Consumers, Editors, and Power Editors at Work: Diversity of Users in Online Peer Production Communities

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Katherine Anne Panciera

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Loren G. Terveen

August 2014

Acknowledgments

It feels like I've been writing this dissertation and working on the research forever, possibly because it's taken me nine years to get to this point. There are lots of people who helped me out along the way, more than can be mentioned here. Thanks!

Loren welcomed me from my first visit to UMN, encouraged me to ask big questions and think about things from different perspectives, and helped me become a better student and researcher.

This work was conducted in cooperation with Aaron Halfaker, Reid Priedhorsky, Mikhil Masli, and Loren Terveen. They are wonderful people to bounce ideas off of, resistant when need be, but willing to have their minds changed by a persuasive argument.

The Cyclopath users who shared their time and ideas with us are very much appreciated. In adddition, the work could not have been done without the support of the Cyclopath development team.

I assembled a committee a little late in the game, but many thanks go to Ching Ren, John Carlis, and Joe Konstan for helping me out, answering questions, and encouraging me to "just finish writing the thing."

Throughout my graduate career, I've struggled with headaches and migraines, and I've been lucky to have been able to work with outstanding medical professionals who don't give up on me. Ann Rechtzigel, CNP, in particular has helped make this dissertation possible in more ways than she can likely imagine. My friend, Katie Kumler, also helped me tolerate the migraines and helped out by proofreading numerous papers, posters, and proposals.

Finally, to my family. My parents, Anne and Mike, have always encouraged me to do whatever would make me happy and satisfied. My brother, Anthony, rejoiced in my success and helped me keep my focus. I didn't expect to end up where I am today, but a good part of it is all due to the family I grew up in and the values they instilled in me.

And to Ben, who's only ever known me as a Ph.D. student. Thanks for understand-

ing and putting up with the long nights, never ending paper deadlines, and grumpy moods. For getting that I didn't want to cook, think, or move after I got home from work. And for doing so many dishes and loads of laundry. But most importantly, thanks for keeping me laughing.

Abstract

Many people rely on open collaboration projects to run their computer (Linux), browse the web (Mozilla Firefox), and get information (Wikipedia). Open content web sites are peer production communities which depend on users to produce content. In this thesis, we analyze three types of users in peer production communities: consumers, contributors, and core contributors. Consumers don't edit or add content while contributors add some content. Core contributors edit or contribute much more content than others on the site. The three types of users each serve a different role in the community, receive different benefits from the community, and are important to the survival of a community.

For the purposes of this thesis, we look at users in two communities: Wikipedia and Cyclopath. Wikipedia is the largest and most well-known peer production community. The majority of the work in this dissertation is from Cyclopath, a geowiki for bicyclists developed by GroupLens. Since we built Cyclopath, we have access to data that allowed us to delve much deeper into the divide between the three types of users.

First, we wanted to understand what the quantitative differences between core contributors and contributors were. On Wikipedia and Cyclopath, core contributors start editing more intensely from their first day on the site. In addition, core contributors have higher edit quality and namespace diversity starting on day one. Following the first day, all users saw a drop off in their edit quantity and edit quality. On Cyclopath we were able to look at pre-registration activity and found equivocal evidence for "educational lurking". In addition the geographic nature of the Cyclopath data allowed us to look at the correlation between editing and viewing activity and we found that Cyclopath core contributors may reveal some information about their viewing activity through their editing activity.

Building on this quantitative analysis, we turned to qualitative questions. By surveying and interviewing Cyclopath users, we learned what motivates them to participate and what benefits they derive from their participation. While consumers and contributors both benefited by receiving routes, contributors were more likely to say they registered in order to edit. (Registration was not required to edit.) In

addition, we discovered user attitudes and thoughts, the effect of the surrounding ecology of related systems, and users' technical and cycling histories. Here, we found that the Cyclopath core contributors aren't who you might think. They are not the most dedicated bicyclists, but they are committed to the values of open content. In addition, Cyclopath evolved in an open ecosystem and the usage of the site was in part determined by the existence of other sites serving other needs of similar users. By providing a holistic view of users on Cyclopath and by looking at Wikipedia editors quantitatively, we discovered opportunities for new forms of participation, such as an outlet for subjective comments and annotations, as well a key to motivating people to contributing objective information, highlighting flaws and easy fixes in the system. We hope that future community builders and researchers will be able to incorporate these findings into their systems.

Contents

List of Tables

List of Figures

A note on the design of this document

This thesis is designed using standard book design guidelines. Readers who prefer a version conforming to the University of Minnesota dissertation formatting requirements (notably, double-spaced and suitable only for single-sided printing) should contact the author.

Chapter 1

Introduction

Social computing technologies have revolutionized the way people connect, communicate, and work together. Past research has focused on the use of these technologies for social production [23, 54]: loosely connected individuals freely coming together to produce information and artifacts of value [36]. Open source software systems and Wikipedia are prototypical cases. For example, tens of thousands of people from around the world have written over three million articles on Wikipedia making it the largest encyclopedia in history. Five of the top ten websites in the world (by traffic) rely on information from individuals: Facebook, YouTube, Wikipedia, Twitter, and LinkedIn [1]. These communities are also known as peer production communities, as your peers are producing the information that you are consuming.

1.1 Peer Production Communities

In this dissertation, I focus on online communities where users produce and share information of value. Sites that we study or compare our research to include Wikipedia, Cyclopath, Open Source Software (OSS), MovieLens, and Everything2. Note that we're not looking at proprietary systems such as Facebook or YouTube. While we compare our results to sites such as Everything2 and MovieLens, the specific focus of this dissertation is on wiki platforms (namely Wikipedia and Cyclopath). One of the factors required for these sites to succeed is active and committed members. This results in content, produced by members, with regular turnover and consistent appeal.

1.1.1 Definition

For the purposes of this dissertation, the peer production communities that we focus on are online communities that depend on user generated content or "any form of content such as video, blogs, discussion form posts,... and other forms of media that was created by consumers or end-users of an online system or service and is publicly available" [5]. These communities are meritocracies where anyone can contribute. In some communities no specific knowledge is required (e.g. Galaxy Zoo) whereas in others specific knowledge or skills are required (e.g. Wikipedia, Open Street Maps, StackExchange).

1.1.2 Failure and Success

Within this online ecosystem, some communities fail while others succeed. Examples of failures include iParent [92], Ping [25], Friendster, Diaspora*, Orkut, and MySpace [75]. While some of these sites haven't closed their websites yet, their user bases have plummeted from their one-time highs. Perhaps closer to this dissertation are the failures of Nupedia [37] and Citizendium [70]. Nupedia was the primary forerunner of Wikipedia, but had a more stringent peer-review process than Wikipedia did [47]. Citizendium is an attempt at creating an online encyclopedia, but requiring all editors to submit their real-life credentials in order to do so [70]. All of these examples are big sites, compared to many on the web, some of which were quite popular before there was a mass exodus to another platform.

In contrast, there are a number of successes in this space, Wikipedia, Reddit, StackExchange, and YouTube being just a handful of the bigger successes. There are also smaller successes, such as Cyclopath, MovieLens, and Everything2 (introduced more in depth in Section 2.1). And yet, even looking at these success stories, there are still problems. On SourceForge, 31% of open source projects are abandoned before their first release [134]. Newgrounds, a collaborative animation site, has 87.4% of collaborative animations listed as incomplete [87]. In August 2012, on English Wikipedia, 1,641 pages were marked as needing expansion[1]

1.2 Users

Within these communities, there is a wide spectrum of activity. Many (or most) users only consume information from the site. These **consumers** are often called "lurkers" and don't edit, but may benefit from content on the site. However, as they are often invisible to researchers, this group is rarely studied.

[1] http://en.wikipedia.org/wiki/Category:Articles_to_be_expanded_from_August_2012

Other users are occasional or **average contributors**. Often called posters or editors, these users add information to the site and are recognized as being distinct from consumers (although contributors also consume) [101].

The final type of user that we look at is the **core contributor**, sometimes called the power editor. These users do more than others in the community A study by Bryant et al. on Wikipedia found differences between average contributors and core contributors [26]. A quantitative study by Priedhorsky et al. found that the top 0.01% of editors contribute 44% of the value of Wikipedia [113]. Edits by these users are also less likely than others to be changed [73].

From site to site, the ratio of these three groups may differ, but they usually all exist. These three types of users serve different roles in the community and receive different benefits from participating in the community and all three are important to the survival of a community. In the open source community, consumers benefit from getting to use the software, whereas developers (contributors) get to develop their own skills [105].

Previous research has looked at different aspects of the spectrum such as just the core contributors [26], all contributors [129, 21, 113], and contributors and consumers [39, 46, 63, 78, 135].

This research focuses on the core contributors, although we also look at the core contributors in relation to the average contributors and consumers. We primarily researched core contributors in part because community survival in a peer production community is dependent on the existence of content. Therefore the existence and retention of core contributors in online communities is crucial to the survival of the community.

1.3 Dissertation Goals

This dissertation is designed to provide a holistic view of core contributors in the Cyclopath community. I present quantitative and qualitative data about these core contributors and offer some insight into their offline, as well as online, activity and patterns.

This dissertation also compares these core contributors to other users of the system, both consumers and average contributors, again, quantitatively and qualitatively. This includes when they contribute, how much they contribute, how much they view the site, and their retention rate, discussed in Chapter 5. Qualitatively, in Chapter 6, I compare core contributors to others in why they registered for the site, how they benefit from the site, and how others may benefit from their participation on the site.

I also present a quantitative understanding of core contributors on Wikipedia in Chapter 3, designed to be paired with an existing qualitative study [26]. This investigates the editing patterns of Wikipedia core contributors, their contribution quality, where they edit, and how involved they are with policy and vandalism patrolling.

Finally, this dissertation, when paired with [26], allows quantitative and qualitative comparison of core contributors on Wikipedia and Cyclopath as well as some comparisons to other online communities.

The penultimate chapter of this dissertation, Chapter 8, provides design guidelines and ideas for community creators to help them make sites that will better engage all their users.

This dissertation does not predict whether or not users will be core contributors, contributors, or consumers. It also does not attempt to find ways to get all consumers to be active contributors or find ways to get all contributors to be core contributors. We believe that many consumers and contributors have more to offer the community, knowledge that only they would be able to contribute. Our goal is to get them to contribute this knowledge, but not to insist that all users, even those without valid contributions, contribute. Some of the ideas presented in Chapter 8 are designed to encourage, but not force, these contributions.

Chapter 2

Related Work

The work in this dissertation relies on a significant amount of prior work ranging from work on the psychology of volunteering to in depth research on the development of Wikipedia policy. While each of the studies presented in this dissertation originally had separate related work sections, the overlap in material and related work suggested that the related work would be better suited in a single chapter, referred to throughout the dissertation.

In this chapter first I outline a number of online peer production systems that are referenced throughout the dissertation. Then I focus on the existence of core contributors throughout a number of these sites or systems and on the differences between average contributors and core contributors. Finally, I focus on research that has been done on consumers.

2.1 Systems

There are a number of peer-production online communities that have been studied. In this dissertation, the focus is on work done on Wikipedia and Cyclopath but we compare results to several other sites as well. They are introduced briefly below and will be referred to throughout the dissertation.

2.1.1 Wikipedia

Wikipedia was founded in 2001 and is consistently one of the top ten websites in the world [8]. In August 2012 alone, Wikipedia had over 492 million readers [2] and over 33,000 active editors [6]. While critics have argued that the quality of

Wikipedia could not be close to that of an expert-written encyclopedia, a study in 2005 judged the quality of Wikipedia as similar to the quality of the Encyclopedia Britannica [55].

WikiProjects are, in essence, teams that come together to improve specific Wikipedia articles. They have official pages to help coordinate work and there are over 1,700 WikiProjects that have been active in the past year [12]. One example of a WikiProject is the WikiProject Minnesota, a group with 120 members, which is trying to improve the quality of Wikipedia articles about Minnesota. On their page, they list core articles by class and quality with comments including praise and hints as to what needs to be done on each page [13].

2.1.2 Cyclopath

Cyclopath is a geographic wiki for bicyclists. It serves as a bike-friendly routefinder while also allowing users to make edits within the map itself or to notes or tags attached to the map. The site was created by the University of Minnesota GroupLens research lab and has been the basis for a number of research papers [1]. It is discussed in more depth in Chapter 4.

2.1.3 MovieLens

Movie Lens is a movie recommendations engine and online community. People who join are asked to rate 10 movies and then can receive recommendations, rate more movies, edit movie data, review movies, or engage with others in a Q&A forum. The site was created by the University of Minnesota GroupLens research lab and has been the basis of a large number of research papers [2]

2.1.4 Everything2

Everything2 is a peer production site that, while originally envisioned in 1999 as an open source encyclopedia, is now a writing site. People can write articles ("write-ups"), communicate with others, and create larger topic pages, called Nodes [91].

[1] http://www.cyclopath.org
[2] http://www.movielens.org

2.2 Core Contributors Exist

Much of what is discussed in this dissertation relies on the existence of core contributors. Many researchers have analyzed users by their level of participation. Often they find that there is a long tail distribution of participation [17] which often follows a power law [14].

2.2.1 Use of Term

People who use software and use lots of the special and complex features are often called "power users". On Wikipedia and other sites that allow editing, the users who contribute the most may also be referred to as "elite editors".

2.2.2 Wikipedia

There has been extensive work done that studies core contributors on Wikipedia. Much of the work discussed was done prior to the work presented in this dissertation.

On Wikipedia, the small group of elite editors completes the majority of the work [73] and also produces the majority of the value [113]. In fact, the top 0.01% of editors (by number of edits) contribute 44% of the value of Wikipedia [113]. Butler et al. argue that there is a hierarchy on Wikipedia (Readers, Editors, Administrators) and that this hierarchy exists in part due to bureaucracy created by core contributors [29].

Bryant et al. interviewed some expert Wikipedia editors about their habits as novices and their current habits. They found that novices transition into what the authors termed Wikipedians. As they gained experience, these editors use different tools and become more engaged in the community. They also start out seeing Wikipedia as a collection of pages and later view it as a community. Eventually these users aren't just part of the community, but they are members of an elite group within the community [26]. Additional research by Forte and Bruckman showed that as these Wikipedians get more involved in the community, they more easily recognize other members of the community. They also feel like they get respect and recognition from others [47].

Work by Krieger et al. found three types of Wikipedia editors: Beginning Editors, Advanced Editors, and Administrators [77]. The uniqueness of Wikipedia administrators is also studied by Burke and Kraut. In one of their studies, Burke and Kraut looked at what it took to become a Wikipedia administrator and found that it wasn't just edits, but rather that administrators needed to demonstrate diverse experience on Wikipedia and coordination efforts [27].

This divide also exists in WikiProjects. A study by Chen et al. found that an increase in tenure diversity led to improved group performance. That is a mix of newcomers and old timers was ideal, however high tenure diversity increases conflict [30].

2.2.3 Cyclopath

This dissertation focuses in Chapter 5 on the existence of core contributors on Cyclopath, but work has also been done by Masli et al. that shows that as Cyclopath editors get more experienced, they start to specialize in editing specific types of content.

2.2.4 Open Source Software

Open Source Software (OSS) is another area where there is evidence that core contributors exist.

Ghosh et al. repeatedly mention elite developers in the OSS community. They found that these elite contributors had more connections and worked on more projects than other users [53].

Nakakoji looked at OSS as having a hierarchy, not elite and non-elite. Users of software are the most prevalent, but as roles get more involved there are fewer people performing those tasks. From least to most involved, Nakakoji defined the groups as: Passive User, Readers, Bug Reporters, Bug Fixers, Peripheral Developers, Active Developers, Core Members, and Project Leaders [98]. Ducheneaut reports a similar structure [41]. While the definition of core contributors in OSS would be somewhere along this spectrum, it's unclear where, exactly, it would lie. What is evident is that there are contributors who do more than the average contributor.

2.2.5 Others

Core contributors exist in many other venues, such as Everything2. Researchers classified contributors to the site Everything2 into four groups based on their primary motivations. The groups were Status Builders, Personal Relationship Builders, Community Builders, and Human Capital Builders [129]. While not defining users as core contributors or not, this study focuses on types of core contributors.

Finally, Preece and Shneiderman defined their Reader to Leader Framework to help describe stages that users can go through on sites like Wikipedia. They describe the

majority of users as readers, but say that the minority go on to other positions to contribute, work with others, and create policy [111].

2.3 Core Contributors are Different

Once it has been established that core contributors exist, many researchers have studied various ways that core contributors may differ from other contributors. This ranges from contribution quantity to contribution quality or motivations.

2.3.1 Contribution Quantity

Across the web, user contributions often fit a power law distribution where the minority of participants provide the majority of the content. This has been observed in Usenet postings [46, 133] and Marks noted this same distribution when analyzing bloggers in 2003 [88]. This same power law distribution was also found in 2005 in del.icio.us tags by Golder and Huberman [56].

Over the life of Wikipedia, Wikipedia contributions have increased dramatically. Kittur et al. [73], Almeida et al. [16], and Ortega et al. [106] all found that contributions had increased. In addition the number of editors has also increased dramatically [123].

2.3.2 Contribution Quality

Contribution quality is another area in which core contributors are different. Unfortunately, because quality is hard to measure, it is not as frequently studied.

Quality definitions can vary according to context. On Slashdot, quality is determined by moderation [83] on sites like reddit, quality is determined by community vote. On Wikipedia, quality is determined by editors.

In an early study of Wikipedia, scientists judging the accuracy of scientific Wikipedia articles found that Wikipedia's accuracy rivaled the accuracy of similar articles in the Encyclopedia Britannica [55].

Another, more scalable measure of quality uses word persistence as a content quality proxy. The reasoning behind this is that if the article is not correct or has other problems, subsequent editors will make changes as necessary to "fix" the article. Similarly if an article is not changed by subsequent editors then that counts as approval from those other editors.

Adler and Alfaro developed a reputation metric that measured how long an editor's changes lasted over time [15]. Priedhorsky et al. found that the top 10% of editors (about 4,400 people) contribute 44% of the value of Wikipedia.

Since the work presented in Chapter Chapter 3 was completed more researchers have begun focusing on figuring out how to measure quality on Wikipedia. These metrics include both editor-based assessments (similar to the work presented in this dissertation) and article-based assessments where a score is assigned to articles instead of editors [132].

Like these previous researchers, in Chapter 3 we look at persistence of words from an editor's perspective.

2.3.3 Community Work

Another aspect that we looked at was community work. This could include maintenance, policy development, or other types of work that benefit the community.

Diversity and community norms influence the experience, especially for newcomers. Dugan et al. studied social network profiles and found that diverse profiles positively influenced the number of friends the user had [43]. Ducheneaut researched community norms within the Python developer community and learned that these norms affect the newcomer experience and joining the community requires completion of several rites of passage [41].

Within Wikipedia, much of the community work takes place outside of the encyclopedia articles in other "namespaces" which serve different purposes. Each content namespace has an accompanying Talk namespace designed for discussion and coordination. Researchers are very interested in Wikipedia Talk pages as they are often used for coordination and policy discussion [29]. Kittur et al. found that in 2001 (early in Wikipedia's life), 90% of edits were done on the main namespace of Wikipedia, the encyclopedia pages. But by June 2006, only 70% of the work was being done in Main [72]. In interviews with elite Wikipedia editors, Bryant et al. found that these editors broadened their interests over time and started to assume more community maintenance work [26].

Wilkinson et al. found that articles with more discussion on their Talk page were ranked higher in quality according to article ratings [135]. Viegas et al. dug into this further, looking at topics and themes on Talk pages and found that half of Talk page comments were requests for coordination and 8% of Talk page comments were policy invocation. Due to these findings, Viegas et al. believe that Wikipedia consists, in part, of a strong and supportive community.

In Chapter 3, we, like Kittur, look at the percentage of edits that are done in

different namespaces on Wikipedia, but we do not analyze the discussions on the talk pages. In our Cyclopath research, we focus more on the qualitative reports of community work as Cyclopath is structured such that there aren't any explicit community tasks to be measured.

2.3.4 Retention

Another important part of online communities is retention. Many online communities struggle not just with attracting members, but also with getting those members to stick around and keep contributing. This is an issue that is attracting attention in the Wikipedia research community in particular [58].

Zhang et al. found that the median lifetime of a Wikipedia editor was 53 days and that there are two critical time periods when editors are at higher risk of becoming inactive, 0-2 weeks and 8-20 weeks [136]. This is interesting and confirms some of the findings in Section 3.2.2. While this research is useful, the authors admit that while we can empirically measure retention and inactivity, we still don't know **why** Wikipedia editors drop out.

The Wikimedia Foundation also doesn't know why editors leave, but one of their suspicions is that newcomers are being rejected, either explicitly or implicitly, and they are wanting to change the newcomer experience, to increase both participation and retention. One of their efforts is the Teahouse [10], an area of Wikipedia that includes outreach and social support and is designed to be a nurturing environment for new editors. In an early iteration of the Teahouse, participants made more edits and were much more likely to keep editing Wikipedia than non-participants [94].

Also on Wikipedia, Chen et al. found that in the domain of WikiProjects, increased diversity in Wikipedia experience decreased member withdrawal, up to a certain extent. Differences in member interest decreased member withdrawal linearly [30].

Outside of Wikipedia, Dabbish et al. looked at turnover in groups of Tetris players and how that turnover may alter the community. As part of that research, the authors found that retention was positively influenced by turnover when there was a common group identity. In particular, the study by Dabbish et al. suggested that turnover increased social presence which led to increased participation [40].

2.3.5 Motivations

Open collaboration communities exist and have value only because people add content. Therefore, issues of motivation are critical. Researchers have been studying motivations within open collaboration communities since this type of community

appeared, despite the fact that motivation is hard to understand and requires more than log data. Much of the research has been done within the OSS and Wikipedia communities.

Some offline research has also been influential in the online realm. Clary et al. studied motivations for people to volunteer (offline). These motivations are listed and briefly described below.

- Values: Allows for expression of altruistic and humanitarian concerns
- Understanding: Allows for new experiences and use of existing skills and abilities
- Social: Allows for opportunities to spend time with friends or to participate in an activity that is valued by others
- Career: Allows for preparation for new career or maintenance of existing career-related skills
- Protective: May reduce guilt over personal fortune and allow for dealing with personal problems
- Enhancement: Allows for maintaining or enhancing positive affect

In addition to identifying these motivations, the researchers created and validated a tool, the Volunteers Function Inventory (VFI), the Volunteer for assessing motivations. They found that volunteers who receive individual benefits are more satisfied and likely to continue volunteering than those who felt they did not personally benefit [34].

A follow up study by Fugelstad et al. administered the VFI online to users signing up for MovieLens. They found that self-oriented motives are negatively correlated with participation on MovieLens [52]. However, it is possible that MovieLens isn't a site that would match with certain motives.

OSS

In a survey of Wikipedia and OSS contributors, Oreg and Nov looked for differences between the two groups as well as relationships between contributors' psychological dispositions and motivations. The survey of 185 SourceForge participants used the portrait values questionnaire and found that the top motivation was self-development, followed by altruism and reputation building [105].

An earlier study by Lakhani and Wolf surveyed participants on SourceForge and found that while a minority of developers were paid, 55% of developers contributed

during their day jobs with about 40% of the population contributing during their day job with their supervisor's knowledge. Paid developers averaged four hours a week more of OSS work than volunteers. The top motivations for developers were enjoyment, improving skills, ideological reasons, an obligation to give back, and a strong sense of group identity in the hacker community [81].

Finally, Hars and Ou took a similar approach to Lakhani and Wolf and surveyed 79 people on OSS mailing lists and newsgroups. They found that 88.3% said they were motivated by skill improvement, 79.7% had self-determination as a motive. In addition, 43% were motivated by peer recognition. The study distinguished between internal and external motivations, where external motivation included future rewards and personal need [62].

Wikipedia

One of the most cited studies of the motivations of Wikipedia editors was done by Nov who used a modified version of the Volunteer Functions Inventory (VFI) created by Clary et al.. The VFI was modified to include factors of fun and ideology Nov emailed requests to participate to a random sample of 370 Wikipedia editors and received 151 valid responses. He found that fun was correlated with higher (self-reported) Wikipedia participation, but fun and adherence to principles of open source were the top motivators. These were followed by values, understanding, enhancement, protective, career, and social in that order. Unlike with the OSS community, career motivations were not common [102].

Other research into Wikipedia motivations focused more on qualitative responses from non-editors. Kuznetsov surveyed students at New York University and asked them why they thought Wikipedia editors participated and the students said altruism, reciprocity, community, reputation, and autonomy. These students had also all indicated that they'd be willing to contribute to Wikipedia [80].

Finally Antin interviewed Wikipedia readers to learn more about who the readers think the editors are and why they think they edit. He found that the readers assumed that the editors' motivations included giving back to the community and gaining intrinsic rewards [18].

Other Systems

Motivations have also been studied in other online communities. Researchers have surveyed users on the online encyclopedia and creative writing site, Everything2. In that research, they found that the top theory-derived motives for participating in the site were entertainment, sense of belonging (to the site), and value from adding information to the site [84]. Following that research, the researchers conducted

semi-structured interviews with Everything2 contributors and found four types of motivation. These were: increasing their status on the site, building relationships with other users, improving the community and increasing it's value as a resource, and building their individual skill sets [129].

The site Kassi is a Finnish social exchange site where users can request and perform favors as well as give, trade, or sell goods. Suhonen et al. found that users came to the site for fun, but frequent users added information to help others. Many of the contributors also mentioned reciprocity as a reason to participate [124].

Finally, Fugelstad et al. did a study on MovieLens using a slightly different modification of the VFI and found that on MovieLens, self-oriented motivations were negatively correlated with logins and ratings [52].

Summary

While there are different factors influencing motivations for contributing to different online communities, studies in new domains, such as Cyclopath, can help to extend prior work and build better generalization. In addition, motives can help us understand better techniques to increase participation and contribution, especially among users who aren't actively contributing.

2.4 Consumers

In addition to contributors, core and otherwise, consumers also make up an important group of users in almost any online community. Online communities and social media have been studied extensively, from both the contributor and the consumer perspective, in part due to the great resources that they leave behind. These rich activity traces may include: messages posted and replied to, user profile settings, and Wikipedia policies debates and articles edited. Researchers have analyzed this data to draw pictures of online life, such as the dynamics of large scale discussions [66, 133] or the effectiveness of distributed moderation for information filtering [83]. This section explores research on consumers, including both quantitative analyses of some of the aforementioned activity traces and analyses of qualitative data.

One of the more pervasive notions of consumers is that they are lurkers. While the original definition of a lurker is "one who lurks or lies concealed," in both the original usage and the usage within the computer culture, lurkers and lurking were viewed as negative [3, 100].

In 1996, Kollock and Smith discussed the notion of lurkers as free riders. They were the people who posted questions but never wrote answers, people who gathered

information and never distributed it, who read and didn't respond. In the opinion of Kollock and Smith, this was dangerous and may lead to the failure of entire communities [74].

In 1998, Katz wrote a commentary on Slashdot which described lurkers similarly as people who are vocal in 1:1 email discussions, but not comfortable in a public space. He also thought that lurkers get the best of the web, the content, and skip the remainder, which he described as insults and attacks. But, he then introduced some mechanisms which he believed could help get the lurkers more involved, namely the banning of anonymous posting, moderated comments and discussion, and a ban on insults and attacks [68].

Towards the end of the 1990s, there started to be a trend away from the more negative notions of lurkers and towards the notion of Legitimate Peripheral Participation or LPP. This concept was introduced by Lave and Wenger [85]. Their work investigated how people enter "communities of practice," such as midwifery or butchery. The notion was that these communities usually have a type of apprenticeship to help newcomers understand the norms and community standards. As the newcomers come to know the community, they participate more.

The seminal research on consumers as lurkers and LPP was conducted by Blair Nonnecke and Jenny Preece at the University of Maryland in the late 1990s and early 2000s. Much of their work attempted to quantify the lurkers and understand how their presence influenced the contributors.

Nonnecke and Preece found that lurking was more common in some communities than others. In particular, medical discussion lists had fewer lurkers than software support discussion lists. Lurking was also more common in larger discussion lists. However in this work, Nonnecke and Preece discussed lurking not as free-riding, but rather as an acceptable and beneficial behavior, legitimate peripheral participation as introduced by Lave and Wenger [85]. The researchers also believed that value derived from the discussion lists might spread to others through the lurkers as well as the posters [100].

Two of their 2004 papers studied online mailing lists, looking in particular at non-active contributors. Their survey was taken by 218 lurkers and 970 posters. They found that both the lurkers and the posters came to the message boards with a goal of getting a general understanding of the topic. The lurkers weren't posting because, for 53.9% of them "just reading/browsing is enough". They also mentioned that they were still learning about the group, shy about posting, or thought that they had nothing to offer. The expectations of the posters were met better than the expectations of the lurkers. The posters felt that they received more benefit from the community and had a greater sense of membership. The lurkers had less respect for the posters than vice versa, but the posters were more likely to consider the lurkers to be members of the community than the lurkers themselves were [101].

The same survey also found that the top five reasons why consumers lurked on MSN bulletin boards were:

- They didn't need to post
- They wanted to find out more about the group
- They thought they were being helpful by reducing clutter or noise in the group
- They couldn't make the software work
- They didn't like the group

After seeing these reasons, Preece and Nonnecke suggested several changes. Better usability support for newcomers, help for new users to cope with the high volume of messages, ability for posters to retain anonymity, ways to eliminate fear of insults and aggression, and some specific measures to encourage newcomers to post. These were moderators, policies for posting, and rewards for posting [110].

While much of the work on consumers has come from Nonnecke and Preece, others have also done some research on the topic. In particular, there is part of this work that focuses within enterprise communities, in part because consumer data is likely more readily available in those systems. Soroka et al. studied lurking within enterprise communities and found no evidence of legitimate peripheral participation (also known as educational lurking) within that community. They did find, however, that lurking was more common among non-native English speakers in the enterprise community of an American company [122].

Subsequent work within an enterprise file-sharing community by Muller et al. classified users into three groups. Uploaders uploaded files, contributors added metadata, shared files, and created file collections, while consumers used the metadata and/or files. Users were classified by the most involved actions they took, so uploaders may also have contributed or consumed, but were classified as uploaders. The research found that users who uploaded were the most active consumers. In addition, users could be part of different user groups and the research found that while in different user groups, users displayed different behavior, even when lurking [96].

Muller later did another study within an enterprise community looking at different sub-communities within the system. He was interested in whether lurking was a personal trait or part of situational disposition. The research found that the majority of contributors lurk within other sub-communities. This argued against the notion of lurking as a personal trait and for the idea that personal traits that may be altered by the attachment a user has to the community in question. Like Soroka et al., Muller found little support for social learning [95].

More generally, Preece and Shneiderman's reader to leader framework sees activity as a spectrum in which everyone has to start as a reader, and may or may not advance to other activity. They acknowledge that users anywhere on the ladder from reader to leader are "active participants" and that readers have value, though Preece and Shneiderman imply that the primary value of readers is their potential to contribute [111].

