The Impact of Community Happenings in OpenStreetMap—Establishing a Framework for Online Community Member Activity Analyses

Moritz Schott [1,*] ⬤, **Asher Yair Grinberger** [2] ⬤, **Sven Lautenbach** [3] ⬤ and **Alexander Zipf** [1,3] ⬤

1   Institute of Geography, GIScience, Heidelberg University, 69120 Heidelberg, Germany
2   The Department of Geography, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel;
     yair.grinberger@mail.huji.ac.il
3   HeiGIT at Heidelberg University, 69120 Heidelberg, Germany; sven.lautenbach@heigit.org (S.L.);
     zipf@uni-heidelberg.de or alexander.zipf@heigit.org (A.Z.)
*   Correspondence: moritz.schott@uni-heidelberg.de; Tel.: +49-6221-54-5547

Abstract: The collaborative nature of activities in Web 2.0 projects leads to the formation of online communities. To reinforce this community, these projects often rely on happenings centred around data creation and curation activities. We suggest an integrated framework to directly assess online community member performance in a quantitative manner and applied it to the case study of OpenStreetMap. A set of mappers who participated in both field and remote mapping-related happenings was identified. To measure the effects of happenings, we computed attributes characterising the mappers' contribution behaviour before and after the happenings and tested for significant impacts in relation to a control group. Results showed that newcomers to OpenStreetMap adopted a contribution behaviour similar to the contribution behaviour typical for the respective happening they attended: When contributing after the happening, newcomers who attended a remote mapping event tended to concentrate on creating new data with lower quality but high quantity in places foreign to their home region; newcomers who attended a field mapping event updated and enhanced existing local data with high accuracy. The behaviour of advanced mappers stayed largely unaffected by happenings. Unfortunately, our results did not reveal a positive effect on the community integration of newcomers through happenings.

Keywords: online communities; user centric analyses; contribution patterns; community events; mapping events; volunteered geographic information; OpenStreetMap

1. Introduction

The paradigmatic shift towards Web 2.0 technologies enables online information users to become also producers. This, along with the ubiquitous availability of the internet, had sparked the evolution of collaborative online projects. Today, these projects and their products are widely used in many domains. Free and open source software (e.g., from the GNU-Project https://www.gnu.org/ (accessed on 2 September 2019)) and wikis like Wikipedia (https://www.wikipedia.org/ (accessed on 2 September 2019)) are prominent examples that have a high economic value and are fundamental in the private and public domains by providing a variety of information at low cost (see for example Ghosh et al. [1] or Feick and Roche [2]). This widespread usage creates a need for quality, quantity, reliability, etc. of the products provided by self-organised communities. For example when using crowdsourced geodata during disaster response, quality is a key influence on the usability of the data [3]. In this context, it is important to note that crowdsourced data are social products whose nature depends on the character of the users creating them and the contexts within which they were produced [4]. Hence, studying crowdsourced data without considering the processes of production can lead to an incomplete understanding of a project.

The need to consider and understand the relations between contributors, production processes, and social contexts has led to studies analysing various aspects of participation in crowdsourced projects. For example, Bryant et al. [5], in one of the first studies of online communities, found that through increasing engagement, Wikipedia users developed an understanding of Wikipedia as a community and not just a random collection of sites. Studies of OpenStreetMap (OSM), a collaborative online mapping project in which anyone can add or edit entities within an open database of the world [6], have identified geographical biases in the locations of contributors [7] and different regional data development trajectories [8]. Other studies of OSM have attempted to categorise users into groups based on their behaviours [9–11] or to depict contribution trajectories over time, identifying different behavioural patterns and stages over users' 'life cycles' [12,13].

Nevertheless, such studies tend to focus on individuals and their engagement with the project. Less attention (at least in quantitative terms) has been given so far to the contexts users operate in, such as the organisation of the project itself and interactions with the community and their effect on the data production process (for qualitative studies see Lin [14], Palen et al. [15]). This study aims to contribute knowledge on the subject through a study of a specific social context and case study—social data contribution events (community happenings) in OSM—exploring how the behaviour of OSM users had changed after participating in such events. To answer this, we develop a novel conceptual and formal analysis framework presenting measures for assessing the extent of change in the type and location of contributions, as well as changes in users' level of engagement with the project. We consider the relations between these changes and variables such as level of experience and type of happening, hence adding an additional level of information. Through this multilayered analysis, we are able to expose how the intertwined relations between users, social contexts, and data unfold.

The following section briefly sketches the current knowledge on community happenings in OSM. The paper then proceeds as follows: In Section 2, the materials and methods of data extraction and analyses are described starting with the conceptual model (Section 2.2) and its respective user metrics (Section 2.3). The necessary used data sources are described in Section 2.4 and the section concludes with a description of the experimental setup (Section 2.5). Section 3 describes the results of the analyses starting with a view on activities during events followed by the impacts on newcomers to the OSM project (Section 3.1) and then on advanced mappers (Section 3.2). We interpret and discuss our findings in Section 4 before the paper concludes with a summary and outlook to further research potential and open questions (Section 5).

Related Work

The OSM project had not yet celebrated its first anniversary when its first "mapping party" took place in central London in 2005. During the day, a small group of interested mappers used an improvised infrastructure and Global Positioning System (GPS)-devices provided by the pioneers of the OSM project to create a free and open map of the surrounding [16]. Since then, different types of happenings have emerged out of this online community with different intentions, methods and results. To acquire and train new mappers as well as to strengthen the community bonds and practice the social aspect of OSM, social events are regularly organised and announced in the central and open calendar [17]. These events may include mapping parties where users meet for on-the-ground mapping of a small area; mapathons, which are time-limited events in which contributors map from afar a certain area in a synchronous manner; and platform mapping events, which are open-ended and done in an asynchronous manner (see Table 1). In this paper, we relate to the first two types, using the terms Community Field Mappings (CFMs) for mapping parties and Community Remote Mappings (CRMs) for mapathons. In CFMs, a considerable effort is made to survey the area by foot or by bike and collect data that then later gets digitised. CRMs, in contrast, are often referred to as "armchair mapping" and do not include the possibility to survey the to-be-mapped area. Nevertheless, these remote mapping practices

have proven to be powerful tools for filling data gaps in the OSM project that exist mostly in areas with low or medium human development [18].

Even though Coast [19], one of the founding members of OSM, sees happenings as a key factor of success for the project, the first scientific analysis comes to a mixed result [20]. The authors take an overall positive resume of their conducted happening especially in terms of the amount of the newly collected geodata, mainly GPS traces, and the interest of participants that were new to the project. It is however marred by the lack of achievements in the aftermath of the event where neither further mapping nor social follow-ups could be observed. Similar findings are reported by Mooney et al. [21].

Hristova et al. [22] asses the effects of 94 mapping parties in central London. Based on two metrics calculated over a period of six months, the authors show a strong positive impact of happenings on novice and less experienced users during and directly after the event in terms of engagement. Nevertheless, the analysed events fail in motivating new users to attend further happenings. Concentrating on disaster mapathons, Dittus [23] also looked deeper into the newcomer retention. He revealed a complex task design to be deterrent to newcomer retention while peer feedback and a social setting can be beneficial, even if the execution of the event is hindered by minor technical issues.

Describing the activities of the mapping network YouthMappers (https://www.yout hmappers.org/ (accessed on 2 September 2019)), Coetzee et al. [24] asses CRM outcomes beyond the pure data production and local social interaction. Concentrating on the personal development of the organisers and attendees as well as the events impact on academic practices and programs, the authors note similarities and country specific differences between the evaluated mapathons. Attendees and organisers gained interest and knowledge on foreign places and people's lives which could even lead to the inspiration of thesis on that topic. The impact of mapping events can go as far as changing universities software and data practices to be more open software and open data aware.

In 2016, Ebrahim et al. [25] took the idea of an open platform of untrained volunteers to a new level by introducing so called "Minimapathons". Students aged around ten years old were introduced to OSM and taught the necessary skills to edit the map. In a CRM setting, the children contributed a large amount of fairly accurate data while learning about new topics of geography and enjoying the event.

The studies referenced above provide evidence on how different social event settings and contexts affect the data product, the behaviour of users and even organisational practices. Nevertheless, each of these does this while relating to a specific context and event type. The literature is still missing a systematic analysis combining the presented user characteristic analyses and comparing the impacts of different types of events. Below, we present a conceptual and formal approach for assessing the effects of events and utilise it to compare the impacts of CRM and CFM events. We do so by analysing changes in behaviour for mappers participating in these events in relation to "control" mappers who did not take part in such happenings.

2. Materials and Methods

2.1. OpenStreetMap Dataset and Ecosystem

The OSM data structure is based on three types of spatial elements [26]—nodes representing single points (e.g., a post box), ways which are collections of nodes forming a line or a polygon (e.g., a road), and relations grouping multiple nodes, ways and relations together (e.g., a bus route consisting of bus stops and roads). Each element can be enriched with non-spatial information through tags consisting of "key = value" pairs (e.g., highway = residential or maxspeed = 50). Edits made by a single user over (usually) a short period of time are grouped in changesets [27], which can contain additional textual information (e.g., data source, details on events). Users can also post notes on the map in which they communicate information [28], e.g., about missing or erroneous data. These notes have been used in the past to estimate data quality and community activeness [29]. Besides the comment-text of a changeset or note written by the user these elements also have the

possibility to discuss the respective object using sub-comments. The project also includes a diverse and vibrant plethora of communication and organisation platforms used by active mappers and interested users, e.g., mailing lists or chats. One of the most central organising platforms is the Wiki (https://wiki.openstreetmap.org (accessed on 2 September 2019)) which can be edited by any registered user and which provides information about the project and guidance for mapping activities.

Contributions are made through various editors, frequently developed as part of the project, such as the iD editor available on the project's website or the Java OpenStreetMap Editor (JOSM) desktop application (https://wiki.openstreetmap.org/wiki/Editors (accessed on 2 September 2019)). The project is maintained by the OSM foundation, yet the foundation does not control the data. In fact, many institutions are involved in OSM activities, including governments and municipalities, commercial corporates, and humanitarian organisations. These, along with the individual contributors and the foundation and its multiple working groups, form a dynamic community (or a web of communities) in which social interactions and events, such as mapping events, work meetings and conferences, play an important role [30].

A single isolated contribution to OSM, meaning any modification of the database, can be categorised through the two dimensions of target region and type of social interaction [24]. The general contribution is done on an individual basis (see Table 1). Note that other contribution options like bulk imports of external databases into the project exist and that social interaction within the project is not limited to physical interaction but can also take place on digital platforms such as the Humanitarian OpenStreetMap Team (HOT) Tasking Manager (https://tasks.hotosm.org/ (accessed on 2 September 2019)). Both settings are excluded within this analyses. Edits can be characterised over a variety of factors (see below). We compare users edit-characteristics before and after they attended a community happening.

