The Life of the Party: Impact of Social Mapping in OpenStreetMap

Desislava Hristova[1], Giovanni Quattrone[1], Afra Mashhadi[2] and Licia Capra[1]
[1] Department of Computer Science, University College London, UK
[2] Bell Labs, Alcatel-Lucent, Dublin, Ireland

Abstract

OpenStreetMap has grown rapidly into an exemplary geowiki, where contributors collectively build an open map of the world. Official 'mapping party' events are organized on a regular basis to invite users to socialize, map and to engage new-comers locally. Here, we measure the direct and indirect impact of mapping parties on user contributions in both the short and long terms. We question whether this social mapping is a cause for users to become highly committed through a social bond, or is an effect of the mappers' need to find a common social ground. We show that mapping parties have distinct effects on different types of users, with a more profound direct impact on weaker contributors and a longer term effect on heavy contributors.

Introduction

Wikis, most famously exemplified by Wikipedia, have proven successful in harvesting user-generated content into practical collective knowledge. *Geowikis* are a sub-category of these, engaging contributors in collectively gathering geographic knowledge, in the form of free and editable maps of the world. OpenStreetMap (OSM) has been described as *the Wikipedia of maps* and is one of the most successful examples of this model (Coast 2011). This and other open mapping projects such as Wikimapia,[1] Cyclopath[2] and mashups such as FixMyStreet[3] are all reliant on volunteered contributions of local spatial knowledge.

What motivates people to contribute to these user-generated content platforms? *Social contact* has been identified as a powerful motivator by many successful online communities (Kraut and Resnick 2012). Hackathons, mapathons and other social events are often organized, in order to bring together people with similar technical skills and interests to accomplish collaborative projects. One example is the GNOME open-source software project,[4] which brings its developers together several times a year to *meet, plan, party*. Socializing during collaborative tasks has been associated with a greater desire to contribute more (Kraut and Resnick 2012). In this paper, we explore whether this holds true for organized local events in OSM.

Contributors to OSM worldwide are invited to attend the annual 'State of the Map' conference; furthermore, local social events, dubbed *mapping parties*, are organized throughout the year, so that contributors can "get together to do some mapping, socialize, and chat about making a free map of the world!".[5] These mapping events have the double goal of improving map coverage in certain areas, and engaging new-comers on a local level. Historically important in the UK since their initiation in 2005, they continue to happen on a fortnightly basis in urban areas like London (Haklay and Weber 2008). Steve Coast, founder of OSM, emphasizes the social side's importance: "A big aspect of getting OSM off the ground was the mapping parties: getting drunk and arguing with people" (GISPro 2007).

In this work, we quantify the success of mapping parties in eliciting contributions and sustaining participation. In user-generated content communities, contributions notoriously follow a long-tail distribution. The success of a platform is thus reliant on the committed adoption of a small group of users: in the Cyclopath geowiki, 5% of its users are responsible for the majority of its content (Panciera et al. 2010), while in Wikipedia the highly committed contributors are only about 2.5% (Panciera, Halfaker, and Terveen 2009). OSM is no exception, with 95% of contributions made in the area of London, UK attributed to less than 10% of users, as we shall illustrate later.

Interestingly, we find that this small group of users actively producing content in OSM is mostly made up of people who attend mapping parties; as a consequence, one may wonder how big the role of these social events is in the continued success of OSM. Here, we propose a methodology, grounded on economic theory, to *quantify* the impact of social collaborative events on user-generated content platforms. We measure impact across different user groups, both in the short and long term. We apply our methodology to analyze the role of OSM mapping parties in the area of Greater London, UK. We conclude with a discussion of the implications of our findings on the future of OSM and mapping parties.

[1] http://wikimapia.org/
[2] http://cyclopath.org/
[3] http://www.fixmystreet.com/
[4] http://www.gnome.org/
[5] http://wiki.openstreetmap.org/wiki/ Mapping_parties/

Related Work

A consistent finding in the study of user-generated content platforms has been that a small fraction of users is responsible for the vast majority of contributions. These committed users are, for example, those who do most of the coding (Mockus, Fielding, and Herbsleb 2002) and, in online encyclopedias such as Wikipedia, they are those who take on a more administrative role (Bryant, Forte, and Bruckman 2005). A substantial body of research has thus studied motivation and commitment of participants in user-generated content systems.

