churn prediction models [VMMB11]. Using subscriber contractual information and call pattern changes ex-

tracted from call details, Wei et al. predicted subscriber churn at the contract level for a specific prediction time-period and used multi-classifier class-combiner approach to deal with the high skew in class distribution between churners and non-churners [WC02].

2.1.4 Peer-to-Peer Systems

Peer-to-Peer systems are popular for file-sharing and content distribution. When a user launches the application, an application session is created. Other users called peers can join the application session, contribute some resources while making use of resources provided by the other peers, and leave the application session when the user exits the application. One such join-participate-leave cycle for a peer is called a peer session. Stutzbach and Rejaie studied several Peer-to-Peer systems such as Gnutella[1], BitTorrent[2], and Kad[3] and found that they all have very similar user churn characteristics [SR06] for peer sessions. One of their findings is that a portion of peers turn over quickly and this shows that studying this phenomenon is important.

2.1.5 MMORPGs

Multiplayer Online Role Playing Games also suffer from erosion of their customer base and this negatively impacts the word-of-mouth reports to existing and new customers contributing to further erosion in customer base. Kawale et al use social influence among players and their personal engagement in the game in order to predict user churn in Online Role Playing games and show that the combination of these two factors could lead to a more accurate churn prediction. [KPS09]. Borbora et al show how in terms of recall, ensembles perform notably better than single classifiers in predicting churn [BS12].

[1] http://www.gnutellaforums.com
[2] http://www.bittorrent.com
[3] https://en.wikipedia.org/wiki/Kad_network/

We want to highlight four things from this body of research on customer churn more broadly that are applicable for research on newcomer churn in online peer-production communities. (i) In any business, a portion of customers (users) leave very early. (ii) Customer (User) interactions with a certain system or organization and their interactions with each other within the system or organization are likely to affect user churn. (iii) Early identification of churners enables businesses to target them using specific strategies to increase profits. (iv) Conventional statistical methods and machine learning algorithms are able to predict churners from non-churners with reasonably high accuracy.

2.2 Newcomer Retention in Online Communities

Based on the above observations, we begin our research by studying what percentage of newcomers leave early.

2.2.1 Statistics

68% of newcomers to Usenet groups are never seen after their first post [ABJ+06]. 54% of developers who registered to participate in the Perl Open Source Development Project never returned after posting a single message [Duc05]. Newcomers to Wikipedia have high probability of leaving within few days with only 40% of contributors continuing to use after 500 days. 60% of registered editors in Wikipedia never make another edit after their first 24 hours of participating [PHT09]. 60% of MovieLens users do not return for a second time [KNK16]. 46% of the members of guilds in World of Warcraft leave their group within one month, migrating to other groups [WDX+06]. These statistics show the importance of understanding factors that contribute to early churn of newcomers in online peer-production communities.

Much of prior literature views the lack of new volunteer retention from three different perspectives [CT86, CHW05, QSTC14]. The first and more dominant view is about the negative effects of low retention of volunteers because the online community literature is generally in favor of sustaining a steady group of volunteers for continued production [KRK+12]. The negative effects include the loss of productive volunteers [Str90], the loss of social capital [DS01, Hus95], the cost of training new, inexperienced volunteers [Dar90], and the weakening of knowledge resources of the organization [Hus95] - all of which deplete the available resources, disrupt the routines and established social ties, threaten the cognitive structures, and the eventual sustainability of the group. The second view sees the positive effect: helping screen out under-performing volunteers [KP85]. The third adopts a more neutral view that suggests that new volunteers with new skills and knowledge replace those who leave, maintaining the critical mass, and this may be optimal for long-term performance [DS01]. Hausknecht and Holwerda [HH13] resolve the above perspectives by arguing that the specific details concerning who is being retained is more important than traditional, aggregated measures of volunteer dropouts such as turnover rates as the latter hide variation in the key causal factors that predict retention and performance. [HT11]. For instance, the loss of a productive manager may be more damaging than the loss of an under-performing employee. Also, depending on these details, the same level of retention could have different consequences [HT11, YWL+17]. Therefore, it is important to study factors that predict retention and performance rather than focusing on aggregate measures of turnover.

Accordingly, factors associated with newcomer churn and retention have been studied in a variety of contexts such as Wikis [PHT09], Newsgroups [JK06, ABJ+06] Q&A sites [DPRS12, YWAA10, PAT14], Multiplayer Online Role Playing Games (MMORPGs) [BS12, KPS09], and social networks [BML09]. Below, we review relevant work in each of

these domains.

2.2.2 Open Source Software (OSS) Projects

OSS Projects mostly rely on developers who volunteer their time on these projects. Therefore, it is important to motivate, engage, and sustain new developers to maintain the critical mass needed for building a project [QF11]. Several researchers investigated factors related to new volunteer churn in OSS Projects. Fogel observed that if a project does not make a good first impression, newcomers do not return for a long time [Fog05]. Steinmacher et al. suggest that expectation breakdowns, reception problems, setup misconfiguration, and learning curves may all impact the overall joining process, especially before making their first contribution [SGR14]. For instance, newcomers try to learn about the various social and technical aspects of the project and request specific help by posting questions in project forums and mailing lists or by sending emails to project coordinators and core developers [PJ09, VKSL03] as they begin contributing to the project. At this stage, receiving unpolished answers, or replies that do not offer guidance, or a lack of guidance in general can result in newcomer dropouts [SWG12, SWCG13]. Steinmacher et al. also performed a systematic literature review of of studies related to barriers faced by newcomers to OSS projects and identified 15 different barriers grouped into five categories - social interaction, newcomers' previous knowledge, finding a way to start, documentation, and technical hurdles. They found that barriers related to socialization were most common [SSGR15].

2.2.3 Newsgroups

Joyce and Kraut investigated newcomers' retention across six Usenet newsgroups – (`netscape.public.mozilla.ui`), (`groups-alt.support.diet`, `alt.support.cancer.breast`, `alt.politics.usa.constitution.gun-rights`,

`alt.sports.hockey.nhl.ny-rangers`, and `alt.baldspot`) – and found that newcomers' initial post properties, reply properties, and the probability of posting again are related [JK06] to each other. Arguello et al. conducted a similar study with eight Usenet newsgroups and found that newcomers and oldtimers differ in their ability to get replies and in the ways they write messages [ABJ+06]. Lampe and Johnston explored newcomer behavior in a social news website called Slashdot[4] and found that the ratings received and the way a newcomer's post is moderated affects their probability of returning [LJ05].

2.2.4 Wikis

A decrease in the number of active contributors was observed even on Wikipedia. For instance, a longitudinal analysis of cohorts of 'new' editors (based on a 10-edit milestone) showed that in a certain year, the fraction of those who made at least one edit has decreased from about 40% in 2004 cohorts, to about 10% in the case of 2009 cohorts [Con11]. Being a complex socio-technical system, a number of factors may contribute to the decline in newcomer retention although the most dominant hypotheses attribute this decline to the desire to maintain high quality standards and to fight vandalism [PCL+07], the daunting body of norms and policies the new editors have to go through [BKM08], and sometimes unpleasant social exchanges, despite producing the same rate of good faith contributions as editors from previous years [HGMR13]. Others such as Morgan et al. showed that user-friendly tools, safe spaces and sandboxes for newcomers, and promoting positive interactions between the newcomers and established members can improve newcomer retention [MBWS13] in Wikipedia.

[4] http://slashdot.org

2.2.5 Q&A sites

Most Q&A sites also suffer from new user churn. On three different Q&A sites across various languages and cultures – Yahoo! Answers in English, Baidu Knows in Chinese, and Naver Knowledge-IN in Korean, Yang et al. [YWAA10] looked at length of the first question posted by users to predict longevity. Including a variety of features such as personal information, rate of activity during the first week, and social interactions with other users, Dror et al. trained several classifiers and built a classification model that can successfully discriminate returning users from non-returning users in Yahoo! Answers [DPRS12]. In StackOverflow, Pudipeddi et al. studied the new user churn prediction problem using temporal features, features based on gratitude, quality, consistency, frequency, speed, content, competitiveness, and knowledge level [PAT14] and found that temporal gaps between subsequent posts is the most significant predictor of all factors and decision trees modeled these factors with the best predictive power.

2.2.6 Summary of the above work

The above body work on new volunteer retention in peer-production communities shows that (i) properties of initial activity (such as length of first post, rate of activity, amount of interaction with other users) are predictive of newcomer retention [DPRS12, JK06, YWAA10] (ii) interaction barriers (such as technical hurdles, lack of documentation, or excessive norms and policies, or unresponsiveness from existing members) can contribute to churn of new volunteers [BKM08, SWG12, SWCG13, SGR14, SSGR15] and (iii) unpleasant social exchanges with existing members and critical ratings and moderation of newcomers' initial posts can contribute to new volunteer churn [].

2.3 Limitations of the above work

Much of prior work attempted to build models for retention and productivity by measuring factors for a few weeks or months (e.g.,[DPRS12, PCK12]). This can be a problem since waiting for a few weeks or months could mean losing a lot of users as much of the churn happens at the end of first session. While some of prior work found that newcomers' initial post properties or response to initial post properties are related to longevity [JK06, LJ05, YWAA10], the challenge with such metrics is that they cannot be generalized to understand churn in other communities with entirely different features. To address these challenges, in this thesis, we propose metrics based on just the first-session activity and those that can generalize to other communities with varying activity structures and properties and build regression models for predicting newcomer retention (see Chapter 3).

Prior work attributes the decline in retention of new editors to a daunting body of norms and policies, rigid structures, and interaction barriers [BKM08, PCL+07]. However, there are plenty of communities where such norms do not exist. We, therefore, wondered what retention would be like in a community that doesn't have such daunting norms, policies, and structures and studied if other community-level characteristics are associated with newcomer retention (see Chapter 5).

Also, much of prior work focuses on newcomer activity within the community they join, the norms and structure of the community and the interactions within the community. There is little work that examines characteristics of the newcomers themselves that determine their retention and activity level with these communities. Therefore, in this work, we model newcomers' psychological traits such as personality (see Chapter 4) and newcomers' profile characteristics such as prior experience in other communities to study their associations with newcomer retention and productivity in the community they join (see Chapter 6).

Chapter 3

Effects of Early Newcomer Activity

3.1 Introduction

In the previous chapter, we have noted that 60% of user churn happens right after the first session. This shows that if community moderators do not intervene at the end of the first session, they have already lost 60% of their new users. Building models that can predict user churn based on all available factors at the end of the first session[1] would be highly useful because moderators might want to either a) invest more energy on those who are likely to return (e.g., by relevant offers, greetings) or attempt to "recapture" the interest of those who are not (e.g., by emails).

However, an inherent problem with predicting new user churn is that not much previous activity history is available, and often, demographic information is incomplete. Yang et al. [YWAA10] looked at length of the first question posted by users to predict longevity. But such metrics fail to generalize to communities with other participation types without such attributes as content length.

Some of prior work (mentioned in the previous chapter) used metrics based on activity history for a few sessions, weeks or months, or semantic attributes that capture general mood or immersion of the user across sessions until the point of analysis to predict user longevity.

[1] A version of this work was published as: Karumur, R.P., Nguyen, T.T. and Konstan, J.A., 2016, February. Early activity diversity: Assessing newcomer retention from first-session activity. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (pp. 595-608). ACM. [KNK16]

Since not much previous activity history is available about a new user, these are factors we cannot assess very well at the first session. Waiting for a few weeks of activity to do the analysis would mean running the risk of losing a vast majority of users until the point of analysis.

Other work used demographic information about the user, but this may be either too sparse or not always available as most communities these days have minimal registration barriers with a one step signup process using their Gmail or Facebook accounts. For instance, in MovieLens, age and gender are only available for about 1% of the users. Therefore, they cannot be used in any useful way.

Some of prior work also used overall time spent on site as a predictor. Most users multitask and so, the exact time a user spends on the community of interest is hard to estimate. Also, users switch to other browser tabs or windows; or close browsers or tabs without ever logging out. Therefore, metrics based on time are often inaccurate representations of amount of user activity (despite sometimes showing moderate correlations with it) and therefore cannot be relied upon.

Some work also used social influence of other users to predict a particular user's retention in the community. Again, for new users who are not necessarily well-networked, or for new users in communities which do not have an active social component, metrics based on social influence are not suitable.

Desires to volunteer online, help others, gain reputation, pursue shared values and beliefs, voice humanitarian concerns, develop careers, develop positive attitude and protect oneself from negative feelings; having previous experience; or just enjoying what the community does have all been identified to be factors that motivate contribution (and thus retention) in online content communities [BML10, FDFM+12, KWL12, Nov07, SG06, WF00] . Site policy changes, personal life changes, or a sense of feeling that they can no longer fulfill

their perceived role in the community on the other hand may lead to user churn [VWLB14].

Much of prior work had extensive user data available about these attributes for their analyses because they have analyzed user churn in general. But, in this work, we focus on the specific challenge of new users for whom we have very little to no data about any of these attributes. With de-identified log information, we could not contact individual users who stayed or left the system either. Therefore, we do not have information regarding motivations and prior experience of individual users.

An obvious and easily visible metric that could be used for predicting retention at the end of the first session is the amount of activity by the users during their first session. We call this *early activity*. As we were wondering if there is something else that we can measure that is non-obvious, we looked at some of prior research that shows that newcomers are happier and stay longer if they have a complete picture of the community while joining [BBE+07]. During their early interaction with the community, they investigate and evaluate it on a variety of dimensions to see if it fits their needs. They decide whether to invest effort in it or move on to explore other alternatives. If they find it suitable, they join and remain in it longer [CUC+05, Kri96, KBZJ05, LM94, PNA04]. While some online communities provide access to their archived content without the need to join, others require that the users login to see what it has to offer. In either case (particularly, the latter), it is evident that the breadth of their exploration of the community's features during their first login session can affect whether they leave for good or return for a second session. We therefore wanted to test this hypothesis. We consulted Prof. Mark Snyder, a psychologist who specializes in offline volunteering and learned that there are examples of this in offline volunteering communities such as RedCross [2]. If all a volunteer for RedCross ever gets to do is giving out cookies, if they get tired of that activity, they are likely to leave RedCross. On the other hand, if they are

[2] redcross.org

sometimes giving out cookies, sometimes checking people in for blood drives, sometimes distributing blankets and canned food, and sometimes helping the hurricane victims, if they get tired of one kind of activity, there is something else for them to fall back on that can keep them returning to RedCross for volunteering. We therefore wanted to study a measure of diversity based on the breadth of activities tried by the new user in their first session and see if it was predictive of their longevity.

In order to do this, we introduce a metric called DSCORE to characterize this early (first-session) activity diversity. We had three specific goals that led us to develop a new metric instead of using one of the popular [Col09, Gin12, Jos06, Sha01, Sim49] diversity metrics. First, based on our grounding in the "complete picture", we wanted a diversity metric that focused on exposure and not quantity – the metric should ignore repetitions of an activity and consider only the breadth of activities a user tries. Second, we wanted to measure diversity in a manner that recognizes that different activities may be more or less similar, awarding higher diversity scores to sets of dissimilar activities. Third, we wanted a metric that would generalize to communities with different activity structures, and that would in turn scale to different numbers of activities. The metric we introduce has these desirable characteristics and is based on a distance tree analysis of the online site's activities.

Thus, using DSCORE, in this work, we explore the utility of naturally-occurring early (first-session) activity diversity in assessing new user retention in an online recommender community 'MovieLens 'together with another measure called ASCORE for the amount of activity.

3.1.1 Research Questions

We organize our research around the following questions:

RQ1: *How is early activity diversity (measured using DSCORE) associated with new user longevity?*

We first establish feasibility by showing a correlation between number of distinct activity types tried and new user retention. We then build and test successive models to explore the degree to which early activity diversity is associated with new user longevity, considering a variety of model types and additional factors such as overall quantity of activity. We use the model with the best fit to illustrate the increase in average longevity associated with marginal increases in a new user's first-session activity level and diversity.

RQ2: *How can we most effectively measure early activity diversity for purposes of predicting new user longevity?*

Once we have established the value of DSCORE as a predictor of user retention, it makes sense to examine how it compares with more traditional metrics. We therefore compare the models built using DSCORE with the ones built using the Gini-Simpson index.

3.1.2 The MovieLens Dataset

We conduct this research using log data from the classic version of MovieLens [3] from December 20, 2007 to January 1, 2014. MovieLens allows users to rate and receive recommendations for movies. In addition, they can add, edit or tag movies, add buddies, or participate in other ways such as by answering questions about movies. The presence of multiple activity types and a large pool of users along with their activity logs from their very first interaction with the community (48,784 users in total) made this dataset useful for our research.

[3] http://classic.movielens.org

3.1.3 Contributions

We make the following contributions in this work:

- **Early activity diversity predicts retention of new users**. Based on an analysis of the usage logs of more than 48,000 users of MovieLens, we find that activity diversity in the very first session is a significant predictor of new user retention.We also show that diversity adds significant value when combined with measures of activity level, with both measures helping predict new user retention over 1, 5, and 10 sessions.
- **DSCORE: A new and more effective diversity metric.** We introduce DSCORE, a metric to measure early activity diversity in a general way based on a similarity tree classifying activity types. It is designed to isolate diversity from quantity of activity and can be applied to different sites that support multiple activity types. Also, we find in the context of MovieLens that DSCORE is more useful than traditional measures capturing diversity such as the Gini-Simpson index.
- **Implications for design and research.** We discuss implications both for designers and for researchers. Diversity can be used to customize experience based on predicted retention, or to assess and improve site design for engagement. Further research is proposed to isolate causal factors underlying the relationship between early activity diversity and retention.

3.2 Related Work

3.2.1 Early Activity as a Predictor of Longer-Term Behavior

Even outside questions of churn, people have found value in early activity as a predictor of longer-term behavior. It was observed in an analysis of "power users" of Wikipedia that

users' activity patterns, even in the earliest days, had an ability to predict future amount, quality and frequency of activity [PHT09]. Also, Pal et al. [PCK12] looked at the first few weeks of activity to detect experts in a community. Burke et al. [BML09] found that newcomers' exposure to different features on Facebook through the newsfeeds of their friends' activities moderately affects (positively) their future usage of those features. These works strengthen our interest in studying the relationship between measures based on early activity and future retention.

3.2.2 Interventions

Prior work showed also that responding to a user's first interaction, eliciting feedback from them through lightweight tools, providing assistance and recommendations early on and properly welcoming them into the community improve user retention [BML09, CAKL10, CT15, FJGG09, JK06]. Also, commercial practice suggests that there is an interest in interventions aimed at new user retention. A lot of sites offer additional gifts, e-coupons, membership discounts, special promotions, free premium account access for extended periods, etc., in order to retain users who do not return for long durations of time. We therefore hypothesize user retention may improve when introduced to other types of participation.

3.2.3 Diversity in Online communities

Zhu et al. found that greater diversity in subgroup membership was associated with greater longevity of Wikia members [ZKK12a]. However, specialization in participation type is most commonly found in online communities. Categories like 'lurker', 'Questioner', 'Answer Person', 'Uploader' and 'Contributor' have been identified based on specialization [MEM+12, NP00, TSFW05, WGFS07]. But these works did not look at the question of whether those who chose to specialize did so after being aware of the wide range of pos-

sibilities. Our measure – DSCORE is specifically designed not to penalize people who specialize after being aware of the alternatives. We hypothesize that someone who tires of their specialized activity will be more likely to be retained if they know there are other things they can fall back on.

3.2.4 Diversity and Community Success

In order to accomplish goals that are important to the community, some attempts have also been made to direct users to other opportunities even if they did not match their interest in the context of Wikiprojects. Examining weekly collaborations, Zhu et al. established that explicit setting of goals and implicit social modeling can help diversify a self-identified user's participation in such a way that tasks important to the community may be accomplished [ZKK12b]. So, we understand that diversity is a characteristic that can be nurtured in users, if we find value in it.

3.2.5 Diversity Metrics

Diversity metrics quantify distribution of entities across various available class types and have been studied extensively in biology, ecology and in social and informational sciences. Many diversity metrics have been proposed based on the need of the community under consideration. Richness [Col09], Shannon Entropy [Sha01], Simpson index [Sim49], Gini-Simpson index [Gin12, Jos06, Sim49] are the most widely used ones. Definitions of diversity have varied widely based on what the proponents of those metrics assumed diversity to be. Diversity metrics in general deal with a richness component - characterizing the number of distinct class types the set of interest contains and an abundance component characterizing number of entities per class type - sometimes using both components, and sometimes just one of them.

Richness does not account for class hierarchies or similarities between entity (activity) types. Other metrics such as Entropy, the Simpson Index or the Gini-Simpson index have quantity of activity included in them. Because we are interested in the marginal value of diversity over quantity, we introduce a diversity metric that separates diversity from quantity of activity. To validate the usefulness of our new metric over existing ones, we redo our analyses replacing DSCORE with the Gini-Simpson index (which is popular in social psychology literature).

3.3 Early Activity and Early Activity Diversity

3.3.1 Identifying Activities

An activity is a single interaction with any feature of a particular online community and an ‘activity type‘ refers to one of the several types of activities that exist in the community. For instance, on MovieLens, a user could rate a movie 3.5 stars, or search using the tag “Animation“, or use the “My Wishlist“ feature. Each of these is a different activity type but the user has performed three activities in all. We occasionally use the term ’participation’ to refer to an activity and the phrase ’participation type’ to refer to an activity type.

Based on consultation with the site maintainers and other site experts (more than 20 researchers – faculty, former and current students involved in development of and research on MovieLens), we classified the features of MovieLens into 17 distinct activity types. A brief description of each activity type is shown in Figure 3.1.

3.3.2 Activity Metrics

To answer the question of how early activity diversity is associated with user longevity in the community, we need to separate diversity from quantity. So we introduce two metrics:

1) *Edit profile* – The descriptor indicates that a user edited his profile or visited the edit profile page to make changes to his profile to represent himself to the community.

2) *Create an RSS feed* – User uses this feature to create an RSS feed for herself.

3) *Invite a buddy* – User uses this feature to invite a buddy to Movielens.

4) *Use help pages* – User uses this feature to learn more about and understand different features of the system.

5) *View movie detail* – User uses this feature for viewing complete details about a specific movie such as actors, directors, genres, language, ratings, a brief description of the storyline and tags applied to the movie.

6) *Search by tag* – User uses the feature to browse for a movie using a list of displayed tags.

7) *Search attribute / metadata* – User uses this feature to search for a movie by entering a search phrase or word. Feature 6 is different in that it does not let the user enter anything. The user can only click on existing tags to search for lists of movies.

8) *View "Most Often Rated" list* – User uses this feature to view the most often rated movies on Movielens.

9) *View "Top Picks for you" list* – User uses this feature to see a personalized list of movies recommended to him by Movielens.

10) *View "Newest Additions" list* – User uses the feature to see the list of new movies added to Movielens.

11) *View "Rate Random Movies" list* - User uses this feature to browse through a list of random movies to rate them.

12) *View "Your Wishlist"* – User uses this feature to see all the movies he has added to his wish list.

13) *View "Your Ratings" list* – User uses this feature to see all movies she has rated and their corresponding ratings.

14) *Rate* – User rates a movie on Movielens.

15) *Tag* – User tags a movie on Movielens.

16) *Participate in Q&A* – User uses the feature to participate in a Q&A discussion.

17) *Add / Edit movie* – User uses this feature to add a movie to Movielens or edit a movie on Movielens. The classic version of Movielens was structured such that the same control was used for both purposes.

Figure 3.1: A chart showing a brief description of each activity type on MovieLens

DSCORE and ASCORE.

Early Activity Diversity Score (DSCORE) Early Activity Diversity Score for user u is a metric characterizing the number and degree of dissimilarity of distinct activity types performed by the user u in the *first* session based on the hierarchical ontological relatedness of these activity types. We denote it by DSCORE.

Design Challenge: The available activity types in an online system range from highly related (e.g., rate an item, tag an item) to fairly distant (e.g., invite a buddy, view a movie).

While each is different, we want a measure of diversity that adequately reflects that carrying out three very different activities may have more diversity than carrying out four or five very similar ones. Intuitively, this is the same as we might find with biological diversity: a zoo with five different types of primate does not have as diverse a collection as one with a chimpanzee, a whale, and a lizard.

Our approach to this challenge is to build our diversity metric in a manner that is tied to a hierarchical taxonomy of activities – a taxonomy that is built specifically to group similar activities together and to separate dissimilar activities. Unlike many of the diversity metrics discussed in the related work section, this approach allows us to take hierarchical ontological relationship between entity types into account. We build upon simple 'richness' accommodating various degrees of dissimilarity between different activity types. So we first model the relationship between various activity types in a community.

Modeling relationship between activity types: One could interview the users of the community, analytically look at which participation types go together, or speak to experts or community moderators to understand the degrees of dissimilarity between various activity types. In our case, we engaged the experts in a card sorting activity [Spe09], a standard usability technique used to understand the information architecture of a site. We asked them to cluster the activity types into as many natural clusters as would make sense to them and provide a brief explanation of why they believed in such architecture. The experts were also asked if they would further cluster them into smaller or larger clusters and some of them did. Clusters that emerged in the process are shown in Figure 3.2.

In the end, we had a tree that depicted the relationship between various activity types on MovieLens (Figure 3.3).

In this tree, all distinct activity types appear as leaf nodes. Each internal node of such a tree represents a hypothetical activity type that encompasses all activity types represented

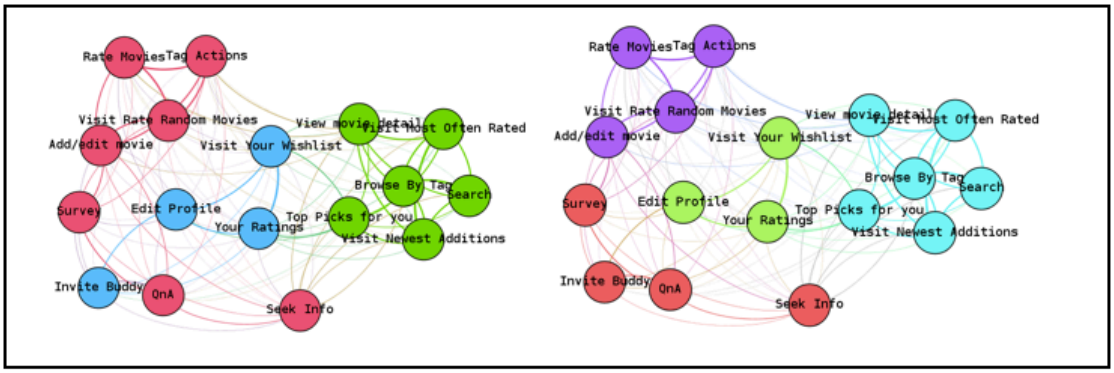

Figure 3.2: Intermediate Clusters that we obtained after plotting our data

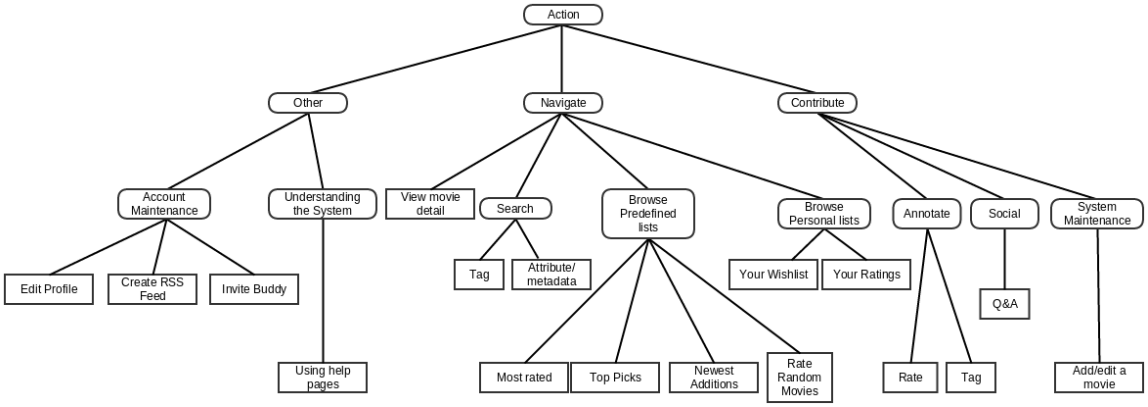

Figure 3.3: Classification of activity types in MovieLens

as its child nodes. We call this hypothetical node an ancestor. There can be multiple levels of ancestors with multiple activity types sharing the same ancestor. Each node in the tree represented in Figure 3.3 is labeled, but one may choose to not label the hypothetical nodes. To make a distinction we depicted all the ancestors with rounded rectangles and the leaf nodes with rectangles. We now use an approach similar to that used in phylogenetics [WSSP03] to study evolutionary relationships between various species of organisms. Note that such a tree need not be a binary tree.

$$\begin{array}{c} \\ p \\ q \\ r \\ s \end{array} \begin{array}{c} \begin{array}{cccc} p & q & r & s \end{array} \\ \begin{pmatrix} 0 & 2 & 3 & 4 \\ 2 & 0 & 3 & 4 \\ 3 & 3 & 0 & 3 \\ 4 & 4 & 3 & 0 \end{pmatrix} \end{array}$$

Figure 3.4: Classification tree and its corresponding distance matrix D. The distance d_{qr} is 3 meaning that there are 3 edges in the path connecting the leaf nodes q and r

We use a distance matrix D to quantify the amount of dissimilarity between any two leaf nodes in the tree. The amount of dissimilarity between two leaf nodes is simply the number of edges in the shortest path connecting them. Let d_{ij} denote the dissimilarity between leaf nodes i and j in the tree. d_{ij} also denotes the ij-th element of the matrix D.