Table 1. Overview of different types of OpenStreetMap (OSM) contributions with respect to their contribution setting and data collection technique. The contribution setting describes the social circumstances of a contribution, i.e., the amount of (possible) direct social interaction. While an individually committed contribution is subject to no (direct) social constraints, a platform like the HOT Tasking Manager offers the possibility for exchange in chats and online discussions. A bulk import committed by a individual mapper is seen as vandalism according to the rules set by the OSM community lacking substantial community buy-in [31].

Contribution Setting	Field Mapping	Remote Mapping	Bulk Import
Individual	"Normal" Contribution		Vandalism
Community Happening	Community Field Mapping	Community Remote Mapping	Community Bulk Import
Digital Platform	Platform Mapping		General Bulk Import

2.2. Conceptual Model and Measurement Framework

Towards the analysis of the impacts of participation in community happenings, we define here a generalised conceptual analysis framework that groups these factors into five contextual dimensions relevant to community project contributions, especially Volunteered Geographic Information (VGI) (see Figure 1):
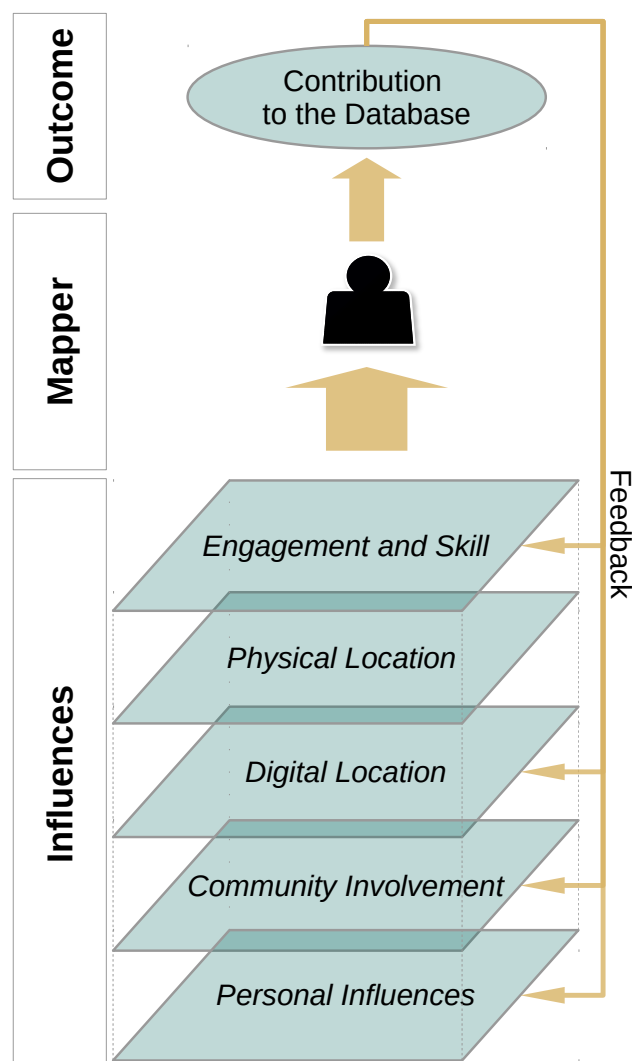
Figure 1. Influences or context a mapper is exposed to while making a contribution, i.e., characteristics of a contribution.

Engagement and Skill describes the knowledge a user has about the project, how experienced she is, and how willing she is to contribute. It can be assumed that skill rises with engagement as stated by Hacar et al. [32].

Physical Location is the geographic space or "real world area" a user is located in at the moment she makes a contribution as well as more generally the socio-economic background of a mapper as analysed by Neis et al. [33], Mashhadi et al. [8].

Digital Location is made up of two parts: the "data area" a user contributes to and, in case of VGI, the respective "geographic space" (compare physical location). The data area describes the digital data space around the contribution. Through their semantic or (geo)spatial proximity, these objects are related and therefore influence each other. Among others, this could be abstracted by attributes like the amount and type of (geo)data already present at a virtual (geo)location at the time of a contribution (see Figure 2).

In contrast, in VGI, the "geographic space" digitally (or virtually) visited fundamentally influences data production, e.g., through aspects like landscape appearance. It can be distant from the physical one (remote mapping) and therefore have different

attributes. A mapper may be located in a coastal European area while editing data describing features in a mountain range in Asia (see Eckle and de Albuquerque [3]).

Community Involvement is the interaction of a mapper with the community at a mappers physical and digital location. It is a vital part of any contribution that has among others been identified by Mooney and Corcoran [34].

Personal Influences are a supplementary "soft" layer of influences. These affect users in a number of different ways from the motivation or goal a user has during participation up to the personal interest, gender, psychological condition, etc. Some of these aspects have already been analysed (see for example Coleman et al. [35], Budhathoki and Haythornthwaite [9], Gardner et al. [36]).

(**a**)　　　　　　　　　　　　　　　　　　　　　　　　　(**b**)
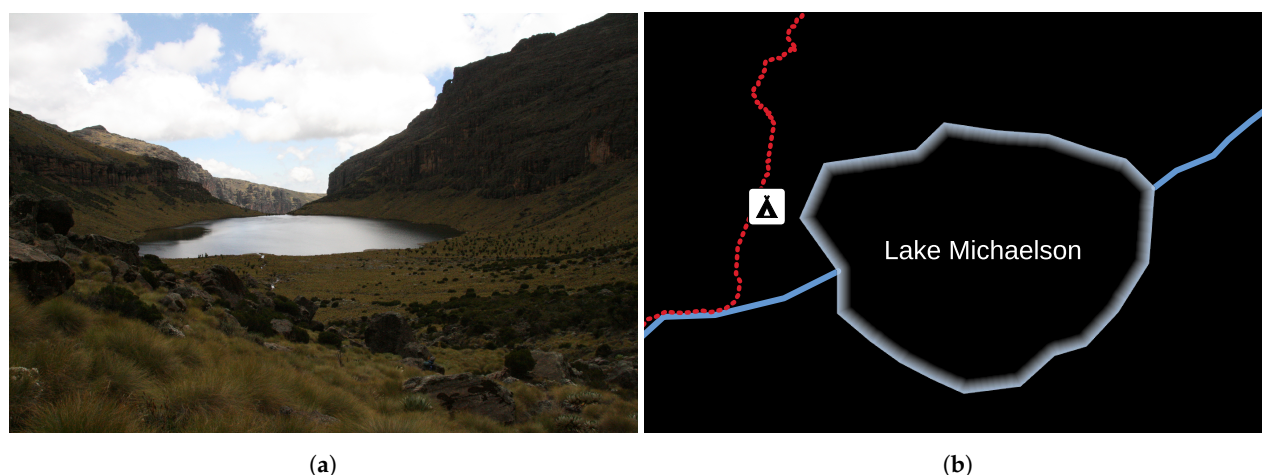
Figure 2. Example digital location. (**a**) The geographic space of a digital location (Mt. Kenya, Kenya). (**b**) The data area for the same digital location as viewed in Java OpenStreetMap Editor (JOSM). Only a small number of objects is present here in the data area with a mixture of natural and human made features.

The interaction and complexity of these influences varies not only between individuals, but also for the same individual over time. A community happening is a distinct event in a user's mapping career highly suitable to assess the influence on mappers behaviours and draw conclusions on the flexibility of users within their mapping routine.

The presented definitions of physical and digital location are tailored towards VGI geodata similar to OSM but can easily be generalised to other aspects of that data or other data in general by replacing the geographic aspect with its semantic one. For example the digital location might be framed as a certain topic the user engages in during a Wikipedia edit while the physical location may in that case incorporate the field the user normally edits or is an expert in. Through this notion, aspects like distance or data area respectively apply to the semantic distance or the amount and shape of contend within that semantic topic. This applies to different types of VGI data accordingly where other aspects of the geographic space play a major role. For example, nature observations from iNaturalist (https://www.inaturalist.org/ (accessed on 2 September 2019)) may put more emphasis on habitat and species aspects while twitter (https://twitter.com/ (accessed on 2 September 2019)) analyses could require a more detailed view of the immediate environments of users. The framework is also flexible in its application. While its usage is demonstrated on the case study of community happenings in OSM, it could be applied to situational or temporal analyses in general as well as to track the effects of any user impacting project changes such as changes in contribution mechanisms or data display.

Figure 3 presents a visual representation of the model outline based on the case study of OSM, depicting certain changes in the mapping behaviour of a hypothetical mapper, taking place after she took part in a happening. A number of hypotheses are thinkable of how mappers may react to events. The figure shows one possible reaction of

a CRM attendee. Before the event, the mapper slowly increases her skill and community involvement but always maps from the same physical location (e.g., from home) and in the same digital location. During the event, the mapper changes this behaviour by largely increasing her engagement and skill, mapping from a slightly different location (the event venue) and at a new digital location. This new impulse translates to a lasting impact on her behaviour, i.e., higher levels of involvement and an increased tendency to explore new digital locations. The personal influence is of random shape and no clear effect by the happening can be observed. The following section describes in detail how these abstract dimensions of effect are translated into concrete measurable indicators.
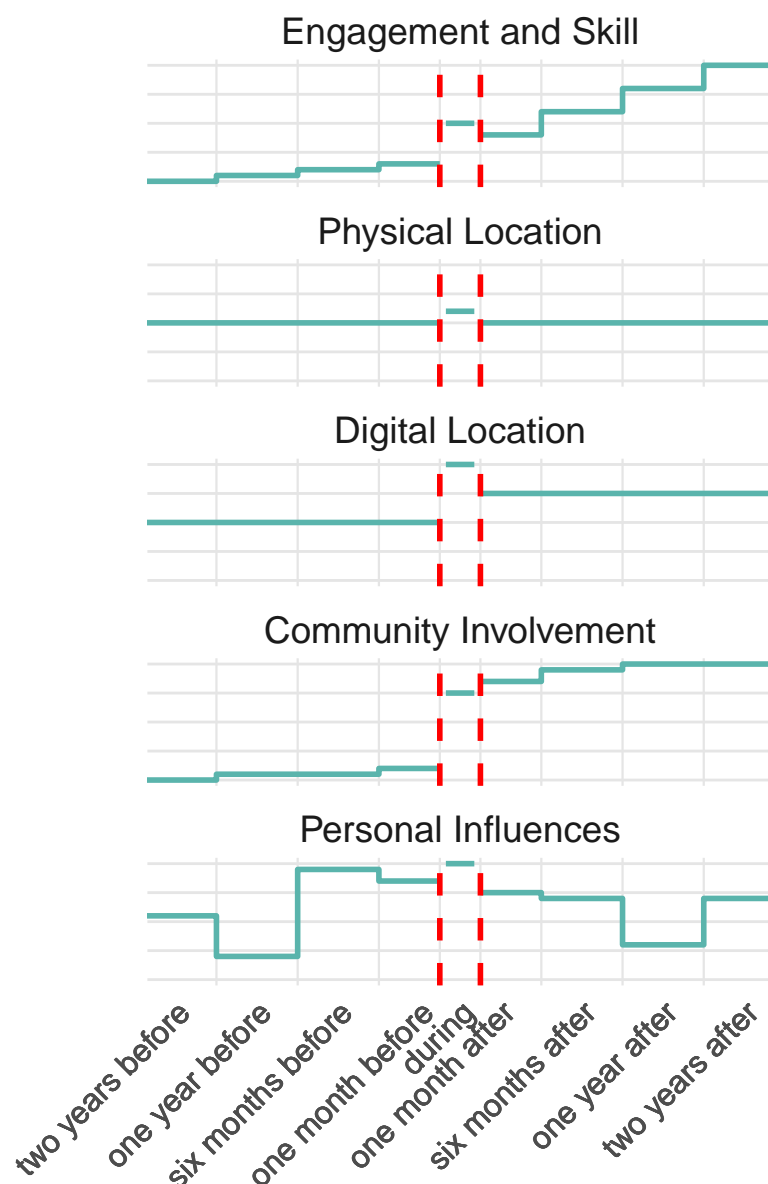
Figure 3. A hypothetical trajectory of mapping behaviour of a model mapper taking part in a happening along the five abstract dimensions of contextual effects. The x-axis represents different time intervals before, during, and after the happening. In this example, the mapper always contributed from the same place (physical location) but her engagement and skill, as well as her community involvement, rose over time and especially after the event. The personal influence seems to be somehow random while her digital location changed through the event. During the event, the mapper diverges from her usual attributes.