Within Wikipedia, Antin and Cheshire found that not all Wikipedia readers are free-riders, as is sometimes thought. Rather, they argue, reading is a form of legitimate peripheral participation and leads to more active participation. Most readers, on a sample survey, indicated that they do not have complete operational knowledge of the site and reading helps them learn about Wikipedia. Finally, Antin and Cheshire argue that there are many paths that can lead to active engagement, there isn't one right path [19].

Antin and Cheshire aren't alone in seeing ways that consumers can add value to a community. Both Muller et al. and Takahashi et al. found that readers propagated information from within a community to outside a community [19, 127]. Ducheneaut et al. also felt that the readers added value to a community precisely because they were able to serve as an audience to those who contributed [42]. Finally, in a study within MovieLens, Harper et al. found that invitations to participate changed users' message reading behavior within the site [60].

2.5 Conclusion

Despite so much research on core contributors and consumers, there are still a number of questions left unanswered. In the following chapters, I will attempt to answer several of these questions, focusing on quantitative differences between average contributors and core contributors (and occasionally consumers) as well as the motivations behind their behavior. Finally, Chapter 7 focuses on core contributors and their experience and development on Cyclopath, similar to the Wikipedia work presented by Bryant et al. in [26].

Chapter 3

Wikipedia and Core Contributors (Quantitative)

We conducted our first research project on Wikipedia, in part because Wikipedia is one of the largest, most successful online communities. When this research was done, Wikipedia was starting to become a popular research platform, due to their open data policy. However, there was also an ongoing debate as to whether work in Wikipedia is done by a small group or the general populous [16, 125, 131, 135]. Since this debate started, researchers found that a small percentage of the editors does the majority of the work [73, 72] and that a small percentage of these editors are the ones producing most of Wikipedia's value [113]. But despite this work, we don't know much about who the editors are and how they work.

This study builds on a previous study by Bryant et al. [26]. They interviewed nine Wikipedians (editors with edits in the thousands) to learn more about the transition from novice editor to expert editor. They specifically looked at how editors branched out to new area and topics, how editors took on more community work, and how the editors' views and thoughts about Wikipedia changed as they gained experience.

We wanted to make use of the massive data logs available from the Wikimedia Foundation to investigate quantitatively some of the claims from the small sample in [26]. This required us to quantify the findings from Bryant. We decided to look at the quantity of work performed, the quality of the work, the work done for the community, and how these three things may change during an editors' lifespan. We decided that looking at the Wikipedia core contributors (in our case Wikipedia editors who had made over 250 edits Section 3.1.2) wasn't enough, we wanted to see if the trends from the Wikipedia core contributors were different from the trends from the rest of Wikipedia editors.

Our goal in this work was two-fold. First, we wanted to investigate whether the

claims from [26] are generalizable to a larger population of Wikipedia editors and second, to see whether the findings from [26] represent differences between core contributors and average editors.

3.1 Data and Methods

3.1.1 Data

This study was conducted in the Fall of 2008. We used the most up-to-date version of the English Wikipedia dump from January 13, 2008.[1] Wikimedia no longer hosts this dump and as of this writing, the dump was no longer publicly available. This dump contained all namespaces for the English Language version of Wikipedia.

When doing the analysis, we excluded edits made by known bots because we wanted to study how the users, not bots edited. We also excluded edits made with AutoWikiBrowser (AWB) which was a semi-automated editor for conducting repetitive tasks. AWB edits were excluded because AWB made abnormal numbers of edits (up to three a minute or 500 edits in 20 hours). Finally, we excluded edits from anonymous editors, because we couldn't track anonymous editors between IP addresses and we wanted to get as complete a view of user activity as possible.

The Wikipedia dump was parsed and dumped to a database. We the excluded the groups of users mentioned above. Charts and statistics were calculated using R.

3.1.2 Who did we count

Wikipedian is a term used within the Wikipedia community. Depending on the use, it can mean any registered editor or a small, select subset of editors. The study conducted by Bryant et al. used the term to include editors who were active for an average of 14 months. In addition, these editors self-reported daily (or near daily) activity [26]. For this study we use the term core contributor to better imply who these users are.

We defined a **Wikipedia core contributor** as an editor who had made at least 250 edits during the course of their lifetime on Wikipedia. Why 250 edits? When this research was conducted, 250 edits was the threshold to use tools such as VandalProof[2]. In addition, 250 edits were half the number of edits required to gain access to the AWB tool.

[1] http://download.wikimedia.org/enwiki/20080113/enwiki-20080103-pages-meta-history.xml.7z Please note that the link is no longer accessible.

[2] http://en.wikipedia.org/wiki/User:AmiDaniel/VandalProof

In order to test if our definition was valid, we varied the threshold from 50 edits to 5000 edits and ran the analyses presented in Section 3.2 to see if trends held. They did, with one exception. When we made the threshold 5000 edits or more, the patterns did change. We present the findings with additional granularity, in part to demonstrate these differences.

We also varied the perspective to see what effect editing lifespan might have. For example, one of our groups was editors with one to two years between their first and last edits.[3] When we ran these analyses, we found that editor activity was similar to the activity of average editors.

As mentioned in the previous section, we only looked at the activity of registered editors where Wikipedia had linked the edit activity to the account name. (If a registered editor is not logged in, edit activity is linked, instead, to their IP address.)

Our analyses include data from all 37,956 users who met our definition of core contributor as of the dump on January 13, 2008. Our average editors are a random sample of 38,975 non-core contributors. These numbers are different because we chose our sample before we cleared hidden bots and AWB users from the Wikipedia core contributors sample. At that time, the computational time required to redo the analyses outweighed the benefits we thought we would receive.

3.1.3 What did we count

Our basic unit for this research is edits per day per editor. Each editor's edits are grouped according to when his or her first edit occurred. If an editor made her first registered edit at 11:15am, we consider her second day of editing to start 24 hours later, the third day 48 hours later, etc. We refer to these as days of the editor's life, although technically they are really the day of the life of the editor's account.

However, it is possible that the editor could have made anonymous edits before registering or that the editor could have made edits with another account. Therefore the first edit by our metric may not be the first edit. We do not rely on an assumption of when the user began editing so this doesn't invalidate any of our results.

Note that we did consider an alternative to the edit: the edit session. The idea behind this alternative is that some editors might make multiple edits to the same article in quick succession. We formalized this as multiple consecutive edits to a single article, with no edits to other articles and no edits by others to the said article, within one hour. This may occur when editors use the save button to preview their edit or after saving their edit realize they made a mistake. Seventy-five percent of

[3]The last edit was the last edit by that editor in our dataset. Note that for many editors this will not actually be their last edit.

editors had an average of 1.4 edits per session or fewer. Due to the rare occurrence of this, we chose, instead, to use the edit as our basic unit, not the edit session.

3.2 Findings

3.2.1 Presentation of Results

All the graphs presented in the results share several key characteristics. First, days of editors' lifespans are on the x axis. The x axis is on a log_2 scale so that differences in the early days are distinguishable. The graphs are presented in color to make them easier to understand. Captions may refer to colors, but the goal is that they are still comprehensible in black and white. Error bars are shown (unless mentioned otherwise) but often errors are so small that the error bars are not visible in the graph.

In some of the graphs, Wikipedia core contributors are shown in black and average editors are shown in red. The other graphs have further breakdowns which are explained in the caption and surrounding text.

When we present graphs with daily averages, those averages include editors who did not make any edits on that day but did a make subsequent edit. After the last of their edits in our data set, they were no longer included in the daily averages. Figure 3.1 shows the percentage of editors who make an edit on any given day.

3.2.2 Wikipedia Core Contributors and Work Quantity

We are interested in looking at the quantity of edits made by Wikipedia core contributors even though we defined Wikipedia core contributors as editors who did more work than other editors. In particular, we wanted to see if the amount of work done changes (or stays consistent) over the editors' lifespans. Is there a difference (on a daily basis) between the quantity of edits by Wikipedia core contributors and that of average editors? If so, does it change or stay constant? We also wanted to learn what early days of editors' lives tell us about their prospects for later activity.

Figure 3.2 shows the number of edits per day for all editors. Both Wikipedia core contributors and average editors start editing with a burst and tail off quickly to a constant level. The Wikipedia core contributors do more work (throughout) than average editors. On their first day of editing, Wikipedia core contributors make 15.1 edits and average editors make 3.5 edits. After two months, Wikipedia core contributors average 3 edits a day and average editors average 0 edits a day. The

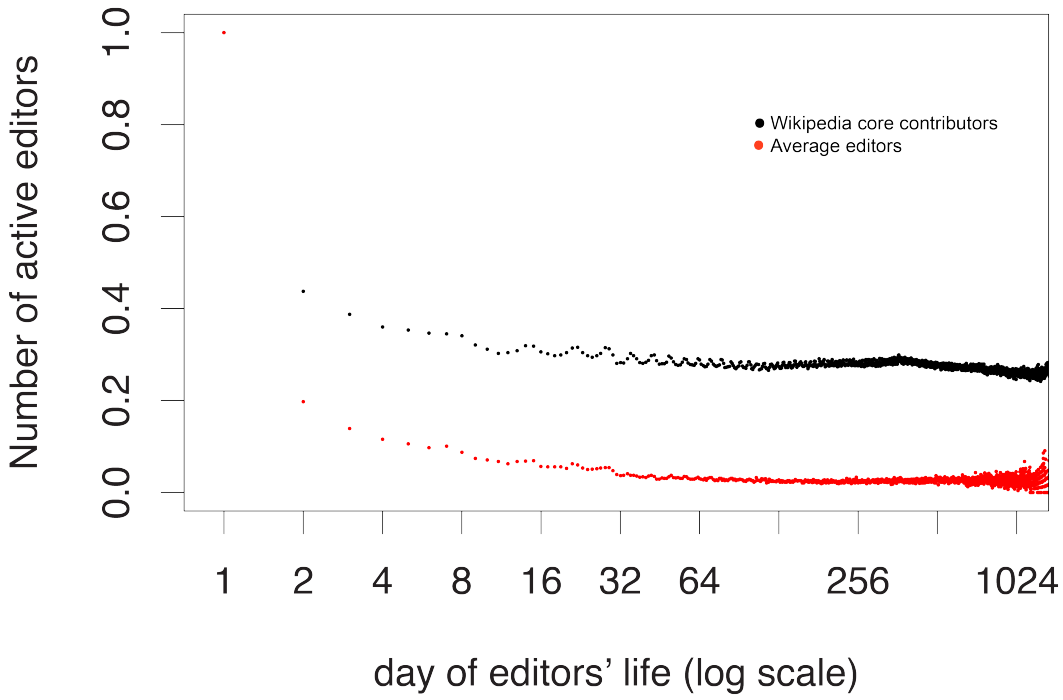


Figure 3.1: Percentage of users who made an edit on the given day. The top black scatter shows the Wikipedia core contributors , while the bottom red scatter shows average editors. Both populations start at 1. Standard error bars are displayed.

pattern we saw was that new Wikipedia editors trying out Wikipedia either do a lot of work over time or do a little work and then lose interest.

This discontinuity can be seen even more clearly in Figure 3.3. In this figure, we've broken down Wikipedia editors into more finely grained segments based on activity. The divisions we use are based on work by Kittur et al. [73] and Ortega and Barahona [106]. Editors are broken down to those who have made less than 100 edits, editors making 101 to 1,000 edits, editors making 1,001 to 5,000 edits, editors making 5,001 to 10,000 edits, and those making over 10,000 edits.

Figure 3.3 shows the average number of edits per day with the aforementioned division of editors. All groups of editors making 5,000 or fewer edits a day showed a pattern similar to that seen in Figure 3.2. The most active editors (3,600 editors making over 5,000 edits) showed a different pattern. Editors who made between 5,001 and 10,000 edits stayed consistent for their first year. Editors making over 10,000 edits increased their activity from the first month through the second year.

Section 3.2.2 shows a different analysis to help understand quantity differences between Wikipedia core contributors and average editors. It indicates that the number of edits made on the first two days is a strong predictor of whether an

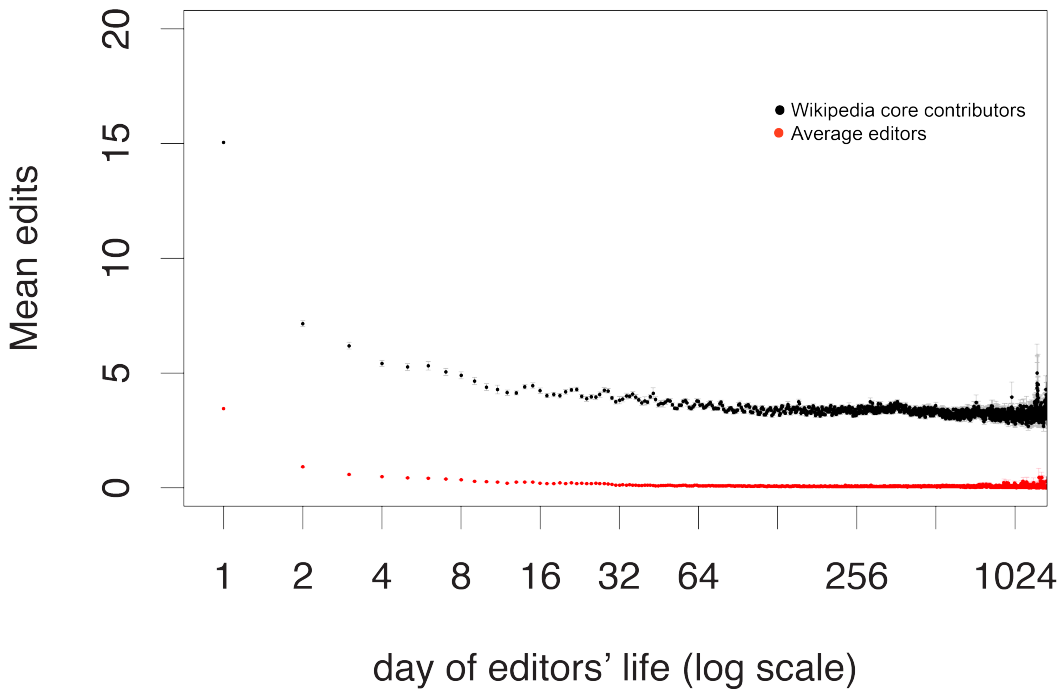


Figure 3.2: Average number of edits by day across all namespaces. The top black scatter represents Wikipedia core contributors and the bottom red scatter represents average editors. Standard error bars are displayed.

editor will become a core contributor. More specifically, fewer than 1% of those making a single edit on their first day will become a core contributor. However 4.5% of those making six to ten edits and 8.27% of those making 11 to 20 edits on the first day will become a core contributor. Looking at how many edits users made in their second 24 hours as editors is even more telling. A single edit between 24 and 48 hours of editing yields almost 6% probability of becoming core contributor and making between six and ten edits in this time period increases the probability to 18%. As a group, Wikipedia core contributors edit 28% of the days between their first registered edit and their last viewable edit. The average group was active only 4.3% of the days.

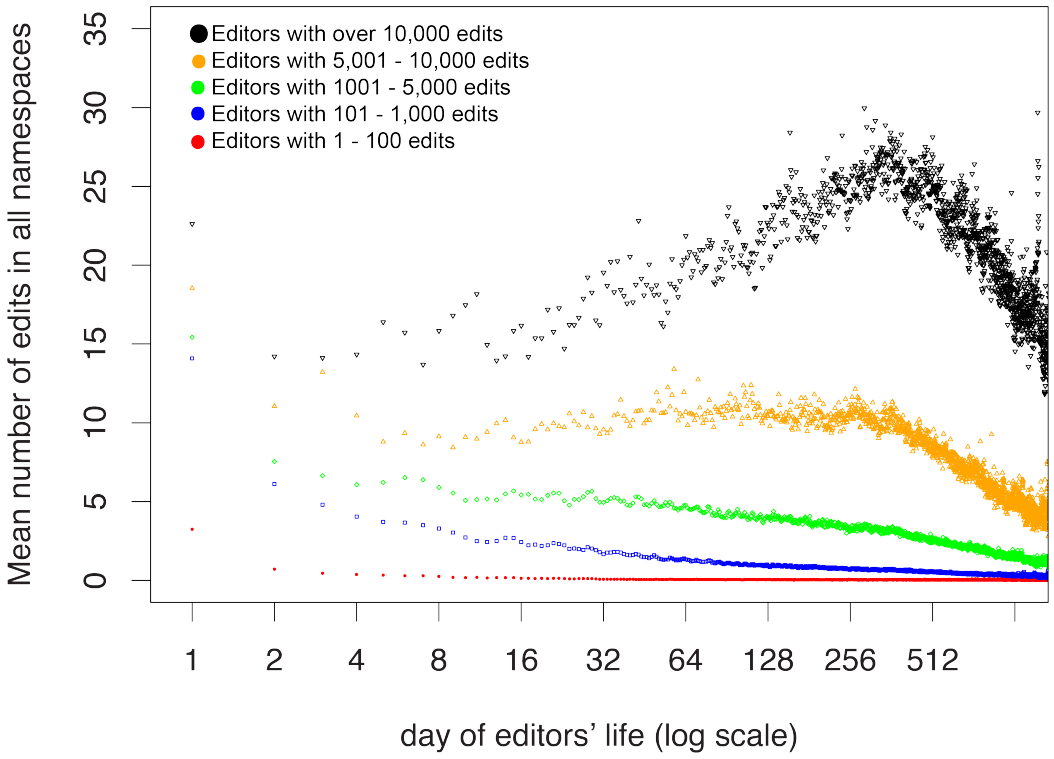


Figure 3.3: Number of edits by day across all namespaces. The top black scatter represents editors with over 10000 edits, the next down orange scatter is editors with 5001-10000 edits, green scatter in the middle is editors with 1001-5000 edits, blue scatter near the bottom is editors with 101-1000 edits, and the red scatter at the bottom is editors with 1-100 edits. Standard error bars are not displayed.

3.2.3 Wikipedia Core Contributors and Work Quality

If Wikipedia core contributors didn't increase the amount of work they did, perhaps they increased the quality of the work they did.

To investigate this, we needed a metric for quality. Unfortunately no universal quality metrics existed at the time for Wikipedia articles. There were, however, some candidates. The most obvious candidate was the Wikipedia assessment ratings. These are ratings from the Wikipedia community themselves and rank pages anywhere from stub (barely an article) to featured article (a comprehensive and complete article). These ratings were used as a quality metric by several researchers [71, 72, 135]. Another option was human coding. This was the approach used by

Number of Edits	First Day		Second Day	
	Proportion of Editors	Likelihood of being core contributor	Proportion of Editors	Likelihood of being core contributor
0 (no subsequent edits)	NA	NA	62.21%	.00021%
0	NA	NA	29.74%	4.47%
1	45.57%	.95%	2.84%	5.62%
2-3	29.11%	1.62%	2.21%	8.83%
4-5	10.25%	2.72%	.93%	12.83%
6-10	8.67%	4.50%	.98%	18.19%
11-20	4.16%	8.27%	.61%	26.75%
21-40	1.59%	15.42%	.33%	37.29%
Over 40	.65%	31.11%	.15%	64.82%

Table 3.1: Likelihood, based on first two days of edits, of a user making a given number of edits and becoming a core contributor. Results are not cumulative and are independent.

Giles when comparing the quality of Wikipedia to the quality of the Encyclopedia Britannica [55]. However human coding is very time intensive.

A third type of quality metric used persistence as a proxy for quality. Similar metrics were used by Adler and Alfaro [15] as well as Priedhorsky et al. [113]. The idea behind these metrics is that the longer the editor's content has lasted and the more of it that has lasted, the higher the quality of the edit.

We decided to use persistence as a proxy for quality. We chose this metric because we wanted to use an editor (not reader) based metric and we wanted to be able to assign and calculate a value for each word added. The Wikipedia assessment ratings (used by Kittur et al. [71, 72] as well as Wilkinson and Huberman [135]) are internal to Wikipedia. In addition they are a reader based metric and the ratings are actually displayed to users. This model is based on the peer review process and ratings are assigned per article, not per edit or per word.

Priedhorsky et al. in [113] created a persistence metric called **persistent word view** or PWV. This was a reader based metric and measured the number of views any given word received. Adler and Alfaro developed an editor based metric that assigned value to words based on how long they lasted [15]. In addition to using the life of the text to calculate the metric, Adler and Alfaro also calculated reorganizations and deletions [15]. Our approach is more basic, in part because we were less concerned about the role of conflict and editor wars.

The metric we used is one that we developed that we call **persistent word revisions** or PWR. Our assumption is that if subsequent editors don't delete words, the words have been approved. This is an editor based metric that is assigned per word. One potential limitation of this metric is that editors don't get points for removing "bad" words, only for adding lasting words. Despite this limitation, the metric still

provided useful information about the quality of work done by editors: namely how much content they added, and how long it persisted in the face of other edits.

How the PWR Metric Works.. Using the sample revisions in Table 3.2, the first thing that we do in processing is that we remove all stop words (a, the, and, for, etc.). To calculate Steve's score, we know that he adds four words, but only three are counted as 'are' is a stop word. He gets one point for every revision that each word lasts. He gets zero points for 'blue', four for 'apples', and three for 'yummy', giving him a net score of 11 PWR. Similarly, Chris has a score of zero since he didn't add any words. Paul has a score of one as 'certainly' lasts for one revision. Robin has a score of zero as 'most' doesn't last. We can't actually calculate a score for Phil because we can't see any subsequent revisions.

Revision	Editor	Text
1	Steve	blue apples are yummy
2	Chris	apples are yummy
3	Paul	apples are certainly yummy
4	Robin	apples are certainly most yummy
5	Phil	apples are nutritious

Table 3.2: Example Revision History

We did an analysis showing that if a word lasts for at least four revisions, 91% of the time it will last for at least 10 revisions and 65% of the time it will last for at least 50 revisions. Therefore, to compute the final score, we count the proportion of the editor's words that last at least five revisions.[4] We also only analyzed edits in the main namespace, the encyclopedic content space on Wikipedia.

Figure 3.4 shows that Wikipedia core contributors do make higher quality edits than average editors. This advantage is large. In a steady state, Wikipedia core contributors average nearly 0.9 PWR where average editors average about 0.7 PWR. However, despite this difference, quality does not increase over time. In fact, Figure 3.4 shows that quality decreases slightly over time.

As we did with quantity, we conducted the same analysis using the buckets proposed by Kittur et al. [73] and Ortega and Barahona [106]. The graph is presented in Figure 3.5. This shows that PWR for users with 100 or fewer edits is similar to the PWR of non-Wikipedia core contributors as seen in Figure 3.4. As one might expect, the average PWR for users with 101 to 1,000 edits is between the users with 100 or fewer edits and users with over 1,000 edits. However, by day 64, editors with 1,001 to 10,000 edits have a slightly higher PWR than editors with over 10,000 edits.

[4]For this analysis, we only considered revisions that would last at least five subsequent edits. e.g. There had to be at least five edits after the edit in question for us to score it.

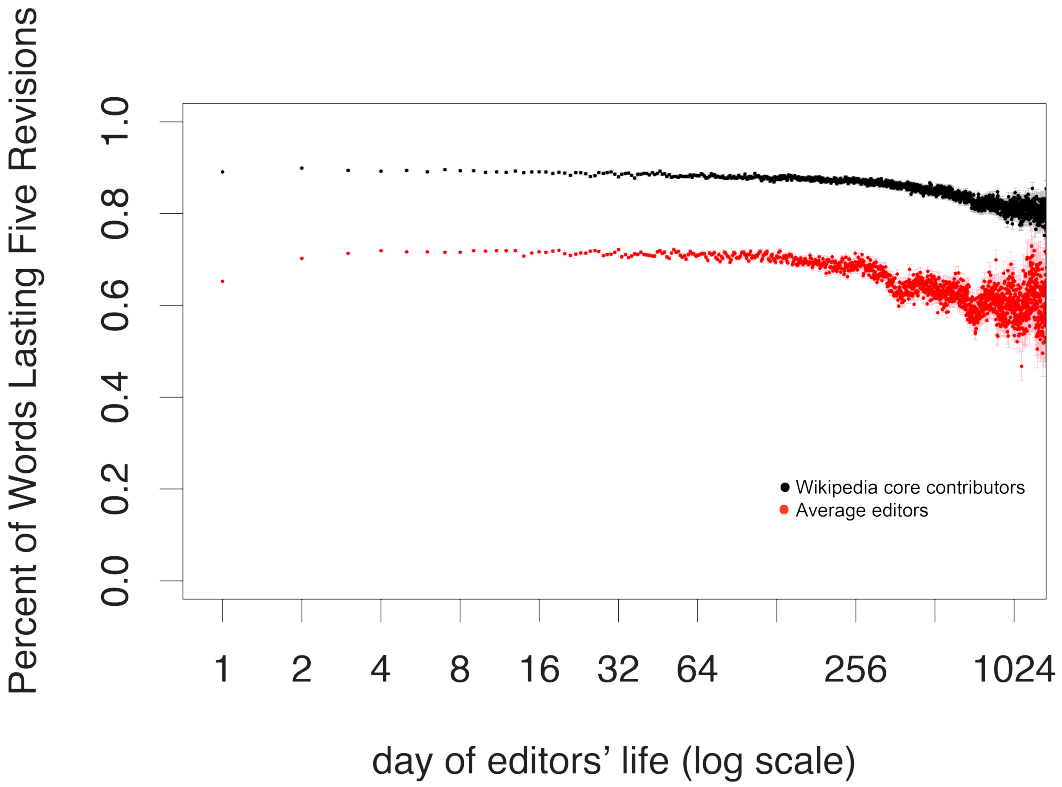


Figure 3.4: Wikipedia core contributors produce higher quality edits than average editors. The top black scatter represents Wikipedia core contributors and the bottom red scatter represents average editors. Standard error bars are displayed.

3.2.4 Wikipedia Core Contributors and Community Work

One of the findings by Bryant et al. was that Wikipedia editors began careers by editing content on topics they knew about. Then, as they aged, they did more community maintenance editing [26]. As with several of their other findings, we quantified these assertions in order to test them. To look at different types of community work, we broke these assertions down. We looked at what namespaces were edited and what percentage of a user's edits explicitly referred to community norms.

Namespace Diversity

Wikipedia has nine publicly editable content namespaces and nine publicly editable communication namespaces. Each of these serves as the location for a different type

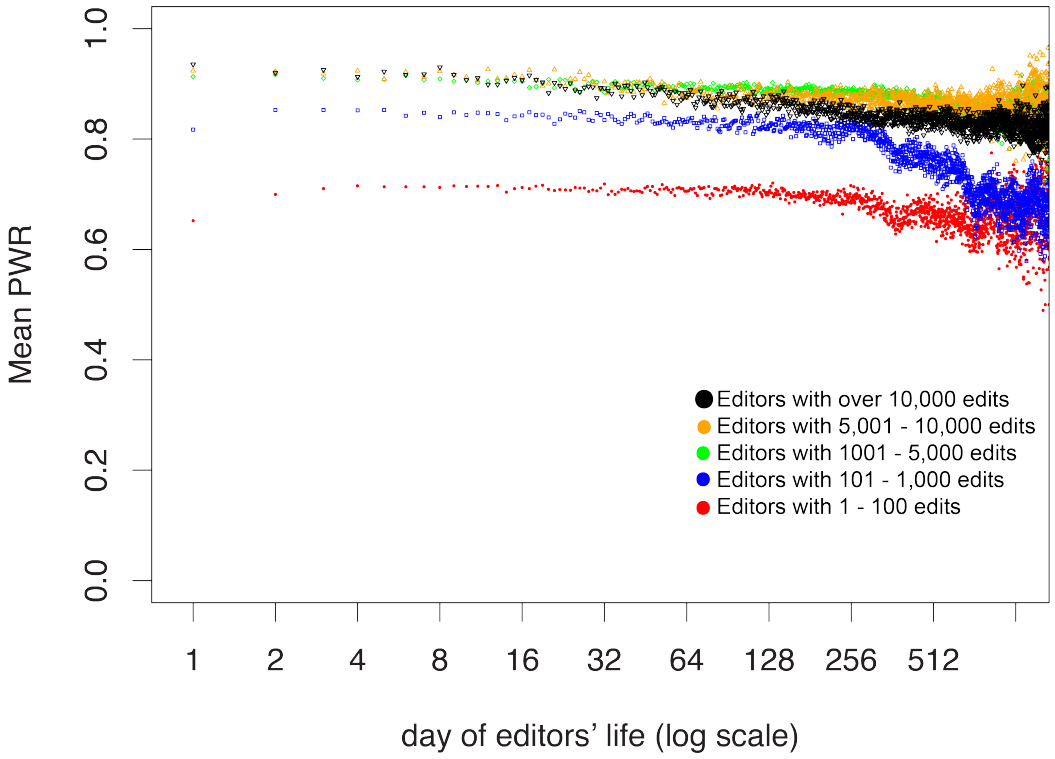


Figure 3.5: Wikipedia core contributors produce higher quality edits than average editor. The top black scatter represents editors with over 10000 edits, the next down orange scatter is editors with 5001-10000 edits, green scatter in the middle is editors with 1001-5000 edits, blue scatter near the bottom is editors with 101-1000 edits, and the red scatter at the bottom is editors with 1-100 edits. This graph is best viewed in color. Standard error bars are not displayed.

of activity. In this work, we looked at three namespaces in particular: Talk, User Talk, and Wikipedia. These namespaces were chosen because they are areas that center around community maintenance activity.

Talk pages are primarily for discussing the content of the encyclopedic articles in Main. There is one Talk page per article in Main and the Talk page is often used for questions and conflicts about facts or for requests for input. Any other activity about content articles would also fit under Talk. User Talk is for conversations about or with editors. Each registered user of Wikipedia can have a User Talk page. This is where messages to welcome new users are posted as well as warnings if the user has an edit that is being deleted for vandalism. These pages can also be used to

discuss edits made by that user and to acknowledge contributions by that user. The Wikipedia namespace is part of the governance of Wikipedia. These pages are where policies are formed and enforced, where users can request to become administrators (a powerful status editors can attain to better the Wikipedia community), where votes are held to elect administrators, and where votes are held to ban editors.

Of these namespaces, 66% of all edits were in Main, 9% of all edits were in Talk, 8.4% of all edits were in User Talk, and 6.5% of all edits were in Wikipedia. These are the top four namespaces and together account for almost 90% of all Wikipedia edits.

Bryant et al. [26] would be supported if Wikipedia core contributors shifted namespaces over time. We expected that as they aged, they would edit less in Main and more in other namespaces. In particular, we expected core contributors to be more active in other namespaces than average editors and for average editors to do almost no community maintenance work (represented in this analysis by edits in the Wikipedia namespace).

Related work by Burke and Kraut studied how likely users were to be selected as administrators. They found that one of the best predictors of promotion was the breadth score, their proxy for diverse experience. This metric gave contributors points based on how many different areas they edited in (out of the 16 they defined). A user who edited a Main page, a User page, and posted in the community newsletter (the Village Pump), would have three points [27].

Results.. The graphs of activity in Main (Figure 3.6), Talk (Figure 3.9), User Talk (Figure 3.8), and Wikipedia (Figure 3.7) don't seem to support these hypotheses. In fact, when we look at the proportion of edits that Wikipedia core contributors and average editors do in Main (Figure 3.6), we find that Wikipedia core contributors do decrease (slightly) the percentage of work done in main. However this means that Wikipedia core contributors can only slightly increase the percentage of work they do in other namespaces.

