Analysing Volunteer Engagement in Humanitarian Crowdmapping

Martin Sebastian Dittus

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Engineering
of
University College London.

Department of Computer Science
University College London

September 15, 2017

I, Martin Sebastian Dittus, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Organisers of large crowdsourcing initiatives need to consider how to produce outcomes, but also how to build volunteer capacity. Central concerns include the impact of the first-time contributor experience, and the interplay of different modes of participation in larger organisations that host multiple strands of activity. How can volunteer capacity be built proactively, so that trained volunteers are available when needed? How important are opportunities for social encounter, either online or in person?

We present four empirical studies of the Humanitarian OpenStreetMap Team (HOT), a novel setting where thousands of volunteers produce maps to support humanitarian aid. Its diversity of settings and activities provides an opportunity to observe the effects of different coordination practices within a single organisation. Participation is online and open to all, however volunteers need to learn specialist tools and workflows. To support newcomers, HOT organises offline events to learn the practice under expert guidance.

Our research is motivated by a dual aim: first, to produce empirical evaluations of novel practices, informed by existing community concerns. Second, to revisit existing theories in social and behavioural science through the lens of this novel setting. We use statistical methods to observe the activity and retention of HOT volunteers. The full HOT contribution history is our primary source of empirical evidence, covering multiple years of activity.

We can demonstrate that coordination practices have a marked impact on contributor retention. Complex task designs can be a deterrent, while social contribution settings and peer feedback are associated with a significant increase in newcomer retention. We further find that event-centric campaigns can be significant recruiting and reactivation events, however that this is not guaranteed. Our analyt-

ical methods provide a means of interpreting key differences in outcomes. We relate our findings to comparable settings, and close with a discussion of the theoretical and practical implications.

Acknowledgements

This work would not have been possible without the encouragement, support, and wisdom of my supervisor Licia Capra. She is a careful and attentive listener, enjoying the process of shared exploration, and her admirable disposition will stay with me for a lifetime. Always ambitious and pushing further, working at a great pace, yet never aggressive or confrontational. Instead patient, constructive, helpful, suggestive rather than imposing; and always looking out for her student. Never assuming that her personal preferences should have any bearing on the matter, instead teaching how to carefully structure your choices.

I am particularly grateful to Andrew Braye and Pete Masters who were my early guides to the Missing Maps and HOT communities, who invited me to team meetings and organiser sessions and generally gave me much access, and who always provided invaluable feedback and support. It has been a great privilege and a great pleasure.

I thank the HOT community for their warm support throughout the research. Nick Allen and Ralph Aytoun who introduced me to the practices of the HOT training and validation communities, but also Ivan Gayton, Dale Kunce, and countless others who offered advice and support.

I would further like to express my appreciation to Yvonne Rogers, Alyssa Wright, João Porto de Albuquerque, and countless others who provided transformative advice in pivotal moments, who helped to develop a nuanced understanding of a complex domain, and to elevate the research in other important ways.

The passions of many individuals helped shape this work, and turned it into a remarkable journey. It wouldn't have been the same without you.

Impact Statement

Overall, the work has already achieved significant impact. Three of the studies have been published in major academic venues, and a fourth publication is forthcoming. One study has been awarded Honorable Mention at a top-tier conference, in recognition of the quality of the work, and of its significance for the field.[1] In addition, over time the work has lead to significant debate and reflection within the HOT community, and on more than one occasion it has already informed specific changes in organiser practice.

Early outcomes of this research were first presented at the inaugural HOT Summit in 2015, the first global gathering of the HOT community. The talk sparked great debate, and participants encouraged me to share findings in a more public manner. Since then, I have maintained a public research diary to share emergent findings with the community.[2] The posts are well-received, and frequently lead to discussions within the community.

The evaluation of HOT task designs in Chapter 4 has informed how HOT organisers structure new projects: organisers simplified previously complex tasks, and make sure to provide more explicit context and motivation for the work. The study of mapathon participation in Chapter 5 has further informed HOT practice. Organisers now place a greater focus on regular public community-building events, rather than private one-off events for employees at partner organisations. These have high overhead, but yield little sustained engagement.

In 2016, Missing Maps organisers published MapSwipe, a micro-tasking mapping application for mobile phones.[3] The application was in part informed by this research. It emerged out of an observation that many HOT contributors are likely

[1] https://uclic.ucl.ac.uk/news-events-seminars/icri-receives-honorable-mention-at-cscw-2017
[2] https://www.openstreetmap.org/user/dekstop/diary

to drop out within the first hour, which may negatively affect contribution quality: people may not stay long enough to learn how to map well. Organisers sought to provide a means of participating in HOT that could yield more useful contributions in such a limited time. Inspired by the dating app Tinder, MapSwipe implements a simple gesture-based workflow that allows volunteers to meaningfully contribute in 15 minutes without any prior training. Within a year, it has already received contributions by more than 15,000 people.

[3]http://mapswipe.org

Contents

List of Figures

List of Tables

Chapter 1

Introduction

On Saturday, 25th April 2015, a large earthquake hit Nepal with a magnitude of 7.8, followed by days of aftershocks. The event lead to the death of thousands, many others were displaced from their homes, and many were in need of shelter, medical help, food, and other aid (OpenStreetMap, 2016f). Within hours, aid organisations started providing emergency relief, as they had done before in other regions of the world. However in this instance, they were supported by an unusual ally: many of their maps were provided by the Humanitarian OpenStreetMap Team (HOT), a crowdsourcing initiative with thousands of volunteer contributors (Meyer, 2016).

On the following Saturday, Sky News presenter Martin Stanford introduces his TV audience to HOT's unusual model of digital volunteering. "*When disaster hits, as it did in Nepal last week, most people's first instinct is to see what they can do to help. Of course aid agencies have asked people not to travel to the country for their own safety, but an online army of volunteers have been supporting efforts without leaving their front rooms*" (Stanford, 2015). Stanford invites Harry Wood, a member of the organisation's board, to explain the HOT process. Thousands of volunteers have produced detailed tracings of maps from satellite imagery, drawing roads and building outlines in great detail. This level of detail was needed by field operations teams on the ground in order to deliver supplies, to plan routes, to share printed maps, or for more advanced geospatial analysis of the data. Wood emphasises that many of the mappers were new to the practice: "*it's a kind of volunteering which anyone can get involved in.*" Within the first week, nearly 4,000 people had mapped for Nepal, of which more than half were first-time contributors (Stanford, 2015).

HOT can be considered an instance of crowdmapping: a crowdsourcing initiative that specifically aims to produce digital maps, often in response to disaster events (Shahid and Elbanna, 2015). It was founded in 2010, and has its origins in an informal network of experts and organisers of the OpenStreetMap community. Its map data is now in increasing demand by a growing number of organisations, and the organisation has gradually refined the necessary processes and technologies that allow it to scale (Soden and Palen, 2014). HOT has coordinated responses to typhoon Haiyan in the Philippines, the West African Ebola crisis, the Nepal earthquake, and many others (HOT, 2017). In addition, it hosts long-term mapping initiatives that are focused on disaster preparedness. In late 2014, several partnering organisations launched a "Missing Maps" initiative where the focus is on proactive mapping, to ensure places are already well-documented before a crisis hits (MM, 2017). Participation in HOT can take different forms: in some cases, regional groups host so-called mapathons to come together in a more social setting, but many contributors simply participate online. More experienced mappers may take part in data validation, reviewing the work of others and providing feedback. (Refer to Section 2.1 on page 29 for a discussion of HOT origins, concepts, and practices.)

In a wider sense, HOT can also be understood as a form of computer-supported mass-volunteerism, a novel phenomenon where large numbers of participants contribute to specific shared outcomes, and where these outcomes have significant societal benefit. Its relatively low bar to entry supports episodic forms of participation, and the commitment of contributors is comparably low, which in principle can allow a larger number of people to become transient participants. In this, HOT represents a glimpse of a potential future for a kind of digital volunteerism that has an immediate and measurable global impact, based on distributed local action, in an impactful arrangement where growing humanitarian need is met with an increase in technical and human capacity.

The promise of its novel initiatives is to achieve a greater capacity for disaster response and humanitarian aid by incorporating the help of a global online volunteer force (Lüge, 2014). However, mapping all the undocumented and crisis-stricken regions of the world is a formidable task: even after many months of work by tens of thousands of volunteers, many places in the world remain unmapped.

A project briefing by Médecins Sans Frontières (MSF) illustrates the scale of this capacity-building challenge: "*To reach our goal, we need the Missing Maps Project to be the biggest instance of digital volunteerism the world has ever seen*" (MSF, 2014). Organisers thus not only need to consider how to produce these maps, but also how to foster a large global volunteer community in the process.

Our research is in part informed by a desire to support HOT's mission, and to contribute to its positive societal impact. We further believe that the theoretical and practical concerns of HOT are likely to recur in future mass-volunteering platforms, and that this makes it a good moment to prepare some conceptual foundations.

1.1 Research Questions

Our research seeks to develop a better understanding of how to best increase volunteer capacity for HOT and comparable volunteering platforms. The purpose of the research is to identify practices that allow HOT and similar platforms to grow volunteer capacity at large scale, and to sustain volunteer activity over longer periods. A particular focus is placed on factors that affect early newcomer retention, including the identification of barriers to engagement in current HOT practice.

The research is motivated by two broad research questions. These are contextualised and refined across four studies, each focused on a particular aspect of the setting. The questions are informed by specific practical concerns in the setting, but also by theoretical concerns that are drawn from a wide range of related literature.

What coordination practices can help build volunteer capacity? A major part of the research aims to discover specific practices that allow HOT to grow. Organisers of large-scale crowdsourcing initiatives need to consider how to produce outcomes with their projects, but also how to build volunteer capacity. The initial project experience of volunteers plays an important role in this, particularly when the contribution process requires some degree of expertise. For example, overly complex task instructions can demotivate some contributors, and lead to early abandonment. Informed by the literature, we seek to identify specific aspects that may improve early newcomer engagement, and foster more sustained participation.

Does the nature of disaster response foster distinct participation behaviours? For example, are disaster events significant recruiting moments? Do they reactivate

volunteers that have previously become inactive? To what extent does this influence campaign outcomes? Arguably characteristic for HOT is the coexistence of both urgent emergency response campaigns and more long-term campaigns without a specific deadline. This provides an opportunity for a comparative study to better understand participation behaviours in each setting.

1.2 HOT as a Novel Setting for Research

As a literature review in Chapter 2 demonstrates, to date there are only few in-depth studies of HOT practices and outcomes. Yet we observe that HOT is a promising setting for research in a number of specific aspects, and its study allows us to produce insights with wide-ranging implications for both the practice and theory of mass-participation platforms. The research presented here forms an important early step in this respect.

From a theoretical perspective, HOT represents an opportunity to study a novel arrangement of familiar structural aspects. In its participation process, it has similarities to existing crowdsourcing and peer production platforms, but also some distinct differences. In its approach to coordinating large numbers of volunteers, it is also comparable to existing citizen science or crowd work platforms. Yet the colocated practice of HOT mapathons, and the event-centric and urgent nature of many of its activities are atypical for many such prior platforms. (For a detailed discussion of these and related concepts, refer to a discussion of HOT practices and related literature in Chapter 2 on page 29.)

HOT is also an attractive research setting because of its great organisational diversity. Its range of initiatives and organisational practices offers many opportunities to compare and evaluate specific organiser choices. Different partner organisations coordinate their own mapping activities on the platform, including volunteer groups by the British and American Red Cross, Médecins Sans Frontières, the Peace Corps, and others. HOT projects can be compared in terms of their purpose, degree of urgency, task design, and other respects. Because of the co-located practice of HOT mapathons, it is possible to observe instances of both online and offline volunteering within the same platform.

Maybe most importantly, HOT is a promising setting for research because it

provides a unique source of readily available observational data. All its map contributions are captured in the OpenStreetMap edit history, a perfect contribution record that is available to the public as an open data set. Further metadata is captured by the HOT Tasking Manager, a coordination platform that provides contextual information for all HOT projects. Together they provide a rich set of annotated empirical evidence which can be used for large-scale observational studies. At the time of writing, the contribution history captures the work of tens of thousands of participants over a multi-year period, across more than 1,000 discrete projects, and dozens of larger initiatives. (Our sources of empirical evidence are described in Section 3.2 on page 55.)

1.3 Methodology

We present a broad programme of study of the specific and generalisable concern of volunteer engagement, based on a research setting that is uniquely amenable to comparative evaluation. Our research takes the form of four large-scale quantitative and observational studies, and primarily relies on computational and statistical methods. The primary source of evidence is the HOT contribution history. In certain instances, we augment it with further metadata where this can enhance our interpretation of outcomes. Our analytical dimensions are grounded in existing theory, and yield operational insights from this readily available public data. This can be considered a forensic approach to researching a socio-technical system: the careful compilation of empirical evidence to uncover underlying processes. It serves as a minimum-effort complement to more invasive evaluation practices such as controlled experiments, A/B tests, and participant observations. The approach was chosen for its ability to support large-scale and long-term observation with comparatively little effort. Our methods allow us to evaluate the outcomes of a wide range of practices in a repeatable and reproducible manner. Furthermore, this process is transferrable to other settings: a minimum requirement is a capacity to observe individual contributions over time. (We provide a theoretical justification for our overall methodology in Chapter 3 on page 51.)

To augment this forensic approach, we spent significant effort preparing each study, taking care to contextualise our observations, identifying and assessing a

wide range of contributing factors, and refining our interpretation of outcomes. This preparation took the form of an iterative process of discovery, starting with early data explorations. It served to reduce a bias of preconceived notions, and open a capacity for new discoveries. We further spent significant time building domain understanding through discussions and interactions with the HOT community, by sharing early evidence, and identifying practitioner concerns in a participatory manner. These participatory–embedded exchanges were not part of the formal research process, however they served to direct our research in important ways. They provided us with an important feedback loop with practitioners, and a means to improve our domain understanding.

As a consequence, our overall methodology can be considered both exploratory-inductive and theory-oriented, with computational and statistical methods at the centre. It represents one possible approach among many, and our focus on quantitative observational methods comes with some limitations. In comparison to fully ethnographic approaches, our findings can lack a detailed understanding of the nuanced and contextual relationships between contributing factors. In comparison to experiment-based approaches, we cannot claim to have identified causal factors. Instead, we are limited to observing the presence or absence of specific associations. In comparison to these, a strength of our chosen approach is its capability to analyse past outcomes at a very large scale.

1.4 Research Contributions

We present four large-scale quantitative studies, covering a broad range of generalisable volunteer engagement concerns through the lens of HOT. The research was executed in the pursuit of a dual aim: to address open questions about HOT practice, but also to capture broader research opportunities offered by the novel setting. The work achieves immediate impact for HOT through empirical evaluations of its novel practices, but also by establishing links between HOT outcomes and existing theoretical concepts. Additionally, it involves more foundational research by revisiting an existing corpus of social and behavioural science through the lens of this novel setting, for example to reproduce prior observations from comparable settings, and to assess whether particular differences in coordination practice yield

differences in outcome. In at least one instance, we introduce new theory to the existing body of knowledge about mass-participation platforms.

In an initial study, we observe how the contribution process itself may affect early participation. How can organisers foster a large community as well as produce outputs? We find that task complexity can inhibit participation by some contributor groups. In addition, we find that newcomers who join during a one-off disaster event are less likely to be retained than those recruited during more long-running campaigns. We further find evidence that participation in mapathons may be linked to more sustained engagement.

In a second study we investigate this latter relationship further, comparing the participation histories of different newcomer cohorts across a range of participation settings. Are mapathons an effective means of building volunteer capacity, even in the absence of urgent causes? We confirm that participation in specific social contribution environments is linked to a significant increase in newcomer retention, however that this is not true for all settings. Our observational methods allow us to identify several contributing factors. Based on these findings, we propose opportunities for further inquiry.

In a third study, we observe the effects of private peer feedback given during a map validation process. We find that certain of these interactions between contributors are linked to a significant increase in newcomer retention. In particular, we observe that such peer feedback may provide an important form of social affirmation, communicating to newcomers that their participation is welcome and appreciated. We do not find that negative performance feedback is harmful to newcomer retention, as is found on platforms where peer feedback is publicly negotiated.

In a fourth study, we compare participation patterns across urgent disaster campaigns and more sustained mission-centric campaigns. We find that HOT in its current form is less characterised as an on-demand task force that springs to action when it is needed, and instead an ongoing stream of activity that is redirected when new needs arise. In a major theoretical contribution we further provide early evidence for a media effect, where public interest in particular topics can affect participation in a manner that significantly affects outcomes. Based on our findings, we recommend that organisers introduce mechanisms to better manage contribu-

tor flows during times of need. This includes the means to reactivate experienced contributors when larger disaster campaigns are launched, but also means to redirect newcomer flows towards more beginner-friendly projects, in particular when public interest is high and data quality is a concern.

1.4.1 Publications

Dittus, M., Quattrone, G., and Capra, L. (2016a). Analysing volunteer engagement in humanitarian mapping: building contributor communities at large scale. In *Proc. CSCW '16*, pages 108–118

Dittus, M., Quattrone, G., and Capra, L. (2016b). Social contribution settings and newcomer retention in humanitarian crowd mapping. In *Proc. SocInfo 2016*, pages 179–193. Springer

Dittus, M., Quattrone, G., and Capra, L. (2017). Mass participation during emergency response: event-centric crowd-sourcing in humanitarian mapping. In *Proc. CSCW '17*, pages 1290–1303

Dittus, M. and Capra, L. (2018). Private peer feedback as engagement driver in humanitarian mapping. In *Proc. CSCW '18*. (Conditionally accepted, expecting publication in November 2017)

1.5 Thesis Structure

A review of HOT concepts and practices is presented in Chapter 2, this also includes an overview of HOT studies to date, and a discussion of the theoretical foundations of contributor engagement in mass-participation platforms. Our general methodology is presented in Chapter 3. This is followed by individual chapters for the four studies in Chapters 4–7, each of which introduces additional theoretical and methodological concerns. We conclude with a summary of the findings and implications of the overall work in Chapter 8, including a critical reflection of our methodology, and recommendations for future work.

Chapter 2

Background

When reasoning about volunteer engagement on mass-participation platforms, it is necessary to understand the catalysts and barriers of sustained contribution activity. In the context of HOT, this includes coordination practices to build volunteer capacity to respond to sudden emergencies, but also to provide sustained proactive work in the absence of urgent causes. We begin by contextualising HOT within existing research discourse of mass-participation platforms, starting with a detailed account of the origins of HOT, and its concepts and practices. We also review existing studies of HOT, including discussions of its potential and limitations. We then present an overview of the theoretical foundations of our research: typologies of mass-participation platforms, and theories of participant engagement. A particular focus is placed on current knowledge about participant motivations. A final section summarises current knowledge gaps in the literature that relate to our research questions, and outlines current research opportunities. Based on this, our own research contributions are developed in the following chapter on page 51.

2.1 Humanitarian Mapping with HOT

HOT has its origins in OpenStreetMap (OSM), a peer production effort to produce a comprehensive map of the world. In many respects, OSM is comparable to Wikipedia: anyone can contribute to the map, and the map data is freely available to all under an open license. People can contribute with a web-based editor hosted on the OSM website, or with a range of mobile and offline editors that are

available for download. The platform's initial focus was on contributors mapping their local neighbourhoods, for example by importing paths that were captured on GPS recorders, and then refining and annotating the resulting traces. Today, mapping activity can have many different starting points, including the tracing of satellite imagery, or importing of geographical data that was placed in the public domain (Haklay and Weber, 2008; Haklay and Budhathoki, 2010).

2.1.1 A brief history

HOT was founded in 2010 by an informal network of GIS experts, technologists, and humanitarian aid workers. The founding group shared an interest in OSM, and wanted to explore its ability to address the geospatial information needs of aid organisations. In particular, they saw potential for a humanitarian geodata commons: a public geospatial database that was shared across aid organisations, built on the platforms and practices of OSM, with enhancements to satisfy the specific requirements of humanitarian aid work. Central to their ambition was an understanding that such a platform can benefit from the contributions of a global volunteer force. Since then, the organisation has gradually refined the necessary processes and technologies that allow it to scale (Soden and Palen, 2014).

The primary role of HOT is the collective creation and maintenance of maps for humanitarian purposes, typically in response to specific needs by aid organisations. To date, HOT organisers coordinated community responses to typhoon Haiyan in 2013, the West African Ebola crisis in 2014, the 2015 Nepal earthquake, and many others (HOT, 2017). During such campaigns, HOT organisers set up projects to address particular information needs, and promote these to potential volunteers. Projects seek to map certain geographic features in a particular region, for example to map settlements so that aid experts can understand population distributions, or to trace roads so that field teams can plan transport routes. Larger HOT activations can consist of dozens of individual projects. The maps produced by HOT volunteers are free for all under a liberal license, and are now in use by experts at MSF, the American and British Red Cross, the World Health Organisation, and a growing number of other institutions (Clark, 2014).

A fundamental challenge for aid organisations is that large parts of the world

have never been mapped, neither by OSM or HOT, nor by other mapping providers. As a consequence, HOT organisers often struggle to keep up with the scale and speed of humanitarian disasters. This has direct impact on the effectiveness of field operations: "*A map that comes post-disaster doesn't save lives*" (Parshley, 2017). To help address this, in late 2014 HOT and several partnering organisations launched a "Missing Maps" initiative with a focus on proactive mapping. Missing Maps seeks to map areas at risk, so that regions are already well-documented by the time a crisis hits (MM, 2017). In comparison to other emergency response activities these are typically more sustained campaigns, covering vast regions over multiple weeks or months. The resulting maps are often the first ever of their kind. Other long-running campaigns are focused on specific geographic regions such as the Congo, Central Africa, Lesotho, or South Sudan (OpenStreetMap, 2016i), or they are umbrella initiatives orchestrated by organisations such as the Peace Corps and MapGive (Michael, 2014).

The promise of HOT and its related initiatives is to achieve a greater volunteer capacity for disaster response and humanitarian aid by incorporating the help of a global online volunteer force. MSF staff member Ivan Gayton summarises one particular appeal of HOT to aid organisations: "*Finally, I can give volunteers something to do that isn't just giving money*" (Michael, 2014). In principle anyone can contribute, and contributors choose their own tasks. A 2014 report by MSF discusses the impact such initiatives can have on the work of aid organisations: "*Many interviewees commented that they were 'amazed' by the speed at which the area was mapped with the help of the volunteers. On his own, the GIS officer would not have been able to produce these base maps during his mission.*" (Lüge, 2014)

2.1.2 Remote Mapping with the Tasking Manager

The first step in the creation of a new HOT map is a **remote mapping** stage, involving the help of dozens to thousands of volunteers. During this remote mapping process, HOT volunteers trace maps from satellite imagery of remote places.

The HOT **Tasking Manager** is a key technology in the contribution process, it emerged out of a need to distribute work across large numbers of remote mappers while reducing edit conflicts (Palen et al., 2015). On the tasking manager, work is

organised into projects, each aspiring to map specific features in a particular region of the world (see Figure 2.1). Multiple projects are often part of larger mapping campaigns, for example covering different areas within a larger region.

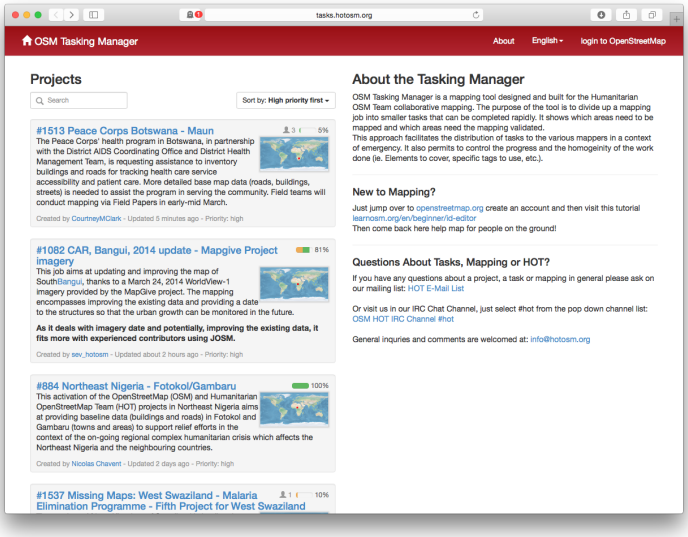

Figure 2.1: The homepage of the HOT Tasking Manager, showing a project listing.

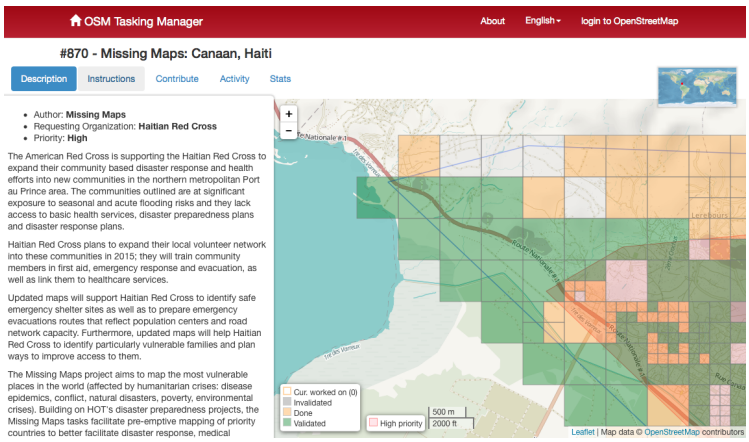

Figure 2.2: A project page in the Tasking Manager, showing a description on the left, and a map with a task grid on the right.[1]

Figure 2.2 shows a screenshot of a project and its description. To facilitate volunteer coordination, project areas are geographically divided into smaller map segments, so-called tasks. Contributors register for a task when they begin their work,

[1] http://tasks.hotosm.org/project/870

and then contribute according to a set of instructions. They mark their task as 'complete' when their work is done.

A typical HOT project may ask volunteers to trace a simple basemap of roads and buildings. These map objects are drawn as line geometries, so-called ways, and annotated with basic metadata to distinguish them. HOT contributors require an OSM account to contribute, and publish their contributions on OSM using map editing tools such as JOSM or iD (see Figure 2.3). Most of their time is spent in these editing tools rather than the Tasking Manager.

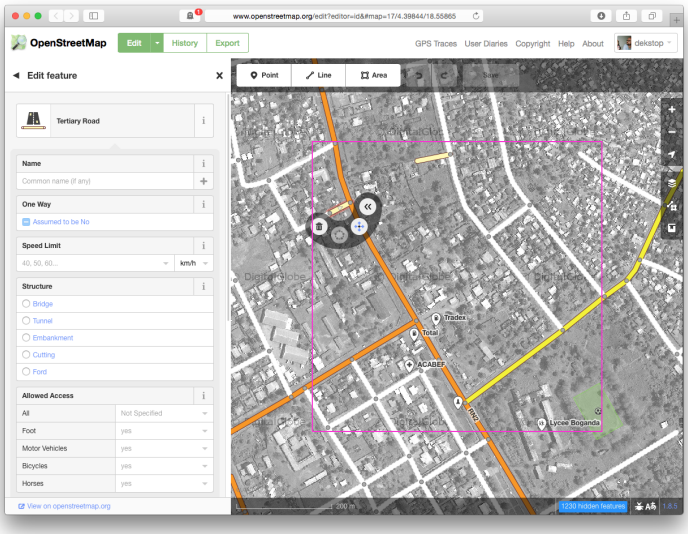

Figure 2.3: Using the web-based iD editor to trace map features from satellite imagery.[2]

Figure 2.4 shows an exemplary outcome of this work, a mapping campaign in rural Nepal after the 2015 earthquake. As the picture shows, in this case there was already some prior map data before the campaign started, however the work by HOT volunteers yielded a much more detailed picture.

Remote mapping predominantly takes place online, however in some cases regional volunteer groups host so-called **mapathons**, social event settings which allow regional community groups to come together in a more social setting, to learn the practice and to socialise. (Chapter 5 on page 83 includes a more detailed description of the format.)

[2]https://www.openstreetmap.org/edit
[3]Screenshot taken from http://osm-analytics.org

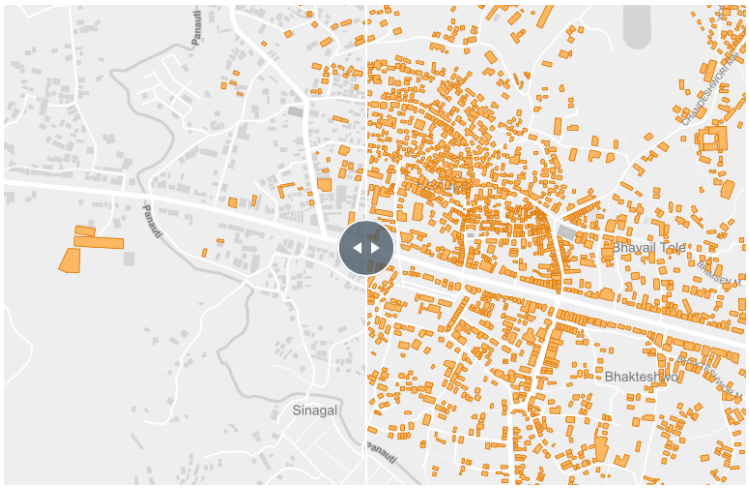

Figure 2.4: Mapped buildings before (left) and after (right) a HOT campaign in rural Nepal.[3]

2.1.3 Quality Control

To improve map quality, HOT has an internal peer review process in the form of a **validation** stage. During map validation, an experienced mapper reviews contributions to a task, makes changes as necessary, and either declares them 'valid' or 'invalid', for example because the work was incomplete. As part of this process, validators may adjust geometries, refine annotations, and delete map objects which were of an insufficient quality. They can also send private messages for the contributors of these submissions. Invalidated work can be picked up by other mappers, who may further refine any existing map data. At the time of writing, any HOT volunteer can validate any task, and there is no vetting process in place for prospective validators. The validation process is currently not standardised, however there are some general guidelines, and much community discourse about different approaches (OpenStreetMap, 2016k,j; Whelan, 2016; Owen, 2015; Reed, 2015).

For select projects, **field mappers** with knowledge of the local area may subsequently augment the maps by adding more detailed annotations, including place names and other details, and by making corrections based on ground observations (Sturgis, 2014). Projects of this nature can emerge out of initial discussions with affected communities about their local needs. An exemplary overview of such local mapping initiatives is provided by Patel (2015), who presents Missing Maps projects with local mappers in Zimbabwe, Tanzania, Rwanda, and South Africa. In

some cases, field mappers collect observational data on paper and in other forms, which are then shared back with remote volunteers who in turn digitise the information and publish it on OSM.

2.1.4 HOT Scale and Growth

HOT has experienced significant growth since its inception in 2010. To illustrate this growth and its current scale, we provide some basic aggregate statistics which were derived from the HOT contribution history. (Refer to Chapter 3 for a description of the process.)

By September 2016, 32,000 contributors have made 180 million map edits across 2,000 HOT projects, involving an estimated total of 240,000 hours of volunteer work. Figure 2.5 visualises the global distribution of all HOT edits. Activities are geographically focused on Central and West Africa, Southeast Asia, but also many other parts of the world.