2.3. User Metrics

The described general framework is meaningless without the possibility for implementation. The following measurable indicators or proxies were implemented to capture the dimensions of contextual effects quantitatively from the OSM data (See Table 2 for an overview of the implemented proxy variables):

Engagement and Skill covers the users activeness and mapping style. The activeness was seen as the number of contributions to the OSM database; her editing style was described by the proportion of creations, tag and geometry changes as well as the diversity of edit types and the complexity of edits. The diversity was calculated using the Shannon Index of biodiversity, framing each edit type as one species (e.g., creation, deletion, etc.) [37]. The complexity on the other hand was measured on a six step ordinal scale from a deletion being the least complex action up to the creation of a very complex multipolygon. In addition, the quality of her edits were calculated based on the conformity with the JOSM validation tool. JOSM is one of the most used and sophisticated editing software available for the OSM database. It is shipped with an extensive rule set for automatic quality assurance that can be run against the data. It will report errors and warnings, where the requirements of the static rule set are not met [38].

Physical Location defines an approximation of the permanent residence of a mapper (home location). It was derived from the location of the event assuming mappers would participate in events near their residence (See Mooney et al. [21] or Danziger [39] for two edge cases where this assumption may be challenged). This information was available as the time and venue of an event are the minimum information required in event announcement texts. For the Control Group (CG) (see Section 2.4) the area with most edits throughout the users mapping career was used as home location, following one of the procedures suggested by Neis and Zipf [7]. The physical location was seen as a static attribute for each mapper assuming that far distance changes of residence are very rare. It also seems extremely unlikely that a mapper relocates as an effect of event attendance. Accordingly, this metric was excluded from the analyses.

Digital Location was defined by the digital region surrounding the edit issued. The digital locations or "data areas" were between 0.02 to 25 km^2 depending on the edits' extent. It was described by the element, tag and mapper density defined by the number of distinct elements and tags that were currently present in the area and the number of distinct mappers that edited the area in the past. In addition, the diversity of present OSM map features (see OpenStreetMap Wiki Contributors [40]) was calculated. Analogous to the contribution diversity the Shannon Index was used, this time grasping each map feature as one species. In addition, a set of distance (i.e., similarity) measures was computed to respect the fact that mapping in an area with similar attributes to the user's home area is easier than mapping in a completely foreign area. These measures were based on dividing the world into areas of equal attributes across the economic, cultural, population density and biome dimensions (see more details in Section 2.4). The implementation calculated the percentage of edits located in areas with different attributes than the home region (see physical location). For each mapper and each dimension, the world was thus separated in "regions that are similar to the home region" and "regions that are distinct to the home region", and the share of edits in the latter regions out of all edits was calculated. The definition of a "remote mapper" therefore changes from its solely geographical or spatial perspective to an attribute perspective here. A mapper was considered a remote mapper if she contributed to areas with different attributes in terms of economic potential, cultural groups, population density or landscape appearance compared to her home region.

Community Involvement refers to the community integration or community work of a mapper and was measured in terms of supplementary project activity (i.e., non

mapping activity). Users allocating time to the OSM project in addition to the actual mapping were seen as more integrated into the community. This dimension was captured calculating the length of changeset comments, changeset discussions as well as OSM map notes and notes discussions issued by a user. The length was defined as the number of distinct words used per comment. These features are of large importance for the OSM community but not mandatory in order to contribute to the database. The more information is provided through these channels, the higher the transparency and possible community interaction outside physical meetings or direct messages.

Personal Influences is a dimension whose measurement in a quantitative manner proved to be a complex task as it has not yet been done in the literature. We therefore omit this dimension but acknowledge its importance and strongly encourage future research in this field.

2.4. Data Sources

The analyses are based on the OSM full history, a feature recording any change to the database since October 2007 [41]. The OSM History Database developed by the Heidelberg Institute for Geoinformation Technology (https://heigit.org/ (accessed on 2 September 2019)) and described in Raifer et al. [42] was used to compute all metrics. It provides a lossless transformation of the original OSM data enabling fast parallel computing in a cluster environment. In addition, a database was setup locally, holding all information on OSM changesets using the ChangesetMD-tool https://github.com/ToeBee/ChangesetMD (accessed on 2 September 2019) and OSM notes using the notes and comments parser https://github.com/mapbox/osm-comments-parser (accessed on 2 September 2019).

Table 2. Overview of measured proxy variables. See Section 2.3 for an explanation of the variables, Section 2.4 for an overview of the involved data sources and Section 2.5 for a detailed explanation on the calculation of the variables. The description in this table refers to the implementation for advanced mappers. Absolute values instead of differences and the change rate were calculated for newcomers.

Group	Variable Name	Description	Source Involved [a]
	Quantity	Change rate of contributions	OSM (Nd, W)
Engagement and Skill	Creations share	Difference in share of creations	OSM (Nd, W)
	Tag-changes share	Difference in share of tag-changes	OSM (Nd, W)
	Geometry-changes share	Difference in share of geometry-changes	OSM (Nd, W)
	Edit diversity	Difference in the Shannon Index over abstract edits	OSM (Nd, W, R)
	Edit complexity	Difference median edit complexity	OSM (Nd, W, R)
	Quality	Difference in JOSM error delta	OSM (Nd, W)
Digital Location	Element density	Difference in element density in mapped areas	OSM (Nd, W)
	Tag density	Difference in tag density in mapped areas	OSM (Nd, W, R)
	User density	Difference in unique users per element in mapped areas	OSM (Nd, W, R)
	Area diversity	Difference in the Shannon Index on map features in mapped areas	OSM (Nd, W, R)
	Economic distance	Difference in the share of mapped regions of different economic status than the home region's status	NE, OSM (Nd, W)
	"Cultural" distance	Difference in the share of mapped regions of different "culture" than the home region's "culture"	HU, OSM (Nd, W)
	Population density distance	Difference in the share of mapped regions of different population density class than the home region's class	EU, OSM (Nd, W)
	Physical geography distance	Difference in the share of mapped regions of different population density class than the home region's class	WWF, OSM (Nd, W)
Community Involvement	Comment size	Difference in unique words per changeset comment used	OSM (C)
	Discussion size	Difference in unique words per changeset discussion used	OSM (C)
	Notes size	Difference in unique words per note action used	OSM (Nt)

[a] Nd—Tagged nodes; W—Ways; R—Relations; C—Changesets; Nt—Notes, NE—Natual Earth Contributors [43], HU—Huntington [44], EU—European Commission and Columbia University [45], WWF—Olson et al. [46].

Events were extracted from the central OSM calendar (https://wiki.openstreetmap.org/wiki/Calendar (accessed on 2 September 2019)). The collection of events requires a considerable amount of manual input as the OSM calendars' machine readability is limited. Most of the time, only the country, date and type of the event can be automatically parsed. All other information like precise venue, start and end times and confirmation that the event indeed took place had to be crawled manually in linked pdf-files or social media events. Maximum effort was undertaken during event data extraction to prevent a situation where the sample would be biased towards better organised happenings.

Event collection took place in early 2018. To ensure a minimum of two years of analysable OSM history after the last event, all events occurring from the 1 January until the 30 June 2016 formed the starting point of the analyses. In total, the calendar recorded 94 happenings in the analysed period. Events that did not provide sufficient information in their announcement text were excluded. Moreover, events that did not create a visible data impact, meaning an abnormal high number of contributions in an area or time interval that could be linked to the event, were excluded. This resulted in a collection of 54 events that were used in the analyses (25 field mapathons and 29 remote mapathons). During the event extraction process, it became apparent that the geographical distribution of events differed between the two types (see Figure 4). CFMs were held globally indicating active local communities in North and South America as well as in Europe and some countries in Asia. In contrast, CRMs were conducted nearly exclusively in Europe and North America, targeting regions in the global south. This distribution is very similar to the distribution of humanitarian aid donors and receivers and may be linked to limiting factors such as access to digital technologies.