Definition: We define early activity diversity score of a set of distinct activity types as the normalized mean value of pair-wise dissimilarity (defined above) between all activity types in the set. More formally, if n is the number of distinct activity types represented in a set S of activity types, then the early activity diversity score of the set S is given by

$$DSCORE = Div(S) = \frac{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} d_{ij}}{n - 1}$$

The proposed metric *Div(S)* has the following properties:

1. *Div(S)* is zero if a user performs activities of only one type.

2. When all leaf nodes (or activity types) have one and only one ancestor, then *Div(S)* is simply what is called ”richness” in biodiversity and ecology literature.

3. *Div(S)* increases as ancestral connection increases. In the figure 2, the set p, s is more diverse than the set *{p, r}* which in turn is more diverse than the set *{p, q}*. In other words, diversity of two leaf nodes whose parent is same is less than diversity of two leaf nodes whose parent is different.

4. As additional distinct leaf nodes are added to a set of activities, Div(S) increases. For example, Div(*p, q, r*) will always be greater than Div(*{p, q}*), for $p \neq q \neq r$.

5. For sets of the same length *Div(S)* attains a maximum value when no two leaf nodes of the set have the same parent and a minimum value when all leaf nodes of the set have the same parent.

Note that we are not interested in proportional abundance of a given leaf node (or activity type), because all we care about is whether the user got an opportunity to use the feature at least once. Using this, we are interesting in predicting user churn, so we formulated the definition such that *{p, q}*, *{p, p, q}*, *{p, p, p, q}* and *{p, p, q, q}* are all equally diverse.

The theoretical maximum for DSCORE for our tree is 41.6.

Early Activity Score (ASCORE) Early Activity Score is defined for user *u* as the quantity of activity performed by user *u* in the *first* session. We denote it by ASCORE.

Design Challenge: In online communities, users engage in different types of activities for different periods of time. So, a simple count of all activities may turn out to be an inaccurate representation of the amount of activity performed by the user for a session as it diminishes the impact of the time- and engagement-intensive multi-step editing and adding activities. We considered two ways to adjust for this imbalance: weighing activities (i) by infrequency of use (so rare activities count more) or (ii) weighing them by time spent on the activity (so more interactive/intensive activities count more). We choose the latter as a better measure of activity that is not connected to diversity (which is related to rarity / dissimilarity / spread). Thus, we define weight of a unit of each activity type factoring in the time duration associated with that activity type. In the related work section, we have stated how time durations can often be inaccurate because users may sporadically drift to other websites for indefinite periods of time during their course of interaction with the community. Therefore, in activity-times data, one might expect to find outliers for certain user activities.

In order to eliminate bias due to such time periods, we pick the median time duration of all users for each activity type as the weight for that activity type for all users.

Definition: If w_i denotes the median time duration for activity i for all users and the user u performed n_i activities of that type in the first session, then the early activity score for the user u is given by

ASCORE = $\sum_i w_i n_i$

3.4 Methodology

3.4.1 Dividing Activity Log into Sessions

Ideally, an activity session is defined to be the time between a user's login and logout. However, for not all users we have the information about the login and logout events. For those users for whom we do have this information, we accurately determine a login session. However, for those users whose login and / or logout events are missing for whatever reason (they stay logged in for an indefinite period or quit the browser or close the tab without logging out, etc) we use the definition of session based on log-scaled inter-activity times [GH13]. For such users, a session is identified to be a set of continuous activities by a user in which any two subsequent activities are within a time difference of 1 hour.

3.4.2 Representing User's Defection from the community

For analyzing the expected active period of a MovieLens user, we model the user's lifetime by defining concepts of "birth" and "defection" in (from) the community as the times at which the user starts his activity and stops his activity for a considerable period of time respectively. For MovieLens, we determine the threshold of inactivity to be 365 days based on activity logs which show a bi-normal distribution with the second normal at about 300

days after the first registration. So if a user does not have an activity for 365 days since they last visited, we consider the user to have dropped out of the community.

3.4.3 Ignoring activities beyond the 365 day inactive period

Based on the above threshold of inactivity, if a user is found to have an activity after 365 days, we have every reason to believe his/her movie-seeking behavior might have changed over the course of this time. So, we consider the activity thereon under a new life instance of the same user. We found a small fraction of users (2,136) with more than one life instance in our dataset. But we have carried out our analysis on 48,784 distinct users for whom only the activities of the first life instance were considered ignoring the activities beyond the 365 day inactive period.

3.4.4 Handling right-censored users

Note that we have ended our data collection on a certain date and therefore we do not have information about some users whether or not they return to the system after 365 days. This concept is identified in survival analyses literature as right-censoring and in these analyses these users are marked as right-censored users. We have 8157 users who are right-censored. We model user churn using two approaches - Survival analysis and Logistic Regression. For modeling using survival analysis, we will use appropriate survival models (Cox-Regression) to handle the right-censored users. Because logistic regression is used for making prediction/binary classification and it does not handle right-censored users, we will ignore those data points in the logistic models.

3.4.5 Computing ASCORE weights

Recall that our ASCORE metric requires a time-measure as a weight for each activity. Based on the timestamps in the log data, we compute the weight as the median time between the start of an activity and the start of the next activity (omitting the final activity in each session). Table 3.1 lists the median time duration in seconds for which users of MovieLens spent time on an activity before moving on to the next one.

Activity Type	**Weight (median time duration in seconds)**
Edit Profile	12
Create RSS Feed	13
Invite a buddy	14
Using help pages	20
View Movie Detail	10
Search by tag	16
Search using attribute/keyword	13
Visit Most Often Rated Movies list	16
Visit Top Picks list	14
Visit Newest Additions list	17
Visit Rate-Random-Movies list	10
Visit "Your Wishlist"	19
Visit "Your Ratings" list	17
Rate a movie	7
Add a tag	9
Q&A	9
Add/edit a movie	13

Table 3.1: Median time duration in seconds spent by MovieLens users for each activity type.

3.4.6 Choosing predictors for the model

The data we have access to has extremely sparse age and gender information with practically no other personal information available. Nor do we have any information about the motiva-

tions or pro-social behavioral history about the users. So, we are unable to use any of these as predictors in our model. We do not use length of first session as a predictor firstly because we believe time durations are inaccurate representations of user activity due to general user drifting behavior and second because we infer activity sessions from activity log data of the user, which does not always contain login and logout. We could choose metrics specific to MovieLens (that may not be generalizable to other systems) such as the number of movies rated by user in the first session (because MovieLens is primarily a movie-recommender website powered by user ratings). However, we found that the number of movies rated has high correlation with amount of activity in the first session and so would not really add much to explaining the model. So, we decided to use amount of activity (which is only about 0.4 correlated with activity diversity) along with activity diversity in our model.

3.5 Results and Discussion

3.5.1 Structure of this section

We present and discuss results in three sections. First, we explore the MovieLens log data to see the distribution of activity diversity and the prevalence of new user churn. Then we try to see how user churn is associated with early activity diversity using varying measures of longevity and different approaches to modeling user churn to establish the robustness of our results. Finally, we validate the usefulness of our DSCORE metric by comparing it with the Gini-Simpson Index in the best-fitting model.

3.5.2 Frequency of activity types on MovieLens

On MovieLens, we find that 37.74% of activity types for all recorded sessions for all users in the data constitute ‘rating movies’ (most often performed) followed by 30.76% of activity

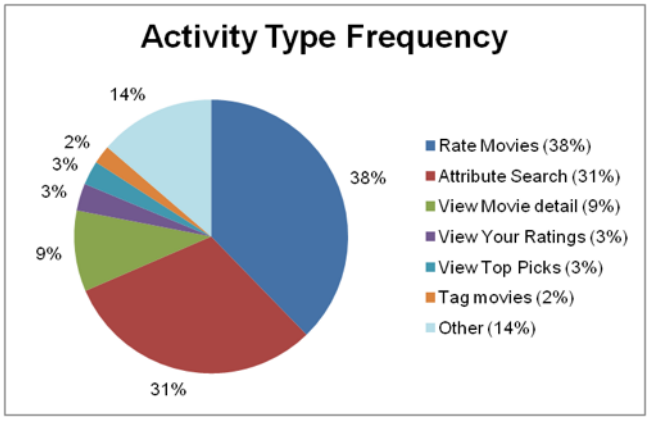

Figure 3.5: Frequency of activity types on MovieLens

types constituting 'search' using attribute/keyword. This is followed by visits to movie detail page constituting about 9.56% of total actions, followed by viewing one's own rated movies, viewing the top picks list and tagging movies accounting for another 8% of activities (See Figure 3.5). The remaining 11 activity types count to only about 14% of the activities on MovieLens. Thus we see that most users are highly specialized in the ways they participate in MovieLens although they have about 17 different activity types to engage in.

3.5.3 Evidence of early user churn on MovieLens

Large numbers of users drop out in their first few sessions (see Figure 3.6, a plot based on our dataset) and particularly significant drop occurs right after the first session. So, we will use ASCORE and DSCORE of a user at the first session.

3.5.4 Relationship between percentage user churn and simple number of activity types tried in the first session

We define percentage user churn after the n-th session to be the number of users who dropped out of the community after the n-th session over the total number of users who

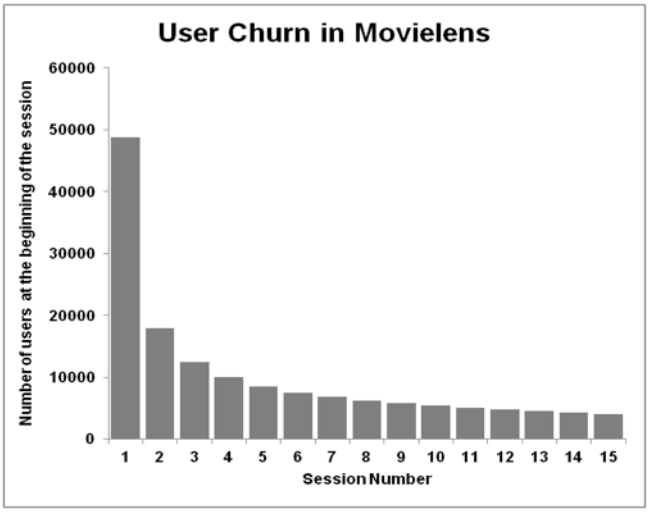

Figure 3.6: User Churn in MovieLens

tried k activity types in the first session where k = 1…15 (although participation in all 17 activity types is theoretically possible, the users in our dataset have participated in at most 15 activity types by the end of the first session) and n = 1,5,10 (we report only for these sessions in Table 3.2. Ignoring the users who are right-censored, we find that the lower the number of activity types tried in the first session, the greater the percentage of user churn. The results are available in Figure 3.7 and its corresponding Table 3.2.

RQ1: *How is early activity diversity (measured using DSCORE) associated with new user longevity?*

Earlier in this paper, we have identified two metrics based on activity: activity score (denoted by ASCORE) and activity diversity score (denoted by DSCORE). We will use the values of these two metrics for the first session of the new user for predicting churn or longevity in the community.

#Activity Types tried in first session	#Users	% User churn after 1st session	% User churn after 5th session	% User churn after 10th session
1	4519	80.7	95.42	97.83
2	4252	79.06	92.48	96.23
3	6508	75.61	91.84	95.71
4	7246	69.74	88.09	93.21
5	6836	63.77	84.54	90.96
6	5798	58.51	81.09	88.13
7	4637	53.35	77.38	86.48
8	3141	49.44	73.95	85.08
9	1842	42.23	68.33	80.97
10	998	35.52	63.73	77.33
11	421	33.44	57.67	71.47
12	219	27.88	56.37	70.3
13	72	25.09	49.06	64.15
14	17	7.69	53.85	84.62
15	1	0	0	0

Table 3.2: Percentage user churn after the first, fifth and tenth sessions.

3.5.5 (Approach 1) Survival analysis using Cox Proportional-Hazards model

In our first approach, we use Survival Analysis using Cox Proportional Hazards because this is ideal in situations where one measures time until an event or hazard (in this case – a user leaving a community) happens with 'Number of Sessions' (continuous measure) as the measure of longevity. Earlier in this paper, we have introduced briefly the concept of right-censoring. Because Cox Regression [Fox02] takes care of right-censored data, we perform survival analysis for all 48,784 users.

We build two models – one consisting only of ASCORE as the predictor and the other having both ASCORE and DSCORE. We find that the difference in log likelihoods of the two models is statistically significant (p-value < 0.0001) with $\chi^2 = 410.6$. Based on likeli-

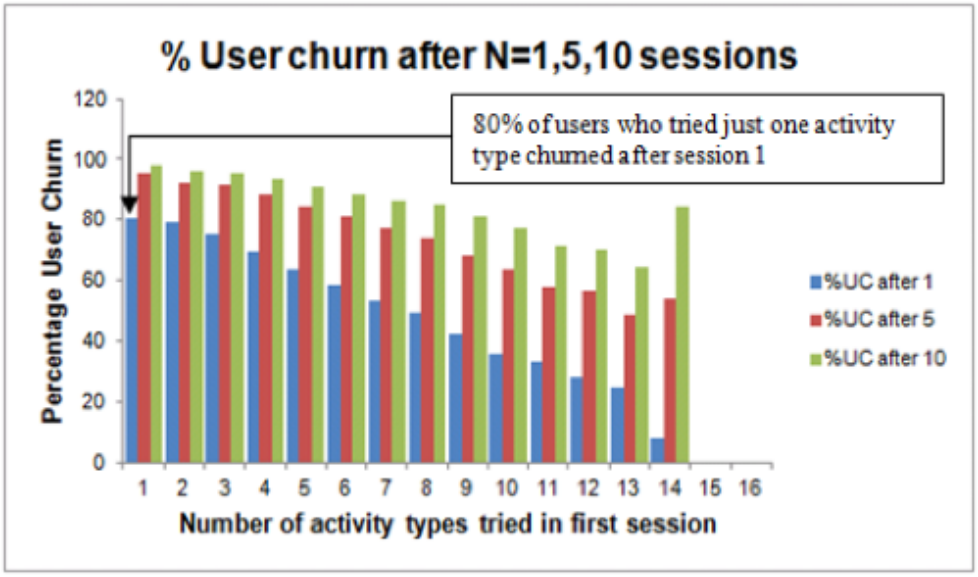

Figure 3.7: User Churn in MovieLens

hood ratio test, this implies that the model that includes DSCORE is better than the model that has only ASCORE.

The corresponding coefficients for the second model are shown in Table 3.3.

	Coef	**Exp(Coef)**
ASCORE	-0.00018***	0.9998
DSCORE	-0.02304***	0.9772

Table 3.3: Coefficients for Cox-Proportional Hazards Model; *** indicates p-value < 0.001.

The values in Table 3.3 indicate that holding the other covariates constant, a unit increase in amount of activity (ASCORE) causes a 0.02% reduction in churn hazard and a unit increase in activity diversity (DSCORE) causes a 2.28% reduction in churn hazard. Given the difference in scales, this is hard to interpret. So we address the quantitative aspect below in our logistic regression model with an illustrative example.

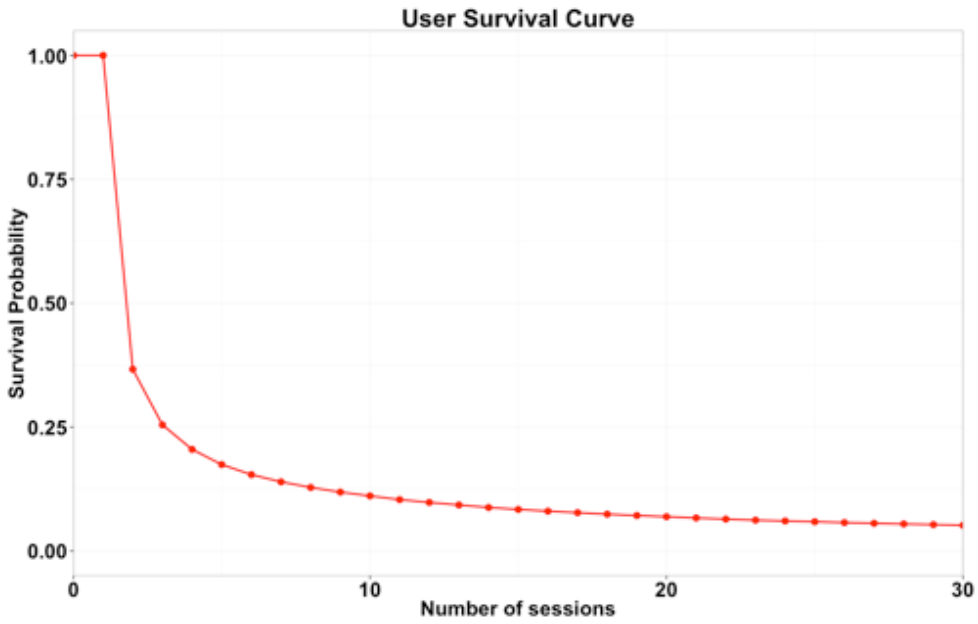

Figure 3.8: User Survival curve plotted to determine a suitable threshold for logistic regression analysis. The graph shows a drop in the probability of survival of users as we proceed from 0 to 30 sessions

3.5.6 (Approach 2) Logistic regression

In our second approach, we use Logistic regression for obtaining a simpler interpretation and a direct estimate of probability of survival past an arbitrary session k. For this we use 'presence beyond session k' (Binary measure) as the measure of longevity. Because logistic regression can be used for prediction, we ignore the users whose survival information is right-censored in our analysis.

Step 1: Longevity measure for logistic regression

The survival curve shown in Figure 3.8 indicates that the probability that the user survives is highest in session one and gradually drops as one moves towards further sessions.

From 3.8, we also see that after using MovieLens for at least 10 sessions (an average

of 2 months), the probability that users continue to use MovieLens is very high. Therefore, we choose N=10 sessions as the measure of longevity to examine how well we can predict if users would stay in the community beyond 10 sessions. (We do a sensitivity check and repeat the analyses with different values of N=1 and 5 sessions, and find consistent results.) So for our logistic regression model, we use a binary response variable with values 1 or 0 indicating the two classes – Class 1 – for 'users who stayed in MovieLens for at least 10 sessions (or 2 months)' and Class 0 – for 'users who stopped using MovieLens after their 10th session'.

Step 2: Analysis

We first build three models – one having only DSCORE, one having only ASCORE and the third having both ASCORE and DSCORE. Table 3.4 shows the corresponding outputs:

	Model 1	**Model 2**	**Model 3**
Intercept	-3.542***	2.602***	-3.352***
ASCORE		0.0003***	0.0002***
DSCORE	0.0939***		0.0604***
AIC	21302	21093	20808

Table 3.4: Summary of the logistic regression models; *** indicates p-value < 0.001.

For a given dataset, AIC (Akaike Information Criterion) measures how one model performs relative to another. The models with smaller AIC have better fit. We find in Table 3.4 that the model having ASCORE alone is better than the one having only DSCORE. However, based on AICs we conclude that the model that includes both ASCORE and DSCORE is better than the individual models. We find also that the likelihood ratio test statistic between the models 2 and 3 has a 2 = 287.04 (p-value ~0) and that between the models 1 and 3 has a 2 = 496.38 (p-value ~0). So again, the model that includes both DSCORE and AS-

CORE is better than the individual models. The results show that both activity and diversity are important, but that retention is more sensitive to smaller changes in diversity.

Using this third model that includes both terms, we note that keeping other terms constant, a unit increase in the amount of activity (ASCORE) produces a 0.03% increase in the odds of survival beyond 10 sessions, while a unit increase in the activity diversity (DSCORE) produces a 6.23% increase in the odds of survival beyond 10 sessions.

We now use the model with the best fit (Model 3 in Table 3.4) to illustrate (see Figure 3.9 the increase in average longevity associated with marginal increases in activity level and diversity. Consider a typical newcomer that we will call Amy with a median ASCORE (288 units) and median DSCORE (12.5 units). A typical profile for such a user would have rating 17 movies, making 5 attribute/keyword searches, using the help feature once, viewing details for 5 movies and using the "Your Ratings" feature twice. Amy's chance of surviving past the 10th session is only 7.31%.

Now let us consider a second user – Ben – who has the same activity pattern as Amy but also performed one additional and fairly different task. Ben invites a buddy to MovieLens. To keep Ben's ASCORE constant, we will also have Ben rate only 15 movies (two fewer than Amy). This changes Ben's DSCORE to 15.4 units while holding his ASCORE at 288 units, but it results in 19.14% higher odds of survival – an increase to 8.6%.

Finally, let us consider a third user – Claire – who starts with Amy's level of activity but we want to increase her ASCORE to the level that would predict the same survival rate as Ben, while holding her DSCORE constant (i.e., by increasing quantity without adding new activity types). Claire would have to increase her ASCORE by 876 units which would involve (for example) 78 additional movie ratings, 10 more attribute/keyword searches, and viewing 20 more movie detail pages. [4]

[4]To come up with this readily interpretable example with substantive meaning, we have used unstandard-

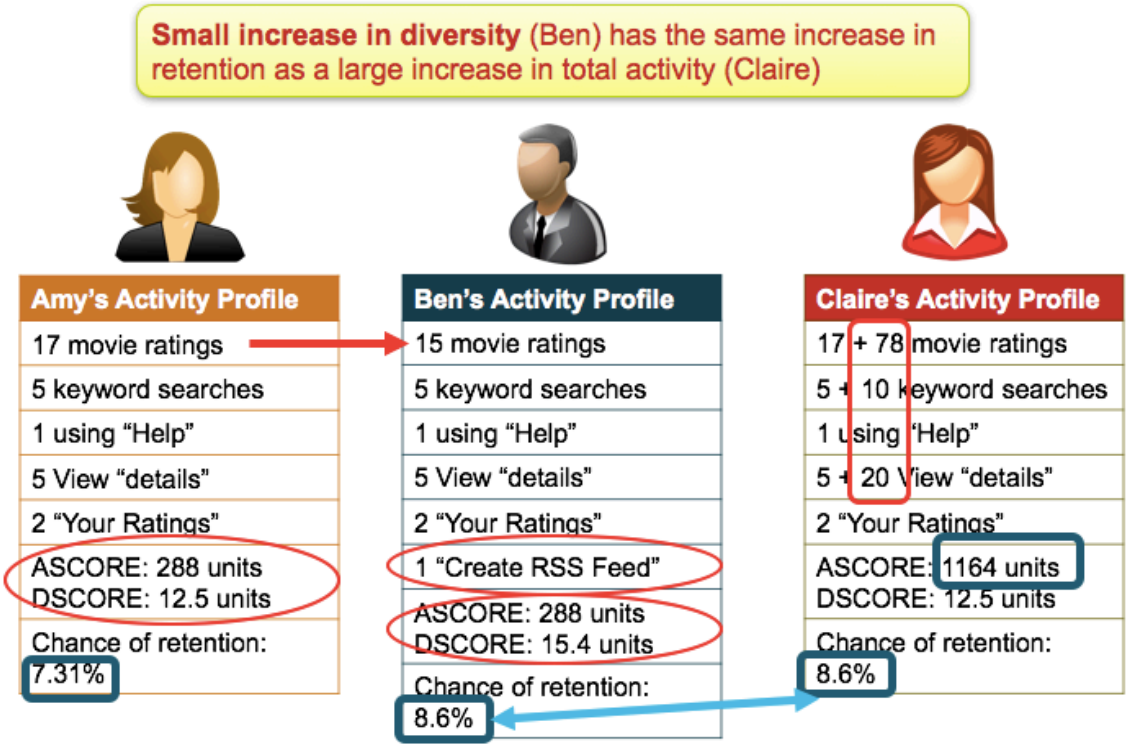

Figure 3.9: An illustration of increase in retention associated with marginal increases in activity level and diversity

In other words, our model shows that performing one activity of a different type is associated with an increase in survival which can only be matched by performing existing activities *many more times each*. We also tested the models at N = 1 and 5 sessions and found consistent results.

Step 3: Prediction accuracy

Because of imbalance in distribution of users in both classes, we do not use precision and F-measure for gauging performance. Instead, we use sensitivity and specificity. Sensitivity,

ized coefficients here. The example shows that about 800 units of unstandardized ASCORE is comparable to roughly 2 units of unstandardized DSCORE in being associated with the same odds of retention.

in our context is the proportion of class 1 users who are correctly classified and specificity is the proportion of class 0 users who are correctly classified.

Because logistic regression gives the probability or log odds that the output belongs to class 1, we need a suitable threshold t to compare the probability obtained using logistic regression p to say if $p > t$ then the user belongs to class 1 else the user belongs to class 0.

We use two approaches to choose an optimal threshold (see Figure 3.10 – the Minimized Difference Threshold (MDT) approach, which minimizes the difference between sensitivity and specificity and the Maximized Sum Threshold (MST), which maximizes the sum of sensitivity and specificity [18]. Note that while these thresholds are not biased towards positives or negatives they do not necessarily give the highest prediction in the model. To make sure our model is not data-dependent, we perform 5-fold cross validation and the average sensitivity and average specificity for the best model were found to be 0.65 using one approach and 0.66 using another.

Step 4: DSCORE: Verifying sensitivity to ontology

To verify DSCORE's sensitivity to the ontology we used, we make slight changes to the existing ontology.

In Figure 3.3, we move Create RSS Feed and Invite Buddy from Account Maintenance to Social and we move all four leaves of Browser Predefined Lists to Search. These changes make the ontology somewhat different but still sensible. We find that the newly computed DSCORE has a correlation of 0.9974 with the old one producing very similar results.

Step 5: DSCORE in the presence of 'Number of activities'

We also include total number of activities into Model 3 and find that it is not significant in predicting survival beyond 10 sessions but is significant in predicting survival beyond 1 and

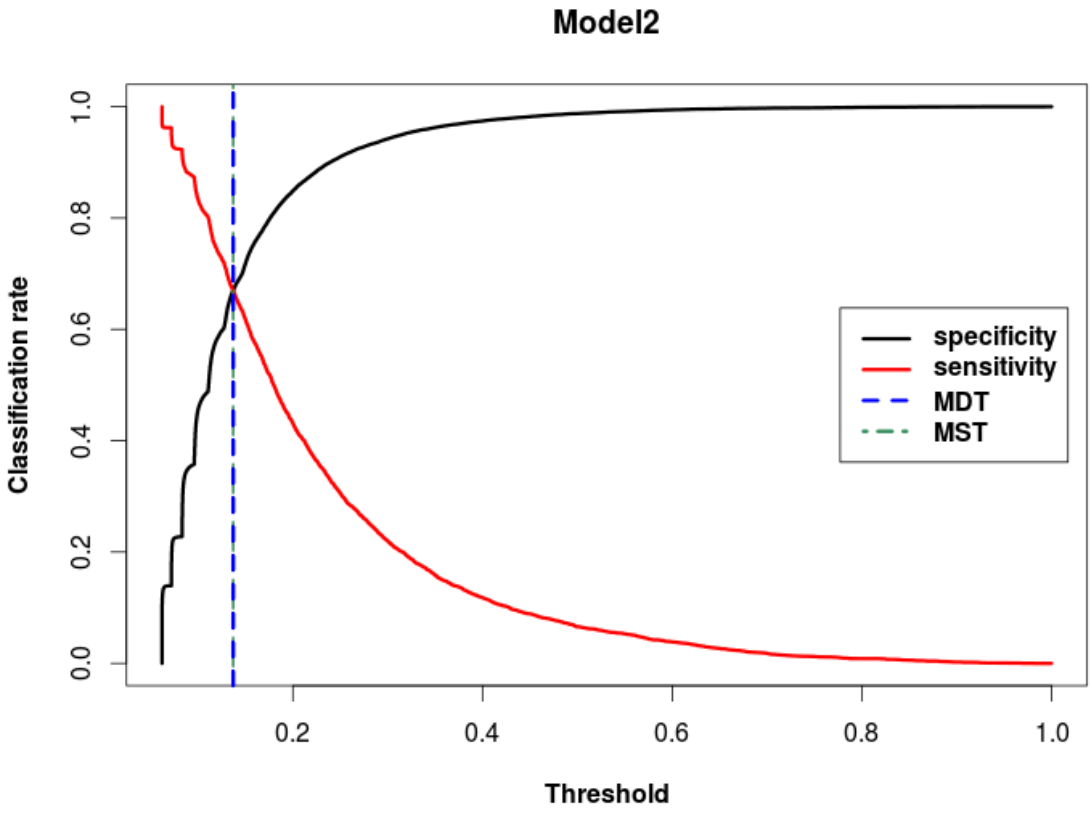

Figure 3.10: An illustration of increase in retention associated with marginal increases in activity level and diversity

5 sessions, but in all three cases (1, 5, and 10) DSCORE is still significant and performs well.

