Social contribution settings and newcomer retention in humanitarian crowd mapping

(Article begins on next page)

09 January 2025

Social Contribution Settings and Newcomer Retention in Humanitarian Crowd Mapping

Martin Dittus[1], Giovanni Quattrone[2], and Licia Capra[2]

[1] ICRI Cities, UCL, London, United Kingdom
[2] Department of Computer Science, UCL, London, United Kingdom

Abstract. Organisers of crowd mapping initiatives seek to identify practices that foster an active contributor community. Theory suggests that social contribution settings can provide important support functions for newcomers, yet to date there are no empirical studies of such an effect. We present the first study that evaluates the relationship between colocated practice and newcomer retention in a crowd mapping community, involving hundreds of first-time participants. We find that certain settings are associated with a significant increase in newcomer retention, as are regular meetings, and a greater mix of experiences among attendees. Factors relating to the setting such as food breaks and technical disruptions have comparatively little impact. We posit that successful social contribution settings serve as an attractor: they provide opportunities to meet enthusiastic contributors, and can capture prospective contributors who have a latent interest in the practice.

Keywords: Humanitarian OpenStreetMap Team, crowdsourcing, crowd mapping, volunteering, mapathon, mapping party.

1   Introduction

Since its inception in 2010, the Humanitarian OpenStreetMap Team (HOT) has coordinated thousands of volunteers in the creation of maps for humanitarian purposes. All maps are published on the online mapping platform OpenStreetMap (OSM), free to use under a liberal license. Contributors have traced satellite images and digitised data collected in the field in response to Typhoon Haiyan in the Philippines, the earthquake in Nepal in early 2015, and other disasters where humanitarian aid teams required updated maps to coordinate their work. Despite these efforts, vast regions of the inhabited world remain unmapped. In November 2014, HOT and partnering aid organisations launched Missing Maps (MM), a proactive effort to produce new maps before they are needed in times of crisis. However, while emergency response initiatives regularly benefit from new contributor influx resulting from widespread media coverage, MM has to learn how to build mapper capacity in the absence of urgent causes.

Since its inception, MM organisers have refined practices that support the collective effort. Among them are so-called mapathons, social event settings which

allow regional community groups to come together in person, to learn the practice and to socialise. Organisers of these events pursue several outcomes: to initiate newcomers to the practice, to have them produce maps over the course of the evening, but importantly also to then retain these new contributors for future activities. The volume of mapping data produced at mapathons can easily be measured with existing tools, however it is not currently clear how many attendees remain active afterwards. Do these events have a measurable impact on contributor retention? There is some evidence that communal event settings can play an important role in fostering sustained contributor activity. In a recent study of HOT contributor engagement, it was found that mapping initiatives which organised mapathons had higher newcomer retention rates [7], however it is not yet known whether this can be attributed to the mapathon format itself.

The present study places a focus on the group experience of HOT mappers: mapathons as social contribution environments, and their impact on newcomer retention. The research addresses two primary concerns, to produce new empirical evidence for the effects of colocated practice in online crowd mapping, and to identify some of the contributing factors. We identify three groups of first-time contributors who physically meet in different social contribution settings, and compare their retention to two groups of online contributors who likely never met in person. We find that participation in mapathons can be associated with a significant increase in newcomer retention. In particular, retention was highest for cohorts that meet regularly, compared to cohorts that only met once, or that likely never attended mapathons. A comparative analysis of different aspects of the setting (such as food breaks and technical disruptions) revealed that these had comparatively little impact on longer-term engagement. The results suggest that organisers may be able to increase newcomer retention by offering regular opportunities for social encounter and peer learning.

2   Related Work

In the study of computer-supported cooperative work, work contexts are often considered along two dimensions: whether participants are colocated (they work in the same place), and whether they operate synchronously (they work at the same time) [1], [12]. According to this model, HOT online practice is asynchronous and remote, and contributors can act entirely independently of each other. HOT mapathons however are synchronous and colocated. In the context of a global online community, colocation may appear an artificial and needless constraint. Yet a range of literature suggests that it can have important benefits for the experience of first-time contributors.