In Figure 3.8 and Figure 3.7, Wikipedia core contributors show an increased proportion of activity in User Talk and Wikipedia very slightly, but they do devote a larger proportion of their work to these namespaces than average editors. However Figure 3.9 is puzzling. Talk is the area where discussions of content occur, where coordination and collaboration are managed [72, 130]. In addition, prior research found that the more discussion on the Talk page, the higher the quality of the Main article will be [135]. Nevertheless, our data shows that Wikipedia core contributors do not increase the proportion of work they do in the Talk namespace, nor do they do a higher proportion of editing here than average editors.

As with previous analyses, we also looked at the editors in different buckets. Figure 3.10 shows the raw number of edits in Talk per day for each group of editors.

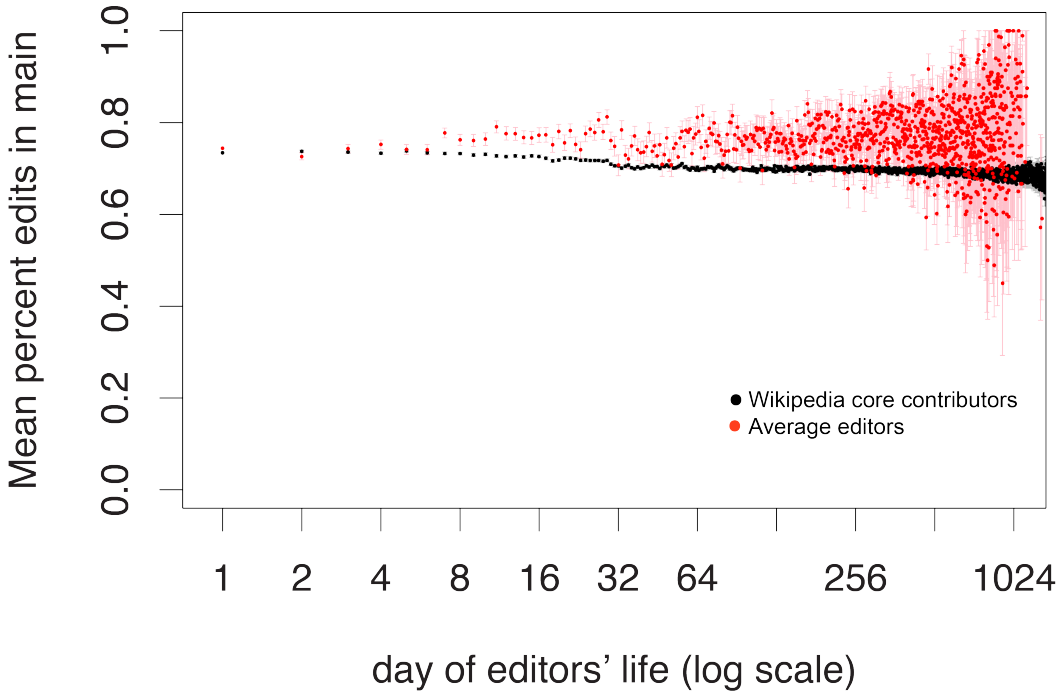


Figure 3.6: The percentage of edits in "Main" for all editors. The lower black scatter represents Wikipedia core contributors and the upper red scatter represents average editors. Standard error bars are displayed.

(Note that this is not a percentage of edits as with other figures in this section, due to decisions made during the writing and the current inability of this same dataset.) However, with this view it is clear that the most active editors do increase the raw number of edits in Talk per day. Still, looking at Figure 3.9 they do not increase the proportion of their work in Talk pages.

Invocation of Community Norms

The other aspect of community work that we were interested in was how the community of editors has evolved norms to govern Wikipedia. For example, the community has created rules like NPOV (Neutral Point of View) indicating that articles in Wikipedia should be free of bias and written from a neutral perspective. The community of editors is also involved in seeking out and reverting vandalism, keeping Wikipedia clean.

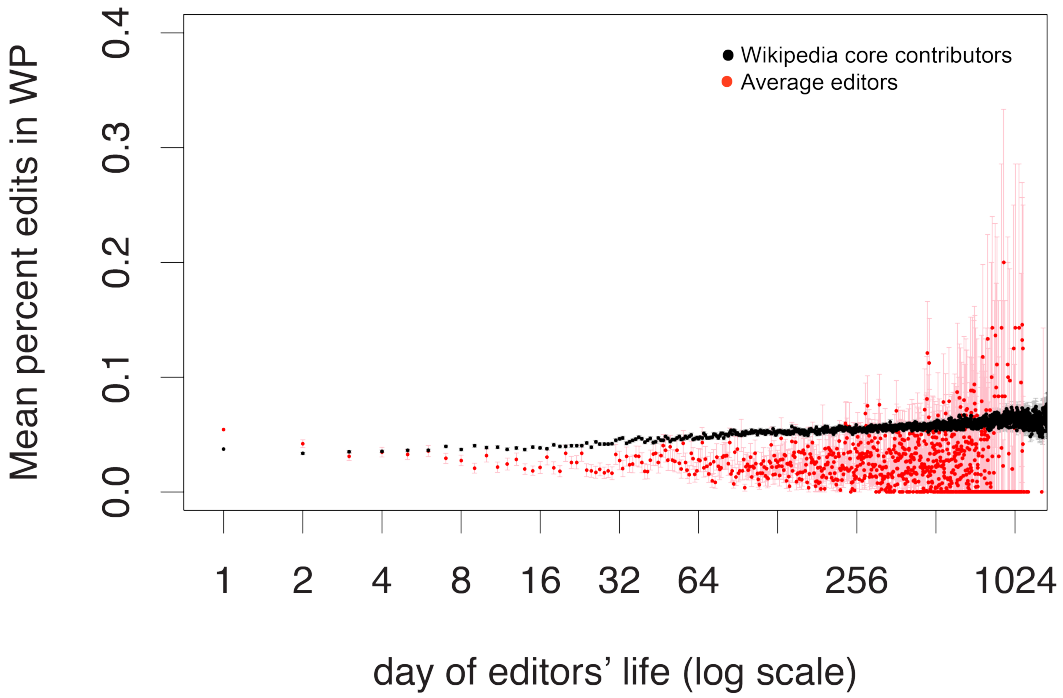


Figure 3.7: The percentage of edits in "Wikipedia" for all editors. The top black scatter represents Wikipedia core contributors and the bottom red scatter represents average editors. Standard error bars are displayed.

To investigate how community norms are invoked by editors, we used the comment field of edits. This is a place where editors can summarize their edit or the point of their edit. For example, "Revert per WP:NPOV" means that the page was reverted (taken back to a previous state) because it didn't follow the NPOV policy.

What we count. In order to be a policy related comment, either "Wikipedia:" or "WP:" are required. In many cases, those phrases are also links to the Wikipedia namespace, so by adding WP:NPOV to an edit comment, editors can also link to the relevant policy.

Our metric for detecting vandalism uses a method introduced by Priedhorsky et al. called "D-Loose" which detects 62% of vandalism [113]. (This counts revisions where the subsequent edit has a comment that suggests an intent to repair vandalism or use of a tool to help fight vandalism) Unlike in Priedhorsky et al., we aren't interested in the revision containing vandalism, but in the following revision. However we can still use the same methods of detection.

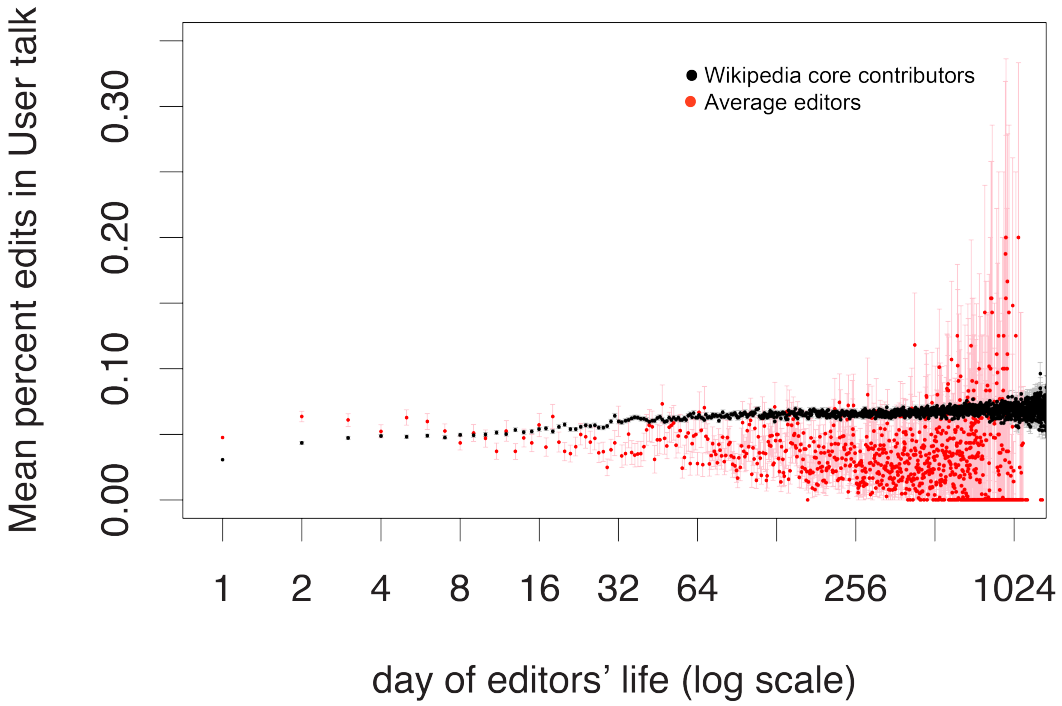


Figure 3.8: The percentage of edits in "User Talk" for all editors. The top black scatter represents Wikipedia core contributors and the bottom red scatter represents average editors. Standard error bars are displayed.

Results. Figure 3.11 shows that Wikipedia core contributors invoke norms (specifically policy and reverting vandalism) more often than average editors. In addition, Wikipedia core contributors become more likely to invoke norms over their lives. By contrast, average editors are not likely to start invoking more norms. This shows a learning effect, however it is not clear what is being learned. It could show that the Wikipedia core contributors are learning more about norm-enforcing and are doing more of it as a result. Or it could show that Wikipedia core contributors are doing the same among of norm-enforcing and are just learning how to cite it in their edit comments. This is an opportunity for additional research.

As with previous analyses, we also looked at the results of norm invocation with five buckets, instead of two. Figure 3.12 shows that this same learning effect is seen in all groups where the users make at least 101 edits.

However, looking at both Figure 3.11 and Figure 3.12, the average number of norm invocations on the first day is not zero. Therefore some editors must be aware of

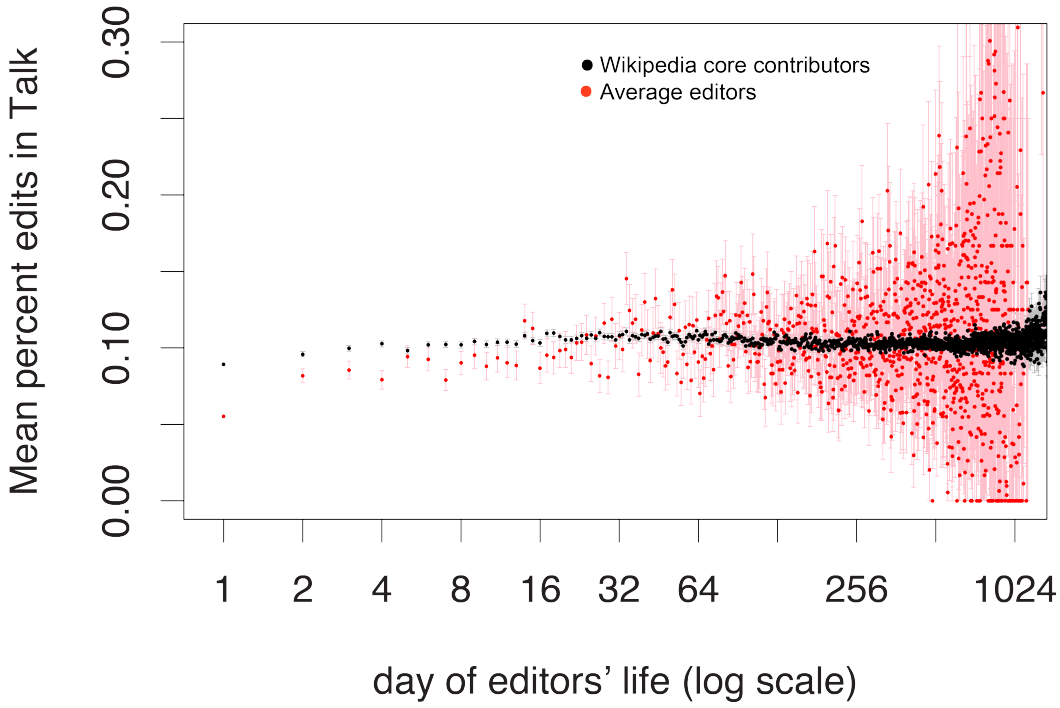


Figure 3.9: The percentage of edits in "Talk" for all editors. The top black scatter represents Wikipedia core contributors and the bottom red scatter represents average editors. Standard error bars are displayed.

community norms from the beginning of their careers.

3.3 Discussion

3.3.1 Wikipedia core contributors and Work Quantity

In Section 3.2.2, results show that nearly all editors begin with a burst and then tail off. An exceptional amount of work is done by editors in their first few weeks (or days). Average editors, by day 16, make almost no edits and the rates of edits for Wikipedia core contributors also drop over time. They don't fall to zero, as with average editors, but they hover around four. The Wikipedia core contributors tend

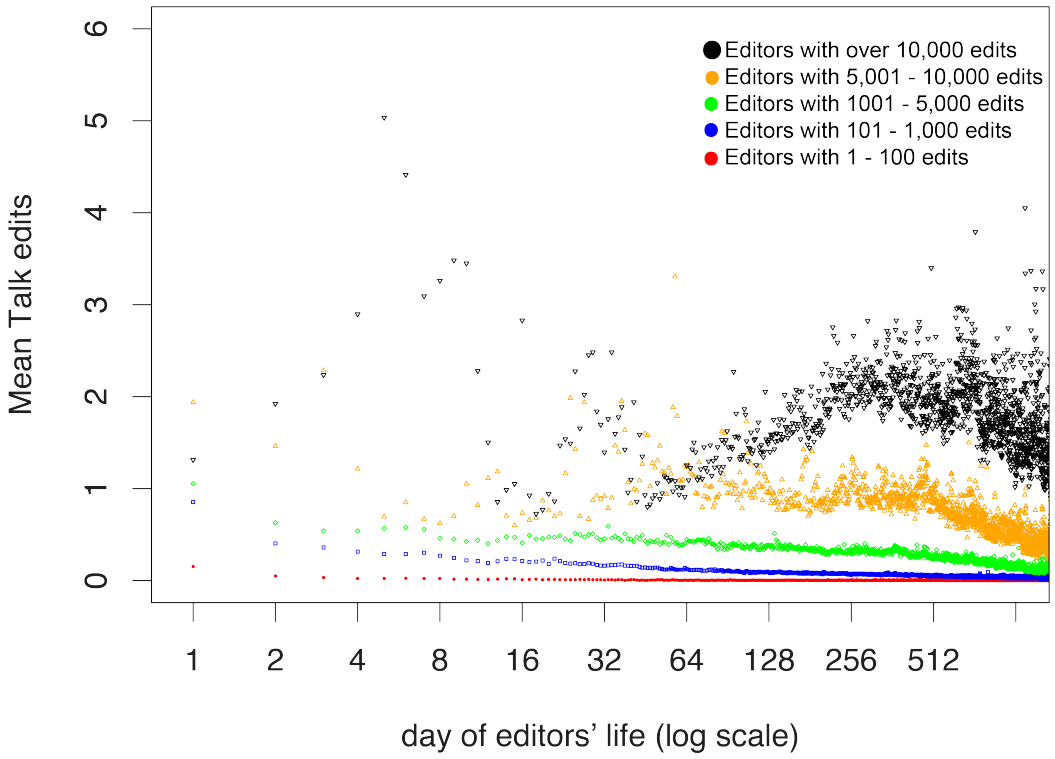


Figure 3.10: Average raw number of edits in "Talk" for all editors. The top black scatter represents editors with over 10000 edits, the next down orange scatter is editors with 5001-10000 edits, green scatter in the middle is editors with 1001-5000 edits, blue scatter near the bottom is editors with 101-1000 edits, and the red scatter at the bottom is editors with 1-100 edits. This graph is best viewed in color. Standard error bars are not displayed.

to remain more active than average editors, but we don't know why they are staying around or what this tells us about how to increase retention on Wikipedia.

One other result that we found of interest is that 60% of registered users never make another edit after their first 24 hours. Why is this? We hypothesize that editors may be scared by negative reactions to their edits (such as removal of their content or reverting their edit). Work by Halfaker et al. has shown that being reverted decreases motivation of continuing editors (at least temporarily) and decreases likelihood that the editor will keep editing [59]. We also wonder if editors don't return because they are not engaged by the existing community. (Note that since this work was published, a number of researchers have investigated the role of socialization and

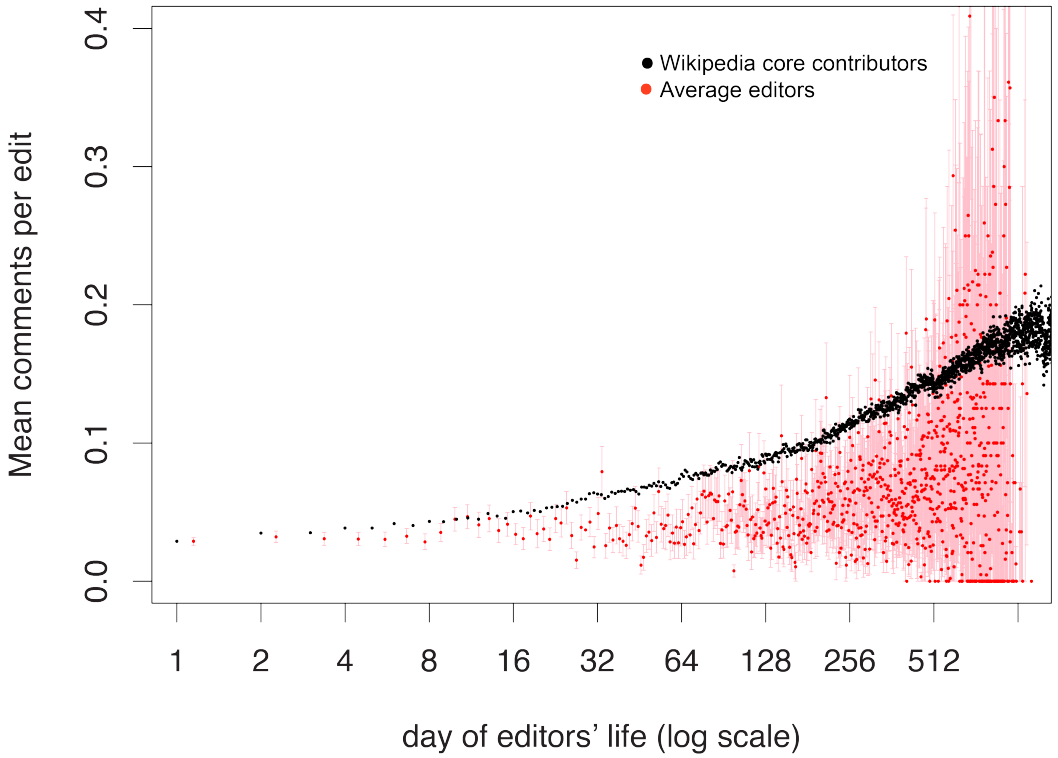


Figure 3.11: Average percentage of edits that include edit comments that reference Wikipedia policy or vandalism. The top black scatter represents Wikipedia core contributors and the bottom red scatter represents average editors. Standard error bars are displayed.

engagement within the Wikipedia community. For more see [82, 97]. There is also a new project on Wikipedia called the Wikipedia Teahouse[5].) Obviously both of these reasons are negative for the community.

A more positive hypothesis is that some editors register in order to complete a one-time task. (Registering is not required for most edits.) This is actually positive for the community as it would demonstrate that users outside the normal editor pool, are contributing, however briefly.

[5] http://en.wikipedia.org/wiki/Wikipedia:Teahouse

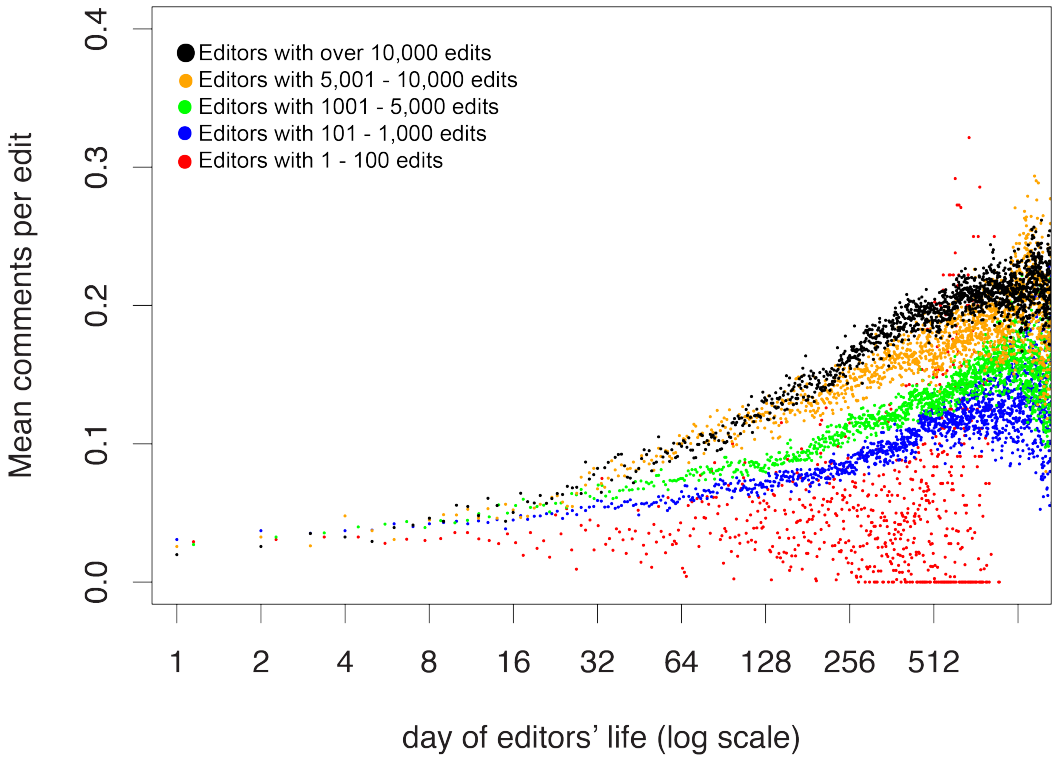


Figure 3.12: Average percentage of edits that include edit comments that reference Wikipedia policy or vandalism. The top black scatter represents editors with over 10000 edits, the next down orange scatter is editors with 5001-10000 edits, green scatter in the middle is editors with 1001-5000 edits, blue scatter near the bottom is editors with 101-1000 edits, and the red scatter at the bottom is editors with 1-100 edits. This graph is best viewed in color. Standard error bars are not displayed.

3.3.2 Wikipedia core contributors and Work Quality

The results shown regarding work quality are consistent with the work quantity results in Section 3.2.2. This finding leads us to the notion that Wikipedia core contributors are born, not made. They begin at a certain level of activity and quality, which, while higher than that of other editors, does not improve over time.

This is somewhat puzzling. In most activities, like skiing or cooking, participants become better as they practice. This leads us to wonder if perhaps editors are becoming more bold over time, thus making more controversial edits, work that has subsequently been investigated by Halfaker et al. [59]. Or perhaps, as in [26],

editors start editing outside their expertise, leading to editing in an area they aren't as confident in. Our final conjecture is that editors do get worse or more lazy over time.

3.3.3 Wikipedia core contributors and Community Work

Bryant et al. wrote "Although none of the interviewees described initial encounters with Wikipedia that involved discussion page or page histories, these features became deeply integrated into their routine activities on the site" [26]. This is something that seems somewhat intuitive and we know that Talk pages (for the Main namespace) are the most frequently used talk pages on Wikipedia. However we don't see a shift in activity towards Talk over an editor's lifespan. We wonder if this might be because we weren't able to capture all communication between editors. After initial Talk page activity, research shows that communication happens outside of Talk pages (in IRC, mailing lists, etc) [48, 69]. It is also possible that users could be coordinating through WikiProjects [30, 93]. But it is highly unlikely that all of the 38,000 users coordinate Wikipedia activity outside of the Talk pages. This is an issue for further research.

We also wondered whether Bryant et al.'s participants shifted topics within Main, not necessarily moving from editing in Main to editing in Wikipedia [26]. We used namespaces as a proxy for diversity, in part because of some of the other findings in Bryant et al. indicated that users would shift namespaces. It may be useful in subsequent studies to find a better proxy for topics within Wikipedia.

Wikipedia core contributors did show an increase in their invocation of community norms. This supports Bryant et al. who stated "Novice users learn the rules and conventions for contributing both through observation and direct coaching from more knowledgeable others". However it is unclear whether this increase comes from editors learning the norms themselves or learning how to cite the norms. Both of these aspects are important.

Burke and Kraut found that the strongest predictors of a successful bid to become an administrator on Wikipedia included the number of edits in the Wikipedia namespace, the number of edits in the Talk namespace, and the number of edit comments that explicitly mention vandalism [28]. Our work has shown differences on these three items exist between Wikipedia core contributors and average editors. Therefore, we believe that Wikipedia core contributors are more likely to have successful bids for and administrator role. This makes sense as administrators are involved more actively in developing and enforcing policy and it's logical that the best people in that role would be people who are more active in the community as a whole.

3.4 Conclusion

Wikipedia core contributors are the essence of the Wikipedia community. Previous work has debated whether Wikipedia was written by the masses or by a small subset of the community. As found by Bryant et al. and Priedhorsky et al., we too find that the majority of work done in Wikipedia is done by a relatively small group of elite editors [26, 112].

Wikipedia core contributors are born, not made. If changes occur in work quantity, work quality, and namespace diversity, they involve Wikipedia core contributors volume, quality, or diversity decreasing. In only one of the aspects we looked at did Wikipedia core contributors increase their work, invocation of community norms. This finding is notable because it is in conflict with some of the findings of Bryant et al. [26], but also because prior work in this domain suggests that lurking or peripheral participation is an important step to becoming a full fledged member of the community [85].

Wikipedia core contributors are consistent. While we observe several shifts in the work of Wikipedia core contributors, as mentioned above, most of these shifts are minor. Wikipedia core contributors appear to be steady in their work and their habits throughout their lives on Wikipedia.

Wikipedia core contributors don't do more over time. We had expected that Wikipedia core contributors would begin at a level equal to other contributors, but would then increase activity, quality, etc., while the other contributors would decrease or remain stable. Instead, we found that these increases don't exist, rather, as mentioned in the previous two points, activity remains stable or has a slight decrease.

In addition to these findings, we have a number of design implications from this study that we present in Chapter 8.

Chapter 4

Cyclopath Platform

Cyclopath was started in 2006 by Reid Priedhorsky and Loren Terveen. The goal was to provide a map that offered bicycle-specific route-finding. While cyclists at the time could get routes, usually routes were tailored towards cars or pedestrians, and not always bike friendly. An additional goal was to have the map be editable by the community and therefore consistently up to date.

The site launched in May 2008. At this point registration was required to view the map, a restriction that was lifted when the public beta launched on August 1, 2008. Since then, the site has become a community resource and a thriving research platform. The research done on the site has been broad, including how work can be elicited [114], how to generate personalized bike routes [115], the use of tags on Cyclopath [128], and task specialization on Cyclopath [89].

4.1 Application Features

The Cyclopath map includes a number of features, both geographic and non-geographic. The geographic features are primarily roads, trails, and points of interest. Non-geographic features in the system include ratings, annotations (both notes and tags), and discussions.

Roads and trails are the foundation of Cyclopath. The original data imports to Cyclopath included road and trail maps from the Minnesota Department of Transportation. Since then, users have edited, added, and deleted roads and trails, making the map more accurate and precise. The other editable geographic objects are points of interest. These can be anything from restaurants and parks to bumps on the trail or water fountains. Points of interest are often used for routing, especially as repeatedly typing in addresses is not optimized with typeaheads, or any other such

features. There are two more types of geographic objects that appear on the map, but are not editable. These are bodies of water and parks/greenspace. Recently routes became geographic objects that could be shared or publicly owned. However while routes are geographic, in most modes routes do not appear on the map unless explicitly requested.

Non-geographic features, while perhaps more hidden, represent much of the value of Cyclopath. Ratings are one of the only objects in the Cyclopath ecosystem that are private. Users can rate individual blocks or trail segments anywhere from impassable to excellent. This data is aggregated and used to estimate the bikeability ratings for other users. Ratings are used in route-finding. If a user hasn't rated blocks, the aggregate ratings are used, but if a user has rated blocks, those individual ratings will be used instead.

Annotations are another important part of Cyclopath. Notes can be attached to blocks (including trail segments) or points of interest. These can be used when searching, but often users refer to notes once they have received a route. Especially when traversing unfamiliar territory, they use the notes to figure out if the route provided is the one they want to take. Notes can provide information about the point of interest (e.g. "Open seasonally from around memorial day until a bit after labor day."), provide information about the block (e.g. "2010 snow season: the top half of the ramp hasn't been plowed these past snowstorms."), or, sometimes, to provide information on routing (e.g. "I've seen a lot of people get confused here. Coming south, go up the hill, take a left along 62 and then another left will wrap you back down to the bike path."). Tags are much shorter (i.e. restaurant, bike path, joggers) and are able to be used for routing. When requesting a route, users can choose to give a bonus or penalty to blocks marked with specific tags or to avoid those blocks entirely.

The final type of non-geographic feature is discussions. These were added to the site in 2010 and allow users to talk directly to each other within the Cyclopath ecosystem. One specific benefit of this is that geographic objects from Cyclopath can be referred to within discussions and a link to the object in question is then provided for all participants to see and use.

4.2 Site statistics

As of March 2012 Cyclopath had over 2,500 registered users and 762 editors. The site had received over 14,000 edits, of which over 12,000 were from registered users. Of the 760 IP addresses that had edits linked to them, there were roughly 2,000 edits. The map contains over 166,000 separate blocks and trail segments and over 3,500 points of interest. During it's tenure, Cyclopath has generated over 100,000

routes. 67% of these route requests came from anonymous users. The data from this dissertation was originally pulled in March 2012 when 96% of all edits were made by users who were registered or IPs that we were able to eventually link to a registered user Section 5.1.[1]

4.3 Underlying Principles

It is important to recognize two differences that exist between Cyclopath and other systems like Wikipedia or Everything2.

First, Cyclopath does not need user input for routing. While edits do dramatically improve routes [114], user input (other than to and from locations) is not needed to get a route. If users are anonymous or haven't rated anything, the system uses an objective metric to compute routes [116]. Therefore, users can benefit from the system without contributing.