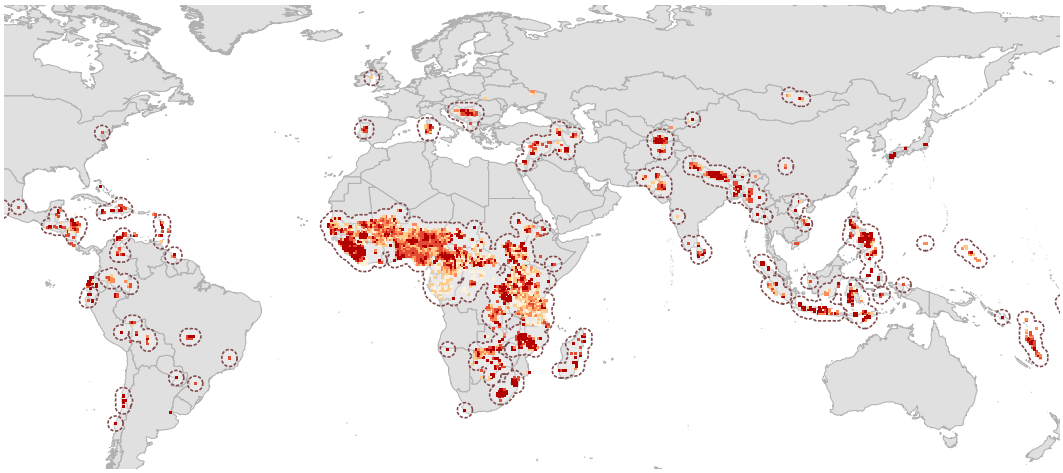

Figure 2.5: Geographic distribution of all HOT contributions.

The cumulative timeline in Figure 2.6 shows steady project growth since the Tasking Manager was launched. Projects were added at an increasing rate, particularly in the first years of activity. This may reflect an increase in needs or organiser capacity.

A corresponding community growth timeline in Figure 2.7 shows that the pool of contributors is growing at a less steady rate. Instead, growth appears to relate to particular disaster events. We marked four events on the timeline which coincide

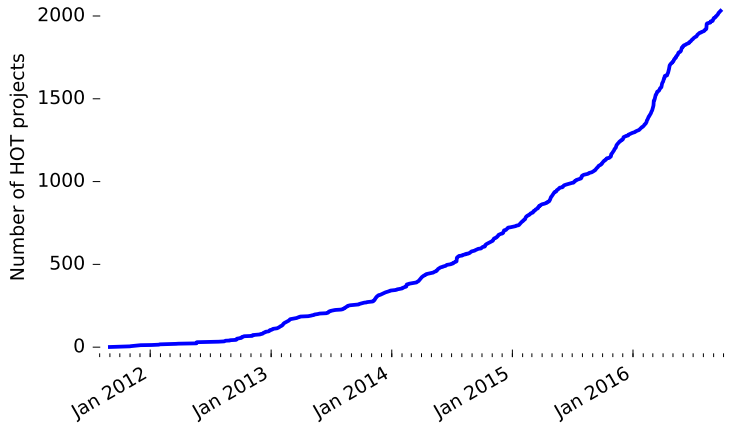

Figure 2.6: Growth of HOT Tasking Manager projects over time.

with particular moments of community growth: Typhoon Haiyan in late 2013, the Ebola response throughout much of 2014,[4] the Nepal earthquake in April 2015, and an earthquake in Ecuador in early 2016.

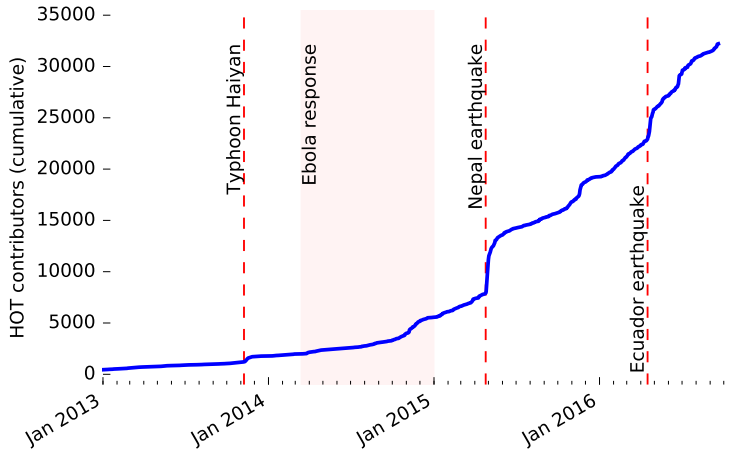

Figure 2.7: Growth of the HOT contributor community over time. The chart highlights three large disaster events and one prolonged emergency response period.

The contributor community is not necessarily active at all times, as is shown in Figure 2.8. Typically only a few dozen people are contributing to HOT projects on any given day, and many contributor accounts are dormant. However on certain days, community activity can involve hundreds of contributors.

[4]Increased community growth towards late 2014 coincides with the launch of Missing Maps in November 2014. Initially, many of the group's activities were focused on the Ebola response.

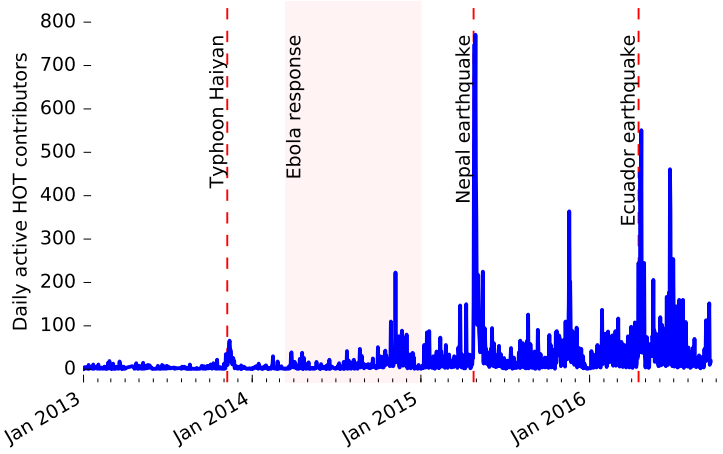

Figure 2.8: Daily number of active HOT contributors over time.

2.2 Studies of HOT

Studies of HOT published to date typically focus on specific HOT initiatives, often with a focus on general descriptions of project outcomes, for example by summarising the volume of data produced. A growing number of studies include assessments of the quality of the data produced for specific projects. To date, only a small number of empirical studies have looked at HOT in a more comprehensive manner across a range of initiatives, and there are only few in-depth assessments of HOT contributor engagement. Broadly speaking, the published research addresses two general concerns: a focus on project outcomes including contribution quality (Section 2.2.1), and a focus on participant engagement (Section 2.2.2).

2.2.1 HOT Campaign Outcomes

A growing body of published work seeks to assess the outcomes of HOT campaigns, with a focus on larger and more well-known event-centric campaigns: a large earthquake in Haiti (Soden and Palen, 2014; Palen et al., 2015), typhoon Haiyan/Yolanda in the Philippines (Palen et al., 2015; Westrope et al., 2014; Dittus et al., 2016a), and a 2015 earthquake in Nepal (Eckle and de Albuquerque, 2015; Anhorn et al., 2016; Poiani et al., 2016). However such studies are typically limited to evaluations of these individual campaigns.

The first prominent example of HOT mapping occurred in the aftermath of a

Haitian earthquake in 2010. In response to the earthquake, large numbers of remote volunteers organised a range of efforts to document the state of urban and rural infrastructure in an effort to help disaster response. Zook et al. (2010) summarise the key projects of that time, including those that involved OSM as a mapping platform. The authors conclude that "*crowdsourcing information and mapping services can greatly enhance the logistical systems upon which relief efforts are ultimately grounded*". In an early study of the outcomes, Roche et al. (2011) illustrate the real impact that OSM volunteer mapping efforts had on the ability of emergency response teams to provide help: "*The road network map of Port-au-Prince, which was almost blank on the evening of the 12th of January, was nearly complete 10 days later*".

Westrope et al. (2014) provide the first detailed quality assessment of one of the largest HOT initiatives, a damage assessment project following Typhoon Haiyan in the Philippines in late 2013. Participant mappers had been asked to document the level of damage to housing based on satellite images. The authors find that there was a large number of contributions, however they acknowledge that it was hard to properly assess building status from satellite pictures. OSM contributors "*overestimated the number of buildings completely destroyed by the typhoon and underestimated the number of buildings that were majorly damaged*". Westrope et al. propose a number of improvements to the process, for example to allow to visually compare the pre-disaster state as a point of reference, and the provision of guidance materials on both OSM and humanitarian damage assessment.

A small-scale experimental study by Eckle and de Albuquerque (2015) assesses the quality of HOT outputs produced by remote mappers for the Nepal region. According to their findings, the most frequent mistakes by remote mappers are caused by an unfamiliarity with local architecture and geography. Eckle and de Albuquerque propose to provide better guidance on how to interpret satellite imagery, but also to cooperate with local mappers who can validate the work and provide feedback.

A further limiting factor to making HOT maps useful are the existing habits of organisations that can benefit from the data. Lüge (2015) presents a detailed report of the impact of HOT mapping in the context of the Ebola epidemic in West Africa, and concludes: "*It is clear that better maps and other GIS products can have a profound*

positive impact on operations. But it is also clear that this very much depends on the GIS officers' skill to proactively communicate what he/she can offer to operations." In other words, convincing operational teams to use these maps instead of existing information sources is a challenge in itself. Shahid and Elbanna (2015) further observe a need to adopt operational processes to HOT practices: HOT maps are not merely an artefact that replaces existing maps, they are also accompanied with a new set of processes and relationships.

2.2.2 HOT Contributor Engagement

Conceptually, we can distinguish three entry points by which participants discover projects, be they first-time contributors or experienced mappers. In some cases, projects are highlighted in public discussion in media, social media, and other forms of external promotion (Meyer, 2013, 2016; Michael, 2014; Clark, 2014). High-profile disaster events may benefit from this in particular. In other cases, organisers may recruit participants through non-public channels, for example as part of partnerships with other organisations, or through the organisation of mapping events (Smith, 2015). Such recruiting channels can play an important role for campaigns focused on community-building, where organisers have a direct relationship with prospective contributors (Smith, 2015; Feinmann, 2014). Finally, contributors may start by browsing the publish project list with or without a particular intention.

It is not currently clear what motivates HOT participation, compared to participation in other platforms. Among early OSM contributors, it was found that an individual's local geographic knowledge was the most significant driver to contribute (Budhathoki, 2010). In comparison, it is conceivable that the global scope and perceived social benefit of HOT attracts different audiences, and fosters a different form of engagement compared to OSM.

In a detailed assessment of OSM contributions after typhoon Haiyan, it is suggested that media coverage may influence contributor turnout, which in turn may affect and mapping outcomes, however the specific relationship is not yet well understood (Westrope et al., 2014). The authors of the study observe that a region which was more frequently covered by news media had been mapped differently than other nearby regions: the contributions were of a markedly lower quality.

The largest HOT effort to date was launched after a large earthquake hit Nepal in spring 2015, it involved thousands of remote mappers. The rapid response by HOT and its contributors benefited from prior efforts to build a local OSM community who knew the local geography well, and from the presence of well-connected organisations who were informed about the earthquake risk and could coordinate mapping activities (Soden et al., 2014). A large amount of new mappers joined HOT to support the cause, and organisers relied on mailing lists and other online platforms to coordinate the work (Poiani et al., 2016). An evaluation of the Nepal campaign finds that more contributions were made by experienced mappers, however that first-time mappers provided small but important contributions, such as the creation of notes for missing information. The authors observe that further research is needed to verify if and how prior contributor experience affects data quality (Poiani et al., 2016).

An important barrier to entry for first-time HOT contributors is likely the fact that mapping with OSM is a complex practice: it requires specialist tools and an understanding of specialist concepts (Schmidt et al., 2013; Brown et al., 2013). To our knowledge, there is no published research on how this affects HOT contributor engagement. It is feasible that peer support during mapathons can provide necessary support for early HOT newcomers, and may increase early newcomer retention. To date, there are no empirical studies of such engagement effects, however prior studies of early OSM mapping parties can provide an early expectation. Mapping parties were similarly structured around social mapping experiences, but typically aimed to map the local area. Haklay and Weber (2008) suggest that mapping parties provided an important social setting for early OSM contributors, and may have had an effect on subsequent contributor engagement. However, two empirical studies of mapping parties in different cities found that such events were unsuccessful at retaining newcomers. Sustained activity after the event was mostly limited to contributors with at least some prior experience, and most first-time attendees stopped contributing after their first event (Hristova et al., 2013; Perkins and Dodge, 2008).

2.3 Theoretical Foundations

2.3.1 Typologies of Mass-Participation Platforms

Although there are only few studies specifically focused on HOT, in its coordination practices it can be compared to several other types of participatory and collaborative systems for which well-developed theory is available. Immediate reference points are the wider OpenStreetMap community (OSM), and similar participatory online communities such as Wikipedia (WP). These are considered instances of peer production and crowdsourcing systems.

Commons-based peer production, also called peer production or social production, is a concept first proposed by Benkler (2002) to describe a third model of production after firm production (production by companies and industry) and market-based production (market economies). Key attributes of peer production are a reliance on information-gathering and exchange as the key activity and output, a more decentralised form of organisation instead of the hierarchical forms of previous models, and the frequent absence of explicit financial compensation for contributors (Benkler, 2002, pp. 375f). Commons-based peer production can be more efficient than firm production and market-based production because knowledge of tasks and capabilities is distributed, and participants can both self-nominate and coordinate between others. Larger groups of individuals can employ larger pools of resources to address the problem at hand.

The term **crowdsourcing** was initially coined by Howe (2006) and has subsequently gained a number of different connotations and popular re-interpretations. An exhaustive literature review of crowdsourcing research is provided by Pedersen et al. (2013), who also provide a more recent definition of the term: "*A collaboration model enabled by people-centric web technologies to solve individual, organizational, and societal problems using a dynamically formed crowd of interested people who respond to an open call for participation*". Wiggins and Crowston (2010) add: "*Initially introduced as a novel alternative business model, more recent popular use of the term has applied it to any form of collective intelligence that draws on large numbers of participants through the internet.*" **Crowdmapping** is a form of crowdsourcing that specifically aims to produce digital maps, for example in response to disaster events (Sutter, 2010; Quaintance,

2014; Shahid and Elbanna, 2015). Crowdsourcing systems in which participants are explicitly motivated by payment and similar incentives are sometimes distinguished as **crowd work** (Kittur et al., 2013). The connotation of outsourced labour shared by these definitions reflects that in a crowdsourcing system, tasks are generally described by a central coordinating party which then procures labour for their fulfilment. In practice, the term is sometimes used more loosely. For example, WP is frequently referenced in crowdsourcing research, although it has no such central body that procures and coordinates labour.

The term **citizen science** is used to describe systems of communal knowledge production that are specifically modelled on notions of scientific practice, which often means they are centrally coordinated by a body of experts, similar to crowdsourcing systems. Wiggins and Crowston (2010) summarise: "*Citizen science is a form of organisation design for collaborative scientific research involving scientists and volunteers, for which internet-based modes of participation enable massive virtual collaboration by thousands of members of the public.*" They state that under the right circumstances, citizen science can not only produce high-quality data, it can also yield unexpected insight and innovations. Some forms of citizen science already involve citizen participants in the design phases of a project, which requires that scientists act as facilitators and coordinators as well as domain experts (Haklay, 2013).

Crisis mapping is a comparatively recent mass-participation phenomenon. Platforms such as HOT and Ushahidi invite large numbers of volunteers to collect geographic data in response to disasters, typically to improve situational awareness, or otherwise support the work of response organisations (Meier, 2011; Collins, 2011; Palen et al., 2010; Okolloh, 2009). HOT volunteers often contribute remotely, and help in the creation of a basemap from satellite imagery without specific awareness of the affected region (Soden and Palen, 2014; Palen et al., 2015). In the case of Ushahidi, volunteers provide geographically referenced situation reports, which are then placed on a public map (Okolloh, 2009).

Further relevant research relates to the concept of computer-supported collaborative work (CSCW, Wilson 1991), and of open source software teams, for example the study of their social structures and coordination practices (Crowston and Howison, 2005; Crowston and Scozzi, 2004). However, while open source software de-

velopment processes rely on expert knowledge by experienced contributors, HOT places no strong skill requirements on its participants, and barriers to entry are comparatively low.

2.3.2 Theories of Participant Engagement

The descriptions of these mass-participation models may already evoke a sense of the varied challenges involved in fostering an engaged contributor community. Dodge and Kitchin (2013) articulate some of the recurring concerns: "*There are, of course, ongoing debates around the worth of these connections and communications, with critics arguing that they are superficial, lacking depth and obligatory commitment. There are also concerns over the unpredictability of crowds, [...] and the extent to which the model is sustainable (will volunteers continue to give 'free' labour next week, next month, and next year; what happens when the crowd disperses?)*" A range of existing literature has looked at these and related concerns.

Kraut et al. (2012) provide a comprehensive discussion of a wide range of empirically tested theories to help interpret engagement effects in the context of online community platforms. Among the many contributing factors, contributors' **motivations to participate** are regarded one of the most important drivers for sustained participation. When they are appropriately considered in platform design, they can provide triggers for action, and support rewarding experiences that can sustain contributor engagement for long periods. However, when the contribution experience comes in conflict with existing contributor motivations, participants are unlikely to remain active. A general model of motivations of volunteer workers is presented by Clary et al. who describe six basic motivational categories, including values such as altruism, social experience, and enhancement-related aspects such as self-improvement and a positive self-image (Clary et al., 1998). In a study of WP contributor motivations, this is amended by two further categories: fun, and ideology (Nov, 2007). Among early OSM contributors, it was found that an individual's local geographic knowledge was the most significant driver to contribute (Budhathoki, 2010). In the context of charitable work, it was further shown that the presence of a community of practice can affect a person's willingness to contribute to charitable organisations, as does the perceived degree of social urgency behind a

particular cause (Schervish and Havens, 1997).

A further psychological model that allows us to reason about the nature of engagement barriers is the framework of **self-efficacy** (Bandura, 1977). According to this model, perceived human efficacy determines if an individual will initiate an activity, how much effort will be expended, and how long the activity will be sustained. Self-efficacy is derived from four principal information sources: performance accomplishments, vicarious experience, verbal persuasion, and physiological states. In other words, task designs that aim to increase self-efficacy must consider how best to bring about experiences of mastery, and to not overwhelm prematurely. This can at least in part also involve social processes: persuasion, observation of others, and feedback. Following the trajectory of more engaged contributors on WP, Bryant et al. (2005) find a mutually reinforcing relationship between **experience and engagement**: as contributors gained more experience they took on more roles and responsibilities, which in turn gave them even more experience. This is evidence that there is a possible progression towards intensified engagement by means of improving a contributor's skills. In other words, systems that allow new contributors to gain experiences quickly and successfully may have a strong impact on contributor engagement.

As a corollary to this, in crowdsourcing and citizen science systems it is found that poor **task design** can be a deterrent for participation. Organisers need to strike a balance between organisational performance and worker satisfaction, "*designing tasks suited to online participation that will elicit valuable contributions while maintaining volunteers' interest*" (Wiggins and Crowston, 2011). A basic design strategy is to split the work into smaller pieces that are manageable by a single contributor in a limited amount of time. Kittur et al. (2013) identify this as one of the key challenges lying ahead for crowdsourcing system designers: "*both we and our surveyed workers have observed many cases where poor quality work arises from poorly designed crowdsourcing tasks*". They further propose to communicate more clearly with contributors, but also to better support their learning experience. "*Workers may need to acquire new skills to perform unfamiliar tasks, before or in the midst of performing the actual work*".

In a general discussion of participant engagement in online communities, Koh et al. (2007) observe that a lack of **social presence** in virtual spaces can present a bar-

rier to engagement, particularly in heterogeneous and dispersed volunteer groups. The authors raise considerations that go beyond immediate concerns of productivity, and include questions of individual and collective purpose. Based on a review of existing literature they recommend the development of strong community leaders and moderators, the articulation of a clear purpose or vision, a clear definition of members' roles, and the provision of both online and offline events. The authors conclude, "*Leaders of robust, sustainable virtual communities find ways to strengthen their members' sense of social identity and motivate their participation in the community's activities*" (Koh et al., 2007).

In the study of CSCW, work contexts are often considered along two dimensions: whether participants are **colocated** (they work in the same place), and whether their participation is **synchronous** (they work at the same time; see Johansen 1988; Baecker 2014). 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. There are analogies to this offline practice in WP edit-a-thons, OSM mapping parties, and meet-ups in open source communities. Studies across these settings suggest that interactions in such colocated synchronous practices can foster increased engagement, although this effect is not necessarily sustained (Lin, 2007; Hristova et al., 2013; Jin et al., 2015; Angelopoulos and Merali, 2015; Farzan et al., 2016).

There is some evidence that the event-centric character of **emergency response** can lead to distinct participation patterns. In social media research, studies have documented the willingness of outsiders to participate remotely during crises, for example to help in information propagation (Starbird and Palen, 2010; Vieweg et al., 2010; Starbird and Palen, 2011). Significant participation bursts have been reported for crisis mapping platforms in response to disaster events, as volunteers help collect geographic data to aid emergency response (Meier, 2011; Collins, 2011; Okolloh, 2009). It is not yet well-understood what factors shape participation in such moments, however there are indicators that it differs from more sustained practice. In the context of WP, it was found that breaking news can lead to intense collective editing activity which surpasses that of most other WP articles (Keegan et al., 2013). Such media events can rally a diverse set of contributors, including many first-time

participants who make some minor changes and never return. In comparison, other articles tend to have a more stable contributor network (Keegan et al., 2013).

2.4 Research Opportunities

Within the wider academic discourse of mass-participation platforms, HOT presents itself as a novel kind of crowdmapping platform with some unique characteristics. Using the concepts introduced in this chapter as reference points, we can characterise HOT as a mixed model across multiple dimensions. On the one hand, HOT relies on the peer production platform OSM as a technical foundation, however its coordination model is more similar to a crowdsourcing or citizen science model: an explicit distinction can be made between organisers (requesting parties, project creators) and project volunteers, and the contribution processes are formal and goal-oriented.

On one hand, this suggests that research about the nature of crowdsourcing, crowd work, and citizen science can also be applicable here. This includes studies of participant motivations, technology platforms and coordination practices, and others. On the other hand, this introduces an opportunity to reproduce existing studies in a novel setting, and compare outcomes with expectations in the theory.

A further novel aspect is its coexistence of different participation modalities. According to the Johansen space-time model of participation, HOT online practice can be considered asynchronous and remote, as contributors can act entirely independently of each other. HOT mapathons however are synchronous and colocated. The colocated practice of HOT mapathons may be comparable to social event formats on other mass-participation platforms, however it is not currently well-understood whether they have comparable effects on newcomer retention. Further characteristic for HOT is the coexistence of both synchronous and urgent emergency response campaigns, and more long-term asynchronous campaigns without a specific deadline. It is currently not well-understood how these structural dimensions affect participation behaviours.

Furthermore, there is a general research opportunity to produce large-scale empirical studies of HOT under consideration of a wider range of practices. To date, HOT studies are typically focused on single initiatives, not on a more comprehen-

sive review across activities. Such research could take the form of comparative studies, for example to observe newcomer retention across different settings, potentially with a comparison of campaign outcomes. Our review so far has revealed several such opportunities to study specific aspects of the HOT contribution process, we will summarise these in the following sections.

2.4.1 Effects of Coordination Practices

Importance of task design. Task design has been identified as a fundamental engagement concern in mass-participation platforms, and a particular barrier to entry. In principle anyone can contribute to any open task, yet it is not currently known to what extent current HOT tasks help build self-efficacy, whether they provide sufficient forms of guidance, or whether they might overwhelm some contributors. This suggests an opportunity for an evaluation study, informed by theories on task design and task complexity. Can task designs pose barriers to entry for first-time mappers? For example, what aspects of task design affect early task abandonment? Which forms of guidance are necessary to successfully turn newcomers into productive mappers? Deeper understanding of such effects can improve the ability of task designers to strike a balance between achieving outputs while fostering an active volunteer community. (Chapter 4 on page 63 includes a study of the engagement impact of certain task design factors.)

Importance of social interactions. Specific social factors have been identified as further important engagement factors. HOT has some unique properties that make it feasible to observe the effects of certain kinds of social encounter on a distributed volunteer group at large scale. In particular, its social contribution environments of mapathons provide an opportunity to compare online and offline activities within the same platform. Do the social contribution settings of mapathons have an impact on contributor engagement? For example, does it make a difference when newcomers attend a mapathon to socialise and get trained in person, rather than simply to participate online? What kinds of settings are suitable for fostering an engaged contributor community? Better knowledge of such effects can help determine where to focus limited organiser capacity and resources. (A study on the engagement impact of mapathons is presented in Chapter 5 on page 83.)

Importance of peer feedback. It is feasible that peer feedback given during validation may have important motivational effects. Are HOT newcomers more likely to be retained when they are given certain kinds of feedback? Should validators be concerned about the engagement impact of their feedback, and spend effort on the content and tone of their messages? (A study of the impact of peer feedback during validation is presented in Chapter 6 on page 101.)

2.4.2 Effects of Disaster Response Context

Importantly, the review also revealed that the theory does not yet offer detailed theoretical expectations of participation behaviours during emergency response, and how they may shape project outcomes. There is an opportunity to contribute further evidence of such activities, and help build new theory. Here, as well, HOT can provide a unique capacity to support comparative studies within the same platform: it hosts both urgent emergency response campaigns, as well as more long-term campaigns without a specific deadline. Are they subject to different patterns of participation?

Event-centric nature. It is currently unclear how the urgent needs of emergency response may influence HOT participation. For example, it is conceivable that widely reported disasters may foster a more transient-superficial form of participation, and that this may affect project outcomes (Dodge and Kitchin, 2013; Keegan et al., 2013). To date there are no studies of such an effect in HOT, or other crowdmapping platforms. (Chapter 7 on page 125 provides early evidence that campaigns relating to public-interest topics can yield significantly different volunteer turnout and project outcomes.)

Reactivation of inactive contributors. Similarly, it is conceivable that event-centric coordination may lead to participation behaviours that are characterised by dormance-reactivation cycles. To what extent does a growing pool of trained but inactive mappers form dormant capacity that can be reactivated for emergency responses? (Chapter 7 on page 125 includes a study of the relationship between emergency response moments and participation behaviours.)

2.4.3 Other Participation Concerns

Improving contribution quality. Any attempt to build volunteer capacity should include a reflection of its potential effects on project outcomes. Do certain training or socialisation interventions have an effect on contribution quality? It is challenging to assess map data quality at global scale, in part because there is no globally available ground truth. To address this, researchers have developed intrinsic quality measures for OSM map data (Haklay et al., 2010; Barron et al., 2014; Senaratne et al., 2017). Could these be applied to HOT? Alternatively, is it possible to develop alternative intrinsic measures? This would open up opportunities for a range of large-scale studies on the effects of organiser practices on project outcomes. (Chapter 7 on page 125 includes proposals for two intrinsic contribution quality measures, these are used to compare outcomes of different campaigns.)

The following chapter will outline our general methodology. This includes the process by which we reviewed these research opportunities, and developed them into discrete studies, and an introduction to our primary data sources.

Chapter 3

Methodology

We will first describe the overall process by which we pursue the research questions outlined in Section 1.1 (page 23). Our research begins with an iterative process of problem discovery. We identify impactful research problems that have significance for both theory and practice, and that can be tackled based on the available evidence. For individual studies we then prepare a theoretical grounding, including a contextual framing based on a deep understanding of the domain. Throughout this process, we engage with the HOT community as a means of grounding our research in practice. When executing each study, we primarily rely on quantitative and statistical methods for analyses. Towards the end of this chapter, we describe our sources of empirical evidence, the process by which we prepare an annotated data set that is suitable for large-scale empirical studies, and statistical methods used throughout our studies. We close with a critical discussion of our methodology and its limitations.

3.1 An Iterative Approach to Research Development

There are two classic strands in the research of Computer-Supported Cooperative Work (CSCW): studies rooted in engineering sciences which are focused on building new systems, and studies rooted in social sciences which pursue empirical descriptions of CSCW phenomena and seek to identify causal relations (Kraut, 2002). Our research is motivated by the latter tradition and grounded in social science theory, however to a large degree it relies on computational rather than ethnographic

methods.

Where classic hypothesis-driven social science may involve costly exploratory experiments, our research makes use of 'big data' methods to interpret existing settings. This constrains our ability to isolate specific causal effects, however it allows us to observe a broader range of concerns with only very limited resources. Furthermore, to an extent the exploratory data-driven aspects of the work helps reduce a risk that is inherent in hypothesis-driven research: a risk of a bias of preconceived notions (McFarland et al., 2015).

Goldberg (2015) discusses such research approaches under the term *forensic social science*, "*the careful compilation of evidence from unstructured digital traces*", and makes a case for data-driven research. According to Goldberg, big data techniques are important exploration methods for research problems in social sciences, in part because they may surface phenomena that are not predicted by existing theory. "*It is precisely this potential for surprise that makes Big Data, and the machine learning tools used to excavate and refine it, an opportunity for theoretical discovery*" (Goldberg, 2015).

However it needs to be acknowledged that statistical methods of categorising and analysing data also embody assumptions about the world, they are "*rhetorical devices disguised as objective representations of reality*" (Goldberg, 2015). Researchers need to be aware of these inherent assumptions when interpreting the data, and be aware of the fallacies they may invite. This includes the fallacy that the sum of captured observations describe the entire population, that the observations are recordings of naturalistic behaviour, or that they can be understood outside of their immediate context (Lewis, 2015). A growing body of literature is outlining a range of further fallacies and other necessary considerations (Ruths et al., 2014; Lewis, 2015). In addition there are a growing number of examples demonstrating the effects of careless interpretation, for example the high-profile case of the Google Flu prediction system (Lazer et al., 2014).