All mappers that could be linked to the selected events were included into the analyses, filtering only obvious bot accounts. Because attendance is not recorded, users had to be linked to events in a separate process. For CFMs, all users editing in a 20 min walkable distance from the event location using isochrones from Openrouteservice (https://openrouteservice.org/ (accessed on 2 September 2019)) within the event time frame until midnight the day after were included adapting the methodology used by Hristova et al. [22]. CRM participants were added either based on their changeset comments mentioning the event itself or if they had a precise peak in contribution during the event. This resulted in 217 CFM participants and 436 CRM participants, meaning that in average CFMs were attended by 8.7 mappers while CRMs were larger events with an average of 15.0 mappers. The actual number of participants extracted for events ranged from three to 51. In addition, we extracted a CG of 500 users formed of randomly chosen mappers that contributed in the time period of the analysed events thereby assuring maximum comparability with the event-users. It was ensured that these mappers never used any happening related changeset comment in their contributions. This was verified by checking for changeset tags such as #hotosm or tags including the words "mapa"(thon) or "party". While this does not guarantee that they have never participated in a happening, it does make this occurrence less likely.
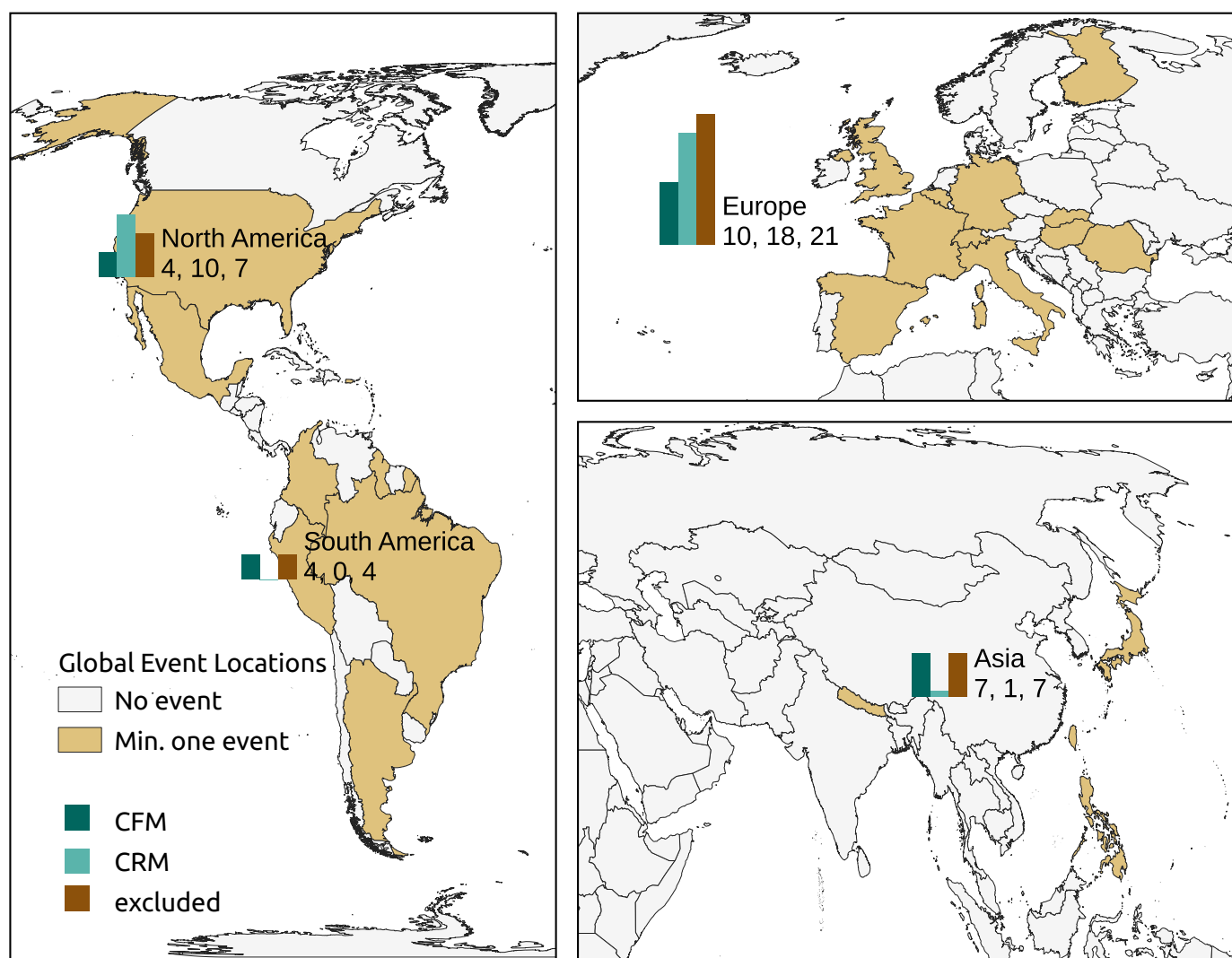
Figure 4. Distribution of event locations by type. While Community Field Mappings (CFMs) were globally distributed, Community Remote Mappings (CRMs) were organised nearly exclusively in North America and Europe. Source: Own work based on Natual Earth Contributors [43]. Projection: Eckert IV.

Table 3 provides an overview of the number or users included in the analysis. After a mapper registers to the project she is considered a potential contributor for any of the following periods. Newcomers are mappers joining the project during or shortly before the event. They are therefore potential contributors in any time frame after the event (see for example the $N = 76$ for CFM newcomers). Nevertheless, newcomer retention is low, and only a small amount of newcomers will stick to the project and continue contributing thereafter. For example, only five out of the 76 newcomers for CFMs actively contributed one year after their respective event (n). This confirms the findings by Dittus [23] that OSM exhibits a low newcomer retention. Advanced mappers on the other hand have a mapping history before the event. The influence of happenings on this group can be analysed in terms of changes in mapping behaviours after the event in relation to pre-event patterns. Advanced mappers were analysed for the maximum time interval of their OSM presence before the event, meaning that an advanced mapper who joined one month before the event was only analysed for the same time frame after the event (N). Just like newcomers, advanced mapper may choose to not contribute to the project for a longer period, resulting in fluctuating values for n, the number of actual contributors to the analysed time intervals before and after the event.

Table 3. Summary of number of analysed users per category for the different statistical tests. *N* is the number of potential attendees qualifying to be included for the respective time frame and *n* the number of actual contributors to the respective time frame. All newcomers that participated at an event are potential contributors thereafter, but since not all users stayed active, *n* is decreasing with time. Advanced mappers were analysed for the maximum time interval of their OSM presence before the event, meaning that an advanced mapper who joined one month before the event was only analysed for the same time frame after the event. Therefore, N is decreasing with time for this group. In addition, advanced users may become inactive for some time intervals as well, leading to a fluctuation of *n*. Users had to be ignored (None) if they were not registered for a full month before their respective event but also did not classify as newcomers.

Seniority	Time Interval	CFM		CRM		CG	
		N	n	N	n	N	n
All Participants	during	217	217	436	436	500	500
Newcomer	during	76	76	214	214	279	279
	one month	76	25	214	57	279	26
	six months	76	14	214	27	279	36
	one year	76	5	214	21	279	33
	two years	76	5	214	21	279	33
None	during	18	18	33	33	3	3
Advanced Mapper	during	123	123	189	189	218	218
	one month	123	84	189	61	218	32
	six months	102	81	139	71	185	65
	one year	93	68	115	58	168	59
	two years	76	63	85	49	133	63

In order to identify changes in the mapping behaviour of users regarding the mapped regions, a global homogeneous dataset was necessary. The analyses targeted changes in terms of real world attributes of the mapped regions. Therefore, a set of descriptive attributes that were estimated to directly influence mapping in a region was used. The economic information from the Natural Earth dataset [43] was extracted alongside with the cultural regions defined by Huntington [44], a one kilometre resolution global population grid [45] and Olson et al. [46] global biome dataset.

The Natural Earth dataset groups countries into four groups from low income to high income combining multiple sources like the World Bank. The cultural dataset was derived from Huntington's [44] system of cultural regions and consisted of ten regions defined by their dominating cultural influence such as Orthodox or Sinic. Huntington's [44] work on cultural regions received considerable criticism, yet this was directed more towards his simplification and stereotyping of "culture" and the resulting conclusions and not towards the system itself. Hence, we choose to use his regions here to delineate cultural regions. For the global population, the dataset provided four groups ranging from sparsely populated areas with less then 300 inhabitants per km^2 to urban clusters (more than 1500 inhabitants per km^2) while the biome dataset was the most diverse with 16 categories such as Tundra or Temperate Conifer Forests. Each contribution was then enriched with these four attributes depending on the location of the contribution.

2.5. Experimental Setup

The users were separated into six groups dividing them by their level of experience (newcomer vs. senior mapper) and the type of event they participated in (CFM, CRM and CG). Newcomers were defined as users having less than ten changesets and being registered for less than three days at the beginning of the event. From a data point of view, contributions and actions in OSM are spontaneous and momentary. They consist of an operation at a precise point in time not directly depicting the possible processes before and

after. Therefore, five fixed complementary time intervals relative to the event were defined (see Figure 5). The described user metrics were calculated and summarised within these intervals of zero, one, five, six and twelve months. The event duration (i.e., t_0 or the zero time interval) is the time of the event for CRMs, the time during the event until midnight the next day for CFMs and the complete first half year of 2016 for the CG.
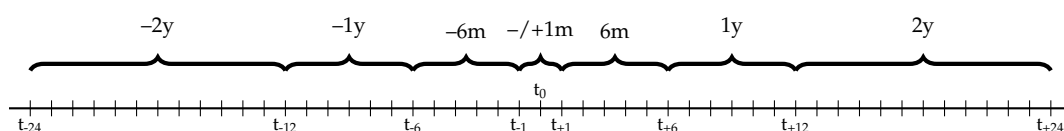
Figure 5. Summary time intervals for analyses. t_0 defines the event. Variables were calculated for the displayed time intervals during the event, one, six, twelve and 24 months before and after.

During evaluation, change (i.e., difference or change rate) in calculated metrics between corresponding time intervals was calculated and compared between groups for senior mappers. Newcomers do not have a mapping history preceding the event wherefore absolute values were used instead. Due to the large data size, the data area was analysed in monthly time intervals prohibiting an evaluation for shorter time intervals like the few hours of the events duration as well as the one month analysis interval. The sample of participants was divided into six groups defined by happening type and seniority of the mapper. For each dimension of contextual effect and each time interval, a separate pairwise comparison between participants of the three happening types within the same seniority group (i.e., between each pair of event groups) was carried out. For the analyses during the event, mappers were not separated by their seniority level.

A general Kruskal–Wallis rank sum test was used to detect effects of happening participation, meaning the verification of the h_0 hypotheses that events have no effect and participants from all three event types have a common population. In case the test indicated a difference among the groups (an effect of happenings), follow up pairwise Wilcoxon rank sum tests were executed to pinpoint the differing groups. This resulted in 432 individual tests (three pairwise tests comparing different happening types, for each of the 18 dimensions of contextual effects, each of the 4 time intervals, and each of the two mapper types: newcomer or experienced). For effects during the event, all mappers were grouped together regardless of their experience, adding additional 54 tests (considering the number of happening types, dimensions, and only one time interval). A 5% level of significance was used to distinguish between significant and insignificant effects. The method introduced by Benjamini and Hochberg [47] to adjust the *p*-value for multiple comparisons was used. The following section describes the results of this procedure where any effect or non-effect reported is significant with an adjusted *p*-value smaller 0.05. To ensure readability of the text, actual *p*-values are not reported but can be found in the supplementary materials. An overview of the implemented metrics, their calculation and the involved data sources can be found in Table 2.

3. Results

In total, 25 CFMs and 29 CRMs were analysed. During the event and mapper acquisition, it became apparent that different happening cultures exist. Aspects like organisation, location, duration, etc. did not only differ between the two happening types but seemed to be also bound to regional needs and preferences. For example CFMs take place globally, attract more experienced mappers and are sparked by an active local community who autonomously identifies incompletely mapped regions, mostly in rural areas. CRMs are nearly exclusively organised in western Europe and the USA but are more attractive for newcomers. Happening types also differed in their practical conduction. While most CRMs had a duration of two to three hours and were conducted mainly on weekday evenings, CFMs lasted for up to a whole day or even week and mostly took place on weekends.