RQ2: *How can we most effectively measure early activity diversity for purposes of predicting new user longevity?*

Because it takes a lot more work to compute DSCORE using a distance tree analysis, we wondered how useful this measure is over a traditional diversity measure such as the Gini-Simpson Index. We therefore replaced DSCORE in our model with the Gini-Simpson and report the results at N = 10 sessions in Table 3.5.

	Coef	Exp(Coef)
(Intercept)	-3.352***	-2.724***
ASCORE	0.0002***	0.0004***
DSCORE	0.0604***	
Gini-Simpson		0.197
AIC	20808	21093

Table 3.5: Coefficients for Cox-Proportional Hazards Model; *** indicates p-value < 0.001.

We find that introducing it does not add significant value over ASCORE in predicting survival beyond N= 10 sessions. We performed a sensitivity analysis by redoing it at N = 1 and 5 sessions as well and we found again that Gini-Simpson is not significant in presence of ASCORE. Also, because some activities are closely related to each other while others may be very different and Gini-Simpson does not account for this as well, we find DSCORE to be more useful in characterizing early activity diversity.

3.6 Conclusion

This work shows a preliminary investigation of new user activity engagement in an online community and is largely intended to describe its effects on new user survival within the community. This work stands out in comparison to previous works addressing the challenge of user retention in that it introduces a novel way of assessing retention using limited amount of information available about new users. We make use of metrics based on activity in the very first session: diversity and amount of activity. We also introduce a way of computing diversity in online communities. The hypotheses were tested on MovieLens, an online community that gives its users an opportunity to participate in a variety of ways – from finding movies they like to rating to tagging movies to answering questions about movies, to inviting buddies and so on. The results of this work are published in [KNK16]

Our findings indicate that 1) the lower the number of activity types tried in the first session, the greater the percentage of users in that category who drop out of the community; 2) early activity diversity measured using DSCORE is a significant predictor of user longevity, and that it remains a significant predictor even in the presence of amount of early activity (ASCORE); and 3) a metric that considers possible similarity between activity types based on a distance-tree is a more useful way of measuring early diversity than traditional metrics. Our results are invariant of measures of longevity and the approaches used to model user churn. We also find that the positive effect of higher early diversity being associated with greater longevity is consistent with prior research [6,24,25].

3.6.1 Limitations and Generalizability

The MovieLens data log limited our ability to assess the relationship of other features to user retention, and of our diversity metric in presence of other features. We are however interested in seeing how this works in other contexts where such data may be available. Nonetheless, we were able to assess user retention and longevity from first-session activity data, which is readily available for all online communities.

We chose MovieLens dataset for our analysis because of its richness in activities offered to its users combined with a long-term longitudinal log of user activity – a log that dates from the user's very first interactions. This work is relevant to community designers, moderators and administrators who wish to understand new user longevity in a variety of contexts: travel sites where diversity of activities (reviews/ ratings) may be with respect to categories: hotels, restaurants and places; Q&A sites and peer-production communities where diversity may be with respect to types of content posted or moderation activities, social networks where diversity may be in types of content shared, on their own profile or others'; product review and retail sites where users may buy/use a variety of products; and so on.

Applying DSCORE to other contexts requires creation of an activity taxonomy. We have not validated its effectiveness in systems with different taxonomies or different types of activity structure, and leave that to future work.

Another interesting scenario is of online communities that unlock features with increase in user reputation such as StackOverflow. New users in such spaces have limited activity choices to explore, and one may have to investigate other approaches for assessing user retention.

3.6.2 Future work

In this work, we have built models that are able to predict about 7/10 users correctly among those who return as well as those who do not. By doing this, we have a better way of identifying these two classes of users based on just their first-session user activity. In the process of building these models, we have also offered a metric that can capture the diversity of user activity. There are two ways in which these might be useful. A community administrator might want to:

(a) **identify users who are more/less likely to return** to invest effort (e.g., relevant offers, mentors, greetings) in those users who are likely to return and attempt to "recapture" their interest in those who are not.

(b) **use activity diversity as a metric to assess overall site engagement**. Apart from simply using for predictions about longevity, an analysis of the activity types usage distribution may lead to further opportunities to engage users. Also, it tells the site administrators what activity types users engage in and what activity types need more visibility.

Based on our observations of commercial site interactions, we expect that some sites may already be employing some of these methods and we look forward to public research results that establish or refute causality. In the next stage of this work, it seems natural to

look at questions of causality including direction – whether users who are longer surviving tend to be more diverse or vice-versa considering even the possibilities of joint causality with other factors of site design.

Chapter 4

Effects of User Personality

4.1 Introduction

In this piece of work, we wanted to look at something more fundamental about the users that we could use to model their behavior the moment they arrive in the system. More than three decades of research in social psychology shows that psychological traits such as a user's emotions, personality, etc. can predict user preferences and behaviors in a variety of contexts. In order to make accurate recommendations on multiple categories which is the usual case in a typical recommender system, it seems essential to capture the fundamental nature of each individual. Since the literature on personality traits seems rich, we decided to use personality to study new user behavior[1].

4.2 Background and Related Work

4.2.1 Personality and The Big Five Model

Research on personality traits in social psychology and computer-mediated communication since the 1990s has shown that personality can predict user preferences and behaviors in all kinds of contexts, ranging from media [KVE05], to activities such as reading books and attending concerts [KVE05], to appreciation for arts such as music and paintings [RG03, ZUM93], to job success [BM91] and marital satisfaction [KC87], and to the amount of internet and social media usage [BKG+12, CHDZ10, MG14, ROS+09, SKT09]. A lot of

studies focused on understanding how internet usage varied among people with different personality type and we summarize them below under each personality type.

The Big Five Model on Personality, also known as the Five Factor Model is a well-researched and widely accepted model of personality traits and is commonly used in studies examining personality and human behavior [CJM92, MJ92, TC92]. This model has been found to be reliable after testing across multiple languages and cultures [SAMBM07]. The Five Dimensions of this Model, often abbreviated using the acronym OCEAN are:

Openness (to experience):

High Openness people tend to be characterized by higher creativity, imagination and ability to ideate. They possess greater intellectual curiosity and appreciation for novelty or variety in experiences and diversity in interests. Low Openness users are more down-to-earth and conservative.

Prior work found positive correlations for Openness and use of internet for entertainment [TB01] and games [Ten08, TB01]. This may be due to their proclivity for new experiences and variety and curiosity. It was also found in [SKT09] that high Openness users stayed online longer. Others found that Openness to experience was positively related to the use of social networking sites and features such as instant messaging [CHDZ10]. High Openness is associated with an interest in more complex and exciting recreational activities [KVE05].

Conscientiousness:

High Conscientiousness people tend to be highly disciplined, organized, consistent, cautious, and dutiful in their behavior, whereas those with low Conscientiousness tend to be

[1] A version of this work was published as: Karumur, R.P. and Konstan, J.A., 2016, July. Relating newcomer personality to survival and activity in recommender systems. In Proceedings of the 2016 conference on user modeling adaptation and personalization (pp. 195-205). ACM. [KK16]

more impulsive, creative, easy-going, and flexible.

Several works report that high Conscientiousness is negatively correlated with general internet use and time spent online on entertainment, leisure and social networking sites [BP08, LL06, RX11]. On the other hand, high Conscientiousness is positively correlated with time spent on academic/work related sites [LL06]. Some researchers [BKG+12, MG14] reason that conscientious people tend to have less interest in activities related to entertainment such as playing games or listening to music as they involve less planned use of time and are more spontaneous activities, which is opposed to their nature of being cautious and self-disciplined, possessing impulse control and having planned behavior [HO97]. Also, in [HP13], low Conscientiousness people were found to rate more items, whereas high Conscientiousness people were found to rate only the required number of items, and such cautious behavior is again characteristic of high Conscientiousness users. Others found that Conscientiousness was negatively related to ability to undertake difficult or unconventional activities [KVE05].

Extroversion:

Extroverts tend to be more sociable, out-going, energetic and desire the company of others and stimulation in external environments. Introverts are more reserved, self-absorbed, low-key, and seek environments in which stimulation is much lower.

Some researchers claim that Extroverts tend to prefer face-to-face interactions while Introverts tend to prefer use of online channels for self-expression [AHWF02]. Amiel et al found high Extroversion to be negatively associated with comfort in online communication [AS04]. Anolli et al. found a negative relationship between Extroversion and use of online chat [AVR05]. Whereas in [CD10], Extroversion was negatively associated with addiction to gaming, Teng [Ten08] found that Extroverts were significantly more into gaming com-

pared to Introverts. Others have found positive associations between high Extroversion and the use of internet for communication and emails [TL10, WD01] as well as more direct face-to-face friendships [TL10]. Yet others have also found Extroversion to be positively correlated with social network usage [CHDZ10, ROS+09]. Some found that Extroverts do a lot of liking, commenting and expressing their appreciation or sympathy for others, befriending a lot of people [BKG+12]. Others suggest that Extroverts may use the internet for more networking and Introverts may use it to escape their offline personas [OF09].

Agreeableness:

High Agreeableness persons tend to be more cooperative, submissive, flexible, adaptable, tolerant, and empathize with others, whereas low Agreeableness persons are more competitive, challenging and tend to exercise their authority over others.

Some works did not find any relationship for Agreeableness with performance and internet use [BM91, MG14]. Others found high Agreeableness to be negatively related to the time spent online [LL06], and activities such as playing online games [CD10, PBB06]. While Agreeableness was negatively associated with ability to undertake unconventional and difficult activities [KVE05], high Agreeableness users were found to be associated with higher number of tags in [BKG+12]. It was found in [HP13] that high Agreeableness users tend to give ratings that are more positive.

Neuroticism:

Users high in Neuroticism tend to be more sensitive, insecure, pessimistic, self-conscious, and are more susceptible to anger, frustration, anxiety, hopelessness and negative emotions. They are more likely to experience stress and depression. People with low Neuroticism, on the other hand, tend to be calmer and more emotionally stable.

Because High Neuroticism users are susceptible to a lot of negative emotions, use of the internet could provide venues to alleviate such emotions, get rid of insecurity/loneliness and find a sense of belonging. A lot of studies found high levels of Neuroticism of users to be associated with higher use and a greater amount of time spent on the internet, in particular on social networks [AHWF02, AS04, APS00, BP08, CHDZ10, MG14, OF09, RX11, WD01]. Some researchers also found activities of leisure such as playing music or watching movies to be attractive for users with high Neuroticism [SHHH02, WD01]. At the same time, other researchers found that High Neuroticism users are less likely to use the internet to seek information [AHWF02, TB01]. One reason for this may be their insecurity and inability to trust any source of information. Another might be due to their nature of lacking hope and being susceptible to frustration.

Some of prior work has connected personality to rating behaviors [EBRT13, HRBL12] in recommender systems, but we are aware of no work that specifically highlighted relationship between personality and newcomer retention, time investment, level and distribution of early user activity in a system. In this work, we are specifically interested in using personality to model newcomer retention and level of activity since newcomer retention and activity are intricately connected to community success [BLW+04, But01, Duc05, KNK16, PHT09, YWAA10].

4.3 Research Metrics

We look at a variety of metrics to address three research questions:

RQ1. ***How is personality related to newcomer retention?*** We measure retention using the following metrics:

- Number of sessions at the end of first month, and at the end of the first four months.

- Odds of returning for a 2nd, 5th or 10th session [2].
- Time to first return.
- Average return time (time between sessions) during the first four months.

RQ2. ***How is personality related to newcomer investment (time committed to early sessions)?*** To answer this question, we measure:

- Length of first session[3].
- Average session length for first four months of activity.

RQ3. ***How is personality related to newcomer intensity of engagement?*** We define level of activity to be number of ratings, number of tags applied, number of items the person adds to their wish list, proportion of tags to ratings, number of page views and so forth. We now measure:

- Level of activity for first-session.
- Average level of activity per session for the first 4 months.
- Aggregate (total amount of) activity for the first four months.

We recognize that the underlying constructs and metrics have overlap. For example, we categorize metrics such as frequency of logging in as retention, but they can also be measures of intensity of engagement. Our goal is to understand user behavior characterized by these metrics. So we have chosen a single organization for our investigation, and report the resulting data to allow others to draw further conclusions.

[2]We choose these sessions to be consistent with prior work [KNK16].

[3]First session in MovieLens is considerably different from other sessions as most users provide a majority of ratings during this session.

In the next section, we discuss the structure and properties of the MovieLens platform. We frame the hypotheses for user behavior in a system like MovieLens based on existing knowledge of personality types. We then present our findings, summarize them and draw implications from them before we conclude this chapter with limitations and future directions.

4.4 Platform, Study Design, And Methodology

4.4.1 MovieLens

MovieLens[4] is a standalone movie recommendation engine which provides an opportunity for its users to express preferences through rating, tagging and wishlisting movies, while allowing them to view movie details at different levels (summary of plot, trailers, posters, etc). With more than 200,000 registered users worldwide, and an average of 50 new user registrations every day, MovieLens is a suitable platform for studying user engagement, participation, retention and commitment in recommender systems.

MovieLens is primarily used for obtaining movie recommendations based on individual taste preferences. Rating is much more common than tagging, both because ratings build user personalization profiles and because the site design permits ratings at every movie display (with a simple click) while tagging requires visiting a detail page and typing. Clicking on a movie brings up a "movie details page" with plot and cast information, the tagging interface, and various other ways to interact with the movie. Users can add movies to a wishlist anywhere they can rate them, but wishlists are a not a widely-used feature. Very rarely, some users suggest movies to be added to MovieLens through an interface for suggesting movies. MovieLens runs several recommendation algorithms, which it calls "The

[4]http://movielens.org

Peasant", "The Bard", "The Warrior", and "The Wizard" and provides different kinds of recommendations depending on what the user selects as their primary recommender. Occasionally, users change their recommenders too. Our data also suggests that occasionally, users view the posters and watch the trailers on the movie details page. Since rating, tagging and wishlisting movies are the three primary activities on MovieLens, and findings on these activities are generalizable to other recommender systems, we mostly focus our analyses on these three activities. However, we do report results on the number of movie detail pages a user visits and the total number of activities the user performs (which may include all the above activities) as well, for completeness.

4.4.2 Dataset

In order to collect personality information for improving recommendations, one of our colleagues, Tien Nguyen administered a questionnaire based on [GRS03] to MovieLens users during the summer of 2015. Users were asked to respond to questions assessing their personality on a Likert Scale with responses ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). Based on these answers, a score for each of the five personality dimensions was computed for each user on the scale 1-7. We use the results of this survey to study retention, early time investment and activity level of new users.

Personality	**# low users**	**# high users**
Openness	62	430
Conscientiousness	33	228
Extroversion	222	87
Agreeableness	34	113
Neuroticism	59	213

Table 4.1: Counts of users in low and high personality types.

We pick 1008 of these users, who registered between 01 July 2015 and 01 October 2015 and extract their activity log for four months along with their personality scores on the scale 1-7 for this study. MovieLens makes it optional for users to enter any profile information and so only a very small fraction of users have some information about their gender and age. We are therefore unable to report summary statistics about age groups, gender, and location of these users.

Finding effect sizes that are small is a known challenge in personality related research methods. In order to circumvent this problem, increase the sensitivity of statistical analyses used, and ensure comparability of results some researchers [ROS+09, SKT09] divide the personality dimensions into thirds in terms of percentiles and compare the users scoring in the higher third with the users scoring in the lower third. We realized that these approaches might have the possibility of users with similar scores (such as a score of 5 on Openness) coming in two different thirds (in this example, the middle third as well as the upper third). So, we partition the users such that those scoring less than or equal to 2 on each dimension are the low personality type, and those scoring greater than or equal to 6 are the high personality type and those with no strong preferences (scoring between 2 and 6) are the medium personality type. Most results reported in the next section are based on a comparison between the users in the low and high personality types. However, since we had too few low Openness users based on this approach to draw statistically significant conclusions, in order to explore the effect of Openness trait in a useful way, we set the threshold for low Openness at 3.5. Since 4 on the Likert scale corresponds to 'Neither Agree Nor Disagree', 3.5 for Openness has the same directional effect as 2. However, since our goal is to also optimize the sensitivity of our analyses, we retained the lower threshold of 2 for the remaining four personality types. We report the counts of users with low and high personality types in Table 4.1.

4.4.3 Hypotheses

Based on the existing knowledge about personality and user behavior, we frame the following hypotheses for newcomer behavior in the context of MovieLens[5]:

Hypotheses for Openness:

Because Openness is characterized by a tendency to seek variety, and a system like MovieLens offers a diverse collection of movies for users to keep returning, we expect high Openness users to last longer. Because Openness is positively associated with use of internet for entertainment and games [Ten08, TB01] and MovieLens does not offer movies to watch, we expect high Openness users to invest shorter durations of time in their visits, maybe just enough to find movies for watching. Because creative activities excite high Openness users [KVE05] and tagging exercises one's creativity we expect high Openness users to tag more. Because high Openness users have greater curiosity and a desire for entertainment, we expect them to have already watched a lot of movies and therefore add less movies to their wish lists compared to low Openness users. Because curiosity is characteristic of Openness users, we expect them to visit more movie detail pages.

O1: Openness is positively correlated with likelihood of retention.

O2: Openness is negatively correlated with time investment per session.

O3: Openness is positively correlated with tagging movies.

O4: Openness is negatively correlated with wishlisting movies.

O5: Openness is positively correlated with visiting movie detail pages.

[5]We do not state all the possible combinations of hypotheses for each personality type because nothing we know of their nature suggests an expected behavior for certain actions for some personality types

4.3.2 Hypotheses for Conscientiousness:

Because Conscientiousness is characterized by self-discipline and planned behavior, we expect high conscientious users to be more judicious with the amount of time they spend on a site aimed at entertainment. So, we expect lower activity and lower number of movie detail views from high Conscientiousness users who are less spontaneous and easy-going. Prior work [HP13] found evidence for negative correlation between Conscientiousness and rating items, and Conscientiousness and ability to undertake difficult activities. So, we have the following hypotheses in relation to Conscientiousness:

C1: Conscientiousness is negatively correlated with likelihood of retention.

C2: Conscientiousness is negatively correlated with time investment per session.

C3: Conscientiousness is negatively correlated with rating movies.

C4: Conscientiousness is negatively correlated with tagging movies.

C5: Conscientiousness is negatively correlated with wishlisting movies.

C6: Conscientiousness is negatively correlated with visiting movie detail pages.

C7: Conscientiousness is negatively correlated with aggregate activity per session.

4.3.3 Hypotheses for Extroversion:

Prior work suggests that extroverts primarily enjoy environments which stimulate them and so would show positive associations in online environments that are social, and help them network or compete with others, but otherwise have negative correlations with online activity in standalone systems like MovieLens. So we make the following hypotheses:

E1: Extroversion is negatively correlated with likelihood of retention.

E2: Extroversion is negatively correlated with time investment per session.

E3: Extroversion is negatively correlated with rating movies.

E4: Extroversion is negatively correlated with tagging movies.

E5: Extroversion is negatively correlated with wishlisting movies.

E6: Extroversion is negatively correlated with visiting movie detail pages.

E7: Extroversion is negatively correlated with aggregate activity per session.

4.3.4 Hypotheses for Agreeableness:

Because high Agreeableness is associated with a tendency to trust others [JI02], we expect more consumption behavior from high Agreeableness users. Because low Agreeableness persons tend to exercise their authority over others, we expect them to actively critique and thus contribute to activities such as rating and tagging movies. Since MovieLens is primarily a rating system, we expect low Agreeableness users to stay longer and offer their critiques. So, we have the following hypotheses in relation to Agreeableness:

A1: Agreeableness is negatively correlated with likelihood of retention.

A2: Agreeableness is negatively correlated with time investment per session.

A3: Agreeableness is negatively correlated with rating movies.

A4: Agreeableness is negatively correlated with tagging movies.

4.3.5 Hypotheses for Neuroticism:

Neuroticism is associated with insecurity and loneliness and a tendency to seek a sense of belonging. So prior work found Neuroticism to be positively related to time spent on social networks and sites with leisure activities such as playing games or watching movies. Since MovieLens is only a movie recommender, we don't necessarily expect any relation to time spent online. Since high Neuroticism users are insecure, there may be a tendency to exercise their opinion on a group of people. So we expect positive correlation with activities such as rating and tagging which annotate the system's items. Since high Neuroticism users

often change their mood, it may be hard to understand their wishlisting behavior and we hypothesize that low Neuroticism users or emotionally stable users have higher activity on tasks such as wishlisting movies. High Neuroticism users are known to be not good at information-seeking, a behavior that may be likely due to their inability to trust any source of information [AHWF02, TB01]. We, therefore, expect negative correlation to browsing pages about movie details:

N1: Neuroticism is positively correlated with rating movies.

N2: Neuroticism is positively correlated with tagging movies.

N3: Neuroticism is negatively correlated with wishlisting movies.

N4: Neuroticism is negatively correlated with visiting movie detail pages.

4.5 Method

To validate the hypotheses, we compute several metrics at several points in time. Due to space constraints, we report only a few of them that typified our results in this paper. We use the term 'session' to mean a normal login period that begins with the user signing in and ends with the user logging out or with the expiration of the cookie. However, since most users multitask (use multiple tabs and switch between them), they make it harder to record their true session length as there is no explicit logout action in the MovieLens data log. So we computed session lengths explicitly as the differences between their first recorded activity and their last recorded activity per unique session ID. The samples in the low and high groups, although independent are not necessarily normally distributed. So, we use the Wilcoxon Mann-Whitney-test to determine whether the users in the low and the high personality type groups differ significantly in terms of their behavior in relation to the metrics listed in the research questions section. In the cases where one of the groups has a lot of

zeros for the metric under consideration (this is mostly the case with the number of tags or the number of movies the user adds to their wishlist), we step away from comparing low and high personality groups and use the personality scores on the original 1-7 scale. We employ the Poisson, Negative Binomial, Zero-inflated Poisson, or Zero-inflated Negative Binomial models, as appropriate, subsequently testing the assumptions for each, to draw conclusions about effect sizes. Our interpretations will therefore follow two different patterns, one directly making a comparison between high and low personality type and the other talking about the change in the metric score associated with an increase/decrease in the particular personality score. We report results that are significant (at 0.05 level) and marginally significant (at 0.1 level) in the Results section.

4.6 Results

First we combine the findings for the three research questions and report the results grouped by each personality type.

In Table 4.2, we report the five summary statistics for some of the measures we use in the results section. In this table, the minimum values for first return time and average return time are zero. Return times have been computed by subtracting the beginning time of a session from the ending time of the previous session. However, a very small proportion of users logged in simultaneously from another device while using MovieLens from one device and for these cases, our approach yields negative return times. In order to resolve this issue, we consider these users to return in "no time" and assign zeros. Also, 44 users did not return after the first session. We exclude these users for the results reported on first and average return times. The user who had the longest inter-session time had only 2 sessions resulting in the same maximum value of 10190000 sec for average return time

Metric	Min	1st Q	Med	3rd Q	Max
Metrics related to newcomer retention (RQ1)					
No. of sessions during first month	1	2	5	11	120
No. of sessions during first four months	1	3	7	19	451
Return time for second session (sec)	0	8863	54960	253500	10190k
Avg. return time between sessions (sec)	0	151200	334000	780700	10190k
Metrics related to newcomer investment (RQ2)					
First session length (sec)	19	860	1945	3907	35860
Avg. session length (sec)	45.25	587.8	963.9	1456	7218
Metrics related to newcomer intensity of engagement (RQ3)					
No. of ratings in first session	0	28	62	430	430
Total no. of movie detail page views in 1st session	1	18	41	87	1753
Total no. of activities during first session	1	59	119	250	3143
Total no. of ratings for the first four months	0	61	143	305	6364
Total no. of activities for the first four months	1	158	352	731	9833
Total no. of movie detail views for the 1st 4 months	1	65	162	360	4689
Avg. no. of ratings/session during the 1st 4 months	0	8	16	35	516
Avg. no. of movie detail page views/session	1	11	18	31	266
Avg. no. of activities/session	1	22	39	74	679

Table 4.2: Summary Statistics for some of the metrics.

between sessions and return time for second session.

4.6.1 Openness

We find a trend of high Openness users having a 21% higher odds of returning for the fifth session compared to low Openness users ($p < 0.14$). We also find a trend of high Openness users having sessions that are 7.2 minutes shorter than low Openness users during the first session ($p < 0.1$). A unit increase in Openness score on the scale ranging from 1 to 7 is associated with a 21% increase in the expected number of tags from them during the first session ($p < 0.05$) and a 28.3% increase in the expected number of tags from them per

session on an average for all the sessions during the first four months ($p < 0.05$) supporting our hypothesis O3. We also find that a unit increase in Openness score on the scale ranging from 1 to 7 is associated with a 156% increase in the odds of producing both nonzero ratings as well as tags on the aggregate during the four month period ($p < 0.05$) and a 177% increase in the odds of producing both nonzero ratings as well as tags per session on an average during the first four months ($p < 0.05$). We find a trend of high Openness users adding an average of 58.4% of total number of movies added by low Openness users to their wish lists during the first session ($p < 0.1$).

Hyp	Results	Data	Summary
RQ1. ***How is personality related to newcomer retention?***			
O1	Not Supported	High Openness users have 21% higher odds of returning ($p < 0.1$)	Marginally significant for fifth session*
C1	Supported	Low Conscientiousness users return 39.2 hours earlier ($p < 0.05$)	Significant per session on average
E1	Supported	Introverts have 33% higher odds of returning ($p < 0.05$)	Significant for fifth and tenth sessions
A1	Supported	Low Agreeableness users return earlier for a second session ($p < 0.05$)	Significant for second session
RQ2. ***How is personality related to newcomer investment (time committed to early sessions)?***			
O2	Not Supported	Sessions for High Openness users are 7.2 minutes shorter ($p < 0.1$)	Marginally significant for first session*
C2	Supported	Low Conscientiousness users last 8.6 minutes longer ($p < 0.05$)	Significant per session on average
E2	Supported	Introverts last 3.6 minutes longer ($p < 0.05$)	Significant per session on average
A2	Not Supported		Not Significant
RQ3. ***How is personality related to newcomer intensity of engagement and distribution of activity?***			
O3	Supported	21-28% more tags per unit increase in Openness score ($p < 0.05$)	Significant for first session, first four months
O4	Not Supported	Low Openness users wishlist 1.6 times more movies ($p < 0.1$)	Marginally significant for first session*
O5	Not Supported		Not Significant
C3	Supported	+42 in first session, +7 per session on average, +63 in all ($p < 0.05$)	Significant for all mentioned periods
C4	Not Supported		Not Significant
C5	Supported	13% less tags/session per unit increase in Conscientiousness ($p < 0.05$)	Significant per session on average
C6	Supported	+15 in first session, +8 per session on average ($p < 0.05$)	Significant for all mentioned periods
C7	Supported	+65 in first session, +18 per session on average ($p < 0.05$), +121 in all	Significant for mentioned periods
E3	Supported	+26 in first session, +52 in all ($p < 0.05$)	Significant for first session, first four months
E4	Supported	29% less tags/session per unit increase in Extroversion ($p < 0.05$)	Significant per session on average
E5	Not Supported	+1 additional movie ($p < 0.1$)	Marginally significant for first four months*
E6	Supported	+30 in first session, +6 per session on average, +81 in all ($p < 0.05$)	Significant for all mentioned periods
E7	Supported	+67 in first session, +10 per session on average, +156 in all ($p < 0.05$)	Significant for all mentioned periods
A3	Not Supported	+25 during first session, +45 in all ($p < 0.1$)	Marginally significant results found*
A4	Not Supported	24% more tags/session per unit increase in Agreeableness ($p < 0.05$)	Significant per session on average
N1	Not Supported	62% higher odds of nonzero ratings/session per unit increase ($p < 0.1$)	Marginally significant per session on average*
N2	Supported	16% more tags per unit increase in Neuroticism ($p < 0.05$)	Significant for first session, first four months
N3	Supported	26% decrease in wishlists per unit increase in Neuroticism ($p < 0.05$)	Significant per session on average
N4	Not Supported		Not Significant

* We saw marginally significant effects ($p < 0.1$) at this amount and we report them as trend evidence; these might deserve further investigation.