Forensic social science can be understood as a constructive synthesis across paradigms, a middle ground that is "*both inductive and theory-oriented*" (McFarland et al., 2015), rather than a purely statistical approach that is removed from theory. In this spirit we base our research on multiple sources of evidence: readily available observational data, existing applicable theory, but also a deep understanding

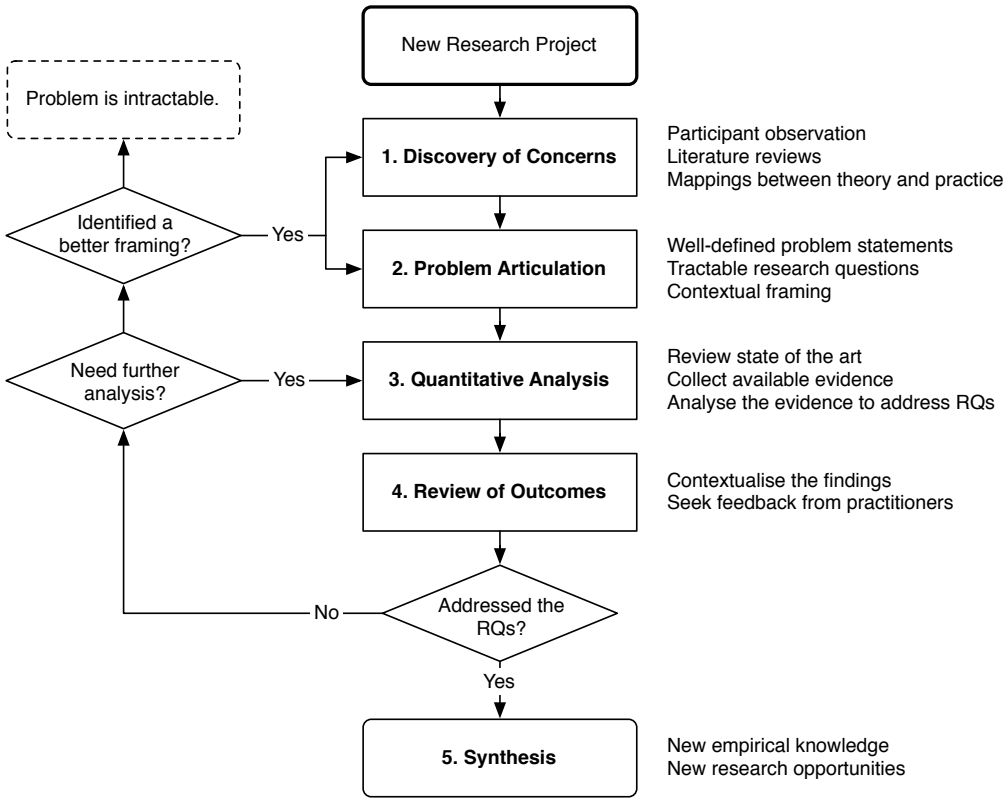

Figure 3.1: Iterative process for designing and executing our studies.

of community concerns and practitioner perspectives.

Importantly, we developed each of our four studies in an iterative manner, working closely with the HOT community. Figure 3.1 provides a schematic representation of the general process for developing our studies. We begin with an initial process of **problem discovery**, this involves different approaches to identify a set of initial concerns. On one hand, we discover existing problems in HOT practice though participatory observations and discussions with the community (refer to Section 3.3 on page 58.) On the other hand, we identify existing problems in theory that we may be able to address through the lens of HOT (refer to Chapter 2, and the literature review sections for each of the studies.) Based on both, we establish a basic mapping between theory and practice, identifying concerns that have relevance for theory and practice.

To turn these into discrete projects, we engage in a process of **problem articulation**. We select a particular volunteer engagement challenge, and describe it as

a set of well-defined problem statements. In order to identify means of addressing the concern, we specify a set of specific research questions or hypotheses. This can include questions about the efficacy of past interventions that were undertaken by the community. These are captured before any analysis takes place. We take care to specify research questions that can feasibly be addressed with the available evidence. Importantly, we also establish a contextual framing to motivate the concern and problem articulation, for example by describing specific scenarios that illustrate the concern. This yields one or more well-specified volunteer engagement problems in HOT.

Once a problem has been defined, our main effort is spent on **quantitative analysis** and general project execution. We assess the existing state of knowledge relating to the problem, this sets a basic expectation for our study. We review and contextualise the available evidence, developing derivative forms that are sufficient to address the stated research questions (see Section 3.2 on page 55). We also identify whether further evidence may be needed. We then analyse the evidence with respect to the research questions, primarily relying on statistical methods. This yields a set of initial findings.

An important stage for each study is a subsequent **review of outcomes**, which may inform further iterations and refinements. We summarise and contextualise analysis findings, relate them to existing theory and to our understanding of HOT practice, and compare them with our prior expectations. Based on this, we assess whether the findings are plausible, and develop an initial understanding of their implications. Importantly, we discuss the most important findings with the HOT community to seek feedback from HOT practitioners. This can take different forms, and can happen online or in person (see Section 3.3 on page 58 for details.) Depending on the outcomes of this review, we may decide to refine our analysis approach, for example to incorporate additional concerns that were overlooked in the first instance. If the review yields an unexpected expression of the problem, we may update our theoretical framing, or even return to a more fundamental process of problem discovery.

The overall outcome of this process is a **synthesis of practical and theoretical concerns**. It yields new empirical knowledge that has a strong grounding in theory

and practice, and that is articulated using a well-specified frame. As a secondary outcome, the process may identify additional knowledge gaps and opportunities for further research.

3.2 Sources of Empirical Evidence

We developed methods to gather empirical evidence of contribution activity, based on readily-available public data, in order to perform both early explorations and the final research. In the following sections we describe our primary sources of evidence, as well as our process of preparing it for analysis. This includes means of aggregating and segmenting the data across fundamental dimensions of concern: attributing contribution activity to discrete HOT projects and campaigns, estimating the amount of time participants spent mapping, and classifying participants by their prior contribution experience.

3.2.1 Map Contributions

A primary data source for the research is the full **OSM editing history**, a public data set which captures the changes OSM contributors have made over time. The data is freely available for download.[5] The OSM data model makes a distinction between three basic geometries: **nodes**, **ways**, and **relations** (OpenStreetMap, 2014). Discrete place markers can be expressed as nodes, while roads and other path objects are expressed using ways. More complex map objects which are groupings of simpler geometries are expressed using relations. All geometry types can be annotated with **tags**, key-value pairs which provide further contextual information, such as names, references to classification schemes, and other annotations. Map objects have a version counter which is increased whenever the object is modified by a contributor. Any creation or modification of a map object is considered an **edit**. Editing tools collect changes made to multiple map objects into so-called **changesets**.

A further data source for the research is the *HOT Tasking Manager*[6], a website which helps coordinate the work of thousands of online contributors (see Section 2.1.2 on page 31). It was first introduced in late 2013 in order to better coor-

[5] `http://wiki.openstreetmap.org/wiki/Planet.osm/full`

dinate the work of large numbers of volunteers (Palen et al., 2015). It presents a list of currently active HOT **projects**, along with contextual information and mapping instructions. Work in each project is divided into smaller geographic regions, each representing an individual task. For the purposes of this research, the tasking manager is an information source in two respects. On one hand it is a rich public archive of remote mapping activities, with task descriptions, boundary polygons for project regions, and other metadata. Additionally it serves as a public record of HOT participation: every project records a list of its contributors.

In a preparatory stage, we identified the map contributions for all Tasking Manager projects. As of summer 2015, editing tools automatically annotate OSM changesets with a HOT project identifier, this makes such an identification straightforward. In cases where this was not provided, edits were instead associated with a HOT project based on their location, date, and contributor. This allowed us to identify the full set of HOT remote mappers, their contributions to HOT, and any further contributions they made to OSM which were not linked to HOT activities.

This combined data set allows us to estimate the size and growth of the HOT contributor community. By September 2016, 32,000 contributors have made 180 million map edits across 2,000 HOT projects, involving an estimated total of 240,000 hours of volunteer work. (Refer to Section 2.1.4 on page 35 for timelines of HOT activity growth, and the geographic spread of its mapped regions.)

3.2.2 Identifying Larger Campaigns

In an effort to better understand the spectrum of HOT activities, we sought to identify key remote mapping campaigns: larger collections of Tasking Manager projects with a shared singular purpose and shared organisational practices. Typically these relate to a distinct disaster event, or reflect long-running efforts organised by a particular group. We identified such collections based on project listings on the HOT homepage[8] and the OSM wiki[9], and based on a review of recurring project identifiers in project names. Approximately 50% of HOT projects could not be allocated to a particular campaign, however these only account for 30% of the total map edits.

[6] `http://tasks.hotosm.org/`
[8] `https://hotosm.org/projects`
[9] `https://wiki.openstreetmap.org/wiki/Humanitarian_OSM_Team`

Figure 3.2 shows a timeline of all campaigns we could identify during this process. The chart highlights the campaign midpoint, which is the date by which 50% of all campaign contributions have been made. The visualisation serves as an early indication that campaigns can follow different participation trajectories: some campaigns are characterised by a very short initial contribution period, which may be followed by a prolonged period of less activity. (This phenomenon is investigated in more detail in Chapter 7 on page 125.)

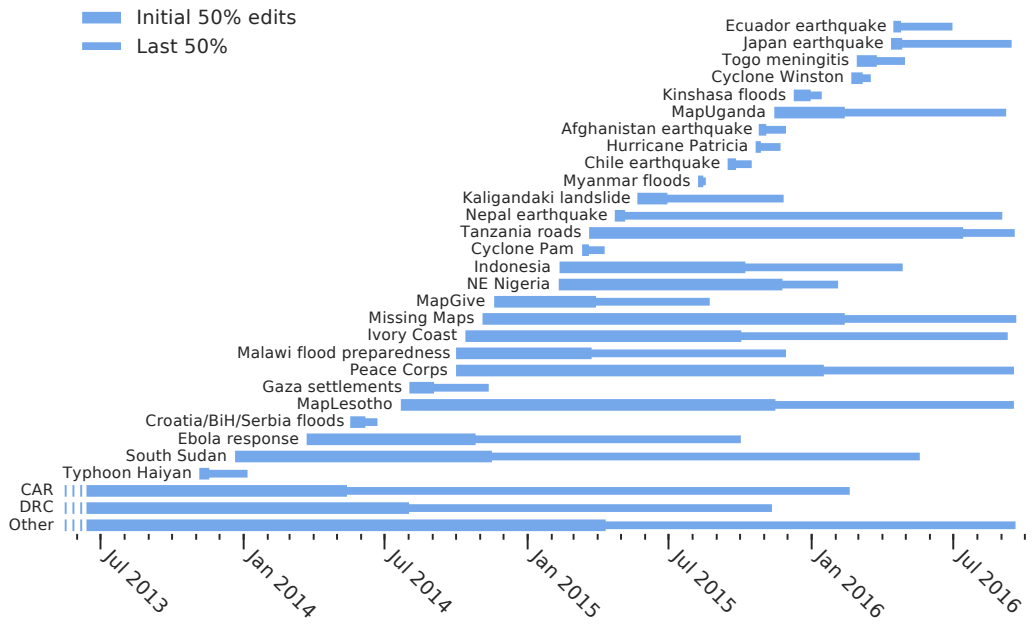

Figure 3.2: Timeline of HOT campaigns. Each line visualises the activity period per campaign, indicating how much time passed until 50% of the overall contributions were made.

3.2.3 Contributor Classification

We use the amount of time an individual spends contributing to HOT as an important means of assessing contributor engagement. This measure can be derived from contribution records. We computed contribution sessions for every contributor based on the timestamps of individual contributions, with a session timeout of one hour. Based on these, we calculated the number of **labour hours** spent on each contribution, using a process described by Geiger et al in the context of Wikipedia contributor analysis (Geiger and Halfaker, 2013). Such time-based activity mea-

sures have some advantages over simple edit counts because they allow to more meaningfully compare contributor effort across different kinds of tasks. In many of our analyses we seek to quantify the amount of time spent contributing, rather than the volume of data that was produced, in order to compare contributor effort across different kinds of contributions.

Figure 3.3 shows a histogram of the total number of hours that HOT contributors spend across their overall participation. 50% of contributors have spent at least 65 minutes contributing to HOT, and 20% for at least 2 hours. The top 5% have contributed for at least 18 hours combined.

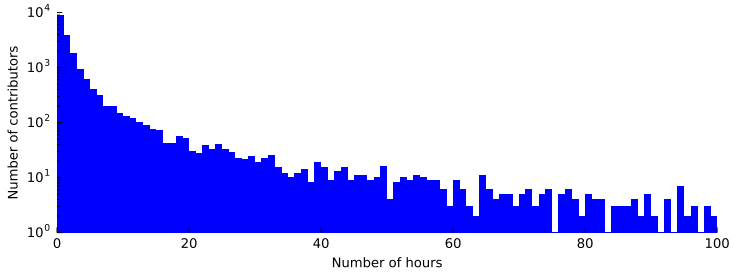

Figure 3.3: Histogram of total labour hours across all contributors.

The edit history also captures any prior OSM contributions that may have preceded the first HOT participation. This makes it possible to quantify each contributor's degree of **prior OSM experience**, for example by counting the number of days with OSM contributions prior to the first contribution to a HOT project. Based on such measures we can distinguish between HOT volunteers who have prior experience in the OSM contribution process, and those who are novices. We found that most HOT contributors are novices: 80% have no prior OSM experience, and 90% have less than 5 prior days of OSM contributions.

3.3 Participatory Research and Community Feedback

The HOT contribution history is a rich source of observational evidence, but it has some important constraints. Importantly, it does not record many contextual factors of the HOT coordination practices and contribution settings that informed contribution behaviours. In order to address this constraint, we engaged with the HOT

community throughout the research. This was not undertaken with an intention to collect further primary evidence. Instead, it can be better characterised as background research, or as a research partnership. The engagement was driven by several motivations: to understand HOT practices and processes, to witness a range of contributor experiences and settings in some detail, and ultimately to better understand the organisational challenges which can inform our research. Finally, it served as an important source of domain feedback throughout our research.

Our encounters with the community took place in a range of online and offline settings, typically in the role of a participant observer (Atkinson and Hammersley, 1994). They involved the observation of and participation in online discourse, including discussions on HOT mailing lists, online chat, and social media. Often participants in these settings are highly-engaged HOT community members and organisers, and their conversations provide a means of understanding current activities and current concerns.

The work also entailed regular participation at mapathons and other events, which provided opportunities for in-person conversations with organisers, trainers, and participants. At MM (Missing Maps) mapathons in London, organisers tend to be affiliated with aid organisations or the local OSM community, while participants are typically young professionals, students, and information workers, but generally cover a range of backgrounds and age groups.

Participation in planning meetings with HOT organisers provided further opportunities for discourse. A starting point for the research was a first workshop with MM organisers in London in early 2015. The aim of the workshop was to discuss current knowledge about HOT community engagement, and to capture open questions which could inform research projects. This yielded a first set of research priorities which was then refined over the following months. Similarly important was the participation in a first global gathering of MM organisers in Toronto in late 2015, a year after the group was founded. Over the course of three days, attendees reviewed community progress over the past year, and discussed plans for the coming year. From a research perspective it was an opportunity to review new knowledge, to discuss interpretations and implications, and to capture further open questions. The event was attended by MM organisers from multiple countries, and

representatives of international partner organisations.

A public research diary has become one of the most important means of gathering community perspectives and seeking community feedback.[1] It was started in June 2015 to share early research findings, data explorations and visualisations, and observations and interpretations of community practice. The writing was intended to inform the community about the research, but also to stimulate debate. Not all posts found resonance with the community, but some have provoked a remarkable range of feedback and observations by practitioners. These responses are not necessarily empirical evidence in themselves, but they expand our realm of experience in important ways, and have regularly challenged our interpretations of a given phenomenon. Commenters are typically HOT organisers and trainers, HOT contributors with strong ties to the OSM community, and other highly engaged mappers who are interested in the respective topic. Posts are automatically listed in an aggregated archive of OSM user diaries,[2] announced using the researcher's personal Twitter account,[3] and regularly shared with the HOT mailing list.[4]

In this open approach to exploring the domain and prioritising the work, the research includes aspects of Action Research and Participatory Action Research (Hayes, 2011), however it does not involve the communal development of an intervention or specific action. Instead it can be characterised as a communal effort to seek deeper understanding.

3.4 Methodological Considerations

3.4.1 Focus on Quantitative Methods

Our quantitative methods are limited in their explanatory power. They can yield empirical evidence for the magnitude of certain effects, and based on this they can help compare the importance of different contributing factors. On the other hand, they only provide a limited capability to reveal models of underlying causal relationships. For this, we rely on existing theory.

[1] https://www.openstreetmap.org/user/dekstop/diary
[2] https://www.openstreetmap.org/diary
[3] https://twitter.com/dekstop
[4] https://lists.openstreetmap.org/listinfo/hot

As Chapter 2 has indicated, prior theoretical and empirical work of contributor behaviour on mass-participation platforms can provide such foundations. Conceptional models have been proposed for a range of comparable settings. This body of theory provides us with a catalogue of potential concerns and proposed causal relationships. We review these in the context of specific studies, and apply them to our body of evidence. Because of this existing wealth of theoretical knowledge, we argue that quantitative methods have a unique power to address our research concerns for this particular setting. When informed by such theory, statistical methods can help us interpret the evidence of the HOT edit history in a principled manner.

3.4.2 Ethical Considerations

Embedded community work as described in Section 3.3 can yield detailed impressions of contributor activity, and provide a greater degree of access to community knowledge and information, however it also requires careful consideration of the constraints and ethical dilemmas of the approach (Labaree, 2002). For example, in the context of a global community such as HOT, where activity takes place across a large number of online and offline settings, it must be recognised that such observations can only capture the activities of a subset of participants in very specific settings. Furthermore, researchers must carefully consider their dual role of a researcher and participant, and disclose their intentions early. An important constraint we placed on our work was to not observe the specific acts of individuals, but rather to capture settings and practices in aggregate.

3.4.3 Limitations

There are some inherent risks and limitations with our approach. Importantly, our observational research approach can only reveal associative rather than causal relationships. To be certain, our findings need to be confirmed with controlled experiments. The approach also limits our work to an assessment of existing practices, rather than an exhaustive assessment of all possible practices. Furthermore, although we underwent significant and systematic efforts to reduce the chance of confounding effects, it is possible that our studies fail to consider important contributing factors.

An important aspect for three of our four studies is a focus on first-time contributors, starting from their initial participation in HOT. In these first hours and days, we can expect that engagement is at least in part shaped by the specific contribution experience of this first project. We can assess whether the initial experience was so discouraging that contributors never returned, or whether it prepared them to contribute for longer periods. A more experienced contributor on the other hand may still be able to contribute to a badly designed task, which would make it harder to identify problematic design choices.

Due to the nature of our primary evidence, we are limited to observations of contribution behaviour in the form of actual map contributions, and many other activities are not captured by our evidence. This includes participation in other online and offline spaces provided by HOT. Except in specific cases, we do not typically have information about a contributor's setting. The edit history also does not capture how contributors were informed of a particular HOT campaign, for example whether they discovered it through public media channels, community discussion forums, by browsing the tasking manager listing, or in other ways.

Our study is specifically focused on contributor engagement. We make an early effort to assess contribution quality in Chapter 7, however a more detailed analysis of contribution would exceed the scope of the work. This is in part the result of a methodological challenge: it is not evident how contribution quality can be measured at the scale of this study. A fundamental barrier is the absence of a reliable ground truth: by definition, the maps produced by HOT are often the first of their kind, and there typically are no other maps to compare against. As a consequence, the few early studies of HOT data quality either rely on specific contextual knowledge of particular geographies, or they assess quality by reproducing the volunteer work with other means, both of which limits their scale (Westrope et al., 2014; Eckle and de Albuquerque, 2015).

Finally, although we seek to develop a more comprehensive understanding of the problem domain, we cannot claim to have covered it exhaustively. Instead, we focus on certain key aspects where research can have an impact on theory or practice, where there is suitable theory to inform a rigorous study, and where evidence is readily available to address the chosen concern.

Chapter 4

Campaign Coordination and Task Design

HOT organisers need to consider how to produce outcomes with their projects, but also how to build volunteer capacity. A review of the literature has indicated that the initial project experience of contributors plays an important role in this, particularly when the contribution process requires some degree of expertise. In a first study, we propose three analytical methods to assess first-time contributor engagement: cohort analysis, task analysis, and observation of contributor performance. Our study shows that different coordination practices can have a marked impact on contributor retention, and that complex task designs can be a deterrent for certain contributor groups. We close by providing recommendations about how to build and sustain volunteer capacity in these and comparable crowdmapping systems.

4.1 Introduction

HOT organisers not only need to consider how to produce maps, but also how to foster a large global volunteer community in the process. HOT campaigns reflect two complementary aspects of this community-building challenge: disaster aid initiatives need to build volunteer capacity to provide quick emergency response, and disaster preparedness initiatives need to sustain volunteer capacity in the absence of urgent causes. While organisers have significant freedom in designing campaigns, it is not clear how they can evaluate their choices in these regards. Furthermore it is not always clear whether certain design choices involve trade-offs.

Existing research has already assessed the quality of HOT outputs, and their impact on the map (Eckle and de Albuquerque, 2015; Westrope et al., 2014). Our initial study will instead focus entirely on engagement aspects: the existence of HOT presents a rare opportunity to compare different coordination practices within the same platform, involving a large number of projects and participants. We evaluate whether individual projects can successfully activate new volunteers (**enrolment**), but importantly also whether HOT can retain them over time (**retention**). Together we define these as **engagement**.

A range of HOT initiatives and organisational practices offer many opportunities to evaluate specific organiser choices. We aim to assess a large number of participations in a consistent manner. To this purpose we propose three analytical dimensions: **cohort analysis** where we compare larger campaigns encompassing multiple projects, **task analysis** where we compare projects in their task complexity, and observation of **contributor performance** relating to the rate of contributions. All rely on readily available public data, and we will demonstrate that they can yield important findings.

On the following pages we first present three research questions motivated by the growth challenge we presented. We then describe key campaigns, present an overview of related work, and introduce our methodology. Finally we address our research questions with a set of analyses based on contributor engagement metrics, and close with a discussion of our findings.

4.2 Research Questions

4.2.1 RQ1: Cohort Analysis

Are different coordination practices associated with different contributor engagement characteristics? In large distributed online communities we are likely to find subgroups with divergent practices. In the case of HOT, these are the different humanitarian causes within the same platform, organised in the form of campaigns.

We expect that different coordination practices and circumstances can affect contributor engagement in different ways. In particular, we believe that the perceived urgency of a cause can act as an attractor and lead to higher enrolment. On

the other hand, we expect that sustained promotion of needs can likely increase retention, however it is unclear how well this works, and how long it can last.

4.2.2 RQ2: Task Analysis

Are different task designs associated with different contributor engagement characteristics? Task analysis serves to assess the impact that contribution mechanics can have on the initial enrolment experience. It allows to distinguish between different tasks in terms of their task complexity.

We expect that a minimum degree of task complexity is needed to yield substantial contributions, however there may be diminishing returns: more complex work can be discouraging. This may be particularly true for newcomers, who may choose to abandon their participation early.

We further expect that such effects can be addressed with better contributor guidance. A gradual learning curve and other forms of guidance can help newcomers build a sense of self-efficacy. However we also expect diminishing reports in this context: too much documentation can be overwhelming as well.

4.2.3 RQ3: Contributor Performance

Is contributor performance an early indicator of retention? A volunteer's rate of contributions, or edit pace, can be considered a key measure of their performance. Do faster contributors tend to stay longer? Can we observe performance improvements during the initial period of participation? Are these measures associated with particular retention profiles?

We expect that faster contributors tend to stay longer, they may be more confident in their abilities. We further expect to see that newcomers start slowly, but then pick up their pace. We further expect that any observed performance improvements are associated with an increase in short-term retention, as a result of increased contributor self-efficacy and enjoyment.

4.3 Related Work

A key barrier to entry for first-time HOT contributors is the fact that mapping with OSM is a complex practice: it requires specialist tools and an understanding of specialist concepts (Schmidt et al., 2013). To our knowledge there is no published research on how this affects HOT contributor engagement, however there is some knowledge in related domains.

In crowdsourcing and crowd work systems, organisers (or "requesters") prepare tasks that are then offered to contributors (see Section 2.3.1 on page 41). Organisers need to strike a balance between organisational performance and worker satisfaction: badly designed tasks can be deterrents for participation (Wiggins and Crowston, 2011). A basic design strategy is to split the work into smaller pieces that are manageable by a single contributor in a limited amount of time.

Kittur et al. further propose to communicate more clearly with contributors, but also to better support their learning experience. "Workers may need to acquire new skills to perform unfamiliar tasks, before or in the midst of performing the actual work" (Kittur et al., 2013). In an evaluation of Mechanical Turk crowd workers, Khanna et al. found that more complex tasks that require a nuanced understanding of the domain can pose barriers to participation. Other barriers included problems related to the user interface, or misunderstandings derived from differences in cultural contexts between coordinator and contributor (Khanna et al., 2010).

There is some evidence that increased activity and increased retention may not always be achievable at the same time. In online citizen science projects it was found that more prolific contributors can have shorter retention periods (Ponciano and Brasileiro, 2014; Ponciano et al., 2014). A similar effect was found for crowdsourcing designs that aim to increase member productivity, and was attributed to either burnout or a sense of a "mission accomplished" (Wang et al., 2012).

Several psychological models allow us to reason further about the nature of engagement barriers, including the framework of self-efficacy (Bandura, 1977), and models of contributor motivations in volunteering and peer production settings (Clary et al., 1998; Nov, 2007; Budhathoki, 2010; Schervish and Havens, 1997). These were introduced in Section 2.3.2 on page 43.

4.3.1 Knowledge Gaps

Despite increased research interest in the topic, to our knowledge there has been no published research on the ability of different HOT project designs to build volunteer capacity, and then successfully retain trained contributors over longer periods.

4.4 Methodology

Our primary source of empirical evidence is the HOT contribution history, this data set and its derivative forms are described in detail in Section 3.2 on page 55.

4.4.1 Study Period

We selected an 18-month enrolment window from mid-2013 to late 2014 to observe first-time contributions, further extended by a buffer period of 180 days for the observation of contributor retention. This time captures the most active period of the Tasking Manager history at the time of study. It excludes an initial early-adopter period, but includes the first Tasking Manager use at a larger scale in late 2013 (Palen et al., 2015).

Figure 4.1 shows the remarkable growth of HOT remote mapping activity in this time: by early 2015, HOT organisers had created almost 1,000 remote mapping projects. Our study period is highlighted in the graph.

The specific timeframe, all dates are inclusive:

- First date of enrolment window: 16th of June 2013
- Last date of enrolment window: 15th of December 2014
- Plus 180-day buffer period: ends 16th of June 2015

4.4.2 Cohort Selection

For this first study of HOT engagement we limited our evaluation to key campaigns: larger collections of Tasking Manager projects with a shared singular purpose and shared organisational practices. We identified such collections based on a process described in Section 3.2.2 on page 56. In order to find collections that are suitable for both cohort analysis and task analysis, we rejected campaigns that involved less

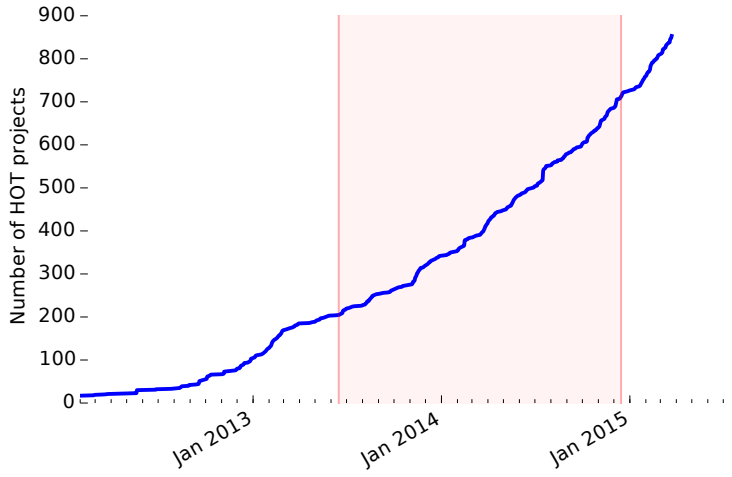

Figure 4.1: Growth of HOT Tasking Manager projects over time. Our study period is highlighted.

than 50 contributors, and less than 10 projects. We identified six larger campaigns as candidates. Of those we chose three as study cohorts based on their profiles: they represent a cross-section of typical HOT coordination practices, from urgent responsive mapping to sustained proactive mapping. Their aggregate size is substantial: they accounted for almost 50% of all first-time contributors in the period, and encompassed 30% of all projects on the Tasking Manager at the time of study.

4.4.2.1 Typhoon Haiyan (Cohort: TH)

Also known as Typhoon Yolanda, TH was a tropical cyclone that devastated the Philippines on November 8, 2013. The HOT projects associated with this campaign were of high urgency: the map data was used in disaster response and humanitarian aid as soon as it became available. TH was probably the first highly promoted HOT campaign to rely on the Tasking Manager for coordination. Volunteer work started only few days before the event, initially with projects to prepare a base map. The focus later switched to damage assessment of the affected areas. The OSM wiki lists 22 mapathons which were organised around TH in November 2013 in cities around the world. Typically these were one-off events (OpenStreetMap, 2016l).

4.4.2.2 Ebola Response (Cohort: ER)

The Ebola outbreak began in Guinea in early 2014. Aid organisations needed maps to locate and treat those infected, yet many of the affected areas were not documented on any existing maps. Initially this campaign was treated as an urgent one-off event, and activity stopped soon. However coordination was picked up again as the epidemic spread to neighbouring countries, and the strategy changed to a more sustained effort covering larger regions. The campaign lasted until early 2015. Mapathons were organised in many parts of the world to train newcomers and coordinate volunteers, including monthly events in several cities (OpenStreetMap, 2015). Project descriptions indicate that many activities were coordinated by representatives of aid organisations.

4.4.2.3 Missing Maps (Cohort: MM)

This effort launched in November 2014: regions vulnerable to crises are mapped early so that maps are already available when a crisis occurs (MM, 2017). In contrast to TH, this campaign is proactive rather than reactive, and continuity of contributor engagement is more important than a quick response. Where TH and ER were more ad hoc in their coordination, MM organisers set up structures to support more sustained engagement: there are regular mapathons in many parts of the world, often organised as monthly events. A signup form for volunteers allows contributors to get informed of upcoming campaigns and mapathons (MM, 2017). Similar to ER, many of the MM projects were initiated by representatives of aid organisations.

We identified instances where new first-time HOT contributors joined one of these projects. These are our study cohorts: the newcomers who joined during one of the three campaigns. We then selected contributions relating to each campaign in the study period, the final data set derived from this is summarised in Table 4.1, with a total of 1,570 first-time contributors across 99 projects.

A timeline of contributor enrolments per cohort is shown in Figure 4.2, this plot illustrates a number of key aspects of our data set. The cohorts were active at different periods: TH in late 2013, ER throughout much of 2014, and MM from late 2014 onwards. The contributor distribution also indicates that many first-time contributors already had existing OSM user accounts, as indicated by their user

Cohort	*Projects*	*First-time contributors*
TH	23	481
ER	65	881
MM	11	208
Total	**99**	**1,570**

Table 4.1: Project cohort sizes.