In distance learning and online education, it was found that **colocated practice** can augment online settings in important ways. The proximity of real-world social interactions can have important benefits for the learning experience of participants, in part by allowing for different forms of knowledge exchange [10]. Similarly, studies of communal software development settings found that social encounters within a community of practice create opportunities for mentoring

and learning, provided there is a mix of experiences among attendees [8], [18]. Such events can allow newcomers to become expert contributors through situated learning, or so-called peripheral participation, and this can become an important motivation to continue participating in the community [13]. However it was also found that there are tradeoffs between the mix of experts and novices, and task interdependence: events at which experts contribute to independent tasks may yield outputs, but contribute less to community growth. Conversely, events with a larger share of newcomers may contribute to community growth [18]. For open source development groups it was suggested that project attractiveness and individual motivations play an important role in the decision of a newcomer to join a project, however that there can be many hindering factors that lead to an aborted onboarding process [17]. In a study of sustained open source participation it was found that newcomers can particularly benefit if the nature of an early task fosters interactions between participants [8].

In **online practice**, there is evidence that the socialisation experience of first-time contributors can increase their contributions and long-term retention: in an evaluation of Wikipedia socialisation tactics, it is observed that early user retention was increased by the use of welcome messages, assistance, and constructive criticism [4]. On the other hand, invitations to join yielded a steeper decline in contributions by new editors. A further study confirms that a successful early socialisation experience among Wikipedia contributors is associated with and can sometimes predict increased contributor engagement [5]. However the authors also observe that the causal structure between socialisation, motivation, and participation is not entirely clear. Further studies identify similar effects [3, 4], [9].

To our knowledge there are no quantitative studies of colocated practice in a global **crowd mapping** project, and of its effects on newcomer retention. There are early studies of OSM mapping parties, these are similarly structured around social mapping experiences, however they typically aim to map the local area and involve colocated practice by necessity [11], [15]. It is not clear how contemporary HOT mapathons compare to these earlier settings. Among early OSM contributors, it was found that an individual's local geographic knowledge was the most significant driver to contribute [2]. On the other hand, it is conceivable that the global scope and perceived social benefit of HOT mapathons attracts different audiences, and fosters a different form of long-term engagement compared to OSM mapping parties.

3   Research Questions

RQ1: Does mapathon attendance improve newcomer retention?

According to existing research, colocated practice in social contribution environments can be associated with improved newcomer retention rates. Qualitative literature in social psychology, online community studies, and related domains provide support for such an association [6], [14], [16], [19]. However, there is no

empirical evidence available to confirm such an effect in a crowd mapping context. We seek to establish such evidence by analysing contributor activities of the HOT crowd mapping community.

RQ2: Which specific factors contribute to increased retention?

Organisers of HOT mapathons have some influence on the setting and format of the events, however to date they have no basis to justify certain choices. In particular, it is not currently known which factors of the setting may affect subsequent newcomer retention. We seek to identify specific aspects of social contribution settings that have an impact on newcomer retention, based on the observation of HOT mapathons and an empirical analysis of their outcomes.

4 Methodology

All our analyses are based on a public record of HOT contributions. We captured the contribution activity of first-time mappers belonging to two separate mapathon cohorts in London, observing a total of 14 events. Subsequent newcomer retention is compared with that of an online control group, a set of first-time online contributors who did not attend the London mapathons. We further compared these cohorts to a second control group of participants of the Arup "Mappy Hour", at an employee-initiated regular mapping event. Furthermore, we co-developed a set of mapathon features in a workshop with MM organisers. Participation outcomes of the 14 mapathons were then compared in relation to these features. The following sections explain these steps in more detail.

4.1 HOT Contribution History

A primary data source for the research is the HOT Tasking Manager, a website which helps coordinate the work of thousands of online contributors while reducing edit conflicts.[1] It presents a list of currently active HOT projects, along with contextual information and mapping instructions. Within each project, work is divided into smaller tasks. Contributors start by selecting a specific project and task, and then contribute to the map using OSM tools. The Tasking Manager also serves as a public record of HOT participation: every project records a list of its past contributors. A further data source for the research is the full OSM edit history, a large public data set which captures OSM map contributions over time. All HOT mapping activity takes place on OSM, and the map contributions by HOT volunteers are contained in this edit history. The full data set is freely available for download.[2]

In a preparatory stage, we identified the map contributions for every Tasking Manager project. Since summer 2015, OSM editing tools automatically annotate

---

[1] `http://tasks.hotosm.org`
[2] `http://planet.osm.org/planet/full-history/`

changesets with a HOT project identifier, which makes such an identification straightforward. In cases where this was not provided, edits were instead identified based on their location, date, and contributor. This provided us with the full set of HOT mappers, their contributions to HOT, and any further contributions they made to OSM which were not linked to HOT activities.

