Lurking? Cyclopaths? A Quantitative Lifecycle Analysis of User Behavior in a Geowiki

Katherine Panciera[1], **Reid Priedhorsky**[1], **Thomas Erickson**[2], **Loren Terveen**[1]

[1]GroupLens Research
Department of Computer Science and Engineering
University of Minnesota
Minneapolis, Minnesota, USA
{katpa,reid,terveen}@cs.umn.edu

[2]IBM T.J. Watson Research Center
PO Box 704
Yorktown Heights, NY 10598 USA
{snowfall}@acm.org

ABSTRACT

Online communities produce rich behavioral datasets, e.g., Usenet news conversations, Wikipedia edits, and Facebook friend networks. Analysis of such datasets yields important insights (like the "long tail" of user participation) and suggests novel design interventions (like targeting users with personalized opportunities and work requests). However, certain key user data typically are unavailable, specifically viewing, pre-registration, and non-logged-in activity. The absence of data makes some questions hard to answer; access to it can strengthen, extend, or cast doubt on previous results. We report on analysis of user behavior in Cyclopath, a geographic wiki and route-finder for bicyclists. With access to viewing and non-logged-in activity data, we were able to: (a) replicate and extend prior work on user lifecycles in Wikipedia, (b) bring to light some pre-registration activity, thus testing for the presence of "educational lurking," and (c) demonstrate the locality of geographic activity and how editing and viewing are geographically correlated.

Author Keywords

Wiki, geowiki, open content, geographic volunteer work, volunteered geographic information, lurking

General Terms

Human Factors, Design

ACM Classification Keywords

H.5.3 [*Group and Organization Interfaces*]: Collaborative computing, computer-supported cooperative work, web-based interaction.

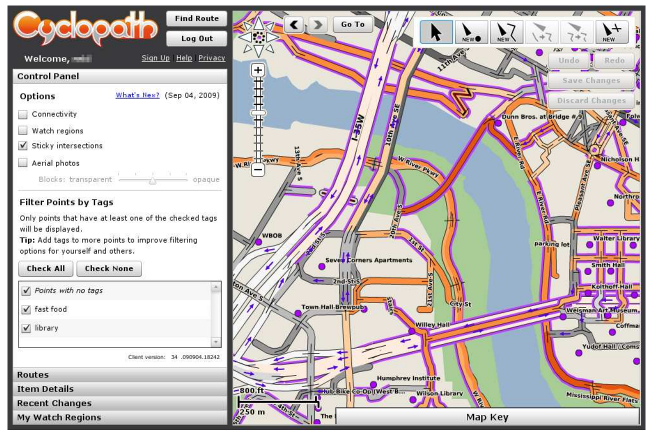


Figure 1. Screenshot of the Cyclopath application.

INTRODUCTION

The recent explosion of *open content* systems like Wikipedia, Flickr, and Stack Overflow has led to a new industry of online knowledge production and organization, carried out by distributed volunteers. The value of this new way of work is clear, and we, as many other researchers do, seek to understand how these systems work and how they can be nurtured.

This paper is concerned with one such open content system. Cyclopath (shown in Figure 1) is a web-based mapping application for bicyclists in the Minnesota cities of Minneapolis and St. Paul, a metro area of roughly 8,000km^2 and 2.3 million people. It serves as a standard web map, offering cyclist-centric route-finding, but as a *geowiki*, it goes beyond, offering wiki editing of the entire map, from the geometry and topology of the road/trail network to points of interest and annotations like notes and tags.

Cyclopath is of interest to researchers for three reasons. First, as an open content system it furthers our understanding of this class of system by allowing us to compare and contrast it with other systems like Wikipedia (which is well-studied). Second, its design enables the study of phenomena – for example, viewing behavior by "lurkers" – which cannot be studied with other open content systems. Third, Cyclopath is an example of a relatively new class of online system, geo-communities, that offers a chance to explore how users consume, explore, and edit geographic data.

Previous work has described Cyclopath's design and rationale [23], demonstrated its effectiveness under laboratory conditions [21], and discovered mechanisms for encouraging work [24]. These enabling efforts have established Cyclopath as a viable research platform. This is the first work to report on Cyclopath's use "in the wild"; as of this writing, Cyclopath has been operational for 16 months and has sufficient users and data to provide a portrait of its use and some lessons learned.

We offer two key contributions. First, we have quantitatively analyzed the lifecycles of users in an open-content system, addressing in particular the question of pre-registration anonymous lurking. Second, and again quantitatively, we analyze the kind of geographic work that is being done, focusing on how the geographic nature of the system affects what work is done and how public and hidden actions relate.

The remainder of this paper is organized as follows. We first present related work, followed by data sources and methods. We then present the bulk of our two contributions: first an analysis of the lifecycle of Cyclopath users, and then of geographic viewing and editing. We close with some lessons learned and conjectures for other systems.

RELATED WORK

Geo-community networks. Online communities have a long history in CHI and CSCW. One strand of that history concerns community networks – online communities focused on a particular city or region, such as Berkeley's Community Memory Project in the 1970s or Montana's Big Sky Telegraph in the 1980s [25]. This interest has continued in sites like Blacksburg Electronic Village [4] and is today manifested in Web 2.0-style sites such as EveryBlock.com and FourSquare.com that provide services for and tap into the local knowledge of the inhabitants of a particular place.

One interesting trend is that geographic information has begun to be incorporated into community network-style sites. This draws on geographic information systems (GIS), the field concerned with visualizing, analyzing, and manipulating geographic and spatial data [17]. While GIS has traditionally been the province of experts using specialized (and expensive) software, more recently, the field has expanded to consider what it refers to as *volunteered geographic information* [11, 6].