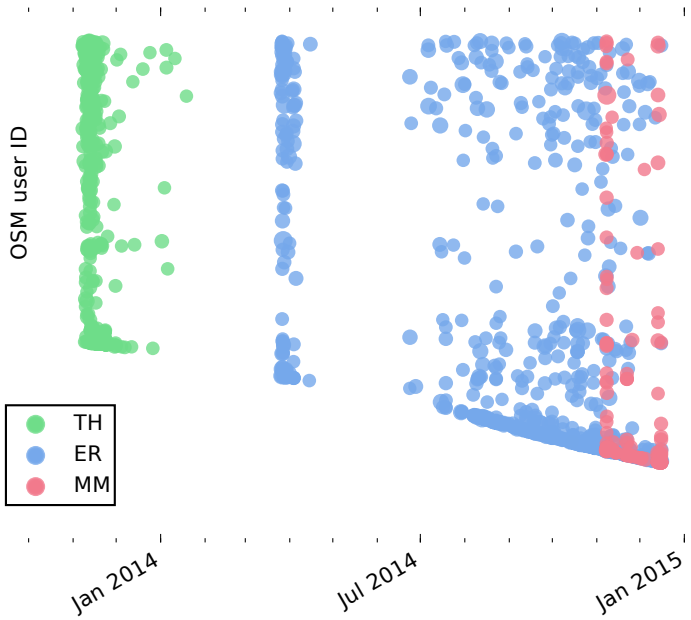

Figure 4.2: Timeline of contributor enrolment, broken out per cohort. Contributors are arranged vertically in order of ascending user ID.

IDs. The plot also reveals a participation gap for ER between May and July 2014, this reflects the period of inactivity between its initial conclusion in early 2014, and a subsequent resurgence of activity later in the summer.

4.4.3 Contributor Observation Periods

We collected all contributions by first-time contributors in the first 180 days after their initial contribution. We paid particular attention to three different timeframes per contributor:

- The initial enrolment period of the first 48 hours. This period is used to observe initial contributions, and to assess how many contributors returned immediately on the second day. The short initial observation period of two days was chosen based on the median difference between the first and last moment of contribution of all first-time contributors, which is only 20 hours.
- A 90-day retention period from the moment of enrolment. This was chosen based on an analysis of average contributor lifetimes: almost 90% of contributors cease contributing after 90 days of their initial participation.
- A 180-day survival period from the moment of enrolment, to identify the last known moment of contribution. This was used for survival analysis: we considered contributors 'dead' if they had been inactive for at least 90 days by the end of this survival period. The last known date of contribution before that point marks their 'death event'.

In summary this means that we only consider the first 90 days of contribution activity for our analyses, however we observe for a full 180 days after enrolment to establish abandonment with some degree of certainty.

4.4.4 Engagement Metrics

Based on these relative observation windows, we computed engagement metrics for every first-time contributor we identified. We computed contribution sessions based on the timestamps of individual contributions, with a session timeout of one hour. Based on these, we computed the number of labour hours spent on each contribution, using a process described by Geiger et al in the context of Wikipedia contributor analysis (Geiger and Halfaker 2013; see also Section 3.2.3 on page 57).

A first set of engagement metrics are *measures of activity in the enrolment period*. We chose time-based activity measures rather than simple edit counts because they allow us to more meaningfully compare contributor effort across different kinds of tasks (Geiger and Halfaker, 2013). We seek to quantify the amount of time spent contributing, rather than the volume of data that was produced, so we can compare contributor effort across different map object types.

More specifically, we captured the number of labour hours l_{48h} in the first 48 hours. These are also calculated separately for the first and second day in this initial

period: l_{d1}, l_{d2}. We further capture the rate of contributions in the firth 48 hours c_{48h}, measured in edits per hour, and contribution rates for the first and second day c_{d1} and c_{d2}. These allow us to determine a change in pace between the first and second day to test for the presence of performance improvements. This change in pace is described by the ratio c_{d2}/c_{d1}.

A second set of metrics are *measures of retention*. These are calculated per project: what share of first-time HOT contributors could later be retained for further activities on any HOT project? To quantify short-term retention per project we determine R_{d2}, the percentage of first-time contributors that are still active on the second day of their participation. Additionally we calculate long-term retention metrics R_{m2} and R_{m3}, the retention rates in the second and third month after enrolment. These 30-day periods were chosen to reflect monthly mapathon cycles observed by some HOT campaigns.

4.4.5 Quantifying Prior Domain Experience

We further quantified each contributor's *degree of prior OSM experience* as d_{pre}, the number of days on which they had contributed to OSM before they joined their first HOT project. In certain analyses we used this measure as a control variable: contributors with prior OSM experience may find it easier to contribute to HOT.

In an initial assessment of the impact of prior experience we correlated d_{pre} with our engagement measures. We found that less experienced users contribute for less hours during their enrolment (Spearman correlation coefficient $\rho_S = 0.19$), that they contribute at a lower pace ($\rho_S = 0.23$), and that they are retained less often than others in the second month ($\rho_S = 0.12$) and third month ($\rho_S = 0.13$, all with $p < 0.001$).

Comparison across cohorts shows that the three groups have different constituencies: the most experienced group was TH (median: 5 days of prior OSM contributions, mean: 113 days), the most inexperienced group was MM (median: 0 days, mean: 23 days), and ER was in between (median: 0 days, mean: 52 days). Table 4.2 visualises the distributions as a histogram. Pairwise comparison of these distributions with Kolmogorov-Smirnov statistic (KS) confirmed that these are statistically different populations ($p < 0.001$), with the greatest difference measured

Experience d_{pre}	0 days	1-9 days	10-99 days	≥100 days
TH	30.9%	22.2%	20.3%	26.6%
ER	52.8%	24.3%	11.7%	11.2%
MM	72.8%	18.4%	5.3%	3.4%

Table 4.2: Distribution of prior OSM experience per cohort, in days with contributions (d_{pre}).

between TH and MM (KS statistic: $\alpha = 0.41$).

4.4.6 Task Analysis

We rely on the fundamental assumption that some tasks are more challenging than others, and that this affects contributor engagement. We expect that we can identify this effect by observing participation across different tasks. We specifically aim to assess how easy it is for a newcomer to become a productive contributor.

In "Task Complexity: Definition of the Construct", Wood distinguishes between the complexity of the required work itself, and the amount of information cues and guidance necessary to produce it (Wood, 1986). These can be quantified as the number of acts and number of information cues involved in the work. Both are measures of task complexity: simple tasks require processing fewer cues than complex tasks (Payne, 1982). They reflect that task designers must consider how to bring about experiences of mastery without overwhelming their participants.

Table 4.3 lists the six task design features we considered for this study, including motivational factors, visual complexity of satellite imagery, task complexity, and forms of guidance. These were labelled in an iterative process, involving a detailed study of a large number of projects. We categorised all requested map features by their geometries and semantic role: natural features, roads and highways, settlement boundaries, buildings, urban infrastructure, and other features. We similarly classified and counted the distinct number of information cues per project: stated priorities, descriptions of map object types, explicit sequences of work steps, and external links to coordination pages, reference documents, or instruction manuals.

These features were then used to compute task design feature vectors for each

Aspect	Variable	Description
Motivation	*has_context*	Does the project description state an explicit purpose?
Visual complexity	*urban_density*	Is the mapped region rural (simple), mixed, or urban (complex)?
Task complexity	*num_concepts*	How many different types of map objects are to be mapped?
Task complexity	*building_trace*	Are buildings to be mapped as points (simple) or polygons (complex)?
Guidance	*num_cues*	Number of information cues provided in the documentation?
Guidance	*num_tag_ex*	Number of tag examples listed?

Table 4.3: Task design feature vector produced by the task analysis.

project, they are summarised in Table 4.3. We standardised these feature vectors using z-scores, so that all variables have a mean of 0 and a standard deviation of 1. We found no multicollinearity across the standardised variables: the condition index of all variables is 2.33, and we found no near-zero eigenvalues in their cross-correlation matrix.

4.5 Findings

4.5.1 RQ1: Cohort Analysis

Table 4.4 shows median engagement statistics per cohort for the enrolment period of the first 48 hours. TH and ER have similar enrolment profiles, while MM contributors appeared to work longer but also contributed more slowly than the other cohorts. We confirmed the pairwise difference of these distributions with a KS statistic: with one exception these differences in distribution were statistically significant ($p < 0.01$). The only instance where a difference could not be asserted were the distribution of labour hours in the TH and ER cohorts, however their rate of contributions differed.

Table 4.5 shows the corresponding median retention rates per cohort, broken out for short- and long-term retention periods. These measures indicate that TH and ER have the highest short-term retention, with almost 30% of contributors returning on day 2. However the long-term retention rates of TH are the lowest of all cohorts. In contrast MM starts with the lowest short-term retention of only 10% returning on second day, followed by the most stable long-term retention of all cohorts: almost

Cohort	l_{48h}	c_{48h}
TH	1.14	640.4
ER	1.16	620.0
MM	1.29	529.2

Table 4.4: Median contribution activity by cohort: labour hours (l_{48h}) and contribution rate (c_{48h}) in the first 48 hours.

Cohort	R_{d2}	R_{m2}	R_{m3}
TH	26.8%	4.2%	4.6%
ER	27.2%	13.6%	8.9%
MM	10.1%	9.6%	8.7%

Table 4.5: Median retention for day 2 (R_{d2}), and months 2 and 3 (R_{m2} and R_{m3}). These indicate the share of HOT contributors who are still active in the respective period.

9% contributors are still retained by month three. We computed survival functions for each cohort based on a Kaplan-Meier estimate with a 95% confidence interval. A pairwise logrank test confirmed statistically significant differences in survival rates between ER and TH ($p < 0.001$), no other pairing was found significant.

When limiting survival analysis to OSM newcomers only, where the first HOT contribution was also the first OSM edit, we observed a statistically significant difference in survival rates between TH and MM, as well as TH and ER (both $p < 0.001$). The corresponding survival plot for OSM newcomers in Figure 4.3 illustrates this: TH has the lowest overall retention. ER and MM differ in the short term, with MM having lower initial retention, yet then achieve similar long-term retention characteristics. According to the plot MM may even have the highest long-term retention, yet this difference was not confirmed by the logrank test.

4.5.2 RQ2: Task Analysis

We prepared a regression model to observe the impact of the six task design features on early abandonment, measured in labour hours l_{48h} as the dependent. We included two control variables: prior user experience d_{pre}, and the size of the project (its number of tasks) to account for goal-setting effects with larger projects.

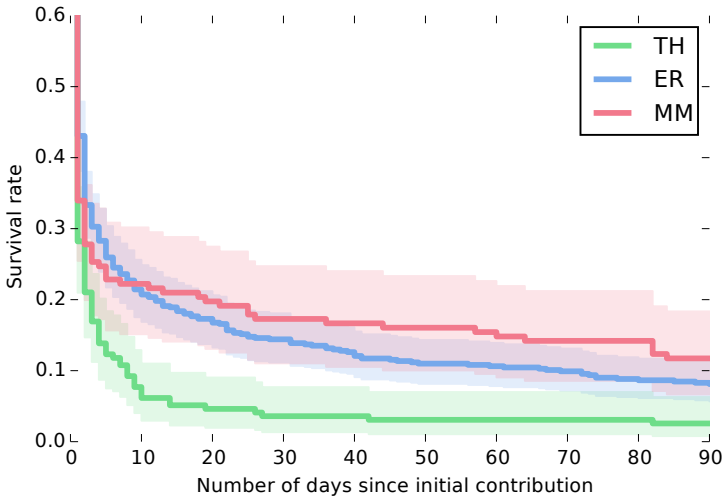

Figure 4.3: Survival functions for each cohort, indicating the rate at which participants with no prior OSM experience ceased contributing. Shaded regions show 95% confidence intervals.

Regression analysis was performed on all data, and for each cohort separately. The only significant model was found for the MM cohort (using ordinary least squares: adjusted R^2=0.37, F-statistic of 18.72, with $p < 0.001$ at 7 degrees of freedom). According to this model, individuals contributed for longer periods if they had prior OSM experience ($\beta_{d_{pre}}$=0.5), and were given fewer map object types to map ($\beta_{num\_concepts}$=-0.15, with a 95% confidence interval between -0.232 and -0.064).

These coefficients are based on standardised z-scores. The negative coefficient $\beta_{num\_concepts}$ indicates that the strongest response is below the mean value of *num_concepts*, which before standardisation is at 2.7. This result suggests that people remained active for longer on tasks that involved the mapping of less than three distinct map features.

Other models yielded no improvements in fit, including models which only included OSM newcomers, or only involved the two significant features d_{pre} and *num_concepts*. No regression model could explain the contribution rate c_{48h}.

4.5.3 RQ3: Contributor Performance

We found that faster contributors tend to remain active for longer: a correlation analysis found that the initial contribution rate c_{d1} is associated with longer participation in the enrolment period (Spearman correlation coefficient $\rho_S = 0.15$), and

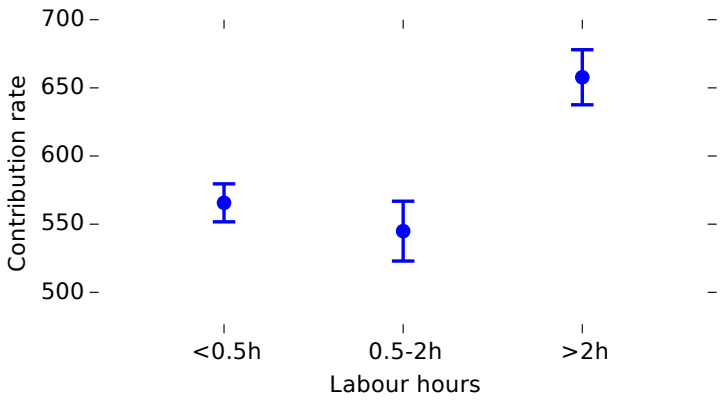

Figure 4.4: Mean contribution rate (c_{48h}) segmented by contributor engagement in the first 48h, with 95% confidence intervals.

with increased retention in month 2 ($\rho_S = 0.15$, both with $p < 0.01$), however this effect is slightly reduced in month 3 ($\rho_S = 0.11$, $p < 0.05$).

During a further analysis we found that contributors who abandoned tasks early and those who stayed the longest tended to contribute more quickly than those in between. We segmented contributors into three engagement classes based on their initial labour hours l_{48h}, this variable follows a normal distribution. Segmentation at the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentile yields three groups of short-term (under 30 minutes), average, and long-term contributors (2 hours or more). Figure 4.4 shows the distribution of contribution rates for each engagement class. Short-term contributors tended to contribute at a faster pace: their median contribution rate (650 edits per hour) was almost 30% higher than that of the average group (515 edits per hour). This effect was found in all cohorts. A pairwise comparison across the three bins with a KS statistic over the respective distributions of l_{48h} was significant for all pairs ($p < 0.001$).

On the other hand it cannot be said that there always are performance improvements within the first sessions: the median change in pace is around 0% between the first and second day of enrolment across cohorts. We found no correlation between performance improvements and long-term retention.

4.6 Discussion

In the following section we will summarise our results, highlighting the key effects we observed. We close with a set of implications and design recommendations.

The three cohorts have different volunteer constituencies. A comparison of their distributions of prior experience revealed markedly different distributions. This may be a result of a difference in engagement strategies: TH required a quick response by an existing community, while ER and MM could build volunteer capacity in a more sustained manner over longer periods. Furthermore they were active at different periods: by the time ER and MM were active, HOT had already gained some prominence outside of the OSM community.

The three cohorts have different contribution profiles during enrolment. The initial contribution profiles of TH and ER are quite similar, although TH contributors contribute at a slightly higher pace. In contrast, MM volunteers have the lowest rate of contributions, yet were working longer hours during initial enrolment than the other cohorts. The lower pace may be an effect of relative inexperience, while increased initial activity may indicate a greater motivation to contribute.

Coordination practices have a marked impact on contributor retention. Long-term retention was highest for the ER and MM cohorts which were specifically set up as more sustained campaigns, and which relied on a range of volunteer engagement practices such as mapathons and social media use. It was further found that for the MM cohort, short-term retention was lowest yet long-term retention was highest: contributors do not tend to come back on the second day, however they are more likely to remain active a month or two later. This may indicate a greater reliance on mapathons, where contributors do not necessarily return a day later, but at the time of the next mapathon event. In contrast to this, TH had a large number of contributors yet the lowest retention rates, likely because of an absence of any such sustained engagement practices. Its contributors may have been attracted by a perceived urgency around the disaster event, however they were not successfully retained for any subsequent HOT activity.

Complex task designs can be a deterrent for certain contributor groups: projects involving three or more map object types saw shorter activity periods for the MM cohort. In other words, more complex task requirements may be demotivating to

first-time contributors, regardless of their prior OSM experience. We found no evidence of an impact of documentation and guidance on engagement.

Most first-time HOT contributors tend to operate at a fairly steady pace, however contributors with prior OSM experience tend to work faster and stay a little longer. This effect is consistent across cohorts.

Early abandonment is associated with higher contribution rates. Volunteers who stopped contributing within the first 30 minutes tended to contribute at a slightly higher pace than those who remained active for an hour longer. This may suggests an instance of a burnout effect, where some first-time contributors begin their engagement at a relatively high pace but then lose motivation quickly. The effect was found in all cohorts.

Performance improvements are not associated with increased retention: contributors whose performance increased within the enrolment period were not necessarily retained for longer, suggesting that the presence of performance improvements is not an early indicator of increased long-term engagement.

4.6.1 Implications

Our aim is to understand engagement factors in crowdsourcing communities that help organisers reason about how to build and retain volunteer capacity. We have identified some initial aspects that can inform the design choices of HOT and related large-scale crowdsourcing systems.

With one exception, the task designs encountered in our study were found to be remarkably consistent in their engagement characteristics. One cohort saw a reduction in activity during the enrolment period for tasks that requested contributions of a higher complexity, which suggests that organisers should limit the number of map feature types requested per project.

This finding of relatively consistent contributor performance across designs may also indicate a limitation of our observational study: we do not compare radically different task designs, and instead merely observe existing tasks which rely on the same tools and interfaces for the actual contribution process. Many of these designs have already been informed by years of prior experience (Palen et al., 2015). Consequently, we encourage HOT organisers to also experiment with new task de-

signs to identify alternative strategies for further improvement. An example of this could be a micro-tasking interface that offers smaller and simpler tasks, which could allow newcomers to become productive more easily and quickly.[1]

An alternative interpretation is that current HOT contribution mechanics do not have a major impact on engagement, and that other aspects may be more important. In particular, our review of prior work suggests that intrinsic participant motivations, participant enjoyment, association with particular humanitarian causes, and social aspects of the HOT contribution experience may be more important to contributor engagement than the specifics of HOT contribution mechanics.

A further conclusion from our findings is that the best means of increasing output capacity is to grow the volunteer base, particularly considering the vast amount of uncharted territory that HOT aims to map. We could observe improvements in the performance of individual contributors, however an increase in contributors would raise output capacity more quickly.

Our findings suggest that the capacity-building strategies of the ER and MM campaigns work well, and we encourage organisers of other campaigns to adopt their practices: a combination of highly promoted projects over a sustained period, a steady stream of new efforts, regular mapathons and other training environments, and the use of email notifications in the MM campaign as a means of notifying interested contributors of new causes. In these two campaigns, a larger share of contributors kept coming back. In contrast to this, the greater urgency of the TH campaign (a response to a discrete crisis event) may have attracted many volunteers, however it did not contribute to an increase in retention.

These retention effects may further be affected by the marked difference in cohort constituents, attributable to self-selection effects and the presence of existing social ties. According to project descriptions, both ER and MM campaigns were largely coordinated by representatives of aid organisations, while TH projects were coordinated by OSM community members. This may have affected how the campaigns were promoted, and to whom. It is possible that ER and MM participants already had existing connections to aid organisations as part of other outreach ef-

[1] After this study was first published, Missing Maps organisers published MapSwipe, a micro-tasking mapping application for mobile phones. This was in part inspired by the work presented here.

forts, and as a result had a higher motivation to remain engaged.

It remains open what constitutes good training conditions for absolute newcomers. Organisers who seek to foster sustained newcomer engagement may choose to place newcomers in projects where they have a higher likelihood of being retained. In our case this would be the ER and particularly MM campaigns: projects that are specifically set up as long-term campaigns. Additionally, there are indications that particularly the MM campaign was successful at retaining and training absolute newcomers with no prior OSM experience.

4.7 Conclusion

We presented an initial study to assess the relationship between project designs and contributor engagement, with a specific focus on the experience of first-time contributors. We compared project designs along three analytical dimensions: cohort analysis, task analysis, and observation of contributor performance. Under consideration of these aspects we evaluated different projects in their ability to foster contributor engagement in the short and long term. The analytical dimensions yielded plausible findings: we found that different coordination practices did have a marked engagement impact, and that differences in task design can have an impact for some groups. Additionally, we found that prior domain experience in first-time contributors is likely to increase their engagement.

We argue that these findings have some external validity: we encounter similar contributor engagement across different cohorts, but also some differences, and we posit that these differences can in part be attributed to the observable factors discussed above. For this reason, we suggest that our findings are transferrable to other crowdsourcing systems.

The nature of this work and the available evidence however limit on our ability to build more nuanced understanding, and there are many aspects which we cannot gauge from the available data. For example, the contributor context is generally not known: in which situation does mapping take place? Yet the large scale of this study allow us to identify some general engagement trends that tend to be shared across different kinds of projects. In particular, our findings suggest that the participation setting may play an important role in contributor engagement: mapathons

and other social learning environments can play an important role in early contributor experiences. To investigate this relationship further, we present a study of the impact of mapathons on newcomer engagement in the following chapter.

Chapter 5

Mapathons

Findings from our first study suggest that the social contribution settings of mapathons might provide important support functions for newcomers. Such an effect would be supported by the theory, yet to date there are no empirical studies of it. We present the first study that evaluates the relationship between colocated practice and newcomer retention in a crowdmapping 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.

5.1 Introduction

Since its inception, MM (Missing Maps) organisers have refined practices that support the collective mapping effort. Among them are so-called mapathons, social event settings which allow regional community groups to come together in person, to socialise and to learn the practice. Organisers of these events pursue multiple 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. While the volume of mapping data produced at mapathons can be measured with existing tools, it is not currently clear how many attendees remain active afterwards. Do these events have a measurable impact on contributor reten-

tion? There is some evidence that communal event settings can play an important role in fostering sustained contributor activity. In our first study of HOT contributor engagement, we found that mapping initiatives which organised mapathons had higher newcomer retention rates (refer to Chapter 4 on page 63), however it is not yet known whether this can be attributed to the mapathon format itself.

Our second 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 crowdmapping, 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.

5.2 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), see Baecker (2014); Johansen (1988). 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 (Haythornthwaite et al., 2000). 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 (Fang and Neufeld, 2009; Trainer et al., 2014). 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 (Lave and Wenger, 1991). 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 (Trainer et al., 2014). 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 (Steinmacher et al., 2014). 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 (Fang and Neufeld, 2009).

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 (Choi et al., 2010). 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 (Ciampaglia et al., 2015). However the authors also observe that the causal structure between socialisation, motivation, and participation is not entirely clear. Further studies identify similar effects (Burke et al., 2009; Choi et al., 2010; Farzan et al., 2012).

5.2.1 Knowledge Gaps

To our knowledge there are no quantitative studies of colocated practice in a global crowdmapping 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 (Hristova et al., 2013; Perkins and Dodge, 2008). It is not clear how contemporary HOT mapathons compare to these earlier settings.

5.3 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 (Clary et al., 1998; Nov, 2007; Schervish and Havens, 1997; Wilson, 2000). However, there is no empirical evidence available to confirm such an effect in a crowdmapping context. We seek to establish such evidence by analysing contributor activities of the HOT 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.

5.4 Methodology

A primary source of empirical evidence is the HOT contribution history, this data set and its derivative forms are described in detail in Section 3.2 on page 55.

We captured the contribution activity of first-time mappers belonging to two separate mapathon cohorts in London, observing a total of 14 events. To assess their overall impact in relation to a baseline expectation, we compared their subsequent newcomer retention with that of an online control group, a set of first-time online contributors who did not attend the London mapathons. We further sought to isolate the specific impact of the mapathon format, as opposed to general colocated practice. To this purpose we further compared the mapathon cohorts to a second control group of participants of the Arup "Mappy Hour", an employee-initiated regular mapping event that does not strictly follow the same format. 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.

5.4.1 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.

5.4.2 Study Cohorts

5.4.2.1 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.

[3] `http://www.missingmaps.org/events/`

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

Table 5.1 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.

5.4.2.2 Online Control Groups: Matched MM and Nepal

As first online control group we identified HOT contributors who engaged in comparable work, 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 crowdmapping 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. Our prior work found that such contributors are less likely to be retained for long periods (refer to Chapter 4 on page 63). 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.

5.4.2.3 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] `http://doggerel.arup.com/mapping-the-worlds-most-vulnerable-regions/`

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 5.2: Mapathon features collected per event.

5.4.3 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. 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 5.2. 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.

5.4.3.1 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 identified 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.

5.4.3.2 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.

5.4.3.3 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.

5.4.3.4 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.

5.4.4 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 5.3.

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'.

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 5.3: Attendee features computed per mapathon newcomer.

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 5.4.3. It seeks to explain the newcomer activation and retention measures listed in Table 5.3 by considering the event setting, attendee mix, and attendee tool choice as shown in Table 5.2. 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.

5.5 Results

5.5.1 RQ1: Newcomer Retention

Activation and revisit rates across the 14 mapathons are shown in Table 5.4. 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

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 5.4: Activation and repeat attendance rates among first-time mapathon attendees.

Kaplan-Meier estimate with a 98% confidence interval. A corresponding survival plot is shown in Figure 5.1, 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 5.2 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.

5.5.2 RQ2: Newcomer Retention Factors

Table 5.5 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

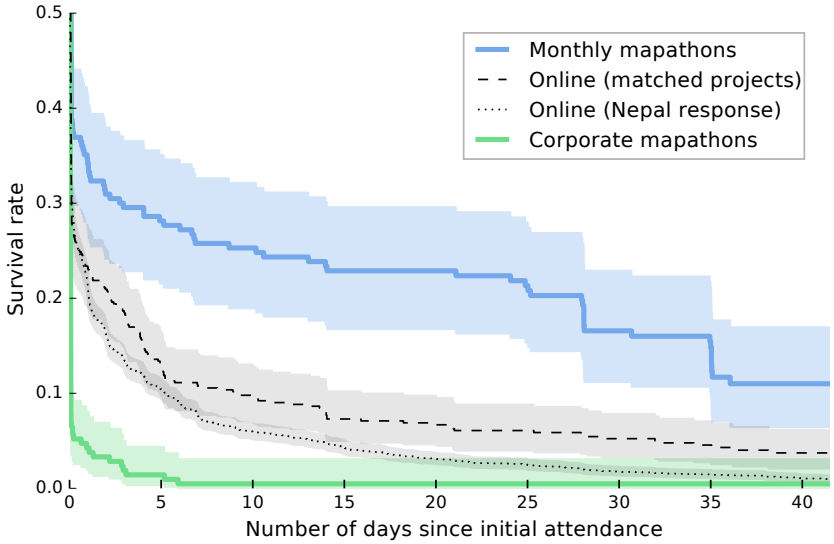

Figure 5.1: Newcomer survival rate for mapathon and online cohorts, with 98% confidence intervals.

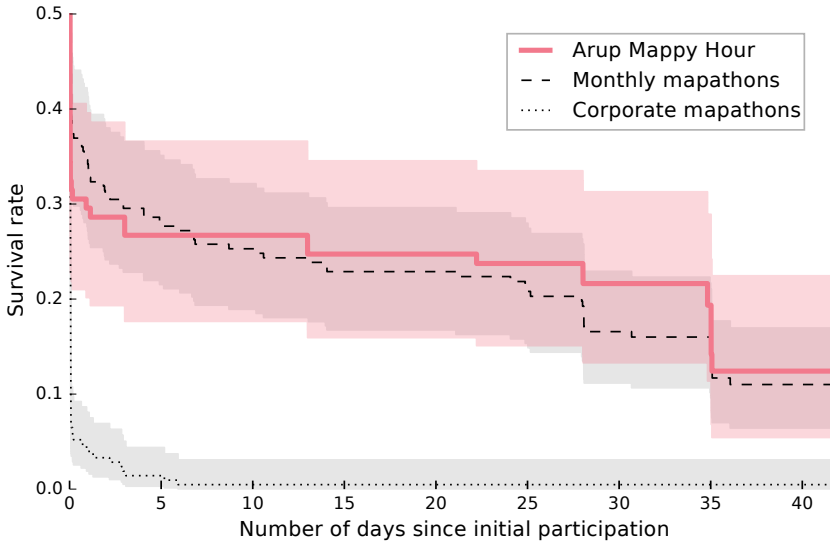

Figure 5.2: Newcomer survival rate for mapathon and Arup Mappy Hour cohorts.

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 5.5: Significant correlations between mapathon features and event outcomes, for monthly mapathons only.

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 bring about 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.

5.6 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 5.1 (page 96) and Figure 5.2 (page 96), 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.

There are indications for further contributing factors that relate to the setting, however many of the causal relationships remain unclear. A correlation analysis found that technical problems at the venue may harm task completion rates by newcomers, however this did not measurably affect subsequent retention. Attendee mix and social food breaks had no measurable effect on retention. Neither of these effects could be confirmed with a regression model, which suggests that other unobserved factors were more important.

5.6.1 Implications

Our results indicate that the most highly engaged cohorts were groups that had a regular meeting schedule, and that incorporated some aspects of peer learning. Furthermore, we found evidence that there may be self-selection effects, or effects involving mastery experiences: people who finished tasks in their first session were more likely to remain engaged, regardless of the setting. This is a common phenomenon in online communities and other social settings: people who are already engaged are more likely to remain engaged. In other words, highly prolific contributors may be "born, not made" (Panciera et al., 2009). If this is the case here, then differences in cohort outcomes may relate to choices in recruiting strategies by organisers, and self-selection effects among participants.

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.

5.7 Conclusion

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.

Organisers can design interventions on the basis of our findings. 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.

Many causal factors remain unclear. Based on the study outcomes we posit that there may be important effects related to participant recruitment and self-selection. Social contribution settings for crowdmapping communities may in part be successful because they serve as an attractor: they provide and promote opportunities to meet enthusiastic contributors, and by doing so they can capture prospective contributors who have a latent interest in the practice.

We further hypothesise that important unobserved effects may relate to access to peer learning, social validation and imitation, mastery experiences, and other aspects relating to the early learning trajectory of individual participants. We investigate these concerns in the following chapter, in a study of the impact of peer feedback during HOT validation.

Chapter 6

Private Peer Feedback

The previous study has shown that social contributions environments can provide important support for newcomers, however the specific reasons are unclear. We hypothesise that an important aspect of such settings is the provision of a supportive learning environment that allows individuals to attain mastery experiences through peer support and social validation. We investigate this relationship further with a study of the effect of peer feedback during HOT validation, a structured interaction that takes place online. Are newcomers more likely to be retained when they are given certain kinds of feedback? Should validators be concerned about the engagement impact of their feedback, and spend effort on the content and tone of their messages? To address these questions, we observe the impact of different classes of feedback, including performance feedback, corrective feedback, and verbal rewards, and control for factors relating to the individual and the contribution context.