Cyclopath is also different because contributors may benefit directly from their own contributions. In other systems, like Wikipedia, contributions do not directly benefit the contributor. If you add new information to a Cyclopath page, you, by definition, already know the information so you don't learn anything new about that topic. Instead, the benefits gained are indirect, like gaining reputation among editors. In Cyclopath edits and ratings by a user may influence the routes that that user gets. (Cyclopath is like a recommender system in this way.) In fact, sometimes users enter ratings or edit the map after they've received a route. If they don't like an aspect of the route, they might contribute and then request the route again to get an improved route.

4.4 Cyclopath vs. Wikipedia

There are several differences that differ between Cyclopath and Wikipedia that are important for understanding the comparisons between the analyses in Chapter 3 and Chapter 5.

Cyclopath serves seven counties and most of the editors are local. Therefore it is more possible that editors could know fellow editors and other users of the site. In contrast, Wikipedia serves the entire world and editors are spread out. We believe that this hyperlocality may affect the Cyclopath ecosystem.

[1] Due to changes in site infrastructure, comparable current statistics are no longer available.

Editing on Cyclopath is inherently more difficult than textual editing. Throughout our research we have heard anecdotal evidence to this effect. (See Chapter 7 for more details about editing difficulty.) We believe that this raises the barrier of entry for Cyclopath and discourages some users from contributing.

As mentioned above, Cyclopath produces computational output. The site uses edits to compute routes which allows contributors to benefit from their own edits. On Wikipedia, if you contribute, you rarely benefit from those edits. We believe that this provides an added incentive to edit that is absent on Wikipedia.

The goal of the Wikimedia Foundation (parent organization of Wikipedia) is "to empower and engage people around the world to collect and develop educational content under a free license or in the public domain, and to disseminate it effectively and globally" [4]. In contrast, the primary goal of Cyclopath is to be a route finder. While this is still empowering and engaging for our users (see Chapter 6), the method of reaching this goal is not collecting and developing content, but helping users find routes. The wiki-nature of the site helps allow it to be a better route finder, but the geo-wiki is secondary to the route-finding.

Finally, Cyclopath is a much smaller site than Wikipedia. While Wikipedia has over 492 million readers a month [2], Cyclopath had only 2,483 unique visitors during May 2013. Cyclopath has 2500 registered users and Wikipedia had 630,017 on English Wikipedia alone in Sept. 2010 [61]. Yet, despite these differences, Cyclopath and Wikipedia are similar. They both are wikis and information resources. They also both require minimal expertise from contributors.

4.5 Cyclopath for Research

While many differences and similarities between Cyclopath and other systems have been discussed above, Cyclopath is being used as a research platform for this work for three primary reasons.

First, Cyclopath is interesting as an open content system. It allows us to further our understanding of this type of system in part by letting us compare and contrast results with other, more well-studied systems such as Wikipedia.

Second, the design of Cyclopath enables us to study phenomena that often can't be studied in other system. An example of this is viewing behavior by "lurkers." This data is unavailable to researchers from Wikipedia and is often not logged in other research platforms.

Finally, Cyclopath is part of a new, emerging class of systems called geo-communities that let us learn how users consume, explore, and edit geography.

Chapter 5

Cyclopath and Core Contributors (Quantitative)

This work builds on the previous research into how online communities work and how they can be nurtured. Instead of continuing research on Wikipedia, however, this work is on a new platform, Cyclopath, introduced in Chapter 4.

Why a new platform? As mentioned previously, there are restrictions to the data sets available from the Wikimedia Foundation. In particular, we have no information available on registration, we only know when registered users make their first edit. There is also the problem that there is no information on viewing to enable us, as researchers, to discover whether educational lurking is taking place. Instead, we chose to investigate further on Cyclopath, a platform we had access to. This study was designed to be a parallel study to that presented in Chapter 3. The goal was to look at the same issues and see if the trends also existed in Cyclopath. In addition, if these trends existed in Cyclopath, we wanted to trace users to registration or earlier to see if core contributors act differently than other editors because they spent more time lurking in advance of their first edits or registration.

This work doesn't always use direct analogues to the work in Chapter 3, for several reasons. First, we do have an analogue for the quantity of edits and we did look at this and report on it. In terms of quality, quality in a geographic wiki is very hard to judge. While a ground truth may exist, other editors may not be aware of ti or familiar enough to judge it. In addition, editing is a much more difficult task, so a person finding an error may not know how to fix it. Also in Cyclopath there is a much smaller group of editors, so it is less likely that editors will see and review the work of their peers. Finally, community work is of a different nature in Cyclopath. In contrast to the vandalism that is omnipresent on Wikipedia, Cyclopath has had almost negligible vandalism, so cleaning up after vandals isn't required. Cyclopath

also lacks the strict policies that Wikipedia has, so policy is rarely cited.

However, despite these differences, there are things we can look at in Cyclopath that aren't possible in Wikipedia. We were able to look at the quantity of viewing, retention, the relationship between viewing and editing, and whether or not educational lurking exists. In addition, we suspected seasonal trends in Cyclopath usage due to Minnesota winters that are cold, snowy, and fairly hostile for cyclists.

When originally published, this was the first work to report on the use of Cyclopath "in the wild". At the time, Cyclopath had been operational for 16 months and had sufficient users and data to provide an idea of its use and lessons learned.

We see this work has having two primary contributions. First, we quantitatively analyzed the lifecycles of users in an open-content system, particularly pre-registration anonymous lurking. Second, we quantitatively analyzed the kind of geographic work that's being done with specific focus on how the geographic nature of the system affects what work is done and how public and hidden actions relate.

5.1 Data and Methods

5.1.1 Data Sources

This research is based on two streams of data that we log: the wiki work that is done and the requests the browser application makes to our web server.

Wiki Work. As most wikis do, Cyclopath keeps a history of changes that are made to each object in the system. On Cyclopath the object can be either a geographic object or an item attached to a geographic object. For example, Washington Avenue between 17th and 18th Streets and a note attached to Washington Avenue between 17th and 18th Streets are both valid objects.

This data can be analyzed at two levels. **Edit actions** are logged when a user changes one type of information on one object. This could mean creating a new block, changing the geometry of a block, or changing the name and speed limit of a block. Each of the above is one edit action. For each edit action, we record:

- The item that was changed
- When the change occurred
- The user who made the change (if logged in)
- The IP address of the client application

Since there is no direct comparison to Wikipedia edits, we have chosen to look at edit actions as they are the most granular data available on edits on Cyclopath.

The other level we can analyze the data at is a **revision**. A revision is the group of edit actions that are saved atomically when the save changes button is clicked. This is analogous to a revision on Wikipedia.

Cyclopath application HTTP stream. When a user starts the Cyclopath client application, HTTP "chatter" happens between the client and the server. This includes requests for data (lists of revisions, geographic objects, notes, etc.) and logs of application usage (which objects are clicked on, etc.).

In this analysis, we use interaction as an activity indicator. The client doesn't talk to the server except when there is a user interaction, therefore chatter presence indicates application use as opposed to idleness We call each of these HTTP requests a **use event**. We assume that each use event indicates Cyclopath was used from the event until 30 seconds after the event. We combine overlapping 30 second micro sessions to estimate the start and end of user sessions in Cyclopath.

In addition, we can employ use events that request geographic objects to know exactly what part of the map is on screen. This is because when the map is zoomed in beyond 4km square, the use events are generated every time a user pans or zooms. This method is described in more detail in [116].

5.1.2 Uncovering Hidden Activity

Research on consumers is often limited by what researchers can "see" and the majority of research analyzes only this visible data, though it often does mention the existence of hidden data. Some research has been done to try to "de-anonymize" anonymous data sets [50, 126], but it has been limited.

Data commonly used for analyzing online communities has two potential problems:

1. Only visible actions are included.

   Most users of online communities are readers and viewers, not editors or posters. For example, Wikipedia has 492 million readers a month [2] but only 130,000 editors a month. In addition only 300,000 editors, over the course of Wikipedia's life, have edited the site more than ten times. And yet Wikipedia has 17 million registered users [11]. The problem with excluding all viewing is that it eliminates the ability to investigate relationships between viewing and editing. It is easy and tempting to assume that editing and viewing are correlated, but very little research has been done on this. The only known

study that compares Wikipedia view data with edits found that there was no correlation between the two on a page level [113].

2. Actions by users who aren't logged in (anonymous actions) aren't linked with action made while users are logged in.

   For users, this is a potential privacy benefit. From an analytical perspective, this is a severe limitation [1] One concern with research that draws conclusions about early stages of users' lives, such as that presented in Chapter 3 is that early stages may be hidden. So perhaps the core contributors did not enter the community and start editing with a bang, but rather did a fair amount of anonymous editing, learned about Wikipedia conventions, and then created an account.

In Cyclopath, we collect a comprehensive log of Cyclopath user activity. This log does not have the first problem. We log viewing behavior, so we can analyze viewing behavior and it's relation to editing behavior. However, the second problem exists. Out of four million use events, 2.2 million of these events are performed by non-logged in users, but for some of these activities, we can estimate which user was responsible. We record client IP addresses for all events on Cyclopath and we also record usernames if the user was logged in. Therefore we can infer that some IPs are likely associated with a single user, we call these **identified IPs**.

The procedure:

1. Create a table with IP address, username (can be null), and number of events with those IPs and usernames
2. Classify IPs (not users) as follows
   (a) IPs that co-occurred with exactly one known user are labeled **identified**. All events from that IP are assumed to be due to that user, regardless of whether the user was logged in. [2]
   (b) IPs that co-occurred with more than one known user are considered **ambiguous**
   (c) IPs that co-occurred with no known users are considered **anonymous**

Following this procedure, we can now consider all users as either:

[1] We're not interested in publishing these actions to the world, but rather using them for analytical access. We have user permission to collect and analyze the user data. To be clear, we do not dismiss privacy concerns. However, research on computer security has shown that demonstrating possible privacy attacks is a necessary step to preventing loss of privacy.

[2] Note that we used a probabilistic, not a binary, approach. For example, if an IP is used at least 90% of the time by one user, we say it is an identified IP for the user. Our results were similar for all reasonable thresholds, so we used the simpler approach.

1. Unambiguous - These are users who only access Cyclopath from identified IP addresses. We can identify all work done by these users, regardless of whether they are logged in. There are 1,172 unambiguous users on Cyclopath.

2. Ambiguous - These users access Cyclopath from at least one ambiguous IP address. Therefore we can identify only their work done when they are logged in. There are 440 ambiguous users on Cyclopath[3]

Following these procedures, we can assign an additional 450,000 use events to known users, 20% of the events generated by non-logged in users. Subsequent analyses look only at Logged-In and Identified data, unless mentioned otherwise.

5.1.3 The Effect of "Remember Me"

Many systems that allow logins have a "Remember Me" feature which lets the users stay logged in across time and browsing sessions. Cyclopath uses this to an extreme, on Cyclopath logins never expire when Remember Me is selected. (However logins are not portable across IP addresses.) Other systems are different. For example, Wikipedia logins expire after 30 days, which is a typical amount of time. In addition, Wikipedia logins are portable across IPs.

We wanted to be able to compare Cyclopath more robustly with other systems. To do this we wanted to understand how our design choice, to make Remember Me not expire, affected user login decisions. Our experiment turned off Remember Me functionality in Cyclopath for two weeks. (This was in August 2009, which was peak riding season.) The Remember Me check box was still available, but had no effect. If you closed Cyclopath, you were completely logged out and not even your username was retained.

When Remember Me was active, 47% of use events and 90% of revisions were completed while users were logged in. While Remember Me was turned off, 24% of use events and 62% of revisions were done by logged in users. We would expect that Wikipedia and many other systems would lie between these two cases. This experiment may have impacted our data (and subsequently our analysis), but due to our work linking IP addresses to users Section 5.1.2, many of the edits and use events were able to be linked to registered users after the fact minimizing the potential affect.

[3]Two groups of users were excluded from this and subsequent analyses, unless mentioned otherwise. First was the research team, along with test accounts, spouses, etc, which made up 64 users. We also excluded the 429 users who registered but never logged into the Cyclopath application. This is due to a system quirk where registration occurs outside of the Cyclopath application making it possible to register without seeing the site.

This small experiment resulted in interesting implications: a small change can have a big effect on login behavior. Specifically for this research project, Cyclopath encourages users to view and edit while logged in and this property is reflected in our results. (In addition to the Remember Me feature, features such as rating roads and trails and creating watch regions are not available unless logged in [116].)

5.1.4 What is a Cyclopath core contributor?

As mentioned in Chapter 2, core contributors exist on Wikipedia [73, 107, 113], Usenet [133], and Cyclopath [89]. From our experience, we know that these users exist on Cyclopath as well. They are the users who have made 100s or 1000s of revisions. They also will contact us to request additional functionality or to let us know that something is broken.

In order to extend our work on Wikipedia core contributors , we needed a similar group of Cyclopath contributors. Wikipedia core contributors were the top 2.5% of all registered editors on Wikipedia that made the most revisions [107]. We wanted to use a similar definition here, but we increased the percentage to 5% in order to have 22 core contributors, a more substantive number than the 11 we would have had with 2.5%.

In Wikipedia there's one intuitive way to rank editors, by number of revisions. In Cyclopath there are different types of editing, so we could rank by number of revisions, number of editing actions[4], geographic actions (moving roads or points, adding roads), non-geographic actions (tagging, changing names, commenting), use events, block ratings, watch regions, cumulative personal activity (block ratings and watch regions,)

To decide which definition we should use for our Cyclopath core contributors, we needed a definition that would accurately represent the system and system activity. To do this, looking at overlaps between different rankings was useful. Table 5.1 shows the overlap between the top 5% of users, given the ranking listed. Based on this data, we chose to rank users by the number of editing actions. This provides a reasonably close comparison to previous work while remaining faithful to the structure and aims of Cyclopath. **Cyclopath core contributors** are the 22 editors that form the top 5% of Cyclopath editors, when ranked by number of editing actions.

[4]When we counted editing action, we counted all actions, whereas the geographic and non-geographic editing actions don't include deletions, just additions or other alterations to the corpus

Type of cyclopaths	Revisions	Edit Actions	Geo Edit Actions	Non-Geo Edit Actions	Use Events	Personal Actions
Revisions	-	16	16	16	12	10
Edit Actions	16	-	19	19	12	13
Geo Edit Actions	16	19	-	22	11	11
Non-Geo Edit Actions	16	19	22	-	11	11
Use Events	12	12	11	11	-	9
Personal Actions	10	13	11	11	9	-

Table 5.1: Overlap for different definitions of cyclopath core contributor.

5.2 Findings

5.2.1 Lifecycles of Cyclopath Users

Since we can look at a user's viewing as well as editing and we can link some anonymous activity to registered users, we are better able to investigate initial stages of user activity on Cyclopath, find patterns of activity over the course of users' involvement with the system, discover how different activities influence user retention, and whether seasonal effects exist.

Educational Lurking?

Learning more about the users' initial experiences is useful to researchers. Many systems don't collect identifiable data until after a user account is registered. This allows users to try out the system without leaving traces. But this also means that important data may be missing from previous analyses of early user experiences. Educational lurking is where users explore the system for a time before deciding to join. A similar idea is that of Legitimate Peripheral Participation (LPP), introduced by Lave and Wenger as when newcomers to a community undertake simple tasks that let them better understand how the community works. These ideas have been explored by previous researchers in qualitative studies [19, 26, 122, 124]. However, by linking the anonymous data with existing user accounts, we can see if educational lurking exists, quantitatively.

Time period	View	Edit
More than 1 month before registering	192	3
1 week - 1 month before	105	1
1 day - 1 week before	85	4
1 hour - 1 day before	69	0
Less than 1 hour before	606	10
Less than 1 hour after	41	126
1 hour - 1 day after	33	51
1 day - 1 week after	14	33
1 week - 1 month after	14	22
More than 1 month after	13	36
Total	1172	286

Table 5.2: Time from first view and first edit to registration for Unambiguous users. For example, 606 users registered within an hour after they first viewed the Cyclopath application, and only 18 users edited before registration.

One way that we believe that educational lurking may work is as follows:

1. Users do read-only activities (On Cyclopath this would be getting routes and browsing the map.)
2. Users participate anonymously (On Cyclopath this might involve adding points, editing the geometry of blocks, tagging, and doing other map editing.)
3. Users register an account and continue to deepen their participation with that identity (Cyclopath gives additional functionality to registered users, allowing them to rate blocks and create watch regions.)

We want to know if this process occurred. By analyzing the unambiguous users, we can identify their pre-registration activity.

We report the amount of time from first view to registration (Table 5.2), from first edit to registration (Table 5.2), and from first view to first edit (Table 5.3).

Editing to Registration. As seen in Table 5.2, only 18 of 286 editors made their first edit prior to registration. Of these 18, 10 edited immediately before registering. We see no pattern of people editing anonymously and then deciding to join. Therefore there is no evidence for educational lurking.

Viewing to Registration. Looking at Table 5.2, the picture is mixed. Over half the users first view immediately before registering and the modal pattern is for people to look an join immediately. However 25% of users made their first view a week or more before registering. Since a meaningful proportion of users use Cyclopath for a

Time period	Users
First edit less than 1 hour after first view	99
1 hour - 1 day	31
1 day - 1 week	29
1 week - 1 month	45
More than 1 month	82
Total	286

Table 5.3: Time from first view to first edit for Unambiguous users who edited. For example, 82 users waited one month or more after their first view before editing.

while before joining, this suggests some form of educational lurking. In this context it is possible that users are figuring out what Cyclopath is, beginning to understand how Cyclopath works, and learning what benefits they might receive by registering.

Viewing to Editing. When looking at time from first view to first edit Table 5.3, the data is bimodal. While 45% of editors edit within one day of their first view, 44% of editors wait more than a week to edit. This latter group (that looks before editing) is consistent with the notion of educational lurking.

Cyclopath User Lifespan

After deanonymizing the pre-registration user behavior, we looked at the activity of Cyclopath users over time. We compared the progression of user activity in Cyclopath to our previous findings from Wikipedia to see if user lifecycles in the two systems exhibited similar or different patterns. [107]. In that research, shown in Chapter 3, Wikipedia core contributors and average editors begin with a burst of activity, then tail off to a much lower level. In addition to editing, in this work we also analyzed viewing behaviors.

Figure 5.1 shows the average number of edit actions per user, according to days since users' first view. (Note that Chapter 3 and [107] count days since users' first registered edit.) This figure shows users segregated into core contributors and non-core contributors, directly analogous to Wikipedia core contributors and Wikipedia average editors. Here, as well, we see the same patterns as with Wikipedia, the core contributors edit more from the beginning. In addition, all editors start with an initial burst of activity. Following this burst, all tail off to a lower, relatively constant level. However in Cyclopath, the tailing off is more gradual than on Wikipedia. Wikipedia core contributors leveled off by their sixteenth day, but Cyclopath core contributors don't reach a steady state until after their fiftieth.

Figure 5.2 is a similar analysis, but looks at viewing behavior, not editing behavior. Here we can add another group of users, registered users who have never edited. All

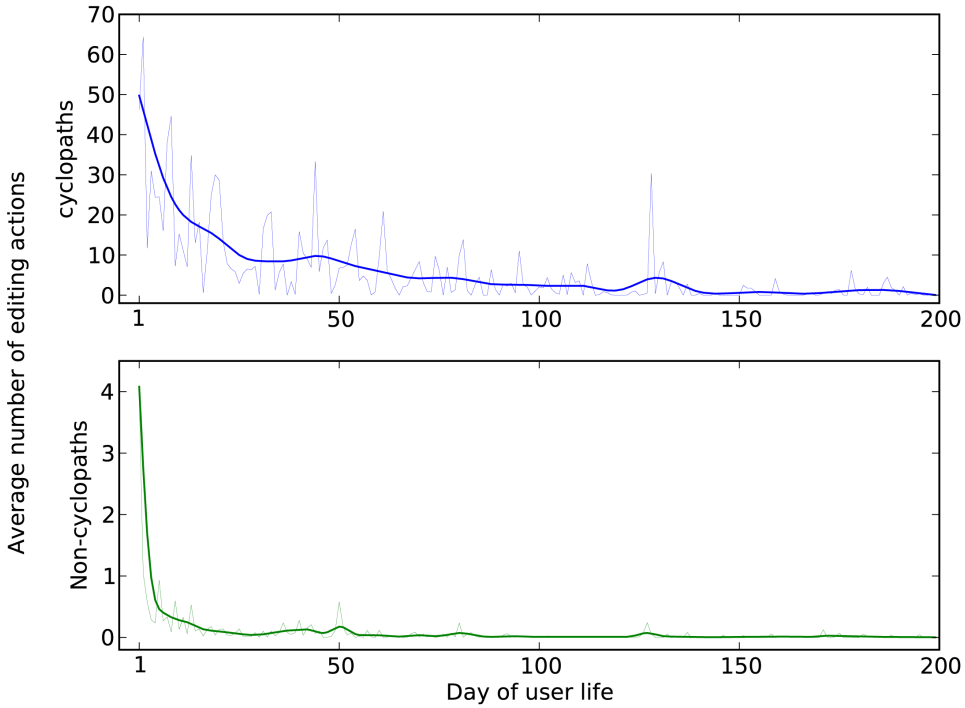


Figure 5.1: Number of edit actions per day of user life, for cyclopaths and non-cyclopaths who edit. For example, a typical cyclopath on his or her 40th day after first viewing is making about 9 editing actions daily, while a typical non-cyclopath is making about 0.2.

users have an early burst of viewing activity, but the size of that burst is indicative of subsequent activity.

Figure 5.3 allows us to look at user activity over time in a different way, by showing retention. A user is retained on day x if he or she visits Cyclopath x days after his or her first view or on any subsequent days. The patterns in this graph are similar to those seen in Figure 5.1 and Figure 5.2. But here the difference between non-core contributors who edit and consumers is clearer. After 100 days in the system 50% of non-core contributors will return, but only 30% of consumers will return.

Location Matters

Online communities change over time, for a variety of reasons. For example, Wikipedia grew exponentially from their founding in 2001 until 2007, when growth began to slow [123]. During this time rich policies evolved which are part of the reason why

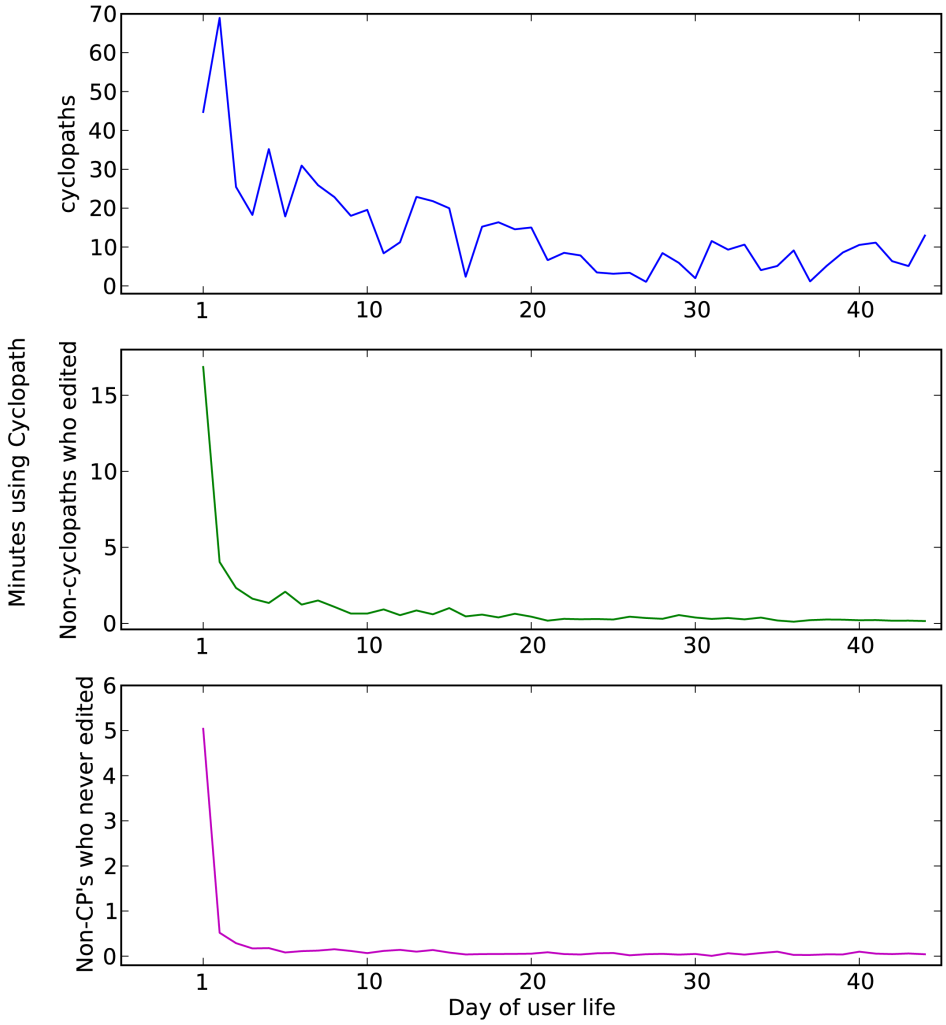


Figure 5.2: Minutes spent using Cyclopath, for cyclopaths, non-cyclopaths who edit, and non-cyclopaths who have never edited.

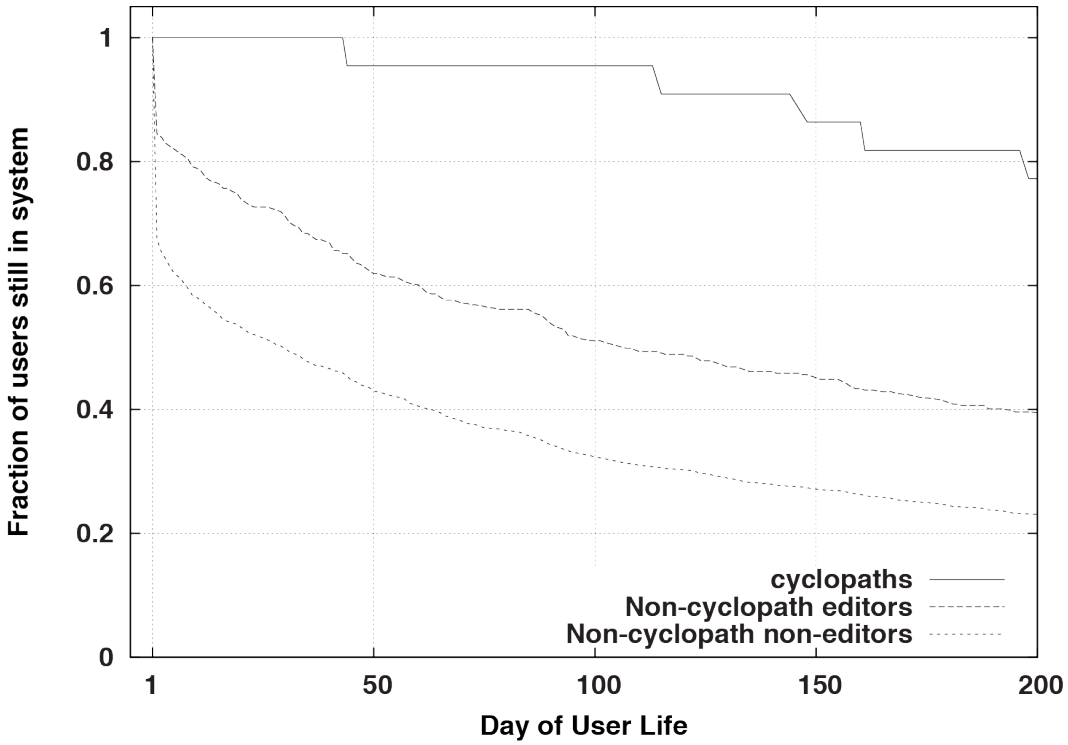


Figure 5.3: Retention, measured as the proportion of of users whose last visit to Cyclopath is n days after their first view. We end this graph at 200 days of user life: as Cyclopath is only 16 months old, very many users are still active at present and have not yet reached the natural end of their activity.

the environment for Wikipedia editors differed substantially between 2002, 2006, and 2009 [58]. We believe that seasonal effects could also be involved. This has been found to be the case on Facebook where there are different patterns of usage during national and cultural holidays [76].

There are many communities with clear "seasons". One example is students and teachers on an academic calendar. They have different educational and social activities at different times of the year. Cyclopath is designed for cyclists in Minnesota. Minnesota is known for having distinct seasons, particularly cold and snowy winters. A minority of cyclists do still ride year round, but most don't. (Even on the worst winter day, 20% of Twin Cities cyclists continue to bike. Nicer winter days may have up to 36% of riders on the roads and trails [33].) In part because of this, Cyclopath has distinct seasonal activity levels.

Figure 5.4 shows user activity between May 2008 and August 2009. In this figure, user activity in 2008 begins to decline in November and increases in mid-March. This trend exists with both Cyclopath core contributors and non-core contributors

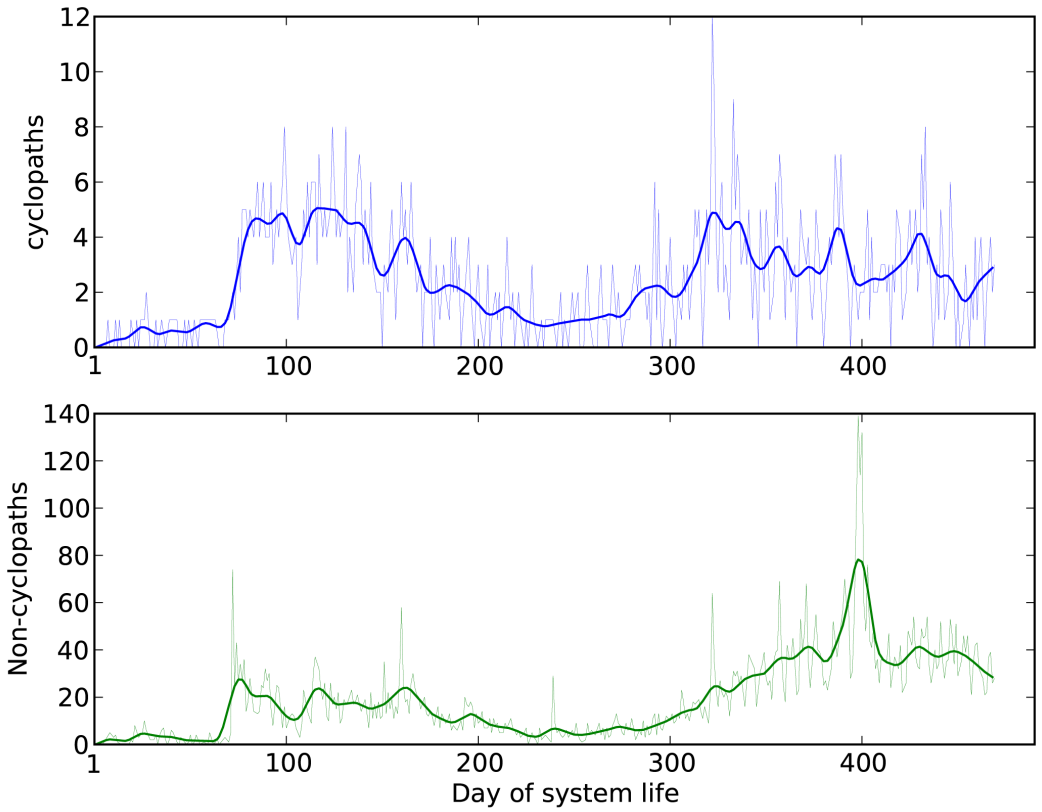


Figure 5.4: Number of users who view Cyclopath on a given day after Cyclopath went live. Day 100 is Aug. 15, 2008, Day 200 is Nov. 23, 2008, Day 300 is March 3, 2009, and Day 400 is June 11, 2009

5.2.2 Geographic Viewing and Editing

Seasonal usage is one way to understand localization, but looking at the activity in Cyclopath as geographic activity is too. On most systems, choosing where to edit means choosing topic. On Cyclopath choosing where to edit means choosing an area. We want to understand the geographic shape of editing and viewing and the relationship between editing and viewing. This relationship is unusual, in part because editing is public, but viewing is private. Therefore if viewing and editing are correlated, users' private activity can be deferred.