For the effect analyses, mapper attributes were first compared between groups during the event realisation. Mappers behavioural changes after the events could then be interpreted in relation to the behaviour during the events. Figures 6–8 display a subset of the most important findings.
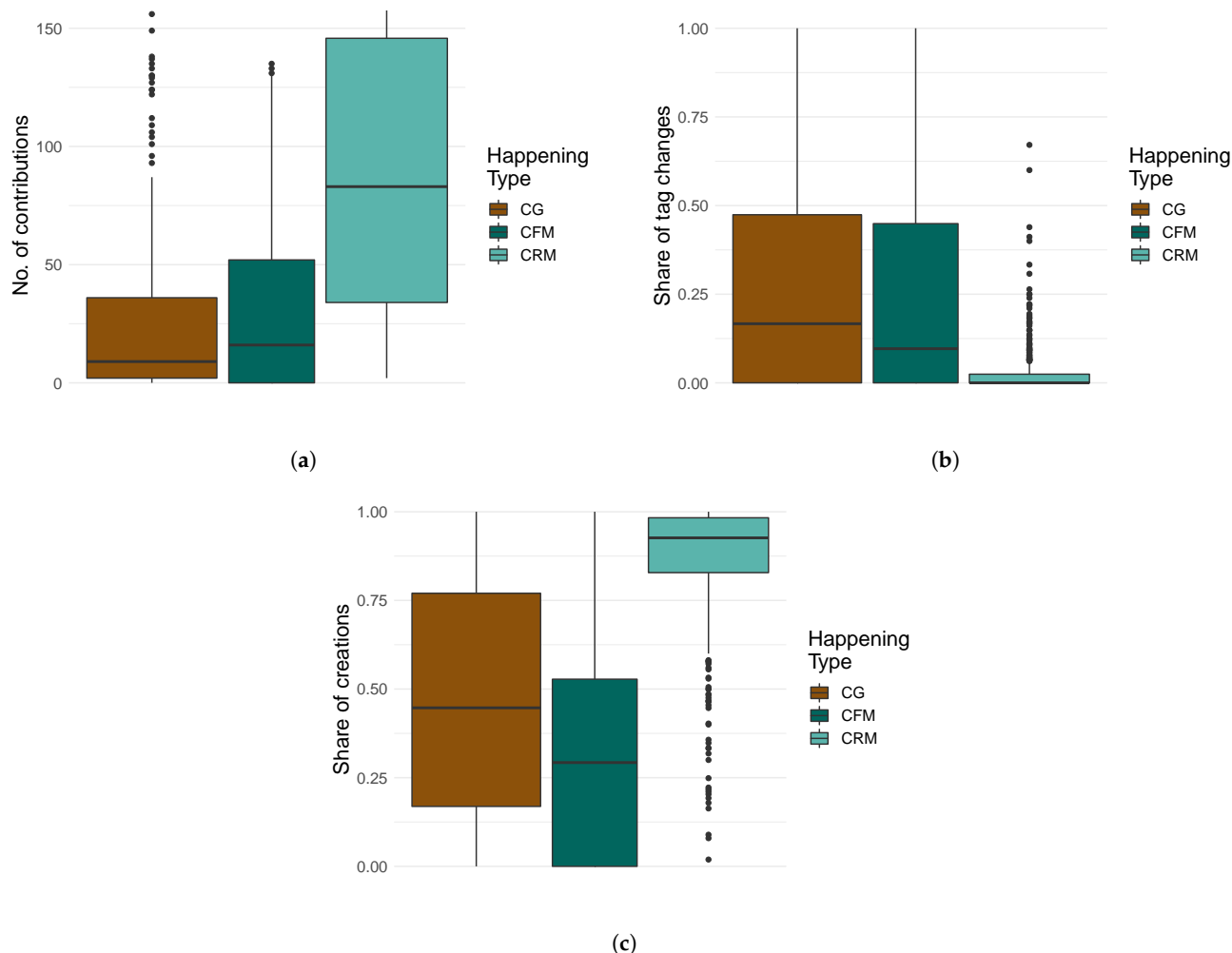
Figure 6. Three of the 14 analysed metrics during events: (**a**) contribution quantity, (**b**) tag changes share and (**c**) creations share. The median quantity of edits and the share of creations were significantly higher for CRM mappers while having a lower tag change rate. CFMs created less data. The width of the boxes is relative to the number of mappers in the respective class (see Table 3). Graph (**a**) was cropped at 150 contributions, but the total maximum was 90,107.

Engagement and Skill. Generally, CFM attendees showed similar behaviour to CG mappers. CRM mappers on the other hand presented unique behaviours, making more contributions (83, CFM: 16, CG: 9, Figure 6a), focusing on creations (86% of edits vs. 33%, 47% for CG and CFM, respectively, Figure 6c) more than on tagging (2% vs. 23%, 28%, Figure 6b), and showing lower diversity (0.77 vs 1.17, 1.23) and more complexity in contributions (average of 2.14 vs. 1.34, 1.84). The edit diversity was generally low among all three groups (maximum measured: 3.53, maximum possible: 8.30). One area in which CFMs were different from the CG was the quality of data, with CFMs producing on average 0.04 errors per contribution, while the CG produced 0.11 errors (CRMs were the least accurate with 0.33 errors per edit).

Digital Location. As the name indicates, CRMs mapped abroad with a mean of around 98% of edits in foreign economies, cultures and physical geographies and 84% in areas with different population densities compared to their home region. Both other groups

rather tended to map "at home" with around 84% to maximum 98% of mean edits in regions of similar nature compared to their home region.

As mentioned in Section 2.5 the digital area analysis was only available in monthly resolution and is therefore not analysed for the short time interval during the event as well as the time interval of one month after the event.

Community Involvement. The community measures applied showed that the CG users had longer changeset comments (mean of 3.44 unique words per changeset) while CFMs had only 0.07 words more per changeset compared to CRMs (CFM: 1.56, CRM: 1.49). No distinction could be found for the discussion size while notes were more actively used by the CG (mean 0.85 unique words) and CFM users (0.28) but not at all by CRM users.
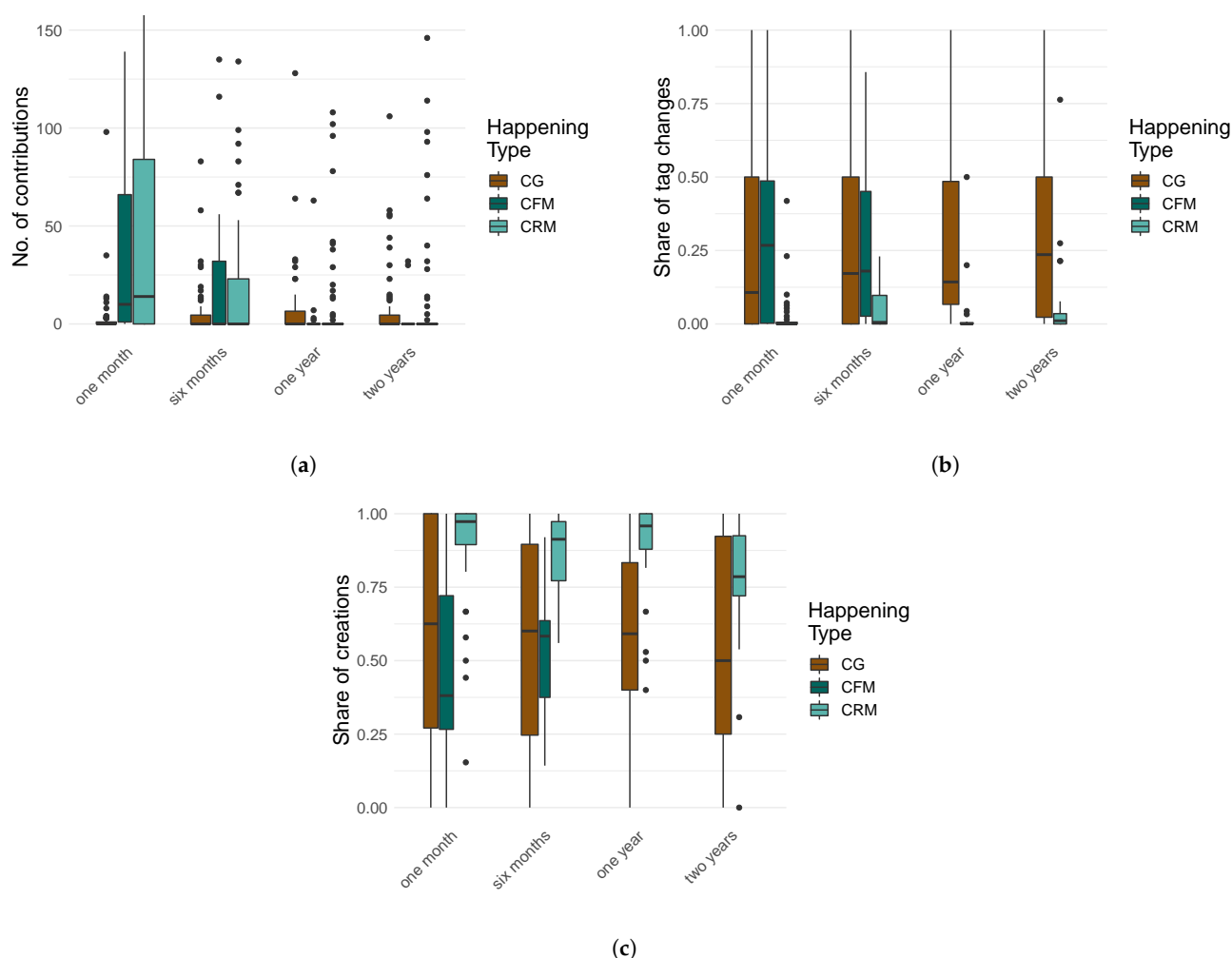
Figure 7. Three example graphs showing the effects of happenings on newcomers: (**a**) contribution quantity, (**b**) tag changes share and (**c**) creations share. The quantity was higher in median for the one month interval but faded later. The share of creations and the share of tag changes seemed to be preserved from the respective happening. The width of the boxes is relative to the number of mappers in the respective class (see Table 3). Graph (**a**) was cropped at 150 contributions, but the total maximum was 21,582. In (**b**) and (**c**) CFM users could not be analysed for the one and two years time period due to the sample size falling below ten.

3.1. Effects on Newcomers

Engagement and Skill. Although the quantity of data produced one month after both event types was ten to 14 edits higher in median than for the CG, the mean values were for both types lower. This high median quantity could not be maintained on the long term with more than 50% of the mappers not contributing after one month.

Field happenings tended to have a much lower mean mapping activity volume among their participants than the CG in the one year time interval, meaning six to twelve months after the event (see Figure 7a). Individuals' activity sparked within the different time intervals with only a handful of mappers continuously contributing over multiple intervals.

Those CFM mappers who stayed active had a distinct style of contributing with a higher share of geometry changes and a higher diversity of edits. Remote mappers stayed active in creating data and doing complex edits (see Figure 7c). These effects survived for most of the analyses time period. The same applied for the tag changes share that stayed very low compared to the CG (see Figure 7b). The geometry changes share for CRM mappers alternated around the CG in mean but stayed slightly higher in median for the one and six months interval. The quality of data produced by CRM mappers was at best equal to the other groups.

Digital Location. The data area edited did, in some time intervals, differ when users had attended a happening. The element density, meaning the number of elements per area surrounding the edit issued, was lower within the period of one to six months after the event with only 67 (CFM) and 90 (CRM) elements per km^2 (CG: 670), but the groups then started to blend in with the CG. The tag density on the other hand was lower only for CRMs in the one to six months and one to two years periods after the event with only 1.4 tags per element for the six months period (1.1 for the two years period, CG: 2.3 in both intervals). The mapper density did not differ between the groups.