Table 4.3: Summary of findings (selected results listed for each hypothesis).

4.6.2 Conscientiousness

We find that low Conscientiousness users return by a median of 39.2 hours earlier for the next session on an average for all session return times during the first four months ($p < 0.05$) and also a trend of returning 5.4 hours earlier for the second session ($p < 0.1$) compared to high Conscientiousness users supporting our hypothesis C1 that low Conscientiousness users show more likelihood of retention compared to their counterparts. We find that low Conscientiousness users last longer per session by a median of 8.6 minutes on an average for all sessions during the first four months compared to high Conscientiousness users ($p < 0.05$) confirming our hypothesis C2 on time investment per session. Low Conscientiousness users rate a median of 42 more movies during the first session ($p < 0.05$), 7 more movies on an average per session for all sessions ($p < 0.05$) and 63 more movies on the aggregate for the first four months ($p < 0.05$) compared to high Conscientiousness users. These findings support our hypothesis C3 on rating movies. We do not find statistically significant difference between number of tags produced by users in the high and low Conscientiousness groups. A unit increase in Conscientiousness is associated with a 13% decrease in the number of movies wishlisted on an average per session for all sessions ($p < 0.05$) supporting our hypothesis C5. We find a trend of low Conscientiousness users viewing a median of 15 additional movie detail pages during the first session ($p < 0.1$) and a statistically significant median of 8 additional movie detail pages per session on an average for all sessions during the first four months ($p < 0.05$) compared to their counterparts. This supports our hypothesis C6 on visiting movie detail pages. We find that low Conscientiousness users perform a median of 65 more activities during the first session ($p < 0.05$), 18 more activities on an average per session for all sessions ($p < 0.05$) and a trend of 121 more activities on the aggregate for the first four months ($p < 0.1$) compared to high Conscientiousness users. These

findings support our hypothesis C7 on overall activeness of low Conscientiousness users.

4.6.3 Extroversion

Introverts visit more frequently by a median of 1 additional session during the first month ($p < 0.05$). We also find a trend of introverts visiting more frequently by a median of 1 additional session on the aggregate four month period ($p < 0.1$) compared to extroverts. Introverts have 34.5% higher odds of returning for the fifth session ($p < 0.05$) and 33.5% higher odds of returning for the tenth session ($p < 0.05$) compared to extroverts. We find a trend of Introverts returning a median of 3.2 hours earlier than extroverts for a second session ($p < 0.1$). All these confirm our hypothesis E1 that Introverts are more likely to retain in the community compared to extroverts. Introverts last for a median of 215 seconds more on an average per session for all sessions during the first four months compared to extroverts, supporting our hypothesis E2 on investment. Introverts rate a median of 26 more movies during the first session ($p < 0.05$) and 52 more movies on the aggregate for the first four months ($p < 0.05$) compared to extroverts, supporting our hypothesis E3 on relationship between Extroversion and rating movies. A unit increase in Extroversion on the score ranging from 1 to 7 is associated with a 40% decrease in the expected number of tags during the first session ($p < 0.05$) and a 29% decrease in the expected number of tags per session on an average for all the sessions during the 4 month period ($p < 0.05$). These findings support E4. We find a trend of Extroverts wishlisting an average of about 55.4% of the total number of movies wishlisted by Introverts during the first session ($p < 0.1$) and Introverts wishlisting a median of 1 additional movie on the aggregate during the entire four month period compared to extroverts ($p < 0.1$). Introverts view a median of 30 additional movie detail pages during the first session ($p < 0.05$), 6 additional movie detail pages on an average per session for all sessions ($p < 0.05$) and 81 additional movie detail pages on

the aggregate for the first four months ($p < 0.05$) compared to extroverts supporting our hypothesis E6. Introverts perform a median of 67 additional activities ($p < 0.05$) during the first session, 10 additional activities per session on an average for all sessions during the first four months ($p < 0.05$) and 156 additional activities on the aggregate for the first four months ($p < 0.05$) compared to extroverts, supporting our hypothesis E7.

4.6.4 Agreeableness

Low Agreeableness users show a trend of visiting more frequently (by a median of 3 sessions more) during the first month ($p < 0.1$) and having a 35% higher odds of returning for the fifth session ($p < 0.1$) compared with high Agreeableness users. We find that low Agreeableness users return for the second session 4.7 hours earlier than high Agreeableness users ($p < 0.05$). We find a trend of low Agreeableness users rating a median of 25 more movies during the first session ($p < 0.1$) and a median of 45 additional movies on the aggregate during the first four months ($p < 0.1$) compared to high Agreeableness users. A unit increase in Agreeableness is found to be associated with a 24.3% increase in the expected number of tags per session on average for all sessions during the first four months ($p < 0.05$). Here we find a direction opposite to the assertion we made for hypothesis A4. One reason for this might be that these users are mostly producing tags similar to what others have produced before just by adding existing tags, which is characteristic of Agreeableness users (to agree with others). This may also be a reason why we do not find any statistically significant relationship between Agreeableness and early time investment. Both high and low Agreeableness users might be investing in different activities (rating and tagging). Bachrach et al (2012) find Agreeableness to be a hard trait to predict using Facebook profile features and report very low R 2 for their model (0.01) [BKG+12]. Others[BM91, MG14] do not find any relationship between Agreeableness and internet use. So, it is not surprising that many

of our results are only significant at 0.1 instead of 0.05.

4.6.5 Neuroticism

We find a trend of a unit increase in Neuroticism being associated with a 61.5% increase in the odds of having both nonzero ratings and tags per session on an average during the first four months ($p < 0.1$). A unit increase in Neuroticism on scale with scores ranging from 1 to 7 is associated with a 16.5% increase in the expected number of tags during the first session ($p < 0.05$). This finding supports our assertion in hypothesis N2 on the relationship between Neuroticism and tagging activity. A unit increase in Neuroticism is found to be associated with an average decrease of 26.4% in the number of movies wishlisted per session for all sessions during the first four months ($p < 0.05$). Low Neuroticism users wishlist a median of 2 additional movies on the aggregate for the first four months compared to high Neuroticism users (p 0.05). These findings support our hypothesis N3. We do not find any statistically significant results to support our assertion on visiting movie detail pages. This may again be due to opposite behaviors on rating and tagging, and wishlisting.

We summarize and report selected findings grouped by the research questions in Table 4.3.

4.7 Discussion

The above results suggest that different personality types use the system differently. Specifically, we find that users with certain personality types (low Extroversion, low Agreeableness) have a higher likelihood of returning to the community compared to their counterparts; users with certain other personality types (low Extroversion and low Conscientiousness) are more active in a system like MovieLens compared with their counterparts; users with some

other personality types show different activity preferences (low Agreeableness users are more likely to rate and high Agreeableness users are more likely to tag); and low and high personality types can show a preference towards consumption vs contribution (ex: high Openness users and high Neuroticism users contribute more compared to their counterparts). All in all, our results show that the challenges of newcomer churn and activity levels can be approached by making use of their personality information.

Implications

Our findings show that there is value in using a stable trait such as personality in deciding how to adapt a recommender system and customize interaction for specific personality types, which features to present to them or how to nudge them towards various existing features, who to recruit at cold-start (e.g., personality types that contribute more annotations), who to recruit for specific tasks (e.g., rating vs tagging), whether to invest particular efforts in them, or how to retain them.

4.8 Limitations and Future Work

In this paper, we investigate the relationship between newcomer retention and activity, and their personality. We expand the theory on personality traits and online behavior by contributing our hypotheses and findings of user activity in one recommender system, MovieLens.

4.8.1 Limitations

MovieLens has the common features of a standalone recommender system with primarily anonymous features. It is not representative of all recommender systems. In particular, it is

not a social system. There are limitations in the kind of data that we have and the kind of activities people can do on MovieLens.

4.8.2 Future Work

One future direction would be to exploit this idea in a wider variety of systems (e.g., that are not standalone, or those which are not anonymous) with different types of social affordances. High Conscientiousness users might use Amazon differently. Extroverts might use social systems differently. We leave all such investigations to future work.

There is also future work to be done in customizing the interface to match personality where it is known. Tkalcic and Chen [TC15] explore other ways in which personality can be used to improve performance of recommender systems such as determining whether or not to present novel, diverse items, improving performance of collaborative filtering algorithms, improving group recommendations and so forth. We focus here on issues of newcomer retention and feature usage which were not explored earlier using personality, but we wish to explore some of these in future.

We had few low Openness users in our dataset. So, in order to explore the effects of Openness trait in a useful way, we set a different lower threshold for Openness. Future work should explore whether finding few low Openness users is endemic to recommender systems or just an artifact of MovieLens. Also, in this work we analyzed personality traits in isolation from each other based on their theoretical independence. Future work, however, should explore ways in which the combination of traits found in each individual can be used to look at relationships with user retention, investment, intensity of engagement, and distribution of activity in various domains.

Chapter 5

Effects of Community Level factors

5.1 Introduction

In order to understand community level factors that affect retention and productivity of newcomers, we look at user logs of new developer collaborations on projects from GitHub, a community for developers who come together to collaborate on projects.

Traditionally, developer collaboration among distributed teams in enterprises was challenging. First, developers could not work on new areas without invitations and were typically assigned tasks to work on. Sometimes, they were even restricted to work only on assigned tasks. This resulted in a lack of awareness of inter-dependencies and changes outside their area of code [VKC10]. Second, there were communication breakdowns and code conflicts, which led to frequent build failures and longer resolution times [CH13, CHC08, Her07]. Third, there were few central portals that recorded individual project activities. To overcome these challenges, many companies are increasingly moving toward *social* coding environments such as GitHub within their enterprise [KDB+15, Kno14, Met15, Pau15, Zak13].

Such collaboration tools have been known to lead to quicker release of products, products that are more reliable and feature-rich, and shorter lead time to fixes. For instance, in 2014, Google declares that because of such tools, over 15,000 of its engineers are able to collaborate on more than 4,000 of its projects (services) executing about 800,000 builds a

day, and deliver many reliable services to the world [MI14].

GitHub (`https://github.com/`), popular as a "Facebook for programmers" [Wei15] is one of the largest social coding platforms which facilitates developers' collaboration on projects. On GitHub, project activity updates are available on a public time line. Because of this transparency, developers can browse through interesting projects, and bookmark (*watch*) and receive updates from them. Developers on GitHub can also join and collaborate on multiple projects simultaneously.

However, from an analysis of public GitHub projects, we find that there are some projects that succeed in attracting a lot of watchers and developer collaboration, and there are others that don't. As enterprises are increasingly adopting environments like GitHub as a common collaboration environment for source code projects, it becomes important to gain a deeper understanding of when projects are likely to succeed in gaining popularity and when developer collaborations sustain in these environments. Therefore, in this work, we set out to understand what project characteristics are likely to attract developers to collaborate on them.

The number of watchers is an indication of the project's quality level [BBS13, DSTH12, VYW+15]. Early identification of project popularity (measured by number of watchers) can have potential benefits for the enterprise such as cutting down further investment on projects less likely to become popular and moving on, or identifying factors leading to such low levels of popularity and guiding them toward attracting watchers and collaborators.

5.1.1 GitHub - The Platform

As developers constantly add new content, features or fix bugs, code changes occur. Version control systems (VCSs) are systems that record such changes over time and are used to revert to earlier versions if and when needed. Centralized Version Control Systems (CVCSs) store

code centrally on a server allowing multiple developers to check out the code and make changes. The risk however, is that, there is just a single point of failure. Distributed version control systems (DVCSs) such as Git (`http://git-scm.org`) help deal with this issue where each developer has a full copy of the entire code repository.

GitHub is a popular code hosting service for projects that primarily use Git as their VCS. In addition, it provides social networking features that allow a community of developers to collaborate and build the codebase. Using a process called *forking*, a developer can make a *clone* of an entire repository of interest, make changes (called *commits*, which could be proposed code changes, enhancements, or bug fixes) remotely, *push* changes to this local repository and send a *pull request* back to the owner of the repository asking them to pull these changes back into the original repository. Forks therefore represent copies of the main codebase made at different points in time by different developers. The number of forks is a rough indicator of the number of people intending to contribute to, or use parts of the repository's source code. An owner or core developer may review such requests and accept only the proposed changes that they find useful. Any *repository* that is publicly available within an *organization* can usually be watched or forked by anyone within that organization.

New features, enhancements, or bugs are typically logged as *issues* to be resolved. A developer may assign issues to themselves or fellow developers working on the same project. Developers involved in handling the issue converse with each other through *issue comments*, although it is not uncommon to use an external platform for communication. Similarly, when individual commits require discussion before being merged into the main codebase, developers converse through *commit comments*. *Watching* is a term used for following the project, and is a rough indicator of interest in the project. When a developer watches a project, they receive updates on events such as issues, new commits or pull requests. Each developer also maintains their own profile or home page which others can view to follow

them and their activities.

5.2 Related Work and Hypotheses

In the context of Open Source Software (OSS) projects, it was found that the a project's environment affects the rate at which developers join it, as well as the likelihood that they become its long-term contributors [ZM11, ZM12]. We therefore hypothesize that a project's environment is likely to affect its popularity. In this work, we characterize a project environment as comprising four components - structure, people, activity and coordination.

Since watching a project indicates potential interest in its activity [DSTH12], prior work used 'number of watchers'on a project as a proxy for its popularity and quality. This was also found to be highly correlated with 'number of forks'[LRM14, SBK+14, VYW+15].

In a variety of online contexts (Facebook, MovieLens, Wikipedia, etc.), prior research shows that specific goals are more likely to attract users to tasks than general goals [BKJ09, LBL+05, ZKK12b]. In GitHub, this means that including the set of planned or desired features in the project description is more likely to attract people to it than a general, rather vague description of the project. GitHub developers typically record the structural features in the forms of test cases [PSL+13] and non-bug issues [VYW+15] (issues that are not related to bugs) and these have been used as proxies for the project's interestingness in prior work. Therefore, we hypothesize that the greater the number of non-bug issues and/or test cases, the more interesting the project is (i.e., there is more information about the project's activities) and is therefore likely to become more popular (attract more watchers).

H1. *Number of non-bug issues is positively related to project popularity.*

Others [VKSL03] have found that the size of the developer community affects project success. The number of core developers working on the project is correlated with the level

of interest, the volume of activity, and the ability to handle the incoming volume of pull requests [VYW+15, VKSL03]. In concert, these contribute to the project's popularity [MDH13]. While it was found in [BBS13, DSTH12] that users learn of interesting projects by watching activity of users with high status, developer status did not affect their activity [LRM14]. Therefore, we hypothesize 'number of developers with prior experience'as a 'people'component affecting the quality of the project, and thus its popularity. Similarly, we hypothesize 'willingness to accept new changes'as a 'coordination component'affecting popularity.

H2. *Number of developers with prior experience is positively related to project popularity.*

H3. *Pull request acceptance rate (willingness) is positively related to project popularity.*

In GitHub, the existence of extensive comments surrounding a commit indicates potential controversy about the commit. As most projects lack a formal documentation (many have terse READMEs or Wikis which are automatically created by GitHub), developers use attributes such as comments and developer activities [SBK+14] to learn more about the project. They consider the amount [DSTH12] and frequency [SSC+14, TS10] of activity in deciding which projects to contribute to. Also, commits, which characterize the volume of activity, provide information on the developer's intentions, the style and pace of their coding, the contributors to the various parts of the code, and the progress of the project through its various phases of development [DSTH12]. Based on these, we include the number of commits, the number of commit comments and the frequency of commits as the 'activity'-based components in our analyses.

H4. *Number of commits is positively related to popularity.*

H5. *Number of commit comments is positively related to popularity.*

H6. *Commit frequency is positively related to popularity.*

While some of these explore the impact of particular activities and behaviors based on qualitative studies, they do not contemplate the interaction and relationship between these factors. Prior research has shown that the early distribution of activity is indicative of longevity and long term success [KNK16]. However, to our knowledge, there is no existing research that makes use of a project's early environment to predict its long-term popularity. Also, much of prior work restricted their research to large and active GitHub projects [Gou13, SBK+14]. Since most projects on GitHub are small and do not have many commits [KGB+14], their findings may not generalize to these projects. In this work, we look at several of these factors in conjunction and understand how they, based on the project's early time period, affect its eventual popularity. In our analysis, we also pick equal samples of popular, moderately popular and less popular projects to complement the findings from prior work.

5.3 Platform, Study Design and Methodology

5.3.1 GitHub

Typically, enterprises do not allow public access to their proprietary code repositories. However, according to Kalliamvakou et al [KDB+15], enterprise projects have practices very similar to OSS projects. We therefore use publicly available GitHub data to draw conclusions for the enterprise context. GitHub has gained popularity as a tool for collaborative development (social coding) because its ease of use, low cost, and presence of social features such as timelines, following, and trending projects. As of December, 2013, GitHub boasts more than 3.5 million users and 10 million repositories. Moreover, a number of enterprises are also adopting GitHub for development within their organization [KDB+15, Kno14, Met15, Pau15, Zak13].

5.3.2 Dataset

Using a REST API, the GHTorrent Project [Gou13] periodically extracts project and developer activity data from GitHub's public event timeline. Because GitHub disabled the API end point for retrieving membership to organizations in November 2014, there are inconsistencies in user activity timestamps in later dumps. We therefore use the MySQL database dump of August 2014 for our study. The curated data consists of information on users, projects, issues, commits and pull requests, issue comments, pull request comments and commit comments produced by these developers from the beginning of their time in GitHub. This forms our initial dataset. A full schema of the database is available in [Gou13].

5.3.3 Method

Language trends on GitHub reveal that certain languages are consistently popular during our analysis period [La15]. We believe that projects belonging to other less popular languages may not have much information particularly in the case of projects with very few watchers. We further assume that the popular projects among the less popular languages may have characteristics similar to popular projects in popular languages. We therefore consider only projects implemented in the seven languages JavaScript, Java, Python, Ruby, C, C++ and PHP in our analysis.

Because GitHub gained popularity recently, we select projects that were created no earlier than January, 2012. Since we want to assess project popularity (as the watcher count) after one year, we select projects with creation dates no later than June, 2013 (the last recorded activity date in our dump is Aug 18, 2014). Accordingly, we end up with 152,147 projects with at least two watchers from January 2012 to June 2013 and follow them through one year and count their popularity, measured by number of watchers after one year.

Based on prior work, we partition these projects into three classes: those with 2-9 watchers, those with 10-99 watchers and those with 100-999 watchers [BB14]. We exclude projects with only one watcher as we assume that may be personal work which is not intended to be shared. We also exclude the projects with a watcher count of more than 1000 as outliers. External factors such as advertisements of the projects on HackerNews, etc. may have contributed to such popularity and it is challenging to analyze such factors. Of the remaining projects, 101,733 have between 2-9 watchers, 26,246 have between 10-99 watchers and 1,948 have between 100-999 watchers. From each of the three classes, we pick random samples of 1,500 projects each, and assemble a dataset of 4,500 projects for answering our first research question.

Less popular projects suffer from low activity levels and have lesser numbers of developers as well. We therefore wish to avoid the effect of project popularity as a factor in our analysis for the second research question. So, we pick the 73,490 users belonging to the class of projects with more than 100 watchers for answering our second research question.

5.4 Results and Discussion

RQ1. *Do early characteristics of a project relate to its long-term popularity? If so, how?*

In order to answer this question, we plot the number of watchers as a function of week from project creation. We observe that the number of watchers start differing widely from as early as 10 weeks for projects belonging to each of these three classes and different languages. We therefore fix the early period to 3 months for all the 4,500 projects. Based on related work and hypotheses H1 through H6, we consider the following characteristics during this period for all projects - structural (number of non-bug issues), people (number of developers with prior experience in GitHub), activity (number of commits, number

of commit comments, frequency of commits) and coordination (willingness to accept pull requests) as predictors.

Predictors:

Number of non-bug issues: We use the number of non-bug issues as a structural aspect of early project environment based on [40]. In order to identify the number of non-bug issues in the first three months, we use regular expressions that eliminate issues with words 'bug', 'wontfix', 'fix', 'issue', 'error'and 'invalid'in the issue's title.

Number of core developers: In order to get a more accurate representation of core developers, prior work considers core developers to be those who have direct commit access to the master repository and those who can merge pull requests into the master repository [Gou13]. We characterize them similarly and consider core developers during the first three months.

Number of experienced developers: is the total number of core developers with prior experience in GitHub who made contributions during the first three months. We measure prior experience in terms of presence of at least one prior commit in GitHub.

(Including both variables - *# core developers* and *# experienced developers* in our models, we found that the *# core developers* variable did not prove to be a useful and significant predictor in the presence of *# experienced developers* due to their correlation.)

Number of commits: is the total number of commits made by various developers in the project during the first three months from its creation.

Number of commit comments: is the total number of commit comments made by various developers in the project during the first three months from its creation.

Frequency of commits: is the median value of the time between two subsequent commits in the project during the first three months from its creation.

Willingness to accept PRs: This is the ratio number of pull requests accepted/merged

out of the total number received.

Language: Because project popularity may vary by language, we include this in addition to the above predictors in our models.

Response:

Our response variable is the number of watchers for the project at the end of one year.

Analysis and Evaluation:

Note that most of our predictors as well as our response variable are counts. Generalized linear models such as Poisson Regression models or Negative Binomial Regression models are used to model non-negative integer responses and would therefore be appropriate for modeling this kind of data. Our data is also characterized by over-dispersion (i.e., the variance is much higher than the mean) in some of the variables, particularly the response. Therefore, Negative Binomial Regression would be most appropriate [BM16].

The model's fitness to the data can be determined either by comparing actual values with the predicted values using the model, or by comparing the model with other competing models [BM16]. Comparison with competing models seems more appropriate in this context since the outcomes are over-dispersed counts. We use the Akaike Information Criterion (AIC) to evaluate the goodness of fit for each of these models. AIC rewards goodness of fit of the model to the data, while penalizing complexity (i.e, more predictors). AICs are always compared with each other and individual AIC magnitudes are not interpreted by themselves as they are affected greatly by sample size. In general, the smaller the AIC among a set of candidate competing models, the better the model.

Using the predictors above, we consider thirteen candidate models by sequentially including predictors related to structure, activity, coordination and people. Parameters are mean-centered as and when necessary. The parameters included in a particular model along with the model's corresponding AIC values are listed in Table 5.1. From the set of candi-

M. No.	*Parameters included*	*AIC*
1	Non-bug issues	47246
2	Non-bug issues, commit comments	47239
3	Non-bug issues, commit comments, language	47127
4	Commits, commit frequency	47187
5	Non-bug issues, commits, commit comments, commit frequency	47172
6	Non-bug issues, commits, commit comments, commit frequency, language	47031
7	Non-bug issues, commits, commit comments, commit frequency, willingness	46440
8	Non-bug issues, commits, commit comments, commit frequency, willingness, language	46277
9	Non-bug issues, commits, commit comments, commit frequency, willingness, prior exp. users	46415
10	Non-bug issues, commits, commit frequency, willingness, prior exp. users, language	46247
11	Non-bug issues, commits, commit comments, commit frequency, willingness, prior exp. users	47006
12	Non-bug issues, commits, commit comments, commit frequency, prior exp. users, language	46850
13	Commits, commit frequency, willingness, prior exp. users, language	46249

Table 5.1: Models considered for determining project popularity from early project characteristics.

date models in Table 5.1, we see that models 10 and 13 seem to have the lowest AIC with model 10 having the additional factor 'non-bug issues'. Although this factor does not have a large effect size (see Table 5.2) and has little impact on the model, we decide to use the effect sizes for model 10 for interpretation for, it has the lowest AIC. The values are shown in Table 5.2.

From Table 5.2, we see that language, number of users with prior experience in GitHub (`priorExp`) and willingness to accept new pull requests (`willingness`) affect popularity with significant large effect sizes. Number of commits (`numCommits`) affects popularity but is only marginally significant. Commit frequency (`commitFreq`) and number of non-bug

Parameter	*Coeff (Std. Error)*	*significance*
(Intercept)	3.926 (0.077)	***
langC++	0.114 (0.118)	
langJava	0.049 (0.098)	
langJavaScript	0.638 (0.084)	***
langPHP	-0.046 (0.097)	
langPython	0.211 (0.093)	*
langRuby	0.292 (0.094)	**
priorExp	0.057 (0.001)	***
numNonbug	0.0003 (0.0001)	**
numCommits	0.00008 (0.00005)	.
commitFreq	1.6e-06 (4.2e-07)	***
willingness	1.222 (0.047)	***

Table 5.2: Output of the Negative Binomial regression model predicting project popularity at the end of one year with 'C'as the reference level for language. ($p < 0.001$: *** $p < 0.01$: ** $p < 0.05$: *)

issues (numNonbug) significantly affect popularity, but the effect sizes are too small to be of practical value. The coefficients in a Negative Binomial regression model do not have a simple linear interpretation. They have an additive effect for the response variable in the logarithmic scale and a multiplicative effect for the response variable in its linear form. The individual coefficients need to be exponentiated to interpret them. For instance, the variable priorExp has a coefficient of 0.057. This means that a unit increase in the number of developers with prior experience is associated with a 0.057 increase in the expected log count of number of watchers. Putting it differently, this means that there is a exp(0.057) = 1.06 times the number of watchers or a 6% increase in the number of watchers.

All in all, our findings indicate that the presence of each additional developer with prior development experience (priorExp) in GitHub is associated with 6% more watchers, and, a unit increase in the proportion of pull requests accepted (willingness) is associated with a 240% increase in the number of watchers. We separate the effect of language on popularity

Lang1	*Lang2*	*% greater*	*significance*
JavaScript	C	89.7%	***
JavaScript	C++	68.8%	***
JavaScript	Java	80.3%	***
JavaScript	PHP	98.1%	***
JavaScript	Python	53.3%	***
JavaScript	Ruby	41.4%	***
Python	C	20%	*
Python	Java	17.4%	*
Python	PHP	29.3%	**
Ruby	C	40%	**
Ruby	Java	27.1%	**
Ruby	PHP	40.2%	***

Table 5.3: Table showing % by which number of watchers in language 2 are greater than language 1 at the end of one year. ($p < 0.001$: *** $p < 0.01$: ** $p < 0.05$: *)

into Table 5.3. The data shown in Table 5.2 is the output of the regression model that takes C as the reference level for language. By changing the reference level to other languages in turn, we re-run the regression seven times in order to generate differences in number of watchers between projects belonging to different pairs of languages. A further elaboration on how popularity varies by language is shown in 5.3.

Table 5.1 includes significant differences in number of watchers at the end of one year in projects belonging to different languages. The '% greater' column shows by how much the number of watchers in language 1 is more than language 2 at the end of one year. For example, JavaScript has 89.7% more watchers at the end of one year compared to C, and Python has 20% more watchers at the end of one year compared to C. Others in the table follow similar interpretation. The numbers themselves may not be important, for they may vary over time. However, the essence is that projects belonging to different languages attract varying numbers of watchers.

Discussion:

We only find trend evidence for hypotheses H1, H4, H5 and H6 due to small and/or marginally significant effect sizes. It is likely that features are documented in other ways such as in README files and number of non-bug issues may not be representative of the project's features. Number of commit comments is not part of the model under consideration. Because most projects on GitHub are small, the number (and frequency) of commits is less. So, it makes sense to see small and/or marginally significant effect sizes for these two variables. We accept hypotheses H2 and H3 due to their large and statistically significant effect sizes from Table 2. We conclude that early characteristics (in our case, project characteristics in the first three months) such as number of prior experienced developers, rate at which pull requests are accepted and language do affect long-term popularity (number of watchers after one year). This means that if a project is low in its popularity, recruiting experienced developers into the team is likely to raise the popularity. Similarly, if a project's acceptance rate of pull requests is low, one might evaluate factors leading to it and try to increase the acceptance rate to see an associated increase in the project's popularity. The choice of language may be dependent on the nature of the end product. The projects that use less popular languages may benefit from additional experienced developers to better handle the volume of pull requests.

5.5 Conclusion and Future Work

In this work, we investigate early factors leading to a project's long-term popularity and developer commitment in GitHub. To the best of our knowledge, this is the first work to make use of early factors from a project's environment to predict longer term success in collaborations in the context of social coding environments. Also, this is the first work examining a number of these factors in conjunction. We identify factors based on correlations and

qualitative studies from prior research and build models using these factors for predicting projects that succeed and those that fail in attracting watchers. Our findings indicate that the choice of primary language of development, number of experienced developers and rate at which pull requests get accepted during the first three month period are indicative of the project's long-term popularity.