Wikipedia is the most studied wiki to date (Lampe et al. 2010). In (Nov 2007), the authors highlight that fun and ideology are the two most important factors that are linked strongly to editor's motivation in Wikipedia. They further suggest that other user-generated systems that seek to recruit and retain volunteering contributors should likewise focus their marketing, recruitment, and retention efforts on the fun aspects of contributing. In (Bryant, Forte, and Bruckman 2005), the authors show that, as participation becomes more frequent, Wikipedians adopt new goals as well as new roles, caring less about editing individual articles and more about maintaining high the quality of Wikipedia content as a whole. Their role thus changes from contributor to administrator, for example in the form of 'watchdog', monitoring community activities and looking for opportunities to help and correct mistakes. 'WikiProjects',[6] are additionally organized, directing work of the open Wikipedia community towards editing specific topics that are considered high priority. This form of 'directed crowd-sourcing' has been shown to be an effective means of eliciting active participation online (Cosley et al. 2006; Harper et al. 2007; Beenen et al. 2004).

Online open-source software communities have also been subject of investigation. In (Ducheneaut 2005), the authors found that, despite the openness of the platform, with developers being able to take part in any project they wished, in practice a very small group of developers actually contributed. This was not just a matter of possessing the right technical knowledge; socialization also played a crucial role, with the integration of new-comers being a politically driven process with rites of passage, making it difficult to become a member of the 'tribe'.

OSM contributors' motivation has been studied in (Budhathoki and Haythornthwaite 2012), with the highest motivation factor being ideological (i.e., the willingness to contribute to the community that is behind the success of the OSM project). The authors also conclude that geography itself is a community-wide motivator, with both casual and serious mappers being motivated by an overall concern with contributing to local geography, so to build a local complete and accurate map of their area. Similar motivation has been found within other geowikis such as Cyclopath, where contributors are first attracted mainly to fix a specific problem, but then continue to contribute to aid the cycling community as a whole (Panciera, Masli, and Terveen 2011; Priedhorsky, Masli, and Terveen 2010).

[6] http://en.wikipedia.org/wiki/Wikipedia: WikiProject

This 'sense of community' is a recurring theme in the qualitative studies on motivation in (geo)wikis conducted to date. However, there has been no attempt to *quantify* the impact that community-boosting social activities, such as mapping parties, really have in terms of contributions. One exception is the work presented in (Perkins and Dodge 2008), where the authors studied the success of one specific weekend mapping party they organized in Manchester, UK. Interestingly, they found that the event was very successful in terms of number of attendees, but rather unsuccessful in improving coverage of the Manchester map itself. These findings however, cannot be generalized to all mapping parties, as only one event was studied. Furthermore, 'success' was measured in terms of the contributions that were made during the party; however, studies conducted in other online social communities (e.g., LinkedIn) show that activity (i.e., new connections) rises significantly in a short period (within 10 days) *after* these events have taken place (Gomez Rodriguez and Rogati 2012), thus calling for an investigation of their impact in the immediate, short as well as long term. In the following sections, we present a methodology to measure the impact of mapping parties on different user groups, over different time spans. The methodology, though applied in the specific context of OSM mapping parties, could be used to measure the impact of similar social events in wiki-style platforms.

Hypotheses & Metrics

The purpose of mapping parties, as defined by the OSM wiki pages, is to *map*, *socialize* and *engage newcomers*. We aim to quantify their success in light of these goals, and therefore put forward the following three hypotheses:

Hypothesis 1 *Mapping parties cause users to map more than usual during the collaborative event.* □

Hypothesis 2 *Mapping parties cause users to map more than usual as a result of the collaborative event both in the short and long term.* □

Hypothesis 3 *Mapping parties foster commitment and therefore retain participants over time.* □

To test the first two hypotheses, we quantify the direct (immediate) and indirect (subsequent) impact of a mapping party using the *Abnormal Returns* (AR) model. The model is used in economics to measure the impact of a specific event on the value of a firm, by observing its pre and post market price, within a given time window (MacKinlay 1997). ARs are triggered by events: the higher the abnormal return, the higher the impact of the event on the value of the firm.