6.1 Introduction

The mapping process with HOT includes a peer review workflow in the form of a validation stage (refer to Section 2.1.3 on page 34). During validation, contributors can self-nominate to review the work of other contributors. They can mark the work as accepted or rejected, and optionally send a private message to the contributor. This validation stage can provide an important opportunity to receive early guidance and encouragement from a peer, possibly a more experienced contributor. This may have important engagement effects, for example by improving newcomer self-efficacy. Conversely, it is feasible that negative feedback and the rejection of

contributions may be discouraging to newcomers.

Studies of peer feedback in social media and online community platforms have shown that public feedback mechanisms can have unintended long-term consequences. Strongly negative contribution feedback may be discouraging to newcomers (Halfaker et al., 2011), and public peer ratings and commenting systems can yield harmful feedback effects affecting contributor behaviour (Cheng et al., 2014; Michael and Otterbacher, 2014). Such effects may arise in part because public peer feedback fosters a desire to build reputation and social standing (Tausczik and Pennebaker, 2012; Parnell et al., 2011). Do these outcomes differ on platforms where peer feedback is given in private?

We analyse the effects of private HOT validation feedback on newcomer engagement, and relate our findings to known effects of other platforms. Are HOT newcomers more likely to be retained when they are given certain kinds of feedback? Should validators be concerned about the engagement impact of their feedback, and spend effort on the content and tone of their messages? Are there opportunities to use validation feedback strategically to improve newcomer retention? How do these outcomes relate to platforms where feedback is given in private?

6.2 Related Work

Prior work on the engagement impact of contribution feedback can be found in a wide range of settings. Research in behavioural psychology can provide general expectations for the impact of different kinds of feedback. In the more specific context of social media and online platforms, further effects have been identified for social review sites, Q&A sites, and peer production platforms. Research on the role of feedback in education can provide additional expectations for settings where feedback is not socially negotiated.

6.2.1 Contributor Motivations as Engagement Drivers

Kraut et al. (2012) provide a comprehensive discussion of a wide range of empirically tested theories to help interpret engagement effects in the context of online community platforms. Among the many factors that are discussed, contributor mo-

tivations are one of the most important drivers for sustained participation. When they are appropriately considered in platform design, they can provide triggers for action, and support rewarding experiences that can sustain contributor engagement for long periods. However, when the contribution experience comes in conflict with existing contributor motivations, participants are unlikely to remain active.

In the motivational theory, a distinction is made between intrinsic motivations by the participant, such as curiosity or a desire for social standing, and extrinsic motivations that are provided to the participant by others, for example in the form of rewards or punishment. The relationship between the two is complex, and subject to a number of moderating factors (Cameron et al., 2001). **Verbal rewards** such as expressions of appreciation and gratitude by peers have been shown to strengthen intrinsic motivations. Conversely, tangible rewards such as badges or access to higher social status do not always have the desired effect. For tangible rewards, **performance-contingent rewards** which are linked to performance are found to be more effective than task-contingent rewards, which are merely linked to completion of a task (Harackiewicz, 1979; Cameron et al., 2001).

In volunteer communities, *constraints in the capacity and ability* of participants may pose further barriers to sustained participation. Newcomers can be demotivated more easily by strongly negative feedback, compared to participants who are more experienced (Halfaker et al., 2011; Zhu et al., 2013; Wohn, 2015). It may be possible to address this difference with targeted interventions to improve self-efficacy, for example by nurturing and educating new contributors, and by fostering a belief that participation is possible and will be welcomed (Bishop, 2007).

6.2.2 Feedback and Ratings on Online Platforms

In social media and online community platforms, motivational aspects are often further moderated by social factors. This is particularly the case when feedback for contributions is made public, or when it is socially negotiated. As a result, contribution feedback in these settings can have unintended long-term consequences (Cheng et al., 2014).

Empirical findings to date reveal a complex picture. On one hand, peer feedback can be a basis for important *social learning effects*. A study of Q&A sites finds

that votes, favourites, and answers to questions provide a form of contributor feedback which can improve future question quality. The study further finds that effects are dependent on the content of such feedback, not merely the volume of feedback (Ahn et al., 2013). 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 (Choi et al., 2010). An effort to provide easy access to peer advice in the form of a Q&A forum yielded a significant increase in newcomer retention (Morgan et al., 2013). Other Wikipedia efforts to provide opportunities of social encounter yielded comparable outcomes (Kittur et al., 2009; Musicant et al., 2011). This suggests that to new participants, access to social encounter can make a significant difference in their decision to keep contributing.

On the other hand, organisers need to consider a *trade-off between the encouragement of participants and improvements to contribution quality*. Positive and social feedback can increase participant motivation (Zhu et al., 2013), however strongly negative feedback can harm motivation. On Wikipedia, reverts to article modifications can improve article quality, but they are powerfully demotivating to newcomers (Halfaker et al., 2011). As a consequence it may be advisable to provide more nuanced feedback, rather than an outright rejection of an entire contribution. When negative feedback is less strong and more directive, for example by providing recommendations for future improvement, it can improve task performance without harming motivation (Zhu et al., 2013).

In some settings, negative feedback can have *long-term social effects relating to community regulation*. In a rating system for online news, early negative feedback was linked to significant behavioural changes that are detrimental to the community. Negatively reviewed authors became more likely to review others negatively, while positive reviews had no such effect. Authors who received no feedback were most likely to leave a community, suggesting that the social affirmation of positive feedback may be an important prerequisite for sustained engagement (Cheng et al., 2014). Similar feedback effects were found on an online travel site. Here, early negative reviews of a venue affected the tone of subsequent reviews, resulting in a herding effect (Michael and Otterbacher, 2014).

Public feedback can further affect the *reputation and social standing* of partici-

pants. On Q&A sites, a pursuit of higher social recognition can be an important driver for participation, and reputation-building strategies are widespread across a range of participants (Tausczik and Pennebaker, 2012; Parnell et al., 2011). This is even more pronounced in online market places, competitive environments where feedback is given strategically in order to gain a market advantage (Resnick et al., 2000; Dellarocas, 2003).

6.2.3 Feedback Models in Education

Hattie and Timperley (2007) provide an overview of contemporary theories of feedback in education. Of particular relevance to this third study are empirical expectations of learner engagement after receiving different types of feedback (Hattie and Timperley, 2007).

Performance feedback includes feedback about how well a task is being accomplished, for example by distinguishing correct and incorrect answers (Harackiewicz, 1979). Positive performance feedback on a task of high interest to the learner can increase future motivation (Deci et al., 1999), or can increase self-efficacy for learners who showed low initial performance (Hattie and Timperley, 2007). In contrast, negative performance feedback can reduce motivation for participants with low self-efficacy or low initial performance (Moreland and Sweeney, 1984; Brockner et al., 1987).

Corrective feedback includes specific and directed suggestions for further actions, for example by helping to distinguish correct from incorrect answers, and by providing more information that aids understanding and helps build procedural knowledge. When it corrects faulty interpretations, corrective feedback can improve self-efficacy and enhance existing motivation (Lysakowski and Walberg, 1982; Tenenbaum and Goldring, 1989).

The **timing of feedback** can play an important role, both in micro and macro scales. On the micro scale of individual tasks, immediate feedback was found to be a more effective driver for improved future task performance (Hattie and Timperley, 2007). On the macro scale of a learner's journey, feedback was found to be more impactful at the beginning: newcomers are particularly receptive to external feedback, and more eager to adopt social norms and practices of their new environ-

ment (Ashforth and Saks, 1996).

In online education, certain types of **private peer feedback** are associated with higher attainment. In a study across multiple online education platforms, students reported an increase in course enjoyment from peer feedback, and stated that the resulting exposure to alternative problem-solving approaches improved their learning experience (Adamopoulos, 2013). Among high school students interacting through an online platform, positive affective peer feedback such as socio-emotional support was shown to increase student motivation and self-efficacy, which in turn can increase their performance. However, peer grading and task feedback on the platform had no measurable effect on either motivation or future performance (Lu and Law, 2012).

6.2.4 Knowledge Gaps

Much existing research of peer feedback in social media and social knowledge production platforms is focused on social feedback mechanisms that are public, such as peer ratings and commenting systems. It was shown that public feedback mechanisms are subject to social feedback effects including herding, and that they can foster a desire to build reputation and social standing.

In comparison, in HOT the review process is closer to a teacher-learner model: feedback is private and from a single reviewer, rather than socially negotiated. As a consequence, it is feasible that the absence of strong social feedback mechanisms changes the effects of the received feedback. To our knowledge, no existing empirical studies assess the relationship between peer feedback and subsequent newcomer retention in such a novel setting. Do outcomes differ from the online and education settings observed in previous studies?

Our large-scale study contributes new knowledge which can support the analysis and design of social knowledge production platforms. In particular, we provide an empirical expectation of the effects of private peer feedback on newcomer retention, and discuss how this relates to public peer feedback. Better understanding of such effects can support the design of improved feedback mechanisms, for example by avoiding negative social feedback loops.

6.3 Research Questions

We regard validation feedback as an intervention that may have an effect on subsequent newcomer retention, and seek to determine under which circumstances the intervention becomes effective. We distinguish between the *validation verdict*, which is the decision whether a given contribution was accepted or rejected, and the *validation message* which can accompany a verdict with more detailed feedback. Together they constitute the validation feedback.

6.3.1 RQ1. What is the Impact of the Validation Verdict?

How important is a positive validation outcome for newcomer engagement? Can an early rejection discourage further participation? The theory suggests that when intrinsic motivation is present, positive task feedback can foster repeat participation (Deci et al., 1999; Hattie and Timperley, 2007). However such feedback should be performance-contingent, not merely task-contingent: positive feedback is only effective if it is linked to contribution performance (Harackiewicz, 1979; Cameron et al., 2001). In contrast, negative feedback can reduce motivation when initial self-efficacy is low (Moreland and Sweeney, 1984; Brockner et al., 1987). Are these effects observable for newcomers in HOT?

6.3.2 RQ2. What is the Impact of the Validation Message?

We consider the content of validation messages beyond the verdict alone, and distinguish between feedback directed at the task performance, and feedback directed at the self. In relation to task performance we observe uses of positive and negative performance feedback (Deci et al., 1999; Hattie and Timperley, 2007), and of corrective feedback (Lysakowski and Walberg, 1982; Tenenbaum and Goldring, 1989). We further observe uses of verbal rewards (Cameron et al., 2001), a kind of affective feedback which includes signs of appreciation and encouragement. Do these different types of feedback have an impact on subsequent newcomer engagement?

6.3.3 RQ3. What is the Impact of Delayed Feedback?

Does the timing of the verdict matter? According to the theory, task-specific feedback is more likely to be effective if it is given early (Hattie and Timperley, 2007). Is overall retention affected when feedback is delayed?

6.4 Method

6.4.1 Dataset

Our evidence is derived from two data sources. The annotated participation history developed in Section 3.2 forms the basis of our analysis. In addition, we make use of the validation history recorded by the HOT Tasking Manager, including validation verdicts and validation messages. This data is not made public, it was provided to us by HOT for the purpose of an evaluation study.

This yields an annotated participation history which forms the basis of our analysis. In an initial stage, we identify a set of basic features for every contributor. The feature *join_day* records the date of a newcomer's first HOT contribution, in days since 1st January 2014. It can serve as control variable to capture changes in newcomer retention over time.

6.4.2 Study Population

We observe first-time HOT contributors with no prior OpenStreetMap experience. We start our observation period in October 2014, the first month where HOT contributors were shown a notification for all validation messages in the Tasking Manager interface. We only include newcomers who join at least 90 days before the end of the available edit history at the time of study. This allows us to observe their retention over time. In particular, we select first-time contributors to HOT who:

- first contributed between 1st October 2014 and 20th June 2016 (both inclusive),
- have no more than one prior day of OpenStreetMap contribution experience outside of HOT,
- have a task validated within the first 45 days of joining,
- were sent a validation message along with their verdict.

Almost 1,300 contributors match these criteria. For each newcomer, we identify the first validation message they received, these form the basis of our study. For each of the 1,300 messages we determine *delay*, the delay between submission and feedback message in days, and the binary indicator *low_delay* which marks messages that were received before the median delay of 28 hours.

For these initial cases, the average task acceptance rate is 34%. We do not consider any subsequent messages newcomers may have received: our study is focused on the impact of early interventions during a newcomer's participation history, in part to avoid confounding factors in later participation stages, but also to ensure consistency in our analysis. Furthermore, it is comparatively rare to receive multiple messages: only 15% of participants received more than one message throughout their participation lifetime, and only 3% received more than 5.

6.4.3 Message Classification

We characterise the content of the 1,300 validation messages by means of a thematic analysis. Informed by a review of the literature we distinguish multiple complementary aspects of learner feedback: positive and negative performance feedback (Deci et al., 1999; Hattie and Timperley, 2007), corrective feedback (Lysakowski and Walberg, 1982; Tenenbaum and Goldring, 1989), task-contingent rewards (Harackiewicz, 1979; Cameron et al., 2001), and the use of verbal rewards (Cameron et al., 2001). They are not mutually exclusive: each message can contain combinations of these.

Validation messages tend to be short, at an average of 100 characters, which makes it feasible to manually label the full corpus. Each message is reviewed to determine whether it satisfies the specified criteria for one or more of these feedback techniques. This labelling process is only concerned with the message content. We further discard 74 messages which were written in languages other than English.

To ensure labelling consistency, in an initial labelling stage we iteratively develop a codebook that is then used as a reference. This codebook includes a definition and example phrases for every code. To build it, we first label a random sample of 100 and then 500 messages, collecting example phrases for each code, and clarifying our definition statements when necessary. We repeat the process un-

til we reach a saturation point where the definition statements and example phrases allow an unambiguous distinction between all codes across the samples. We then manually label each message in the full corpus, recording the presence or absence of each code. To increase confidence, a second rater labels a randomly sampled subset of 500 messages, after receiving one hour of training. We measure rater agreement with Cohen's Kappa for each individual code. Across the labelled codes, we find an agreement between 0.82 and 0.97 among raters, these Kappa scores fall within the range of almost perfect agreement (Landis and Koch, 1977).

These codes will be discussed in turn. Three codes describe *feedback that is focused on the task*:

- Positive Performance Feedback (PPF). Phrases that express a *positive assessment of the merit* of the current contribution. Example phrases include: good job, great work, looks good, this is high quality, good interpretation of the imagery, nice mapping, etc.
- Negative Performance Feedback (NPF). Phrases that express a *negative assessment of the merit* of the current contribution, for example by highlighting omissions or mistakes. Example phrases include: doesn't look complete, missing, needs improvement, not done yet, not mapped, not very accurate, still not traced, untagged, needed squaring, etc.
- Corrective Feedback (CF). Phrases that provide *procedural guidance* about the contribution process, typically to improve future work. This can include explicit requests for specific acts, including detailed instructions, or references to external documentation. It can include implicit request for action, for example a negative merit assessment. It can also include invitations to consider a particular concern, clarify community expectations, or other means of fostering deeper understanding. Example phrases include: could you, can you correct, don't use, double check that, read the instructions, learn how to, look for, make sure, please complete, missing, untagged, please do not, please trace, should be drawn, should be mapped, etc.

During early iterations, we further observed a significant number of instances where validators accepted contributions which contained errors or were otherwise incomplete. Instead of rejecting them, they corrected the mistakes themselves,

marked the task as 'accepted', and attached a message where they summarised their refinements. This can be seen as an example of a *task-contingent reward* (Cameron et al., 2001), in an attempt to avoid rejecting a newcomer's early work. We introduced an additional code for such cases to assess their impact:

- Generous Acceptance (GA). Tasks that have been marked as accepted, but where the validator message includes either CF or NPF.

A correlation analysis confirmed that GA places additional burden on the validator: newcomers whose validation messages is labelled with GA had a larger share of their contributions removed by a validator (Spearman correlation coefficient $\rho_S = 0.32$, $p < 0.001$). In comparison, "strict" acceptance without CF or NPF also incurred such deletions, but to a lesser degree ($\rho_S = 0.10$, $p < 0.001$).

Finally, we include *verbal rewards* as a feedback category that is more broadly focused on the contributor, and which can help strengthen self-efficacy in early learners. For the purpose of this study, we consider verbal rewards to be any feedback that shows appreciation and encouragement, including PPF:

- Verbal Rewards (VR). Phrases that express *positive assessment of merit, gratitude, or encouragement*. This includes any instance of PPF. Further example phrases include: thank you, thank you for mapping, thanks for your contribution, keep mapping, keep up the good work, please don't hesitate, please keep mapping, etc.

Figure 6.1 shows a frequency distribution of all feedback techniques for the full message corpus. The distributions show that most messages are corrective: CF and NPF are the most widely used feedback types, they are included in approximately 80% of all messages. NPF almost always is used as part of a correction (i.e., along with CF), rather than merely by itself. This suggests an overall constructive tone of the peer feedback. GA is the least widely used technique, at 28%. Only 4% of messages include none of the feedback techniques and remain unlabelled, many of these are instances where submitters accidentally marked a task as completed.

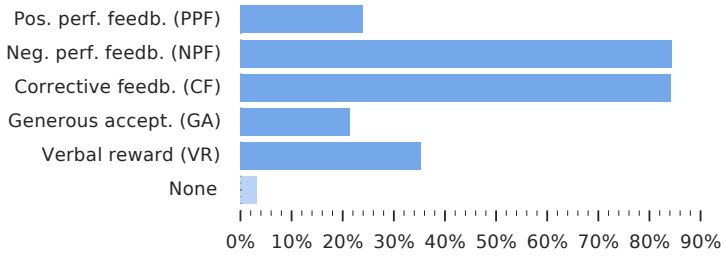

Figure 6.1: Frequency distribution of feedback techniques throughout the message corpus.

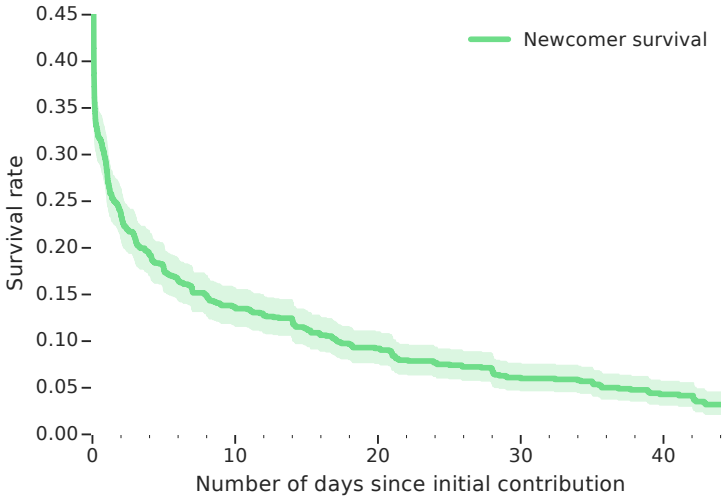

Figure 6.2: Base survival function for all study participants, with 95% confidence interval.

6.4.4 Newcomer Retention

Newcomer retention is measured by means of a survival analysis, capturing activity during an observation period of 45 days. We record the relative date of the last observed contribution as feature *last_day*, measured in days since the initial contribution. We confirm death events (censure) by ensuring that there is no activity within the next 45 days after this initial observation period, this is recorded as binary feature *has_died*.

We fit a Kaplan Meier survival model to establish base survival rates, these provide a basic expectation of overall newcomer retention. The resulting survival plot is shown in Figure 6.2. The base expectation of survival is low: only 14% of newcomers are projected to still be active 10 days after their initial contribution, 9% after 20 days, and 6% after 30 days. The 95% confidence interval for this model is

within ±3% throughout the observed period, suggesting a high model fit. In other words, the base expectation is that newcomers will not be retained, regardless of any intervention they may experience.

6.4.5 Contextual Factors

While we primarily seek to observe the impact of validation feedback, we also need to account for any other factors that can influence newcomer retention in a significant manner. This includes contributor-specific factors such as intrinsic motivation, and any project-specific factors which may have an effect on subsequent retention.

Much prior work on online communities has shown that intrinsic motivation and commitment to goals is a strong predictor of future engagement (Kraut et al., 2012). We cannot easily capture contributor motivation at large scale, however we can estimate it based on initial contribution behaviour. To this purpose we measure the *initial effort* as *initial_hours*, the number of hours for which each newcomer contributes to HOT during their first 24 hours of activity. A correlation analysis confirms that this measure is highly correlated with the subsequent survival period *last_day* (Spearman correlation coefficient $\rho_S = 0.50$, $p < 0.001$). This makes it a suitable proxy measure for intrinsic motivation. We further determine *high_initial_hours*, a binary indicator to identify newcomers whose initial effort was above the median of 74 minutes.

Chapter 4 and Chapter 5 have shown that urgent disaster-related HOT campaigns can be significant recruiting events, but tend to have lower average retention rates than other more sustained campaigns. To account for this, we include a final control variable *disaster_campaign* to capture whether a newcomer's first HOT contributions were during such a campaign. (A principled empirical justification for this effect is provided in Chapter 7.)

6.4.6 Analysis

We regard validation feedback as an intervention that may have an effect on subsequent newcomer retention, and seek to determine under which circumstances the intervention becomes effective.

A Cox proportional hazards model is used to explain the survival rates we ob-

Aspect	Variable	Description
Project	*disaster_campaign*	Disaster event or more sustained campaign?
Contributor	*join_day*	Join date, in days since 1st January 2014
	initial_hours	Initial effort: hours spent mapping during initial 24h
	high_initial_hours	Was the initial effort above the median (74m)?
Validation outcome	*accepted*	Was the contribution accepted?
	delay	Delay between submission and feedback, in days
	low_delay	Was feedback delay below the median (28h)?
Message content	*PPF*	Positive performance feedback?
	NPF	Negative performance feedback?
	CF	Corrective feedback?
	GA	Generous acceptance?
	VR	Verbal rewards?
Contributor retention	*last_day*	Last active day, in days after joining
	has_died	No further activity for at least 45 more days?

Table 6.1: Features collected per first-time contributor.

served, based on a set of features and control variables. This is comparable to a regression analysis, but specifically intended to model participant survival. In the context of this study, the term 'hazard' is a synonym for the risk of abandoning HOT participation. A hazards model yields a rate of risk for each covariate, denoting the relative increase in hazard per unit increment. The model fit for such a model is described by its concordance, where 0.5 is equivalent to the performance of a random predictor, and 1.0 denotes perfect prediction accuracy. We further compute 95% confidence intervals for all covariates. In cases when an effect cannot be found, we employ a power calculation method proposed by Freedman (1982) to determine whether our models have sufficient statistical power to identify an effect. A Cox model relies on two basic assumptions: first, that all covariates are multiplicatively related to the hazard, capturing an effect relative to an initial baseline hazard. Second, that all observed effects are time invariant: all covariates are constant throughout the observation period. In the present study, this assumption can be justified based on a conceptual observation: we seek to model the impact of a specific intervention at the start of the observation period.

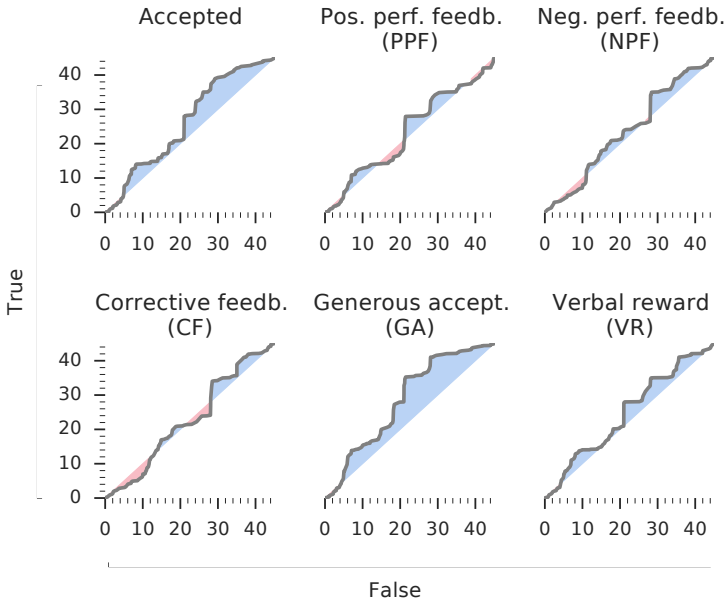

Figure 6.3: Q-Q plots of survival rates for different feedback types. Each plot compares the distribution of those who have received a particular type of feedback (*True*), and those who have not (*False*). Survival is measured in days after joining.

We derive a feature vector per first-time contributor using the available features and control variables. Table 6.1 on page 114 shows the full feature vector per newcomer which is used for our analysis, different subsets of these features are used to test specific effects. All models include the covariate *disaster_campaign* to control for project-specific factors, and *join_day* and *initial_hours* to control for engagement factors related to the contributor. These variables have been discussed in previous sections. This basic model has a condition index of 2.5, indicating low multicollinearity, and a concordance of 0.68. A 10-fold cross-validation shows that the model is stable when presented with different subsets of the data: the median concordance across iterations is 0.68, with a low standard deviation of 0.02.

To set a basic expectation of potential outcomes, the Q-Q plots in Figure 6.3 visually compare the survival distributions associated with different message classes. These figures show that acceptance of a task, generous acceptance, and verbal rewards are associated with longer survival periods, while the remaining features are not clearly associated with a single outcome. These are mere associations, hazard models in later sections will confirm whether they indicate real relationships.

6.4.7 Limitations

Although we make an effort to reduce the chance of confounding effects, this observational study can only reveal associative rather than causal relationships. We are further limited to an assessment of existing practices, rather than an exhaustive assessment of all possible practices.

Our study is specifically focused on contributor engagement, and an analysis of the impact on task performance is out of scope, including an assessment of the impact on future contribution quality. Analyses of such effects are confronted with methodological challenges that exceed the scope of the present study. In particular, it is not evident how contribution quality can easily be measured at the scale of this study. A fundamental challenge is the absence of a reliable ground truth: by definition, the maps produced by HOT are the first of their kind, and there typically are no other maps to compare against. As a consequence, the few early studies of HOT data quality either rely on specific contextual knowledge of particular geographies, or they assess quality by reproducing the volunteer work with other means, both of which limits their scale (Westrope et al., 2014; Eckle and de Albuquerque, 2015).

Due to a limitation in the setting, we are not able to determine whether certain messages would reactivate contributors who have previously been inactive. Messages sent during the validation process do not result in an email notification, instead they are displayed prominently on the Tasking Manager the next time a contributor logs in. In the current form, HOT messaging requires contributors to actively log in in order to receive a new message. As a consequence, we can only assess the impact of validation messages on contributors who return at least once.

6.5 Findings

6.5.1 RQ1. Validation Verdict

In order to assess the engagement impact of the validation verdict, we build multiple Cox hazard models to compare the effects of three different validation verdicts: acceptance, generous acceptance, and rejection.

An initial hazard model compares the impact of acceptance and rejection outcomes across the full study population. The feature vector includes all control vari-

Aspect	Variable	Hazard	p
Project	*disaster_campaign*	133.6%	< 0.001
Contributor	*join_day*	100.0%	< 0.001
	initial_hours	82.8%	< 0.001
Feedback	*accepted*	(94.6%)	0.43

Table 6.2: Proportional hazards model for the effect of task acceptance on newcomer retention, when compared with task rejection.

ables, and *accepted* as a predictor.

The impact of GA is tested with two separate models. GA is first compared with 'strict' acceptance, these are instances where the validation message did not include instances of NPF or CF. This test is computed across those in the study population whose contributions have been accepted, and *GA* is used as the predictor. A second test compares GA with rejection, here the test is computed across those who have received a GA or rejection verdict, and *GA* is the predictor.

For each of these three basic models, we further derive three variants: the full model, and versions of the model that only include high-effort or only low-effort contributors, as determined by *high_initial_hours*. This allows us to identify effects that are dependent on intrinsic contributor motivation, as expressed by their initial contribution effort.

We find that none of the models show an effect associated with the respective predictor. Power analysis suggests that we may simply not have sufficient data to establish an effect: in each model, the statistical power of the feature of interest is below 0.2 ($p<0.05$). Instead, the strongest effects involve contributor-related factors: in all models, the control variable *initial_hours* is significant. In some models, the control variable *disaster_campaign* is significant, indicating engagement effects relating to contributor recruitment.

Table 6.2 shows an example of such an outcome, in this case for a model comparing acceptance and rejection, with a concordance of 0.679. In this model, the intervention covariate *accepted* is not found to be significant. This covariate only has a statistical power of 0.14 ($p<0.05$), which suggests there is insufficient evidence

Aspect	Variable	Hazard	p
Project	*disaster_campaign*	168.9%	< 0.001
Contributor	*join_day*	(100.0%)	0.14
	initial_hours	68.1%	< 0.01
Feedback	*VR*	78.0%	< 0.05

Table 6.3: Proportional hazards model for the effect of verbal rewards (VR) on newcomers who contributed for less than 75 minutes on their first day.

available to identify a significant effect.

In other words, we find no evidence that the task verdict has an impact on newcomer retention. In particular, we find no evidence that rejection of work harms retention, or that task-contingent rewards through GA improve retention.

6.5.2 RQ2. Validation Message

To assess the impact of the validation message, we determine the engagement effect of each feedback type. We build a separate model for each PPF, NPF, CF, and VR, using the respective feedback type as a predictor. To identify interactions for sub-populations, we further build variants of these models to only include those who have been accepted, those who have been rejected, and those who showed high or low initial effort.

We find a significant effect for VR, at a concordance of 0.644. The full model is shown in Table 6.3. According to this model, for newcomers whose initial contribution effort is below the median, VR reduces the hazard rate to 78%. In other words, among contributors who contributed for less than 75 minutes on the first day, those who received a message with a verbal reward were more likely to remain active in HOT. Compared to the baseline, their likelihood to stop contributing drops to 78%. The 95% confidence interval for this feature provides further support for the presence of an effect (lower threshold at 62.4% hazard rate, upper threshold at 97.6% hazard rate).

For models involving PPF, NPF, and CF, the predictors are not significant. In all other cases, power analysis for the feature of interest never exceeds 0.3 ($p<0.05$),

Aspect	Variable	Hazard	p
Project	*disaster_campaign*	133.3%	< 0.001
Contributor	*join_day*	100.0%	< 0.01
	initial_hours	82.4%	< 0.001
Feedback	*low_delay*	78.6%	< 0.001

Table 6.4: Proportional hazards model for the effect of early feedback on all newcomers.

suggesting a lack of sufficient evidence to identify a significant effect.

Models for pairwise combinations of all features are similarly inconclusive, with two exceptions: for newcomers whose initial contribution effort is below the median, the pairwise combinations of VR/CF and VR/NPF both reduced the hazard rate to 77% and 76%, respectively ($p<0.05$, both at concordance 0.646). They can be regarded as more specific extensions of the previously observed effect involving only VR.

6.5.3 RQ3. Timing

To assess the impact of feedback timing, we first test whether early feedback was associated with improved retention. Across the validation messages analysed for the study, feedback was sent relatively quickly. 40% of messages were sent within 12 hours after task submission, the median was at 28 hours. The 25% slowest validation responses were sent a week or more after submission.