4.2   Study Period

From November 2014, MM organisers in London started hosting regular mapathons that were open to the public. These provided us with an opportunity to observe contributor retention over time. The first MM mapathon marks the start of our study period. We seek to study participants after their first attendance for a subsequent period of up to 90 days. Our evaluation is based on a snapshot of the OSM edit history that was published on 11th of January 2016, which means the cutoff date for the inclusion of an event is 13th of October 2015. The study considers newcomer activity between the date of the first mapathon on 24th of November 2014, and the last mapathon held on 6th of October 2015. In this period, a total of 14 mapathons took place.

4.3   Study Cohorts

Mapathon Cohorts: Monthly and Corporate   The organisers of Missing Maps mapathons in London are affiliated with the British Red Cross and Médecins Sans Frontières. Throughout the study period, two types of mapathons were organised by this team. Ongoing *monthly mapathons* are open to the wider public, and hosted at a different venue every month. Event sizes are limited by venue capacity rather than interest, and typically vary between 50 and 100 people. Events start in the early evening on a weekday, and typically last three hours. From early 2015, MM further organised a number of *corporate mapathons* for staff members at large corporations, these are one-off events that are not open to the public. The setting and format is comparable to monthly mapathons, however the attendee mix differs in some important ways. Typically all attendees are first-time contributors, and training is limited to basic mapping techniques. According to organisers, participants tend to be office workers and highly computer literate.

These form our mapathon cohorts:

– 11 monthly mapathons between 24th November 2014 and 6th October 2015.
– 3 corporate mapathons on 12th February, 15th May, and 6th October 2015.

For our analysis we seek to identify newcomers who attended these events, and then observe their activity in the subsequent days and weeks. However, there is no public register of HOT mapathon attendance. Instead, we estimate event attendance based on a limited set of information that is readily available: event dates and times, and the list of HOT projects which were worked on during each event. Event dates are generally made public, for example on the

MM homepage[3]. Project lists were collected by participating in the events, or by consulting with organisers after the fact, and in total comprise 19 HOT projects across the 14 mapathons. Since MM mapathons involve proactive HOT mapping initiatives rather than urgent crises, their projects are not listed in a prominent position on the Tasking Manager homepage. Instead remote mappers need to make a conscious effort to find them, either by paging through the listing of active projects, or by searching for them by name. As a result, mapathon activity is clearly visible in the contribution timelines of these projects, and there are only few contributors in the hours before or after a mapathon.

We further identified HOT newcomers among these attendees, first-time mappers with at most one prior day of OSM contributions. The threshold of one day was chosen because attendees are generally asked to sign up to OSM and make some test contributions before the event. As a result, it is plausible that many newcomers may already have made some minor contributions to the map before they attend their first event.

Date	Cohort	Attendees	Newcomers	% newcomers
2014-11-24	Monthly	64	37	57.8%
2014-12-15	Monthly	58	24	41.4%
2015-01-27	Monthly	52	16	30.8%
2015-02-12	Corporate	50	44	88.0%
2015-02-24	Monthly	49	25	51.0%
2015-03-31	Monthly	62	29	46.8%
2015-04-28	Monthly	51	19	37.3%
2015-05-15	Corporate	191	174	91.1%
2015-06-02	Monthly	27	6	22.2%
2015-07-07	Monthly	51	15	29.4%
2015-08-04	Monthly	87	49	56.3%
2015-09-01	Monthly	41	15	36.6%
2015-10-06	Corporate	30	28	93.3%
2015-10-06	Monthly	69	24	34.8%

Table 1. Estimated attendance at the 14 mapathons under study, including the number and share of first-time attendees.