The regional distance on the other hand stayed strongly influenced by events. CRM newcomers stayed remote mappers through the analysed time frame, and field mappers mapped exclusively at home and were indistinguishable from the CG which also mapped mostly at home.

Community Involvement. The community integration of newcomers that took part in events was mostly equal to the CG. Medians and means alternated between 1 and 5 unique words per changeset. Changeset discussions and notes were nearly unused by newcomers in general. Insignificant spikes existed where single users used these features.

3.2. Effects on Advanced Mappers

In general, advanced mappers were less affected by happenings than newcomers. Less significant group differences could be found. While the number of effects rose with time their occurrence seemed irregular.

Engagement and Skill. Advanced happening mappers displayed an increase in mapping in the one month period issuing 294% (CFM) to 476% (CRM) more edits in mean than before the event while the CG had an increase of only 2%. Median values for the CG and remote mappers were 0% and for field mappers 100% due to the large number of non-recurring contributors (see Figure 8a). The mapping type of advanced event attendees did not shift at all. All groups slightly decreased their creations share by around 0.01 to 0.04 (see Figure 8b) while tag changes and geometry changes alternated around 0. The edit complexity did not differ among groups while the edit diversity seemed to generally decrease in all time intervals.

Digital Location. The preferred digital area was only in one case affected by events, when participants after one year turned more towards areas with less users then the CG (CG: −0.01 mean editors per element; CRM: −0.68; CFM: −0.49). Generally, the mapped data area became more and more dense in all time intervals with all groups having mostly positive mean values of up to +274 elements per km^2. The regional distance was mostly not affected by happenings.

Community Involvement. No effect on the community involvement could be observerd for advanced mappers.
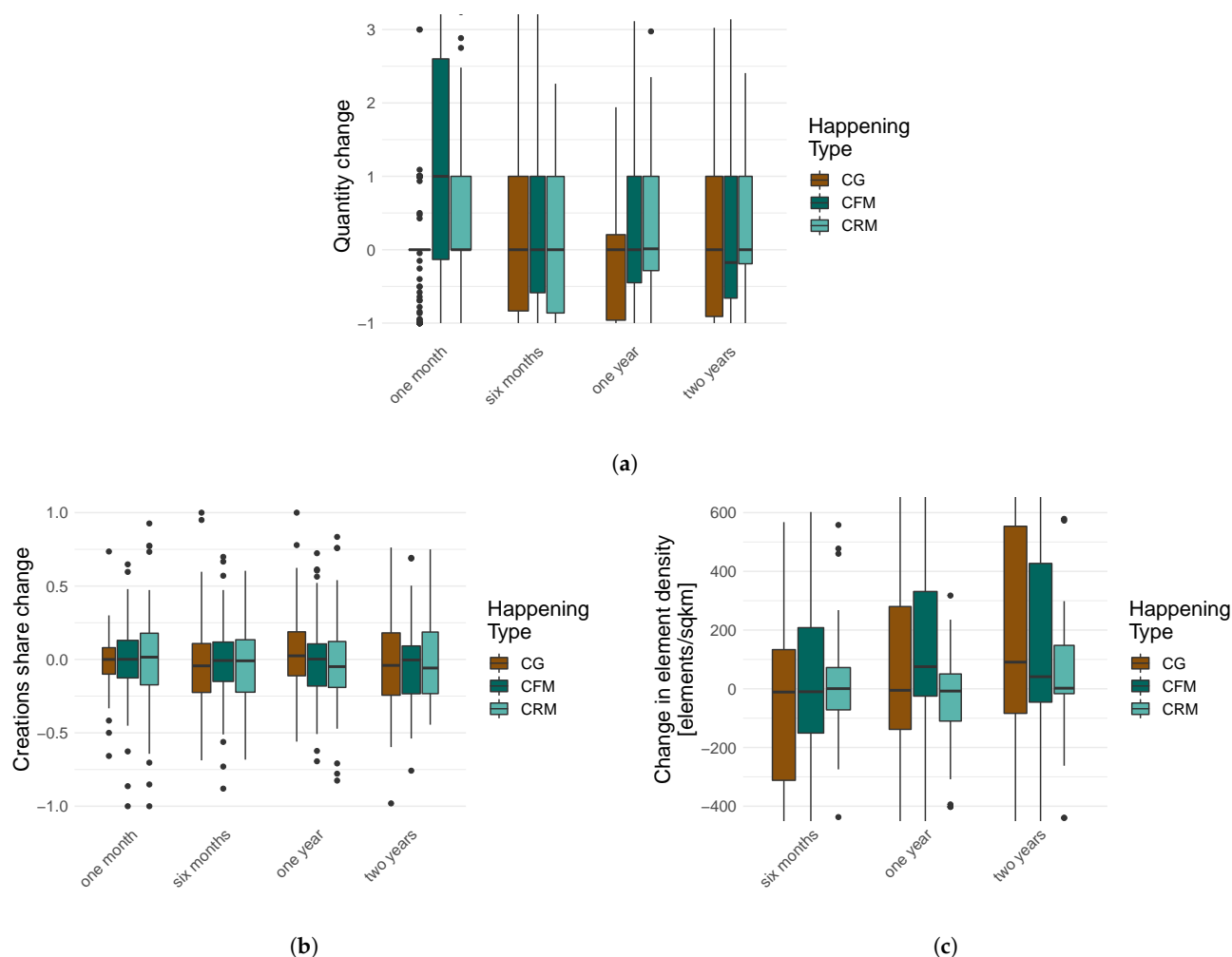
Figure 8. Three example graphs showing the effects of happenings on advanced mappers: (**a**) contribution quantity change, (**b**) creations share change and (**c**) element density change. In contrast to the analysis of newcomers, differences in values before and after the event for matching time intervals are shown here. The quantity produced by happening attendees increased for one month while a large part of the CG decreased their quantity in all time intervals. While the share of creations did not change through an event, the element density mostly increased for all groups over time. The width of the boxes is relative to the number of mappers in the respective class (see Table 3). Graph (**a**) was cropped at 3, but the total maximum was 22,191. Graph (**c**) was cropped between −400 and 600, but the total range was −12,885.92 to 1032.89.

4. Discussion

The combination of the organisation, attendee and mapping types that were distinguishable between event types resulted in different editing schemes during events. Apart from the obvious discrepancies in mapping distance between CFMs and CRMs (local vs. remote), data production differed from the CG in many ways.

CRMs were wells of new OSM data with a clear focus on the fast creation of map elements. As most tasks asked for building or highway data, their contribution diversity was low. This concentration on buildings and highways resulted in relatively complex edits with more than ten vertices per linestring or polygon. However, data already present at the mapped location might stay outdated as geometry changes were rare, and tags did not get updated frequently. The low quality might be linked to the high attraction of newcomers who even continue to contribute low quality data after the event. Another explanation could be the missing semantic information that would be needed to satisfy the static JOSM validator filters. Necessary information like the address of buildings is hardly mappable from afar. However, this lack of semantic details enabled mappers to

create large amounts of data when compared to both of the other groups during a relatively short time interval with many mappers having 50 or more contributions, mainly creations. These findings extend the new insights gained by Herfort et al. [18] by not only confirming the notion of CRMs as sources for new data but also analysing in detail what type of data of which quality is created. During the event collection, a strong bias in the geographical distribution of CRMs was observed. While Herfort et al. [18] showed how that bias impacts data production, more detailed research is needed on the biases created through CRMs within the OSM community and how these biases can be overcome.

Participants in CFMs showed a mapping type that differed only in few cases from the CG. Considering that the CG is a representation of "the general" OSM mapper, CFM attendees were a rather non-specialised or diverse group compared to CRM users. They did not explode in activism during their happenings due to the complicated data collection process but practiced a balanced mapping type with high quality of resulting objects. Nevertheless, participants in CFMs managed to create the same quantity of edits as those of the CG, notably in a much shorter time frame.

The community measurements clearly showed a lack of sensitisation of attendees of both happening types for extensive commenting and feedback. Discussions on problematic or controversial edits might have taken place on the spot at the event before uploading. This might explain the observed short comments and few changeset discussions. The feedback and personal interaction available during the event could lead to carelessness towards external OSM contributors who might visit the data area later finding vacuous changeset comments. In addition, the vast number of intended edits as well as sometimes proposed default changeset comments may hinder users to think of longer comments.

Our findings therefore support the distinction of the three types of happenings whose attendees can be summarised as follows:

CFM A group of skilled users mapping rural areas close to their home region, editing the local data or creating new data as necessary but concentrating on quality rather than on OSM community interaction.

CRM A group of new and advanced users digitising buildings and highways all over the world in large amounts with lower quality but higher complexity.

CG The general OSM mapper mapping mostly at home but sometimes abroad and having a balanced creation to change ratio and a medium quality of contribution describing well what she is doing.

Turning towards the effects of the happenings, it can be said that newcomers were far more affected by happenings than advanced mappers. The community event motivated them to have more contributions for at least one month. Unfortunately, this effect did not last, and event attendees then blended in with the crowd or even got demotivated.

This is partly due to low retention rates creating an enormous set of users that do not contribute at all. One could expect that event newcomers might be retained less often because, instead of solely and actively deciding to join the project, like for the CG, numerous attendees of mapathons presumably came because of a social pressure by their peers or took the advantage to "have a look", but even for the CG, the retention rate is devastating, it seemingly being a project internal issue. Knowing that online communities are highly shaped by extreme contributors [7], the nativity of at least one strong mapper from the CG can be seen as a success the other event types lack. Assuming a positive impact of these strong mappers on the project, it is, according to Thebault-Spieker et al. [13], important to attract them as they are hardly made through happenings or other activities. The role of happenings for newcomer activation and retention therefore might be overestimated by the literature and the projects community. Our findings show that a large share of users join the project outside these events on an individual basis (see Table 3).

The lack in community integration might be due to the limited time available at happenings. Learning the mapping process itself is already time consuming and teaching

newcomers about the importance of the community might just not be doable. Another explanation could be the use of in-person discussions or other discussion tools (like the HOT Tasking Manager) that hinder a transparent discussion on the OSM platform, but the missing community integration might also be linked to the observed low retention rate and negative tendency in quantity.

Those newcomers that nevertheless decide to join the project for a longer period are highly influenced by the mapping type they first came in contact with. They continue to create data remotely for CRM newcomers or edit local data for CFM newcomers. Both mapper groups focus, at least for six months, on sparsely mapped areas after the event. It can be presumed that this is also the case during the event as well as for the one month period but cannot be verified due to the technical limitations described (see Section 2.5). Sensitising new mappers in that way would be a great accomplishment by happenings towards a healthier user and database. Apparently newcomers are motivated by events to help overcoming one of the major problems of OSM being the data gap between rural and urban as well as high and low income areas [7].