In our analysis, we adopted the AR model as follows: whenever a mapping party took place, for each mapping party participant i and time period τ after the party, we measured the *actual* returns R_i^τ as the average number of contributions per unit of time Δt made by user i during period τ. We also computed the *expected* returns E_i^δ as the average number of contributions made by the same user i per unit of time Δt during a period δ prior to the event. We then calculated the abnormal returns $AR_i^{\delta\tau}$ per unit of time Δt of each user i as:

$$AR_i^{\delta\tau} = R_i^\tau - E_i^\delta. \quad (1)$$

The higher the AR, the higher the actual returns are compared to the expected, therefore the higher the impact of the mapping party on the users' contributions. As an example, if we choose to compute our expected returns based on $\delta = 6$ months prior to the event, and $\tau = 1$ month after the event, with unit of time $\Delta t = 1$ week, having AR equal to 100 means that the user, in the month following the event, is performing on average 100 edits more per week than in the six months preceding the event.

To test the third hypothesis, we compute the *Retention Ratio* metric on a per mapping party basis: that is, the number of attendees n_j^τ at the event j that also attend another mapping party within the time window τ following the event, *relative to* the number n_j of those who attended the event j under exam:

$$RR_i^\tau = \frac{n_j^\tau}{n_j}. \quad (2)$$

The higher the retention ratio, the more successful mapping parties are at committing members to attend again.

Dataset

OpenStreetMap. We verify the above hypotheses, using measures of abnormal returns and retention rate specifically on the OpenStreetMap dataset. OSM is a popular geowiki, where registered users can contribute spatial content to the global OSM database, thus collectively building a free, openly accessible, editable map of the world. Spatial objects can be one of three types: nodes, ways, and relations. Nodes broadly refer to points-of-interest, ways are representative of roads, and relations are used for grouping other objects together. To reduce the dataset to a manageable size, we chose to restrict our study to the area of Greater London, UK: London is the birthplace of OSM and mapping parties; it is exemplary of organic crowdsourced geographic data, and is prosperous in terms of edits and events (Haklay and Weber 2008). The dataset we use spans from the beginning of 2007 (when the first large-scale mapping party took place in London), up until June 2011; during this period, there have been 2,736 users making 2,459,705 edits overall in the London area encircled by the M25 highway.

Mapping Parties. In order to conduct impact analysis of mapping parties on users' contributions, we need to derive two further pieces of information: when/where mapping parties took place, and who took part in each of these. Official mapping events on OSM are recorded on the OSM wiki pages, which are edited by organizers and attendees. We therefore manually constructed a dataset of all events that took place in London by extracting location, time and date information from the wiki. We then assigned geographic coordinates to each mapping party, according to the meeting point for the event (typically a pub or a subway station). We recorded 94 mapping parties for the period under examination.

Inferring Participation. We do not have ground truth about who took part in what event. A list of users who 'intend to attend' an event is often included in the wiki pages; [7] however, wishing to attend does not mean they will do so after all. Vice versa, users may take part without pre-registering their intention. We thus devised an inference method to determine participation, based on users' editing activity 'near' the mapping event, both from a *spatial* and a *temporal* perspective.

- *Spatial processing.* Mapping parties start at a precise location (e.g., a pub or station, which serves as a meeting point). During the party, an area near the meeting point is explored usually on foot and mapped. Such areas are divided into 'cake diagrams', allowing efforts to be split among participants. We do not have boundary information about these areas; however, we know the meeting point of each event. We thus compute the *ward* (i.e., the UK primary administrative and electoral geographic unit) within which the meeting point falls, and consider as *party area* the one covered by this ward plus all adjacent ones. We chose to use wards as the spatial unit of analysis, as they are defined not only by population density but also, and more importantly for this study, by geographic morphology, with physical dividers such as highways, rivers and parks having been taken into account. Intuitively, as mapping party goers explore an area on foot, they restrain themselves to contained areas.
- *Temporal processing.* OSM mapping is a three step process, of which only the first step is data collection in the field – using GPS receivers to record traces and survey the area.[8] The process then requires data storage and rendering, typically done using specialized desktop software and hardware. The last step is data upload, when the edits are uploaded to the map server. These three steps do not need to be conducted at the same time. Indeed, even though mapping parties are designed to encourage participants to upload their edits *during* the event, with laptops and Internet connection secured during the party in recent years, data rendering and upload may take place shortly after the event itself. We chose to define *party time* as the temporal window that goes from the time of the party up until the midnight of the day after.