A hazards model for early feedback is shown in Table 6.4. It includes *low_delay* as a predictor, this feature denotes whether a newcomer received their first feedback within 28 hours after submitting. The model shows that early feedback matters: it yields a reduction of the hazard rate to 79%, at a model concordance of 0.672. The 95% confidence interval for this feature further supports this (lower threshold at 69.9%, upper threshold at 88.3%).

A separate model for *delay* as predictor showed that each additional day of delay incurs a hazard increase of 1.4% (95% confidence interval of 0.7% to 2.1%).

6.6 Discussion

Across all models we computed, the most consistently predictive factor for high contributor retention was a contributor's initial contribution effort, as measured by *initial_hours*. The second-most frequent predictive factor was *disaster_campaign*, controlling for the particular recruitment effects during event-centric disaster campaigns which can yield lower average newcomer retention (refer to Chapter 7). Both support the general expectation in the literature that intrinsic motivation is one of the strongest predictors for future participation (Deci et al., 1999; Cameron et al., 2001; Panciera et al., 2009; Dittus et al., 2016a). Two further aspects related to peer feedback were found to have a powerful effect on newcomer retention: the use of verbal rewards in validation messages, and feedback that was immediate.

Verbal rewards lead to a significant increase in retention (a reduction in hazard rate to 80%) among newcomers who had contributed for less than the median 75 minutes during their first day, possibly indicating low intrinsic motivation or self efficacy. The finding suggests that verbal rewards may increase self-efficacy, which may in turn increase retention. In comparison, newcomers who already start with a high degree of self-efficacy may not require such affective-supportive feedback to remain engaged.

The literature offers a range of possible interpretations for this outcome: the importance of fostering a belief that participation is possible and will be welcomed (Bishop, 2007), the importance of positive affective feedback such as socio-emotional support (Lu and Law, 2012), and the importance of positive interactions as a baseline behaviour instead of harmful silence (Cheng et al., 2014) suggest that the social affirmation of positive feedback may be an important prerequisite for future participation. The process of contributing to HOT online can be considered a depersonalised form of interaction: it is often focused on the task, rather than the learner. In the absence of other prominent social cues, small phrases of support may have a large effect.

This interpretation is further supported by the second significant effect, the importance of early feedback. Peer feedback that is sent a week after a contribution is significantly less likely to still have a motivational impact. In comparison, feedback that is sent within 28 hours or less yielded a reduction of the hazard rate to

78%. Any additional day of delay increased the hazard rate. These findings suggest that the absence of any feedback message likely also increases hazard, supporting a prior observation that contributors who receive no feedback are more likely to leave (Cheng et al., 2014). However, this does not necessarily mean that delayed feedback cannot also be effective. Rather, this is in part a limitation in the setting: in the current form, HOT messaging requires contributors to actively log in in order to receive a new message. It is feasible that delayed feedback can still be effective when it is coupled with an email notification mechanism, so that contributors receive feedback messages even after they have stopped actively contributing.

In contrast to prior studies involving public feedback, we could not confirm that negative feedback has a negative impact on retention. This includes both the rejection of contributions, and the use of negative performance feedback. These forms of negative feedback in HOT are comparatively mild compared to other platforms: for example they do not necessarily result in a discarding of all contributions. Furthermore, because the feedback is private, it does not incur a risk of reputational damage or other social feedback effects.

We can confirm that task-contingent rewards do not improve future engagement, which confirms prior expectations in the motivational literature (Harackiewicz, 1979; Cameron et al., 2001). These were issued by some validators in the form of generous acceptance. Additionally, task acceptance in itself appears to have no effect. We further cannot confirm an engagement effect for corrective feedback, possibly because such feedback is targeted at task performance rather than contributor motivation. However, the evidence basis for these tests is low, which suggests we may simply not have sufficient data to establish an effect. In this respect, it should be highlighted that an initial survival analysis of all participants established the base expectation that newcomers will not be retained, regardless of any intervention. It is further worth pointing out that we encountered a highly imbalanced distribution for certain kinds of feedback: both NPF and CF are used in the vast majority of messages (>80%), suggesting there may not be sufficient counter-examples to capture their effect.

These findings also raise the question whether we successfully captured the most important contributing factors. Our model captures existing theory of the

effects of peer feedback well, which is reflected in our detailed labelling process. However, many factors relating to the more general concern of participant retention are not included, in large part because they are unobservable to us. For example, we cannot currently capture prior familiarity with mapping practice outside of OSM, nor the extent to which participants experience peer interactions outside of the observed setting, be it online or in person.

6.6.1 Implications

We believe that our findings are relevant for a wide range of social knowledge production platforms. HOT participation as observed here is open to a lay audience, and the observed feedback effects are independent of the specific mechanics of the HOT contribution workflow. Further, our study places a focus on a transferrable concern: the extent to which private peer interactions can help foster sustained newcomer engagement.

Based on our findings, we recommend that system designers consider the role of peer feedback beyond quality assurance: verbal rewards in peer feedback can have important effects on newcomer retention. To better support such effects, system designers need to distinguish between two separate concerns: on one hand, assessing contribution quality, which is a concern of global task coordination; on the other, guiding newcomers in early skills development, which is a concern of community capacity-building and newcomer training.

Many online platforms conflate these two concepts, so that a single mechanism is responsible for both quality assurance and learner feedback, for example in the form of peer ratings and commenting systems. While the resulting public recognition may be motivating to some, a wide range of studies has shown that negative feedback in such public settings can be demotivating to participants (Resnick et al., 2000; Dellarocas, 2003; Tausczik and Pennebaker, 2012; Parnell et al., 2011; Michael and Otterbacher, 2014; Cheng et al., 2014; Halfaker et al., 2011; Zhu et al., 2013; Wohn, 2015). Our findings demonstrate that private peer feedback is an alternative design option to effectively support community capacity-building while avoiding this particular risk of newcomer discouragement.

In the specific case of HOT, we find that current validation workflows conflate

these two concerns, which in turn affects the tone of validation feedback. In the current form, more HOT validation messages include negative performance feedback than verbal rewards, suggesting that validators focus on regulating contribution quality, and that they may not be aware of the motivational impact of their feedback. To address this, we recommend that improved validation interfaces should regard quality control and learner feedback as separate concerns. Organisers may further seek to provide guidance on how best to articulate impactful feedback. For example, they could emphasise that the specific wording of a feedback message matters, and that the recipient of a validation message may appreciate encouragement by a peer. Furthermore, we see potential to introduce further workflows to accompany mappers throughout their early experiences, and to help identify contributors who might benefit from early feedback, ensuring that newcomers have access to peer contact and mentoring.

Based on the available evidence, we have not found an engagement effect related to generous task acceptance. This validation strategy may be beneficial when timely completion of a project is of importance, however it places a burden on validators to finish the mapping work themselves. In comparison, other strategies may be similarly acceptable when timely completion is not a concern, including the rejection of low-quality contributions when accompanied with a supportive message. This is an area that warrants further research and experimentation.

From a theoretical perspective, we see opportunities for further research in this novel setting. Our findings suggest that private peer feedback on HOT may not suffer from the negative social feedback loops observed on platforms where peer feedback is publicly negotiated. For example, HOT contributors may be less concerned about building a public reputation. However the full implications of this difference are not yet evident, and warrant further research.

6.7 Conclusion

Contributing to HOT online can be considered a depersonalised form of interaction: the process is often focused on the task, rather than the learner. In light of this, it is not surprising to find that the social affirmation of peer feedback can be an important factor in future participation. In the absence of other prominent social cues,

small phrases of support can have a large effect. In particular, we find evidence that verbal rewards and a timely response can significantly improve newcomer retention. On the other hand, we cannot find evidence that feedback related to task performance has an effect on newcomer retention. In particular, we do not find that negative feedback harms retention, as is found on platforms where feedback is publicly negotiated, and where negative feedback can harm one's reputation.

As a consequence of these findings, we propose that public and private peer feedback can be considered complementary mechanisms which can foster different kinds of outcomes. Better understanding of such effects can support the design of improved feedback mechanisms, for example by avoiding negative social feedback loops where they may be harmful to newcomer retention.

Chapter 7

Disaster Campaigns

The previous studies have successfully identified specific coordination practices that can help foster sustained newcomer engagement. Yet it is not yet well-understood to what extent the general aspect of the disaster response context itself may inform participation behaviours. Is HOT building a dormant task force that springs to action when it is needed? Alternatively, do initiatives mainly rely on the recruitment of new contributors during disaster events, possibly at the expense of contribution quality? We seek to develop a better understanding of these relationships, comparing the outcomes of urgent event-centric campaigns with more long-running mission-centric campaigns. In a large-scale quantitative study, we assess the outcomes of 26 campaigns with almost 20,000 participants. We find that event-centric campaigns can be significant recruiting and reactivation events, however that this is not guaranteed. Our analytical methods provide a means of interpreting key differences in outcomes.

7.1 Introduction

Characteristic for HOT is the coexistence of two broad types of campaigns. **Event-centric campaigns** are initiatives that seek completion within days or even hours, typically in the context of urgent emergency response after a natural disaster. This is a synchronised kind of activity, in that contributors participate in the specific moment when a particular urgent need arises. **Mission-centric campaigns** on the other hand are mapping initiatives without a particular deadline. They may seek to proactively map certain unmapped areas, or update existing maps. Some of these may be long-running initiatives covering vast geographic areas, or larger umbrella

initiatives orchestrated by organisations such as the Peace Corps, MapGive, and Missing Maps. Mission-centric activity can be characterised as asynchronous, in that individuals largely contribute at their own leisure.

Does the coexistence of both event-centric and mission-centric activity have implications for contributor engagement? It is currently not clear how contributor capacity for these different activities is constituted. On one hand, event-centric campaigns may be important growth events: media coverage during crisis events may attract many new contributors. On the other hand, event-centric campaigns might also invite a kind of contributor engagement that is characterised by dormancy-reactivation cycles: an experienced yet passive membership that only reawakens when it is needed. However there are no existing studies of such a general effect.

A further important organiser concern is the potential tradeoff between community growth and data quality. A quick response often matters when a disaster strikes. According to FEMA surveys, the value of updated ground surveys to coordinating aid teams decreases with every additional day (Meier, 2016). Modern coordination technologies make it possible to satisfy such timelines with the help of a large global volunteer force. However for the data to be useful it also needs to be accurate, which makes it a concern who participates in these mapping efforts.

To address these concerns, we observe participation outcomes by almost 20,000 HOT volunteers across 26 HOT campaigns, including both event-centric and mission-centric campaigns. We compare HOT community activity in two complementary respects.

Reactivation. Is HOT building a volunteer task force that springs to action in response to specific external events? To what extent are dormant contributors reactivated by urgent emergency responses? What is the impact of such dormant capacity on campaign outcomes?

Recruiting. Are disaster events important growth moments because they attract more newcomers, compared to other campaigns? What is the impact of these newcomers on campaign outcomes, compared to more experienced volunteers? Do they join future campaigns?

We propose a measure of campaign burstiness to quantitatively distinguish these two kinds of campaigns. To assess campaign outcomes we further develop

two intrinsic measures of contribution quality: the share of untagged new objects, and the share of new objects that are subsequently deleted by other contributors.

On the following pages we first provide an overview of related work on event-centric crowdsourcing and HOT contributor engagement. We then outline our research questions, and describe our methodology. Finally we address our research questions with a set of analyses of campaign participation and outcomes, and close with a discussion of our findings.

7.2 Related Work

7.2.1 Event-centric Crowdsourcing

To our knowledge, there are no published studies on the impact of event-centric coordination on the *recruitment and reactivation* of HOT contributors over time. However, work in related domains can introduce some initial expectations. In social media research, studies have documented the willingness of outsiders to participate remotely during crises, for example to help in information propagation (Starbird and Palen, 2010; Vieweg et al., 2010; Starbird and Palen, 2011). It was further found that participation in such events may lead to an interest in becoming more deeply involved in future initiatives (Starbird and Palen, 2013; Cobb et al., 2014). In the context of Wikipedia, it was found that breaking news can lead to intense collective editing activity which surpasses that of most other Wikipedia articles (Keegan et al., 2013). Such media events can rally a diverse set of contributors: some may have contributed during a previous event, while others may be first-time contributors to Wikipedia who make some minor changes and never return. In comparison, other articles tend to have a more stable contributor network (Keegan et al., 2013).

There is some early knowledge about how these different coordination practices may affect the HOT *contribution flow* over time. A recent MSc thesis compares OSM and Wikipedia edit patterns during major disaster events (Esworthy, 2016). The study finds two patterns of behaviour across both platforms: contribution activity after earthquakes and hurricanes is characterised by large initial spikes followed by a long decay period, while a multi-month response to the West Africa Ebola epidemic was characterised by more sustained activity levels. It was further

found that OSM activity was much reduced in instances where it emerged informally, rather than being coordinated by HOT. In other words, there is evidence that HOT coordination can amplify and even foster increased community activity that otherwise would not have taken place. According to the study, organised mapping campaigns have a large impact on contribution patterns (Esworthy, 2016).

7.2.2 HOT Engagement and Outcomes

To date, few studies have tried to assess how HOT contributor engagement is constituted over time. A growing body of published work seeks to assess the outcomes of HOT campaigns, with a focus on larger and more well-known event-centric campaigns: the Haiti earthquake (Soden and Palen, 2014; Palen et al., 2015), typhoon Haiyan/Yolanda (Palen et al. 2015; Westrope et al. 2014; and our initial study in Chapter 4 on page 63), the Nepal earthquake (Eckle and de Albuquerque, 2015; Anhorn et al., 2016; Poiani et al., 2016). However such studies are typically limited to evaluations of these individual campaigns, and focused on questions of process and data quality.

An evaluation of the Nepal campaign finds that more contributions were made by experienced mappers, however that first-time mappers provided small but important contributions, such as the creation of notes for missing information. The authors observe that further research is needed to verify if and how prior contributor experience affects data quality (Poiani et al., 2016).

Our initial study in Chapter 4 (page 63) compares HOT contributor engagement across three large campaigns, typhoon Haiyan, the Ebola response, and Missing Maps. We find that newcomer retention is significantly lower during the former event-centric campaign than during the latter two mission-bases campaigns. This may be attributable to self-selection effects related to different recruiting practices, as well as differences in community-building practices between the campaigns.

In a detailed assessment of OSM contributions after typhoon Haiyan, Westrope et al. further suggest that media coverage may influence mapping outcomes, however the specific relationship is not yet well understood. The authors of the study observe that a region which was more frequently covered by news media had been mapped differently than other nearby regions: the contributions were of a markedly

lower quality (Westrope et al., 2014).

7.3 Research Questions

7.3.1 RQ1: Contributor Recruitment and Reactivation

- Do event-centric campaigns attract a larger share of newcomers than mission-centric campaigns?
- Do they reactivate a larger share of dormant contributors?
- What is the aggregate contribution impact by these groups, and how does this differ across campaign types?
- How long do newcomers remain active in HOT after they joined a particular kind of campaign?

We expect that event-centric campaigns are both recruiting and reactivation moments, however the relative proportions are not clear. We further expect that experienced mappers provide more contributions in aggregate than newcomers, regardless of campaign type. Finally, based on the reviewed literature we expect that newcomers who join during an event-centric campaign are less likely to be retained than those who join during mission-centric campaigns.

7.3.2 RQ2: Contributor Performance

- Do the observed recruiting and reactivation patterns have an impact on the nature of incoming contributions?
- Do different contributor segments produce work at a different rate, or of a different quality? For example, how do the contributions by new recruits compare to those of more experienced contributors?
- How does this affect the outcomes of particular campaigns?

We expect that some contributor groups may be motivated to spend more time on event-centric campaigns. We also expect that on average, newcomers spend less time contributing, produce work at a lower pace, and of a lesser quality than experienced contributors.

7.4 Methodology

Our primary source of empirical evidence is the HOT contribution history, this data set and its derivative forms are described in detail in Section 3.2 on page 55.

We identified a list of HOT campaigns based on a process described in Section 3.2.2 on page 56. We restricted our analysis to tasking manager projects with at least 50 participants. This process yielded 29 campaigns. The start of the first campaign in early 2012 marks the beginning of the observation period for our study. At the time of the study, the last available date in the contribution history was 4th July 2016. Our evaluation includes an analysis of newcomer retention, for which we chose an observation period of 90 days. This threshold was chosen to be significantly longer than 30 days: some regional HOT communities organise monthly mapping events (Smith, 2015), and attendees of such events who do not map at home should still be regarded active contributors.

This determines our study period:

- First observation date: February 2012
- Last inclusion date: 18th March 2016
- Last observation date: 16th June 2016

We only considered campaigns where at least 75% of contributions were made before the cutoff date of 18th March 2016. After this process, 26 campaigns remained part of our study. We determined all campaign participants whose first contributions were before the last inclusion date, amounting to a total of approximately 19,000 participants. They represent a majority share of 87% of the almost 22,000 recorded HOT participants before that date. Their contribution history is the basis for our study, it represents 76% of all HOT edits since early 2012, approximately 100 million edits.

Figure 7.1 shows a timeline of all campaigns we identified during this process, classified by type. The classification process is explained in the following section.

7.4.1 Campaign Classification

On the tasking manager, projects may refer to specific events or long-term missions in their documentation, however event-centric and mission-centric campaigns are

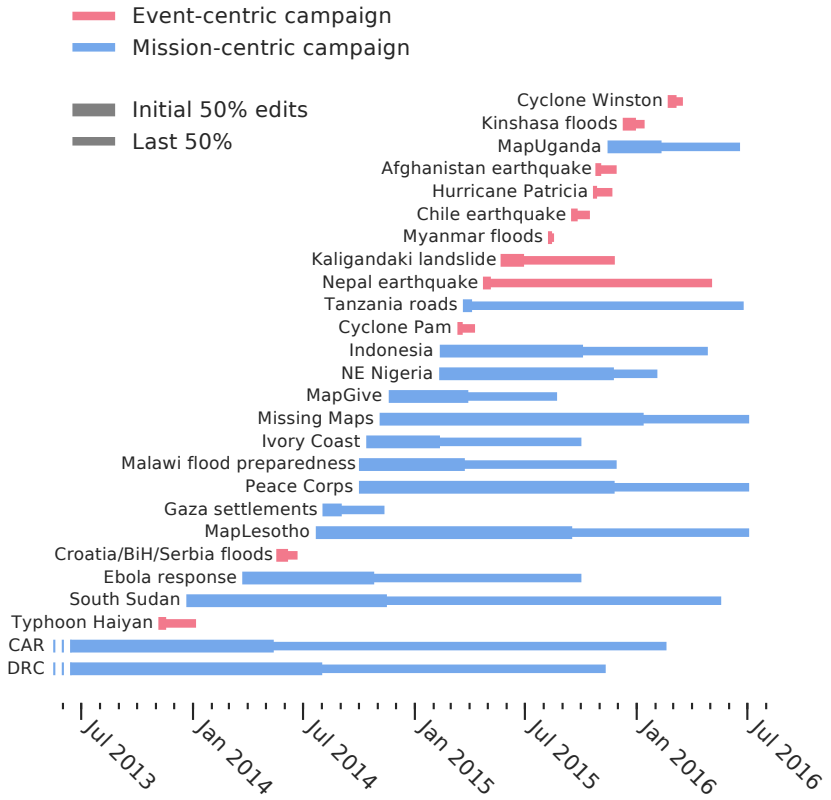

Figure 7.1: Timeline of event- and mission-centric campaigns. Each line visualises the activity period per campaign, indicating how much time passed until 50% of the overall contributions were made.

otherwise not explicitly labelled as such. For our study, we instead manually labelled all campaigns based on their primary cause, as stated in project titles and documentation. 11 campaigns relating to an external event such as a natural disaster (a flood, earthquake, landslide, typhoon, hurricane, or cyclone) were labelled as event-centric. The remaining 15 campaigns were labelled as mission-centric.

In addition to these manual labels, we sought to develop a quantitative classifier which discriminates between campaign types based on participation activity over time. Such a classifier could later be used to replicate our findings in larger studies where manual labelling is infeasible, and in studies of different systems. In particular, the classifier should be suitable for studies of systems which may have event-centric participation characteristics, but that do not make the same explicit distinction between urgent event-centric and more long-running mission-based campaigns. For such cases, we sought to derive the distinction from collective

participation activity over time. Is a campaign an urgent response to an external event, or does it entail a sustained period of activity?

During a review of campaign contribution timelines we encountered two distinct temporal patterns. Event-centric campaigns were characterised by a single, short, and large burst of initial activity, followed by a longer decay period of minimal activity. In some cases this decay period lasted multiple weeks or even months. More long-term mission-centric campaigns were characterised by a sustained period of activity, typically over multiple weeks or months. Their contribution timelines may include multiple intermediate bursts of activity. This matches an observation by a recent study of HOT contribution flows across different campaigns (Eswor-thy, 2016).

After some iterations we chose the *campaign midpoint* as a discriminator between campaign types. The midpoint is the time that has passed until 50% of all contributions to a campaign have been made. We use this measure as an indicator of *campaign burstiness*. The median midpoint across all campaigns is 60 days: half the campaigns finished within or before this time. The remaining campaigns lasted much longer, some spanning many months of activity.

This median midpoint can be used as a threshold for binary classification. Almost all campaigns below this threshold were event-centric campaigns, with only two false positives. All campaigns above the threshold were mission-centric, with no false negatives. Overall this classifier has a false discovery rate of 7.7% of all campaigns (2 out of 26). This misclassification represents an effective error rate of 1.4% of the total edit volume, and 1.5% of all participants. We consider this an acceptable classification error for the purpose of this study.

The two false positives were both comparatively small-scale projects. A campaign to map the Tanzania road network (OpenStreetMap, 2016n) started with a single large automated import, which was succeeded by a longer period of mission-centric mapping. It represents less than 1% of total edits in the study, involving less than 1% of all study participants. A mission-centric effort to digitise buildings in informal settlements in Gaza (OpenStreetMap, 2016b) was a relatively popular campaign with limited geographic scope, characterised by a bursty contribution flow and a short lifespan. It was similarly small in scale, at approximately 1% of total

edits, involving 1% of study participants.

Since the classification error is so low, we decided to use the classifier in parts of our quantitative evaluation. Specifically, we used it for correlation analyses between campaign type (as measured in burstiness) and campaign outcomes. In our discussion of such analyses, we will refer to bursty campaigns with the term 'event-centric campaigns', unless the meaning of these terms diverges in a way that would affect the interpretation of our findings. Non-bursty campaigns will be called 'mission-centric campaigns'.

7.4.2 Contribution Profiles

For a comparative study of campaign outcomes we sought to determine who participated in particular campaigns, how much they contributed, and how well they contributed. We will first outline in a general manner how these aspects were derived, and then explain key aspects in more detail.

We first computed a session history for every study participant, using a method introduced for Wikipedia contributor analysis (Geiger and Halfaker 2013; see also Section 3.2.3 on page 57). This yielded the number of edits per session, and an estimate of the time spent on these contributions, also called *labour hours*. The division of these yields a contributor's *edit pace*, the rate at which they contributed during the session. We further computed each participant's *campaign history*, starting from their first HOT campaign, and recording any subsequent campaign they joined.

Using these measures as a basis, we computed *contribution profiles* for every instance where a study participant contributed to a new campaign. These contribution profiles were later used to evaluate the outcomes of the different campaigns. They include measures across different areas of concern:

- The participant's prior activity at the time of initial campaign participation. Are they a first-time mapper, were they already recently active in other campaigns, or have they been inactive for a longer period? We call this the *contributor segment*.
- The participant's *contribution activity*, measured in number of edits, labour hours, and edit pace.
- The participant's *contribution quality*: the share of untagged new objects, and

the share of objects which are later deleted by other users.

- For newcomers: their *retention* in HOT after the initial contribution.

Retention was measured with a survival analysis over the duration of the observation period (90 days after the initial contribution). Contributor segment classification and measures of contribution quality are developed in more detail in the following sections.

7.4.3 Contributor Segments

When can a contributor be considered dormant? We analysed the frequency and duration of contributor inactivity periods to inform our choice of an inactivity threshold: the time between the last edit of the previous campaign, and first edit of the following. Inactivity periods follow a long-tail distribution, as is shown in Figure 7.2. 3,500 contributors joined a second campaign (18%). Of these, the median time between engagements is 13 days. 40% engagements involved an intermittent dormancy period of 29 days or more, 30% of 62 days or more.

We sought an inactivity period long enough so that it can arguably be regarded as dormancy period, rather than a temporary interruption. In particular, it should be significantly longer than 30 days: some regional HOT communities organise monthly mapping events (Smith, 2015), and attendees of such events who do not map at home should still be regarded active contributors. On the other hand it should also be short enough to capture a significant number of occurrences.

After some early trials we chose 60 days of inactivity as a threshold for dormancy. This is a fairly high threshold above the 70th percentile, yet it still captures a large number of samples. More than 1,600 study participants (9%) at one point in their contribution history became inactive for at least 60 days, but then returned for a future campaign. They represent approximately half of the study participants who had contributed to more than one campaign.

Using this inactivity threshold, we classified all study participants at the time they joined a new campaign into one of three segments:

- If this is their first HOT contribution: *Newcomer*.
- If they contributed in the last 60 days: *Already active*.
- If they previously contributed, but not in the last 60 days: *Previously dormant*.

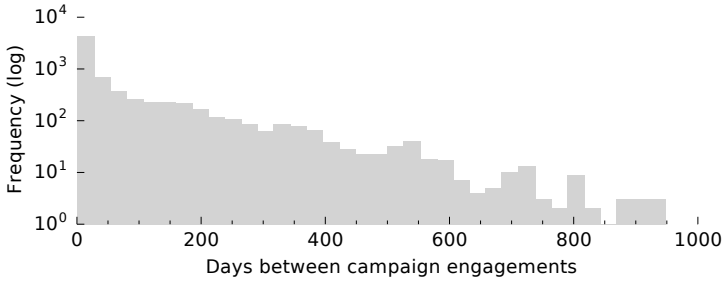

Figure 7.2: Inactivity period between campaign engagements, for all contributors who participated in more than one campaign.

7.4.4 Assessing Contribution Quality

For our evaluation we further sought to compare the contribution quality of different contributors across different mapping activities. To this purpose we required simple indicators of a mapper's contribution quality. These measures needed to be suitable for the HOT context, and manageable within the scale and scope of this study, encompassing many thousands of contributors, and millions of edits.

At this scale, ground truthing of contributions is infeasible. Instead we sought to develop *intrinsic* measures of contribution quality. In the OSM literature, there is a range of widely used intrinsic measures of map quality (Barron et al., 2014), however many are not suitable in this context. As an example of this, the frequently-used indicator measure of Linus' Law (Haklay et al., 2010) assumes that regions are refined by multiple mappers over time, whereas the HOT contribution process assigns only a single mapper to each region.

Outcomes from the HOT validation process could in principle serve as an indicator of contribution quality, but they are unfortunately not made public. Map edits by validators are public, but not clearly attributable to validation, and indistinguishable from contributions by other mappers.

After a review of these options we decided to incorporate two complementary aspects. We computed the share of *untagged new objects* as an instance of easily identified mistakes during the mapping process, and the share of *objects that are eventually deleted* as example of modifications made to new objects some time after they have been added. Both measures can easily be derived from the edit history. They are discussed in more detail in the following sections.

7.4.4.1 Untagged Map Objects

Prior research of the OSM contribution process suggests that beginners do not always annotate their map objects, which renders their contribution unusable (Arsanjani et al., 2013; Neis and Zipf, 2012). There is some dispute in the literature about whether this is a regular occurrence, or even unique to newcomers: a study that assessed HOT contributions after Nepal finds that mappers with less experience do not necessarily produce less well-annotated map objects (Anhorn et al., 2016).

We included the share of untagged new objects as an evaluation criterion in our study to determine whether this reportedly low occurrence would still be sufficient for our large-scale study of contribution behaviour. However based on the prior evidence we expected a low rate of untagged new objects.

To compute the measure, we determined the number of map objects created by study participants which had no annotations at the end of each contribution session. We excluded annotations from this analysis which are automatically added by editing tools (`created_by` and `source`). In total, only a low 0.6% of map objects created during the observation period were left untagged by the creator. 4,700 study participants (26%) had created at least one such untagged map object.

7.4.4.2 Map Object Persistence

Studies on Wikipedia contribution quality introduced the notion of *contribution persistence*, also called transience. The concept describes the extent to which a contribution survives subsequent review by other contributors(Adler and De Alfaro, 2007; Panciera et al., 2009; Priedhorsky et al., 2007; Wöhner and Peters, 2009). It maps well to the HOT contribution process: validators may delete contributions they consider of a low quality.

A review of the HOT edit history showed that a significant number of new map objects were deleted by their creator, often within the same edit session. This indicates that deletions can also be a normal part of the contribution process, for example to fix mistakes as they occur. For our analysis we thus ignore instances where objects are deleted by the creator, and only observe deletions by different contributors. The median delay between object creation and deletion was 29 days, the 75th percentile 213 days. Based on this we chose an evaluation threshold of 90

days, matching our observation window.

During the study period, 3.1% of newly created map objects were deleted by a different contributor within 90 days. This is still a low proportion, but higher than the share of untagged new objects. Approximately 56% of study participants have created at least one object that was deleted by someone else within the next 90 days.

A correlation analysis across contributor records showed that the two contribution quality measures are not mutually correlated. Furthermore, neither of them is correlated with campaign burstiness or campaign start date. On the other hand, the account age of contributors at the time of object creation is negatively correlated with the rate of deleted objects (Spearman coefficient $\rho_S = -0.16$, $p < 0.001$), and weakly correlated with the rate of untagged new objects ($\rho_S = -0.04$, $p < 0.0001$), suggesting a relationship between contributor experience and the two contribution quality measures.

These factors indicate that the measures are useful for an evaluation of contributor engagement: they capture different user behaviours that are not obviously interrelated, and not obviously biased by campaign-specific processes. Instead they can serve as indicators of the contribution quality of individual contributors at particular moments in their contribution history.

7.5 Findings

7.5.1 RQ1: Contributor Recruitment and Reactivation

Table 7.1 shows the distribution of participants across both event-centric and mission-centric campaigns. According to these numbers, emergency response does benefit from a clear reactivation effect. For event-centric campaigns, the share of previously dormant contributors doubles compared to mission-centric campaigns. A correlation analysis between campaign burstiness and reactivation rate confirms this effect: mission-centric campaigns involve a lower share of reactivated dormant contributors ($\rho_S = -0.74$, $p < 0.0001$). However in comparison to the other contributor segments, overall participation by this group was relatively rare.

The inverse applies to newcomers: the share of first-time contributors during event-centric campaigns is markedly smaller than during mission-centric cam-

% contributors	Newcomer	Prev. dormant	Already active
Event	38.4%	16.1%	45.2%
Mission	61.2%	7.7%	33.3%
Overall	50.2%	10.1%	38.3%

Table 7.1: Median share of participants, by contributor segment and campaign type.

paigns. Mission-centric campaigns tend to attract a larger share of newcomers ($\rho_S = 0.40,\ p < 0.05$). However manual inspection showed two exceptional outliers, both large event-centric campaigns, each with approximately 80% newcomers among their participants. These were emergency response campaigns to typhoon Haiyan in the Philippines in 2013, and the Nepal earthquake in 2015. In comparison, all other event-centric campaigns had only between 10 and 50% newcomer participants.