Within the CHI and CSCW communities, there has also been growing interest in geographic-based collaboration and online communities, such as the systems for supporting map-based collaborative planning developed and studied at Penn State [5]. Similarly, in the Web 2.0 world, we see open mapping sites like OpenStreetMap.org, which aims to build a global street map from scratch, and community-oriented sites such as FixMyStreet.com and SeeClickFix.com that enable locals to plot the location of potholes and similar problems on a map. Major online mapping vendors also participate in this space; for example, Google My Maps lets users create and share arbitrary maps with low process overhead.

Analyzing traces of online behavior. Internet-based social media has been much studied because it creates rich activity traces – messages posted and replied to, user profile settings, Wikipedia policies debated and articles edited, etc. Researchers have analyzed data from many systems to draw compelling pictures of life online. For example, Whittaker et al. [31] and Jones et al. [14] revealed the dynamics of large scale discussions; Lampe and Resnick investigated the effectiveness of distributed moderation for information filtering [16].

Types and lifecycles of users. Of most relevance to this paper is work examining the types and lifecycles of online community members. One prominent line of research seeks to assign users to different interaction roles, often through applying social network analysis [1, 7, 30].

Closer to the current work, other research analyzes users by their level of participation. Famously, when the level of participation is visualized, it typically exhibits a "long tail"[2], such as by following a power law [1]. Such relationships have been observed in many different kinds of online communities, including Usenet discussions [31], tagging [10], and Wikipedia edits [22, 15]. Users near the top of the curve often are called "power users" or "elite editors." In Wikipedia, a small group of elite editors does the majority of the work [15] and produces the majority of value [22].

Several researchers have studied how user behavior in online communities changes over time, especially among the power users. Bryant et al. [3] interviewed a number of experienced Wikipedia editors and identified a broadening of interests and gradual assumption of community maintenance activities. Our earlier work [20] extended this work with quantitative analyses investigating the lifecycles of elite and non-elite Wikipedia editors. Specifically, we found that elite editors edited more than non-elites immediately upon appearing, and that all editors' activity was characterized by an initial burst of intense activity followed by gradual decline to a fairly low, constant level of activity.

Activity spectrum: identified, anonymous, hidden. A key difficulty of online community research is that it is limited by what researchers can see, which is not everything. Acts of "participation" or "editing" are traditionally visible, while acts of "viewing" or "consumption" are not. Thus, in discussion groups, posted messages, replies, and threads are visible, but reading of messages is hidden; similarly, in Wikipedia, edits to articles (and talk pages, user pages, etc.) are visible, article reading is hidden, etc. The vast majority of research, e.g. [31], analyzes only visible data, while pointing out the problem of "hidden" data. Those who do not participate but only view are called *lurkers* [19].

A few research efforts have been able to study lurkers. Nonnecke and Preece [19] quantified the amount of lurking in discussion lists. Soroka et al. [27] studied the amount of lurking within a corporate discussion system and whether an initial lurking period served to educate users about system conventions. Harper et al. [12] analyzed how participation invitations influenced users' message reading behavior in a movie discussion system. However, these few efforts are the exception to the general analytical invisibility of viewing data.

Another type of data remains not exactly unseen, but usually unanalyzed: anonymous participation. A meaningful fraction of Wikipedia revisions is made by users who either have not registered, or, if they have registered, chose to edit or post while not logged in. Such edits are associated with the IP address of the machine from which the edit was made. While these edits are visible, they usually are ignored by analysts because (a) a single IP could represent multiple users, (b) a single user might use multiple IPs, and (c) IPs may or may not correspond to known users. There is a large stream of research that investigates computational techniques to "de-anonymize" supposedly anonymous data and defensive techniques to preserve anonymity [29, 8].

This Research. We go beyond prior work in several ways. First, we are able to analyze a greater fraction of user activity (in Cyclopath) than prior research typically has. We are able to consider all "hidden" (viewing) activity and match a meaningful fraction of anonymous work to known (registered) users. Second, this lets us analyze lifecycle issues that have been impossible for previous work to consider, such as: How do users' pre- and post-registration activities compare? How do patterns of viewing and editing activity compare? Finally, we reveal aspects of user activity in a geographic context, an emerging and very little studied area.

As any system would, studying Cyclopath leads to some limitations. For example, the userbase is measured in hundreds, not millions, and Cyclopath is young, only 16 months old, so the set of basic features is still growing, which leads to startup effects which may not be typical. Lastly, the map editing tools and concepts are complex, more so than text wiki tools, so use patterns may be atypical and the tools themselves are in flux due to usability improvement efforts. Nonetheless, Cyclopath is a much-used open-content system with obvious parallels to other systems, which leads us to believe that our results may be generally applicable.

METHODS

Data Sources

This work is based on two basic streams of data logged while cyclists use Cyclopath: the wiki work they do, and the requests the browser application makes to our web server.

Wiki work. As in all wikis, Cyclopath keeps a full history of the changes made to each object in the system. In Cyclopath, the unit of change is the geographic object or an item attached to it: e.g., Washington Avenue between 17th and 18th Streets is one object and a note attached to that block is another object.

We analyze this editing history at two levels. An **edit action** is changing one type of information on one object: for example, creating a new block, changing the geometry of a block, or changing the name and speed limit of a block each constitute one edit action; one edit action corresponds very loosely to changing one word in Wikipedia. For each edit action, our database records:

- 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.

A **revision** is the group of edit actions saved atomically when the *Save Changes* button is clicked.

Cyclopath application HTTP stream. Once a user has started the Cyclopath client application in his or her browser, there is a variety of HTTP "chatter" which takes place between the client and server. This includes both requests for data (tags, lists of revisions, geographic objects, etc.) and logs of application usage (which objects are clicked on, etc.).

For the purposes of this work, we treat most of this interaction as simply an activity indicator: because the client does not talk to the server except in response to user interactions, we can treat the presence of the chatter as an indicator of application use as opposed to idleness. Thus, we call each such HTTP request a **use event**. Specifically, we assume that each use event indicates Cyclopath was in use from the event until 30 seconds after the event and then combine these overlapping "microsessions" to estimate the start and end of user sessions in Cyclopath.