We infer participation to a mapping party by observing if users edited within the *party area* during the *party time*. Figure 1 illustrates the heatmap of edits made around London over the 48 hours during and after the Isle of Dogs mapping party on June 18th, 2008. The map on the left-hand side shows that the activity for the whole of London during this time frame is indeed concentrated in the area of the mapping party. Zoom-in on the right hand shows the individual edits done in the central (dark highlighted) and adjacent (lighter highlighted) wards of the event.

[7] http://wiki.openstreetmap.org/wiki/London/Summer_2008_Mapping_Party_Marathon/2008-05-21

[8] *Armchair mapping* is an alternative way of creating content, which does not require the user to physically collect the data and is representative of methods such as aerial imagery tracing.

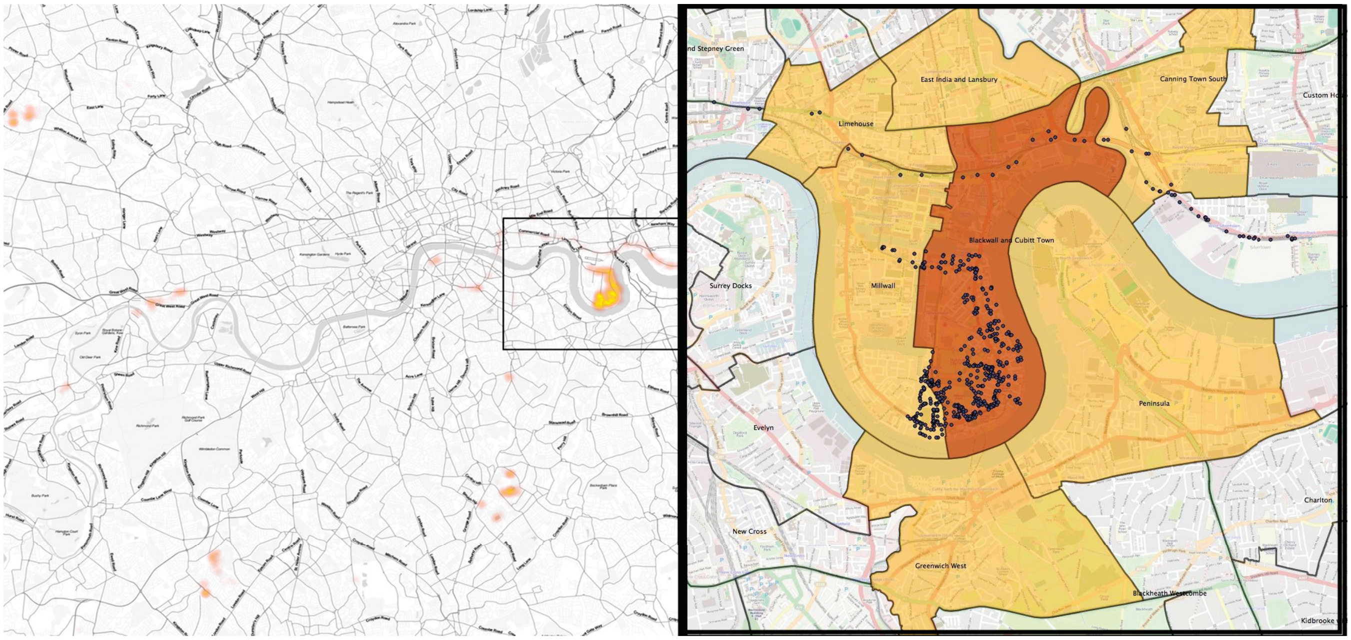


Figure 1: Map of edits made around London over 48 hours during and after the Isle of Dogs mapping party

To gain confidence in the spatio-temporal technique we use to infer participation, we further inspected a random sample of 10 mapping parties, in a fashion similar to Figure 1. In 8 cases, the spatial filter we use (i.e., meeting point ward plus adjacent ones) appears to cover more than 90% of the edits that could be attributed to the event; in 2 cases, we captured around half of the edits (with the remaining edits happening in further away wards). We tried a spatial processing of two-hops of adjacent wards from the meeting point, but this caused a large number of clear false positives (i.e., edits done away from the heart of the party) to be mistakenly considered as part of the event. We thus set on the spatial processing presented above.