Table 7.2 lists the campaigns with the largest share of newcomers. The top ranks include Haiyan and Nepal, but otherwise only mission-centric campaigns, including Missing Maps, Indonesia, Ebola, Maplesotho, PeaceCorps and others, all with a newcomer share between 60-85%. In absolute terms, mission-centric campaigns recruited twice as many mappers as event-centric campaigns (11,600 vs 6,800).

The bar charts in Figure 7.3 further illustrate this relationship between campaign type, user segment, and participation: the newcomer share varies widely across campaign types, while participation by dormant reactivated contributors is low across all campaigns.

Table 7.3 shows how the contributions of these segments vary by campaign type. Contributors who had already been active were the most prolific, accounting for more than half of the total edits. The picture is more varied for newcomers. While they are the second-largest group in event-centric campaigns, they only contribute a low proportional share of edits: on average, they represent 40% contributors yet only provide 11% of edits. Here again, Nepal and Haiyan are an exception, their 80% newcomers produced around 80% of the overall campaign work. Correlation analysis confirmed that dormants provide a higher share of contributions during event-centric campaigns than during mission-centric campaigns ($\rho_S = -0.46$,

Type	Campaign	% newcomers	# newcomers
Event	Nepal earthquake	84.7%	5,072
Mission	Missing Maps	84.4%	6,280
Mission	Tanzania roads	80.2%	134
Event	Typhoon Haiyan	76.4%	482
Mission	Ebola response	76.2%	2,184
Mission	Indonesia	75.5%	240
Mission	MapLesotho	68.4%	360
Mission	MapUganda	66.5%	177
Mission	Peace Corps	65.5%	898
Mission	MapGive	61.2%	112

Table 7.2: The 10 campaigns with the largest newcomer share.

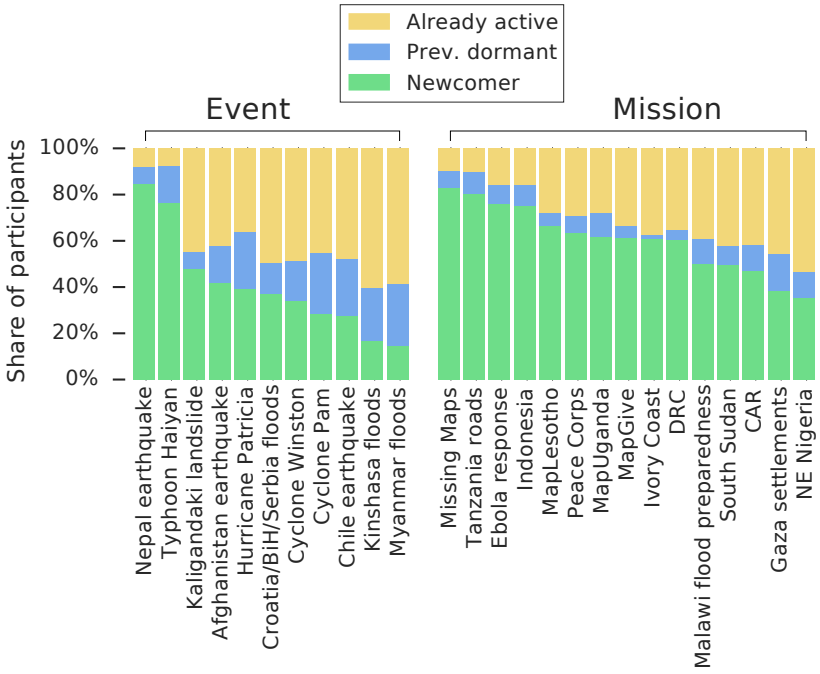

Figure 7.3: Campaign participation per contributor segment. Every vertical bar represents a campaign.

% edits	Newcomer	Prev. dormant	Already active
Event	11.0%	9.9%	67.4%
Mission	40.3%	5.5%	43.6%
Overall	29.9%	8.9%	53.9%

Table 7.3: Median share of contributed edits, by contributor segment and campaign type.

$p < 0.02$), but was not significant for the other groups.

Overall, 80% of newcomers never joined a second campaign, this did not vary by campaign type. For a retention analysis we computed a survival analysis with Kaplan-Meier model, observing newcomer activity over 90 days after initial contribution. This revealed that newcomers who joined during an event-centric campaign have approximately a third of the retention rate: after 45 days, only 3.0% were still active, whereas for longer campaigns 9.6% were still active. For Nepal and Haiyan in isolation, retention rates were even lower (2.3% after 45 days), whereas newcomer retention in other event-centric campaigns was closer to mission-centric campaigns (6.2% after 45 days). Log-rank tests confirmed that the survival rates were statistically different between these groups ($p < 0.02$).

In other words, retention rates were between 5-10% across most contributor segments and campaign types, with the exception of newcomers who joined during the Haiyan and Nepal campaigns. For these campaigns, average newcomer retention was only at 2.3%.

The share of already active contributors was not correlated with campaign type (as measured in burstiness), suggesting their choice to participate may be less related to campaign-specific considerations. However a closer review of their temporal contribution behaviour showed instances where these contributors joined large events as they occurred, and then resumed mission-based campaign work after the events were over. Figure 7.4 shows such contributor flows for two period of intense emergency activity in April-May and October-November 2015. In both cases, a large number of contributors first participated in mission-centric campaigns in earlier months, but then joined event-centric campaigns as they appeared. After the emergency work was completed, many who had participated in the events re-

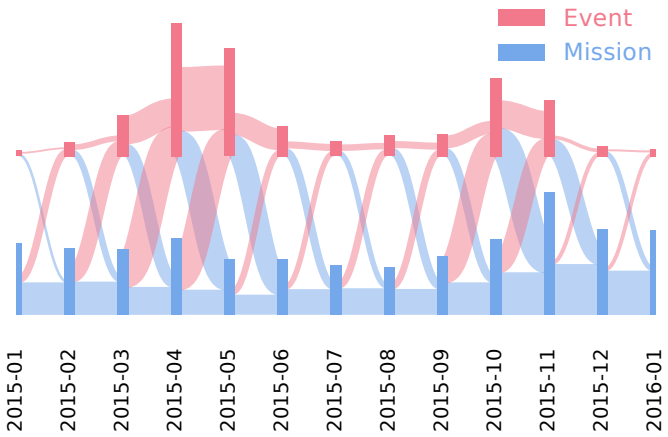

Figure 7.4: Alluvial flow visualisation of temporal contributor flows between event-centric (top) and mission-centric campaigns (bottom). Vertical bars indicate the monthly contributor count, edges the number of contributors who move between campaign types month over month.

turned to mission-centric work.

7.5.2 RQ2: Contributor Performance

Overall, people spent more time and contributed more work during mission-centric campaigns, with a median of 5 hours compared to 3.3 hours during event-centric campaigns, and 2,500 edits compared to 1,600 edits. The average contribution rate is fairly stable across campaign types, at 630 edits per second. However on average, newcomers tended to work more slowly and be less prolific than the other contributor segments.

Average contributor performance according to our evaluation measures is further broken down in Table 7.4. This allows a more detailed comparison of contributor profiles across campaign types. The table shows that newcomer performance was lowest during event-centric campaigns: on average, they contributed less edits and spent less time contributing than any other group. On the other hand, already active contributors tended to increase their effort slightly during event-centric campaigns. Both effects are illustrated by the corresponding distributions of average contribution volumes in Figure 7.5.

Deletions happened rarely, as can be seen in Table 7.4. The median percentage of non-persistent objects is 0.0% across mission-centric campaigns, and around 1%

	Newcomer		Prev. dormant		Already active	
	Event	**Mission**	**Event**	**Mission**	**Event**	**Mission**
Edits	1,252	2,508	3,178	2,741	2,340	2,996
Labour hours	2.8	5.4	5.2	4.4	4.1	5.2
Edits / hour	603	594	634	634	633	633
% **untagged objects**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
% **eventually deleted objects**	2.5%	0.33%	0.36%	0.00%	0.30%	0.00%

Table 7.4: Contribution profiles. Median contribution activity and contribution quality, by contributor segment and campaign type.

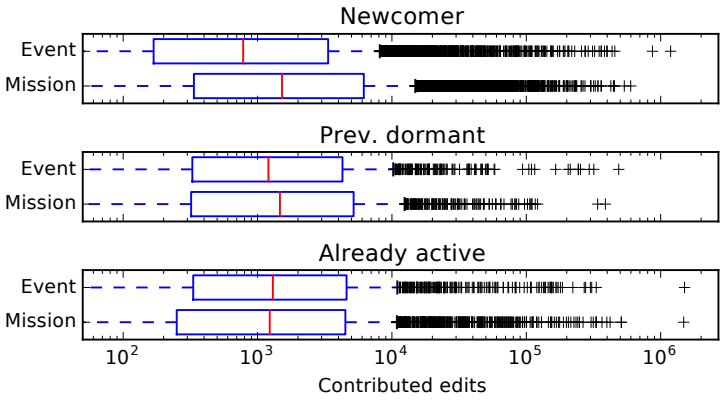

Figure 7.5: Contribution volume. The number of edits per contributor, by contributor segment and campaign type.

for event-centric campaigns. However the distribution of deletion rates in Figure 7.6 illustrates that these outcomes vary significantly by contributor segment: newcomers who joined during event-centric campaigns created a larger share of non-persistent objects than any other group (2.5%), including newcomers who joined during mission-centric campaigns (0.33%). A Mann-Whitney U test confirmed that this difference is significant ($p < 0.01$). Further inspection of the data showed that this effect can mainly be attributed to the outlier campaigns of cyclone Winston and Nepal, where object deletion rates for newcomers were around 4%. All other campaigns had deletion rates of 1% or less.

Untagged objects were created even less frequently, and statistical tests com-

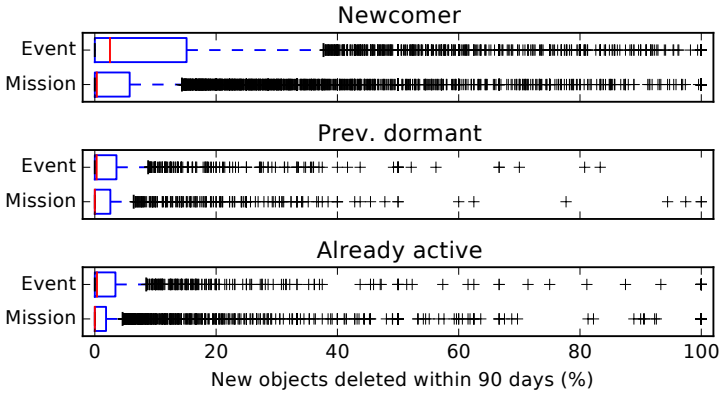

Figure 7.6: Contribution quality. The percentage of new objects per contributor which are deleted within 90 days.

paring average rates per contributor were inconclusive. However we found that on average across the campaigns, a marginally higher share of newcomers created untagged objects (24%) compared to dormant contributors (20%). This difference is statistically significant (Mann-Whitney U test with $p < 0.05$).

7.6 Discussion

Our findings provide a nuanced picture of several interrelated phenomena. They suggest that synchronous and asynchronous modes of campaign coordination are in fact strongly associated with different modes of engagement. However they also reveal the presence of further effects which are not fully captured by the data. We will first summarise the key observed effects and outline how they relate, and then discuss their implications.

Most recruiting happened outside of event-centric campaigns, against our initial expectation. On average, only 40% of participants in disaster campaigns were newcomers, compared to 60% in more long-running mission-centric campaigns. In absolute terms, mission-centric campaigns recruited twice as many mappers as event-centric campaigns. The top recruiting campaigns included Missing Maps and the Ebola response, which are specifically focused on sustained outreach and community-building (refer to Chapter 4, in particular Section 4.4.2 on page 67). Mission-centric campaigns had a significant overall impact on HOT recruitment.

There are indicators for a reactivation pattern. The share of reactivated dormant

contributors was twice as high during event-centric than mission-centric campaigns, and on average these contributors were more prolific than any other group. This suggests that there is interest among a subset of the community to specifically help out during urgent emergencies, which matches observations from comparable participatory emergency response systems (Starbird and Palen, 2013; Cobb et al., 2014). However reactivation only happened at a small scale, on average only 16% of contributors to disaster campaigns exhibited this behaviour. As a result, the overall impact of these reactivations was low. The comparatively small effect may indicate that experienced but dormant mappers are not easily reactivated for synchronous on-demand participation, or that there is a need for better reactivation mechanisms.

Contribution quality was lowest for newcomers during event-centric campaigns. For this group, a median of 2.5% map objects are deleted within 90 days, compared to 0.3% for more experienced mappers during disaster response, and 0.3% or less for any contributor group during mission-centric campaigns. The effect was particularly pronounced during emergency responses for the Nepal earthquake and cyclone Winston: during these campaigns, 4% of contributions by newcomers were eventually deleted. This suggests that timeliness may have taken preference over data quality, although it is unclear whether this was a deliberate organiser choice.

Campaign burstiness is a useful activity-based classifier for event-centric campaigns. We demonstrated that emergency response campaigns can be identified with a simple heuristic measure, the number of days that passed until 50% of the work of a campaign was completed. A binary classifier based on this measure had a low false discovery rate of 7.7%, with 2 wrongly classified campaigns out of 26.

Map object persistence is a useful intrinsic measure of HOT contribution quality. Object persistence as measure of contribution quality was first introduced in research of Wikipedia contribution quality, and maps well to the HOT contribution process. It may be unsurprising that inexperienced participants can produce work of a lower quality, yet to our knowledge this is the first study to observe such an effect for HOT activity at this scale, and across this range of activities. However because deletions were rare overall, the measure could only be used to characterise a small percentage of contributors. The merit of measuring untagged new objects is less clear. Statistical tests involving the average rate of untagged new objects were generally incon-

clusive.

7.6.1 Differences in Event-centric Campaign Outcomes

The Nepal and Haiyan campaigns were noteworthy exceptions by almost all measures. These event-centric campaigns were significant recruiting events, each with 80% of first-time participants. Nepal in particular managed to recruit several thousand first-time mappers. These had a significant effect on campaign outcomes, accounting for the vast majority of all work in both campaigns. On the other hand, they were much less likely to be retained: only 2.3% were still active after 45 days, compared to 6.2% among newcomers who joined during any other event-centric campaign, or 9.6% during any mission-centric campaign. To our knowledge, this is the first published study to observe such a difference in outcomes among event-centric HOT campaigns.

The campaign characteristics observed so far do not offer a good means to reason about the difference. However, it is noteworthy that the two campaigns had the largest numbers of participants among event-centric campaigns, which may be indicative of other unobserved effects. For example, it is feasible that there were differences in promotion and recruiting strategy across campaigns, differences in media coverage, or differences in public interest.

To our knowledge, there is no public record of HOT promotion strategies for these campaigns. However, community members have collected references to international media coverage for some campaigns, and published them on the respective coordination pages. The pages for both Nepal and Haiyan each list more than 40 references per campaign, including coverage of HOT activities by the New York Times, the BBC, The Guardian, The Atlantic, and others (OpenStreetMap, 2016g,m). Where coordination pages exist for the remaining campaigns, they include few to no such references (OpenStreetMap, 2016d,e,h,a,c). While such lists are not a good reflection of actual media coverage, they indicate a community interest to observe such coverage for Nepal and Haiyan. They may suggest that for these two campaigns, media promotion was perceived as important.

As a measure of public interest, we can compare the search volume relating to each campaign's disaster event as captured by Google Trends. A timeline of world-

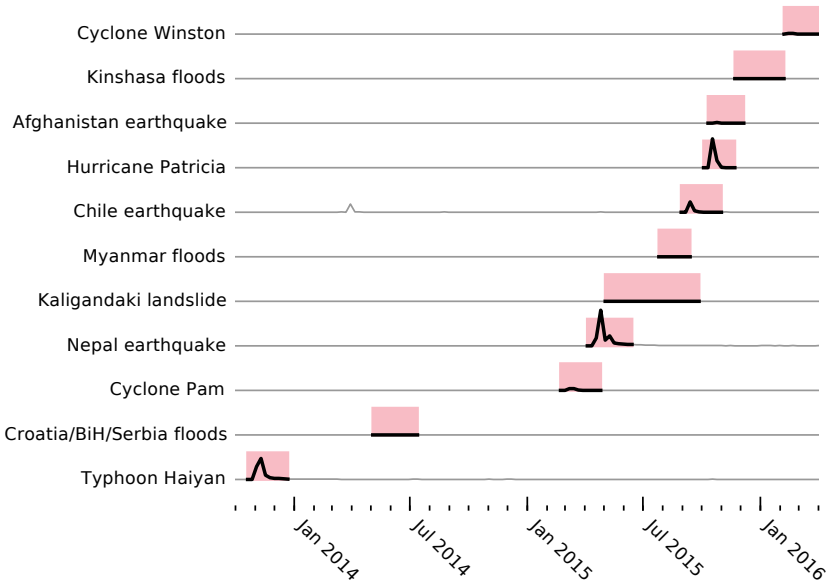

Figure 7.7: Google Trends: relative search volume per disaster event. HOT campaign periods for each event are highlighted.

wide search activity is shown in Figure 7.7, and associated with each campaign. By this measure, three disaster events were most prominent: the Nepal earthquake, typhoon Haiyan, and hurricane Patricia which hit Central America in October 2015. Search volumes for the remaining disaster events were significantly lower.

These observations can serve as initial indication that promotion practices and media coverage may play a role in explaining campaign outcomes, in particular when they are accompanied by increased public interest. However, further research is needed to better understand the relationships between these factors. We offer some suggestions in a Future Work section in Chapter 8. In the meantime we advise caution when applying lessons from Nepal and Haiyan to other disaster campaigns. Experiences gained in these campaigns may not always be transferrable to other emergency response settings.

7.6.2 Implications

HOT as an on-demand task force for emergency response. During event-centric campaigns, a large part of participants was already active in recent weeks. Only a comparatively small share of participants were reactivated dormant contributors. This

suggests that in the current form, experienced HOT contributors are less characterised as a dormant task force that can be activated on demand, and more a stream of continuous activity which can be redirected. Most of the observed on-demand capacity during emergency response was instead provided by inexperienced newcomers who tended to produce contributions of a lower quality. In response, in the following paragraphs we provide some recommendations to organisers for capacity management during emergency campaigns.

Promotion of urgent disaster campaigns, and newcomer training. During some event-centric campaigns, quality of contributions may be just as important as expediency. In order to improve contribution quality for such cases, we recommend to promote upcoming disaster campaigns to the existing community first, and only promote more widely if a quick response is needed and a decrease in overall quality is acceptable. Furthermore, newcomers should be encouraged to first make early experiences in mission-based campaigns where they can receive feedback on their work, and where outcomes are not as time sensitive.

A need for an explicit reactivation mechanism? HOT community coordination takes place on high-traffic mailing lists, online chat, and wiki pages (Soden and Palen, 2014; Palen et al., 2015; Shahid and Elbanna, 2015; Anhorn et al., 2016). Monitoring these on an ongoing basis can require significant time investment. There currently is no means of being informed about urgent emergency responses except by being actively engaged in the community, which is reflected in the low reactivation rates of dormant contributors observed during event-centric campaigns. A dedicated opt-in alert mechanism may allow experienced mappers to join urgent initiatives without having to actively follow daily community discourse, and increase the participation of experienced mappers during emergency response.

A need for a personalised project listing? The urgent HOT responses after the Nepal earthquake and typhoon Haiyan were largely completed by inexperienced contributors. While an influx of newcomers contributed to the completion of the responses, it also resulted in a marked decrease in contribution quality. This is a systemic limitation, and organisers have limited ability to affect such outcomes: in cases where high map quality is an important concern, they cannot currently divert newcomer flows to less critical projects, instead all participants are presented

the same list of projects. In such cases, organisers may benefit from an ability to personalise the project listing, for example to emphasise different kinds of projects depending on a contributor's prior experience.

In summary, we found that event-centric campaigns can be significant recruiting and reactivation events, however that this is not guaranteed. There is evidence that dormant contributors can be reactivated by event-centric campaigns, yet this was rare, possibly due to the lack of a coordinated reactivation process. In two cases, on-demand capacity was largely provided by first-time contributors, which had a negative effect on contribution quality. The causes for this difference in recruiting outcome are not clear, however we provide suggestions for further research at the end.

Theoretical implications. Our findings have theoretical implications for crowdsourcing platforms with event-centric contribution characteristics, including emergency response initiatives that involve public participants (Meier, 2011; Collins, 2011; Palen et al., 2010; Okolloh, 2009), platforms which solicit contributions in response to breaking news (Keegan et al., 2013; Väätäjä et al., 2011), and other event-centric platforms where participants contribute towards a shared outcome. Organisers of such platforms should consider emergent effects relating to recruiting, reactivation, and how they may affect the rate and quality of incoming contributions. This is a particular concern during events which draw large public audiences, and where contribution quality matters. On one hand, platform designers should consider means to manage increased flows of first-time contributors, for example by emphasising less quality-sensitive contribution opportunities. On the other hand, overall contribution quality may benefit from the provision of opt-in notification mechanisms that reactivate more experienced contributors during key events. Designers of opinion-gathering platforms further need to consider how public interest around large events may affect the recruitment of participants, and how this may in turn influence the distribution of responses. This includes online petition platforms (Yasseri et al., 2013; Schmidt and Johnsen, 2014).

7.7 Conclusion

We presented the first large-scale study of a crowdsourcing system to investigate the relationship between event-centric coordination practices, contributor engagement, and campaign outcomes. Using the example of the HOT volunteer community, we compared outcomes of 26 different humanitarian mapping campaigns, involving almost 20,000 participants. We distinguished two types of campaigns: event-centric campaigns which seek to provide outcomes quickly, and mission-based campaigns which proactively map larger regions in the absence of urgent causes. A particular focus was placed on engagement effects relating to the reactivation of inactive contributions, and recruiting of newcomers. We further assessed the quality of contributions by these volunteer segments, making this the first large-scale HOT study to establish a link between prior contributor experience and contribution quality.

Overall, HOT appears to be a highly engaged community of significant proportions: every campaign studied here benefited from low hundreds to thousands of participants. However in most cases, HOT emergency response does not give the impression of an on-demand task force. Instead, most of the contributions were made by participants who were already active at the time. The two high-profile exceptions to this were instances where synchronous activity originated from outside of the community. During the Nepal and Haiyan campaigns, the vast majority of contributions were provided by first-time mappers. While this shows that HOT can successfully channel popular concern around humanitarian crises, it also raises data quality concerns. It may be preferable for newcomers to make early experiences in a less time-sensitive setting. Furthermore, while most mappers are recruited during more long-term campaigns involving proactive mapping efforts, few experienced mappers are then reactivated during emergency response, suggesting a potential for more explicit synchronous coordination.

Chapter 8

Discussion and Conclusions

The steady HOT community growth observed in Section 2.1.4 suggests that there is a seemingly unlimited capacity to rally new HOT volunteers. However our research also shows that in many cases, new contributors are not retained. Our research sought to bring deeper understanding to such questions of sustained contributor engagement. Why is it that almost anyone can learn how to map, but not everyone will stay? Throughout this work, we have considered and evaluated a set of complementary factors.

8.1 Research Themes

Campaign coordination and task design. Are different coordination practices associated with different contributor engagement characteristics? For example, can the perceived urgency of a cause act as an attractor and lead to higher enrolment? Can more sustained promotion of needs increase retention? Once participants have joined, are different task designs associated with different contributor engagement characteristics? For example, is more complex work potentially discouraging to newcomers? Can this be addressed by providing better guidance?

Mapathons as social contribution environments. Does mapathon attendance improve newcomer retention, for example because these social contribution settings provide important newcomer support? Which specific factors contribute to increased retention? According to existing research, colocated practice in social contribution environments can be associated with improved newcomer retention rates,

however there is no empirical evidence available to confirm such an effect in a crowdmapping context.

Private peer feedback during validation. How important is access to a supportive learning environment that allows individuals to attain mastery experiences through peer support and social validation? Are newcomers more likely to be retained when they are given certain kinds of feedback? Should validators be concerned about the engagement impact of their feedback, and spend effort on the content and tone of their messages?

Disaster campaigns as activation events. It is not yet well-understood to what extent the event-centric nature of some HOT campaigns may inform participation behaviours. Is HOT building a dormant task force that springs to action when it is needed? Alternatively, do initiatives mainly rely on the recruitment of new contributors during disaster events, possibly at the expense of contribution quality? How does this impact HOT's ability to grow volunteer capacity over time?

8.2 Research Methods

Our overall methodology relied on three important elements: a readily-available source of empirical data, participation in the setting, and iterative research development. The use of a readily-available source of empirical data allowed us to observe community activities at large scale, across a wide range of practices and settings, and over long time periods. Participation in the setting and regular interactions with practitioners provided important access to domain knowledge, allowing us to better contextualise the participation phenomena emerging from the observational data. Throughout the work, these exchanges have yielded important feedback from practice. Our iterative approach to research development allowed us to discover, articulate, and prioritise important problems, in coordination with the community, and informed by theory. Participation and iteration yielded invaluable support for this work, making it possible to research a complex and emerging practice that previously had received little academic attention.

Once a set of important problems was articulated, the major part of our work was in the execution of discrete research projects. This included the development of a theoretical foundation from the literature, refinement of our conceptual under-

standing in relation to a specific concern, and quantitative analyses of the evidence. The limitations to our analytical methods are discussed in Section 3.4.3 on page 61. Most importantly, our research was limited to observational rather than experimental studies. Our findings are broadly supported by the theory, however there is a risk of mis-interpretation of the evidence. To increase confidence in our findings, they should be confirmed with controlled experiments. On the other hand, the available evidence allowed us to produce research at a scale and thematic breadth that would be hard to achieve in an experimental setting.

This can be considered a forensic approach to researching a socio-technical system: the careful compilation of empirical evidence to uncover underlying processes. It serves as a minimum-effort complement to more invasive evaluation practices such as controlled experiments, A/B tests, and participant observations. The approach was chosen for its ability to support large-scale and long-term observation with comparatively little effort. Our methods allow us to evaluate the outcomes of a wide range of practices in a repeatable and reproducible manner. Furthermore, this process is transferrable to other settings: a minimum requirement is a capacity to observe individual contributions over time.

8.3 Summary of Findings

Our studies identified specific coordination practices which can help build volunteer capacity, and they provide evidence that the event-centric nature of some HOT campaigns can yield markedly different participation behaviours. We summarise the most important findings, grouped into emergent themes.

8.3.1 Impact of Coordination Practices

Campaign practices have a marked impact on contributor retention. Long-term retention was highest for initiatives which were specifically set up as more sustained activities, and which relied on a range of volunteer engagement practices such as mapathons, social media use, and email notifications. A perceived urgency around disaster events may attract new contributors, however they are unlikely to be retained if this is not also accompanied with such sustained coordination efforts. (See

Chapters 4, 5.)

Complex task designs can be a deterrent for certain contributor groups. At the same time, it was found that many differences in task designs yielded comparable engagement outcomes across the observed mapping initiatives. (See Chapter 4.)

8.3.2 Importance of Social Factors

There is evidence that the presence of more experienced attendees can improve newcomer activation at mapathons, however the specific causal factors are unclear. The presence of a community of practice may be of importance to participant motivations, and the enjoyment of a shared practice. The finding may also indicate the importance of access to peer support, which necessitates the presence of other contributors with some prior experience. (See Chapter 5.)

In the absence of other prominent social cues, small phrases of peer support can have a large effect. In our study of peer feedback, we found evidence that verbal rewards and a timely response can significantly improve newcomer retention. This matches existing observations in the literature, and suggests that the social affirmation of peer feedback can be an important factor in future participation. For this reason, these interactions can be considered a potential substitute for impactful encounter when in-person attendance at a mapathon is not feasible. (See Chapter 6.)

The causal structure of engagement effects relating to the individual and the social setting is still unclear. There is evidence that long-term engagement relates to factors relating to the individual, this aligns with findings in the literature. For example, engagement tends to be higher when motivations to participate align with project needs, and when participation is voluntary. A likely prerequisite is the availability of free time. According to participant observation, prior experience in the use of certain software and internet platforms is a further potential prerequisite. However there is also evidence of the importance of social ties, particularly the presence of an existing community of practice. This may be a particularly important motivational factor for contributors who are already embedded in such a community, for example GIS experts or aid volunteers. It is feasible that in online spaces, this can be supplemented with other forms of peer contact. (See Chapters 4–7)

8.3.3 Impact of Recruiting Practices

There is evidence that highly engaged contributors can be identified early. Contributors with much prior OSM experience are likely to become prolific HOT contributors, however they do not necessarily remain engaged for long periods. Increased early activity is associated with long-term engagement. For example, new mapathon attendees who become active mappers at home are likely to come back to a subsequent mapathon. On the other hand, attendees who abandon their work early during a first mapathon are unlikely to continue mapping. (See Chapters 4, 5, 6.)

Promotion practices, media coverage, and public interest may affect campaign recruiting and outcomes, however the specific relationships are not yet well-understood. In Chapter 7, we observed that most recruiting happened outside of event-centric campaigns. The marked exception were disaster campaigns that were accompanied by significant public interest, as observed by increased media coverage and online search behaviours. These were significant recruiting moments, but also yielded contributions of a markedly lower quality. The findings illustrate the importance of promoting activities to the right audiences, in part also to reduce certain risks: a flood of untrained newcomers may decrease quality of outcomes. (See Chapters 4, 7.)

8.3.4 Impact of Participation Context

HOT offers a wide range of participation contexts. HOT activities span a range of initiatives, from urgent emergency responses to more sustained proactive mapping activity. Some HOT initiatives use email notifications and social media to keep participants informed about current activities. Many contributors participate online, however a growing number of mapathons in many cities provide social contribution environments for the practice. Mapathon attendees are often affiliated with aid organisations or other professional networks, and some events are specifically organised for members of a particular organisation. (See Chapters 4, 5, 7.)

Different participation contexts are associated with different volunteer constituencies. In Chapter 4, we found that urgent emergency response initiatives had a larger share of experienced contributors, while a proactive long-term initiative which focused on sustained community building was shown to build new contributor capacity from a

pool of relatively inexperienced mappers. In Chapter 5, corporate mapathons were one-off events for employees of large organisations, often organised for captive audiences, while the regular monthly mapathons attracted and retained a growing number of volunteers from a range of backgrounds. (See Chapters 4, 5, 7.)

Different participation contexts are associated with different engagement profiles. Emergency response initiatives in Chapter 4 benefitted from a large volume of participants, however many of them were only active for short periods. Similarly, the corporate mapathons in Chapter 5 yielded few active contributors. In contrast to this, the availability of regular mapathons organised by the MM community was associated with increased long-term engagement in both studies. (See Chapters 4, 5, 7.)