Use events which are requests for geographic objects have an additional use: when the map is zoomed in beyond a viewport of approximately 4km square, these requests are generated each time a user pans or zooms (see [21] for details on this technique), thus letting us know exactly which part of the map is onscreen.

A quirk of Cyclopath's architecture is that some use events – most notably requests for map tiles in the zoomed-out zoom levels – bypass our software and are served directly by Apache without authentication, and thus are always anonymous. We exclude these use events from analysis unless otherwise noted.

Uncovering Hidden Activity

As stated earlier, the data available for analysis of most online interactions is deficient in two ways:

1. Only visible actions are included. Most people, most of the time, are readers/viewers, not participants (editors or posters). Whether and how viewing and participating are related is a potentially rewarding question. Indeed, when only participation data are available, it is tempting to assume that participation and viewing are correlated. For example, some Wikipedia research has taken amount of editing an article receives as a reasonable proxy for how much the page is viewed. However, the only work we know that had access to Wikipedia view data, and thus

could compare viewing and editing, found that there was essentially no correlation between the two on a page level [22], but no research is available on a user view level.

2. Actions by users who are not logged in – *anonymous* actions – cannot be linked with those made while users are logged in. In principle, this has privacy benefits for users. However, from an analytical perspective, this is a severe limit.[1] For example, one concern with research that draws conclusions about the early stages of users' lifecycle in a system – as we did for Wikipedia [20] – is that the *early stages of a user's lifecycle may be hidden*. In other words, perhaps elite editors, or "Wikipedians" are not "born, not made." Instead, perhaps Wikipedians were users who did a fair amount of anonymous editing, learned something about Wikipedia conventions, decided they were interested in continuing in Wikipedia, and only then decided to create an account.

Since we collect a comprehensive log of Cyclopath user activity, the first problem does not arise: we can examine viewing behavior and how it related to participation. The second problem still exists. Indeed, of the 4 million use events we analyzed, 2.2 million of them were performed by not-logged-in users. However, the nature of the logs gives an opportunity to estimate which user was responsible for some of these 2.2 million.

Specifically, recall that we record the client IP address for all events and the username if the user was logged in. This pairing lets us infer that certain IP addresses are most likely to be associated with a single user: we call these *identified* IPs. We use the following procedure, the results of which are shown in Figure 2.

1. Create a table recording the IP address, username (perhaps null), and the number of events which occurred with that pairing.
2. Classify IPs according to the following procedure.
   (a) If an IP co-occurred with precisely one known user, classify the IP as **Identified**, and assume that all events from that IP are due to that user regardless of whether the event was **Logged-In**.
   (b) If an IP co-occurred with more than one known user, classify the IP as **Ambiguous**.
   (c) Otherwise, if the IP co-occurred with zero known users, classify the IP as **Anonymous**.

Similarly, this leads to a classification of registered users: **Unambiguous users** access Cyclopath only from Identified IP addresses – thus, we can identify all work done by these users whether they log in or not. There are 1172 of these unambiguous users. On the other hand, **Ambiguous users** access Cyclopath from at least one Ambiguous IP address, and we can identify only work done by these users while logged in. There are 440 ambiguous users in Cyclopath.[2] (We investigated a probabilistic rather than a binary approach – e.g., if an IP is used at least 90% of the time by one user, classify it as Identified for the leading user – but found that the results were very similar for all reasonable thresholds, and so we reverted to the simpler approach.)

[1] We're not interested in publishing these actions to the world, but rather using them for analytical access. We have user permissions 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.

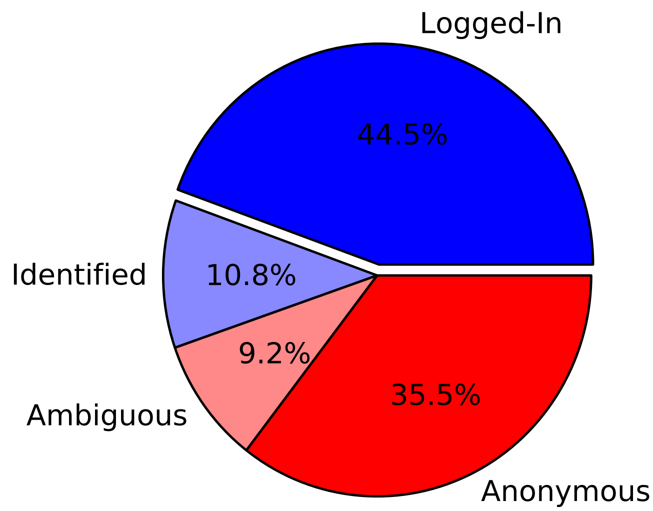


Figure 2. Frequency of different event IP classifications. Previous work was able to analyze only the Logged-In events, while we analyze everything in blue, the Logged-In *and* Identified work.

Following this procedure, we are able to assign 450,000 additional events – 20% of the events generated by non-logged-in users – to a known user. Subsequent analyses use only the Logged-In and Identified data unless otherwise noted.

The Effect of "Remember Me"

Many systems which allow logins have a feature called "Remember Me" which lets users persist a login session across time and browsing sessions. Cyclopath lies at an extreme – Cyclopath logins never expire, though they are not portable across IP addresses. But other systems with which we would like to compare Cyclopath are different. Most notably, Wikipedia logins expire after 30 days and are portable across IPs (this is a fairly typical expiration time).

To make comparisons with other systems more robust, we needed to explore the relationship between this design choice and user login decisions. Thus, we turned off the Remember Me feature in Cyclopath for two weeks: the checkbox was still available but had no effect – users who closed the Cyclopath application were completely logged out, with no username or any other state retained.