5.5.1 Limitations and Future Work

In this work, we limit our analyses to seven languages that remain consistently popular during our analysis period. Future work may study these effects by including projects that use other languages. Due to resource limitations, we were unable to download all the commits and extract features based on test cases, code written, style, etc. Future work should explore some of these features and consider building models that include these.

Watchers on a project include people who are interested in the project as the end product. It is likely that factors such as presence of instructions for installing, configuring and running the code, license/copyright information, names and brief descriptions of all sub-modules / libraries of the project, information on troubleshooting, contact information (email addresses, website, etc) may affect the project's popularity. Future work should consider including some of these features under structural aspects, and see how they might affect the models.

Developers join projects for variety a of other reasons such as an interest in the end product, building reputation, learning new technologies and so on [LW03, LH03]. While we cannot directly obtain this information without interviewing/surveying, prior research shows that incoming developers decide whether a project helps them accomplish these goals using their social contacts [GUSA05, HMZ08, US05]. Also, with prior contacts, the greater level of mutual trust and familiarity with each other's expertise leads to suc-

cess in terms of achieving better coordination, complementing expertise of each other [FS00, MT06], greater efficiency and decision making [HPB98]. A number of works also show that prior social ties play a key role in acceptance of contributions of new developers [DSTH12, MDH13, PSL+13, TDH14, YWF+15]. Future work should therefore include models encompassing the number and degree of social connections of these developers as well in predicting their commitment.

Our work is a preliminary attempt in the direction of using early activity to predict long-term popularity and developer commitment in GitHub. We therefore limited our analyses to fixing the initial period at three months for RQ1 and one week for RQ2. Future work should explore how varying these time periods affects the respective outcomes. Also, a more thorough understanding based on investigations conducted on similar other platforms such as BitBucket and GitHub Enterprise may be needed to inform the design of large-scale social coding environments.

Chapter 6

Effects of Prior Experience

6.1 Introduction

Volunteer groups have existed for a long time in settings such as local non-profits, NGOs, and charity organizations. The explosive growth of computer technology and near-universal access to the Internet have enabled the growth of new forms of volunteer contributions and groups at unprecedented scales. Examples include Wikipedia, OpenStreetMap, Open Source Software projects, product review forums, technical Q&A sites, citizen science projects, and online fund-raising, to name but a few. Bringing in people who actually do work and stick around long enough is a common problem to all these groups.

It is also common for volunteers in these groups to come in with varying levels of prior experience that shapes their activity in the group and perhaps, their success within the group. Prior work on new volunteer retention and productivity in online and offline groups, however, suggests that the effect of prior experience[1] on newcomer success is complicated, with some studies showing positive effects [BF05, BK08, MGMZ13, MHE+10, WTWT15] and others indicating negative ones [COK14, Deu59, LaR09]. Because prior experience is something that is usually visible and can be objectively measured, it is both theoretically and practically important to understand how what volunteers carry from their prior work experience affects their performance in the context of a new group. In this paper, we want

[1] A version of this work was published as: Karumur, R.P., Yu, B., Zhu, H. and Konstan, J.A., 2018, April.

to unpack prior experience and resolve conflicts in prior work. We, therefore, ask the following Research Question:

How does a new volunteer's prior experience affect their early retention and productivity in the group they join?

To answer this question, we review prior literature and identify three types of prior experience: (i) generalized prior work-productivity experience, which is prior experience associated with normal production activities (i.e., non-leadership activities) in other similar volunteer groups, (ii) prior leadership experience, which is the experience of organizing activities and managing people in other volunteer groups, and (iii) localized prior work-productivity experience, which is the amount of work a volunteer would invest in a group before joining the group (in other words, the "internship" experience). Because early identification of group failures can help community moderators intervene in a timely manner and shape the group for success [KNK16], we identify two early outcomes: retention and productivity at the end of the first quarter after joining a group.

We explore the effects of prior experience in the specific context of WikiProjects. WikiProjects are subgroups in Wikipedia, which are intended to help organize volunteer effort around building and improving articles in specific topic areas. WikiProjects often share structure and volunteer membership. Additionally, the volunteers' editing records on Wikipedia are visible to the public, which gives us an opportunity to explore whether and how their prior record is predictive of future contributions to a group they join.

Our findings, indeed, show mixed effects of prior experience on retention and productivity in the group they join:

Content is King, Leadership Lags: Effects of Prior Experience on Newcomer Retention and Productivity in Online Production Groups. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 506). ACM. [KYZK18]

- Generalized prior work-productivity experience is positively associated with retention but negatively associated with productivity.
- Prior leadership experience is negatively associated with both retention and productivity.
- Localized prior work-productivity experience is positively associated with both retention and productivity within that focal project.

Using our findings, we hope to advance knowledge about the behavior of members in online production groups, the nuanced effects of prior experience, and inform the designs of early interventions aimed at shaping group success in online social collaborative knowledge systems.

The rest of the paper is organized as follows. In the next section, we discuss the theory and hypotheses concerning retention, productivity, and prior experience within online groups. We then discuss our platform, dataset, and study methodology. We state our research methods and describe our findings followed by a brief discussion of the results. We then conclude by discussing how prior experience may be more broadly effective in supporting the design and management of online social collaborative knowledge systems.

6.2 Theory and Hypotheses

6.2.1 Retention and Productivity

A number of online groups face the problems of lack of early retention and productivity from new users. 46% of members of guilds in World of Warcraft leave their group in less than a month, migrating to other groups within the game [WDX+06]. On MovieLens, 60% of new users do not come back after the first session [KNK16]. In The Pearl Open Source

Development Project, more than half of the newly registered developers never showed up after their first post [Duc05]. In Usenet groups, 68% of newcomers did not return after their first post [ABJ+06]. Half of the social, hobby, and work mailing lists had no traffic over a 130-day period [But99] and even in active mailing lists, less than 50% of subscribers posted even a single message in a 4-month period [But99].

Two outcome measures have been extensively studied in offline work groups as determinants of group success: the member retention to contribute to the group effort, and the quantity of work output [ABJ+06, CRR10].

Prior organizational science literature views the lack of volunteer retention from three different perspectives [CT86, CHW05, QSTC14]. Because the online community literature is generally in favor of sustaining a steady group of volunteers for continued production [KRK+12], the first and more dominant view is about the negative effects of low retention of volunteers. These include the loss of productive volunteers [Str90], the loss of social capital [DS01, Hus95], the cost of training new, inexperienced volunteers [Dar90], and the weakening of knowledge resources of the organization [Hus95] - all of which deplete the available resources, disrupt the routines and established social ties, threaten the cognitive structures, and the eventual sustainability of the group. The second view sees the positive effect: helping screen out under-performing volunteers [KP85]. The third adopts a more neutral view that suggests that new volunteers with new skills and knowledge replace those who leave, maintaining the critical mass, and this may be optimal for organizational performance [DS01]. To reconcile the above arguments, Hausknecht and Holwerda [HH13] argue that the traditional, aggregated measures of volunteer dropouts such as turnover rates hide variation in the key causal factors that predict retention and performance and so, the specific details concerning who is being retained and who is not are more important than the level of turnover itself [HT11]. As an example, the loss of a productive manager may

be more damaging than the loss of an under-performing employee. Depending on these details, the same level of retention could have different consequences [HT11, YWL+17].

While it is ideal that volunteer groups should achieve both retention and productivity simultaneously, often, there may be a tension between the two [Deu59, WCRR12]. For instance, it is likely that volunteers may "free-ride" i.e., stay but not contribute, in some groups [AVF85]. Or, core members who take the lion's share of workload may eventually "burnout" and leave the group [BVDZLD07, CD93]. Or, core contributors might feel they have accomplished their mission by contributing everything they know and might stop contributing as further contribution may require more research and effort [Tay09]. Or, the presence of multiple experienced core members with clashing interests might lead to conflicts that can erode each others' energy and enthusiasm causing them to leave the group. Accordingly, prior research has examined a number of factors influencing productivity and retention in online production communities. Some of them include members' personality [KK16, KNK17], socialization tactics used [CAKL10, FDKP11, FKPK12, FKL+12, KPK09], members' ability to identify with and integrate into the group [CFG13, RKK07, YRZ17], the diversity of the subgroups they belong to [CRR10], the availability of activities to perform [KNK16], the leadership behaviors within the group [ZKK12a], the feedback [ZZH+13], and the type and amount of social support they receive [WKL12] from other members of the group.

There is also some work suggesting that prior experience is predictive of future retention and productivity in groups [BK08, PHT09, PPET10, YRZ17]. However, a closer examination of the theory on retention of newcomers in offline and online groups suggests a more complicated relationship of prior experience with retention and productivity, with some types showing positive effects and others negative ones. And, to our knowledge, there is no work that either makes a clear distinction between the various kinds of prior experience that

a volunteer can potentially possess or draws any conclusions about them either individually or together with other group and individual level factors. In this work, we treat *prior experience* more systematically and examine its effects on a new volunteer's early retention and productivity after they join a group.

6.2.2 Prior Experience

Because volunteers frequently move in and out of groups, it is useful to learn about the impact their prior experience in other groups has on their retention and productivity in the future groups they join.

Much of prior work in online communities suggests that prior experience has a positive impact on both the individuals and the communities as a whole. For example, the theory on Legitimate Peripheral Participation (LPP) widely used to describe the newcomer experience in online communities [BF05] suggests that newcomers' initial peripheral participation is important for them to be acquainted with the tasks, vocabulary, and organizing principles of the community. Experience gained in using editing tools, organizing activities, and communicating and collaborating with other members could positively affect their future performance. Also, prior experience is positively predictive of future productivity and administrative behaviors [BK08, PHT09, PPET10, YRZ17].

On the other hand, research based on analysis of employee behavior in offline organizations suggests that prior experience might have a negative impact on people's performance in a new context. For example, experienced employees are also likely to leave due to mismatch in expectations [DWR09, MG16], the need to suppress their perspectives [Bro15, COK14], unfavorable group structures [MG16], or stress and exhaustion [CD93].

In this paper, we want to resolve the conflicts in prior work by unpacking prior experience. In online peer production groups, we identified three types of prior experience:

(i) **generalized prior work-productivity experience**, which is generalized prior experience associated with normal content production in all the other similar volunteer groups.

(ii) **prior leadership experience**, which is prior experience associated with the tasks of coordination and organization in other similar volunteer groups, and

(iii) **localized prior work-productivity experience**, which is the amount of work a volunteer would invest in a group, as they identify with it, before joining the group.

It is unusual for volunteers to have leadership experience on a group before joining it and so we do not consider the fourth kind of prior experience i.e., localized prior leadership experience.

Based on prior literature in online and offline groups, we now propose hypotheses about the primary effects of the different dimensions of prior experience.

6.2.3 Research Hypotheses

Effects of Generalized Prior Work-Productivity Experience

Prior work-productivity experience is usually positively associated with retention [BK08, PHT09, PPET10, YRZ17] and the lack of it is associated with withdrawal [AF82, Art94, GHG00]. However, prior work-productivity experience is also associated with a decrease in productivity. A majority of workers in offline work groups eventually reach a plateau in their contributions or decrease them [Stu03]. The initial motivation to produce more could be the desire to learn or grow within the organization [Kol99] and such a motivation may not exist after they have accomplished their goals [Tay09]. For instance, in university settings, faculty often shift their focus from research to administrative service work after promotion to full-professorship [RR]. In online subgroups too, prior work has found that users' motivations change as they become more engaged in the community [BF05]. The initial

motivation could be the desire to contribute what they know or to gain reputation. With an increase in contributions along with experience, they move into more caretaker roles. Accordingly, their contribution levels might change although they stick around. For instance, individuals in the GNOME project[2] increased their coordination work and decreased their technical contributions to specific projects after moving to more lateral authority roles such as board directors [DO11]. Members with such longer tenures tend to contribute less to subgroups and more to the larger community [WCRR12]. Some others who start strong, begin to decline in their contributions later due to a potential buildup of stress and exhaustion [CD93].

As we read these together, there is an interesting conflict. The more experienced someone is, the less likely they are to leave [BK08, PHT09, PPET10] but their contribution to an individual workgroup within the organization is likely to decrease with change in motivation or roles [WCRR12, DO11, Stu03], buildup of stress and exhaustion [WCRR12, CD93] or because they have contributed everything they know and accomplished their goals [Tay09]. We, therefore, believe that a change in motivation or roles that comes with experience is likely to affect future productivity.

In order to test this, we frame the following two hypotheses regarding the effect of past work productivity experience on future productivity and retention in the new online groups they join within a larger online community:

Hypothesis 1a: *Higher prior generalized work-productivity experience is associated with greater retention in a focal group.*

Hypothesis 1b: *Higher prior generalized work-productivity experience is associated with lesser work-productivity in a focal group.*

[2]GNOME is a desktop environment composed of free and open-source software that runs on Linux and most BSD derivatives, and the GNOME project refers to the community behind it which consists of all the software developers, artists, writers, translators, other contributors, and active users of GNOME.

Effects of Prior Leadership Experience

Broadly speaking, the prior literature suggests two perspectives to understanding the effects of prior leadership experience on retention and productivity within groups.

The first perspective suggests that when members gain more leadership experience, they are likely to be involved in many interactions outside of the group, and these are likely to pull them away from the focal subgroup [MPD92] affecting both their performance as well as retention. As we have seen in the examples of faculty promotion to a full professorship and individuals moving to administrative roles [DO11, RR], increase in administrative activities and leadership behavior is strongly associated with a decrease in performance.

The second perspective suggests that leaders can find it challenging to adjust to a group for various reasons. Prior work in online groups found that users' perceptions of their roles change as they become more engaged in the community [BF05]. Those who are power users and administrators see themselves as caretakers, as leaders with an established reputation, identities, organizational perspectives, mental models, and existing modes of practice. According to the Social Cognitive Theory (SCT), their self-efficacy (belief in one's capabilities and the ability to complete various actions [Ban97]) in tasks such as knowledge sharing [EL00, HJYC07, LaR09] is high. And, the more familiar they are with a domain, the higher their self-efficacy is [CLYL11]. When they join a new group, they usually also carry their established reputation, mental models, organizational perspectives and modes of practice from their previous groups [BH02, Bro15, CPVB06, GELA10, ZFBL09]. Often, existing members of a group vouch for native patterns and structures to protect native knowledge hierarchies and resist new, innovative ideas, differing practices, or past-reputation-based leadership of these experienced folks until they establish their identities independently in the new group [WTWT15]. As a result, for succeeding in their new role

in the new group, they may need to modify or suppress their perspectives, innovations, practices or role identities [Bro15, COK14]. Sometimes, their performance in the new role may not match their prior performance, their own expectations, or the expectations of the new group [DWR09, MG16]. At other times, the layers of structures, bureaucratic requirements, and oppositional rigidities in the new group may serve as barriers for their contributions and practices and leave them frustrated [MG16]. Often, they themselves tend to make judgments about the level of disparity that exists between their old and new settings, colleagues, and practices [Kor09]. Certain of their attributes or practices may be oppositional to the established knowledge and practice structures and frameworks in their new setting [GELA10] and even generate counter-productive responses among new colleagues [KMWRS13, MHE+10]. For instance, volunteers tended to get bolder and increased the likelihood of having their work rejected [BF05, HKKR09] in Wikipedia.

Thus, prior leadership experience can create barriers to fit, adaptation and integration [DWR09, Gro10]. Consequently, they often experience lesser satisfaction and high degrees of frustration and conflict in their attempts to connect with others in a way similar to their previous setting for needing support for their performance [Kor09]. In the online groups of Wikipedia and del.icio.us, researchers find that there was a dramatic shift in workload from power users to the common user [KCP+07]. We, therefore, posit that:

Hypothesis 2a: *Higher prior leadership experience is associated with lesser retention in a focal group.*

Hypothesis 2b: *Higher prior leadership experience is associated with lesser work productivity in a focal group.*

Effects of Localized Prior Work-Productivity Experience

Prior research concerning the transition of potential members from outsiders to organizational members shows that volunteers who strongly identify with a topic area or a group tend to more positively evaluate it, are willing to become more active, and exert more effort than those who don't. And, as they become more active, they tend to contribute more [RJ03, OICB89]. Also, during the evaluation period, those who see that the group fits their needs join it and remain in it longer, whereas those who don't see it as a fit leave (see [KNK16] for a review). Thus, those who join after preliminary experience with a group are likely to remain longer and contribute more. Similar research examining the hypotheses concerning the effects of college internships on individuals shows a strong support for future employment with the organization for individuals with internships [CB04, Tay88]. We therefore hypothesize:

Hypothesis 3a: *Higher localized prior work-productivity experience is associated with greater retention in the focal group.*

Hypothesis 3b: *Higher localized prior work-productivity experience is associated with greater work productivity in the focal group.*

It is unusual for volunteers to have leadership experience on a group before joining it and so we do not consider the fourth kind of prior experience i.e., localized prior leadership experience.

6.3 Methods

6.3.1 Study Platform

We study membership and editing contributions in Wikipedia through WikiProjects[3]. Wikipedia is best known for its *articles* – community-edited pages devoted to specific topics and collectively forming an encyclopedia – but it also has other pages devoted to collaboration (*talk pages* and *project pages*), to people (*editor pages*), and to policies and guidelines. Individual units of contribution are called *edits*, and such edits can be made on any of types of page. Any internet user can contribute content to Wikipedia's pages and is called an *editor*.

WikiProjects are subgroups within Wikipedia where editors come together to improve Wikipedia's coverage of a particular topic. Usually, this is done by organizing a group of related articles under one heading. A typical organization effort might include gathering all pages related to a particular topic under one heading, expanding the content of these articles, aligning articles to the same style of writing, and ensuring the articles meet certain quality standards. A typical main page of a project called the *project page* includes a brief description of the project and its scope, a list of members volunteering to contribute to the project, the list of tasks to be done, and guidelines and policies adhering to which members should work toward content production. Discussions regarding project maintenance and resolution of issues within the broader scope of the project are done in dedicated pages called *project talk pages*.

We choose WikiProjects as our research platform for three reasons. First, prior work identified WikiProjects as an example of Ostrom's *nested organizational structures* with clear goals [FLB09]. Second, there is rich historical data available about editor activities in

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

Wikipedia as a whole as well as various WikiProjects which helps us explore the concept of prior experience and design various metrics around it. Third, because WikiProjects span a large topical scope, we feel conclusions drawn from WikiProjects are more likely to be generalizable than those drawn from narrower communities such as health and technical forums.

6.3.2 Dataset

We use the English Wikipedia data dump of June 2, 2015, downloaded from the site[4] hosted by the Wikimedia Foundation. The dump data contains the complete revision history of all the pages in English Wikipedia. We use an open source Python package[5] to pre-process the dump files and extract the revision information stored in the HTML format. To construct the WikiProjects for our analysis, we parse the project templates on articles' talk pages which included the information about which WikiProjects an article belongs to. We include articles that belonged to multiple WikiProjects in all those WikiProjects. This resulted in an initial set of 1,949 WikiProjects. From these, we exclude projects that never grew to more than three members (which is the minimum size of a group) as we want to understand this in a collaborative context. Further, we exclude projects that are not related to specific topics such as *WikiProject: Articles for creation*. This resulted in a final dataset of 1,054 WikiProjects.

Many editors edit the pages without being aware of any projects. So, it would not make sense to look at edits to any page in the scope of the project randomly. Also, we want to explore the notion of pre-joining contributions for which we want to explicitly identify volunteer membership in the groups. Two approaches to identifying volunteer membership

[4] https://dumps.wikimedia.org/enwiki/20150602/

[5] https://github.com/earwig/mwparserfromhell

in projects are common in prior literature: declared membership, based on voluntary sign-up on the project page, and participatory membership, where an editor is considered to have joined a project when they made their first edit to either the project page or the project talk page. Morgan et al. compared the two approaches and found no significant difference [MGMZ13]. In this work, we choose the participatory approach. This yielded a total of 88,427 members of the projects in our sample (excluding the bots) who contributed a total of 44,135,006 edits over 14 years.

6.3.3 Operationalization

Definition of Joining: In this work, we operationalize *joining* as the first explicit *project or project talk-page edit*. This definition is not original to our work; it is used in prior work by others including [UD10, YRZ17, ZKK12b].

Independent Variables

Generalized Prior Work-productivity Experience: We count the total number of edits an editor made on the article (and the corresponding talk) pages in Wikipedia before joining the focal project except edits on the focal project as their generalized prior work-productivity experience as these represent efforts in individual article content production.

Prior Leadership Experience: We count the total number of edits an editor made on the project pages and the corresponding talk pages in Wikipedia before joining the focal project as their prior leadership experience as these represent organizational behavior at the project level.

Localized Prior Work-productivity Experience: We count the total number of edits an editor made on the main article pages and the corresponding talk pages of articles within

the scope of the project before joining it as the localized prior work-productivity experience as these represent efforts on content production for individual articles for a specific project.

The explicit joining action indicates the editor's first point of awareness of a larger community of members and of a collection of pages beyond the page (or pages) they are editing. This is the point where they *begin* documented project-level collaborations and *begin* exhibiting different behaviors with group members compared to non-group members [MGMZ13]. Prior research shows that those that explicitly join groups share a strong sense of group identity [RKK07], establish group norms and common repertoires [LW91], may exhibit in-group favoritism [Duc05, WOI98], which non-members may not. We find that even employees who join a company, despite interning many times, are considered *new* and go through new-employee training. For all these reasons, we consider those with localized experience also as *newcomers*.

Dependent Variables

Early identification of group failures can help community moderators intervene in a timely manner and shape the group for success [KNK16]. And prior work that studied WikiProjects longitudinally collected project-related measures for quarters i.e., 90-day periods, as a quarter captures a regular editing cycle on Wikipedia [QSTC14, WCRR12]. Since we are interested in studying the effects of prior experience on early retention and early productivity, we measure our response variables at the end of the first quarter for all the volunteers.

Early Retention: We measure this as a binary variable. Consistent with prior work [YRZ17], we regard an editor as having withdrawn from a project if they have not made any edits for a continuous six-month period at the end of the first quarter in any of the article, the article talk pages, the project or the project talk pages.

Early Productivity: We measure early productivity in terms of the number of edits made

Descriptive Statistics						
Variables	Min	25%ile	50%ile	Mean	75%ile	Max
1. Project Scope	0	1016	5179	37455	18119	1143441
2. Project Size	0	65	194	482	517	5248
3. Project Age	0	22	47	52	77	167
4. Editor Tenure	0	3	16	26	39	165
5. Interest Match	0	0.1	0.1	0.1	0.2	1.0
6. No. of Simul Projects	0	1	5	39	22	1251
7. Prior Gen. Work Exp.	0	97	1664	17948	12418	1285322
8. Prior Leadership Exp.	0	0	6	209	94	23451
9. Prior Loc. Work Exp.	0	0	6	128	50	134810

Table 6.1: Descriptive Statistics of Variables.

Correlations									
Variables	1	2	3	4	5	6	7	8	9
1. Project Scope	1.00								
2. Project Size	0.41	1.00							
3. Project Age	0.16	0.52	1.00						
4. Editor Tenure	-0.03	0.05	0.36	1.00					
5. Interest Match	0.01	-0.04	-0.04	0.01	1.00				
6. No. of Simul Projects	-0.04	-0.09	0.01	0.18	-0.03	1.00			
7. Prior Gen. Work Exp.	-0.03	-0.04	0.11	0.37	-0.01	0.29	1.00		
8. Prior Leadership Exp.	-0.04	-0.06	0.04	0.25	-0.02	0.38	0.35	1.00	
9. Prior Loc. Work Exp.	0.17	0.09	0.06	0.10	0.05	0.01	0.14	0.02	1.00

Table 6.2: Correlations of Variables.

[CRR10, KCP+07, KPK09, WCRR12, YRZ17] on all articles within the scope of the project during the first quarter after joining.

Control Variables

Prior work shows that a number of other factors are likely to influence outcomes of members' successful collaborations in WikiProjects [CRR10, WCRR12, YRZ17]. We, there-

fore, explore our three dimensions of prior experience along with all of these factors to see if prior experience measures provide an additive value over these in determining early retention and productivity of new volunteers in the focal project. We have operationalized many of these in ways consistent with prior work:

Project Scope: This is a count of the number of articles within the scope of the project [CRR10, WCRR12, YRZ17].

Project Size: This is a count of the number of editors who participated in the focal project before the focal editor joined [YRZ17].

Project Age: This is a count of the number of months from the project's creation until the focal editor joined [YRZ17]. This variable is used to control for the project maturity, which may affect the ease with which new members could integrate into and contribute to the project.

Editor Tenure: This is a count of the number of months from the registration of the editor in Wikipedia to the time they joined the focal project.

Interest Match: This measures the interest match between an editor and the focal project. Following prior work [YRZ17], we create a topic vector for the editor based on their prior edits on articles, another topic vector for the project based on the articles within the scope, and compute the cosine similarity between the two vectors.

Number of Simultaneous Projects: This is a simple count of the number of projects the editor has any edits in during the time he is a member of the focal project.

6.4 Analysis and Results

6.4.1 Analysis Strategy

We present the descriptive statistics and correlations among all our variables in Tables 6.1 and 6.2.

Tables 6.1 and 6.2 suggest that most of the variables have a heavily right-skewed distribution. We, therefore, log-transform all the above variables (except *Interest match*, which is between 0 and 1) to stabilize the variance and improve the fit of the models in which we will use them as predictors. We also standardize all of them (i.e., normalize to mean zero and unit standard deviation) for ease of comparing their relative importance (i.e., the coefficients across the predictors in the models we build). Most of the correlations between the variables are low. Nonetheless, in order to examine and remove any potential multi-collinearity between the individual predictors, we compute the VIFs (Variance Inflation Factors) for all the variables included in the model and find that removing the variables *Project Size* and *Editor Tenure* from the set of predictors achieves a set with all individual VIFs sufficiently below 5, the recommended maximum for behavioral sciences data [CCWA13] (including these two gave at least two values very close to 5). The VIFs for all predictors used in our models are shown in Table 6.3.

The standard errors are small and we have seen that the predictor variables do not change signs when we try to remove variables further from the remaining set of predictors here, indicating that this set of predictors do not pose problems of multicollinearity.

Each project can have multiple editors and an editor can belong to multiple projects. Our data, therefore, is cross-nested between WikiProjects and individual editors. We, therefore, use random-effects regression models to take care of potential correlations across observations that are nested within a level (e.g., editors nested under projects). For our first outcome

Variable	VIF
Project Scope	1.51
Project Age	1.11
Interest Match	1.12
No. of Simul Projects	3.04
Prior Gen. Work Exp.	3.16
Prior Leadership Exp.	3.68
Prior Loc. Work Exp.	1.87
Mean VIF	2.21

Table 6.3: Collinearity diagnostics on all the Independent Variables after log-transforming and standardizing.

measure, i.e., determining the early retention, we use a binary response variable that measures whether or not an individual volunteer remains in the project by the end of first quarter. For our second outcome measure, i.e., determining the early productivity, we see that our dependent variable is the total number of edits made in the first quarter after joining which is a count variable with over-dispersion (i.e., the variance is much higher than the mean). We, therefore, use a negative binomial regression model to handle this scenario.

We control for the effect of variables examined in prior literature (namely, Project Scope, Project Age, Interest Match, and Number of Simultaneous Projects) while examining the additive effects of the three Prior Experience variables (namely, Generalized Prior Work-productivity Experience, Prior Leadership Experience, and Localized Prior Work-productivity Experience).

6.4.2 Results

Out of our initial dataset, we use a sample of 30,000 editors along with all their edits in all the WikiProjects they participated. To examine whether the prior experience variables have additive value over and above the variables we are controlling for, we build three sep-

	Models for Early Retention					
Variables	**Model I**		**Model II**		**Model III**	
	Coeff	**S.E.**	**Coeff**	**S.E.**	**Coeff**	**S.E.**
Project Scope	1.012***	0.024	0.576***	0.021	0.518***	0.02
Project Age	-0.511***	0.01	-0.447***	0.011	-0.41***	0.011
Interest Match	0.61***	0.010	0.305***	0.011	0.301***	0.011
No. of Simul Projects	0.912***	0.01	1.605***	0.019	1.706***	0.02
Prior Gen. Work Exp.			0.449***	0.016	0.091***	0.029
Prior Leadership Exp.			-1.381***	0.02	-1.169***	0.02
Prior Loc. Work Exp			1.046***	0.014	1.142***	0.015
Gen. Work × Leadership					-0.294***	0.013
Gen. Work × Loc. Work					-0.390***	0.019
Leadership × Loc. Work					0.378***	0.017
AIC	100434		83859		82936	
χ^2			16581.42***		929.25***	

Table 6.4: Results of the effects of prior experience on Early Retention (Models I through III). We use the following notation in tables for p-value significance ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, ns: $p > 0.05$

arate models each for retention and productivity: base models (I and IV) containing just the control variables, the models (II and V) containing the control variables as well as the prior experience variables, and the models (III and VI) that also include potential 2-way interactions among the prior experience variables. We do not include 3-variable and higher interactions for they not only make interpretation considerably more complex but also do not significantly improve our understanding of interactions between the variables. The results of the random effects logistic regression for retention (Models I, II, and III) at the end of the first quarter and those of the random effects negative binomial regression for productivity during the first quarter are shown in Table 3.