As for the validity of the temporal processing, we measured the volume of daily edits in the party areas for all mapping parties in the week following the event, and compared it to the average daily activity for the week before. We found that in 40% of mapping parties the peak of activity was on the day of the event, while in 89% of cases the peak activity was within 30 hours after the party. In 99% of cases, the peak activity was within 48 hours, after which the daily edits stabilize to the norm previously observed. We thus set on the temporal threshold described above, that covers from the party date up until midnight of the day after (thus, on average, 36 hours).

Using this method, we discovered 150 ‘social mappers’ in the Greater London area. These are users who attended at least one of the 94 mapping parties we recorded. We use the word ‘attendance’ in the rest of the paper as the result of our inference method. Figure 2 shows the number of events that took place on a quarterly basis, and how many users took part in each. As shown, after the first mapping event that took place in the first quarter of 2007, a year passed

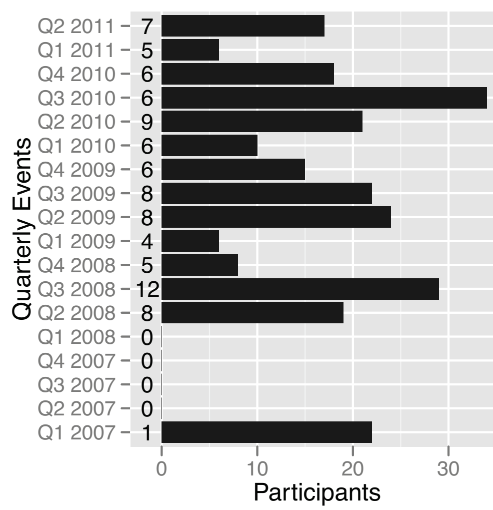


Figure 2: Number of events and user attendance per quarter

before mapping parties became ‘regular’ events, with between 4 and 12 events being organized each quarter. Mapping party attendance appears to be seasonal, with decreased attendance in the first and last quarters of each year (organizationally, London winter events are typically dubbed ‘pub meet-ups’, while summer events are more often referred to as ‘mapping parties’, reflecting the relaxed winter activity).

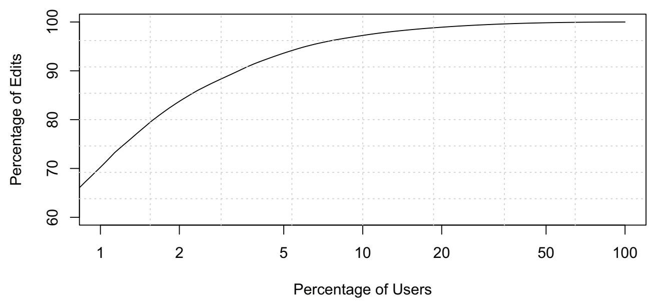


Figure 3: Participation inequality in OSM in London

	#users	*#edits*	*total*
social	150	1,723,168	70%
non-social	2,586	736,537	30%

Table 1: Proportion of social vs. non-social contributions

Results

Preliminary Analysis

Before testing the validity of each hypothesis on the dataset above, we report on the results of a preliminary analysis we conducted in order to gain deeper insights into users' contributing behavior in OSM, especially comparing social mappers versus non-social ones. We begin with a bird's eye view, and compute the overall user participation in OSM. As Figure 3 illustrates, about 2% of London contributors produce more than 80% of content, and 10% of users generate near 95% of it. We then look at the aggregate number of contributions made by the 150 users that have attended at least one mapping party during their lifetime. We find that the identified 150 social mappers (that is, 5% of the 2,736 OSM users present in our dataset) are responsible for 70% of OSM content, as further detailed in Table 1. We also measured repeated users' participation in mapping parties, and found that this also follows a skewed distribution (albeit more balanced than the general participation shown in Figure 3); more precisely, the majority of social mappers (69%) participate in one event only, while the remaining 31% take part repeatedly. The latter are responsible for 55% of the total content in our OSM dataset, suggesting a relationship between the frequency of attendance and the amount of contributions. Interestingly, only 1.8% of the 1.7M edits attributed to social mappers are actually placed during an event, suggesting that mapping parties might have a more profound indirect (follow up) effect than the immediately quantifiable contributions. We will explore the validity of these statements when verifying our hypotheses. Before doing so, we look into the activity of the 150 social mappers in more details, to have a better understanding of their behavior, and in particular whether (and how) they differ from the other OSM contributors. We do this in terms of: the amount of edits they do (Figure 4), the date when they join the OSM community (Figure 5), and their longevity, measured as time passed between their first and last contribution (Figure 6).