With Remember Me active, 47% of use events and 90% of revisions were logged in; but with Remember Me disabled,

[2] We exclude from analysis two additional groups of users: those closely affiliated with the Cyclopath team, test accounts, etc. (64 users), and those who registered but never logged in to the Cyclopath application (429 users). The latter, apparently impossible group is the result of a quirk of the Cyclopath registration process: it takes place on an external system which does not record the necessary user/IP pairs.

this dropped to 24% of use events and 62% of revisions logged in. We would expect Wikipedia and most other systems to lie between these two cases.

The implications of this are interesting: a very small change can have a dramatic effect on login behavior. Specifically for the current work, Cyclopath encourages people to view and edit while logged in (in addition to the very persistent "Remember Me" feature, other features like rating roads and trails and creating watch regions [21] are not available unless logged in), and this property is reflected in our results.

What is a cyclopath?

We know from prior research that "power users" or "elite editors" exist in systems such as Wikipedia [22, 20, 15] and Usenet [31]. We also know, from experience managing it, that such users seem to exist in Cyclopath. These are users who have made hundreds or thousands of revisions and often contact us to request more advanced functionality or let us know if something is broken.

To extend our work on Wikipedians [20], we needed to be able to classify a similarly elite group of contributors to Cyclopath. In [20] we defined Wikipedians as the 2.5% of all registered editors on Wikipedia that make the most revisions; here, we use this definition but increase the percentage to 5%, which yields 22 **cyclopaths**.[3]

However, in Wikipedia there is one intuitive way to rank editors: by number of revisions. There are, however, different namespaces which measure different *types* of revisions that users are making. For example, the edits in "Talk:" are discussion oriented while "Wikipedia:" provides a forum for policy discussion and arbitration.

Similarly, in Cyclopath, there are different types of editing that an editor can do. We could rank editors by number of revisions or by number of editing actions. We could break down the editing actions: geographic (e.g., moving roads or points) versus non-geographic (e.g., tagging, changing name, or commenting).[4] We could identify cyclopaths by looking at use other than editing, perhaps by counting use events. Registered users also have the ability to rate blocks and create watch regions, items that could be used to create a more personalized experience in Cyclopath; these could then be used to rank users by personal activity.

To decide which definition of cyclopath to use, we wanted to find one that accurately represented the system and activity within it, and examining the overlap between different rankings is helpful. Table 1 shows the overlap between top 5% of users when ranked using various methods. We chose to rank by number of editing actions, as this provided a reasonably close comparison to previous work while also remaining faithful to the structure and aims of Cyclopath. Thus, for this work, we define **cyclopaths** as the 22 editors who are the top 5% when ranked by number of editing actions.

[3]Note that we will use the term Cyclopath to refer to both the system and the singular elite user from this point on. We will try to distinguish between them with our language, but the system will always be capitalized and the user will always be lower case.

[4]When we counted editing actions, we counted all actions, whereas the geo and non-geo edit actions don't include deletions, but just additions 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 1. Overlap for different definitions of cyclopath.

LIFECYCLE OF CYCLOPATH USERS

As noted above, our novel analysis can consider viewing as well as editing and links some anonymous activity to registered users. What is gained? In this section, we investigate the initial stages of user activity in Cyclopath, patterns of activity over the course of users' involvement with the system, and how different activities influence user retention. We also note seasonal influences on usage.

In The Beginning: Educational Lurking?

Users' initial experience with a system is valuable to researchers, but in most systems, identifiable data for an individual is not collected until after that individual registers a user account. Thus, a user can try out a system without leaving any traces, meaning that previous analysis of initial user experiences has lacked an important source of data. We can rectify this. Specifically, we can study the extent of "educational lurking" [18, 26], where users explore the system for some period before deciding to join.

A potential process of educational lurking is:

1. Users engage in "read-only" activity; in Cyclopath, this includes browsing the map and requesting routes.
2. Users participate anonymously; in Cyclopath, they can add points of interest, edit the geometry of road and trail segments, tag objects, and any other map editing.
3. Users register an account, then continue and deepen their participation using this identity.

To investigate the extent to which this process occurred, we analyzed the Unambiguous class of users – these are the users whose pre-registration activity we can identify. We report the amount of time (a) from their first *view* to registration and (b) from their first *edit* to registration (Table 2), as well as (c) the amount of time from their first view to their first edit (Table 3). We observe several patterns.

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 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.

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 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.

Editing to Registration. Very few editors – only 18 of 286 – made their first edit before registering, and 10 of these edited immediately before registering; i.e., there was no pattern of people editing anonymously for a while, then deciding to join. Thus, we see here no evidence for educational lurking.

Viewing to Registration. For viewing, the picture is intriguingly mixed: well over half of users first viewed immediately before registering. The modal behavior pattern for those who join is: look, then join immediately. On the other hand, 25% of users (297 of 1172) made their first view at least a week prior to registration. Thus, a meaningful proportion of users did use Cyclopath for a while before deciding to join, suggesting some educational lurking.[5]

Viewing to Editing. Here, the data are more or less bimodal. About 45% of editors edited immediately or within a day of their first view, but about 44% waited more than a week to edit. The latter group's pattern – look around for a while before making the first edit – is also consistent with educational lurking.

Discussion. Our findings offer equivocal evidence about educational lurking. Thus, we present several observations

[5]This table also shows that our identification process is noisy – it seems highly unlikely that anyone would register before seeing the system, yet there is nonzero data in that portion of the table. We conjecture that much of this is for two reasons: (a) Cyclopath's quirky registration system, which takes place on an external system which does not record the user/IP pairs we need for the identification process, and (b) in the early days of the system, accounts were created by Cyclopath staff, not users directly.