Table 1 summarises our attendee estimates per mapathon, accounting for both newcomers and more experienced participants. In total, more than 600 distinct attendees participated across the 14 events. Among these, we identified 505 newcomers, approximately evenly split between the two cohorts. They represent 82% of all attendees across the 14 events. The data shows that the share of newcomers differs significantly between the event cohorts: corporate events on average are attended by 90% newcomers, while the monthly mapathons are attended by 40% first-time attendees.

Online Control Groups: Matched MM and Nepal As first online control group we identified HOT contributors who engaged in comparable work,

---
[3] http://www.missingmaps.org/#events

but likely never attended a mapathon. We identified new HOT contributors in the study period who started with one of the same 19 MM projects used for mapathons, but who were not among the attendees identified for these events. In total, 550 first-time HOT contributors matched these criteria. Some of these may have attended mapathons in other cities, however the large sample size and long study period makes it likely that a significant share of this group started as online contributors. This group comprises our "matched" online cohort: contributors who started out doing the same work as MM mapathon attendees, but who were unlikely to have done so at a mapathon. That is, they likely joined the crowd mapping platform online.

As a further control group we added a second online cohort of newcomers who started mapping during an urgent disaster event, contributing to a different set of projects than the other groups. This group was included so we could compare the previous settings to a different kind of stimulus which may feasibly attract new engaged mappers. We selected HOT newcomers who joined to help with the Nepal emergency response in April 2015, their initial contributions were to urgent projects that were focused on emergency response mapping. None of these volunteers attended a MM mapathon in London when they first started mapping. In total, 4,518 first-time HOT contributors fall into this group, they comprise the "Nepal" online cohort.

Arup Mappy Hour As a final point of reference we chose Arup "Mappy Hour" participants, these are staff members of the multinational engineering consultancy Arup who regularly meet to contribute to HOT. Their office setting may be comparable to that of corporate mapathons, although their events are peer-organised by staff members, not external organisers.[4] According to organisers, Arup Mappy Hour emerged in early 2015 out of the independent activities of multiple staff members. Mappy Hour groups are comparatively small (under 20 attendees). Mappy Hour attendees can be identified in the HOT contribution history because they annotate their HOT contributions with an $\#Arup$ tag. Based on these annotations, we found that 135 HOT newcomers had attended a Arup Mappy Hour session in the observation period. These are contributors who had at most one day of prior OSM contribution experience before they attended their first Mappy Hour.

4.4 Mapathon Features

In order to address RQ2, we sought to identify specific aspects of the mapathon format that may have an impact on participant engagement and retention, with a focus on aspects that are under organiser control. This analysis is restricted to monthly and corporate mapathons only. For these cohorts, it was possible to observe hundreds of participants over the course of the observation period.

---

[4] http://doggerel.arup.com/mapping-the-worlds-most-vulnerable-regions/

In comparison, Arup Mappy Hour events are held in a non-public setting. Online cohorts are excluded from the analysis because the contribution setting of participants is not known.

We organised a workshop with MM organisers to identify aspects of a mapathon that may plausibly encourage or discourage continued engagement. Workshop participants developed a set of hypotheses of potential mapathon aspects that may affect participant engagement. Based on these we developed a set of event features which are easily observed, comparable across events, and were identified as potentially important factors because they can affect the actions of and interactions between attendees.

These features are summarised in Table 2. In the following sections we will describe each of the features in turn, discuss our motivation to include it in the study, and describe the associated data collection process.

Aspect	Variable	Description
Cohort	*cohort*	Monthly or corporate mapathon?
Attendees	*hot_mappers*	% with prior HOT contributions
	home_mappers	% who mapped at home
	osm_experts	% with > 50 days of OSM activity
	prev_attendees	% repeat attendees
Setting	*social_food*	Food served in separate area?
	tech_issues	Larger technical disruptions?
Tool use	*josm_learners*	% newcomers learning JOSM?

Table 2. Mapathon features collected per event.

Attendee Mix The attendee mix was considered an important aspect of the attendee experience: according to organisers, attendees who are experienced in mapping can provide important peer support, and the presence of an existing community of practice may affect the motivation of newcomers to keep coming back. Mapathon features relating to the attendee mix were derived from the OSM edit history. We identifyied the share of attendees with prior contributions to HOT, and separately those who contributed to HOT outside of a mapathon ("at home"). We further identified OSM experts with more than 50 days of prior OSM contributions, this approximately captures the 10% most experienced attendees across all events. Finally, we identified repeat attendees: the share of attendees who have been to at least one previous mapathon. This share of repeat attendees can be regarded as an indicator of the presence of a community of participation.