	Models for Early Productivity					
Variables	**Model IV**		**Model V**		**Model VI**	
	Coeff	**S.E.**	**Coeff**	**S.E.**	**Coeff**	**S.E.**
Project Scope	1.659***	0.045	0.805***	0.034	0.774***	0.034
Project Age	-0.297***	0.009	-0.173***	0.008	-0.146***	0.008
Interest Match	1.103***	0.013	0.61***	0.01	0.583***	0.01
No. of Simul Projects	0.246***	0.010	0.637***	0.013	0.686***	0.013
Prior Gen. Work Exp.			-0.276***	0.013	-0.431***	0.015
Prior Leadership Exp.			-0.808***	0.013	-0.771***	0.014
Prior Loc. Work Exp			1.496***	0.009	1.584***	0.010
Gen. Work × Leadership					-0.159***	0.011
Gen. Work × Loc. Work					-0.26***	0.011
Leadership × Loc. Work					0.274***	0.001
AIC	579852		548927		548091	
χ^2			30931.46***		841.93***	

Table 6.5: Results of the effects of prior experience on Early Productivity (Models IV through VI). We use the following notation in tables for p-value significance ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, ns: $p > 0.05$

Choosing the best model

The model's fitness to the data can be determined either by comparing actual values with the predicted values using the model or by comparing the model with other competing models. Comparison with competing models seems more appropriate in this context since the outcomes are over-dispersed counts. We use the Akaike Information Criterion (AIC) to evaluate the goodness of fit for each of these models. AIC rewards goodness of fit of the model to the data while penalizing complexity (i.e, more number of predictors). AICs are always compared with each other and individual AIC magnitudes are not interpreted by themselves as they are affected greatly by sample size. In general, the smaller the AIC among a set of candidate competing models, the better the model. Using the AIC, we note

that the models including both the prior experience variables and their interactions (Model III for early retention and Model VI for early productivity) are better. We find also that the difference in log likelihoods of the base model with Model II is statistically significant ($p < 0.001$) with $\chi^2 = 16581.42$ and of Model II with Model III is statistically significant ($p < 0.001$) with $\chi^2 = 929.25$, indicating again that Model III is better than Model II and Model I. Similarly, we find Model VI is better than Models V and IV. We also find that prior experience variables have coefficients that are comparable in magnitude to the control variables.

We interpret Models III and VI to understand the impact of various kinds of variables. Note that the above variables are log-transformed (with e as the base of the logarithm) and normalized to mean 0 and a standard deviation of 1. This makes it easier to understand the impact of different predictors with respect to each other. First, we note that all the predictors are significant and the effects of control variables are largely consistent with prior work. We, therefore, focus on interpreting only the variables of interest (i.e., the prior experience variables) on the linear scale to understand the actual impact of prior experience.

Overall Effects of Prior Experience Variables

Based on Model III, we find that holding all the other variables constant, an e-fold (i.e., roughly 2.7 times) increase in generalized prior work-productivity experience (in terms of number of prior article and article talk page edits) is roughly associated with an overall 3% increase in the odds of retention, whereas an e-fold increase in prior leadership experience (in terms of number of prior project and project talk page edits) is roughly associated with an overall 62% decrease in the odds of retention, and an e-fold increase in localized prior work-productivity experience (in terms of pre-joining article edits to the focal project) is roughly associated with an overall 70% increase in the odds of retention. And based on Model VI,

we find that holding all the other variables constant, an *e*-fold increase in generalized prior work-productivity experience is roughly associated with a 17% decrease in productivity (i.e., the expected count of number of edits made) during the first quarter, an *e*-fold increase in prior leadership experience is associated with a 37% decrease in the expected count of number of edits and an *e*-fold increase in localized prior work-productivity experience is associated with a 108% increase in the expected count of number of edits [6].

The above overall percentages include the effects of interactions within them. In order to tease out the effects of individual interactions, we plot the interaction plots for the two response variables for low and high values of various prior experience variables. Below, we present and discuss a couple that are interesting.

Interaction Effects of Prior Experience

Figure 6.1 suggests that the retention is the highest when prior leadership experience is low and generalized prior work-productivity is high. A potential scenario for this could be when the volunteers have not yet moved into leadership roles and are continuing to enjoy non-administrative level contributions to individual projects. We see that the retention is the lowest when both generalized prior work-productivity and prior leadership experience are high. Potential reasons for these could be a burnout effect or a challenge of adjustment with the group.

Figure reffig:figure62 suggests that the early productivity in the focal project is the lowest when both prior leadership and generalized prior work-productivity experiences are high. indicating a case of potential burnout effect or an adjustment issue. We see that the early productivity is the highest when both generalized prior work-productivity and prior

[6] Since the minimum values showed in Table 6.1 are 0, this *e*-fold increase is applicable only from the point where they gain a non-zero experience.

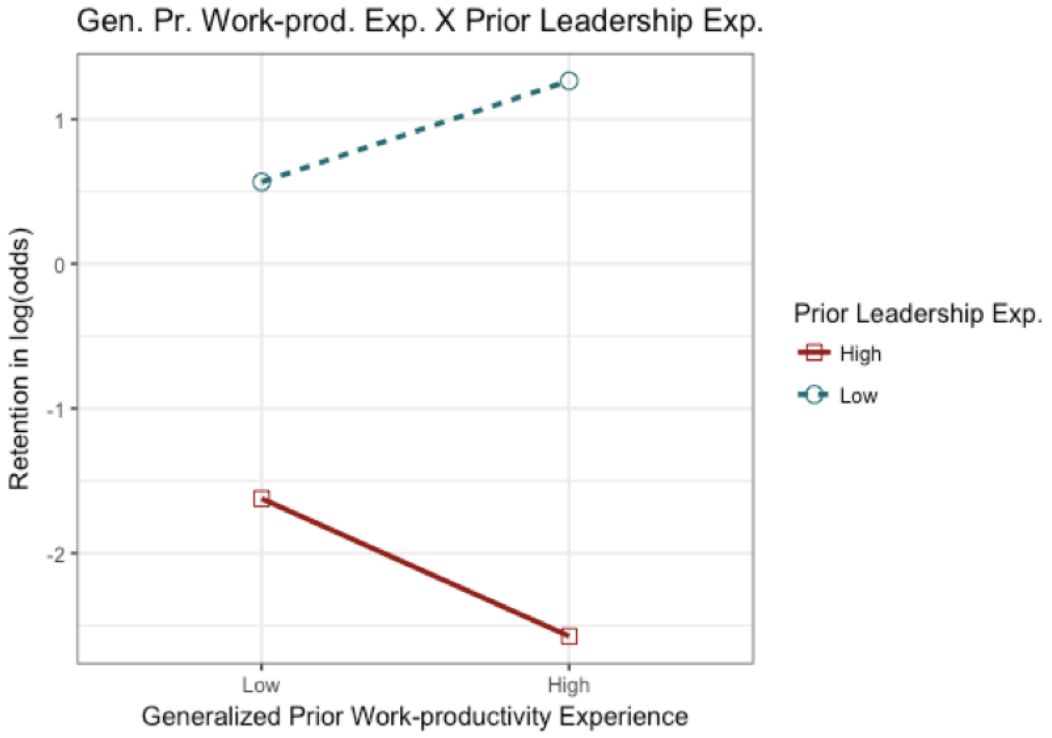

Figure 6.1: Figure showing the interactions between low and high values of prior leadership experience and prior generalized work-productivity experience for the effects of Retention.

leadership experiences are low, indicating that editors with low workloads as well as lower administrative overhead are likely to be more productive.

Figures 6.3 and 6.4 show how generalized prior work productivity experience interacts with localized prior work-productivity experience. We see that the retention and the productivity are the lowest when both generalized prior work-productivity and localized prior work-productivity experiences are high. One potential scenario for high localized work-productivity is when the volunteers have already contributed everything they know and contributing more would require much more research and effort. A high generalized prior work-productivity experience might be indicating a potential burnout effect due to stress or exhaustion - the combination of which is possibly associated with the low retention and

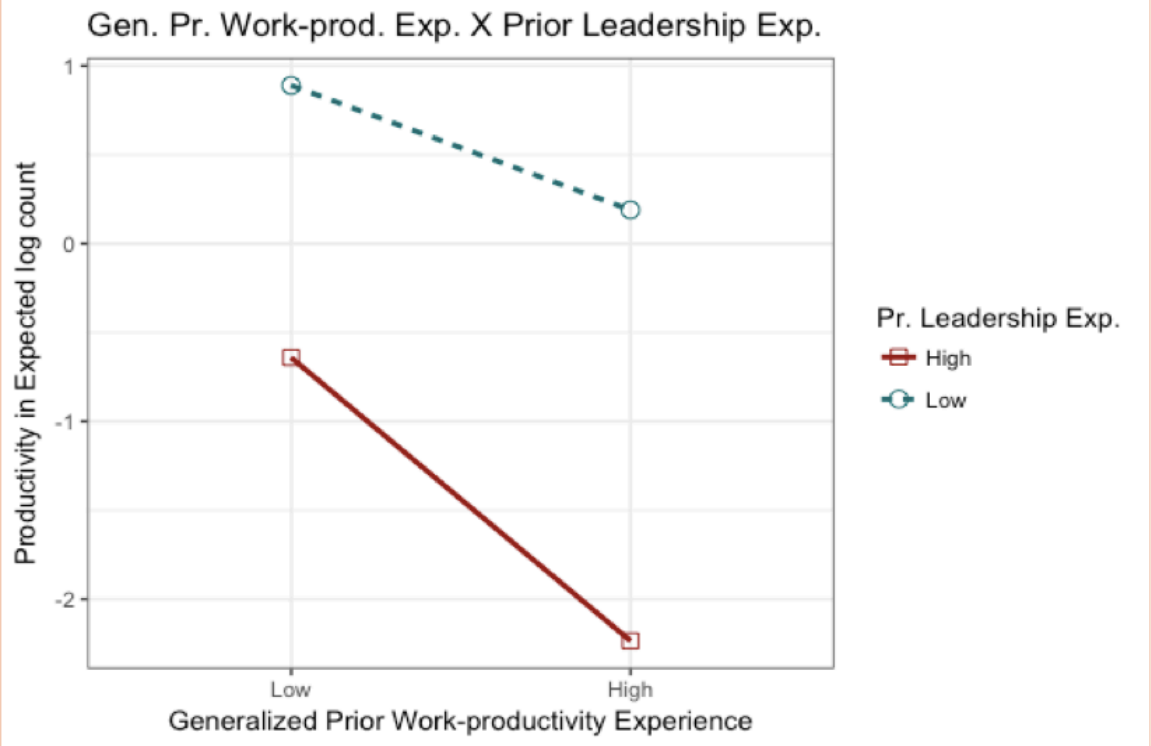

Figure 6.2: Figure showing the interactions between low and high values of prior leadership experience and prior generalized work-productivity experience for the effects of Productivity.

productivity in the focal project. On the other hand, we see that the retention and productivity are the highest when generalized prior work-productivity experience is low - an example of this is a situation where a potential burnout has not yet happened and the high level of attachment associated with the high level of localized prior work-productivity experience is potentially responsible for high retention and high productivity in the focal project.

Figures 6.5 and 6.6 show the interactions between prior leadership experience and localized prior work-productivity experience. We find that the retention and productivity are the highest when prior leadership experience is low and localized prior work-productivity experience is high One scenario where the involvement of volunteers in administrative tasks is less and their interest in a particular project is indicated by the high level of localized

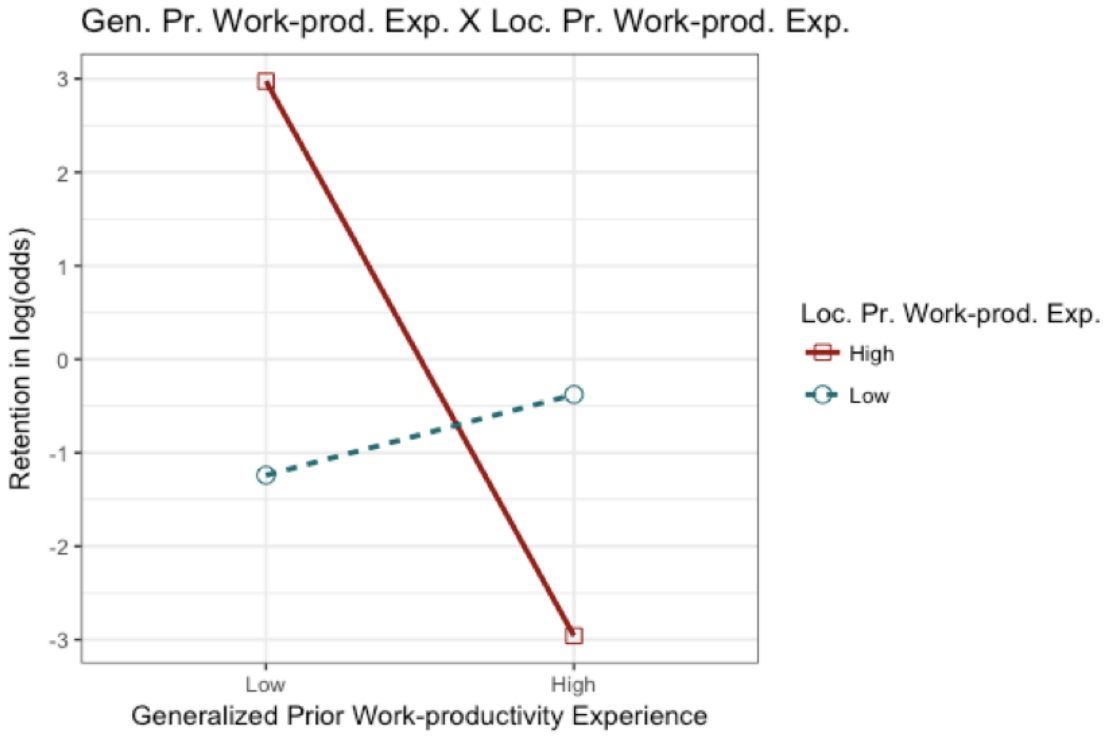

Figure 6.3: Figure showing the interactions between low and high values of the generalized and localized prior experience variables for retention.

prior work-productivity experience. We find that the retention, as well as productivity, are low when both the experiences are high – an example of this could be a situation where the administrative responsibility at the community level pulls the editors off of individual projects combined with a feeling that they have already contributed everything they know.

6.5 Discussion

First, our findings show that generalized prior work-productivity experience is positively related to retention and negatively related to the productivity confirming our hypotheses **1a** and **1b**, prior leadership experience is negatively related to both the retention as well as the

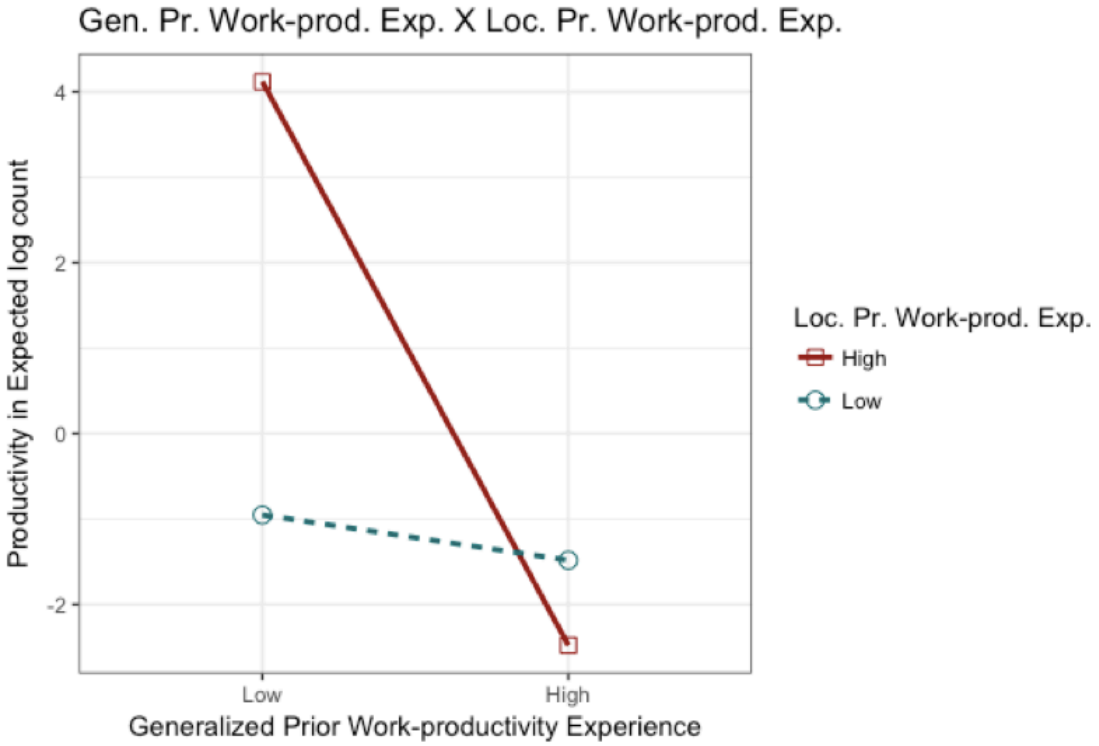

Figure 6.4: Figure showing the interactions between low and high values of the generalized and localized prior experience variables for productivity.

productivity confirming our hypotheses **2a** and **2b**, and localized prior work-productivity experience is positively associated with both the retention and productivity confirming our hypotheses **3a** and **3b**.

Second, while prior work shows only a positive relationship between metrics based on prior experience and future productivity and administrative behaviors [BK08, PHT09, PPET10, YRZ17], our work confirms that the relationship is, indeed, much more complicated, with some types showing positive effects and others negative ones. Even with the caveat that we are talking about productivity in its simplest form i.e., edit count, our work shows that prior experience, in general, is worse for productivity although better for retention.

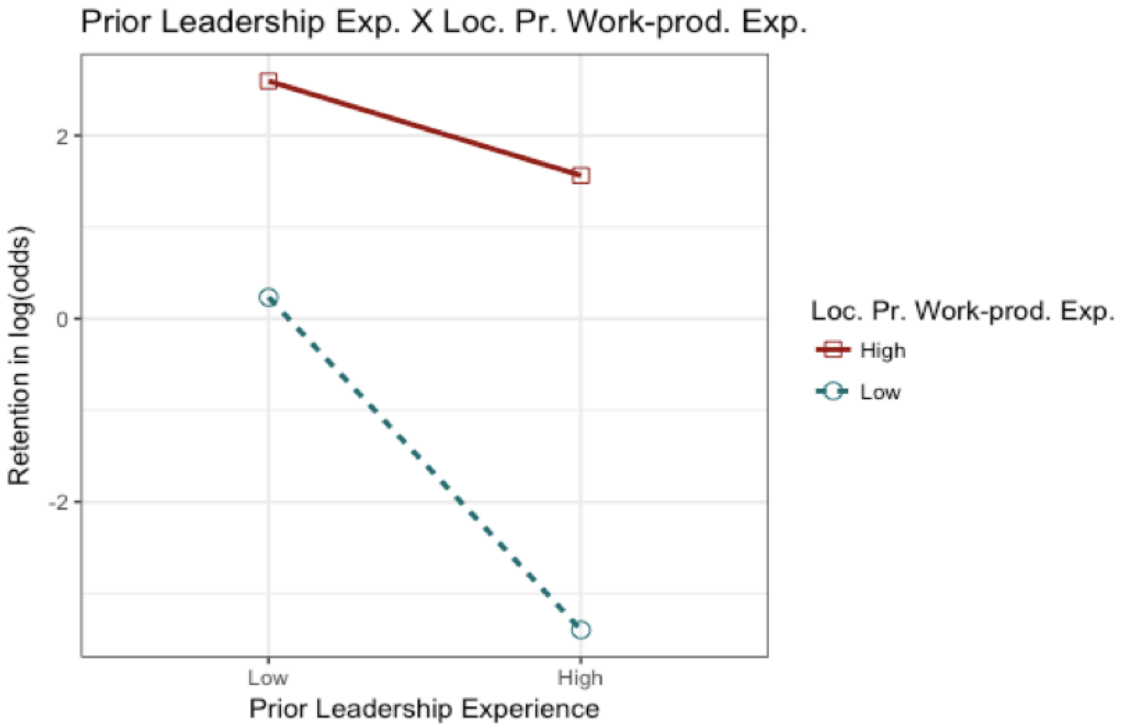

Figure 6.5: Figure showing the interactions between low and high values of the leadership and work-productivity prior experience variables for retention.

Third, it is interesting to observe some of these interactions. Consider the interactions between localized prior work-productivity experience and prior leadership experience (Figures 6.5 and 6.6). These could be understood in two ways: (1) Localized prior work-productivity in a specific topic area has a huge positive effect that it dampens any of the negative effects of prior leadership experience. OR (2) The benefits of localized work-productivity get cut down, the more someone has overall prior leadership experience. However, the net effect of localized work-productivity still remains positive (see Table 3.3. Hence, *content is king, and leadership lags*. The effect of generalized prior work experience in the presence of interactions is pretty small (e.g., compare models II and III). This means that generalized prior work experience is useful and positively predictive of retention

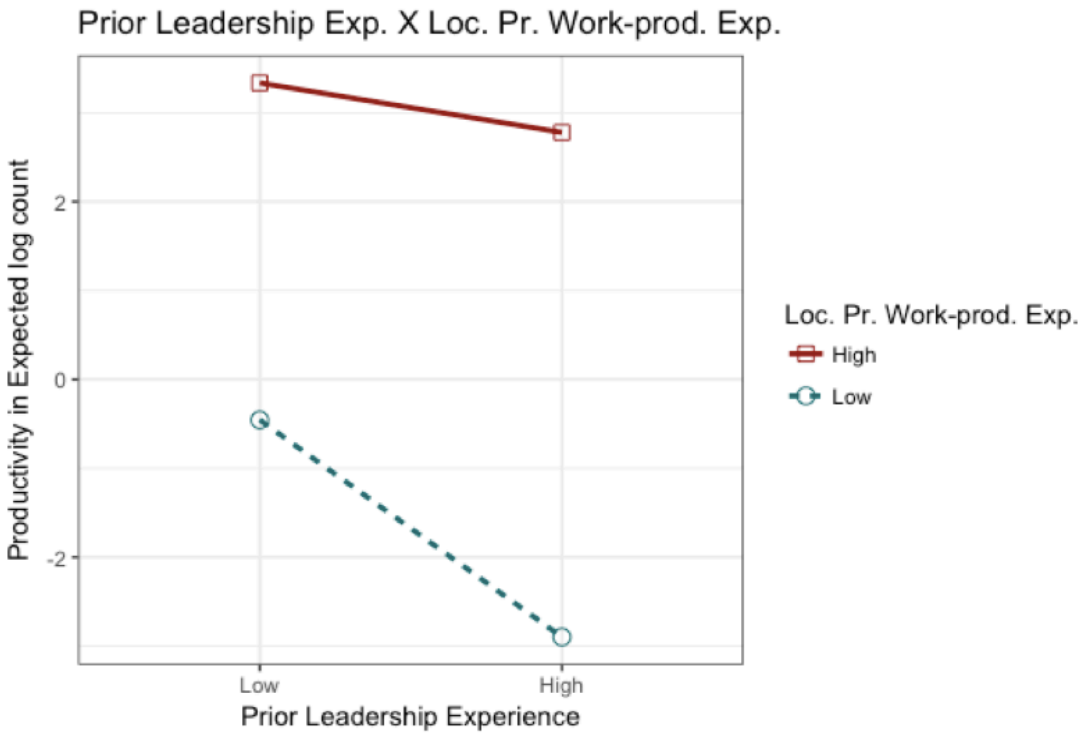

Figure 6.6: Figure showing the interactions between low and high values of the leadership and work-productivity prior experience variables for productivity.

only when the volunteers do not have localized or leadership experiences. The benefits of localized work productivity experience which are substantial get cut down significantly if you have too much overall experience (see Figures 5 and 6). Future experiments along these directions could reveal interesting insights about causal relationships.

6.5.1 Theoretical Implications

While the notion of *prior experience* has been explored before, in this work, we show for the first time, that *different kinds* of prior experience reveal and predict more interesting and nuanced effects in volunteering groups. We show, for instance, the importance of identifying and distinguishing between prior work productivity experience and prior leadership

experience, and between generalized and localized experiences. We think these concepts might generalize not just to other online peer production communities but to volunteer or organizational management more generally. Given our findings, it would be useful to study other domains (and revisit prior studies of newcomer contribution/retention) through the lens of different types of prior experience.

6.5.2 Practical Implications

The practical implication we have for WikiProjects is that we can improve the success of recruiting and retaining productive contributors. Looking only at the primary effects, the first implication is that the localized prior work-productivity experience is the most effective indicator of whether the new volunteer will become a productive and dedicated member. Our findings suggest that WikiProjects looking to recruit and retain productive workers should focus their recruitment foremost on those who have already demonstrated a commitment to the specific work or cause of the project.

However, projects that are smaller or in their initial phases may not have a lot of candidates with a demonstrated commitment to the project to recruit. In these cases, project recruiters can recruit those with generalized prior work-productivity experience who are likely to stay in the project longer but not necessarily be very productive. Recruiting experienced leaders is much trickier, as leadership experience is generally associated with low productivity and retention. When leaders are sought (e.g., to bolster leadership in a new group), WikiProject organizers may want to both verify commitment to the cause/topic and consider specific re-orientation/transition plans to help the leader better integrate and be successful in the WikiProject.

6.5.3 Generalization

The framework, metrics, and hypotheses we provide in this paper apply not just to WikiProjects, but also to other peer production groups such as OSS projects in GitHub, communities in StackExchange, groups in GoodReads or projects in OpenStreetMap. Considering prior experience can reveal more interesting and nuanced effects, one might even consider broader kinds of prior experience available in these specific volunteer groups. For instance, in the case of projects in OpenStreetMap, these results might suggest using their prior communication patterns with project members within projects along with prior map creation and leadership activities to predict new members' retention and productivity. We encourage future research in these communities.

We believe similar effects of prior experience may appear in offline groups. Consider a mosque looking to start a homeless shelter. As it recruits a collection of volunteers to staff the shelter, how much should it draw on top volunteers in other efforts (interfaith outreach, study sessions), how much on leaders of other efforts, and how much on volunteers who have worked in other homeless shelters and similar projects, even outside the mosque? Of course, much additional research is needed to validate the generalizability of our findings.

6.6 Conclusion

In this study, we explore the effects of prior experience of new volunteers on their early retention and productivity in the group they join with the understanding that early identification of group failures can help community moderators intervene in a timely manner and craft the group for success. We found that certain kinds of prior experience have positive effects on newcomer retention and productivity whereas other kinds of prior experience have negative effects.

Specifically, we carried out the study on a sample of 30,000 new editors to 1,054 WikiProjects, which are groups dedicated to building content around specific topic areas. This platform allowed us to measure prior experience in multiple dimensions and potential interactions between them which could generalize to other communities with similar structures. Also, WikiProjects have been ideal for such an exploration owing to their well-established and shared structures, shared membership and publicly available historical data about each volunteer. Through our analysis, we found that (i) generalized prior work-productivity experience (measured by overall prior article and article talk page edits) is positively associated with retention, but negatively associated with productivity within the focal group, (ii) prior leadership experience (measured by overall prior project and project talk page edits) is negatively associated with both retention and productivity within the focal group, and (iii) localized prior work-productivity experience (measured by pre-joining article edits on a focal group) is positively associated with both retention and productivity within the focal group.

6.6.1 Limitations, Future Work and Potential Impact

In this study, we made a preliminary investigation of the effects of prior experience on early retention and productivity within the subgroups of a larger community in order to understand if examining prior experience has any value and we found that considering the prior experience of a member adds value over other group-level metrics such as composition and structure, and, even within prior experience, some kinds of prior experience have positive effects on group outcomes whereas other kinds have negative effects. However, prior experience might also vary with factors such as project age and the number of simultaneous projects. Future work should look at these variations in order to gain a deeper insight into the effect of prior experience.