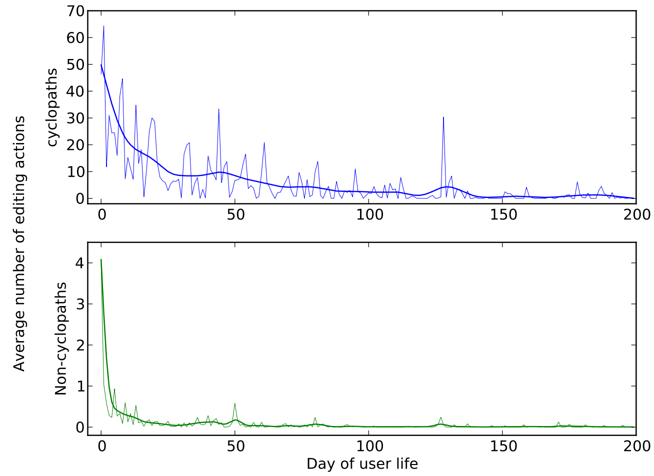


Figure 3. Number of edit actions per day of user life, for cyclopaths and non-cyclopaths who edit. For example, a typical cyclopath on his or her 40th day after first viewing is making about 9 editing actions daily, while a typical non-cyclopath is making about 0.2.

and ideas for related future study. First, different systems reserve different benefits for registered users. Sometimes, certain actions can only be done by registered users – only registered Wikipedia users may create new articles. Sometimes, anonymous users' contributions are pejoratively labeled – Slashdot's "Anonymous Coward." In Cyclopath, the key benefit of registering and logging in is rating road and trail segments and thus receiving personalized route recommendations. Regardless, users cannot develop an identity or gain reputation without registering and creating a persistent username [9]. The particular benefits a system offers to registered users create different incentives to register.

Second, while our quantitative results *suggest* educational lurking in Cyclopath, follow-up qualitative research could sharpen our understanding. For example, we could interview users to ask why they joined or to describe their stages of activity in the system, or we could gather this information using some form of contextual experience sampling [13], perhaps by presenting a window with questions after registration or the first edit. A clearer understanding of why people join would inform the design of Cyclopath and related systems, perhaps by emphasizing benefits that already encourage registration or suggesting additional benefits.

The Cyclopath User Lifespan

Once users have registered, what patterns do we see in their behavior? We traced the activity of Cyclopath users over time, using our previous findings from Wikipedians [20] as a baseline. That research found that both elite and non-elite Wikipedia editors began with a burst of activity, then tailed off to a much lower and relatively constant level. We present an analogous analysis of Cyclopath, extending the analysis to viewing, not just editing.

Figure 3 shows the average number of edits per user according to days since users' first views, segregated into elite (cyclopaths) and non-elite editors (non-cyclopaths). This analysis is directly comparable to [20] and shows the same pat-

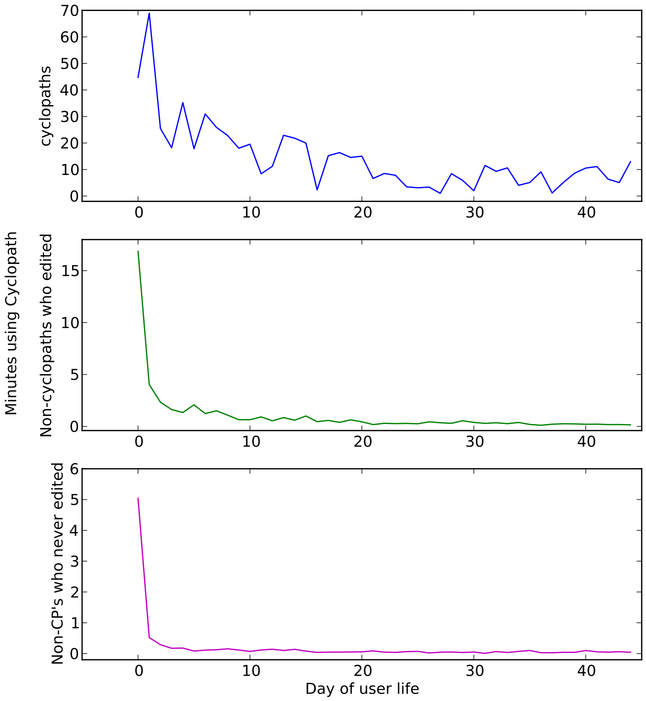


Figure 4. Minutes spent using Cyclopath, for cyclopaths, non-cyclopaths who edit, and non-cyclopaths who have never edited.

tern: elite editors are elite *from the beginning*, and all start with an initial burst of activity, then tail off to a lower and relatively constant level. This tailing off of activity for cyclopaths is much slower than observed in [20] for Wikipedians. While Wikipedians leveled off by their 16th day (of editing), it isn't until after their fiftieth day in the system that cyclopaths begin to level out.

Figure 4 shows a similar analysis for viewing behavior. Here we include a third group of users: registered users who have never edited. The story is similar: all users have an early "burst" of activity, and the size of that burst is predictive of subsequent activity.

Figure 5 lets us look at user activity over time in a somewhat different way: it measures *retention*. A user is retained at day n if he or she visits Cyclopath n days after his or her first view or on any subsequent day. The patterns are similar to those noted above; however, the difference between non-cyclopaths who edit and non-cyclopaths who do not is clearer. For example, after 100 days in the system, 50% of non-cyclopaths who edit will return, but only 30% of non-editors will.