Food Served in Separate Area Organisers further debated the role of food as social catalyst. Attendees are always provided with free food. At most events,

food is served at the desks, and attendees can resume work while they eat. At some events, however, the setting is more conducive to social interactions between attendees. Organiser experience showed that food that is served in a separate room may disrupt the work, but it also tends to encourage mingling. At three of the 14 mapathons, food was served away from desks, for example as a buffet in a separate room, introducing opportunities to socialise.

Larger Technical Disruptions Technical issues at mapathons can have a negative impact on the overall event experience when they disrupt the contribution process of many attendees. In some cases this merely interrupts the contribution process for a short time. However more severe disruptions can lead to frustrating experiences for both organisers and attendees, for example when earlier work is lost as a result, and has to be repeated. Four events were disrupted by technical issues that affected all attendees.

Share of Newcomers Learning JOSM There was further debate among organisers on the topic of tool use. Mapathon organisers train most newcomers in the use of iD first. This editor is web-based, simple to learn, and does not require the installation of software. However, some attendees start by learning JOSM, this editor is more complex but also more powerful, and it allows for faster mapping. It needs to be installed on the attendee's laptop, and this process can take some time. In conversation, organisers stated that contributors with professional GIS and IT backgrounds tended to prefer JOSM over iD, however there is a concern that some newcomers may be discouraged by the more complex interface. Annotations in the OSM edit history allow us to determine which editor was used for a particular contribution. Based on this data we computed the share of JOSM learners at each event.

4.5 Approach

In order to address RQ1, we compare newcomer retention across three event cohorts and two online cohorts, involving the study of hundreds of first-time mappers during their initial activity period of 45 days. We distinguish three aspects of participation: the initial learning of the contribution practice during the event (*initiation*), subsequent mapping at home over the following days (*activation*), optional repeat attendance at a mapathon (*revisit*). Measures of initiation capture whether the participant started with the JOSM editor or iD, whether they abandoned their session within the first 30 minutes, and whether they completed at least one task. A contributor is considered activated if they contribute to any HOT project in the subsequent 7 days following the mapathon event. A revisit takes place when a contributor attends a subsequent mapathon in the 45 days following their first attendance. The full set of newcomer features is listed in Table 3.

We further compare longer-term *retention* across all five cohorts by means of a survival analysis. We observed contribution activity by HOT newcomers after

their first attendance for a period of 90 days to identify the last known moment of contribution. We considered contributors 'dead' if they had been inactive for at least 45 days by the end of this survival period. The last known date of contribution before that point marks their 'death event'.

Our analysis of potential causal factors for RQ2 makes additional use of the observational data collected at monthly and corporate mapathons as described in Section 4.4. It seeks to explain the newcomer activation and retention measures listed in Table 3 by considering the event setting, attendee mix, and attendee tool choice as shown in Table 2. All analyses were computed on a per-user basis, first with a pairwise Spearman correlation, and finally as a logistic regression model to explain the particular outcomes for all first-time attendees. In regression models we further included aggregate outcome measures as control variables: the share of attendees who have been successfully initiated, activated, or retained at each event (*initiation_rate*, *activation_rate*, *retention_rate*). Before analysis we standardised numerical features using z-scores, so that all variables have a mean of 0 and a standard deviation of 1.

Phase	Variable	Description
Initiation	*josm*	Started with JOSM?
	abandoned	Active for less than 30 minutes?
	completed	Submitted at least one task?
Activation	$active_{7d}$	Active in the first week?
Revisit	*revisit*	Repeat mapathon attendance?

Table 3. Attendee features computed per mapathon newcomer.

5 Results

5.1 RQ1: Newcomer Retention

Activation and revisit rates across the 14 mapathons are shown in Table 4. On average, at monthly mapathons 11.9% of newcomers are activated in the following 7 days, and 4.6% attend a subsequent mapathon. In comparison, at corporate events on average only 3.6% newcomers are activated and 2.0% revisit. These numbers indicate that the two mapathon cohorts have markedly different retention profiles.