8.3.5 Creating or Discovering Great Contributors?

Our findings also suggest that there is unlikely to be a clear activation moment for most newcomers, nor a clear technical barrier to entry, and that sustained community engagement is only partially a matter of optimising the contribution process. Current contribution mechanics and mapathon settings do not appear to be strong barriers to entry, however neither are they catalysts for sustained engagement in themselves. Instead, factors relating to the individual and their participation context are likely just as important.

In all studies, factors relating to the individual were more indicative of future engagement than any particular intervention in the setting. Furthermore, there is evidence that online and offline peer support, regular social mapping events, and other social factors are important prerequisites for sustained engagement. In contrast, online contributors who join HOT during a widely publicised emergency response are unlikely to be retained. For these reasons, it appears less likely that prolific mappers are "created" by a particular process, and a pursuit of indiscriminate growth would likely be an inefficient use of organiser resources. However the evidence shows that newcomers who are likely to become highly engaged contributors can be *discovered*, and they may need some support in order to get started.

8.4 Future Work

8.4.1 A Taxonomy of Participation Contexts?

Research findings suggest that different initial participation contexts are associated with different forms of engagement. Long-term engagement trajectories varied between attendees at regular mapathon events and corporate mapathons, remote participants with and without prior experience, and contributors of several larger HOT initiatives. Often the same tools and processes were used, yet the long-term engagement trajectory of these groups varied consistently. Our findings suggest that this difference in long-term outcome is in part attributable to a recruiting effect: different participation contexts can attract different participant cohorts.

There currently exists no comprehensive theoretical model that describes and explains these effects in a crowdmapping or crowdsourcing context. However in some of the scenarios involved in our studies, enough is known about the initial participation context to derive interpretations based on existing theory, including theories of participant motivations. For example, a mapathon participant who enjoys the social experience of the event, but not the act of mapping itself, is likely to return to a future event but may not feel motivated to also contribute at home. A new mapper who learned about HOT through a high-profile emergency response on the news may initially be driven by a desire to help, however is unlikely to be retained unless additional factors foster more sustained motivations. Captive audiences of highly skilled professionals may produce high-quality outcomes for the duration of an organised event, but are unlikely to remain engaged once the institutional framing is removed.

Is it possible to better match projects with suitable participation contexts based on an understanding of such effects? Can this be achieved in a systematic manner, for example based on a taxonomy of participation contexts? The projects discussed so far span a range of needs, from initiatives that involve the completion of a large number of simple but urgent tasks, to proactive initiatives that require a steady stream of mappers in the absence of urgent causes. Some projects may place further requirements on contributors, for example domain knowledge of a certain geography and architecture, or experience with a particular technical process, and

may consequently aspire to recruit more experienced volunteer mappers.

8.4.2 "Piggybacking" on Existing Communities

Our findings and a review of prior HOT studies (Section 2.2 on page 37) also emphasise the importance of social embeddedness in crowdmapping. Successful project execution relies not merely on the ability to find a sufficient number of remote volunteers. Equally important is the participation of organisers and volunteers who are socially connected to the right places. This may mean they have knowledge of a local territory, or they understand the concerns and practices of aid organisations and other participating institutions. It may also mean that they are embedded in institutions which can provide prospective volunteer capacity.

In our experience, many of the organisers and most engaged contributors in Missing Maps are either professionally or socially connected, for example through one of the participating aid organisations. How important were such existing networks in building the active contributor communities of Missing Maps and similar HOT initiatives? In at least one documented case, a small number of enthusiastic mappers launched their own HOT mapping group within a large organisation. A team of GIS experts at the engineering consultancy Arup are now meeting regularly to contribute to HOT projects, and to introduce others to the practice.[1] The group started in London, and is gradually expanding its efforts to Arup offices in other parts of the world. In our experience, many of the UK mapathons organised outside of London are launched by university faculties or student societies.

Are there further organisations and networks which can provide more contributor capacities? Is this a feasible trajectory to grow the contributor community in a more sustainable manner?

8.4.3 A Media Effect?

Is there a media effect of increased newcomer recruitment and reduced contribution quality? In Chapter 7, we observed that contribution quality was worst for newcomers during Haiyan and Nepal, the disaster campaigns with the highest new-

[1] http://www.imeche.org/news/engineering/arup-backs-digital-cartography-project-10041509

comer recruiting rates, and the lowest newcomer retention. The combination of these factors suggests the effect may relate to the specific contexts and coordination practices of these campaigns, including their recruiting practices. Both campaigns received more mainstream media coverage than other emergency response campaigns under study. It is feasible that audiences attracted by such coverage were curious about the phenomenon and interested in supporting a worthwhile cause, rather than specifically motivated to produce high-quality maps. While research to date may be indicative of such a media effect, further study is needed to establish whether this was in fact the case.

8.4.4 What is the Size of the Community?

What is the size of the HOT contributor community? More broadly, our findings in Chapter 7 suggest that campaign-based organising is associated with distinct modes of contributor engagement. This affects how we might conceptualise community size and growth. In particular, the size of the currently active contributor community may not necessarily be indicative of the number of participants who will join a new campaign. There is a need to also consider passive capacity that may be reactivated, as well as recruiting effects during highly promoted campaigns, and campaigns that focus on outreach. For this reason we see an opportunity to develop predictive models of *potential capacity* that may help in campaign planning.

8.4.5 Limits to Growth?

Responding to the HOT aspiration of growing a vast volunteer community, a mapathon organiser expressed his alternative view in conversation during an event. The organiser had observed that a small number of mappers will become exceedingly prolific, while many others contribute comparatively little. According to his assessment, the many casual mappers joining during highly promoted initiatives tended to produce maps of a lower quality. Based on this, the organiser proposed that maybe there was no need to grow the community indiscriminately.

Should HOT be an ever-growing community that is open to all, or are smaller groups of well-trained and highly active mappers preferable to a large crowd? It remains open how best to reconcile the two opposing perspectives. Thanks to grow-

ing public interest, HOT is in the fortunate position that it could pursue either strategy. Research outcomes to date suggest that both options can conceivably deliver satisfactory project outcomes, provided they are carefully managed.

Our research can provide evidence to evaluate these options, and can provide procedural guidance once a particular path has been chosen. We have discussed why different forms of volunteer capacity may be needed to address different kinds of project needs. Organisers should also be aware that differences in coordination practice can yield differences in contribution quality. Future research may help develop a more nuanced understanding of the many contributing factors that shape contribution outcomes.

Yet it is also worth emphasising that the matter of growth-based and expertise-based strategies involves a strategic choice. How important is it to HOT that anyone can contribute? How important is it that the outcomes are of a particular quality? These questions reflect a tension at the core of all volunteer-based knowledge production. How to best resolve this fundamental tension is not simply a question for science, it is also a matter of organisational culture, of organisational priorities, and different organisations might resolve it differently. In this sense, research can merely provide some context for discourse, and can then inform strategies to grow a community in either direction. But the basic tension is unlikely to be resolved by science; rather, it will be decided by the people who show up to participate, to organise, and to make change.

Appendix A

Credits

OpenStreetMap Edit History

OpenStreetMap data (c) OpenStreetMap contributors. The data is made available under the Open Data Commons Open Database License (ODbL).
`http://www.openstreetmap.org/copyright`

HOT Tasking Manager

Project metadata for this research was extracted from the public HOT Tasking Manager, a mapping tool designed and built for the Humanitarian OSM Team. Further information is available at `http://tasks.hotosm.org/about`

World Borders Dataset

The World Borders Dataset is provided by Bjorn Sandvik under a Creative Commons Attribution-Share Alike License. The data is available at `http://thematicmapping.org/downloads/world_borders.php`

It is derived from free mapping data provided by the Mapping Hacks website at `http://www.mappinghacks.com/data/`

Bibliography

Adamopoulos, P. (2013). What makes a great mooc? an interdisciplinary analysis of student retention in online courses. In *ICIS*.

Adler, B. T. and De Alfaro, L. (2007). A content-driven reputation system for the wikipedia. In *Proc. WWW '07*, pages 261–270. ACM.

Ahn, J., Butler, B. S., Weng, C., and Webster, S. (2013). Learning to be a better q'er in social q&a sites: social norms and information artifacts. *Proc. ASIS&T '13*, 50(1):1–10.

Angelopoulos, S. and Merali, Y. (2015). Bridging the divide between virtual and embodied spaces: Exploring the effect of offline interactions on the sociability of participants of topic-specific online communities. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, pages 1994–2002. IEEE.

Anhorn, J., Herfort, B., and Albuquerque, J. P. d. (2016). Crowdsourced validation and updating of dynamic features in openstreetmap an analysis of shelter mapping after the 2015 nepal earthquake. In *Proc. ISCRAM '16*.

Arsanjani, J. J., Barron, C., Bakillah, M., and Helbich, M. (2013). Assessing the quality of openstreetmap contributors together with their contributions. In *Proc. AGILE '13*.

Ashforth, B. K. and Saks, A. M. (1996). Socialization tactics: Longitudinal effects on newcomer adjustment. *Academy of management Journal*, 39(1):149–178.

Atkinson, P. and Hammersley, M. (1994). Ethnography and Participant Observation.

Baecker, R. M. (2014). *Readings in Human-Computer Interaction: toward the year 2000*. Morgan Kaufmann.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2):191.

Barron, C., Neis, P., and Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18(6):877–895.

Benkler, Y. (2002). Coase's Penguin, or, Linux and The Nature of the Firm. *Yale Law Journal*, 112(3):369–446.

Bishop, J. (2007). Increasing participation in online communities: A framework for human–computer interaction. *Computers in human behavior*, 23(4):1881–1893.

Brockner, J., Derr, W. R., and Laing, W. N. (1987). Self-esteem and reactions to negative feedback: Toward greater generalizability. *Journal of Research in Personality*, 21(3):318–333.

Brown, M., Sharples, S., Harding, J., Parker, C. J., Bearman, N., Maguire, M., Forrest, D., Haklay, M., and Jackson, M. (2013). Usability of Geographic Information: current challenges and future directions. *Applied ergonomics*, 44(6):855–65.

Bryant, S. L., Forte, A., and Bruckman, A. (2005). Becoming Wikipedian : Transformation of Participation in a Collaborative Online Encyclopedia.

Budhathoki, N. R. (2010). *Participants' motivations to contribute geographic information in an online community*. PhD thesis, University of Illinois at Urbana-Champaign.

Burke, M., Marlow, C., and Lento, T. (2009). Feed me: motivating newcomer contribution in social network sites. In *Proc. SIGCHI '09*, pages 945–954. ACM.

Cameron, J., Banko, K. M., and Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst*, 24(1):1.

Cheng, J., Danescu-Niculescu-Mizil, C., and Leskovec, J. (2014). How community feedback shapes user behavior. In *Proc. ICWSM '14*.

Choi, B., Alexander, K., Kraut, R. E., and Levine, J. M. (2010). Socialization tactics in wikipedia and their effects. In *Proc. CSCW '10*, pages 107–116. ACM.

Ciampaglia, G. L., Taraborelli, D., and Francisco, S. (2015). MoodBar: increasing new user retention in Wikipedia through lightweight socialization. In *Proc. CSCW '15*, pages 734–742.

Clark, L. (2014). The race to contain west africa's ebola outbreak - wired uk. `http://www.wired.co.uk/news/archive/2014-04/11/ebola-open-street-map`.

Clary, E. G., Snyder, M., Ridge, R. D., Copeland, J., Stukas, A. A., Haugen, J., and Miene, P. (1998). Understanding and assessing the motivations of volunteers: a functional approach. *Journal of personality and social psychology*, 74(6):1516.

Cobb, C., McCarthy, T., Perkins, A., Bharadwaj, A., Comis, J., Do, B., and Starbird, K. (2014). Designing for the deluge: Understanding & supporting the distributed, collaborative work of crisis volunteers. In *Proc. CSCW '14*, pages 888–899. ACM.

Collins, S. (2011). Conflict and disaster management in a hyper-connected world: Cooperative, collaborative, real time. *Journal of Urban and Regional Research*, 25(2):227–52.

Crowston, K. and Howison, J. (2005). The social structure of free and open source software development (originally published in volume 10, number 2, february 2005). *First Monday*.

Crowston, K. and Scozzi, B. (2004). Coordination practices within floss development teams: The bug fixing process. *Computer Supported Acitivity Coordination*, 4(1):21–30.

Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6):627.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science*, 49(10):1407–1424.

Dittus, M. and Capra, L. (2018). Private peer feedback as engagement driver in humanitarian mapping. In *Proc. CSCW '18*. (Conditionally accepted, expecting publication in November 2017).

Dittus, M., Quattrone, G., and Capra, L. (2016a). Analysing volunteer engagement in humanitarian mapping: building contributor communities at large scale. In *Proc. CSCW '16*, pages 108–118.

Dittus, M., Quattrone, G., and Capra, L. (2016b). Social contribution settings and newcomer retention in humanitarian crowd mapping. In *Proc. SocInfo 2016*, pages 179–193. Springer.

Dittus, M., Quattrone, G., and Capra, L. (2017). Mass participation during emergency response: event-centric crowd-sourcing in humanitarian mapping. In *Proc. CSCW '17*, pages 1290–1303.

Dodge, M. and Kitchin, R. (2013). Crowdsourced cartography: mapping experience and knowledge. *Environment and Planning A*, 45(1):19–36.

Eckle, M. and de Albuquerque, J. P. (2015). Quality assessment of remote mapping in openstreetmap for disaster management purposes. *ISCRAM 2015*.

Esworthy, G. (2016). Comparing openstreetmap and wikipedia edit patterns during major events. Master's thesis, George Mason University.

Fang, Y. and Neufeld, D. (2009). Understanding sustained participation in open source software projects. *Journal of Management Information Systems*, 25(4):9–50.

Farzan, R., Kraut, R., Pal, A., and Konstan, J. (2012). Socializing volunteers in an online community: a field experiment. In *Proc. CSCW '12*, pages 325–334. ACM.

Farzan, R., Savage, S., and Saviaga, C. F. (2016). Bring on board new enthusiasts! a case study of impact of wikipedia art+ feminism edit-a-thon events on newcomers. In *International Conference on Social Informatics*, pages 24–40. Springer.

Feinmann, J. (2014). How msf is mapping the world's medical emergency zones. *BMJ*, 349:g7540.

Freedman, L. S. (1982). Tables of the number of patients required in clinical trials using the logrank test. *Statistics in medicine*, 1(2):121–129.

Geiger, R. S. and Halfaker, A. (2013). Using edit sessions to measure participation in Wikipedia. In *Proc. CSCW '13*.

Goldberg, A. (2015). In defense of forensic social science. (December):1–3.

Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing Geographic Knowledge*, pages 105–122. Springer.

Haklay, M. and Budhathoki, N. (2010). Openstreetmap–overview and motivational factors.

Haklay, M. M., Basiouka, S., Antoniou, V., and Ather, A. (2010). How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *Cartographic Journal, The*, 47(4):315–322.

Haklay, M. M. and Weber, P. (2008). OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, 7(4):12–18.

Halfaker, A., Kittur, A., and Riedl, J. (2011). Don't bite the newbies: how reverts affect the quantity and quality of wikipedia work. In *Proc. WikiSym '11*, pages 163–172. ACM.

Harackiewicz, J. M. (1979). The effects of reward contingency and performance feedback on intrinsic motivation. *Journal of personality and social psychology*, 37(8):1352.

Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1):81–112.

Hayes, G. R. (2011). The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction*, 18(3):1–20.

Haythornthwaite, C., Kazmer, M. M., Robins, J., and Shoemaker, S. (2000). Community development among distance learners: Temporal and technological dimensions. *Journal of Computer-Mediated Communication*, 6(1):0–0.

HOT, H. O. T. (2017). Featured hot projects. `http://hotosm.org/projects`.

Howe, J. (2006). The rise of crowdsourcing. *Wired magazine*, 14(6).

Hristova, D., Quattrone, G., Mashhadi, A., and Capra, L. (2013). The life of the party: impact of social mapping in OpenStreetMap. In *Proc. ICWSM '13*, pages 234–243.

Jin, L., Robey, D., and Boudreau, M.-C. (2015). The nature of hybrid community: An exploratory study of open source software user groups. *The Journal of Community Informatics*, 11(1).

Johansen, R. (1988). *Groupware: Computer support for business teams*. The Free Press.

Keegan, B., Gergle, D., and Contractor, N. (2013). Hot off the wiki: Structures and dynamics of wikipedia's coverage of breaking news events. *American Behavioral Scientist*, page 0002764212469367.

Khanna, S., Ratan, A., Davis, J., and Thies, W. (2010). Evaluating and improving the usability of Mechanical Turk for low-income workers in India. In *ACM DEV*.

Kittur, A., Nickerson, J., and Bernstein, M. (2013). The future of crowd work. *Proc. CSCW '13*.

Kittur, A., Pendleton, B., and Kraut, R. E. (2009). Herding the cats: the influence of groups in coordinating peer production. In *Proc. WikiSym '09*, page 7. ACM.

Koh, J., Kim, Y.-G., Butler, B., and Bock, G.-W. (2007). Encouraging participation in virtual communities. *Communications of the ACM*, 50(2):68–73.

Kraut, R. E. (2002). Applying social psychological theory to the problems of group work.

Kraut, R. E., Resnick, P., Kiesler, S., Burke, M., Chen, Y., Kittur, N., Konstan, J., Ren, Y., and Riedl, J. (2012). *Building successful online communities: Evidence-based social design*. Mit Press.

Labaree, R. V. (2002). The risk of 'going observationalist': negotiating the hidden dilemmas of being an insider participant observer. *Qualitative Research*, 2(1):97–122.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

Lave, J. and Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of google flu: traps in big data analysis. *Science*, 343(14 March).

Lewis, K. (2015). Three fallacies of digital footprints. (December):1–4.

Lin, H.-F. (2007). The role of online and offline features in sustaining virtual communities: an empirical study. *Internet Research*, 17(2):119–138.

Lu, J. and Law, N. (2012). Online peer assessment: effects of cognitive and affective feedback. *Instructional Science*, 40(2):257–275.

Lüge, T. (2014). Gis support for the msf ebola response in guinea 2014. Médecins Sans Frontières, Logistics Department. `http://preparecenter.org/resources/gis-support-msf-ebola-reponse-guinea-2014`.

Lüge, T. (2015). Gis support for the msf ebola response in liberia, guinea and sierra leone. Médecins Sans Frontières, Logistics Department. `http://sm4good.com/2015/11/30/maps-helped-fight-ebola-part-ii/`.

Lysakowski, R. S. and Walberg, H. J. (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal*, 19(4):559–572.

McFarland, D. A., Lewis, K., and Goldberg, A. (2015). Sociology in the Era of Big Data: The Ascent of Forensic Social Science. *The American Sociologist*.

Meier, P. (2011). New information technologies and their impact on the humanitarian sector. *International review of the Red Cross*, 93(884):1239–1263.

Meier, P. (2016). The value of timely information during disasters (measured in hours). `https://irevolutions.org/2016/04/06/the-value-of-timely-information-during-disasters-measured-in-hours/`.

Meyer, R. (2013). How online mapmakers are helping the red cross save lives in the philippines. `http://www.theatlantic.com/technology/archive/2013/11/`

how-online-mapmakers-are-helping-the-red-cross-save-lives-in-the-philippines/281366/.

Meyer, R. (2016). The internet mapmakers helping nepal. http://www.theatlantic.com/technology/archive/2015/05/the-mapmakers-helping-nepal/392228/.

Michael, C. (2014). Missing maps: nothing less than a human genome project for cities - the guardian. http://www.theguardian.com/cities/2014/oct/06/missing-maps-human-genome-project-unmapped-cities.

Michael, L. and Otterbacher, J. (2014). Write like i write: Herding in the language of online reviews. In *Proc. ICWSM '14*.

MM, M. M. P. (2017). Missing maps - putting the world's vulnerable people on the map. http://www.missingmaps.org.

Moreland, R. L. and Sweeney, P. D. (1984). Self-expectancies and reactions to evaluations of personal performance. *Journal of Personality*, 52(2):156–176.

Morgan, J. T., Bouterse, S., Walls, H., and Stierch, S. (2013). Tea and sympathy: crafting positive new user experiences on wikipedia. In *Proc. CSCW '13*, pages 839–848. ACM.

MSF, M. S. F. (2014). Msf: the missing maps project. http://www.msf.org.uk/missing-maps-project.

Musicant, D. R., Ren, Y., Johnson, J. A., and Riedl, J. (2011). Mentoring in wikipedia: a clash of cultures. In *Proc. WikiSym '11*, pages 173–182. ACM.

Neis, P. and Zipf, A. (2012). Analyzing the contributor activity of a volunteered geographic information project–the case of openstreetmap. *ISPRS International Journal of Geo-Information*, 1(2):146–165.

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

Okolloh, O. (2009). Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory learning and action*, 59(1):65–70.

OpenStreetMap (2014). Elements – openstreetmap wiki. http://wiki.openstreetmap.org/wiki/Elements.

OpenStreetMap (2015). 2014 west africa ebola response - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/2014_West_Africa_Ebola_Response.

OpenStreetMap (2016a). 2010 chile earthquake - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/2010_Chile_earthquake.

OpenStreetMap (2016b). 2014 gaza strip - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/2014_Gaza_Strip.

OpenStreetMap (2016c). 2014 southeast europe floods - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/2014_Southeast_Europe_floods.

OpenStreetMap (2016d). 2015 eastern afghanistan earthquake: Media coverage - openstreetmap wiki. http://wiki.openstreetmap.org/wiki/2015_Eastern_Afghanistan_Earthquake#Media_coverage.

OpenStreetMap (2016e). 2015 hurricane patricia: Media coverage - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/2015_Hurricane_Patricia#Media_coverage.

OpenStreetMap (2016f). 2015 nepal earthquake - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/2015_Nepal_earthquake.

OpenStreetMap (2016g). 2015 nepal earthquake: Write-ups and in the news - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/2015_Nepal_earthquake#Write-ups_and_in_the_News.

OpenStreetMap (2016h). 2015 pam cyclone response: Media coverage - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/2015_03_Vanuatu_Pam_Cyclone_Response#Media_coverage.

OpenStreetMap (2016i). Humanitarian osm team - openstreetmap wiki. http://wiki.openstreetmap.org/wiki/Humanitarian_OSM_Team.

OpenStreetMap (2016j). Josm/validator - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/JOSM/Validator.

OpenStreetMap (2016k). Osm tasking manager/validating data - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/OSM_Tasking_Manager/Validating_data.

OpenStreetMap (2016l). Typhoon haiyan - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/Typhoon_Haiyan.

OpenStreetMap (2016m). Typhoon haiyan: Media coverage - openstreetmap wiki. https://wiki.openstreetmap.org/wiki/Typhoon_Haiyan#Media_coverage.

OpenStreetMap (2016n). Wikiproject tanzania - openstreetmap wiki. http://wiki.openstreetmap.org/wiki/WikiProject_Tanzania.

Owen, L. M. (2015). Global validation procedure. https://www.openstreetmap.org/user/Lisa%20Marie%20Owen/diary/35522.

Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., and Grunwald, D. (2010). A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters. In *Proc. ACM-BCS '10*, page 8. British Computer Society.

Palen, L., Soden, R., Anderson, T. J., and Barrenechea, M. (2015). Success & scale in a data-producing organization: the socio-technical evolution of OpenStreetMap in response to humanitarian events. In *Proc. SIGCHI '15*, pages 4113–4122.

Panciera, K., Halfaker, A., and Terveen, L. (2009). Wikipedians are born, not made: a study of power editors on wikipedia. In *Proc. GROUP '09*, pages 51–60. ACM.

Parnell, L. D., Lindenbaum, P., Shameer, K., Dall'Olio, G. M., Swan, D. C., Jensen, L. J., Cockell, S. J., Pedersen, B. S., Mangan, M. E., Miller, C. A., et al. (2011). Biostar: an online question & answer resource for the bioinformatics community. *PLoS Comput Biol*, 7(10):e1002216.

Parshley, L. (2017). Faultlines, black holes and glaciers: mapping uncharted territories - the guardian. https://www.theguardian.com/science/2017/feb/07/faultlines-black-holes-glaciers-mapping-uncharted-territories.

Patel, D. (2015). 2015 was a good year for creating the world's 'missing maps' with openstreetmap. `https://opensource.com/life/15/12/creating-worlds-missing-maps-openstreetmap`.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, 92(2):382.

Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., Xiong, J., Deng, W., Oh, O., and de Vreede, G.-J. (2013). Conceptual Foundations of Crowdsourcing: A Review of IS Research. *2013 46th Hawaii International Conference on System Sciences*, pages 579–588.

Perkins, C. and Dodge, M. (2008). The potential of user-generated cartography: a case study of the OpenStreetMap project and Mapchester mapping party. *North West Geography*, 8(1).

Poiani, T. H., dos Santos Rocha, R., Degrossi, L. C., and de Albuquerque, J. P. (2016). Potential of collaborative mapping for disaster relief: A case study of openstreetmap in the nepal earthquake 2015. In *Proc. HICSS '16*, pages 188–197. IEEE.

Ponciano, L. and Brasileiro, F. (2014). Finding volunteers' engagement profiles in human computation for citizen science projects. *Human Computation*, 1(2).

Ponciano, L., Brasileiro, F., Simpson, R., and Smith, A. (2014). Volunteers' engagement in human computation astronomy projects. *Computing in Science & Engineering*.

Priedhorsky, R., Chen, J., Lam, S. T. K., Panciera, K., Terveen, L., and Riedl, J. (2007). Creating, destroying, and restoring value in wikipedia. In *Proc. GROUP '07*, pages 259–268. ACM.

Quaintance, K. (2014). Concepts to know: Crowdmapping. `https://kimoquaintance.com/2011/09/04/concepts-to-know-crowdmapping/`.

Reed, S. (2015). Missing maps training video suggestions. `https://lists.openstreetmap.org/pipermail/hot/2015-May/008865.html`.

Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Communications of the ACM*, 43(12):45–48.

Roche, S., Propeck-Zimmermann, E., and Mericskay, B. (2011). GeoWeb and crisis management: issues and perspectives of volunteered geographic information. *GeoJournal*, 78(1):21–40.

Ruths, D., Pfeffer, J., et al. (2014). Social media for large studies of behavior. *Science*, 346(6213):1063–1064.

Schervish, P. G. and Havens, J. J. (1997). Social participation and charitable giving: a multivariate analysis. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 8(3):235–260.

Schmidt, J.-H. and Johnsen, K. (2014). On the use of the e-petition platform of the german bundestag.

Schmidt, M., Klettner, S., and Steinmann, R. (2013). Barriers for contributing to VGI projects. In *Proc. ICC '13*.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., and Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1):139–167.

Shahid, A. R. and Elbanna, A. (2015). The Impact of Crowdsourcing on Organisational Practices: the Case of Crowdmapping. *ECIS 2015 Proceedings*, pages 1–16.

Smith, S. (2015). Prefer typing to running a marathon? take part in a mapathon. `http://blogs.redcross.org.uk/international/2015/10/prefer-typing-to-running-a-marathon-take-part-in-a-mapathon/`.

Soden, R., Budhathoki, N., and Palen, L. (2014). Resilience-building and the crisis informatics agenda: Lessons learned from open cities kathmandu. In *Proc. ISCRAM '14*.

Soden, R. and Palen, L. (2014). From crowdsourced mapping to community mapping: the post-earthquake work of OpenStreetMap Haiti.

Stanford, M. (2015). Sky news interview - harry wood on hot nepal earthquake mapping. `https://www.youtube.com/watch?v=DTH31Qgtu7o`.

Starbird, K. and Palen, L. (2010). Pass it on?: Retweeting in mass emergency. In *Proc. ISCRAM '10*.

Starbird, K. and Palen, L. (2011). Voluntweeters: Self-organizing by digital volunteers in times of crisis. In *Proc. SIGCHI '11*, pages 1071–1080. ACM.

Starbird, K. and Palen, L. (2013). Working and sustaining the virtual disaster desk. In *Proc. CSCW '13*, pages 491–502. ACM.

Steinmacher, I., Gerosa, M. A., and Redmiles, D. (2014). Attracting, onboarding, and retaining newcomer developers in open source software projects. In *Workshop on Global Software Development in a CSCW Perspective*.

Sturgis, S. (2014). Open-source mapping the world's most vulnerable regions will save lives. `http://www.citylab.com/design/2014/12/open-source-mapping-the-worlds-most-vulnerable-regions-will-save-lives/383337/`.

Sutter, J. D. (2010). Ushahidi: How to 'crowdmap' a disaster. `http://edition.cnn.com/2010/TECH/innovation/10/25/crowdmap.disaster.internet/index.html`.

Tausczik, Y. R. and Pennebaker, J. W. (2012). Participation in an online mathematics community: differentiating motivations to add. In *Proc. CSCW '12*, pages 207–216. ACM.

Tenenbaum, G. and Goldring, E. (1989). A meta-analysis of the effect of enhanced instruction: Cues, participation, reinforcement and feedback and correctives on motor skill learning. *Journal of Research & Development in Education*.

Trainer, E. H., Chaihirunkarn, C., Kalyanasundaram, A., and Herbsleb, J. D. (2014). Community code engagements: summer of code & hackathons for community building in scientific software. In *Proc. ACM GROUP '14*, pages 111–121. ACM.

Väätäjä, H., Vainio, T., Sirkkunen, E., and Salo, K. (2011). Crowdsourced news reporting: supporting news content creation with mobile phones. In *Proc. MobileHCI '11*, pages 435–444. ACM.

Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L. (2010). Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proc. SIGCHI '10*, pages 1079–1088. ACM.

Wang, L. S., Chen, J., Ren, Y., and Riedl, J. (2012). Searching for the Goldilocks zone. In *Proc. CSCW '12*, page 989.

Westrope, C., Banick, R., and Levine, M. (2014). Groundtruthing OpenStreetMap Building Damage Assessment. *Procedia Engineering*, 78:29–39.

Whelan, J. (2016). So you've thought about becoming a validator. `https://lists.openstreetmap.org/pipermail/hot/2016-May/011925.html`.

Wiggins, A. and Crowston, K. (2010). Developing a conceptual model of virtual organisations for citizen science. 1(2005):148–162.

Wiggins, A. and Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. *Proc. HICSS '11*.

Wilson, J. (2000). Volunteering. *Annual review of sociology*, pages 215–240.

Wilson, P. (1991). *Computer Supported Cooperative Work: An Introduction*. Springer.

Wohn, D. Y. (2015). The effects of feedback and habit on content posting in an online community. *iConference 2015 Proceedings*.

Wöhner, T. and Peters, R. (2009). Assessing the quality of wikipedia articles with lifecycle based metrics. In *Proc. WikiSym '09*, page 16. ACM.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37(1):60–82.

Yasseri, T., Hale, S. A., and Margetts, H. (2013). Modeling the rise in internet-based petitions. *arXiv preprint arXiv:1308.0239*.

Zhu, H., Zhang, A., He, J., Kraut, R. E., and Kittur, A. (2013). Effects of peer feedback on contribution: a field experiment in wikipedia. In *Proc. SIGCHI '13*, pages 2253–2262. ACM.

Zook, M., Graham, M., Shelton, T., and Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy*, 2(2):6–32.