Discussion. These lifespan and retention findings suggest several future possibilities, building on the fact that many users continue to use Cyclopath even if they participate minimally or not at all. First, their currently-hidden presence could be made visible in the aggregate: e.g., showing the "audience size" for a particular region could motivate additional work by current editors and entice other users to begin editing. Second, we should explore the details of non-editors' and low-editors' activity: e.g., to what extent do

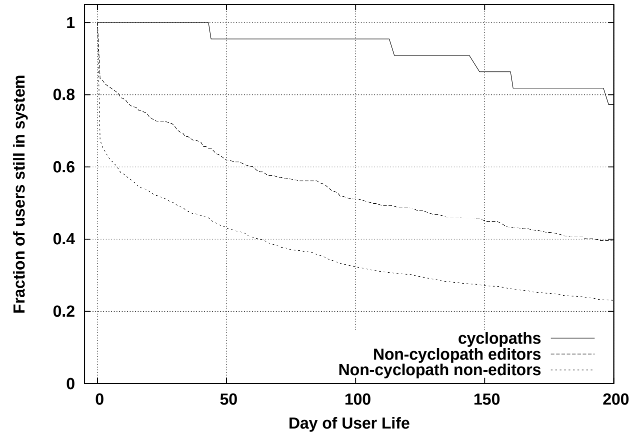


Figure 5. Retention, measured as the proportion of of users whose last visit to Cyclopath is n days after their first view. We end this graph at 200 days of user life: as Cyclopath is only 16 months old, very many users are still active at present and have not yet reached the natural end of their activity.

they request routes vs. browse the map? Third, this reveals an opportunity to learn more about the motivations of and barriers to editing with interviews or contextual surveys as mentioned above. Finally, we could build on the results of these steps to design interventions to attract greater participation [24]: e.g., if non- or low-editors make use of Cyclopath's route finding mechanism, feedback making clear that user edits lead to better routes might motivate editing.

In addition to the steps outlined above, in the future we are interested in learning more about the lifecycles of users who never register, purely-anonymous users. However we do not believe that the current IP-based technique for identifying users is sufficiently robust when identifying purely anonymous users so we are pursuing other techniques for this purpose.

Location matters: It's hard to bike in the snow

It is widely recognized that online communities change over time. For example, Wikipedia grew exponentially from its founding in 2001 to about 2007, and then growth began to slow [28]; moreover its now-rich policies have evolved from very simple beginnings. Thus, the environment for Wikipedia editors differed substantially between (say) 2002, 2006, and 2009.

Less recognized are seasonal effects. However, there are many communities with clear "seasons." For example, students follow the academic calendar, with different educational and social activities at various times of year.

Cyclopath is for cyclists located in Minnesota, a state known for having distinct seasons and cold, snowy winters. While a small minority of cyclists ride year-round, most do not. Therefore, it is unsurprising that Cyclopath has pronounced seasonal activity levels. Figure 6 shows that user activity in 2008 began to decline at the beginning of November and then increase in mid-March. This was true for both cyclopaths and non-cyclopaths.

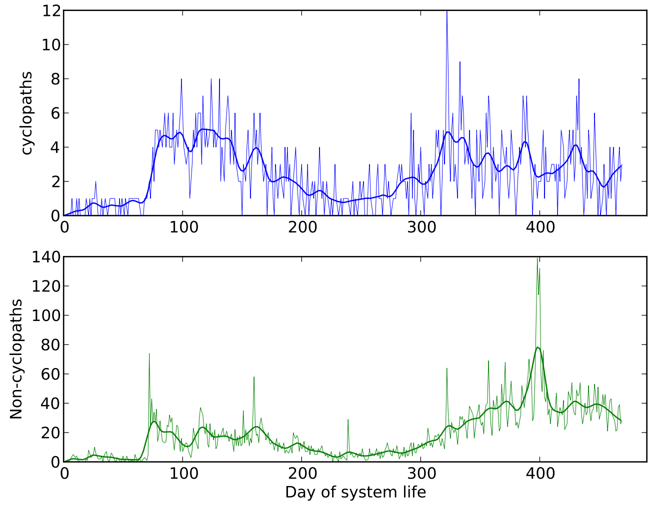


Figure 6. Number of users who view Cyclopath on a given day after Cyclopath went live. Day 100 is Aug. 15, 2008, Day 200 is Nov. 23, 2008, Day 300 is March 3, 2009, and Day 400 is June 11, 2009

Discussion. If the activity supported by a system has a seasonal nature, then designers should consider what features and activities are seasonally appropriate. For example, in the cycling off-season, Cyclopath could promote work campaigns to improve the quality of the map, cycling-related discussions, notifications of cycling-related events, reminders about off-season maintenance, etc.

GEOGRAPHIC VIEWING AND EDITING

Seasonal usage is one way the essentially localized nature of Cyclopath is revealed. An even more obvious way is in the geographic nature of Cyclopath activity. In most systems, selecting "where" to view or edit means choosing a *topic* of interest, but in Cyclopath, it means choosing an *area*. We are interested in the geographic "shape" of editing and viewing, and in how they relate. The relationship between editing and viewing is unusual because editing is public behavior but viewing is private. Therefore, to the extent that viewing and editing are correlated, users' private activity may be inferred.

The analyses of this section consider all the logged-in and identified activity of the 400 registered Cyclopath users who have saved at least one revision. The metrics used in this section are:

1. **Number of revisions**. How many times has a user saved their edits?
2. **View compactness**. How geographically concentrated (or dispersed) is a user's viewing behavior? We measure this using all the viewports of size 4km square or smaller for each user. (Larger viewports reveal little map detail and thus are not good indicators of a user's interest.) We measure compactness by computing the geometric centroid of each user's viewports and then aggregating the average distance from each such viewport to the centroid.
3. **Edit compactness**. This is the analogue of view compactness, but for editing. Centroids are compute by taking the centroid of objects modified in revisions. (Note that this measure is zero when a user has saved only one revision.)

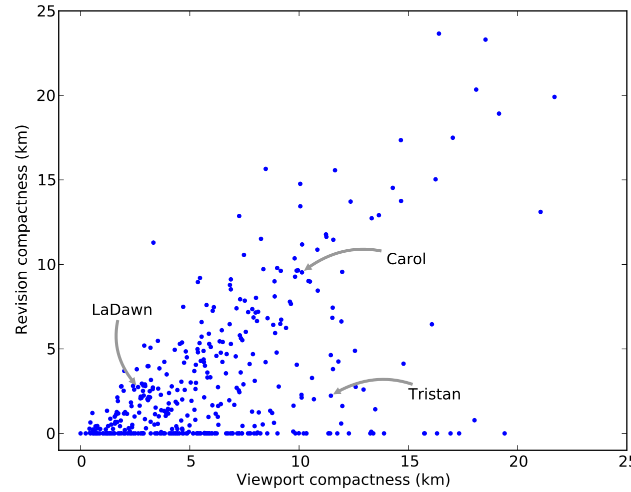


Figure 8. Scatterplot showing the relationship between view compactness and edit compactness.