To confirm this we computed survival functions for each cohort based on a Kaplan-Meier estimate with a 98% confidence interval. A corresponding survival plot is shown in Figure 1, this also includes the two online cohorts. The monthly mapathon cohort had the highest predicted retention rates, with a 20% chance of newcomer survival after 28 days. In contrast to this, the corporate mapathon cohort had the lowest retention rates, with a near-zero likelihood of survival

in the same amount of time. In comparison, the matched online cohort had a survival rate of 6%, and the Nepal online cohort 2% after 28 days. A pairwise logrank test confirmed that the four cohorts have distinct survival distributions ($p < 0.001$).

Compared to online cohorts, the Arup Mappy Hour cohort has a significantly higher retention rate, as the survival plot in Figure 2 shows. Retention of this group is comparable to the highly engaged monthly mapathon group: after 28 days, almost 25% of first-time contributors are still actively contributing to HOT.

Date	Type of event	% activated	% revisits
2014-11-24	Monthly	13.5%	0.0%
2014-12-15	Monthly	8.3%	8.3%
2015-01-27	Monthly	12.5%	6.3%
2015-02-12	Corporate	6.8%	0.0%
2015-02-24	Monthly	16.0%	4.0%
2015-03-31	Monthly	3.4%	3.4%
2015-04-28	Monthly	5.3%	0.0%
2015-05-15	Corporate	1.7%	2.3%
2015-06-02	Monthly	50.0%	16.7%
2015-07-07	Monthly	6.7%	6.7%
2015-08-04	Monthly	14.3%	6.1%
2015-09-01	Monthly	20.0%	6.7%
2015-10-06	Corporate	10.7%	3.6%
2015-10-06	Monthly	8.3%	4.2%

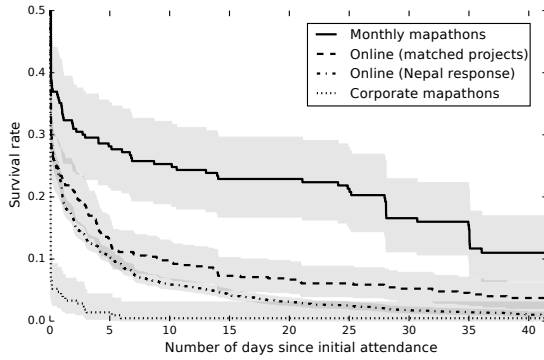**Table 4.** Activation and repeat attendance rates among first-time mapathon attendees.

Fig. 1. Newcomer survival rate for mapathon and online cohorts, with 98% confidence intervals.
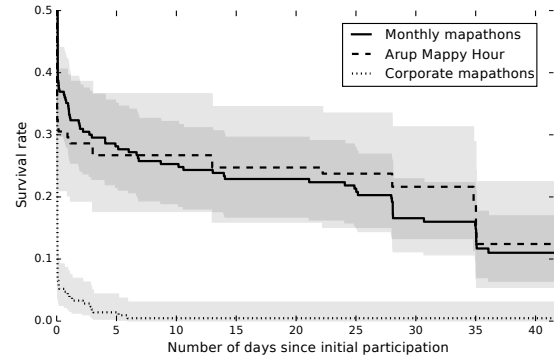

Fig. 2. Newcomer survival rate for mapathon and Arup Mappy Hour cohorts.

5.2 RQ2: Newcomer Retention Factors

Table 5 shows significant associations between mapathon features and event outcomes, as identified with a pairwise correlation analysis. According to these results, only two of the hypothesised mapathon features have a significant link to participant activity during the event. First-time mappers at monthly mapathons with a higher share of first-time JOSM users were more likely to complete at least one task ($\rho_S = 0.18$, $p < 0.01$). On the other hand, technical problems during a monthly mapathon negatively affected newcomer task submission ($\rho_S = -0.16$, $p < 0.02$). In neither case was there a significant association to subsequent activation or retention outcomes. In other words, during monthly mapathons, technical problems affected performance during the event, but they did not necessarily carry any longer-term engagement effects. Only one feature was significantly associated with a subsequent activation outcome: newcomers who contributed at least one task were more likely to contribute on at least one more day in the following week ($\rho_S = 0.16$, $p < 0.02$). Pairwise correlation found no significant associations for newcomer retention.