Our analyses included all logged-in and identified activity from all 400 registered Cyclopath editors. We used the following metrics:

1. **Number of revisions.** How many times has a user saved edits?
2. **View Compactness.** How geographically dispersed or concentrated is the

users' viewing behavior? We measure this using all the viewports of size 4km square or smaller for each user. (Larger viewports don't have much detail and therefore aren't good indicators of a user's interest.) To measure compactness, we computed the geometric centroid of each user's viewports. We then aggregate the average distance from each viewport to this centroid.

3. **Edit compactness.** This is a metric similar to view compactness, but for editing. We computed centroids by taking the centroid of objects modified in each revision. (Note that this measure will be zero if a user has only saved one revision.)

4. **Viewport coverage.** This is the fraction of a user's viewports that are intersected by *any* object modified by that user. This is our estimate for the proportion of viewing activity that can be predicted from editing activity.

Figure 5.5 shows these metrics with data from three users. These users have differing levels of editing activity as well as different patterns of geographic activity.[5] These maps show that public editing may offer good indicator of a user's private viewing activity. To test this further, we did some global comparisons.

Figure 5.6 shows a positive relationship between view and edit compactness, as we expected. If you view a wide variety of places, you likely edit a wide variety of places (assuming you are an editor).

Figure 5.7 shows viewport coverage increases with the number of revisions done. In fact, viewport coverage approaches completeness for the most prolific editors. This shows that the more the user edits, the more they are revealing about their private viewing activity.

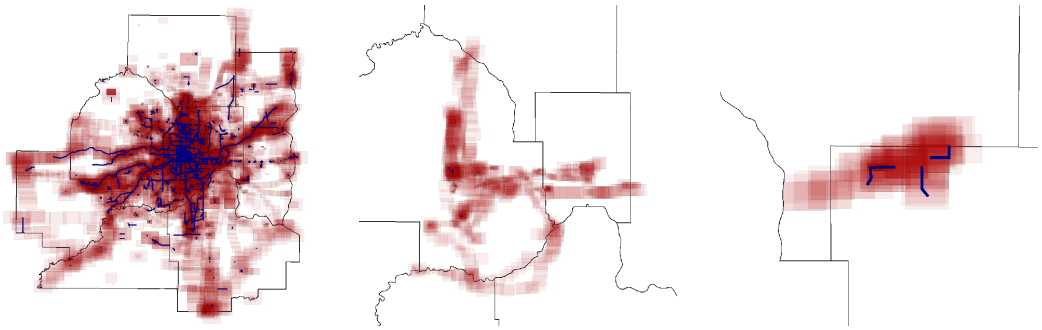
5.3 Discussion

5.3.1 Educational Lurking

Our research found equivocal evidence about educational lurking. Therefore we present several ideas for further study. First, because our results aren't definitive, follow-up qualitative research is needed to help us learn more about the role of educational lurking in the development of users. The results of this are presented in Chapter 6 and Chapter 7

We could also do contextual experience sampling [64]. For this study, we would present users with questions after their first edit or registration. This qualitative

[5]This data is shown with the permission of the users.



(a) Carol: 2,291 revisions, view compactness of 10.1km, edit compactness of 9.5km, and viewport coverage of 0.92. The image is 110km square.

(b) Tristan: 5 revisions, view compactness of 2.6km, edit compactness of 2.7km, and viewport coverage of 0.10. The image is 63km square. Note that even though coverage is low, the edits are still at the "core" of viewing.

(c) LaDawn - 5 revisions, view compactness of 11.4km, edit compactness of 2.2km, and viewport coverage of 0.82. The image is 27km square.

Figure 5.5: Example of editing and viewing activity for three users (identified by randomly assigned pseudonyms). The red layer is a heat map of viewing: darker red indicates areas viewed more often, the blue overlay is revisions. Black lines show the boundaries of the counties in our metro area as context. The metro area has a radius of about 50km, and the map contains over 150,000 editable road and trail segments. Note that the different maps are at different scales.

understanding of the user lifecycle would inform Cyclopath design and design of related systems. The findings may suggest that emphasizing benefits to encourage registration or suggesting additional benefits to users may help engage them.

Second, we know that different systems offer different benefits to registered users. For example, on Wikipedia if you want to create a new article you must be a registered user. In some communities, anonymous users and their contributions are looked down upon. Slashdot was designed to allow anonymous posts, but research has shown that these anonymous posts generally have lower value [83]. In addition, the author of an anonymous post on Slashdot is listed as "Anonymous Coward." These anonymous posters also can't develop an identity or gain any community reputation without registering and then using a persistent username [51].

In addition to registered users being more highly regarded than anonymous users, many systems offer additional benefits. On Cyclopath, registered users can rate roads and trails, enabling them to receive personalized route recommendations. These benefits that a system may offer create different registration incentives. This

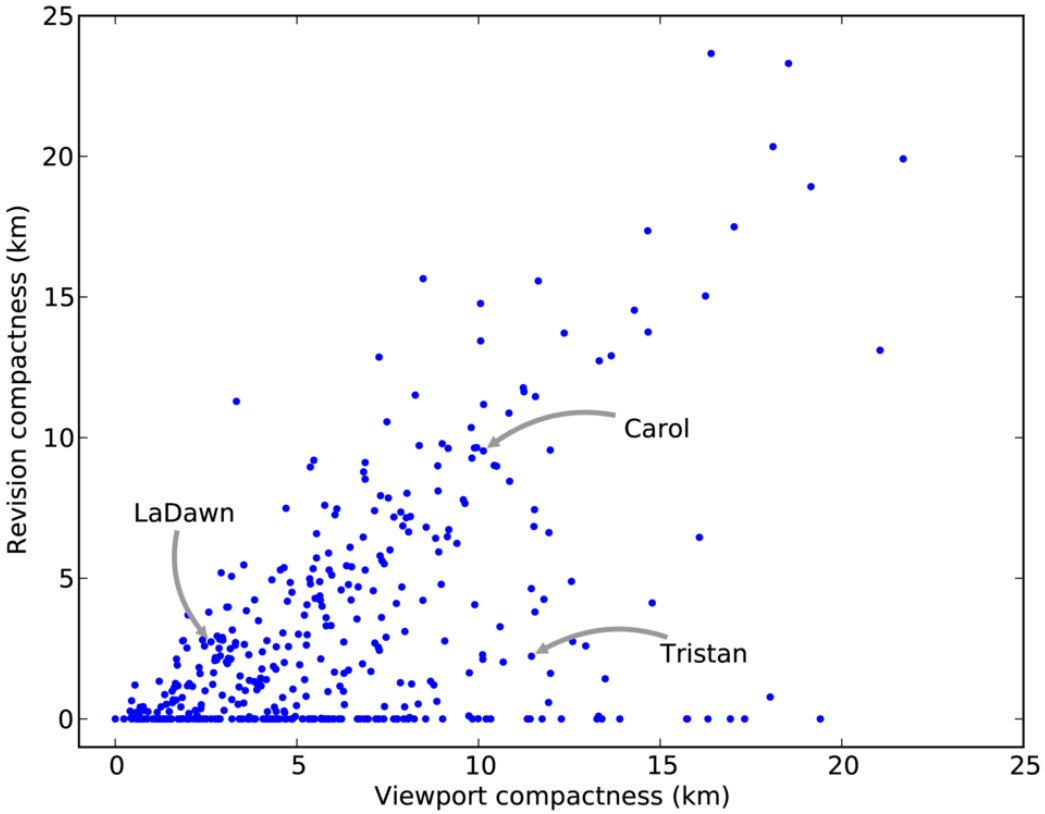


Figure 5.6: Scatterplot showing the relationship between view compactness and edit compactness.

notion is discussed in more depth in Chapter 6.

5.3.2 Cyclopath User Lifespan

We are interested in learning what the motivations are for users to edit and what barriers to editing might exist. In addition to the survey we conducted that is discussed in Chapter 6 and the interviews presented in Chapter 7, we could also run contextual surveys. In addition, we could then build on the results of these steps by designing interventions to attract greater participation [114].

We are also interested in learning more about purely-anonymous users who never register. Our current technique that we used in this paper is not robust enough to identify purely anonymous users. We could, and in fact did, introduce a change in the Cyclopath logging system that makes this possible. However this change was only added following the cycling season in 2009, so it only allows this data to be

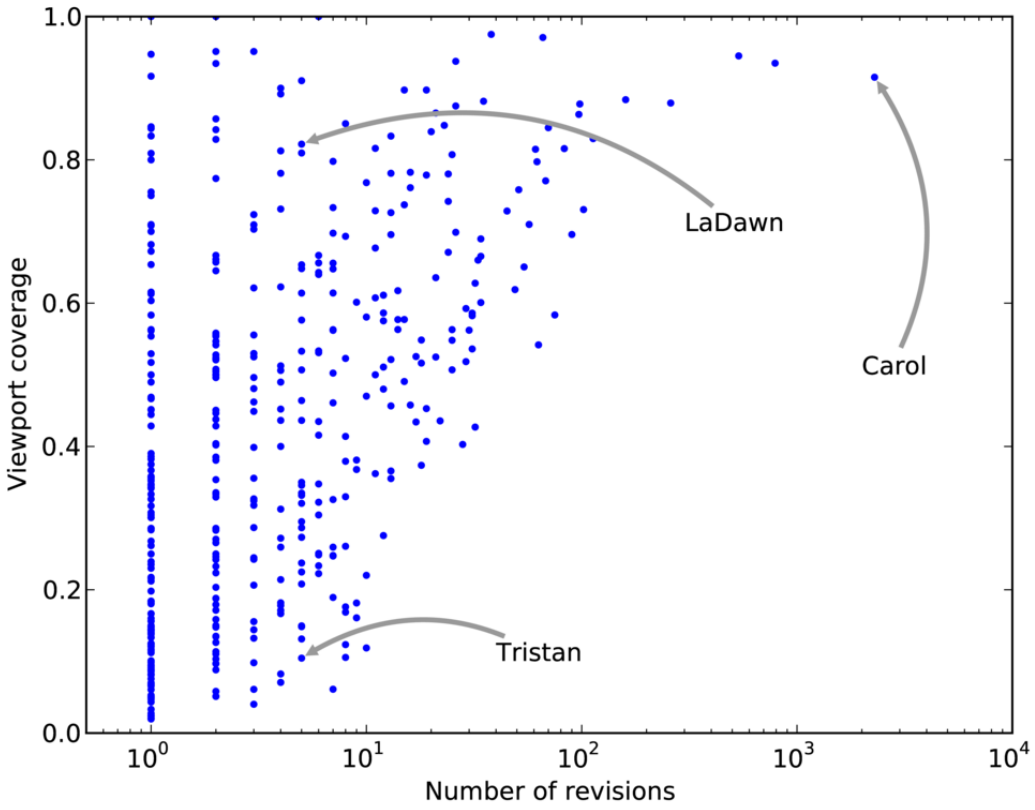


Figure 5.7: Scatterplot showing the relationship between number of revisions and viewport scale. Note that the x axis is on a log scale.

studied with newer users.

5.3.3 Location Matters

While we can conjecture ways to engage users throughout the off-season, it may be useful to do more studies (or analyze existing studies) to learn how users use the system throughout the winter and why they visit the site when they aren't biking. Mikhil Masli has done a number of studies in the off-season, in part to try to get users to remain engaged. These studies have all started with an email request to users to come to the site and complete an activity. These email requests may help drive users to the site when they might not be otherwise checking the site.

5.3.4 Geographic Viewing and Editing

Regarding Figure 5.7, it is not clear that these results should change the way we think about viewing and editing. It is likely that the areas someone views most are close to potentially sensitive places like their home, work, or commute. To understand more about these issues, we need to conduct follow-up studies with users. This would allow us to learn how sensitive users feel that their viewing activity is. In particular, we would be interested in learning how viewing data compares with other private data like library books checked out or web pages viewed. We are also interested in studying whether this relationship holds in other systems. Do a user's edits on Wikipedia predict the articles that they view? (Note that this is currently untestable due to lack of available data.)

5.3.5 Future research

This work on Cyclopath user lifespans leads to several future possibilities which build on the idea that users may continue to use Cyclopath even if they do not make any edits. Currently the presence of consumers is hidden. One experiment we could conduct would be to show the presence of consumers in the aggregate. So perhaps showing the audience size for a specific map region. This may motivate additional work by current editors as well as enticing other to begin editing.

We could also look at details of the activity of consumers and average editors. By learning to what extent they request routes vs. browse the map, we have the potential to better tailor the application for them.

If we found that non-editors and average-editors made use of Cyclopath's route finding mechanism, we could give them feedback showing that user edits lead to better routes. An intervention to this effect was conducted on Cyclopath following the work discussed in [114], but it was not heavily used and the results have not been analyzed. (Users were presented with a "Region of the Day" to focus their editing efforts and were encouraged to complete work in order to help either themselves, the Cyclopath system, or the community.)

If the activity supported by a system has a seasonal nature, designers should consider what features and activities are seasonally appropriate. On Cyclopath, we have promoted work campaigns to improve the quality of the map in the middle of winter. We could also do things lie facilitate cycling related discussions, let users know about cycling events, or remind users of essential off-season maintenance.

The results in Figure 5.6 showing the relationship between viewing and editing suggest that the diversity of viewing and editing may influence each other. Previous work shows that editing can be encouraged in areas that are basically arbitrary [114].

This may be a useful tool if more diverse viewing is desired. In addition, users may then begin "watching" changes in more of the map. If more diverse editing is desired, viewing campaigns like "ride of the week" or "do you know what's in neighborhood x" may be effective, especially to fix errors in areas that have gotten little attention.

5.4 Conclusion

There is some evidence of educational lurking on Cyclopath. Some users viewed at least a week before registering and almost half of all editors waited at least a week between their first view and their first edit. There is more research to be done on this topic, including looking at how much people are viewing before registering and editing. In addition, learning about users motivations for registering or editing may help us understand this more. This is looked at more in depth in Chapter 6.

Cyclopath core contributors act like Wikipedia core contributors. The Cyclopath core contributors do the bulk of the work and are more active from their first day. They also remain involved with Cyclopath longer than other users. This mirrors the Wikipedia core contributors. This is surprising, in part, because a small group of local bicyclists wouldn't normally be expected to act like a huge distributed community producing an encyclopedia.

There is a relationship between locality of viewing and editing. This suggests that viewing may be able to be guided by focusing editing behavior in certain geographic regions (demonstrated in [114]). In addition, collective editing may be able to be steered by interventions that encourage viewing in specific areas.

People who edit extensively reveal some information about their viewing behavior. This finding raises some potential privacy implications and suggests that more study could be done to learn about user attitudes towards privacy of geographic behavior. Similarly, this raises questions about editing on Wikipedia and whether editing behavior may reveal information about viewing behavior.

In addition to these findings, we have a number of design implications from this study that we present in Chapter 8.

Chapter 6

Cyclopath Motivations

Our previous work has shown how users behave in both Cyclopath and Wikipedia, however we haven't learned *why* consumers and contributors act the way they do. There has been a significant amount of previous research on the motivation of users (see Section 2.3.5 for more). In particular, there has been work done on identifying factors that motivate people to participate in open collaboration systems [62, 84, 102, 105]. Other work has looked into how users transition to become contributors [26, 111], how users' initial participation relates to subsequent participation [107, 109], and gender imbalance in participation [20, 82]. However much of the work about motivation ignores both the difference between consumers and contributors and the difference between core contributors and average contributors.

Understanding how users are motivated and how they (and the community) benefit from their participation is important because research shows that motivation is a factor in volunteer retention [34] as is belief that you benefit from volunteering [34, 104] and belief that your participation benefits the community [57, 103]. In addition, even successful sites struggle with getting all the needed work done. For example, in November 2012 alone, Wikipedia had 1,587 articles marked as needing expansion.[1] On SourceForge, 31% of OSS projects are abandoned before a first release [134]. Finally, 87.4% of collaboratively made animations on the user generated Flash game and animation site Newgrounds[2] are incomplete [87]. If these successful sites have problems with contribution, those problems are likely inevitable on smaller sites.

This research looks at differences between Cyclopath consumers (readers) and contributors (editors) specifically with regards to motives and benefits. We also make some distinctions between average contributors and core contributors. Note that in this work we are not looking at psychological motivations, but at system based

[1] http://en.wikipedia.org/wiki/Category:Articles_to_be_expanded
[2] http://www.newgrounds.com

motivations. This means that the motivations we're interested in are things like finding routes or fixing streets, not "understanding" or "values". We look at these motivations because these allow us to directly understand what changes we should make to the system instead of implying changes that may or may not be be impactful for those users.

Our research questions are as follow:

RQ1. Registration:.

Why do Cyclopath users register for the community?

RQ2. Benefits:.

What personal benefits do consumers and contributors believe they receive from their participation?

RQ3. Motives:.

Why do contributors begin contributing and why do they continue contributing?

6.1 Data and Methods

6.1.1 Data

Survey. The survey was designed to learn about different aspects of Cyclopath users usage and behavior on the site. There were also questions about bicycling habits, social media usage, printed bike maps, and more. Of the 61 questions, 28 were multiple choice, ranking, or selection questions.

A link to the survey was emailed to all 2,100 registered Cyclopath users. The link was also posted on Twitter. Just over 400 users took the survey, in whole or in part. Of those, 342 identified themselves as Cyclopath members, but we were only able to link 290 respondents to existing Cyclopath accounts. This chapter only looks at answers from those 290 users. People were not required to answer any questions, so the number of responses for a given question varies and we report on the number of responses for each question we analyzed.

Behavioral Data. Cyclopath logs information about users. This includes viewing, rating, and editing information as well as route requests. In addition to these actions, we also store usernames, IP addresses, timestamps, and any data that may be relevant to the action. In this chapter we only look at behavioral data from registered users who completed the survey. The Cyclopath log data is from March 17, 2011.

6.1.2 Behavioral Analysis

Before conducting behavioral analysis, we applied the technique from Chapter 5 to deanonymize actions performed by survey participants. Recall that this method allows us to link actions that were done anonymously with registered users. Unlike with previous work, in addition to analyzing users that we could link to anonymous activity, we also included users where we couldn't link anonymous activity. In all, this research separated the 290 survey takers into 153 consumers (who had not edits and no ratings) and 137 contributors (who had edited, rated, or both).

Following this, we processed usage logs to separate contributors from consumers. We then segmented contributors based on the number of edits they had. This was because our prior research had shown significant differences in the behavior of casual contributors vs. core contributors Chapter 3, Chapter 5. We report on the segmented results when the results display significant differences.

When broken down beyond contributor vs. consumer, the contributors are segmented as follows. There were 63 users classified as low contributors. These users had 1-19 edits and/or ratings each. Forty-five users were classified as medium contribution, having 20-199 edits and/or ratings each. Finally, high contributors were the 29 users who had over 200 edits and/or ratings each.

6.1.3 Survey Analysis

For this research we analyzed five open ended questions. The open ended questions we analyzed deal with the user's motivations and the perceived benefits of participation.

The open ended questions we analyzed were as follows:

> Why did you register for Cyclopath?
>
> How do you feel the cycling community has benefited from your use of and contributions to Cyclopath?
>
> How do you feel you have benefited from using and/or contributing to Cyclopath?
>
> Why did you start editing Cyclopath?
>
> Why do you contribute to Cyclopath?

Analysis process. With the help of Loren Terveen and Mikhil Masli, I coded responses to the questions, using a grounded theory approach. First, the three of us

independently coded all the responses, then we came together and mutually decided on standard codes for each question (based on individual codes). Then we coded 20 responses (per question) as a group to ensure that definitions were clear and would be applied similarly by all three authors. We then independently coded all five questions using our standardized codes and our mutually agreed upon rules. Multiple codes could be applied for each response. We applied codes before we formed hypotheses, however some codes were collapsed after the fact.

Agreement. We calculated Fleiss' Kappa for exact agreement and used a strict calculation. If two coders coded a response "routing" and the third coded the response with "routing" and "general knowledge", this was not considered agreement. All questions were in the range of "moderate to substantial" agreement, using Fleiss' Kappa. Why did you register for Cyclopath? was .62. How do you feel the cycling community has benefited from your use of and contributions to Cyclopath? was .68. How do you feel you have benefited from using and/or contributing to Cyclopath? was .70. Why did you start editing Cyclopath? was .59. And Why do you contribute to Cyclopath? was .71. This shows that our coding was reliable. The following analyses report using the majority of coders (2 of 3 had to agree, unlike the strict requirements we used when calculating Fleiss' Kappa).

6.2 Findings

6.2.1 RQ1 Registration

Registration Motivation

As mentioned above, we coded and analyzed responses to the question "Why did you register for Cyclopath?". Our goal was to identify the users' main motivations. We also wanted to see if the motivations listed by contributors differed from the motivations listed by consumers. The results are shown in Table 6.1.

We used Fisher's Exact Test to see if the differences between consumers and contributors were significant, and they were ($p = 0.047$). A follow up test for equality of proportions showed a significant difference for the edit code ($p = 0.007$).

We also looked at the differences between low, medium, and high contributors. Table 6.2 shows the codes that had significant differences between levels of contributions.

Results

In this section, we explain the codes used and discuss interesting patterns.

Find Routes. The most common reason cited for registering for Cyclopath was to find bike routes. This is not surprising as this is the main point of the site. Some responses emphasized finding routes in unfamiliar areas:

> To find routes to parts of town where I'm not used to riding.

Other responses came from people new to the Twin Cities or new to the Twin Cities bike scene.

> New to the Twin Cities, wanted to find out how to get to bike trails
> ....

There was a trend showing that the more people contributed, the less likely they were to cite "Find Routes" as a reason for joining, although this relationship was not statistically significant ($p = 0.2$). This result made us think that perhaps people who became contributors came to the site intending to contribute (e.g. more like [107] and [109] than [111]). This idea is also consistent with the responses to RQ3 shown in Section 6.2.3.

Edit. The second most popular reason for joining Cyclopath was contributing information to the system. The response pattern was almost opposite that from "Find Routes". The more one contributed, the more likely it was that one cited "edit" as a reason for joining. Unlike earlier, this difference is significant ($p = 0.02$).

The following are some examples of responses with nuances noted.

> In order to add locations to the map like restaurants

This response identifies a desire to add useful points of interest to the map. Prior research shows that this is a common type of initial edit [89].

> To edit things under a user name rather than an IP address ...

This suggests that the user wants the edits to be visible. This could be in order to gain reputation (a popular motivation for contributing). For more on this motivation, see [105] or [129].

> to edit the map focusing on NE Minneapolis

An interest in editing in a specific region suggests two factors in the collective effort model [67] as promoting participation in group activities.

1. Caring more about the outcome of the activity (e.g. the map). The user cared about the region and wanted it well represented on the map.
2. Believing that one has a unique contribution to offer. The user may have felt that he/she was the only Cyclopath member to have the interest and ability to edit this region.

Customization. Customization was the third most common reason that users had for registering. In particular, users mentioned a desire to save system settings and personal preferences for a better user experience.

> . . . to save riding preferences for better routes

is an example from one users. As with "Find Route", this reason was more frequently cited by consumers than contributors. However this difference is also not statistically significant.

I like it and Ideology. Some people mentioned general support of Cyclopath and the idea behind Cyclopath.

> I thought it was a much-needed service for cyclists in the Twin Cities.

> It was obviously an amazing tool. It was an easy sell, and it seemed to be engineered correctly – that is, it appeared to work.

Others mentioned support for a resource for bicyclists that was open content. These were coded as "ideology".

> I wanted to participate in the exchange of information and network of "scouts" that Cyclopath entails.

In these areas there were no systemic differences between consumers and contributors, but prior research has found that agreeing with the open collaboration ideology can help to motivate contribution [62, 102].

Remaining Codes. The other seven codes we used were applied sparingly, were not interesting, or both. "Misc." encompassed vague and uncategorizable responses. "Not Sure" was used when users said they didn't know why they registered. "Required" applied to cases when the user said the feature(s) they wanted to access required registration. (Note that this is similar to customization, but the two codes

can be disjoint. In particular, all alpha users were required to create an account in order to access any part of Cyclopath.) "Get General Knowledge" was used when the respondent mentioned a desire to obtain bicycle knowledge in general (not connected with routes or a locale). "Benefit Cyclopath" included responses such as

> I wanted to help Cyclopath.

Finally, "Testing" was used when respondents said that they wanted to try Cyclopath out.

Code	**Consumers**	**Contributors**
Find routes	27 (23.48%)	26 (18.57%)
Edit	12 (10.43%)	34 (24.29%)
Customization	16 (13.91%)	17 (12.14%)
I like it	11 (9.57%)	15 (10.71%)
Misc	12 (10.43%)	11 (7.86%)
Not Sure	14 (12.17%)	5 (3.57%)
Required	5 (4.35%)	9 (6.43%)
Ideology	6 (4.35%)	7 (5%)
Get general knowledge	8 (6.96%)	5 (3.57%)
Benefit Cyclopath	4 (3.48%)	5 (3.57%)
Testing	0 (0%)	6 (4.29%)
Num. Applications	115	140

Table 6.1: Coding results for the survey question "Why did you register for Cyclopath?" The table shows the number of responses to which each code was applied (using the 66% agreement rule). Multiple codes could be applied to each response; thus we report the total number of code applications for users in each participation category. 255 codes from 217 distinct survey respondents are shown.

Code	**Low**	**Medium**	**High**
Find routes	14 (24.56%)	9 (16.98%)	3 (10%)
Edit	8 (14.04%)	14 (26.42%)	12 (40%)
Customization	13 (22.81%)	2 (3.77%)	2 (6.67%)
Nam. Applications	57	53	30

Table 6.2: Additional results for the survey question "Why did you register for Cyclopath?" The contributors in this table are divided into low, medium, and high contribution groups.

Summary

Consumers and contributors tend to cite different reasons for registering for Cyclopath. In addition, the least prolific contributors are more similar to consumers than they are to the more prolific contributors.

The reasons that users cite for registering for Cyclopath are similar to those found in prior research on Wikipedia and OSS. These results help to deepen and generalize other work.

Finally, different tiers of editors may have differing motives. Therefore it may be that different motives may be tied to the number of edits that the user makes.

6.2.2 RQ2 Benefits

In this section, we present results that include the ways people said that they personally benefited as well as their perceptions of how the community benefited from their use of or participation in Cyclopath.

Individual Benefits

Gaining individual benefits from participation in an activity is important. Research on people who volunteer shows that if the volunteer in question feels no personal benefits from volunteering, they will be less likely to continue volunteering [104]. As contributing to and participating in open content communities can be viewed as a type of volunteerism, we wanted to know if Cyclopath users could identify ways that they benefited from participating in Cyclopath. We also were interested in learning if consumers feel differently than contributors.

Results from the survey are shown in Table 6.3. As a reminder, the question users were asked in the survey was "How do you feel you have benefited from using and/or contributing to Cyclopath?".

The results in Table 6.3 show some differences between consumers and contributors, but all of these differences are modest and none are significant. (Fisher's Exact test results in $p = 0.23$.)

Most users cited benefits such as gaining knowledge and improving as cyclists. However, as with the previous section, we can break this down further by type of benefit.

Code	Consumers	Contributors
Routing	84 (59.15%)	86 (58.5%)
Safety	19 (13.38%)	14 (9.52%)
No benefit	13 (9.15%)	9 (6.12%)
Ride Attitude	7 (4.93%)	11 (7.48%)
Bicycling Community	6 (4.23%)	7 (4.76%)
Misc	3 (2.11%)	8 (5.44%)
Riding Habits	3 (2.11%)	9 (6.12%)
General knowledge	7 (4.93%)	3 (2.04%)
Num. Applications	142	147

Table 6.3: Coding results for the survey question "How do you feel you have benefited from using and/or contributing to Cyclopath?" The table shows the number of responses to which each code was applied (using the 66% agreement rule). Multiple codes could be applied to each response; thus we report the total number of code applications for users in each participation category. 289 codes from 223 distinct survey respondents are shown.

Results

Routing. The most common benefit was improved routing. This could be finding better routes, finding routes more easily, finding routes in unknown areas, or finding faster routes.

> I have found a better and faster route for my commute.

> It helps me get the feel for a ride before I go out to physically ride. It also helps me figure out what route would be best.

> Tips on routes to take through neighborhoods I'm not very familiar with... what roads to avoid, what streets allow for faster commute, etc.

> I've quickly learned how to find "child-friendly" bike routes to different locations.

Safety. Safer routes are a subclass of better routes that we chose to display because it was so prominent. Cyclopath does not claim to produce safe routes, but tends to compute routes that use more trails, bike lanes, and quiet streets. There are also fewer recommendations that go on busy roads. It seems that users seem to consider the routes given by Cyclopath to be safe, or safer than alternatives.

> I've discovered new routes that have made getting around easier, faster, and safer.

> It's like having an advisory panel of people who bike more than I do to ask "hey how should I go from __ to __ without getting killed?"

In addition, some users have added notes warning about dangerous areas. The following is the note added to a point of interest a user created entitled "Dangerous Problem".

> This area is quite dangerous, especially as one comes down the hill from Warner Road. There is a low spot in the paving where the trail curves that accumulates sand and detritus. At least two serious accidents requiring ambulance rides have happened at the exact spot. As of 8/28/10, the area is marked with cones and spray paint. Messages have been left with DOT.

Ride Attitude and Riding Habits. Some people had a change in attitude to bicycling, particularly more confidence or more fun.

> This is a great service that helped me commute more confidently via bicycle

> The routes are locally focused and fun to use.

Other people say that the additional knowledge they gained encouraged them to ride more or to ride to new places.

> I commute more often by bicycle and share routes with friends.

> I've found a few new places to ride, and had fun editing too.

Bicycling Community. Some users said they felt more connected to the cycling community by participating on Cyclopath.

> I feel good about it. I like to help my community and this is just another way of doing so.

Summary

Consumers and contributors mentioned similar individual benefits from participating in Cyclopath. Gaining knowledge to help them become better cyclists was the most popular benefit received. The results shown here may differ from those on other sites, in part because Cyclopath allows all users to benefit without needing to contribute and any contributions made by a user may end up benefiting that user directly.