Advanced mappers already found their mapping type and speed and are therefore less influenced by happenings. The results show that advanced mappers fall under general editing trends more than they are affected by events. Our results nevertheless show that they are highly motivated to increase their mapping after the event. This effect might though be a residue of the methodology attributing any contribution after the event to the one month period. Advanced mappers might tend to stay longer at events than the duration mentioned in the invitation, and therefore, their contributions may fall into the one month time interval although they are still contributing at the event itself. The general decrease in edit diversity may be due to senior mappers concentrating on a specific topic or edit type they feel comfortable with. This notion is supported by the fact that users increase their quality over time getting experts in their field of mapping. These findings are in favour of Bégin et al. [12] over Thebault-Spieker et al. [13], claiming that users develop their mapping preferences over time.

Surprisingly, the observed spike in contribution amount for advanced mappers shortly after the event did not lead to any change in their editing behaviour. The mappers seem to be mostly inert to the influences of happenings and will even continue eventual cleanup work after the event in their usual style. This non-effect may indicate that advanced mappers take part in happenings that suit their own mapping routine meaning that, for example, advanced remote mappers take part in CRMs. In conjunction with the findings on newcomers, this would mean that the first event decides not only on the mapper type that emerges but also on the type of happenings that will be visited thereafter. Each of these happenings may then only strengthen the mappers bond to that specific mapping style.

The effects of general editing trends can also be seen in the element density of advanced mappers (see Figure 8c) where the element density in the region mapped in two out of three time intervals increased. This is most certainly due to the fact that the overall number of elements in the OSM database, and therefore, the "global element density" increases over time. If advanced mappers continue to fill in missing elements instead of moving to regions less densely mapped, this would lead to the observed pattern. On the contrary, very active advanced mappers might also be the cause for low user density in the areas they map by being one of their main contributors and reducing the mapping opportunities for newcommers.

Many happenings rely on advanced mappers besides the organisers to train newcomers. Looking at the community measures during the events that are partly due to these seniors, it is no wonder the newcomers did not learn how to integrate into the community. As stated, a low community integration was found during events which results in an overall low community integration of new mappers joining through events. The same is true for the advanced mappers themselves who did not get any hang to a more communicative behaviour though the events.

Most findings are in line with the current literature on mapathons. Hristova et al. [22] have equally found more activity shortly after the event and a larger effect on inexperienced mappers. The newly gained insights show that this effect mostly does not last for a longer period. The low newcomer retention rates and fast detachment from the project stated by Neis and Zipf [7], Dittus [23] could equally be confirmed. More knowledge on the type of mapping practised by those newcomers that could be motivated was gained in our analysis. Our results attest to the value of the analysis framework developed in this work, uncovering unique behavioural patterns across multiple dimensions, as reflected in Figure 3. For example, our findings show that in contrast with our hypotheses in that figure, trajectories of advanced mappers remain rather stable and flat-lined.

Some inevitable sources of error need to be considered though, when drawing conclusions from the presented results. The bias in users collected for the different happening types can have unforeseen influence on the result production. Before the event even starts, it already draws the attention of distinct mapper groups. Trends within groups and subgroups of users may interfere with the effects from happening attendance. Advanced field mappers, for example, may simply have more edits over time, indifferent of their happening participation. The different lengths of time frames may also influence the results in the same way, possibly covering event effects. In addition, mappers may choose to visit multiple happenings during the analyses period whereby effects overlap or interfere.

The methodology used to attribute a certain home region to CG mappers may cause their home activity to be overestimated. Each of their analysed contributions within a certain region increases the possibility for that region to become their home region possibly resulting in circular reasoning for some mappers. Furthermore, the sporadic nature of OSM contributions may cause the metrics in a certain time interval to be highly influenced by only a few mappers and only seldom by the same set of mappers. Although, e.g., 426 users are incorporated in the six months analyses interval for advanced mappers, the number of actual contributors in both relevant time frames is only 217, and the number of contributions influencing the metrics might be very low. A vast majority of users might decide to not contribute at all making the result less robust. In addition, if the number of contributions for a single mapper is low, the percentage value (e.g., share of creations) can highly fluctuate creating a near random outcome in the results. A similar issue is linked to the event duration of the CG extending over six months wherefore the retention rate can be underestimated. A newcomer registering on the 1 January 2016 may have contributed one month later, but the contribution was classified as being during the "event". In the same way, other relative metrics like the contributions during or the contributions one month after the "event" may be over or underestimated.

Future analyses might tackle these issues by adjusting the measuring methodology. We will turn our research interest towards some possible adaptions that we present in the following section. Nevertheless, the stringent and explainable results of the presented study in line with the current literature are an indicator for a high quality data basis and an expedient methodology.

5. Conclusions and Outlook

The presented work defines a new conceptual framework for analysing user behaviour within online communities. By directly providing a possible implementation through distinct proxy variables and closely analysing the outcomes, the presented work paved the way for a comparable and comprehensible methodology across studies. The open and abstract definition of the framework facilitates application to other online communities if user contributions are recorded in a similar manner, yet the sporadic and spontaneous nature of contributions limits the direct temporal analyses. To overcome this issue, new methodologies need to be researched. For example, measuring progress through units of contribution volumes rather than absolute temporal units (e.g., dates, hours) could increase the robustness of statistical tests. By comparing equal quantities of contributions between mappers, the data can be split into comparable blocks. In that case, quantity would be

defined as the speed at what these equally sized blocks of contributions were achieved by each mapper.

In addition, the presented use-case extended the current knowledge on the role of social events within crowdsourced projects by drawing a more detailed picture on effects during and after their conduction. It also presented one of the first comparisons between CFMs and CRMs. With the security of statistical tests, by applying a vast range of mapper properties and by analysing the effects on senior mappers with respect to their pre-happening mapping behaviour, happening organisers as well as OSM data analysts may use these insights to draw conclusions on how to improve their own work. For example, organisers may revise their happening organisation scheme in order to attract a more diverse user and newcomer base for their happenings. Data analysts may use the attendance of a CFM happening as a quality indicator knowing that these events create high quality data and mappers. Other research fields concerned with projects relying on a community-like user base may use this research as reference to gain more insights into the "cultural information" shaping their "technological information" (see Sieber and Haklay [4]).

It can be concluded that happenings shape the OSM database and community in multiple facets. The community is enriched with newcomers that might not have been attracted by the project itself, but at the same time, the community integration of those newcomers largely fails. The database is enhanced by happenings concentrating on areas that have not been well mapped so far, yet users seem to not stick to these areas. In addition, CRMs themselves and their influence can be a great source of new data while CFMs enhance the database by updating and refining existing data. Thereby happening attendees and "normal" OSM users create a synergy improving OSM as a whole.

The exploratory character of the methodology nevertheless poses more questions on the effects of happenings and the composition and integration of the community. We will direct our future research towards an automatic event detection and attribution mechanism. The results show that happening attendance has a positive influence on a mapper's contribution quality, turning mapping attendance into a valuable quality proxy. More future research interest should also be directed towards regional schemes in happening culture. We found direct links between the event location and its organisation, conduction and the like. For example, event announcements in the US focused on the humanitarian aspect of the events while Spanish invitations provided little to no information presumably relying on word-of-mouth advertising. A more profound understanding of these schemes and their success may help future event organisers to establish new and more diverse events within their local community.

Apart from the location of the happening itself, socio-economic attributes of mappers may influence their reaction to happenings. For example the outliers among the CRM participants in Figure 6 may indicate that happening attendees can be further subdivided into groups that present distinctive behaviours during or after events. Yet this information is not available in a quantitative study as it can neither be inferred from the data itself nor gathered through quantitative methods. Future qualitative studies should therefore analyse how aspects like gender, age, or formal level of education of mappers influence the reaction to happening participation. Other project intrinsic attributes like the degree of specialisation are also thinkable. This would not only enable happening organisers to provide attendees with the support that fits their needs but also would open new advertising opportunities for happenings targeted directly towards mappers that would be most affected by participation.

The new metrics proposed here might also be of use in other contexts but need more investigation to ensure their usability. A general global or regional evaluation of the behaviour of users within these metrics can lead to insights into, e.g., the way mapping changes in general. Does, for example, the suggested assumption on the global data area density frequented by mappers hold true (see Section 4)? Are mappers actually mapping in more and more dense areas as more and more elements, tags and mappers are present in the database, or does mapping rather move to other areas where finding new elements

to map is easier? Links to the underlying data would expose how mappers adapt to a changing data world.

Through this work the virtue of happenings has been deprived in many aspects, but many other spells lying on OSM contributors need to be disenchanted to fully understand the trends and effects of the different data creation and editing schemes. More research on all OSM user groups from the data user, the registered user and the potential mapper all the way to the senior user, the heavy contributor or even the founder will strengthen this and other community projects for a future with more free and open information. The application of the developed framework to other community driven projects can yield similar and comparable insights and unlock the potential to learn from each other and enhance community driven projects as a whole.

Supplementary Materials: The related source code for the analyses can be obtained from the public git repository at https://github.com/SlowMo24/Effects-of-Happenings. The OpenStreetMap data extracted during the analyses potentially contains personal data. It is therefore available from the authors only. The presented study is partly based on a masters thesis by Moritz Schott titled "Under the Spell of Community Happenings" available from the author.

Author Contributions: Conceptualization, Moritz Schott and Asher Yair Grinberger; methodology, Moritz Schott and Asher Yair Grinberger; software, Moritz Schott; validation, Moritz Schott, Asher Yair Grinberger and Sven Lautenbach; formal analysis, Moritz Schott; investigation, Moritz Schott; resources, Alexander Zipf, Moritz Schott, Asher Yair Grinberger and Sven Lautenbach; data curation, Moritz Schott; writing—original draft preparation, Moritz Schott and Asher Yair Grinberger; writing—review and editing, Moritz Schott, Asher Yair Grinberger, Sven Lautenbach and Alexander Zipf; visualization, Moritz Schott and Sven Lautenbach; supervision, Asher Yair Grinberger, Sven Lautenbach, Alexander Zipf and Moritz Schott; project administration, Moritz Schott; funding acquisition, Alexander Zipf. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Klaus Tschira Stiftung.

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to the data being sourced from an open platform. Any mapper contributing to the dataset accepts its licence that clearly states that any use of the data is allowed.

Informed Consent Statement: Mapper con-sent was waived due to the data being sourced from an open platform. Any mapper contributing to the dataset accepts its licence that clearly states that any use of the data is allowed. It is up to the single mapper and allowed to obstruct their personal identity. No personal information can be retrieved from this publication or the supplementary materials.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here:

- OpenStreetMap History Data Base converted extract of the OSM data base: http://downloads.ohsome.org/ (accessed on 2 September 2019).
- Note and changeset data taken from https://planet.openstreetmap.org/ accessed on 2 September 2019).
- Global Human Settlement Population Grid by [45]: https://data.jrc.ec.europa.eu/dataset/jrc-ghsl-ghs_pop_gpw4_globe_r2015a (accessed on 2 September 2019).
- The Natural Earth dataset [43]: https://www.naturalearthdata.com/ (accessed on 2 September 2019).
- The terrestrial ecoregions of the world data set [46]: https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world (accessed on 2 September 2019).