However it is not certain that these associations are indicators of real effects. We sought to assess the relationship of the effects by means of logistic regression models. Our models included the observations for both monthly and corporate events, taking into account all mapathon features and engagement outcomes. A model to explain activation of individual newcomers confirmed the relationship between task completion and activation, however had a bad model fit ($pseudoR^2 = 0.146$), and none of the other parameters were found significant. A further logistic regression model to explain newcomer repeat attendance had bad model fit, with no regression parameters found significant.

Aspect	Variable	Outcome	ρ_S	p
Setting	*tech_issues*	*completed*	-0.16	< 0.02
Tool use	*josm_learners*	*completed*	0.18	< 0.01
Initiation	*completed*	*active_{7d}*	0.16	< 0.02

Table 5. Significant correlations between mapathon features and event outcomes, for monthly mapathons only.

6 Discussion and Implications

The monthly mapathon format appears to be working well, many attendees were retained for longer periods. On average, 50% of monthly mapathon attendees were repeat visitors. Around 10% of first-time attendees subsequently mapped at home in the first week, and a similar proportion returned to a future event. Retention rates of Arup Mappy Hour participants were similarly high. However

there are clear differences across the remaining cohorts. Newcomers at corporate mapathons contributed to the same projects, but were rarely retained. Only few attendees at these events were activated in the first week. Similarly, online participants who contributed to the same projects were approximately three times less likely to be retained for future work than the monthly cohort.

A potentially important difference between the settings is the frequency at which social events are held. The two cohorts with the highest retention rates hosted regular events, the remaining cohorts were one-offs or online cohorts. The survival curves of the top cohorts show drops in retention after 28 and 35 days, visible as steps in Figure 1 and Figure 2, which would correspond to a monthly event frequency. This suggests that such future events may foster continued newcomer participation. However our observational data does not allow us to isolate event frequency from other factors, such as the attendee mix.

Furthermore, event frequency alone does not account for the fact that a difference in newcomer activation rates is immediately apparent within the first days after each event: activity levels for corporate and online cohorts already drop within the first few days. The reasons for this consistent difference across cohorts are currently unclear.

In addition to these effects, our results provide empirical evidence that sustained engagement relates to factors of the individual. People who managed to complete at least one task during their first mapathon were more likely to remain engaged in future activities. This effect was found regardless of the setting.

6.1 Implications

Based on this research we make the following recommendations to organisers:

- The cohort that held regular monthly events has the highest newcomer retention. Based on this, we suggest that organisers of other HOT groups provide regular opportunities for existing and latent enthusiast contributors to come together. This is difficult to achieve at large scale, however there may be opportunities to provide similar social experiences in an online setting.
- In particular, we suggest to experiment with forms of online support that imitate the mapathon experience: expert guidance, peer support, the presence of a community of practice, and other aspects.
- The most engaged mapathon cohort also had the most diverse mix of attendees in terms of prior experience, from newcomers to highly experienced mappers. It is feasible that newcomers can be swayed by others' enthusiasm, even if they are selected from a cohort that is unlikely to be retained. We recommend to experiment with the attendee mix, for example by organising events where corporate and monthly mapathon groups come together.
- Corporate mapathons has the lowest newcomer retention, although their event format was comparable to other mapathons. This difference in outcome may be related to the recruiting strategy: some participant groups may be more likely to become engaged mappers that others. We recommend to carefully evaluate the outcomes of such recruiting efforts, and to focus on groups that are more likely to be interested in sustained participation.

7 Conclusion and Future Work

A growing body of evidence suggests that sustained contributor engagement also has to do with aspects of the participation context and attendee selection, rather than just specific details of the setting and contribution process. In particular, there is evidence for the importance of regular social events and the presence of an existing community of practice.

Findings from this study can form a basis on which organisers can design interventions. Introducing newcomers to an existing and active community of practice may have a longer lasting effect than just the demonstration of the contribution process alone. Furthermore, instead of aspiring for indiscriminate growth, it may be advisable to identify prospective contributors who already have a propensity for the practice, and who may already be embedded with existing contributor communities.