Our data analysis, although providing us with key insights into the interactions of various dimensions of prior experience and their effects on retention and productivity, provides only limited support for understanding *why* the association between prior experience and outcomes exist and in what ways they are causal. In addition, we see interactions that at this point we don't have data to explain. Further qualitative studies could reveal more insights into this which we leave for future work to explore. Based on our findings, future studies could also run field experiments with varying on-boarding processes for volunteers with different kinds and/or levels of experience. For instance, in groups with no opportunities for pre-joining contributions, volunteers in one condition might require going through a probationary period where they achieve a certain level of contributions before they become members and be compared with volunteers in another condition where there is no such requirement and both may be measured on their retention and productivity.

We do not have information regarding the amount of workload of these editors in their personal lives or in other online communities with similar skillset and we believe high contributors online are also potentially very knowledgeable in their respective fields which might also affect their performance once they undertake too many activities online. Again, conducting qualitative studies might help us gather this information and further insights into the interplay. We leave this also for future research to explore.

Consistent with much of prior work, our study used productivity and retention as measures of group success. Future research could extend this by incorporating more nuanced measures such as the quality of the artifacts produced or rate or amount of progress toward group/community goals.

In this study, we examine the phenomenon of near transfer, i.e., how prior experience is associated with group outcomes in case of groups having similar structures. However, knowledge, usability experience, and human capital may be easier to transfer across groups

with similar structures than they would be across groups with dissimilar structures within the same community or within different communities. Future work should consider extending these findings to more heterogeneous environments with different structures and affordances.

The *potential impact* of this work lies in three areas. First, we have demonstrated the importance of considering diverse types of prior experience in predicting the longevity and productivity of experienced newcomers. This result makes a theoretical contribution to our understanding of newcomer behaviors in online groups. Second, this work is directly applicable to WikiProjects where it can be used to identify individuals to recruit and to plan pre-joining activities to test and/or build commitment to a project. Third, while our results have not yet been tested outside Wikipedia, we provide a framework for extending this research into new domains, including generalized hypotheses and research methods that can be used for systematic research and exploration.

Chapter 7

Conclusion and Future Work

7.1 Introduction

We are living in the age where we use online communities for most tasks we accomplish from social and professional networking, to browsing and purchasing of products, to getting our doubts clarified and in turn, sharing the knowledge we have with others. This thesis began with a list of challenges every online community faces and the need for addressing these challenges for the community to be successful in accomplishing its goals. Specifically, this dissertation has listed key challenges with respect to starting a new online community, encouraging contribution from members as well as non-members, encouraging commitment of existing members who display low participation levels, regulating member behavior on the site for quality content and healthy interactions, and dealing with newcomers. This dissertation then focused on the specific challenge of dealing with newcomers and explored this challenge in greater detail listing six basic problems related to newcomers that every community has to solve in order to be successful. Among these, this dissertation focused on the challenges of newcomer retention and productivity and explored these in greater detail in the context of online peer-production communities.

The work presented in this dissertation began with studying newcomer retention patterns on various online communities. Based on prior research showing that on an average 60% of newcomers do not return for a second session in most online peer-production sites and the

average number of non-returning users rises to about 80% in the case of Android apps, and that early contribution level is predictive of future contribution level and which newcomer is likely to be successful going forward [PHT09], in this work, we decided to learn about the specific factors contributing to newcomer retention and productivity and build models that can predict these outcomes in the context of online peer-production communities.

7.2 Key Contributions

First-Session Activity Diversity substantially affects Newcomer Retention

We studied prior literature underscoring the importance of this problem in both online and offline settings and documented key research findings relevant to our research in Chapter 2 of this thesis. We then presented our research in our first piece of work studying the effects of early activity in an online site on newcomer retention in Chapter 3. Through this first piece of work, we made three key contributions to the theory of newcomer modeling and success in online sites. First, we showed that the diversity of early activity (i.e., breadth of features explored by a new user) is an effective predictor of newcomer longevity (even in the presence of amount of activity) and that, a small increase in this diversity has the same increase in retention as a large increase in total early activity. Second, we showed how one could use participatory design techniques such as card sorting in understanding site architecture and build metrics from it. Third, we offer two metrics based on early activity - ASCORE and DSCORE – that can be used to measure the amount of activity and breadth of activity of users – that are generalizable to sites with any architecture.

Personality affects Newcomer Activity, Preferences, and Retention

Following this, we explored the effect of more fundamental factors at the user level such as personality in predicting newcomer longevity, level of activity, and preferences in an online movie recommender site by running a large scale survey gathering user personality early on, and showed that a new user's personality could be effectively used to predict all of these. Our work is the first work showing relationships between newcomer behavior and personality. This piece of research provides an easy solution to recommender systems suffering from the cold-start problem for recommending items as well as experiences early on when they lack any user data.

Community Interactions affect Newcomer Retention

After this, we explored the effect of community level factors in determining a newcomer's retention using individual codebases on GitHub and showed how factors such as language of development and willingness to accept new contributions by existing members are likely to affect newcomer contributions and retention.

Prior Experience Affects Newcomer Retention and Productivity

Because a lot of online users have experience using similar tools or working with similar environments elsewhere in the online world, we wanted to study how their prior experience might affect their retention and performance in the context of a similar new online community. In this work, we showed for the first time, that prior experience is not be evaluated as a whole but that *different kinds* of prior experience reveal and predict more interesting and nuanced effects in online communities. We showed, for instance, the importance of identifying and distinguishing between prior work and leadership experiences, and between

work and internship experiences. We think these concepts might generalize not just to other online peer production communities but to volunteer or organizational management more generally.

7.3 Lessons Learned in Modeling User Behavior

After weeks, months, or even years of data collection, one gets the data needed for performing analysis. Often, the data analysis stage is equally or even more laborious than the data collection stage. Many critical decisions need to be made while analyzing the data such as the type of statistical technique to be used, the threshold to pick to obtain confidence intervals, the significance and interpretation of results in context, and so on. Incorrect choice of methods or inappropriate interpretation of the results can lead one astray and waste high-quality data collected arduously. Before I conclude this dissertation, I want to document some of the key steps, challenges, and gotchas that I have learned, used, and suggest for future researchers trying to model user behavioral data.

7.3.1 Data Preprocessing

It is important to pre-process the data correctly and there are three reasons for this. First, improper data results in erroneous, and often noisy outcomes. Second, the collected data might be in a form that is too primitive and might need higher level coding of themes for more meaningful interpretations. Third, sometimes the specific statistical technique (e.g., arguments to methods, or parameters) or software used might require organization of the data into specific formats or pre-defined layouts, and this might require some knowledge of the language used in these tools and techniques for representing various statistical parameters so that one can input them into the software and algorithms correctly.

Data can be improper due to a variety of reasons such as data going missing, inconsistent formatting, network errors, accidental typos, and so on. Often, plotting simple plots such as histograms can reveal most of the inconsistencies in the data. There might also be data in more than one format (some of it numeric, and some of it in characters. For instance, one person's age might be recorded as 19, and another's as 'nineteen'. Such inconsistencies are often, results of badly designed user interfaces, and they also need to be processed in order to use a field such as 'age' in a meaningful way.

7.3.2 Obtaining Descriptive Statistics

After preprocessing and cleaning the data, it is important to run a number of basic statistical tests to understand the nature of the dataset such as the range in which the data falls, the means, medians, standard deviations, and so forth.

7.3.3 Assumptions

For applying parametric tests, it is often assumed that the errors of the data points are independent of each other, identically distributed, and normally distributed. If these are not met, one ends up making incorrect inferences. It is important, therefore, to either transform the data using suitable transformations, or apply non-parametric techniques where such requirements are relaxed.

7.3.4 Building Models

Correlation analysis allows the study of only two variables at a time, and so, to deal with data at scale and investigate the relationship between one dependent variable and multiple independent variables (which is often the case in user behavioral data), regression analyses are used for model construction and predictions. Depending on the specific research ob-

jective, an appropriate regression procedure needs to be adopted. In the simplest case of finding the relationship between one dependent variable and many independent variables, a simple linear regression is used. Using this approach, one can understand the percentage of variance in the dependent variable explained by the independent variables together. This procedure is useful but insufficient if one is interested in the impact of each individual independent variable. In the latter scenario, a hierarchical regression procedure is more appropriate. It is important to note that the order of entry of the independent variables is determined by the pre-defined theoretical model. The independent variables entered first into the equation in the model fall into two classes. The first class of variables are those that are considered important from prior literature or observation. The second class of variables (called the covariates) are usually of no interest to us, but significantly impact the dependent variable. In this case, it is important to exclude the variable's impact on the dependent variable before one studies the variables that one is interested in. To put it simply, entering the covariates first allows one to remove the variances in the dependent variable that can be explained by the covariates, making it simple to identify significant relationships for the variances in which one is interested.

A mixed model is similar in many ways to a linear regression model. It estimates the effects of one or more independent variables on a dependent variable. The output of a mixed model will give you a list of explanatory values, estimates and confidence intervals of their effect sizes, p-values for each effect, and at least one measure of how well the model fits. A mixed model is appropriate when there is dependence among data points. For instance, in the Wikipedia data we handled in the previous chapter, each editor could potentially contribute to multiple projects, and so, when we collect editorial data on multiple WikiProjects, each user could potentially appear under multiple projects in the data, and the data points are, therefore, not independent.

Before one proceeds with mixed models, one must also think about the structure of the random effects. The random effects are said to be nested when each user produced a data point, and no two users produced the same data point. In the previous example, each user is nested under WikiProjects. Then one must find a probability distribution that best fits the data. Whereas the Negative Binomial and Gamma distributions can only handle positive numbers, the Poisson distribution can only handle non-negative whole numbers. The Binomial and Poisson distributions are different from the others because they are discrete rather than continuous, which means they quantify distinct, countable events or the probability of these events. Following finding a probability distribution, one can fit the model.

If the data is normally distributed, a simple linear mixed model (LMM) would suffice. It is important to specify whether the mixed model will estimate the parameters using maximum likelihood or restricted maximum likelihood. If the random effects are nested, or there is just a single random effect, and if the data are balanced (i.e., similar sample sizes in each factor group) REML must be set to FALSE and maximum likelihood can be used. If your random effects are crossed, the REML argument usually defaults to TRUE and nothing needs to be set here.

One complication one might face when fitting an LMM is a ”failure to converge” error. I have had this issue multiple times in my projects on Personality as well as Prior Experience. This usually means that the model has too many factors with insufficient sample size, and cannot be fit. One strategy here is to drop the fixed effects and random effects from the model one at a time, and compare to see which fits the best. Running an ANOVA might tell about important differences between the models. Or one might use measures such as AIC to compare models.

If your data are not normally distributed, the REML and maximum likelihood methods for estimating the effect sizes in the model make assumptions of normality that do not apply

to the data, and so, one must use a different method for parameter estimation. There are many alternative estimation methods and one must decide which one suits better.

In conclusion, one cannot really know which analyses are right for one's data unless one gets acquainted with the data sufficiently. Following some of these guidelines can significantly save time and improve one's ability to obtain and interpret outcomes.

7.4 Future Work

We now conclude this dissertation by offering some directions for future research.

First, as much as we strived for exploring similar factors on various peer production sites, we found that each online site we explored had entirely different structures and challenges surrounding them and so we were not able to consistently study the same metrics on various online communities. In a community that has all these various factors available, a natural future piece of research would be to put them all together and understand the effects of each of these factors in the presence of the other, and how they all, together, predict newcomer behavior. For instance, would users high in Openness display greater early activity diversity, and thus show more longevity? Would extroverts engage more in leadership activities leading to lower retention levels? Would conscientious users or users high in agreeableness be more tolerant of other newcomers' contributions? Future research should explore some of these questions.

Second, in this work, we presented several relationships between various factors. As part of immediate future work, one could run subsequent analyses and online experiments on these and other similar platforms to confirm causality and generalizability of our findings.

Third, in this dissertation, we have shown some promising results about newcomer behavior suggesting a few things we can do to improve newcomer experience in three contexts -

a recommender system, an open source portal, and a Wiki – all of which are peer-production communities. However, we see that the number, and type of online communities are growing. There are communities for education such as MOOCs (Massive Open Online Courses), for exchanging information about topics such as games, parenting, cooking, and knitting, for patients with life-threatening diseases such as Cancer and their caregivers to share and discuss their journeys, dating sites for people to bond with each other, sharing economy sites such as TaskRabbit, sites offering homestay for travelers such as Airbnb and Couchsurfing, communities for those recovering from addictions such as intherooms.com, and e-commerce sites such as Amazon.com – each presenting very unique environments and first time experiences and the type of activities one might have to be doing and the type of benefits one might derive from these platforms. And then, there are specific populations such as children, teenagers, and older adults who come with their own unique challenges. There are challenges of privacy, safety, literacy and simplicity. A systematic examination of newcomer behaviors on these different platforms could increase our understanding of newcomer behavior in general and (even specific to each different type of site) and lead us to develop better and more rewarding first time experiences that gets them to stay longer and turn into active contributors, and solve one of the key challenges of building new online communities. Future work could research on one or more of these themes.

Fourth, it is likely that context affects newcomer outcomes. For instance, how often does a new user want a certain kind of item? Do all users like personalization at all times? etc. Answering questions such as these requires addressing the challenges of eliciting context from new users and the mapping of the context of items with the context of users. This may require modeling of additional things such as interests (e.g., genres), personas (e.g., student vs parent vs husband), preferences (e.g., exercise, transport mode), locations (route places), personal attributes (fitness level, financial level, gender, socio-economic background etc.),

occupations (dentist, nanny, etc.). We leave these questions for future research to explore.

7.5 Conclusion

We began this dissertation with an exploration of key challenges faced by every online community and dived deeper into one of the challenges i.e., Newcomer Retention and Productivity in the context of online peer-production communities. Exploring three different platforms with entirely different structures - (i) MovieLens, a movie recommender system, (ii) GitHub, a social-collaboration platform for developers, and (iii) Wikipedia, a community where anyone can freely edit content and showed that factors such as early activity, personality, community interactions,and prior experience affect newcomer retention, activity levels, and preferences within online sites. We leave questions of interactions between these various factors, exploration of some of these in the presence of context as a factor, and questions of causality for future research.

Bibliography

[ABJ+06] Jaime Arguello, Brian S Butler, Elisabeth Joyce, Robert Kraut, Kimberly S Ling, Carolyn Rosé, and Xiaoqing Wang. Talk to me: foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 959–968. ACM, 2006.

[AC92] Deborah Gladstein Ancona and David F. Caldwell. Demography and design: Predictors of new product team performance. *Organization Science*, 3(3):321–341, 1992.

[AF82] Hugh J Arnold and Daniel C Feldman. A multivariate analysis of the determinants of job turnover. *Journal of applied psychology*, 67(3):350, 1982.

[AH00] E. Adar and B.A Huberman. Free riding on gnutella. *First Monday*, 5(10), 2000.

[AHWF02] Yair Amichai-Hamburger, Galit Wainapel, and Shaul Fox. " on the internet no one knows i'm an introvert": Extroversion, neuroticism, and internet interaction. *CyberPsychology & Behavior*, 5(2):125–128, 2002.

[AKR08] Dudyala Anil Kumar and V Ravi. Predicting credit card customer churn in banks using data mining. *International Journal of Data Analysis Techniques and Strategies*, 1(1):4–28, 2008.

[APS00] Lynette Armstrong, James G Phillips, and Lauren L Saling. Potential determinants of heavier internet usage. *International Journal of Human-Computer Studies*, 53(4):537–550, 2000.

[Art94] Jeffrey B Arthur. Effects of human resource systems on manufacturing performance and turnover. *Academy of Management journal*, 37(3):670–687, 1994.

[AS04] Tel Amiel and Stephanie Lee Sargent. Individual differences in internet usage motives. *Computers in Human Behavior*, 20(6):711–726, 2004.

[AVF85] R. Albanese and D.D. Van Fleet. The free-riding tendancy. *Academy of Management Review*, 10(2):244–255, 1985.

[AVR05] Luigi Anolli, Daniela Villani, and Giuseppe Riva. Personality of people using chat: An on-line research. *CyberPsychology & Behavior*, 8(1):89–95, 2005.

[Ban97] Albert Bandura. *Self-efficacy: The exercise of control*. Macmillan, 1997.

[BB14] Marco Biazzini and Benoit Baudry. "may the fork be with you": Novel metrics to analyze collaboration on github. In *Proceedings of the 5th International Workshop on Emerging Trends in Software Metrics*, WETSoM 2014, pages 37–43, New York, NY, USA, 2014. ACM.

[BBE+07] Talya N Bauer, Todd Bodner, Berrin Erdogan, Donald M Truxillo, and Jennifer S Tucker. Newcomer adjustment during organizational socialization: a meta-analytic review of antecedents, outcomes, and methods. *Journal of applied psychology*, 92(3):707, 2007.

[BBS13] Andrew Begel, Jan Bosch, and Margaret-Anne Storey. Social networking meets software development: Perspectives from github, msdn, stack exchange, and topcoder. *IEEE Software*, 30(1):52–66, 2013.

[BF05] SL Bryant and A Forte. Bruckman. a. 2005. becoming wikipedian: Transformation of participation in a collaborative online encyclopedia. *Proceedings of GROUP'05*, pages 1–10, 2005.

[BH02] Janice M. Beyer and David R. Hannah. Building on the past: Enacting established personal identities in a new work setting. *Organization Science*, 13(6):636–652, November 2002.

[BK08] Moira Burke and Robert Kraut. Mopping up: modeling wikipedia promotion decisions. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 27–36. ACM, 2008.

[BKG+12] Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. Personality and patterns of facebook usage. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 24–32. ACM, 2012.

[BKJ09] Moira Burke, Robert Kraut, and Elisabeth Joyce. Membership claims and requests: Conversation-level newcomer socialization strategies in online groups. *Small group research*, 2009.

[BKM08] Ivan Beschastnikh, Travis Kriplean, and David W McDonald. Wikipedian self-governance in action: Motivating the policy lens. In *ICWSM*, 2008.

[BLW+04] Gerard Beenen, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick, and Robert E. Kraut. Using social psychology to motivate contributions to online communities. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pages 212–221, New York, NY, USA, 2004. ACM.

[BM91] Murray R Barrick and Michael K Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26, 1991.

[BM04] Anita L Blanchard and M Lynne Markus. The experienced sense of a virtual community: Characteristics and processes. *ACM Sigmis Database*, 35(1):64–79, 2004.

[BM16] A Alexander Beaujean and Grant B Morgan. Tutorial on using regression models with count outcomes using r. *Practical Assessment, Research & Evaluation*, 21, 2016.

[BML09] Moira Burke, Cameron Marlow, and Thomas Lento. Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 945–954. ACM, 2009.

[BML10] Moira Burke, Cameron Marlow, and Thomas Lento. Social network activity and social well-being. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1909–1912. ACM, 2010.

[BP08] Sarah Butt and James G Phillips. Personality and self reported mobile phone use. *Computers in Human Behavior*, 24(2):346–360, 2008.

[Bro15] Andrew D. Brown. Identities and identity work in organizations. *International Journal of Management Reviews*, 17(1):20–40, 2015.

[BS12] Zoheb H Borbora and Jaideep Srivastava. User behavior modelling approach for churn prediction in online games. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 51–60. IEEE, 2012.

[But99] Brian S. Butler. When is a group not a group: An empirical examination of metaphors for online social structure. In *Carnegie Mellon University*, 1999.

[But01] Brian S Butler. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information systems research*, 12(4):346–362, 2001.

[BVdP05] Wouter Buckinx and Dirk Van den Poel. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual fmcg retail setting. *European Journal of Operational Research*, 164(1):252–268, 2005.

[BVdP07] Jonathan Burez and Dirk Van den Poel. Crm at a pay-tv company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2):277–288, 2007.

[BVDZLD07] A.B. Bakker, K.I. Van Der Zee, K.A. Lewig, and M.F. Dollard. The relationship between the big five personality factors and burnout: A study among volunteer counselors. *The Journal of Social Psychology*, 146(1):31–50, 2007.

[CAKL10] Boreum Choi, Kira Alexander, Robert E Kraut, and John 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.

[CB04] Gerard Callanan and Cynthia Benzing. Assessing the role of internships in the career-oriented employment of graduating college students. *Education+ Training*, 46(2):82–89, 2004.

[CCWA13] Jacob Cohen, Patricia Cohen, Stephen G West, and Leona S Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, 2013.

[CD93] C.L. Cordes and T.W. Dougherty. A review and an integration of research on job burnout. *Academy of Management Review*, 18(4):621–656, 1993.

[CD10] John P Charlton and Ian DW Danforth. Validating the distinction between computer addiction and engagement: online game playing and personality. *Behaviour & Information Technology*, 29(6):601–613, 2010.

[CFG13] Luis V Casaló, Carlos Flavián, and Miguel Guinalíu. New members' integration: Key factor of success in online travel communities. *Journal of Business Research*, 66(6):706 – 710, 2013. International Tourism Behavior in Turbulent Times.

[CH13] Marcelo Cataldo and James D Herbsleb. Coordination breakdowns and their impact on development productivity and software failures. *IEEE Transactions on Software Engineering*, 39(3):343–360, 2013.

[CHC08] Marcelo Cataldo, James D Herbsleb, and Kathleen M Carley. Socio-technical congruence: a framework for assessing the impact of technical and work dependencies on software development productivity. In *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*, pages 2–11. ACM, 2008.

[CHDZ10] Teresa Correa, Amber Willard Hinsley, and Homero Gil De Zuniga. Who interacts on the web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2):247–253, 2010.

[Che15] Andrew Chen. " new data shows losing 80best apps do better", 2015. Retrieved July 7, 2016 from `http://andrewchen.co/new-data-shows-why-losing-80-of-your-mobile-users-is-normal-and-that-`

[CHW05] Galyen N. Chandler, Benson Honig, and Johan Wiklund. Antecedents, moderators, and performance consequences of membership change in new venture teams. *Journal of Business Venturing*, 20(5):705 – 725, 2005.

[CJM92] Paul T Costa Jr and Robert R McCrae. Neo personality inventory–revised (neo-pi-r) and neo five-factor inventory (neo-ffi) professional manual. *Odessa, FL: Psychological Assessment Resources*, 1992.

[CLYL11] Jengchung Victor Chen, Chinho Lin, David C Yen, and Kyaw-Phyo Linn. The interaction effects of familiarity, breadth and media usage on web browsing experience. *Computers in Human Behavior*, 27(6):2141–2152, 2011.

[COK14] Samantha A Conroy and Anne M O'Leary-Kelly. Letting go and moving on: Work-related identity loss and recovery. *Academy of Management Review*, 39(1):67–87, 2014.

[Col09] Robert K Colwell. Biodiversity: concepts, patterns, and measurement. *The Princeton guide to ecology*, pages 257–263, 2009.

[Con11] Wikimedia Strategic Planning Contributors. Wikipedia: Editor trends study, 2011. Retrieved February 14, 2018 from `https://strategy.wikimedia.org/w/index.php?title=Editor_Trends_Study`.

[CPVB06] Jon C. Carr, Allison W. Pearson, Michael J. Vest, and Scott L. Boyar. Prior occupational experience, anticipatory socialization, and employee retention. *Journal of Management*, 32(3):343–359, 2006.

[CRR10] Jilin Chen, Yuqing Ren, and John 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.

[CT86] John L. Cotton and Jeffrey M. Tuttle. Employee turnover: A meta-analysis and review with implications for research. *Academy of management Review*, 11(1):55 – 70, 1986.

[CT15] Giovanni Luca Ciampaglia and Dario Taraborelli. Moodbar: Increasing new user retention in wikipedia through lightweight socialization. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 734–742. ACM, 2015.

[CUC+05] Derek S Chapman, Krista L Uggerslev, Sarah A Carroll, Kelly A Piasentin, and David A Jones. Applicant attraction to organizations and job choice: a meta-analytic review of the correlates of recruiting outcomes. *Journal of applied psychology*, 90(5):928, 2005.

[CVdP08] Kristof Coussement and Dirk Van den Poel. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert systems with applications*, 34(1):313–327, 2008.

[Dar90] Y. Rene Darmon. Identifying sources of turnover costs: A segmental approach. *The Journal of Marketing*, pages 46–56, 1990.

[Deu59] Morton Deutsch. Some factors affecting membership motivation and achievement motivation in a group. *Human Relations*, 12(1):81–95, 1959.

[DMML00] Piew Datta, Brij Masand, DR Mani, and Bin Li. Automated cellular modeling and prediction on a large scale. *Artificial Intelligence Review*, 14(6):485–502, 2000.

[DO11] Linus Dahlander and Siobhan O'Mahony. Progressing to the center: Coordinating project work. *Organization science*, 22(4):961–979, 2011.

[DPRS12] Gideon Dror, Dan Pelleg, Oleg Rokhlenko, and Idan Szpektor. Churn prediction in new users of yahoo! answers. In *Proceedings of the 21st international conference companion on World Wide Web*, pages 829–834. ACM, 2012.

[DS01] Gregory G. Dess and Jason D. Shaw. Voluntary turnover, social capital, and organizational performance. *Academy of management review*, 26(3):446–456, 2001.

[DSTH12] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. Social coding in github: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1277–1286. ACM, 2012.

[DSV+08] Koustuv Dasgupta, Rahul Singh, Balaji Viswanathan, Dipanjan Chakraborty, Sougata Mukherjea, Amit A Nanavati, and Anupam Joshi. Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, pages 668–677. ACM, 2008.

[Duc05] Nicolas Ducheneaut. Socialization in an open source software community: A socio-technical analysis. *Computer Supported Cooperative Work (CSCW)*, 14(4):323–368, 2005.

[DWR09] Gina Dokko, Steffanie L. Wilk, and Nancy P. Rothbard. Unpacking prior experience: How career history affects job performance. *Organization Science*, 20(1):51–68, 2009.

[E+07] Nicole B Ellison et al. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

[EBRT13] Mehdi Elahi, Matthias Braunhofer, Francesco Ricci, and Marko Tkalcic. Personality-based active learning for collaborative filtering recommender systems. In *Congress of the Italian Association for Artificial Intelligence*, pages 360–371. Springer, 2013.

[EL00] Matthew S Eastin and Robert LaRose. Internet self-efficacy and the psychology of the digital divide. *Journal of Computer-Mediated Communication*, 6(1):0–0, 2000.

[FDFM+12] Paul Fugelstad, Patrick Dwyer, Jennifer Filson Moses, John Kim, Cleila Anna Mannino, Loren Terveen, and Mark Snyder. What makes users rate (share, tag, edit...)?: predicting patterns of participation in online communities. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 969–978. ACM, 2012.

[FDKP11] Rosta Farzan, Laura A Dabbish, Robert E Kraut, and Tom 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.

[FJGG09] Jill Freyne, Michal Jacovi, Ido Guy, and Werner Geyer. Increasing engagement through early recommender intervention. In *Proceedings of the third ACM conference on Recommender systems*, pages 85–92. ACM, 2009.

[FKL+12] Andrea Forte, Niki Kittur, Vanessa Larco, Haiyi Zhu, Amy Bruckman, and Robert 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.

[FKPK12] Rosta Farzan, Robert Kraut, Aditya Pal, and Joseph Konstan. Socializing volunteers in an online community: A field experiment. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 325–334, New York, NY, USA, 2012. ACM.

[FLB09] Andrea Forte, Vanesa Larco, and Amy Bruckman. Decentralization in wikipedia governance. *Journal of Management Information Systems*, 26(1):49–72, 2009.

[Fog05] Karl Fogel. *Producing open source software: How to run a successful free software project*. " O'Reilly Media, Inc.", 2005.

[Fox02] John Fox. Cox proportional-hazards regression for survival data. *An R and S-PLUS companion to applied regression*, 2002, 2002.

[FS00] Samer Faraj and Lee Sproull. Coordinating expertise in software development teams. *Management science*, 46(12):1554–1568, 2000.

[FSW06] Danyel Fisher, Marc Smith, and Howard T Welser. You are who you talk to: Detecting roles in usenet newsgroups. In *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, volume 3, pages 59b–59b. IEEE, 2006.

[GELA10] Boris Groysberg, Linda Eling-Lee, and Robin Abrahams. What it takes to make 'star' hires pay off. *MIT Sloan Management Review*, 51(2):57–61, 2010.

[GH13] R Stuart Geiger and Aaron Halfaker. Using edit sessions to measure participation in wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 861–870. ACM, 2013.

[GHG00] Rodger W Griffeth, Peter W Hom, and Stefan Gaertner. A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of management*, 26(3):463–488, 2000.

[Gin12] C Gini. Variabilità e mutabilità: Contributo allo studio delle distribuzioni e relazioni statistiche [variability and mutability: Contribution to the study of statistical distributions and relations]. *Bologna: Tipogr. di P. Cuppini*, 1912.

[Gou13] Georgios Gousios. The ghtorrent dataset and tool suite. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, MSR '13, pages 233–236, Piscataway, NJ, USA, 2013. IEEE Press.

[Gro10] Boris Groysberg. *Chasing stars: The myth of talent and the portability of performance*. Princeton University Press, Princeton, NJ, USA, 2010.