Acknowledgments: We thank Martin Raifer and Rafael Troilo from HeiGIT and the HeiGIT as an institution. Without their technical support and possibilities, only a very small part of the analyses in this paper would have been possible.

Conflicts of Interest: Moritz Schott, Asher Yair Grinberger and Sven Lautenbach are active members of the OpenStreetMap community and regularly engage in various community-related activities. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CFM   Community Field Mapping.
CG   Control Group.
CRM   Community Remote Mapping.
GPS   Global Positioning System.
HOT   Humanitarian OpenStreetMap Team.
JOSM   Java OpenStreetMap Editor.
OSM   OpenStreetMap.
VGI   Volunteered Geographic Information.

References

1.   Ghosh, R.A.; Aigrain, P.; Andradas, R.; Badin, R.; Bernard, R.; Díaz, L.C.; David, P.; Dueña, S.; Dunnewijk, T.; Glott, R.; et al. *Economic Impact of Open Source Software on Innovation and the Competitiveness of the Information and Communication Technologies (ICT) Sector in the EU*; Resreport; European Commission: Brussels, Belgium, 2006.
2.   Feick, R.; Roche, S. Understanding the Value of VGI. In *Crowdsourcing Geographic Knowledge*; Sui, D., Elwood, S., Goodchild, M., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 15–29.
3.   Eckle, M.; de Albuquerque, J.P. Quality Assessment of Remote Mapping in OpenStreetMap for Disaster Management Purposes. In Proceedings of the ISCRAM 2015 Conference, Krystiansand, Norway, 24–27 May 2015.
4.   Sieber, R.E.; Haklay, M. The epistemology(s) of volunteered geographic information: A critique. *Geo Geogr. Environ.* **2015**, *2*, 122–136. [CrossRef]
5.   Bryant, S.L.; Forte, A.; Bruckman, A. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work, Sanibel Island, FL, USA, 6–9 November 2005; pp. 1–10. [CrossRef]
6.   OpenStreetMap Wiki Contributors. About OpenStreetMap. Available online: https://wiki.openstreetmap.org/wiki/About_OpenStreetMap (accessed on 2 September 2019).
7.   Neis, P.; Zipf, A. Analyzing the Contributor Activity of a Volunteered Geographic Information Project—The Case of OpenStreetMap. *ISPRS Int. J. Geo-Inf.* **2012**, *1*, 146–165. [CrossRef]
8.   Neis, P.; Zielstra, D.; Zipf, A. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet* **2013**, *5*, 282–300. [CrossRef]
9.   Coleman, D.J.; Georgiadou, P.Y.; Labonte, J. Volunteered geographic information: The nature and motivation of produsers. *Int. J. Spat. Data Infrastruct. Res.* **2009**, *4*, 332–358. [CrossRef]
10.   Steinmann, R.; Gröchenig, S.; Rehrl, K.; Brunauer, R. Contribution Profiles of Voluntary Mappers in OpenStreetMap. In Proceedings of the International Workshop on Action and Interaction in Volunteered Geographic Information (ACTIVITY) at the 16th AGILE Conference on Geographic Information Science, Leuven, Belgium, 14–17 May 2013.
11.   Quinn, S. Using small cities to understand the crowd behind OpenStreetMap. *GeoJournal* **2017**, *82*, 455–473. [CrossRef]
12.   Bégin, D.; Devillers, R.; Roche, S. The life cycle of contributors in collaborative online communities—The case of OpenStreetMap. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 1611–1630. [CrossRef]
13.   Thebault-Spieker, J.; Hecht, B.; Terveen, L. Geographic Biases Are 'Born, Not Made': Exploring Contributors' Spatiotemporal Behavior in OpenStreetMap. In Proceedings of the 2018 ACM Conference on Supporting Groupwork, Sanibel Island, FL, USA, 7–10 January 2018; ACM: New York, NY, USA, 2018; pp. 71–82. [CrossRef]
14.   Lin, Y.W. A qualitative enquiry into OpenStreetMap making. *New Rev. Hypermedia Multimed.* **2011**, *17*, 53–71. [CrossRef]
15.   Palen, L.; Soden, R.; Anderson, T.J.; Barrenechea, M. Success & Scale in a Data-Producing Organization: The Socio-Technical Evolution of OpenStreetMap in Response to Humanitarian Events. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, Seoul, Korea, 18–23 April 2015; ACM: New York, NY, USA, 2015; pp. 4113–4122. [CrossRef]
16.   OpenStreetMap Wiki Contributors. Map Limehouse Event 2005. Available online: https://wiki.openstreetmap.org/wiki/Map_Limehouse_Event_2005 (accessed on 2 September 2019).
17.   OpenStreetMap Wiki Contributors. Current Events. Available online: https://wiki.openstreetmap.org/wiki/Current_events (accessed on 6 September 2019).
18.   Herfort, B.; Lautenbach, S.; de Albuquerque, J.P.; Anderson, J.; Zipf, A. The evolution of humanitarian mapping within the OpenStreetMap community. *Sci. Rep.* **2021**, *11*. [CrossRef]
19.   Coast, S. The GISPro interview with OSM founder Steve Coast. *GIS Prof.* **2007**, *18*, 20–23.
20.   Perkins, C.; Dodge, M. The potential of user-generated cartography: A case study of the OpenStreetMap project and Mapchester mapping party. *North West Geogr.* **2008**, *8*, 19–32.
21.   Mooney, P.; Minghini, M.; Stanley-Jones, F. Observations on an OpenStreetMap mapping party organised as a social event during an open source GIS conference. *Int. J. Spat. Data Infrastruct. Res.* **2015**, *10*, 138–150. [CrossRef]

22. Hristova, D.; Quattrone, G.; Mashhadi, A.; Capra, L. The Life of the Party: Impact of Social Mapping in OpenStreetMap. In Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, Boston, MA, USA, 8–11 July 2013; pp. 234–243.

23. Dittus, M. Analysing Volunteer Engagement in Humanitarian Crowdmapping. Ph.D. Thesis, University College London, London, UK, 2017.

24. Coetzee, S.; Minghini, M.; Solís, P.; Rautenbach, V.; Green, C. *Towards Understanding the Impact of Mapathons-Reflecting on YouthMappers Experiences*; The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences: Hannover, Germany, 2018; pp. 35–42.

25. Ebrahim, M.; Minghini, M.; Molinari, M.E.; Torrebruno, A. Minimapathon—Mapping the World at 10 Years Old. In Proceedings of the EDULEARN16 Proceedings, IATED, Barcelona, Spain, 4–6 July 2016; pp. 4200–4208. [CrossRef]

26. OpenStreetMap Wiki Contributors. Elements. Available online: https://wiki.openstreetmap.org/wiki/Elements (accessed on 2 September 2019).

27. OpenStreetMap Wiki Contributors. Changeset. Available online: https://wiki.openstreetmap.org/wiki/Changeset (accessed on 6 September 2019).

28. OpenStreetMap Wiki Contributors. Notes. Available online: https://wiki.openstreetmap.org/wiki/Notes (accessed on 11 September 2019).

29. Seto, T.; Kanasugi, H.; Nishimura, Y. Quality Verification of Volunteered Geographic Information Using OSM Notes Data in a Global Context. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 372. [CrossRef]

30. OpenStreetMap Wiki Contributors. Category:Past Events. Available online: https://wiki.openstreetmap.org/wiki/Category:Past_events (accessed on 2 September 2019).

31. OpenStreetMap Wiki Contributors. Import/Guidelines. Available online: https://wiki.openstreetmap.org/wiki/Import/Guidelines (accessed on 2 September 2019).

32. Hacar, M.; Kilic, B.; Sahbaz, K. Analyzing OpenStreetMap Road Data and Characterizing the Behavior of Contributors in Ankara, Turkey. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 400. [CrossRef]

33. Mashhadi, A.; Quattrone, G.; Capra, L. The Impact of Society on Volunteered Geographic Information: The Case of OpenStreetMap. In *OpenStreetMap in GIScience*; Jokar Arsanjani, J., Zipf, A., Mooney, P., Helbich, M., Eds.; Springer: Cham, Switzerland, 2015; pp. 125–141. [CrossRef]

34. Mooney, P.; Corcoran, P. Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Trans. GIS* **2014**, *18*, 633–659. [CrossRef]

35. Budhathoki, N.R.; Haythornthwaite, C. Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *Am. Behav. Sci.* **2013**, *57*, 548–575. [CrossRef]

36. Gardner, Z.; Mooney, P.; De Sabbata, S.; Dowthwaite, L. Quantifying gendered participation in OpenStreetMap: Responding to theories of female (under) representation in crowdsourced mapping. *GeoJournal* **2019**. [CrossRef]

37. Spellerberg, I.F.; Fedor, P.J. A Tribute to Claude Shannon (1916–2001) and a Plea for More Rigorous Use of Species Richness, Species Diversity and the 'Shannon-Wiener' Index. *Glob. Ecol. Biogeogr.* **2003**, *12*, 177–179. [CrossRef]

38. Scholz, I.; Stäcker, D.; other contributors. Validator Window. Available online: https://josm.openstreetmap.de/wiki/Help/Dialog/Validator (accessed on 18 October 2019).

39. Danziger, R. Mannheimer Mapathons. Available online: https://mamapa.org/ (accessed on 18 October 2019).

40. OpenStreetMap Wiki Contributors. Map Features. Available online: https://wiki.openstreetmap.org/wiki/Map_Features (accessed on 9 October 2019).

41. OpenStreetMap Wiki Contributors. Planet.osm/full. Available online: https://wiki.openstreetmap.org/wiki/Planet.osm/full (accessed on 9 October 2019).

42. Raifer, M.; Troilo, R.; Kowatsch, F.; Auer, M.; Loos, L.; Marx, S.; Przybill, K.; Fendrich, S.; Mocnik, F.B.; Zipf, A. OSHDB: A framework for spatio-temporal analysis of OpenStreetMap history data. *Open Geospat. Data Softw. Stand.* **2019**, *4*, 3. [CrossRef]

43. Natual Earth Contributors. Natural Earth. Available online: https://www.naturalearthdata.com/ (accessed on 9 October 2019).

44. Huntington, S.P. *The Clash of Civilizations and the Remaking of World Order*; Simon & Schuster: New York, NY, USA, 1996.

45. European Commission; Columbia University. GHS Population Grid, Derived from GPW4, Multitemporal (1975, 1990, 2000, 2015). Available online: http://data.europa.eu/89h/jrc-ghsl-ghs_pop_gpw4_globe_r2015a (accessed on 25 November 2019).

46. Olson, D.M.; Dinerstein, E.; Wikramanayake, E.D.; Burgess, N.D.; Powell, G.V.N.; Underwood, E.C.; D'amico, J.A.; Itoua, I.; Strand, H.E.; Morrison, J.C.; et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* **2001**, *51*, 933–938. [CrossRef]

47. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* **1995**, *57*, 289–300. [CrossRef]