Community Benefit

In this section, we present responses to the question, "How do you feel the cycling community has benefited from your use of and contributions to Cyclopath?" Here there are clear differences between consumers and contributors, shown in the results in Table 6.4. The difference between consumers and contributors is significant (Fisher's Exact Test $p < 0.001$) and follow up tests for equality of proportions shows significant differences ($p < 0.01$) for the codes "No benefit" and "General Knowledge".

Code	Consumers	Contributors
No benefit	50 (47.62%)	29 (26.13%)
Better navigation	17 (16.19%)	30 (27.03%)
General Knowledge	3 (2.86%)	26 (23.42%)
Spread the word	17 (16.19%)	5 (4.5%)
Bicycling Community	8 (7.62%)	9 (8.11%)
Misc	5 (4.76%)	7 (6.31%)
Safety	5 (4.76%)	5 (4.5%)
Num. Applications	105	111

Table 6.4: Coding results for the survey question "How do you feel the cycling community has benefited from your use of and contributions to Cyclopath?" The table shows the number of responses to which each code was applied (using the 66% agreement rule). Multiple codes could be applied to each response; thus we report the total number of code applications for users in each participation category. 216 codes from 194 distinct survey respondents are shown.

Consumers were most likely to say that the community got no benefit from their participation. Contributors were less likely to say that the community got no benefit and were more likely to say that others benefited through "better navigation" and better "general knowledge" about cycling in the Twin Cities.

> I imagine there are other cyclists who travel to the landmarks I've marked. I'd like to think I've made their rides just a little easier.

> Shared some fun things and expanded tags to make more locations searchable.
>
> Slight improvement in the understanding of the St. Anthony Park neighborhood."

Some consumers were able to articulate a benefit they provided to the community, something that both surprised and pleased us. 15% of consumers and 5% of contributors said they told others about Cyclopath. We termed this "spread the word". These users told others and therefore more people benefited from the knowledge available from Cyclopath.

> I have told others about the site and hopefully that has improved their biking experiences.
>
> I have shared the site with many folks who were curious about the best route to take for various needs.

"Spreading the word" is consistent with the role of "active-lurker-as-propagator" introduced by Takahashi et al. [127].

Summary

Consumers and contributors gave very different responses when asked how their participation benefited the community. Consumers were most likely to say that there was no benefit. But if consumers cited a benefit, it was most likely telling others about Cyclopath. Contributors could articulate benefits that the community received. This is important as believing that your participation in a group effort matters encourages prosocial behavior, including continued participation [57].

6.2.3 RQ3 Motives

In this section, we present the motives (from contributors) for their initial and continuing contribution on the site.

Motives for Contribution

Here we present the results of two questions, "Why did you start editing Cyclopath?" and "Why do you contribute to Cyclopath?". The question "Why did you register

Code	Start	Contribute
Fix Problems	30 (37.04%)	0 (0%)
Benefit Cyclopath	17 (20.99%)	23 (28.05%)
Benefit Others	17 (20.99%)	22 (26.83%)
Ideology	8 (9.88%)	20 (24.39%)
Benefit Self	5 (6.17%)	9 (10.98%)
Misc	4 (4.94%)	8 (9.76%)
Num. Applications	81	82

Table 6.5: Coding results of the survey question "Why did you start editing Cyclopath?" and Why do you contribute to Cyclopath?" The table shows the number of responses to which each code was applied (using the 66% agreement rule). Multiple codes could be applied to each response. 81 codes from 71 distinct users are shown for the former question, 82 codes from 63 distinct users are shown for the latter.

for Cyclopath?" is also relevant, but was discussed earlier. The only users who saw these two questions were users who said they had contributed. The results in Table 6.5 show the answers to both questions, why users initially contributed and why they continue to contribute. These answers were very similar with nearly the same set of codes. The one exception was that users often mentioned that they began editing in order to fix a problem with existing content. In contrast, users never listed this as a reason to continue contributing. Below, we discuss the different response codes.

Results

Fix Problems. Users mentioned two types of problems that spurred them to start editing. First was missing information.

> There weren't many notes on the roads in my neighborhood and I wanted to put in more information.

The second type of problem was incorrect information.

> Some point [*sic*] of interest were off by a block, so I moved them. Some notes were outdated, I.e. road construction being finished, and the note still said it was under construction.

Benefit Cyclopath. Some contributors were motivated to help Cyclopath and improve the site or content. In particular, contributors mentioned improving the site's

content, benefiting the Cyclopath community, and reciprocity (a form of group-generalized exchange [31, 44]).

> It didn't work very well. I wanted it to work better. I figured if there were others like me that collectively we could make it much better.

> I felt I had good stuff to share with other cyclists using Cyclopath - especially since I've found good stuff entered by other people!

Benefit Others. Other users expressed that they contributed (either initially, now, or both) to help others. In particular, several users mentioned that they believed that they had unique knowledge that would benefit others.

> Used my personal knowledge of conditions to make notes such as connections and barriers for other bicyclists.

Ideology. Ideological explanations were given by some users. These responses are consistent with prior research on both OSS [53] and Wikipedia [102]. We identified two types of ideological response. The first involved the open content approach of Cyclopath.

> Love the Wiki-inspired idea of collective knowledge and correction.

The second was identifying with the bicycling community.

> Because of the sense of solidarity and contributing to a collective project. The idea that, as cyclists, we're really "all in this together" in terms of sing an alternative-and often marginalized-form of transit.

Benefit Self. Some users offered self-oriented reasons for contributing, such as:

> To have it give me better directions on the routes I expected to use.

However, many of the responses coded "benefit self" were also coded "benefit others".

> It makes my routes more accurate and helps others find routes in an area they may not be very familiar with.

Summary

While the set of motives that led users to begin contributing to Cyclopath and continues to motivate them today is mostly the same, new editors cited fixing problems as an important initial motivation.

6.3 Discussion

6.3.1 Registration Motivation

Our research shows that consumers and contributors say they joined Cyclopath for different reasons. However, there are some additional events that may affect these results and are worth investigating further. Most notably, for the first three months that Cyclopath was open to the public, registration was required for all users before they could see the product. We believe that this may have affected results. In addition, like most research of this kind, the results display the reasons why people think they registered in retrospect. Were we to instrument the registration form with a "reason you're registering" field, we would be able to capture this more accurately.

We are also interested in learning how reasons for registering have changed over time, more generally than just from the alpha period to the beta period. This survey covers from the alpha period to spring 2010 and while there were significant changes and improvements to the site in that time period, since that time period some notable features have been introduced that had the potential to entice new contributors. (The biggest change in this respect was the introduction of discussions.) Understanding how this issue may change over the history of a site is important as decisions are made about which new features to introduce or whether to publicize different ways to use the site.

6.3.2 Individual and Community Benefits

We found that while both consumers and contributors cited similar individual benefits from using Cyclopath, the perceived benefit to the community differed significantly between the two groups. While we will discuss this further in Chapter 8, it is worth noting that it may be beneficial to create more transparency around the notion of community benefit.

6.3.3 Motives for Contribution

Contributors gave similar, but not identical, reasons for starting to edit and continuing to edit. Most notably, they said that they initially began editing in order to fix problems. This is another area that we are interested in learning more about. Specifically, if we identified each users first edit, we could then pop up a question about the edit following the initial editing experience. This would allow us to capture the detailed information about the edit in consort with the real time reasoning of the editor.

Over thirty-seven percent of the editors in the survey cited fixing problems as the reason they began editing Cyclopath. However, perhaps the map or the product must be in a certain state before this can be the case. That is, if you can't envision the product ever being good enough to give you or others good routes in return, you may not think it worth the time to make the edits. In 2010, following this survey, we launched Cyclopath for the Boulder and Denver areas. The underlying transportation network in our system, however, needed a lot of work, and combined with other issues, we believe played a big role in the overall failure of that product. One way that we could learn more about this issue would be to look at the responses to these questions in chronological order. Our interviews that we present in Chapter 7 found that editors who joined early on had often contributed all they felt they could by 2012, so we would expect that as time progressed, fewer contributors would begin editing to "fix problems".

6.4 Conclusion

Consumers and contributors gave different reasons for joining Cyclopath. Consumers were more likely to sign up to get bicycling information while consumers often signed up in order to edit content. The more users contributed, the more their reasons differed from those of consumers. This shows that perhaps it would be useful to highlight ways that users could contribute as part of the registration or pre-registration process.

Consumers and contributors felt they received similar benefits from participating in Cyclopath but contributors were more likely to think that the community also benefited from their participation. Both groups felt that they had received better bicycling information than without Cyclopath. Many consumers felt that the community had no benefit. Those that cited a benefit also mentioned that they had "spread the word" sharing both bicycling knowledge and information about the site with friends and colleagues. In contrast, consumers could often list specific benefits that community members received like access to better routes and up to date knowledge of road conditions. This suggests that perhaps

it would be useful for both consumers and contributors to highlight the benefit of contributions or the need for additional contributions.

The reasons that contributors gave for starting to edit differed from the reasons they continued to edit. Often users were inspired to start editing in order to fix specific problems they found. Once those problems were fixed, they continued editing to benefit other cyclists and because of their commitment to the cycling community. This finding suggests that while an interface that shows problems that need fixing may be useful in attracting new editors, new editors likely need more than problems in order to become serial contributors.

In addition to these findings, we have a number of design implications from this study that we present in Chapter 8.

Chapter 7

Cyclopath Core Contributors and Activity Theory

Prior research has shown that much of the content and value in communities like Wikipedia is produced by a small proportion of the community members: highly active core contributors. Furthermore, these individuals also take on the majority of community maintenance work, such as welcoming newcomers and ensuring that they follow community norms, policing content contributed by others to verify that it meets quality standards, arbitrating conflict, etc.

We build on and extend prior research on these questions by studying core contributors to Cyclopath. Our results shed light on who the core contributors are, how they use the platform and tools and conduct community maintenance, and how they are embedded in the bicycling community, offline and online, in Cyclopath and other community sites. We also compare our results to findings from studies of other open content systems.

For this work, we used the same activity theory framework as Bryant et al. [26] which allows us to easily contrast results. In doing so, we found different results than Bryant et al. [26]. For example, discussion about rules on Wikipedia is intense and people discuss general policy as well as content. In contrast, Cyclopath users rarely converse and if they do have discussions, they are much calmer than debates on Wikipedia and tend to be focused on a specific instance rather than general policies. Why do these results differ? We believe that they are due to differences in content, community, and the nature of the sites.

These comparisons to other systems also allow us to learn about issues of generalization. Much research on online communities assumes that all communities are the same or at least similar. By contrasting the results from Cyclopath with data from Wikipedia, MovieLens, and Everything2, we seek to show that detailed char-

acteristics of system and communities help to shape the user experiences within the community.

7.1 Data and Methods

7.1.1 Definitions of user categories

In this research, we use several terms when referring to different groups of Cyclopath users. Here we define them for reference.

Cyclopath users are all users of the Cyclopath website, registered and anonymous. **Cyclopath editors** are all users of the Cyclopath website who have edited or annotated the map. **Core contributors** are the 22 Cyclopath editors (the top 2.5%) who have made the most edits. (See Section 5.1.4 for additional information, but note that the number of Cyclopath editors has doubled since that research, so the top 22 are now 2.5%, not 5%.)

In addition, **Survey respondents** are the 396 Cyclopath users who responded to the 2010 survey. However, since no questions were mandatory, it is possible and likely that a question will have fewer than 396 responses. Ten of the 22 core contributors were also survey respondents. **Interview participants** refers to the seven Cyclopath editors who we interviewed in person. They were all core contributors and four of them had also taken the survey.

7.1.2 Usage Log Analysis

Research that required Cyclopath logs comes from the Cyclopath usage history dataset that includes all activity on the site between launch on May 15, 2008 and July 6, 2012. This dataset includes 15,998 edits, 13,123 of which were made by logged in editors. There are 906 editors in this data set. The remaining 2,875 edits were made by anonymous editors.

7.1.3 Survey

This research also relies on survey results from the survey also used for Chapter 6. As a reminder, this survey was developed and administered to Cyclopath users in March and April 2010. The survey had 60 questions ranging from standard scales to custom designed items. No survey questions were required to be answered.

As an example, there was a version of the Volunteer Functions Inventory (VFI) developed originally by Clary et al. [34] and subsequently modified by Fugelstad et al. [52] for use in online community research. The questions were grouped into six categories as described by Clary et al.. Those categories were Values, Understanding, Enhancement, Career, Social, and Protective. Three statements per category were shown to the user, e.g. I feel compassion towards others in need. For each statement, users rated how accurate or important each of the reasons was on a seven point scale.

There were also questions about bicycling habits. In particular we asked what portion of the year the user rode, how often they rode, and how far they rode. We also asked questions about other websites that they used, both bicycling sites (like Minneapolis Bike Love and Bikely) and Open Content Communities (like Wikipedia, Facebook, Twitter, and YouTube).

Finally, we asked the users questions about Cyclopath: how they learned about Cyclopath features, what features they used most often, and when they use Cyclopath.

In addition to the above questions, we also asked users standard demographic questions.

We had 396 users fill out at least part of the survey. The average age of the users was 41. 55% were male and 21% female (the remainder didn't answer the question). The users averaged four hours a day online and an additional five hours a day on the computer offline.

7.1.4 Interviews

To gain more in depth knowledge about core contributors, we conducted seven interviews in spring 2012. These were semi-structured, face-to-face interviews where all participants were classified as core contributors. Details on the participants are shown in Table 7.1. All the interviewees were male and they ranged from late twenties to late sixties.

During the interview we asked about cycling habits and preferences, familiarity of the participant with the Twin Cities and the Twin Cities bicycling community, the participant's use of Cyclopath, their perceived role on Cyclopath, editing idiosyncrasies, and their thoughts on the Cyclopath community. In addition, to allow direct parallels to Bryant et al., we also asked questions used by Bryant et al. for data collection in [26]. These questions were used with the permission of PI Andrea Forte and were adapted slightly for Cyclopath.

A copy of the interview script is provided in Chapter B. Please note that because the interview was semi-structured, all interviews were slightly different and in none of the interviews was every question asked.

Participants	Days Active	Revisions
Participant 1	1218	500+
Participant 2	905	0 - 125
Participant 3	568	250 - 500
Participant 4	643	125 - 250
Participant 5	1148	125 - 250
Participant 6	1064	500+
Participant 7	1015	500+

Table 7.1: Interview Participants. Since Cyclopath revisions are public, we bucketed the umber of revisions to preserve participant anonymity. Number of days active is computed as $date\ of\ last\ view - date\ of\ account\ creation$.

7.1.5 Framing our findings with Activity Theory

Activity Theory is a descriptive theory for studying and interpreting human activity. It has been useful in HCI studies [79] and was the very successful primary framing for Bryant et al.'s Wikipedia study [26].

At an individual level, Activity Theory looks at a *subject* who engages in an activity directed at an *object* mediated by a *tool*. However, Activity Theory also has a framework for a community level which introduces the constructs of *rules*, *community*, and *division of labor*.

We have chosen this framework to present our results because it was a useful was to organize and interpret our findings and to contrast our results with [26]. We report on subject, object, community, rules, and tools. Our interviews, surveys, and usage logs do not adequately address division of labor, so we do not present any results about that, but we mention it in our discussion.

7.2 Core Contributors and Activity Theory

7.2.1 Subject

Core contributors may often be viewed as experts, as in Bryant et al. [26]. We wanted to look at and differentiate between two types of expertise that Cyclopath core contributors may have. The first is the content domain, that is, bicycling. In order to evaluate expertise in the content domain, we looked at whether core contributors were dedicated bicyclists. The second type of expertise is in the technical domain, online open content communities. To evaluate expertise in the technical

domain we looked at whether core contributors were familiar with and engaged in other online open content communities.

Content Domain. Previous work [24, 99] categorizes bicyclists by factors such as distance ridden, frequency of rides, and conditions ridden in. We use the same three factors in our definition. Cyclists are considered dedicated if they meet one of the three conditions below.

> Distance: Ride 20 miles or more on average
>
> Frequency: Ride one or more times a day (during riding season)
>
> Riding conditions: Ride at least nine months of the year (forcing them to ride during at least part of the cold, potentially snowy, months)

These results combine survey data (cycling behavior) with log data (core contributor status). 186 of 299 survey respondents are classified as dedicated cyclists by the above definition There is no statistically significant association between being a core contributor and being a dedicated cyclist ($\chi^2 = 0.0047$, $df = 1$, $p = 0.95$). Even if we look at distance, frequency, and riding conditions independently, there is still no significance.

(A survey by Bike/Walk Twin Cities found that 20–36% of cyclists continued cycling over the winter[33]. This compares to over 50% of our survey takers. We realize that this means that our survey takers, and perhaps Cyclopath participants, aren't representative of the entire cycling population of Minnesota and this shows that our user population consists generally of more dedicated bicyclists.

Technical Domain. We wanted to assess Cyclopath users' familiarity and participation in open content communities. In our survey we asked users whether they were a member of, used, and/or contributed to Wikipedia, Twitter, Facebook, or YouTube. In our user population we found that the only significant differences were with Wikipedia.

A significant number of core contributors used Wikipedia ($\chi^2 = 4.47$, $df = 1$, $p = 0.03$) and contributed to Wikipedia ($\chi^2 = 5.27$, $df = 1$, $p = 0.02$). In addition, core contributors were marginally more likely to be members of Wikipedia ($\chi^2 = 3.46$, $df = 1$, $p = 0.06$).

In our interviews, we were able to discuss this more in depth. One of our interview participants contributed to Wikipedia. Participant 5 edited Open Street Map (OSM), another open content mapping project. He wanted to be able to edit in Wisconsin, because he often bikes there but Cyclopath doesn't cover the area. (Cyclopath actually has measures in place that block users from editing in Wisconsin, which several interviewees mentioned.)

> [While on a trip] I discovered the trail is closed because of some old train wreck... They haven't finished repairing the trail. So I actually went around the tracks and found out that they actually had a detour in place... and I thought, "Well, the maps [Open Street Map] ought to reflect the detour." So I did that. (Participant 5)

While they may not be active contributors to other open online communities, core contributors do value knowledge contributed by their peers. This is consistent with prior research [116] which found that bicyclists used information from other bicyclists frequently The users also valued and trusted information obtained this way.

> In most cases I trust it more because it is people who update it for the most part.... Just knowing that the information has been entered in by people is comforting to me. (Participant 4)

Survey results were consistent with this. All responding core contributors said they used Wikipedia. When asked about how much they trusted services (Cyclopath, Wikipedia, Facebook, and Twitter), all seven core contributor respondents listed Wikipedia or Cyclopath as the most trustworthy site. When asked to rank the value of the same set of websites, five out of the seven said Wikipedia was the most valuable and the remaining two found Cyclopath to be the most valuable.

Identity. We were surprised to find that core contributors saw themselves as users, not contributors or elite users. When asked "How would you describe your role on Cyclopath?", four of the interview participants used the term "users". Three also used qualifications like "engaged user", "user that uses Cyclopath a lot", or "occasional contributor". Only one of the seven said "contributor". One participant called himself a "lurker".

(Note: Some users used multiple terms, or, in a list of terms listed multiple identities, so we describe a subset of these terms.)

7.2.2 Object

In our case, the Cyclopath system is the object. However Cyclopath has two main interaction paradigms. Like Wikipedia, Cyclopath is an open content system based on end users editing content. But Cyclopath is also a computational service: computing bicycle routes [115]. These two paradigms are linked. User input influences routes that are computed. The paradigms have implications for users as well. Like a recommender system, e.g. MovieLens, users benefit personally from data they input: they receive routes that better match their preferences, especially when they input bikability ratings. Note that Cyclopath users don't have to input any data in order

to get a route, unlike MovieLens. Without personal ratings or data, Cyclopath uses aggregate ratings and objective properties of trails and roads. In MovieLens, all movie ratings are private and only displayed in aggregate while edits to the movies or posts in the Q&A sections are public. However, unlike MovieLens, the public effects of user edits are more prominent in Cyclopath. Many types of user edits are public: geographic edits, notes, discussions, and tags.

We were interested in how the user input influencing the routing engine may affect attitudes and goals of core contributors.

Building a Better System. Core contributors participate because they want to build a better computational system. Contributing gives them power to improve the output (e.g. routes) of Cyclopath.

> I just liked the idea of this routing tool that was actually pretty broken, but I could make it work by fixing things. (Participant 1)

> It is not designed to be usable at the outset: it is designed to give powers to users to make it work. (Participant 3)

In some cases, users added information to Cyclopath so that the route-finder (here referred to as Cyclopath), not others, could use it.

> In my neighborhood, there was a little sidewalk that actually gets you to Target and without that there, Cyclopath wouldn't know it existed... So I made that little that path to say it actually goes there so that Cyclopath can actually see it. (Participant 3)

The brokenness of the computation of Cyclopath (its ability, or inability, to compute routes with the information it had) served as a motivator to core contributors to "fix it" by contributing. This was especially true early on. Half of the core contributors started using Cyclopath before the public release on August 1, 2008. The state of the map at that point was poor. There was a lot to be done. A number of interviewed core contributors mentioned that they have reduced their editing over time because there was less to do (or to fix).

> I still think I edit more frequently than anybody else, but I think I probably edit less, a fair bit less, than I did, largely because the area that I know well is already in pretty good shape. (Participant 1)

The amount of information and quality of the information in the map varies from location to location. Parts of the Twin Cities metro area where Cyclopath users are the most interested in and familiar with have great coverage. Outer suburbs and

rural areas have had much less attention paid them, but are also routed through less frequently.

These results serve to supplement the results from Section 6.2.3 which show that 37.04% of surveyed editors began contributing to fix a problem, however no editors have continued contributing solely to fix problems.

7.2.3 Communities

Cyclopath provides features for community interaction in addition to the map and route finder features. When Cyclopath launched in May 2008, a Google Group (mailing list) was also launched. This was used for discussion among users and developers. In particular it was used for bug reporting, troubleshooting, and discussions about governance (see more in Section 7.2.4). In addition to the Google Group, there was also a user editable text wiki that allowed for user and developer documentation. Finally, in April 2010, an in-application discussion system was launched. This system functions as a forum (although the Google Group is still used) and allows users to tag discussions with words or locations (users can tag a certain segment of a certain route).

However, in comparison with Wikipedia, these community resources are rarely used. As mentioned in Section 3.2.4, every Wikipedia article has a talk page where discussions regarding content and organization of the page happen. In addition, every user can create a user page with an associated user talk page. Talk pages are "the most frequently used communication channel on the Wikipedia" [26]. There is also the Village Pump, an area for Wikipedia core contributors to discuss "technical issues, policies, and operations of Wikipedia".[1]

Cyclopath has a potential advantage over Wikipedia in that Cyclopath is closely tied to an offline, geographically local community. From August 1, 2011 to July 31, 2012, Google Analytics shows that Cyclopath had 32,011 views. 27,447 of those were from inside the state of Minnesota, 14,709 from Minneapolis and 4,705 from St. Paul. This means that the person who bikes past you on your commute could be the person whose edits helped you find your route. In Wikipedia the primary page editor could be halfway across the world.

Online Bicycling Communities. We wanted to know if Cyclopath core contributors participated in other online bicycling communities. In our 2010 survey we asked users if they participated in Minneapolis Bike Love (a local bicycling forum), Bikely (a route sharing site), Bikemap (another route sharing site), Gmaps-Pedometer (a route distance calculator), MocaTrails (a local off-road club) , or other online bicycling communities.

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

A chi-squared analysis showed that core contributors weren't more likely than other users to participate in these online communities.

> "I do participate in the Minneapolis Bike Love message board... Mostly just reading it, I usually don't post... I just like to stay up to date and it's always nice to hear what people are up to, interested in.... I tend to focus on threads involving safety or conditions.... It's purely functional use." (Participant 4)

Offline Bicycling Communities. In contrast to the above, core contributors were active in the local cycling community offline. Participant 3 said the following when he was asked if he was attached to the local cycling community:

> Not at all... To me it's just not feasible in one day to spend three hours traveling to an event. (Participant 3)

But has he continued, it was obvious that he was fairly active in the local cycling community.

> I go to events and I am participating in the cycling community that way... but there's no attachment... I went to a bike summit... I have an LCI [League Cycling Instructor meet up] down at Peace Coffee... (Participant 3)

Interviewees who did say they were attached to the local cycling community were often involved in cycling advocacy groups, not just riding groups. Participant 2 talked about what he got out of participating in a local bicycle coalition:

> I think I get more opportunity for community input affecting the city and the neighborhood and where I live and then also kind of get to work to foster change to benefit the community as a whole and provide more options for cycling. (Participant 2)

Participant 1 was the only user we talked to who was active in both online and offline cycling communities.

> Well, besides Cyclopath, Minneapolis Bike Love, the Minneapolis Bicycle Coalition, the bike alliance of Minnesota... (Participant 1)

Reasons for Volunteering. In trying to understand more about core contributors and other contributors, we wanted to understand why Cyclopath editors were editing

and volunteering. To do this, we administered a version of the Volunteer Functions Inventory (VFI) [34] as modified by Fugelstad et al. [52].

In the VFI, we were particularly interested in two factors. The first of these was value. This is a factor that would come up if you were volunteering because you are acting on personal values. This would likely be a factor if you were helping people less fortunate than yourself or doing humanitarian work. The second factor was social. This factor is strong if you are volunteering to strengthen social ties or because your friends volunteer. The higher the number for the VFI factor, the more the survey participant felt that the factor was an important reason for them to volunteer or an accurate description of why they volunteered

We calculated correlation coefficients using the VFI factors against the total number of revisions that the user had made. The number of revisions an editor made was negatively correlated with the value factor ($r = -0.20$, $p = 0.002$) as well as the social factor ($r = -0.15$, $p = 0.032$).[2]

This result shows that the more revisions a user has, the less likely that that user is volunteering due to values or to receive social benefits. This is similar to the results found by Fugelstad et al. as found on MovieLens [52].

Awareness of Others. Perhaps because core contributors are not participating on the site for social reasons, they know few contributors. When asked whether they know anyone of Cyclopath or recognize the usernames of other editors, four interviewees said one to "a handful" of usernames were familiar.

> Sure, I recognize some of the names in Cyclopath. I don't know any of them though. (Participant 1)

Some of the users used this name recognition to identify users in the recent changes list. This allowed them to focus their effort on changes that might be suspicious, instead of on changes that were likely well done.

This differs from the findings of Bryant et al. of core contributors [26] as well as findings of [129] in Wikipedia and Everything2. This research found much higher awareness of other editors. On Everything2, some users were "status builders" or "personal relationship builders" and were very conscious of their audience and had personal relationships with other users on the site.

View of System. We asked people how they would describe Cyclopath to others. The answers we got were all similar to the quotes below.

> Cyclopath is an alternative to Google Maps with editable functions where locals can edit to show its accuracy. (Participant 3)

[2] Regarding correlation coefficients, Rosnow and Rosenthal have argued that small effects can have a powerful impact on outcomes over time, especially in the aggregate. [120]

> Cyclopath is a site that lets you find bike routes. It is also a wiki, community-supported. (Participant 6)

> It is a cycling geowiki... [To non tech-savvy people] it is a way of finding a bike route that you can modify and set parameters. (Participant 7)

We were surprised that none of the participants explicitly described Cyclopath as a community, but the core contributors did not see Cyclopath as a medium for social interaction, but rather as a resource that is modifiable by its users.

7.2.4 Rules

Rules in the context of Activity Theory refer to socially defined and enforced norms, conventions, and social relations [7].

Some of the rules in Cyclopath were derived from other open content communities. For example, Wikipedia has a rule that no private data should exist in a public resource. While there is difficulty in enforcing this type of rule from a system perspective (simply not allowing said edits), this type of rule is enforced, in general, by other users or the users themselves after the fact. This is the reason for 147 of 628 (23.4%) reverts done in Cyclopath. (These are reverts that contain the word "private" in the comment accompanying the revert.)

Establishing Rules and Norms. Cyclopath core contributors want rules to help create a uniform resource and to handle controversies, questions, or problems incurred while editing. One of the recurring discussions regarding rules and norms revolves around terminology used in tags. Tags are used for routing and a uniform vocabulary mask using tags, applying tags, and routing with tags easier. For more information on tags, see Chapter 4 or [128].

The following is a note that appeared in the Cyclopath discussion section in an attempt to try to determine whether the correct tag should be "bikelane" or "bike lane".

> 1) Is one term of other preferred? 2) If so is it documented anywhere? 3) If not, should it be? 4) If so, where? In the [text] wiki? 5) If so, would you expect users like me to just jump in and start proposing (in the wiki) some standards? (Cyclopath Discussions)

As a result of the subsequent discussion, users developed conventions for when the "bike lane" tag should be applied.

> It's been our convention to mark a two-way road as having bike lanes only if it has lanes in both directions of travel. Pleasant has only a general travel lane (marked with poorly placed sharrows) in the southbound direction and a counter-flow bike lane northbound so it doesn't seem a good candidate for the bikelane tag. (Cyclopath Discussions)

One user we interviewed gave another example. He had spent hours simplifying roundabouts, distilling the eight or more segments to a simpler four way intersection. In retrospect though, he wondered if roundabouts needed to appear in detail or if his simplification was acceptable. This is something that is very rare, so the community has not developed any visible guidelines. That said, the user didn't want to overstep or do something that would be considered "wrong" by others.

> I never know if that's acceptable, what the proper way to deal with situations like that is. (Participant 4)

Benefit of Lack of Rules. A lack of system enforced rules allowed for some innovation on Cyclopath. One interview participant added links to pictures of complicated intersections to points near those intersections. This is discussed more in detail in Section 7.2.5.

While some users may find the enforced rules useful, environments with more (and more strongly enforced) rules and policies (e.g. Wikipedia) may not have allowed for some of this innovative behavior, instead squashing such behavior.

Opacity of Editing Rules. Cyclopath rules are, in general, difficult to find and perhaps to follow. This makes it both hard and nerve wracking for users to begin editing.

> I was terrified to edit and then I tried a few things, and my terrors were justified... Well, make a mistake, do something that is wrong that is going to inconvenience someone else, that somebody else needs to clean up your mistakes... There's so many discussions about reversions and tags... I read this and I'm terrified to start, absolutely. (Participant 5)

Another participant had an edit reverted because he didn't know what the community norms were.

> Those [mistakes] are scary. You don't want to edit again for a while. It's like I don't know what I'm doing anymore. (Participant 4)

Dangerous Consequences. Cyclopath is geographic and is used for finding routes, this allows for editing situations that are unique to this system. Interviewees were

worried about getting things on the map wrong. In essence, editors wondered, "what happens if due to my screw up, someone goes on the wrong road and gets hurt". To our knowledge, this has never happened, but core contributors are worried about it and think about it while editing.

> I hesitated for a long time to rate the Washington Ave. Bridge. I know that it's not legal to ride there but I am sometimes willing to do it so I would rate it something better than "impassable". However, when I did that the default rating change [sic] from "impassable" to "poor". Does that make me in a way responsible when Cyclopath routes an unfamiliar rider over that road? (Google Group)

There are places (Google Group, Cyclopath discussion forum) where rules can be discussed. Core contributors were frustrated that rules were often hidden and they weren't able to learn the norms until people (in discussions or revert comments) came and told them explicitly.