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.

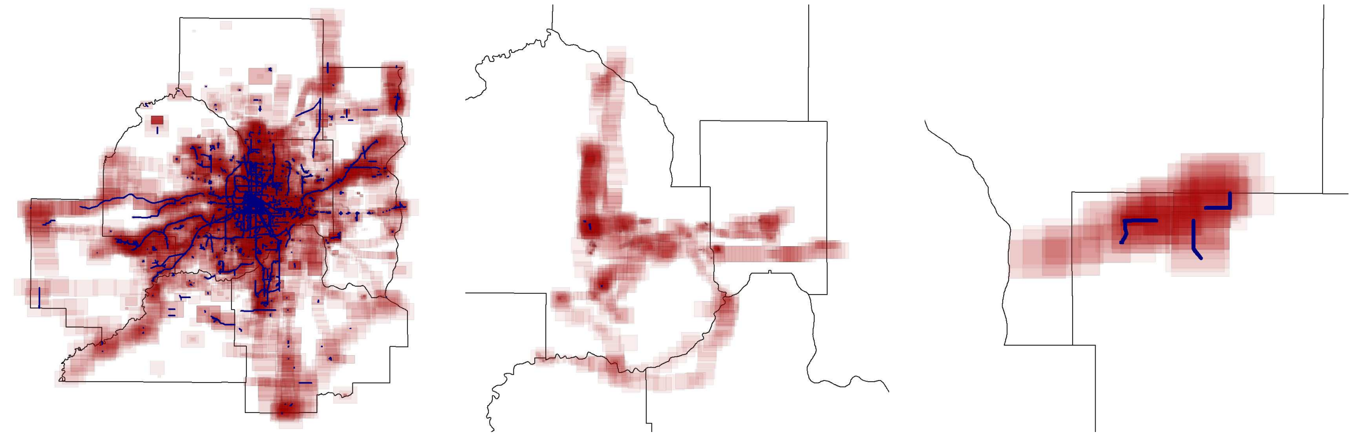
Figure 7 illustrates these metrics with data from three users with differing levels of activity and patterns of geographic activity (these data are shown with permission of the users). The maps suggest that public editing may give good indications of a user's private viewing activity. To test this conjecture systematically, we did several global comparisons.

First, Figure 8 shows a positive relationship between view and edit compactness: users who view in a wide variety of places also edit in a wide variety of places.

These results suggest that the diversity of one's viewing and editing may influence one another. Previous work [24] showed that editors can be attracted to edit in essentially arbitrary areas: this might be useful if more diverse viewing is desired (for example, to encourage more users to "watch" for changes in more of the map). On the other hand, if more diverse editing is desired (for example, to fix map errors in areas that have received little attention), then viewing campaigns such as "Ride of the week" or simply "Do you know what's in Neighborhood X?" might be effective.

Similarly, Figure 9 shows that viewport coverage increases with the number of revisions and approaches completeness for the most prolific editors. In other words, the more users edit, the more they are revealing about their private viewing activity.

But does this actually matter? It seems likely that the areas one views most are close to potentially sensitive places like home, work, and commuting routes. Follow-up studies with users are needed to investigate how sensitive they consider their viewing activity to be. How does it compare to other private data like library books checked out or web pages viewed? We also are interested whether this relation-



(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 7. 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.

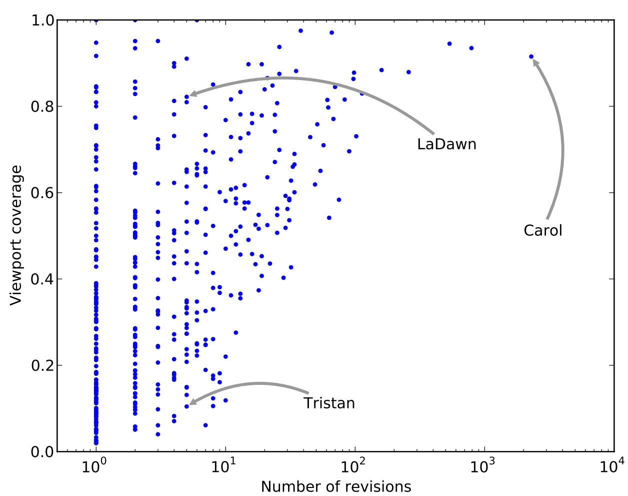


Figure 9. Scatterplot showing the relationship between number of revisions and viewport scale. Note that the x axis is on a log scale.

ship holds in other systems. For example, does the history of articles edited by Wikipedia users predict the articles they view? This particular question is currently untestable due to lack of data, but that might change in the future.

CONCLUSION

The intersection of online community and geographic information creates a new and important area for HCI. In this paper we described a geo-community network called Cyclopath that serves bicyclists in Minnesota. This paper has provided a quantitative analysis of Cyclopath's use during its first year in public. Cyclopath is a rich source of behavioral data. As a wiki, it records a complete record of user edits. It also records users' map viewing activity. Analysis of web server logs allowed 20% of anonymous activity to be linked with registered users, letting us make inferences about what users do before they register. Together, these data let us draw a picture of the types and lifecycles of Cyclopath users and begin to understand the ways in which they interact with geographic information.