Our study captures the presence of potential socialisation triggers at each mapathon, but we do not determine whether people actually made use of them. Attendee surveys could identify which factors were involved in the decision to remain engaged. Studies could further test for a match between participant motivations and project needs, prior experience with the practice outside of an OSM context (such as GIS experience), the presence of social ties, and related motivational factors.

Further research is needed to identify other unobserved factors which can help improve our model fit. For example, the work could be augmented with a comparative study of mapathons in other cities, and organised by different teams. Such a study could seek to confirm the observed relationships between attendee selection, event setting, and subsequent newcomer retention. Additionally it could seek to observe differences in recruiting and organising practices that were not captured by our London-focused study. At the time of writing, HOT enthusiasts are organising mapathons in a growing number of cities around the world, and there is similar growth in informal mapping groups at universities and other institutions.

References

1. Baecker, R.M.: Readings in Human-Computer Interaction: toward the year 2000. Morgan Kaufmann (2014)
2. Budhathoki, N.R.: Participants' motivations to contribute geographic information in an online community. Ph.D. thesis, University of Illinois (2010)
3. Burke, M., Marlow, C., Lento, T.: Feed me: motivating newcomer contribution in social network sites. In: Proc. SIGCHI '09. pp. 945–954. ACM (2009)
4. Choi, B., Alexander, K., Kraut, R.E., Levine, J.M.: Socialization tactics in wikipedia and their effects. In: Proc. CSCW '10. pp. 107–116. ACM (2010)
5. Ciampaglia, G.L., Taraborelli, D., Francisco, S.: MoodBar: increasing new user retention in Wikipedia through lightweight socialization. In: Proc. CSCW '15. pp. 734–742 (2015)

6. Clary, E.G., Snyder, M., Ridge, R.D., Copeland, J., Stukas, A.A., Haugen, J., Miene, P.: Understanding and assessing the motivations of volunteers: a functional approach. Journal of personality and social psychology 74(6), 1516 (1998)
7. Dittus, M., Quattrone, G., Capra, L.: Analysing volunteer engagement in humanitarian mapping: building contributor communities at large scale. In: Proc. CSCW '16. pp. 107–116. ACM (2016)
8. Fang, Y., Neufeld, D.: Understanding sustained participation in open source software projects. Journal of Management Information Systems 25(4), 9–50 (2009)
9. Farzan, R., Kraut, R., Pal, A., Konstan, J.: Socializing volunteers in an online community: a field experiment. In: Proc. CSCW '12. pp. 325–334. ACM (2012)
10. Haythornthwaite, C., Kazmer, M.M., Robins, J., Shoemaker, S.: Community development among distance learners: Temporal and technological dimensions. Journal of Computer-Mediated Communication 6(1), 0–0 (2000)
11. Hristova, D., Quattrone, G., Mashhadi, A., Capra, L.: The life of the party: impact of social mapping in OpenStreetMap. In: Proc. ICWSM '13. pp. 234–243 (2013)
12. Johansen, R.: Groupware: Computer support for business teams. The Free Press (1988)
13. Lave, J., Wenger, E.: Situated learning: Legitimate peripheral participation. Cambridge university press (1991)
14. Nov, O.: What motivates wikipedians? Communications of the ACM 50(11), 60–64 (2007)
15. Perkins, C., Dodge, M.: The potential of user-generated cartography: a case study of the OpenStreetMap project and Mapchester mapping party. North West Geography 8(1) (2008)
16. Schervish, P.G., Havens, J.J.: Social participation and charitable giving: a multivariate analysis. Voluntas: International Journal of Voluntary and Nonprofit Organizations 8(3), 235–260 (1997)
17. Steinmacher, I., Gerosa, M.A., Redmiles, D.: Attracting, onboarding, and retaining newcomer developers in open source software projects. In: Workshop on Global Software Development in a CSCW Perspective (2014)
18. Trainer, E.H., Chaihirunkarn, C., Kalyanasundaram, A., Herbsleb, J.D.: Community code engagements: summer of code & hackathons for community building in scientific software. In: Proceedings of the 18th International Conference on Supporting Group Work. pp. 111–121. ACM (2014)
19. Wilson, J.: Volunteering. Annual review of sociology pp. 215–240 (2000)