Despite these problems and issues, Cyclopath is young and the rules and available resources are evolving as the community ages.

7.2.5 Tools

The Cyclopath web interface has a number of tools for different functions. Users can rate bikability of roads and trails, add tags, add text notes, edit map geography and topography, create new segments, and connect segments with existing segments. (For more details, see Chapter 4.) Yet the number of tools and complexity of the tools make editing in Cyclopath harder than other systems such as Wikipedia.

Learning the Tools. Our survey asked editors how they learned to use the interface and editing tools. (Note that users could list multiple responses.)

321 (81%) taught themselves, 104 (26%) learned from the text wiki, and 43 (11%) learned from existing video tutorials. Of the nine core contributors who took the survey, eight taught themselves (while sometimes referring to the text wiki or video tutorials if they couldn't figure something out).

Using the Tools. When they started using Cyclopath, even the core contributors thought the editing tools were hidden and awkward. Half of the core contributors were early adopters, joining prior to the public launch and at that point the interface was still in flux.

> Earlier on, there were problems. It was quite clunky, especially doing a lot of edits... Like breaking and merging intersections or moving a whole segment and things like that... (Participant 6)

Since this time, core contributors agree that the site has gotten easier to use and the process of editing is less clunky and more robust than it was initially However, this shows the potential importance of a good initial user experience. Had the map itself or the tools been harder to use, adoption may not have been as wide.

Wanting More Tools. We found that core contributors are opportunistic in nature. They are interested in additional tools and often think in terms of Cyclopath when riding. In particular, they think about how they can enter data relevant to where they are at the time and how they can enter information about problems they encounter. One interviewee drew an entire area of the map freehand. He drove around the newly developed area often for other reasons.

> The... area had no satellite imagery and I drew it out freehand just by driving around because I went there for work anyway. Get back home at the end of the day and draw it out. When the satellite imagery actually came out, someone even commented that whoever did the drawing did a really great job. (Participant 7)

Participant 5 edited Open Street Map, in part because Open Street Map is able to import GPS traces and use these traces to automatically make map edits.

> If there's a detour on the trail, I like to ride it so that I get a GPS trace so that I can edit Open Street Map or Cyclopath. Snipping GPS traces and adding info to Open Street Map is easy. (Participant 5)

(Since the interviews, Cyclopath has launched a mobile app that allows users to upload GPS tracks.)

One core contributor wanted keyboard shortcuts to make editing faster. When Cyclopath developers said they didn't have the time, he downloaded the open source Cyclopath software and wrote the code himself. The developers merged the code into the main code base and the shortcuts are still used.

Participants also mentioned that while riding they would discover edits that they wanted to make, but they weren't always able to recall these when they got to their computers to edit.

> On really long rides, it is really hard to remember sometimes and I don't have any fancy technology that allows me to update it while I'm riding, which would be really nice. (Participant 4)

Participant 3 solved this problem by carrying a notepad or camera around to record problems as he encountered them.

> That's why I take pictures... so that I can take a picture of what I see and then come back and [record edits on Cyclopath]. (Participant 3)

While the Cyclopath tool set is not ideal for all users yet, the group of core contributors has learned and embraced the tools and started to request additional new tools.

7.3 Discussion

People familiar with Activity Theory may notice that we have not analyzed core contributors with respect to Division of Labor. We found during our research that while we think of, and refer to, Cyclopath as a community, the people in the community tend to be fairly independent and not interconnected. The core contributors were barely aware of each other and definitely not negotiating anything in regards to division. This is in contrast to a site like Wikipedia, where editors often have defined roles and spend time negotiating changes to Wikipedia pages, among other things [26].

7.3.1 Subject

We found that core contributors were not the most dedicated cyclists, but they were similar to each other in that they had belief and engagement in open content knowledge production, both in Cyclopath and beyond.

This result was somewhat surprising to us as we'd expected core contributors to be more dedicated cyclists than the other survey respondents. However, examining our data, in conjunction with the Bike/Walk Twin Cities data [33], we realize that our sample is likely skewed. If we were to compare our core contributors to the average cyclist in the Twin Cities, it is likely that the core contributors would be much more likely to be dedicated cyclists than non-Cyclopath users.

Belief and engagement in open content is a characteristic we had expected to see in our core contributors and we were not surprised by these results. There may be ways to leverage this trait when building systems relying on user contributions.

7.3.2 Object

Users were motivated by creating an effective computational resource and fixing problems. Fixing problems has been previously found to be a key motivation for

participating in open content systems [102]. In Cyclopath this is definitely compelling. In addition, the Cyclopath routing engine (e.g. computational aspect) is important and sometimes leads to cycles of requesting routes, editing, and then re-requesting the original route.

Many systems requiring user input rely on problem-fixing to draw in users. However this requires a delicate balance between problems and functionality. Without functionality, users may not see the value in contributing or participating in the system. However, if the content is fully formed, users may not realize where the opportunities for contribution lie.

In the Implications chapter Chapter 8, we dwell a little more on the success of a project highlighting problems that need to be fixed, but keep in mind that as systems differ, some projects may benefit more from this sort of approach than others.

7.3.3 Community

Despite the way Cyclopath researchers think about the site, core contributors don't view Cyclopath as a community and don't seek out social interaction. In addition, they are not motivated by the prospect of social interaction.

When Cyclopath was introduced, Minneapolis Bike Love was an existing online discussion community for Twin Cities bicyclists. Cyclopath was an information resource, in contrast to Minneapolis Bike Love's discussion forum and social interaction. In talking to users, it seems like this dichotomy still exists. This is consistent with prior research on MovieLens. When social interaction was introduced on MovieLens experienced users, who saw MovieLens as a film recommendation site, did not respond positively to those features [119].

While many sites today are trying to figure out how to build community and incorporate social interaction, users may not want those features or the features may be conflicting with other systems. The complete ecosystem that a site exists in can be an important factor when designing or changing functionality.

7.3.4 Rules

Some Cyclopath rules are based on those from other open content systems, but Cyclopaths unique features led to unique rules and a desire among core contributors for more rules.

One of the few situations where social interaction mechanisms (Cyclopath discussions) were used was to discuss rules. However there is a difference in the discussion of rules on Wikipedia and Cyclopath. On Wikipedia there's intense discussion about

general policies and content of some articles. On Cyclopath there is causal discussion about general policies. There are several factors that may lead to these differences. The first is scale. Wikipedia is much larger than Cyclopath with thousands of active editors in English Wikipedia alone, whereas Cyclopath has few dozen. In terms of controversy, on Wikipedia different people disagree about what content to include and present. However on Cyclopath there is little to no public disagreement or conflict among Cyclopath editors. Finally, Wikipedia struggles with a vandalism problem. It attracts malicious editors who make damaging edits. Policies and tools are in place to deal with this, though initially they didn't exist. Currently Cyclopath has seen no vandals on the geographic wiki (though the text wiki has been spammed).

While we will touch more on issues of generalization in the design implications chapter Chapter 8, this is an area where communities may differ significantly. For communities like Wikipedia and Cyclopath, letting the community make their own rules has, for the most part, worked. In other communities, such as formspring or reddit, it could be argued that, while they don't have many rules, the lack of rules has led to a much more volatile community.

We feel that the organic community-creation of rules in the Cyclopath community has benefited the community and the system. As the community has developed rules, they have also, as shown in this chapter, begun enforcing their set of rules. This is a necessary part of almost every community, so, as with many of these factors, it is one part of the puzzle.

We heard from our participants that the potential for repercussions in the physical world made them pause and sometimes delayed their editing. While we think it likely that this is a significant differentiator of a geographic wiki, we don't have similar data for Open Street Map, Wikipedia, or other contribution sites.

7.3.5 Tools

Cyclopath has various tools ranging from easy to use (rating a block) to quite hard (dividing blocks into multiple segments). These geographic editing tools are unique to Cyclopath and Open Street Maps. (However the two systems use different tools.) These tools allow users to correct and extend the map. Editing has quantitatively improved the routes Cyclopath generates [114, 115]. The complexity of tools creates an entry barrier.

The core contributors suggest that there's a need for additional tools, in particular tools that would be of use for recording while riding. An example of a tool that may be of use would be Biketastic [118]. Previous work on a citizen science river monitoring application showed a similar need for additional tools [121]. Now that

Cyclopath has a mobile app with GPS, it would be possible to incorporate note-taking functionality into the app though this has not been done yet.

7.4 Conclusion

Cyclopath core contributors aren't who you might think. They aren't the most dedicated bicyclists (riding long distances, year round, or riding multiple times a day), but they are engaged in open content systems and believe in open content as a form of knowledge production. In addition, the Cyclopath core contributors that we interviewed were active in the local cycling community, not just as members or leaders of riding groups but also as participants in advocacy groups.

Characteristics of the community matter. In the case of Cyclopath, the site was launched with lots of issues on the map. (Note that this was not intentional, but a result of messy data.) Many users were motivated to begin contributing to the site because there were problems that needed to be fixed. As the site has evolved, there are fewer problems, but the tools provided by the site have matured and now allow for additional interactions and forms of contribution. In the same way, when Cyclopath started, we didn't launch the site with rules for contributors, instead those rules sprung up from the community itself. Analyzing with respect to Object, Rules, and Tools and contrasting with other existing systems, it is clear that these features and aspects of a site evolve and grow, but also help differentiate between communities.

What is a community?. In all our work on Cyclopath, we discuss the site as a community of users. While we as researchers may believe that Cyclopath is a community, the users do not. Users also don't go to Cyclopath seeking social interaction. This could be for several reasons. First, according to the Volunteer Functions Inventory, the core contributors were not motivated by the prospect of social interaction. Second, the Twin Cities had a well-established online cycling community before Cyclopath started. These likely both contribute to Cyclopath not being seen as a community.

In addition to these findings, we have a number of design implications from this study that we present in Chapter 8.

Chapter 8

Design Implications

Throughout the dissertation research, many design ideas and implications have arisen. These have been placed together in order to better display the potential impact of this research. The implications are focused on helping to build or grow online peer production communities, such as Wikipedia and Cyclopath.

8.1 Wikipedia core contributors

Our work on Wikipedia focused on three different areas: work quantity, work quality, and community work. In all of these areas, we discovered findings which may be useful to the community at large when trying to create or improve online peer production communities.

8.1.1 Core contributors and Work Quantity

We found that while core contributors perform more work than others, by definition, they produce more from their first day as a registered editor. Since 62% of registered editors make one edit on their first day and no subsequent edits, we were interested in ways that communities might be able to entice these users to return or to find work to do.

Encourage new editors to return. One of the changes with a potential to increase work quantity later in an editors life is encouraging new editors to return after their first (or second) day.

There are several ways we envision this happening. First, when posts are made to a user's Talk page, the default doesn't notify the user outside of the Wikipedia

ecosystem. If the default was to notify the user via email, the user may come back to learn more and may then make more edits. Another option would be to automatically add pages to a user's watch list and send email updates when watched pages are updated.

These options are common practice on LinkedIn and Facebook, and while users have occasionally protested the notifications, research has shown that users view notifications as a way to be passively aware of what's happening [65]. If these messages and notifications are well-crafted, they can get users to return and to do work [22, 60].

Direct the initial activity burst. Given the amount of activity that users perform in their first few days, it may also be useful to point new editors to work that they tend to be most successful with. This could mean finding poorly written pages for editors to fix up if they've been improving grammar and spelling or punk music pages for editors who've been working on punk rock articles.

This would have an added benefit, because if users are routed to niches, it is possible for them to feel their contribution is substantial. Social psychology [67] and social computing [22, 86] show that if people feel that they have made an unique contribution it significantly increases the amount of work that they do.

Cosley et al. [35] found that welcome messages help with sustaining an engaged community and increasing retention. They also found that SuggestBot, a recommender system that helped people find work on Wikipedia with a customized to-do list, yielded four times as many edits as a random to-do list [36].

Outside of Wikipedia. This research, while conducted on Wikipedia may also apply to other sites. In particular, being able to notice power users early on may be very useful when picking moderators and administrators. Identifying these core contributors may also help ensure that potential prolific contributors are not ostracized, but rather, embraced.

8.1.2 Core contributors and Work Quality

Our analysis of Wikipedia core contributors and work quality showed that core contributors have higher quality edits than average contributors, but that the quality of the core contributors' edits remained fairly consistent, not improving over time.

Difficulty of Measuring Quality. Quality is something that is difficult to measure. In this research we didn't measure an objective quality, but rather we measured perceived quality by other editors in Wikipedia. Note that the additional metrics we discussed in Section 3.2.3 are also proxies for quality, not true objective metrics. Future research could be done to measure quality using other metrics, potentially

discovering new results.

One option would be time metrics like PWV [113] or persistence over time [15]. However, we expect that using these metrics would give a substantial advantage to early adopters.

Another option is Wikipedia assessment quality. Using this would allow us to learn whether editors become more active editing higher rated articles as they age.

Outside of Wikipedia, experience is often a proxy for quality. One example of this is the work of Ducheneaut et al. [41]. In their research, they found that Python developers get better as they age. In addition, the community itself matters. Some communities have a learning effect that influences quality and some communities remain static or unstable [41].

8.1.3 Core contributors and Community Work

We found that Wikipedia core contributors did not diversify their work and start editing in more namespaces as they became more experienced.

If using Talk pages is important for community function and community building (which we think it is), it may be useful for Wikipedia to promote Talk pages within the interface. For example, when someone opens the content editor for a Main page, there could be a summary of the most recent or most interesting conversations about the page.

Online coordination is an issue for many communities. On Wikipedia, coordination often happens on Talk pages or within WikiProjects [26, 30, 49, 130], but coordination is also an integral part of OSS [38] and many other peer production sites. There may be ways to promote unused or under utilized methods of communication in these communities.

Despite the lack of changes in namespace diversity, we found that Wikipedia core contributors did invoke community norms more as they gained experience.

The tie between communication, invocation of norms, and becoming a community watch member may hold true elsewhere online. If so, it is important and useful to be able to identify potential leaders and enforcers early in order to retain them and help them find positions that suit them. For example, Farzan et al. identified potential leaders in tax-related forums and found that these leaders were ten times more active than others in the forums [45].

8.2 Cyclopath Core Contributors

Our related work on Cyclopath core contributors found some evidence of educational lurking and also found that Cyclopath core contributors act like Wikipedia core contributors. The work on viewing and editing found that viewing and editing are related and that people who edit extensively reveal their viewing behavior (to an extent).

The finding that Cyclopath core contributors act like Wikipedia contributors is an important one for several reasons. First this shows that existing research may be able to benefit people running online communities. Second, it shows that studying small sites may be able to give us insight into larger sites and vice versa. While it is definitely the case that generalization is not always possible, nor accurate (see Section 8.4.1), the idea that some communities can help us gain insight into others is useful. This is especially the case for communities in their early lives and communities with limited access to data or users.

8.3 Cyclopath Motivations

In our research on Cyclopath motivations, we found that consumers and contributors, while receiving similar benefits from Cyclopath participation, joined for different reasons. In addition, Cyclopath contributors were more likely to think that the community had benefited from their personal participation.

The results from this study suggest a number of ways to enhance participation in online open communities. We try to focus on techniques for active contribution as well as enhanced participation methods for consumers.

8.3.1 Active Contribution

The first area of enhancement concerns active contribution, like editing articles in Wikipedia or fixing a bug in Firefox. The following methods suggest general, empirically based techniques which may enhance contribution.

Fix Problems. From our survey, we know that many users get started editing (or continue editing) because they find problems with Cyclopath's maps. In fact further research by Masli and Terveen found that 48% of feedback on Cyclopath routes contained negative evaluations. Their work also found that over half of users who submitted feedback also marked specific problem segments and/or provided an explanation for why they wanted a different route [90].

We could promote editing due to map problems by highlighting potential problems in the map or route-finder and inviting users to help fix these problems. Earlier work by Priedhorsky et al. found that highlighting missing intersections led to users editing the map [114]. We could highlight intersections along the route a user requests in order to make this more relevant to users. On Wikipedia, it may be useful to indicate which parts of an article have problems (e.g. paragraphs that are written at an inappropriate level).

Benefits Self. The current Cyclopath interface does a poor job of conveying to users what benefits they would see from contributing. If we emphasized (and showed) that information entered by a user directly benefits that user, users may be more inclined to contribute.

On Cyclopath, this may be as simple as letting users know that entering ratings results in better, more personalized routes. On MovieLens, entering ratings yields better movie recommendations.

This benefit to self appeal seems like it should work, both on intuitive and theoretical grounds, but it hasn't always worked in prior research [86, 117]. This prior work combined with these survey results show that more experimentation may be useful.

Benefit Cyclopath. Reminding users how they benefit from information entered by other Cyclopath members and then inviting users to reciprocate may also yield more contributions. Appeals to reciprocity, both direct and generalized, are powerful motivators to action [31, 32].

On Wikipedia, this could take the form of a statement on an article. "This article is brought to you by the efforts of over 50 volunteer editors. Please click here to find out how you can share your knowledge to help others."

Benefit Others. While benefiting others is similar to benefiting Cyclopath, concentrating on the benefit to others focuses on other users not the system in general.

Another thing that might be used to help elicit contributions is reminding users that their contributions can directly benefit their friends or other users "like them". Prior work has shown that this can be effective [117].

As to how this might be implemented, on Cyclopath the site could tell users that other cyclists in their neighborhood or others who ride the same trails would benefit from their contributions. On Wikipedia, readers could be told that others who are interested in the topic would benefit if the reader improved the article.

Ideology. Appealing to the shared values of a community may also be helpful in encouraging contributions. For example, Cyclopath appeals could relate to the real-world bicycling community or could appeal to the Cyclopath users who use Wikipedia. Both of these suggest powerful values. Other examples of values Cyclopath users may embrace include health, fitness, and sustainability. Similarly,

Wikipedia users may respond to appeals that feature commitment to open content or information sharing.

8.3.2 Enhanced participation by information consumers

A number of consumers said that their use benefited the community because they "spread the word" about Cyclopath. With this in mind, designers should make it easy for users to share information about the site. This may take the form of an "invite a friend" feature. Other options may let users publicize results (routes) they get from the system or publicize their own activity on the system. Cyclopath users could share a route they got and liked via email, Twitter, Facebook, or other social media. Currently users can get a link to their route, but they have to then go to the outlet of their choice to share it.

This functionality would benefit Cyclopath in several ways. It lets people know about Cyclopath and would hopefully let people know that the sender is positive about Cyclopath. On a site like StackExchange, this may involve users who asked questions being able to share the question and some of their favorite responses.

8.3.3 New forms of participation

In the survey responses, people often told stories to help explain how they had benefited from using Cyclopath.

> "I've quickly learned how to find"child-friendly" bike routes to different locations."

> "My ability to route has improved dramatically. I have a better understanding of how bits of the cities are connected in terms of biking. I have a better understanding of how other people view bits of the cities from the perspective of cycling."

> "A single route search has been of enormous benefit to me – I discovered a shorter, flatter, safer, more scenic route for my daily commute than I had been able to locate on my own, and also happened to discover a park I'd been looking for at the same time."

> "I only knew of one or two ways to get to work before using cyclopath and there is one HORRIBLE hill on my ride that I wanted to avoid. By using cyclopath I could find another path that made my ride so much better by not only avoiding horrible hills, but I could also select better

> biking environments such as safer paths with wider shoulders, a bike path, less traffic etc....surprisingly without adding much to my overall distance."

Bicyclists and others who are passionate about a topic like telling stories about experiences, successes, and problems. These bind the community together and also contain knowledge which, while subjective and situated, may still be beneficial for other users.

Most production-oriented online communities don't have an outlet for users to share stories or other subjective information. Stories may enhance the community and while new, should be designed in such a way that makes it technically easy for people to contribute. This could be within the interface itself (on Cyclopath, on the map) or as part of an integrated forum on the site.

On Cyclopath stories could take the form of users talking about routes they've received and ridden. On Wikipedia, one could imagine stories like "my first edit" or "the first time I was reverted" or "the first time I got in an edit war".

8.4 Cyclopath and Activity Theory

We found that Cyclopath core contributors aren't the most dedicated bicyclists, but they are engaged in open content systems and believe in open content knowledge. Our research also showed that community characteristics matter and help to shape the site and draw (or repel) certain users. Finally, we discovered that while we, the researchers, view and think of Cyclopath as a community, users think of it as a resource and not a destination for social interaction.

This demonstrates yet another reason for people building online communities to understand their users at a deeper level. By understanding primary motivations of core contributors, builders can create communities that better suit the needs of users and also attract core contributors. For example, if users are attracted more to open content than the specific content of that site, make sure that all marketing materials (including the community itself) promote the open content concept. Open content wouldn't have to be promoted to the exclusion of other material, but it should be front and center.

This work also argues for talking to users, especially core contributors, early and often. When Cyclopath was founded, research was done to try and understand the user base [112]. Those initial interviews were incredibly useful and helped form the site that exists today, but they were done before the site was created. Subsequent interviews weren't conducted until spring 2012, when the site had been live for four years. If we'd talked to users in between, we may have discovered some patterns

or learned more about core contributors that would have helped form the site or inform some of the potential design changes.

One other idea to highlight is that of sandboxing. Wikipedia and other wikis have pages called sandboxes [9] that allow users to experiment and try edits out before posting them to the actual websites. This is a concept regularly used in software testing as well to test things before pushing them live. Many of the interview participants we talked to mentioned being nervous about making edits or actually breaking things. For sites like Wikipedia and Cyclopath, providing users with a safe place to experiment may encourage more editing and editing that is more risky, but incredibly valuable for the site.

8.4.1 Generalization

This work also led us to think, in depth, about generalization. As researchers, we want to generalize results and give guidelines to practitioners, however our findings from Chapter 7 suggest several cautions. There are a number of factors that can affect online communities and ways communities can affect users. The specific details of a system and an online community matter. In particular, how the system fits into an ecology of related tools and communities will likely play a role in community dynamics. Another factor is how users view the tool, the community, and their role or interactions with both.

8.5 Conclusion

This work suggests many implications and ideas that could be used or adopted by many new or existing sites. This is not a comprehensive list, nor is it guaranteed that these ideas or principles will work. However these ideas do provide a number of options that may help community builders to create sites that engage users (both consumers and contributors) in a number of different ways and allow users to be creative and help to build a useful resource.

Chapter 9

Conclusion

The initial goals of this research, as laid out in the introduction, were to provide a holistic view of core contributors on Cyclopath, to provide quantitative insight into the role of core contributors on Wikipedia, and to present any differences between core contributors and non core contributors that were found.

To achieve these goals, we conducted quantitative analysis of log data for Wikipedia core contributors and average contributors. That was followed by quantitative analysis of Cyclopath core contributors, average contributors, and consumers, again based on log data. The quantitative analyses presented questions that could not be answered solely by log data, so the next step was surveying Cyclopath users (core contributors, average contributors, and consumers) to try to understand some of their underlying motivations to see if differences existed between them. Finally, we conducted interviews with Cyclopath core contributors to gain deeper insight into who the core contributors were, both on and offline.

Wikipedia Quantiative Analysis. The analysis of log data that we conducted on Wikipedia was sparked by the work of Bryant et al. [26] and aimed to understand the behavioral differences between Wikipedia core contributors and average contributors. We confirmed prior research that showed that the majority of work done in Wikipedia is done by a small group of core contributors. We also found that core contributors differ from average contributors starting on their first day, in regards to work quantity, work quality, and namespace diversity. Wikipedia core contributors are also relatively consistent and, if anything, do less over time, not more.

Cyclopath Quantitative Analysis. The analysis conducted on Cyclopath was able to dig deeper into some of the issues raised in Chapter 3, particularly the potential for users to be editing prior to registering (and, thus, not showing in the Wikipedia logs we had access to). On Cyclopath we found that there is some evidence of educational lurking, but more research is required. We did find that

Cyclopath core contributors act like Wikipedia core contributors in that the are different than average contributors from the beginning in terms of work quantity and time spent on the site. The Cyclopath data also allowed us to discover that there is a relationship between the locality of viewing and editing and that core contributors may reveal some information about their viewing behavior.

Cyclopath Motivations Survey. The Cyclopath survey work was designed to try to learn more about Cyclopath users than we could get from the logs. In particular, we were interested in users' motivations for registering and participating as well as what benefits they thought they or the community received from their participation. We found that consumers and contributors registered for the site for different reasons, but felt that they received similar benefits from their participation in the site. In contrast, contributors were more likely to think that the community had benefited from their participation. Regarding editing, contributors had different motivations for beginning to edit than continuing to edit.

Cyclopath Core Contributor Interviews. Our interviews with Cyclopath core contributors allowed us to more deeply understand who the core contributors were. We were surprised to find that the core contributors we spoke with weren't all dedicated cyclists, but were drawn to Cyclopath by the notion of open content. This work also led us to realize that the characteristics of a community matter. In particular, the state of the map when Cyclopath launched provided functionality with enough bugginess for users to fix problems. Finally we found that while we, the site designers and maintainers, think of Cyclopath as a community, the users are much less inclined to see it as a community and more likely to think of it as a resource or utility.

Challenges. This work has presented a number of challenges. On Wikipedia, we only had partial data, no registration data or consumer data. The site is also very large, making it slow to analyze. The user base is geographically distributed and we don't control it, making it difficult to conduct live experiments. In addition, researchers often find Wikipedia to be a hostile research environment.

On Cyclopath, contribution is seasonal, as cycling is seasonal. The user base is dramatically smaller than Wikipedia and is geographically bounded. However, as the site creators and maintainers, we have access to all the data.

The survey and interview work has a potential for self-reporting bias and potential flaws in memory. There is little or no longitudinal data from users on the survey and interview questions.

Finally, this work doesn't contain any research relating to live experiments. We had considered live experiments, but due to the seasonality of Cyclopath, the timing was suboptimal. Instead, we investigated different questions and relied on log data, survey data, and interviews.

Next Steps. There is continuing work being done by the Cyclopath community and some of it follows in the footsteps of this work. Due to my current position, conducting follow up work on Cyclopath is unlikely, however it may be possible for me to start understanding these issues in the context of Facebook. This would be an interesting site to conduct this research on, in part as the site is seen as a community and the "contributions" can be visible to a small group of users or the entire internet.

Appendix A

Cyclopath Survey (Spring 2010)

Cyclopath Usability Survey - Jan 2010

1. Introduction

Thank you for participating in this survey!

We are asking you to complete this survey to help us learn more about the Cyclopath community. Specifically, the results will:

1) ... help us learn more about you, our users, and your experiences within Cyclopath.

2) ... help us identify users to interview during follow-up studies.

3) ... inform our efforts to obtain more funding.

We estimate that completing the survey will take 20 minutes. The survey will be open until April 10, 2010. At that time we will randomly select seven of the participants to win gift certificates to The Hub Bike Co-op. There will be two winners of $50 certificates and five winners of $20 certificates.

Thanks,

Katie Panciera
Loren Terveen
and the rest of the Cyclopath team
info@cyclopath.org

2.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with Cyclopath or the University of Minnesota. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions:

This study is being conducted by Katherine Panciera and Loren Terveen, Department of Computer Science, University of Minnesota. If you have any doubts at any point during the survey, feel free to contact them at katpa@cs.umn.edu or terveen@cs.umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researchers, you are encouraged to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455, (612) 625-1650.

3. Risks and Benefits of Study

There are no known risks associated with this study. You can choose to provide personal information, but it will be stored securely and is not required to participate.

The benefits to participation are: you will aid in our understanding of Cyclopath and users of Cyclopath and you will help inform our efforts to obtain more funding which would benefit the Cyclopath community as a whole.

Cyclopath Usability Survey - Jan 2010

1. Do you understand the risks and benefits to the study and do you agree to participate?

- Yes
- No

4. Thanks

Thank you for being part of the Cyclopath community!

If you have any further questions, please e-mail info@cyclopath.org.

5. Contact Information

This page asks several questions about your Cyclopath data and your interest in future work. Answering these questions is optional.

2. What is your Cyclopath username? (Not required)

3. Do you want to be entered in the raffle for a gift certificate to The Hub Bike Co-op?

- Yes
- No

4. Would you be interested and willing in participating in a follow up interview?

- Yes
- No

5. If the answer to either of the above questions is Yes, please leave your email address. (We will only use it to contact you in the above circumstances. We will not share it with others.)

6. Riding Habits

This page asks some general questions about your cycling habits.

Cyclopath Usability Survey - Jan 2010

6. When does your riding season (usually) start?

- January
- February
- March
- April
- May
- June
- July
- August
- September
- October
- November
- December
- I ride year round

7. When does your riding season (usually) end?

- January
- February
- March
- April
- May
- June
- July
- August
- September
- October
- November
- December
- I ride year round

Cyclopath Usability Survey - Jan 2010

8. What affects when your riding season starts and finishes?

- Temperature
- Snow
- Road Conditions

Other (please specify)

9. During riding season, how often do you ride?

- Multiple times a day
- Once a day
- Multiple times a week
- Once a week
- Multiple times a month
- Once a month
- Multiple times a season
- Once a season

10. Why do you ride?

- Commute
- Leisure
- Exercise
- Profession
- Other (please specify)

11. How long is your typical ride?

- 0-5 miles
- 5-10 miles
- 10-15 miles
- 15-20 miles
- 20-25 miles
- over 25 miles

Cyclopath Usability Survey - Jan 2010

7. General Cyclopath Use

This page asks some general questions about how you use Cyclopath.

12. How often do you use Cyclopath during riding season?

- Daily
- Weekly
- Monthly
- Once or twice
- Never

13. What do you do on Cyclopath during the riding season?

- Browse the map
- Request routes
- Edit geography of blocks and points
- Edit attributes of blocks and points
- Write notes
- Apply tags
- Rate blocks
- Set up watch regions
- Other (please specify)

14. How often do you use Cyclopath during the off-season?

- Daily
- Weekly
- Monthly
- Once or twice
- Never
- Doesn't Apply (Ride Year Round)

Cyclopath Usability Survey - Jan 2010

15. What do you do on Cyclopath during the off season?

- Browse the map
- Request routes
- Edit geography of blocks and points
- Edit attributes of blocks and points
- Write notes
- Apply tags
- Rate blocks
- Set up watch regions
- Other (please specify)

8. Discovering Cyclopath

This page asks several questions about how you first discovered Cyclopath.