Our research provides us with a new and deeper characterization of our users. There is a cohort of most active users – cyclopaths – who do the bulk of the work, whose degree of activity differs from other users from the very first day, and who stay involved with Cyclopath longer. This finding mirrors results for the cohort of most active Wikipedia editors. It is far from obvious that a relatively small, local community of bicyclists editing a map would exhibit similar activity patterns as an enormous distributed community involved in producing a global encyclopedia. Further, analysis of users' pre-registration and early viewing activity gave some evidence for "educational lurking." Whether this is true for Wikipedia, and what it would mean for research on Wikipedia editors and their lifecycles is an open question.

Focusing in on geographic work, our analyses revealed interesting patterns and relationships between viewing and editing. One finding is that there is a relationship between the geographic locality of viewing and editing. This finding is compelling because it suggests that collective viewing might be steered by focusing editing behavior on particular regions (as demonstrated in [24]), and that collective editing might be steered by interventions that encourage viewing of particular areas. In addition, the finding that those who edit intensively are revealing information about their viewing behavior has potential privacy implications. It suggests further inquiry into user attitudes toward the privacy of their geographic behavior. More generally, this raises analogous questions for Wikipedia, namely what editing behavior reveals about reading behavior.

The work we have described here is just a beginning. Our findings suggest a number of new design directions for Cyclopath and show that qualitative work is needed to understand the motivations and attitudes that attend the large scale quantitative patterns we've observed, and to suggest investigations of topics ranging from educational lurking to what drives the geographic spread of users' viewing and editing behavior. The work also has possibilities for future work beyond Cyclopath, – do other successful open content systems have cohorts that are the equivalent of cyclopaths and Wikipedians?

ACKNOWLEDGEMENTS

We thank the members of GroupLens Research, the Cyclopath team, and the Cyclopath user community. This work was supported in part by the National Science Foundation grants IIS 05-34692 and IIS 08-08692 and in part by a GAANN fellowship.

REFERENCES

1. Adamic, L. A. et al. Knowledge sharing and Yahoo Answers: Everyone knows something. In *Proc. WWW*. 2008.
2. Anderson, C. *The Long Tail: Why the Future of Business is Selling Less of More*. Hyperion, 2006.
3. Bryant, S. L. et al. Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. In *Proc. GROUP*. 2005.
4. Carroll, J. and Rosson, M. Developing the Blacksburg electronic village. *CACM*, *39* (1996), 69–74.
5. Convertino, G. et al. Supporting content and process common ground in computer-supported teamwork. In *Proc. CHI*. 2009, 2339–2348.
6. Elwood, S. Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. *GeoJournal*, *72*, 3–4 (2008), 133–135.
7. Fisher, D. et al. You are who you talk to: Detecting roles in usenet newsgroups. In *HICSS*. 2006.
8. Frankowski, D. et al. You are what you say: Privacy risks of public mentions. 2006, 565–572.
9. Friedman, E. and Resnick, P. The social cost of cheap pseudonyms. *Economics and Management Strategy*.
10. Golder, S. and Huberman, B. The structure of collaborative tagging systems. *Arxiv preprint cs.DL/0508082*.
11. Goodchild, M. F. and Gupta, R. Workshop on volunteered geographic information (2007). http://www.ncgia.ucsb.edu/projects/vgi/.
12. Harper, F. et al. Talk amongst yourselves: Inviting users to participate in online conversations. In *Proc. IUI*. ACM New York, NY, USA, 2007, 62–71.
13. Intille, S. and Ma, X. Eliciting user preferences using image-based experience sampling and reflection. In *Proc. CHI*. 2002.
14. Jones, Q. et al. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Info. Sys. Research*, *15*, 2 (2004), 194–210.
15. Kittur, A. et al. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proc. alt.CHI*. 2007.
16. Lampe, C. and Resnick, P. Slash(dot) and burn: Distributed moderation in a large online conversation space. In *Proc. CHI*. 2004, 543–550.
17. Lo, C. P. and Yeung, A. K. W. *Concepts and Techniques of Geographic Information Systems*. Prentice Hall, 2006, 2nd edition.
18. Nonnecke, B. and Preece, J. Shedding light on lurkers in online communities. *Ethnographic Studies in Real and Virtual Environments* (1999), 123–128.
19. Nonnecke, B. and Preece, J. Lurker demographics: Counting the silent. In *Proc. CHI*. 2000, 73–80.
20. Panciera, K. et al. Wikipedians are born, not made: A study of power editors on Wikipedia. In *Proc. GROUP*. 2009, 51–60.
21. Priedhorsky, R. and Terveen, L. The computational geowiki: What, why, and how. In *Proc. CSCW*. 2008, 267–276.
22. Priedhorsky, R. et al. Creating, destroying, and restoring value in Wikipedia. In *Proc. GROUP*. 2007.
23. Priedhorsky, R. et al. How a personalized geowiki can help bicyclists share information more effectively. In *Proc. WikiSym*. 2007.
24. Priedhorsky, R. et al. Eliciting and focusing geographic volunteer work. In *Proc. CSCW*. 2010.
25. Schuler, D. Community networks: Building a new participatory medium. *CACM*, *37*, 1 (1994), 38–51.
26. Soroka, V. and Rafaeli, S. Invisible participants: how cultural capital relates to lurking behavior. In *Proc. WWW*. 2006, 163–172.
27. Soroka, V. et al. We can see you: A study of communities' invisible people through ReachOut. In *Proc. Communities and Technologies*. 2003.
28. Suh, B. et al. The singularity is not near: Slowing growth of wikipedia. In *Proc. WikiSym*. 2009.
29. Sweeney, L. K-anonymity: A model for protecting privacy. *Uncertain. Fuzziness Knowl.-Based Syst.*, *10*, 5 (2002), 557–570.
30. Welser, H. T. et al. Visualizing the signatures of social roles in online discussion groups. In *Social Structure*. 2007.
31. Whittaker, S. et al. The dynamics of mass interaction. In *Proc. CSCW*. 1998.