[GRS03] Samuel D Gosling, Peter J Rentfrow, and William B Swann. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.

[GUSA05] Roger Guimera, Brian Uzzi, Jarrett Spiro, and Luis A Nunes Amaral. Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308(5722):697–702, 2005.

[Her07] James D Herbsleb. Global software engineering: The future of socio-technical coordination. In *2007 Future of Software Engineering*, pages 188–198. IEEE Computer Society, 2007.

[HGMR13] Aaron Halfaker, R Stuart Geiger, Jonathan T Morgan, and John Riedl. The rise and decline of an open collaboration system: How wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 57(5):664–688, 2013.

[HH98] Michael Hauben and Ronda Hauben. Netizens: On the history and impact of usenet and the internet. *First Monday*, 3(7), 1998.

[HH13] John P. Hausknecht and Jacob A. Holwerda. When does employee turnover matter? dynamic member configurations, productive capacity, and collective performance. *Organizational Science*, 24(1):210–225, 2013.

[Hir70] Albert O Hirschman. *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*, volume 25. Harvard university press, 1970.

[HJYC07] Meng-Hsiang Hsu, Teresa L Ju, Chia-Hui Yen, and Chun-Ming Chang. Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations. *International journal of human-computer studies*, 65(2):153–169, 2007.

[HKKR09] Aaron Halfaker, Aniket Kittur, Robert Kraut, and John Riedl. A jury of your peers: quality, experience and ownership in wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, page 15. ACM, 2009.

[HMZ08] Jungpil Hahn, Jae Yun Moon, and Chen Zhang. Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties. *Information Systems Research*, 19(3):369–391, 2008.

[HO97] Joyce Hogan and Deniz S Ones. Conscientiousness and integrity at work. 1997.

[HP13] Rong Hu and Pearl Pu. Exploring relations between personality and user rating behaviors. In *UMAP Workshops*, 2013.

[HPB98] David A Harrison, Kenneth H Price, and Myrtle P Bell. Beyond relational demography: Time and the effects of surface-and deep-level diversity on work group cohesion. *Academy of management journal*, 41(1):96–107, 1998.

[HRBL12] David John Hughes, Moss Rowe, Mark Batey, and Andrew Lee. A tale of two sites: Twitter vs. facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2):561–569, 2012.

[HT11] Johm P. Hausknecht and Charlie O. Trevor. Collective turnover at the group, unit, and organizational levels: Evidence, issues, and implications. *Journal of Management*, 37(1):352–388, 2011.

[HTRR06] John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymtr Ruta. Churn prediction using complaints data. In *Proceedings Of World Academy Of Science, Engineering and Technology*. Citeseer, 2006.

[Hus95] M.A. Huselid. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *AMJ*, 38(3):635–672, 1995.

[JI02] Timothy A Judge and Remus Ilies. Relationship of personality to performance motivation: a meta-analytic review. *Journal of applied psychology*, 87(4):797, 2002.

[JK06] Elisabeth Joyce and Robert E Kraut. Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, 11(3):723–747, 2006.

[Jos06] Lou Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006.

[KBZJ05] Amy L Kristof-Brown, Ryan D Zimmerman, and Erin C Johnson. Consequences of individuals'fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. *Personnel psychology*, 58(2):281–342, 2005.

[KC87] E Lowell Kelly and James J Conley. Personality and compatibility: a prospective analysis of marital stability and marital satisfaction. *Journal of personality and social psychology*, 52(1):27, 1987.

[KCP+07] Aniket Kittur, Ed Chi, Bryan A Pendleton, Bongwon Suh, and Todd Mytkowicz. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World wide web*, 1(2):19, 2007.

[KDB+15] Eirini Kalliamvakou, Daniela Damian, Kelly Blincoe, Leif Singer, and Daniel M. German. Open source-style collaborative development practices in commercial projects using github. In *Proceedings of the 37th International Conference on Software Engineering - Volume 1*, ICSE '15, pages 574–585, Piscataway, NJ, USA, 2015. IEEE Press.

[KGB+14] Eirini Kalliamvakou, Georgios Gousios, Kelly Blincoe, Leif Singer, Daniel M German, and Daniela Damian. The promises and perils of mining github. In *Proceedings of the 11th working conference on mining software repositories*, pages 92–101. ACM, 2014.

[KK16] Raghav Pavan Karumur and Joseph A Konstan. Relating newcomer personality to survival and activity in recommender systems. In *Proceedings of the 24th ACM Conference on User Modeling, Adaptation & Personalization*, pages 595–608. ACM, 2016.

[KMWRS13] John Kammeyer-Mueller, Connie Wanberg, Alex Rubenstein, and Zhaoli Song. Support, undermining, and newcomer socialization: Fitting in during the first 90 days. *Academy of Management Journal*, 56(4):1104–1124, 2013.

[KNK16] Raghav Pavan Karumur, Tien T Nguyen, and Joseph A Konstan. Early activity diversity: Assessing newcomer retention from first-session activity. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 595–608. ACM, 2016.

[KNK17] Raghav Pavan Karumur, Tien T. Nguyen, and Joseph A. Konstan. Personality, user preferences and behavior in recommender systems. *Information Systems Frontiers*, 2017.

[Kno14] Eric Knorr. "github's new ceo: We're serious about the enterprise", Sep 2014. Retrieved July 26, 2016 from `http://www.infoworld.com/article/2608907/application-development/github-s-new-ceo--we-re-serious-about-the-enterprise.html`.

[Kol99] Peter Kollock. The economies ol online cooperation. *Communities in cyberspace*, 220, 1999.

[Kor09] Russell F Korte. How newcomers learn the social norms of an organization: A case study of the socialization of newly hired engineers. *Human Resource Development Quarterly*, 20(3):285–306, 2009.

[KP85] David Krackhardt and Lyman W. Porter. When friends leave: A structural analysis of the relationship between turnover and stayers' attitudes. *Administrative science quarterly*, pages 242–261, 1985.

[KPK09] Aniket Kittur, Bryan Pendleton, and Robert 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.

[KPS09] Jaya Kawale, Aditya Pal, and Jaideep Srivastava. Churn prediction in mmorpgs: A social influence based approach. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 423–428. IEEE, 2009.

[Kri96] Amy L Kristof. Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel psychology*, 49(1):1–49, 1996.

[KRK+12] Robert E. Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph A. Konstan, Yuging Ren, and John Riedl. *Building Successful Online communities: Evidence-based social design*. MIT Press, Cambridge, MA, USA, 2012.

[KVE05] Gerbert Kraaykamp and Koen Van Eijck. Personality, media preferences, and cultural participation. *Personality and individual differences*, 38(7):1675–1688, 2005.

[KWL12] Sanjay Ram Kairam, Dan J Wang, and Jure Leskovec. The life and death of online groups: Predicting group growth and longevity. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 673–682. ACM, 2012.

[KYZK18] Raghav Pavan Karumur, Bowen Yu, Haiyi Zhu, and Joseph A Konstan. Content is king, leadership lags: Effects of prior experience on newcomer retention and productivity in online production groups. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 506. ACM, 2018.

[La15] Alyson La. "language trends on github", Aug 2015. Retrieved Aug 2, 2016 fron `https://github.com/blog/2047-language-trends-on-github`.

[LaR09] Robert LaRose. Social cognitive theories of media selection. *Media choice: A theoretical and empirical overview*, pages 10–31, 2009.

[LBL+05] Kimberly Ling, Gerard Beenen, Pamela Ludford, Xiaoqing Wang, Klarissa Chang, Xin Li, Dan Cosley, Dan Frankowski, Loren Terveen, Al Mamunur Rashid, et al. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4):00–00, 2005.

[LH03] Karim R. Lakhani and Eric V. Hippel. How open source software works: "free" user-to-user assistance. *Research Policy*, 32(6):923 – 943, 2003.

[LJ05] Cliff Lampe and Erik Johnston. Follow the (slash) dot: effects of feedback on new members in an online community. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pages 11–20. ACM, 2005.

[LL06] Richard N Landers and John W Lounsbury. An investigation of big five and narrow personality traits in relation to internet usage. *Computers in Human Behavior*, 22(2):283–293, 2006.

[LM94] John M Levine and Richard L Moreland. Group socialization: Theory and research. *European review of social psychology*, 5(1):305–336, 1994.

[LRM14] Antonio Lima, Luca Rossi, and Mirco Musolesi. Coding together at scale: Github as a collaborative social network. *arXiv preprint arXiv:1407.2535*, 2014.

[LW91] Jean Lave and Etienne Wenger. *Situated learning: Legitimate peripheral participation*. Cambridge university press, 1991.

[LW03] Karim Lakhani and Robert G Wolf. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. 2003.

[MAN06] Teemu Mutanen, Jussi Ahola, and Sami Nousiainen. Customer churn prediction-a case study in retail banking. In *Proc. of ECML/PKDD Workshop on Practical Data Mining*, pages 13–19, 2006.

[MBWS13] Jonathan T Morgan, Siko Bouterse, Heather Walls, and Sarah Stierch. Tea and sympathy: crafting positive new user experiences on wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 839–848. ACM, 2013.

[MDH13] Jennifer Marlow, Laura Dabbish, and Jim Herbsleb. Impression formation in online peer production: activity traces and personal profiles in github. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 117–128. ACM, 2013.

[MEM+12] Michael Muller, Kate Ehrlich, Tara Matthews, Adam Perer, Inbal Ronen, and Ido Guy. Diversity among enterprise online communities: collaborating, teaming, and innovating through social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2815–2824. ACM, 2012.

[Met15] Cade Metz. "how github conquered google, microsoft and everyone else", March 2015. Retrieved July 25, 2016 from `http://www.wired.com/2015/03/github-conquered-google-microsoft-everyone-else/`.

[MFH02] Audris Mockus, Roy T. Fielding, and James D. Herbsleb. Two case studies of open source software development: Apache and mozilla. *ACM Trans. Softw. Eng. Methodol.*, 11(3):309–346, July 2002.

[MG14] Gloria Mark and Yoav Ganzach. Personality and internet usage: A large-scale representative study of young adults. *Computers in Human Behavior*, 36:274–281, 2014.

[MG16] Claire M. Gardiner. Legitimizing processes: Barriers and facilitators for experienced newcomers' entry transitions to knowledge practices. 06 2016.

[MGMZ13] Jonathan T Morgan, Michael Gilbert, David W McDonald, and Mark Zachry. Project talk: Coordination work and group membership in

wikiprojects. In *Proceedings of the 9th International Symposium on Open Collaboration*, page 3. ACM, 2013.

[MHE+10] Bjørn Erik Mørk, Thomas Hoholm, Gunnar Ellingsen, Bjørn Edwin, and Margunn Aanestad. Challenging expertise: On power relations within and across communities of practice in medical innovation. *Management Learning*, 41(5):575–592, 2010.

[MI14] Melody Meckfessel and Tomas Isdal. "devops at the speed of google", June 2014. Retrieved June 27, 2016 from `https://www.youtube.com/watch?v=Xt9Fc3-wp0E/`.

[MJ92] Robert R McCrae and Oliver P John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, 1992.

[MPD92] J Miller McPherson, Pamela A Popielarz, and Sonja Drobnic. Social networks and organizational dynamics. *American sociological review*, pages 153–170, 1992.

[MT06] Richard L Moreland and L Thompson. Transactive memory: Learning who knows what in work groups and organizations. *Small groups: Key readings*, pages 327–346, 2006.

[MTSC+16] Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. Blissfully happy" or "ready to fight": Varying interpretations of emoji. *Proceedings of ICWSM*, 2016, 2016.

[MWG+00] Michael C Mozer, Richard Wolniewicz, David B Grimes, Eric Johnson, and Howard Kaushansky. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on neural networks*, 11(3):690–696, 2000.

[Nov07] Oded Nov. What motivates wikipedians? *Communications of the ACM*, 50(11):60–64, 2007.

[NP00] Blair Nonnecke and Jenny Preece. Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 73–80. ACM, 2000.

[OF09] Lisa J Orchard and Chris Fullwood. Current perspectives on personality and internet use. *Social Science Computer Review*, 2009.

[OICB89] Charles A O'Reilly III, David F Caldwell, and William P Barnett. Work group demography, social integration, and turnover. *Administrative science quarterly*, pages 21–37, 1989.

[OR95] Dennis W Organ and Katherine Ryan. A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel psychology*, 48(4):775–802, 1995.

[PAT14] Jagat Sastry Pudipeddi, Leman Akoglu, and Hanghang Tong. User churn in focused question answering sites: characterizations and prediction. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 469–474. International World Wide Web Conferences Steering Committee, 2014.

[Pau15] Sawers Paul. "github by the numbers: 32m people visit each month - 74% from outside the u.s., 36% from europe", June 2015. Retrieved July 26, 2016 from `http://venturebeat.com/2015/06/17/github-by-the-numbers-32m-people-visit-each/month-74-from-outside-the-u-s-36-from-europe/`.

[PBB06] James G Phillips, Sarah Butt, and Alex Blaszczynski. Personality and self-reported use of mobile phones for games. *CyberPsychology & Behavior*, 9(6):753–758, 2006.

[PCK12] Aditya Pal, Shuo Chang, and Joseph A Konstan. Evolution of experts in question answering communities. In *ICWSM*, 2012.

[PCL+07] Reid Priedhorsky, Jilin Chen, Shyong Tony K Lam, Katherine Panciera, Loren Terveen, and John Riedl. Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work*, pages 259–268. ACM, 2007.

[Pen09] Parag C Pendharkar. Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. *Expert Systems with Applications*, 36(3):6714–6720, 2009.

[PHT09] Katherine Panciera, Aaron Halfaker, and Loren 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.

[PJ09] Yunrim Park and Carlos Jensen. Beyond pretty pictures: Examining the benefits of code visualization for open source newcomers. In *Visualizing Software for Understanding and Analysis, 2009. VISSOFT 2009. 5th IEEE International Workshop on*, pages 3–10. IEEE, 2009.

[PNA04] Jenny Preece, Blair Nonnecke, and Dorine Andrews. The top five reasons for lurking: improving community experiences for everyone. *Computers in human behavior*, 20(2):201–223, 2004.

[PPET10] Katherine Panciera, Reid Priedhorsky, Thomas Erickson, and Loren Terveen. Lurking? cyclopaths?: a quantitative lifecycle analysis of user behavior in a geowiki. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1917–1926. ACM, 2010.

[Pre00] Jenny Preece. *Online communities: Designing usability and supporting socialbilty*. John Wiley & Sons, Inc., 2000.

[PSL+13] Raphael Pham, Leif Singer, Olga Liskin, Fernando Figueira Filho, and Kurt Schneider. Creating a shared understanding of testing culture on a social coding site. In *2013 35th International Conference on Software Engineering (ICSE)*, pages 112–121. IEEE, 2013.

[QF11] Israr Qureshi and Yulin Fang. Socialization in open source software projects: A growth mixture modeling approach. *Organizational Research Methods*, 14(1):208–238, 2011.

[QSTC14] Xiangju Qin, Michael Salter-Townshend, and Pádraig Cunningham. Exploring the relationship between membership turnover and productivity in online communities. In *Proceedings of the 8th International Conference on Weblogs and Social Media*, ICWSM '14. AAAI, 2014.

[RC05] Shelly Rodgers and Qimei Chen. Internet community group participation: Psychosocial benefits for women with breast cancer. *Journal of Computer-Mediated Communication*, 10(4):00–00, 2005.

[RG03] Peter J Rentfrow and Samuel D Gosling. The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, 84(6):1236, 2003.

[Rhe93] Howard Rheingold. *The virtual community: Finding commection in a computerized world*. Addison-Wesley Longman Publishing Co., Inc., 1993.

[RJ03] Amy E Randel and Kimberly S Jaussi. Functional background identity, diversity, and individual performance in cross-functional teams. *Academy of Management Journal*, 46(6):763–774, 2003.

[RKK07] Yuqing Ren, Robert Kraut, and Sara Kiesler. Applying common identity and bond theory to design of online communities. *Organization Studies*, 28(3):377–408, 2007.

[ROS+09] Craig Ross, Emily S Orr, Mia Sisic, Jaime M Arseneault, Mary G Simmering, and R Robert Orr. Personality and motivations associated with facebook use. *Computers in human behavior*, 25(2):578–586, 2009.

[RR] Yuqing Ren and John Riedl. Helping wikipedia versus helping a wikiproject: Subgroup dynamics, member contribution and turnover in online production communities. Technical report.

[RX11] Tracii Ryan and Sophia Xenos. Who uses facebook? an investigation into the relationship between the big five, shyness, narcissism, loneliness, and facebook usage. *Computers in Human Behavior*, 27(5):1658–1664, 2011.

[RYTS10] Yossi Richter, Elad Yom-Tov, and Noam Slonim. Predicting customer churn in mobile networks through analysis of social groups. In *SDM*, volume 2010, pages 732–741. SIAM, 2010.

[SAMBM07] David P Schmitt, Jüri Allik, Robert R McCrae, and Verónica Benet-Martínez. The geographic distribution of big five personality traits patterns and profiles of human self-description across 56 nations. *Journal of cross-cultural psychology*, 38(2):173–212, 2007.

[SBK+14] Jyoti Sheoran, Kelly Blincoe, Eirini Kalliamvakou, Daniela Damian, and Jordan Ell. Understanding watchers on github. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, pages 336–339. ACM, 2014.

[SG06] Katherine J Stewart and Sanjay Gosain. The impact of ideology on effectiveness in open source software development teams. *Mis Quarterly*, pages 291–314, 2006.

[SGR14] Igor Steinmacher, Marco Aurélio Gerosa, and D Redmiles. Attracting, onboarding, and retaining newcomer developers in open source software projects. In *Workshop on Global Software Development in a CSCW Perspective*, 2014.

[Sha01] Claude E Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[SHHH02] Rhonda J Swickert, James B Hittner, Jamie L Harris, and Jennifer A Herring. Relationships among internet use, personality, and social support. *Computers in human behavior*, 18(4):437–451, 2002.

[Sim49] Edward H Simpson. Measurement of diversity. *Nature*, 1949.

[SKT09] Johann Schrammel, Christina Köffel, and Manfred Tscheligi. Personality traits, usage patterns and information disclosure in online communities. In *Proceedings of the 23rd British HCI group annual conference on people and computers: celebrating people and technology*, pages 169–174. British Computer Society, 2009.

[SMO97] Christine B Smith, Margaret L McLaughlin, and Kerry K Osborne. Conduct control on usenet. *Journal of Computer-Mediated Communication*, 2(4):0–0, 1997.

[Spe09] Donna Spencer. *Card sorting: Designing usable categories*. Rosenfeld Media, 2009.

[SR06] Daniel Stutzbach and Reza Rejaie. Understanding churn in peer-to-peer networks. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 189–202. ACM, 2006.

[SSC+14] Margaret-Anne Storey, Leif Singer, Brendan Cleary, Fernando Figueira Filho, and Alexey Zagalsky. The (r) evolution of social media in software engineering. In *Proceedings of the on Future of Software Engineering*, pages 100–116. ACM, 2014.

[SSGR15] Igor Steinmacher, Marco Aurelio Graciotto Silva, Marco Aurelio Gerosa, and David F Redmiles. A systematic literature review on the barriers faced by newcomers to open source software projects. *Information and Software Technology*, 59:67–85, 2015.

[Sta17] Wikipedia Statistics. Wikipedia: Statistics, December 2017. Retrieved January 10, 2018 from `https://en.wikipedia.org/wiki/Wikipedia:Statistics`.

[Str90] Myra H. Strober. Human capital theory: Implications for hr managers. *Industrial Relations*, 29(2):214–239, 1990.

[Stu03] Michael C Sturman. Searching for the inverted u-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management*, 29(5):609–640, 2003.

[SW14] Jacob Solomon and Rick Wash. Critical mass of what? exploring community growth in wikiprojects. In *ICWSM*, 2014.

[SWCG13] Igor Steinmacher, Igor Wiese, Ana Paula Chaves, and Marco Aurélio Gerosa. Why do newcomers abandon open source software projects? In

Cooperative and Human Aspects of Software Engineering (CHASE), 2013 6th International Workshop on, pages 25–32. IEEE, 2013.

[SWG12] Igor Steinmacher, Igor Scaliante Wiese, and Marco Aurélio Gerosa. Recommending mentors to software project newcomers. In *Proceedings of the Third International Workshop on Recommendation Systems for Software Engineering*, pages 63–67. IEEE Press, 2012.

[Tay88] M Susan Taylor. Effects of college internships on individual participants. *Journal of Applied Psychology*, 73(3):393, 1988.

[Tay09] Lesley .C Taylor. Thousands of editors leaving wikipedia. Article, November 2009. Retrieved September 4, 2017 from `https://www.thestar.com/news/2009/11/23/thousands_of_editors_leaving_wikipedia.html`.

[TB01] Tracy L Tuten and Michael Bosnjak. Understanding differences in web usage: The role of need for cognition and the five factor model of personality. *Social Behavior and Personality: an international journal*, 29(4):391–398, 2001.

[TC92] Ernest C Tupes and Raymond E Christal. Recurrent personality factors based on trait ratings. *Journal of personality*, 60(2):225–251, 1992.

[TC15] Marko Tkalcic and Li Chen. Personality and recommender systems. In *Recommender Systems Handbook*, pages 715–739. Springer, 2015.

[TDH14] Jason Tsay, Laura Dabbish, and James Herbsleb. Influence of social and technical factors for evaluating contribution in github. In *Proceedings of the 36th international conference on Software engineering*, pages 356–366. ACM, 2014.

[Ten08] Ching-I Teng. Personality differences between online game players and nonplayers in a student sample. *CyberPsychology & Behavior*, 11(2):232–234, 2008.

[TL10] Leman Pinar Tosun and Timo Lajunen. Does internet use reflect your personality? relationship between eysenck's personality dimensions and internet use. *Computers in Human Behavior*, 26(2):162–167, 2010.

[TS10] Christoph Treude and Margaret-Anne Storey. Awareness 2.0: staying aware of projects, developers and tasks using dashboards and feeds. In *2010 ACM/IEEE 32nd International Conference on Software Engineering*, volume 1, pages 365–374. IEEE, 2010.

[TSFW05] Tammara Combs Turner, Marc A Smith, Danyel Fisher, and Howard T Welser. Picturing usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication*, 10(4):00–00, 2005.

[UD10] Hang Ung and Jean-Michel Dalle. Project management in the wikipedia community. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, WikiSym '10, pages 13:1–13:4, New York, NY, USA, 2010. ACM.

[US05] Brian Uzzi and Jarrett Spiro. Collaboration and creativity: The small world problem1. *American journal of sociology*, 111(2):447–504, 2005.

[VKC10] Padmal Vitharana, Julie King, and Helena Chapman. Impact of internal open source development on reuse: Participatory reuse in action. *J. Manage. Inf. Syst.*, 27(2):277–304, October 2010.

[VKSL03] Georg Von Krogh, Sebastian Spaeth, and Karim R Lakhani. Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, 32(7):1217–1241, 2003.

[VMMB11] Wouter Verbeke, David Martens, Christophe Mues, and Bart Baesens. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3):2354–2364, 2011.

[VWLB14] Alcides Velasquez, Rick Wash, Cliff Lampe, and Tor Bjornrud. Latent users in an online user-generated content community. *Computer Supported Cooperative Work (CSCW)*, 23(1):21–50, 2014.

[VYW+15] Bogdan Vasilescu, Yue Yu, Huaimin Wang, Premkumar Devanbu, and Vladimir Filkov. Quality and productivity outcomes relating to continuous integration in github. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, pages 805–816. ACM, 2015.

[WC02] Chih-Ping Wei and I-Tang Chiu. Turning telecommunications call details to churn prediction: a data mining approach. *Expert systems with applications*, 23(2):103–112, 2002.

[WCRR12] Loxley Sijia Wang, Jilin Chen, Yuqing Ren, and John 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*, CSCW '12, pages 989–998, New York, NY, USA, 2012. ACM.

[WD01] Uwe Wolfradt and Jörg Doll. Motives of adolescents to use the internet as a function of personality traits, personal and social factors. *Journal of Educational Computing Research*, 24(1):13–27, 2001.

[WDX+06] Dmitri Williams, Nicolas Ducheneaut, Li Xiong, Yuanyuan Zhang, Nick Yee, and Eric Nickell. From tree house to barracks: The social life of guilds in world of warcraft. *Games and culture*, 1(4):338–361, 2006.

[Wei15] Matt Weinberger. "github, the $2 billion 'Facebook for programmers,'has a plan to get even bigger", October 2015. Retrieved July 26, 2016 from `http://www.businessinsider.com/github-universe-2015-ceo-chris-wanstrath-keynote-2015-10`.

[WF00] M McLure Wasko and Samer Faraj. "it is what one does": why people participate and help others in electronic communities of practice. *The Journal of Strategic Information Systems*, 9(2):155–173, 2000.

[WGFS07] Howard T Welser, Eric Gleave, Danyel Fisher, and Marc Smith. Visualizing the signatures of social roles in online discussion groups. *Journal of social structure*, 8(2):1–32, 2007.

[WKL12] Yi-Chia Wang, Robert Kraut, and John 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*, CSCW '12, pages 833–842, New York, NY, USA, 2012. ACM.

[WOI98] KY Williams and CA O'Reilly III. Demography and diversity in organisations: A review of 40 years of research in bm staw and ll cummings (eds) research in organisational behaviour vol. 20. *Jai Pres, Connecticut*, 1998.

[WSD+96] Barry Wellman, Janet Salaff, Dimitrina Dimitrova, Laura Garton, Milena Gulia, and Caroline Haythornthwaite. Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual review of sociology*, 22(1):213–238, 1996.

[WSSP03] Jason Tsong-Li Wang, Huiyuan Shan, Dennis Shasha, and William H Piel. Treerank: a similarity measure for nearest neighbor searching in phylogenetic databases. In *Scientific and Statistical Database Management, 2003. 15th International Conference on*, pages 171–180. IEEE, 2003.

[WTWT15] Etienne .C. Wenger-Trayner and Beverly Wenger-Trayner. *Learning in landscapes of practice: Boundaries, Identity, And Knowledgeability in Practice-Based Learning*, chapter Learning in a landscape of practice: A framework., pages 13–29. Routledge, New York, NY, 1st edition, 2015.

[YRZ17] Bowen Yu, Terveen Loren Ren, Yuqing, and Haiyi Zhu. Predicting member productivity and withdrawal from pre-joining attachments in online production groups. In *Proceedings of the ACM 2017 conference on Computer Supported Cooperative Work*. ACM, 2017.

[YWAA10] Jiang Yang, Xiao Wei, Mark S Ackerman, and Lada A Adamic. Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities. *ICWSM*, 10:186–193, 2010.

[YWF+15] Yue Yu, Huaimin Wang, Vladimir Filkov, Premkumar Devanbu, and Bogdan Vasilescu. Wait for it: Determinants of pull request evaluation latency on github. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*, pages 367–371. IEEE, 2015.

[YWL+17] Bowen Yu, Xinyi Wang, Allen Yilun Lin, Yuqing Ren, Loren Terveen, and Haiyi Zhu. Out with the old, in with the new?: Unpacking member turnover in online production groups. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):117:1–117:19, December 2017.

[Zak13] Nicholas C. Zakas. "github workflows inside of a company", May 2013. Retrieved July 7, 2016 from `https://www.nczonline.net/blog/2013/05/21/github-workflows-inside-of-a-company/`.

[ZFBL09] Robert Zinko, Gerald R. Ferris, Fred R. Blass, and Mary D. Laird. Toward a theory of reputation in organizations. *Research in Personnel and Human Resources Management*, 26(1):163–204, 2009.

[ZKK12a] Haiyi Zhu, Robert Kraut, and Aniket Kittur. Effectiveness of shared leadership in online communities. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 407–416, New York, NY, USA, 2012. ACM.

[ZKK12b] Haiyi Zhu, Robert Kraut, and Aniket 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.

[ZM11] Minghui Zhou and Audris Mockus. Does the initial environment impact the future of developers? In *Proceedings of the 33rd International Conference on Software Engineering*, pages 271–280. ACM, 2011.

[ZM12] Minghui Zhou and Audris Mockus. What make long term contributors: Willingness and opportunity in oss community. In *Proceedings of the 34th International Conference on Software Engineering*, pages 518–528. IEEE Press, 2012.

[ZUM93] Marvin Zuckerman, Roger S Ulrich, and John McLaughlin. Sensation seeking and reactions to nature paintings. *Personality and individual differences*, 15(5):563–576, 1993.

[ZZH+13] Haiyi Zhu, Amy Zhang, Jiping He, Robert E. Kraut, and Aniket Kittur. Effects of peer feedback on contribution: A field experiment in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2253–2262, New York, NY, USA, 2013. ACM.