16. How did you find out about Cyclopath?

- Friend or Family
- Online link
- Newspaper article
- Cyclopath team member
- Flyer (enter location below)
- Other (please specify) or location of flyer

Cyclopath Usability Survey - Jan 2010

17. What did you first do when you came to Cyclopath?

- Read more about Cyclopath
- Explore the map
- Request a route
- Edit
- Register
- Other (please specify)

18. How long have you been using Cyclopath (approximately)?

Years
Months

19. Are you a member of Cyclopath? (Do you have a username?)

- Yes
- No

9. Membership on Cyclopath

20. Why did you register for Cyclopath?

21. How soon after discovering Cyclopath did you register? (Specify units in your answer)

22. Why did you register when you did?

23. What did you do on the site prior to registering?

Cyclopath Usability Survey - Jan 2010

24. When do you feel you have been most active on the Cyclopath website?

- When I first joined
- Recently
- My activity level has remained constant

10. Features of Cyclopath

25. What features do you regularly use on Cyclopath? (By use we mean either using the feature, creating the feature, or looking at the feature)

- Route finder
- Route search preferences
- Rating blocks
- Tagging
- Point filters
- Notes on streets and trails
- Points of interest (coffee shops, grocery stores, parks, etc)
- Watch Regions
- Other (please specify)

26. How did you learn how these features worked?

- Reading the text wiki
- Watching the tutorial videos on the text wiki
- Someone explained it to me through IM, Email, or chatroom
- Someone showed me in person
- I experimented
- Other (please specify)

Cyclopath Usability Survey - Jan 2010

27. How do you learn new skills in general (outside of Cyclopath)?

- I read about them first
- I watch someone else do them first
- I like to hear someone else tell me about them first
- I prefer to just learn by experimenting
- Other (please specify)

28. Have you ever made an edit on Cyclopath? (That is have you ever altered a point or a block, added or deleted tags or notes, or rated blocks?)

- Yes
- No

11. Editing Cyclopath

For this page, unless we specify otherwise, when we refer to edits we are including any changes that you make to the system and save including changes to points, blocks, tags, notes, and ratings.

29. How often do you edit Cyclopath?

- Daily
- Weekly
- Monthly
- Yearly

30. How many edits would you estimate you've made on Cyclopath?

31. Why did you start editing Cyclopath?

32. Tell us about an edit you made. When did you make it? Where was it? How did you find it? How did it improve Cyclopath?

Cyclopath Usability Survey - Jan 2010

33. Why do you contribute to Cyclopath?

34. Do you think you have an editing style?

- Yes
- No

35. If so, what is it?

36. How do you decide where to edit?

37. (Fill in the blanks)
On a typical day that I use Cyclopath, I request ____ routes, make ____ edits, and spend ___ minutes on Cyclopath.

(number of routes)
(number of edits)
(number of minutes)

12. Cyclopath over Time

38. How do you feel you have benefited from using and/or contributing to Cyclopath?

39. How do you feel the cycling community has benefited from your use of and contributions to Cyclopath?

40. How do you feel the wider (non-cycling) community has benefited by your use of and contributions to Cyclopath?

13. Cyclopath and Other Cycling Resources

Cyclopath Usability Survey - Jan 2010

41. Do you ever use printed bike maps? When do you prefer printed bike maps over Cyclopath and vice versa?

42. Have you used Google Maps Biking Directions? If so, please share your thoughts.

43. Which of the following websites have you used before?

- mplsbikelove.com
- bikely.com
- gmap-pedometer.com
- bikemap.net
- mocatrails.org
- None of the above

44. Are there other cycling websites that are local or national that you frequently use? What are they?

Cyclopath Usability Survey - Jan 2010

45. Which of the sites/maps is the best place (of the places listed) to solve the given task?

	Cyclopath	mplsbikelove.com	bikely.com	gmap-pedometer.com	bikemap.net	mocatrails.org	None
Find a route							
Find people to ride with							
Learn about a route (before riding)							
Learn about a route (after riding)							
Share information about a route							
Ask questions about a section of road							
Share information about a section of road							
Find out about road construction							
Learn about cycling events							
Share information about cycling events							

46. How does your activity on Cyclopath differ from your activity on mplsbikelove, bikely, gmap-pedometer, bikemap and mocatrails? (Do you edit as much on Cyclopath as you post on mplsbikelove? etc)

5
6

47. Which of these websites do you feel the most engaged in? Why?

5
6

14. Cyclopath and Other Online Content

48. Please indicate which of the following websites you use, are a member of and/or contribute content to:

	Use	Member of	Contribute Content to
Cyclopath			
Wikipedia			
YouTube			
Google MyMaps			
Facebook			
Twitter			

Cyclopath Usability Survey - Jan 2010

49. Rank the value of these websites to you.

	Least valuable					Most valuable
Cyclopath						
Wikipedia						
YouTube						
Google MyMaps						
Facebook						
Twitter						

50. Why do you consider a website the most valuable?

51. Which of the following websites do you consider the most trustworthy?

	Least trustworthy					Most trustworthy
Cyclopath						
Wikipedia						
YouTube						
Google MyMaps						
Facebook						
Twitter						

52. Why do you consider a website the most trustworthy?

15. Cyclopath Motivation

Cyclopath Usability Survey - Jan 2010

53. Using the scale below, please indicate how important or unimportant each of the following possible reasons for participating in Cyclopath is for you.

	1 Very unimportant	2 Unimportant	3 Somewhat unimportant	4 Neither important nor unimportant	5 Somewhat important	6 Important	7 Very important
I like to rate blocks.							
I want to view routes.							
I want to be part of the Cyclopath community.							
I like to express my opinions.							
I like to contribute to Cyclopath.							
I want to receive accurate routes.							
I want to contribute to the Cyclopath community.							
I like to share my opinions with other people.							
Contributing to Cyclopath is fun.							
I want to help other people receive accurate routes.							

16. Reasons for Volunteering

Cyclopath Usability Survey - Jan 2010

54. Please indicate how important or accurate each of the following possible reasons for volunteering is for you in doing volunteer work. If you have never volunteered, simply do your best to indicate how important or accurate you think each of the following possible reasons would be for you if you were to volunteer.

	Not at all important/accurate 1	2	3	4	5	6	Very important/accurate 7
Volunteering is a good escape from my own troubles.							
I feel compassion toward others in need.							
I can make new contacts that might help my business or career.							
Volunteering is an important activity to the people I know best.							
I can explore my own strengths.							
Volunteering helps me work through my own personal problems.							
Volunteering can help me to get my foot in the door at a place where I would like to work.							
I can learn how to deal with a variety of people.							
I can do something for a cause that is important to me.							
My friends volunteer.							
Volunteering is a way to make new friends.							
Volunteering experience will look good on my resume.							
Volunteering increases my self-esteem.							
Volunteering allows me to gain a new perspective on things.							
I feel it is important to help others.							
Others with whom I am close place a high							

value on community service.							
By volunteering I feel less lonely.							
To get to know people who are similar to myself.							
Because of my concern and worry about the community I am serving.							
To challenge myself and test my skills.							
To get to know people in the community I am serving.							
To gain experience dealing with emotionally difficult topics.							
To help members of the community I am serving.							

17. Sense of Community

Cyclopath Usability Survey - Jan 2010

55. Please respond to this series of statements that ask about your thoughts and feelings toward Cyclopath, using the scale provided.

	1 Strongly Disagree	2 Disagree	3 Somewhat Disagree	4 Neither Agree nor Disagree	5 Somewhat Agree	6 Agree	7 Strongly Agree
I view Cyclopath as a community.							
I identify with the Cyclopath community.							
I do not feel a sense of attachment and belonging to Cyclopath.							
I am similar to members of Cyclopath.							
It is not very important for me to feel connected to Cyclopath.							
I am invested in Cyclopath.							
I am not usually conscious of the fact that I am a member of Cyclopath.							
Being a member of Cyclopath has little to do with how I feel about myself.							
I see myself as being a part of the Cyclopath community.							
I have a lot of pride in Cyclopath.							

18. General Questions

We would appreciate you answering these questions, but they are optional. We emphasize again that all your answers are strictly confidential and they will not be available to external parties in any way whatsoever.

56. Internet and Computer Usage

How many hours a day do you actively use the internet?

How many hours a day do you actively use your computer for activities other than the internet?

Cyclopath Usability Survey - Jan 2010

57. What is your gender?

- Female
- Male

58. What is your Race/Ethnicity/Cultural Background?

- White
- Hispanic
- Black/African-American
- Asian or Pacific Islander
- American Indian/Native American
- Other (please specify)

59. What is your age?

60. What is your profession?

19. Closing

61. Is there anything else you want to tell us?

20. Thanks

Thank you for completing the survey! This information is extremely helpful for us.

If you have any further questions, please e-mail info@cyclopath.org.

Appendix B

Cyclopath Interview Protocol

This is an illustrative set of questions. We did ask questions that are not on this list as follow-up questions in our interview as our interviews were semi-structured so that we could glean the maximum amount of information possible.

Cycling

When during the day (or week) do you ride?

Where do you ride (to work, to the store, etc)?

How far do you ride?

Why do you ride?

Do you ride alone or with others? Who do you ride with (small children on bike carriers, school age children on their own bikes, one other adult, a group of adults, etc)?

How long have you been riding in the Twin Cities?

How long have you been in the Twin Cities?

How well do you feel you know the local trails and roads?

Are you attached to the cycling community in the Twin Cities? How? Why?

What benefits do you gain by being attached to to cycling community?

Have you been active in other cycling communities in the US or abroad? Which ones? Were they similar to the Twin Cities or different? How did that impact your cycling?

Cyclopath

How did you discover Cyclopath?

Why do you use Cyclopath?

Please log into Cyclopath and show us what you do on a regular visit. Let's see you add some information to the map.

If you can't add information, let's talk about that. What are the barriers to you? Have you added information before?

Can you make a different type of edit to the map now? Do you prefer one type of edit to another?

What are some challenges you've faced in Cyclopath or while editing Cyclopath?

What would make you become more active on Cyclopath?

How much time and effort do you think you put into Cyclopath?

How much benefit do you think you get out of Cyclopath?

Do you think you give more than you get or vice versa or is it balanced? Why?

Becoming Wikipedian questions

These questions were modified from questions asked in the interviews that were reported on in [26]. Questions were obtained from Andrea Forte on January 27, 2012 and used with her permission.While Andrea was not the first author on [26], she was the PI on the project.

What is your username on Cyclopath?

How did you first hear about Cyclopath?

How would you describe Cyclopath to someone who had never heard of it?

How long have you been contributing to Cyclopath?

Can you tell me the story of when you first came upon Cyclopath and how you became involved with it?

Can you talk a little bit more about the first time you went in and contributed?

Has your process changed since then?

How would you describe your role on Cyclopath?

Has it changed over time or are you still playing the same type of role?

Do you know a lot of the other users by username?

Can you talk about why you contribute to Cyclopath? What draws you to it?

How do you decide what you're going to edit?

Are you involved in any other online communities?

How would you compare Cyclopath community to those communities?

Is there anything else that seems important about Cyclopath that we didn't talk about?)

Appendix C

References

This thesis includes much or most of [107, 109, 108]. Chapter 7 is a working version of a paper written with Mikhil Masli and Loren Terveen, accepted to OpenSym 2014.

[1] Alexa Top 500 Global Sites. http://www.alexa.com/topsites.

[2] Frequently asked questions - Wikimedia Foundation. http://wikimediafoundation.org/wiki/Frequently_asked_questions.

[3] lurker, n.1. http://oed.com/view/Entry/111323?rskey=d2lKd5&result=1&isAdvanced=false.

[4] Mission statement - wikimedia foundation. http://wikimediafoundation.org/wiki/Mission_statement.

[5] UGC. http://www.webopedia.com/TERM/U/UGC.html.

[6] Wikimedia project at a glance. http://stats.wikimedia.org/EN/SummaryEN.htm.

[7] Wikipedia:activity theory - wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Activity_theory.

[8] Wikipedia.org Site Info. http://www.alexa.com/siteinfo/wikipedia.org.

[9] Wikipedia:Sandbox - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Wikipedia%3ASandbox.

[10] Wikipedia:Teahouse - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Wikipedia:Teahouse.

[11] Wikipedia:Wikipedians - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Wikipedia:Wikipedians.

[12] Wikipedia:database reports/wikiprojects by changes. http://en.wikipedia.org/wiki/Wikipedia:Database_reports/WikiProjects_by_changes, October 2013.

[13] Wikipedia:wikiproject minnesota. http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Minnesota, October 2013.

[14] Lada A. Adamic, J. Zhang, E. Bakshy, and Mark S. Ackerman. Knowledge sharing and Yahoo Answers: Everyone knows something. In *Proc. WWW*, 2008.

[15] B.T. Adler and L. de Alfaro. A Content-Driven Reputation System for the Wikipedia. In *Proc. WWW*, 2007.

[16] Rodrigo B. Almeida, Barzan Mozafari, and Junghoo Cho. On the evolution of Wikipedia. In *Proc. ICWSM*, 2007.

[17] Chris Anderson. *The Long Tail: Why the Future of Business is Selling Less of More*. Hyperion, 2006.

[18] Judd Antin. My kind of people?: perceptions about wikipedia contributors and their motivations. In *Proc. CHI*, 2011.

[19] Judd Antin and Coye Cheshire. Readers are not free-riders: reading as a form of participation on wikipedia. In *Proc. CSCW*, 2010.

[20] Judd Antin, Coye Cheshire, and Oded Nov. Gender Differences in Wikipedia Editing. In *Proc. Wikisym*, 2011.

[21] Judd Antin, Coye Cheshire, and Oded Nov. Technology-mediated contributions: editing behaviors among new wikipedians. In *Proc. CSCW*, 2012.

[22] Gerard Beenen, Kimberly Ling, X Wang, K. Chang, Dan Frankowski, Paul Resnick, and Robert E Kraut. Using social psychology to motivate contributions to online communities. In *Proc. CSCW*, 2004.

[23] Yochai Benkler. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, 2006.

[24] Anna Bergström and R Magnusson. Potential of transferring car trips to bicycle during winter. *Transportation Research Part A: Policy and practice*, 37(8), 2003.

[25] Christina Bonnington. So Long, Ping: Apple Shuttering Failed Social Network Sept. 30. http://www.wired.com/gadgetlab/2012/09/goodbye-ping/, Sep 13 2012.

[26] Susan Bryant, Andrea Forte, and Amy Bruckman. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proc. GROUP*, 2005.

[27] Moira Burke and Robert E Kraut. Mopping up: modeling wikipedia promotion decisions. In *Proc. CSCW*, 2008.

[28] Moira Burke and Robert E Kraut. Taking up the mop: identifying future wikipedia administrators. *Proc. CHI Extended Abstracts*, 2008.

[29] Brian Butler, Elisabeth Joyce, and Jacqueline Pike. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In *Proc. CHI*, 2008.

[30] Jilin Chen, Yuqing Ren, and John Riedl. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proc. CHI*, 2010.

[31] Coye Cheshire. Selective Incentives and Generalized Information Exchange. *Social Psychology Quarterly*, 70(1), 2007.

[32] R.B. Cialdini and V. Griskevicius. Social influence. *Advanced Social Psychology: The State of the Science*, 2010.

[33] Bike Walk Twin Cities. Bike/Walk Pilot Profile: Minneapolis, Minnesota. http://www.railstotrails.org/resources/documents/whatwedo/policy/Pilot%20Projects_Minneapolis_final_ltr_lr.pdf.

[34] Gil Clary, Mark Snyder, R.D. Ridge, J. Copeland, A.A. Stukas, J. Haugen, and P. Miene. Understanding and assessing the motivations of volunteers: A functional approach. *Journal of Personality and Social Psychology*, 74(6), 1998.

[35] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proc. CHI*, 2006.

[36] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. SuggestBot: using intelligent task routing to help people find work in wikipedia. In *Proc. IUI*, 2007.

[37] Matt Cowan. Fail study: Jimmy Wales and Nupedia (Wired UK). May 2011.

[38] Kevin Crowston, Kangning Wei, Qing Li, U. Yeliz Eseryel, and James Howison. Self-organization of teams in free/libre open source software development. *nformation and Software Technology Journal*, 49(6), 2007.

[39] Jonathon N. Cummings, Brian Butler, and Robert E. Kraut. The quality of online social relationships. *CACM*, 45(7), 2002.

[40] Laura Dabbish, Rosta Farzan, Robert E Kraut, and Tom Postmes. Fresh Faces in the Crowd: Turnover, Identity, and Commitment in Online Groups. In *Proc. CSCW*, 2012.

[41] Nicolas Ducheneaut. Socialization in an Open Source Software Community: A Socio-Technical Analysis. *Computer Supported Cooperative Work*, 14(4), 2005.

[42] Nicolas Ducheneaut, N. Yee, E. Nickell, and R.J. Moore. Alone together?: exploring the social dynamics of massively multiplayer online games. In *Proc. CHI*, 2006.

[43] Casey Dugan, Werner Geyer, Michael Muller, Joan DiMicco, Beth Brownholtz, and David R. Millen. It's all 'about you': Diversity in online profiles. In *Proc. CSCW*, 2008.

[44] Peter P. Ekeh. *Social Exchange Theory: The Two Traditions*. Heinemann Educational, 1974.

[45] Rosta Farzan, Robert E. Kraut, Aditya Pal, and Joseph Konstan. Socializing volunteers in an online community: a field experiment. In *Proc. CSCW*, 2012.

[46] Danyel Fisher, Marc Smith, and Howard T. Welser. You Are Who You Talk To: Detecting Roles in Usenet Newsgroups. In *Proc. HICSS*, 2006.

[47] Andrea Forte. Why do people write for wikipedia? Incentives to contribute to open-content publishing. *Proc. GROUP Workshops*, 2005.

[48] Andrea Forte and Amy Bruckman. Scaling Consensus: Increasing Decentralization in Wikipedia Governance. In *Proc. HICSS*, 2008.

[49] 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 *Proc. CSCW*, 2012.

[50] Dan Frankowski, Shilad Sen, Loren Terveen, and John Riedl. You are what you say: privacy risks of public mentions. In *Proc. SIGIR*, 2006.

[51] E.J. Friedman and Paul Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(2), 2001.

[52] 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 *Proc. CSCW*, 2012.

[53] Rishab A Ghosh, Ruediger Glott, Bernhard Krieger, and Gregorio Robles. Free/Libre and Open Source Software: Survey and Study. http://www.math.unipd.it/~bellio/FLOSS%20Final%20Report%20-%20Part%204%20-%20Survey%20of%20Developers.pdf.

[54] Eric Gilbert and Karrie Karahalios. Predicting tie strength with social media. In *Proc. CHI*, 2009.

[55] Jim Giles. Internet Encyclopedias Go Head to Head. *Nature*, 438, 2005.

[56] Scott. Golder and Bernardo A Huberman. The structure of collaborative tagging systems. *Arxiv Preprint*, 2005.

[57] Adam Grant and Jane Dutton. Beneficiary or Benefactor: Are People More Prosocial When They Reflect on Receiving or Giving? *Psychological Science*, 23(9), September 2012.

[58] Aaron Halfaker, Stuart Geiger, Jonathan Morgan, and John Riedl. The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to sudden popularity is causing its decline. *American Behavioral Scientist*, in press.

[59] Aaron Halfaker, Aniket Kittur, and John Riedl. Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work. In *Proc. Wikisym*, 2011.

[60] F. Maxwell Harper, Dan Frankowski, Sara Drenner, Yuqing Ren, Sara Kiesler, Loren Terveen, Robert E Kraut, and John Riedl. Talk amongst yourselves: inviting users to participate in online conversations. In *Proc. IUI*, 2007.

[61] Brandon Harris. Quora: How many registered users does Wikipedia have? http://www.quora.com/Wikipedia/How-many-registered-users-does-Wikipedia-have.

[62] A. Hars and Shaosong Ou. Working for Free? - Motivations of Participating in Open Source Projects. In *Proc. HICSS*, 2001.

[63] Tad Hogg, Dennis M. Wilkinson, Gabor Szabo, and Michael J. Brzozowski. Multiple relationship types in online communities and social networks. In *Proc. AAAI SIP*, 2008.

[64] S. Intille and X. Ma. Eliciting user preferences using image-based experience sampling and reflection. In *Proc. CHI*, 2002.

[65] Shamsi T Iqbal and Eric Horvitz. Notifications and Awareness: A Field Study of Alert Usage and Preferences. In *Proc. CSCW*, 2009.

[66] Quentin Jones, Gilad Ravid, and Sheizaf Rafaeli. Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration. *Info. Sys. Research*, 15(2), 2004.

[67] Steven J. Karau and Kipling D. Williams. Social Loafing: A Meta-Analytic Review and Theoretical Integration. *Personality and Social Psychology*, 65(4), 1993.

[68] J Katz. Luring the Lurkers - Slashdot. http://news.slashdot.org/story/98/12/28/1745252/luring-the-lurkers, December 1998.

[69] Brian Keegan, Darren Gergle, and Noshir Contractor. Hot off the Wiki: Dynamics, Practices, and Structures in Wikipedia's Coverage of the Tōhoku Catastrophes. In *Proc. Wikisym*, 2011.

[70] Brian Kelly. The Failure of Citizendium. http://ukwebfocus.wordpress.com/2011/12/20/the-failure-of-citizendium/, Dec 20 2011.

[71] Aniket Kittur and Robert E Kraut. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proc. CSCW*, 2008.

[72] Aniket Kittur and Bongwon Suh. He says, she says: conflict and coordination in Wikipedia. In *Proc. CHI*, 2007.

[73] Aniket Kittur and Bongwon Suh. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proc. WWW 2007*, 2007.

[74] Peter Kollock and Marc Smith. Managing the Virtual Commons: Kollock and Smith. In *Proc. CMC*, 1996.

[75] Kooz. Kooz Top 5: Top 5 Failed Social Media Sites. http://kooztop5.blogspot.com/2012/08/top-5-failed-social-media-sites_20.html, Aug 20 2012.

[76] Adam D. I. Kramer. An unobtrusive Behavioral Model of "Gross National Happiness". In *Proc. CHI*, 2010.

[77] Michel Krieger, Emily Margarete Stark, and Scott R Klemmer. Coordinating Tasks on the Commons: Designing for Personal Goals, Expertise and Serendipity. In *Proc. CHI*, 2009.

[78] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. Structure and evolution of online social networks. In *Proc. KDD*, 2006.

[79] K Kuutti. Activity theory as a potential framework for human-computer interaction research. In B Nardi, editor, *Context and consciousness: Activity theory and human-computer interaction*, pages 17–44. MIT Press, 1996.

[80] Stacey Kuznetsov. Motivations of contributors to Wikipedia. *SIGCAS*, 36(2), 2006.

[81] Karim R. Lakhani and Robert G. Wolf. Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In Joe Feller, Brian Fitzgerald, Scott Hissam, and Karim R. Lakhani, editors, *Perspectives on Free and Open Source Software*. MIT Press, 2005.

[82] Shyong Tony K Lam, A. Uduwage, Z. Dong, D. Musicant, Shilad Sen, Loren Terveen, and John Riedl. WP:Clubhouse? An Exploration of Wikipedia's Gender Imbalance. In *Proc. Wikisym*, 2011.

[83] Cliff Lampe and Paul Resnick. Slash(dot) and burn: Distributed moderation in a large online conversation space. In *Proc. CHI*, 2004.

[84] Cliff Lampe, Alcides Velasquez, and Elif Ozkaya. Motivations to participate in online communities. In *Proc. CHI*, 2010.

[85] Jean Lave and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge University Press, 1991.

[86] K. Ling, G. Beenen, P. Ludford, X. Wang, K. Chang, X. Li, D. Cosley, D. Frankowski, L. Terveen, A. M. Rashid, P. Resnick, , and R. Kraut. Using social psychology to motivate contributions to online communities. journal of computer-mediated communication. 10(4), 2005.

[87] K. Luther, K. Caine, K. Ziegler, and Amy Bruckman. Why it works (When it works): Success Factors in Online Creative Collaboration. In *Proc. GROUP*, 2010.

[88] K Marks. Power Laws and Blogs. http://homepage.mac.com/kevinmarks/powerlaws.html, 2003.

[89] Mikhil Masli and Loren Terveen. Task Specialization in Social Production Communities: The Case of Geographic Volunteer Work. In *Proc. ICWSM*, 2011.

[90] Mikhil Masli and Loren Terveen. Leveraging the Contributory Potential of User Feedback. In *Proc. CSCW*, 2014.

[91] Wyl McCully, Cliff Lampe, Chandan Sarkar, Alcides Velasquez, and Akshaya Sreevinsasan. Online and Offline Interactions in Online Communities. In *Proc. WikiSym*, 2011.

[92] Richard Millington. Important Lessons From A Failed Community. http://www.feverbee.com/2011/01/important-lessons-from-a-failed-community.html, Jan 14 2011.

[93] Jonathan Morgan. *WikiProjects as Complex Systems*. PhD thesis, University of Washington.

[94] Jonathan T. Morgan, Siko Bouterse, Heather Walls, and Sarah Steirch. Tea & Sympathy: Crafting Positibe New User Experiences on Wikipedia. In *Proc. CSCW*, 2013.

[95] Michael Muller. Lurking as personal trait or situational disposition: lurking and contributing in enterprise social media. In *Proc. CSCW*, 2012.

[96] Michael Muller, N Sadat Shami, David R Millen, and Jonathan Feinberg. We are all lurkers: consuming behaviors among authors and readers in an enterprise file-sharing service. In *Proc. GROUP*, 2010.

[97] David R Musicant, Yuqing Ren, James A Johnson, and John Riedl. Mentoring in Wikipedia: a clash of cultures. In *Proc. WikiSym*, 2011.

[98] Kumiyo Nakakoji, Yasuhiro Yamamoto, Yoshiyuki Nishinaka, Kouichi Kishida, and Yunwen Ye. Evolution Patterns of Open-Source Software Systems and Communities. In *Proc. IWPSE*, 2002.

[99] NJDOT. Introduction to Bicycle Facilities - State of New Jersey. http://www.state.nj.us/transportation/publicat/pdf/BikeComp/introtofac.pdf.

[100] Blair Nonnecke and Jennifer Preece. Lurker Demographics: Counting the Silent. In *Proc. CHI*, 2000.

[101] Blair Nonnecke, Jennifer Preece, and D. Andrews. What lurkers and posters think of each other. In *Proc. HICSS*, 2004.

[102] Oded Nov. What motivates Wikipedians? *CACM*, 50(11), 2007.

[103] Morris Okun and Josef Michel. Sense of Community and Being a Volunteer Among the Young-OId. *Journal of Applied Gerontology*, 25(2):173–188, April 2006.

[104] A. M. Omoto and Mark Snyder. Sustained Helping Without Obligation: Motivation, Longevity of Service, and Perceived Attitude Change Among AIDS Volunteers. 68, 1995.

[105] S Oreg and Oded Nov. Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior*, 24(5), 2008.

[106] Felipe Ortega and Jesus M Gonzalez Barahona. Quantitative analysis of the Wikipedia community of users. In *Proc. Wikisym*, 2007.

[107] Katherine A Panciera, Aaron Halfaker, and Loren Terveen. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proc. GROUP*, 2009.

[108] Katherine A Panciera, Mikhil Masli, and Loren Terveen. "How should I go from ___ to ___ without getting killed?": motivation and benefits in open collaboration. In *Proc. Wikisym*, 2011.

[109] Katherine A Panciera, Reid Priedhorsky, Thomas Erickson, and Loren Terveen. Lurking? cyclopaths?: a quantitative lifecycle analysis of user behavior in a geowiki. In *Proc. CHI*, 2010.

[110] Jennifer Preece and Blair Nonnecke. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, 20(2), 2004.

[111] Jennifer Preece and Ben Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 2009.

[112] Reid Priedhorsky, Benjamin Jordan, and Loren Terveen. How a personalized geowiki can help bicyclists share information more effectively. In *Proc. Wikisym*, 2007.

[113] Reid Priedhorsky, Shyong Tony K Lam, Katherine A Panciera, Loren Terveen, and John Riedl. Creating, destroying, and restoring value in wikipedia. In *Proc. GROUP*, 2007.

[114] Reid Priedhorsky, Mikhil Masli, and Loren Terveen. Eliciting and focusing geographic volunteer work. In *Proc. CSCW*, 2010.

[115] Reid Priedhorsky, David Pitchford, Shilad Sen, and Loren Terveen. Recommending routes in the context of bicycling: algorithms, evaluation, and the value of personalization. In *Proc. CSCW*, 2012.

[116] Reid Priedhorsky and Loren Terveen. The computational geowiki: What, why, and how. In *Proc. CSCW*, 2008.

[117] Al M. Rashid, Kimberly Ling, Regina D. Tassone, Paul Resnick, Robert E Kraut, and John Riedl. Motivating participation by displaying the value of contribution. In *Proc. CHI*, 2006.

[118] Sansank Reddy, Katie Shilton, Gleb Denisov, Christian Cenizal, Deborah Estrin, and Mani Srivastava. Biketastic: Sensing and Mapping for Better Biking. In *Proc. CHI*, 2010.

[119] Yuqing Ren, F Maxwell Haper, Sara Drenner, Loren Terveen, Sara Kiesler, John Riedl, and Robert E Kraut. Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds. *Management Information Systems Quarterly*, 36(3), 2012.

[120] R Rosenthal and R L Rosnow. *Essentials of behavioral research: Methods and data analysis*. McGraw-Hill, 1991.

[121] S Andrew Sheppard and Loren Terveen. Quality is a Verb: The Operationalization of Data Quality in a Citizen Science Community. In *Proc. Wikisym*, 2011.

[122] Vladimir Soroka, Michal Jacovi, and Sigalil Ur. We can see you: A study of Communities' invisible People through ReachOut. In *Proc. Communities and Technologies*, 2003.

[123] Bongwon Suh, Gregorio Convertino, Ed H. Chi, and Peter L. Pirolli. The Singularity is Not Near: Slowing Growth of Wikipedia. In *Proc. Wikisym*, 2009.

[124] Emmi Suhonen, Airi Lampinen, Coye Cheshire, and Judd Antin. Everyday favors: a case study of a local online gift exchange system. In *Proc. GROUP*, 2010.

[125] Aaron Swartz. Who Writes Wikipedia? http://www.aaronsw.com/weblog/whowriteswikipedia, 2006.

[126] Latanya Sweeney. K-anonymity: A model for protecting privacy. *Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 2002.

[127] Masamichi Takahashi, Masakazu Fujimoto, and Nobuhiro Yamasaki. The active lurker: influence of an in-house online community on its outside environment. In *Proc. GROUP*, 2003.

[128] Fernando Torre, S Andrew Sheppard, and Loren Terveen. bumpy, caution with merging: an exploration of tagging in a geowiki. In *Proc. GROUP*, 2010.

[129] Alcides Velasquez, Rick Wash, Cliff Lampe, and Tor Bjornrud. Latent users in an online user-generated content community. *Computer Supported Cooperative Work (CSCW)*, 2013.

[130] Fernanda B. Viégas, Martin Wattenberg, Jesse Kriss, and Frank van Ham. Talk Before You Type: Coordination in Wikipedia. In *Proc. HICSS*, 2007.

[131] Jakob Voß. Measuring Wikipedia. In *Proc. ISSI*, 2005.

[132] Morten Warncke-Wang, Dan Cosley, and John Riedl. Tell Me More: An Actionable Quality Model for Wikipedia. In *Proc. WikiSym*, 2013.

[133] Steve Whittaker, Loren Terveen, W. Hill, and L. Cherny. The dynamics of mass interaction. In *Proc. CSCW*, 1998.

[134] A. Wiggins and Kevin Crowston. Reclassifying Success and Tragedy in FLOSS Projects. In *Proc. OSS*, 2010.

[135] Dennis M Wilkinson and Bernardo A Huberman. Assessing the Value of Coooperation in Wikipedia. *First Monday*, 12(4), 2007.

[136] Dell Zhang, Karl Prior, and Mark Levene. How Long Do Wikipedia Editors Keep Active? In *Proc. WikiSym*, 2012.