The first observation to make is that social mappers are generally *heavy contributors*. Figure 4 shows the frequency distribution of the total number of edits (in log scale) done by social users versus non-social ones. Let us look at non-social users first: almost all of them make less than 10 edits in their lifetime, a small proportion edits more than 100 times, and very few edit more than 1,000 times. Notably, social mappers do not follow the left-skewed distribution of the broader OSM population; their edit distribution is slightly right skewed instead.

Second, many social mappers are *early adopters*. Figure 5 shows the frequency distribution of users' joining date. The majority of social mappers joined before 2008 (that is, even before mapping parties became regular events). This is in stark contrast to the non-social contributors, whose first edits started to be recorded after 2008 mainly.

Finally, social mappers are *long-lived*. As Figure 6 illustrates, many users have recorded activity in OSM for just one day, both among social and non-social mappers. As we look further to the right (note the log scale x axis), the proportion of very long-lived users (more than 10^4 days) is much higher among social mappers than among non-social ones. Social mappers again do not follow the dominant (non-social) OSM population, their distribution being right-skewed, indicating their longevity in the system.

Social mappers are therefore largely composed of heavy editors, who have joined the community in its early days (early adopters), and whose engagement spans a long time period (long-lived). These users are key to the success and long-term sustainability of OSM. However, the analysis conducted so far does not answer the question of whether mapping parties are a *cause* of greater engagement with OSM or an *effect* of the need of a social ground for heavy contributors. To what extent do mapping parties stimulate contributions? Do mapping parties integrate new-comers and foster a committed contributor community? We answer these questions next.

Hypothesis 1: Direct Impact of Mapping Parties

The first hypothesis we test is that users contribute more during mapping parties than outside these events. For each mapping party, and for each user who took part in it, we compute the abnormal returns as per Formula 1, with Δt equal to one day. We further selected δ equal to six months prior to each party, so to have enough history about users' editing behavior, and τ equal to the 'party time' (from the day of the party up until the midnight of the day after).

As our preliminary analysis has illustrated, OSM users greatly differ in terms of the amount of contributions they make, and over what timespan. In order to quantify the impact of mapping parties on different types of users, we have grouped them based on the number of contributions they made in the six months prior to each party. We do so on a log scale of 10 as in the above pre-analysis, and split users in five distinct groups - *Group 0* (just 1 edit); *Group 1* (from 1 up to 10 edits); *Group 2* (from 10 up to 10^2 edits); *Group 3* (from 10^2 up to 10^3 edits); *Group 4* (from 10^3 up to 10^4 edits). An additional group of newly joined users (*Group NA*) is considered, consisting of those who make their first edit in the system either during the mapping party or in less than six months preceding it (thus not having sufficient editing

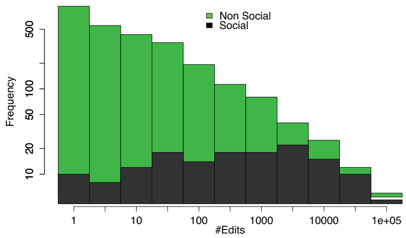


Figure 4: Distribution of edits for social and non-social users

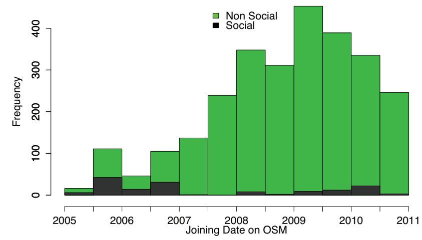


Figure 5: OSM joining date distribution of social and non-social users

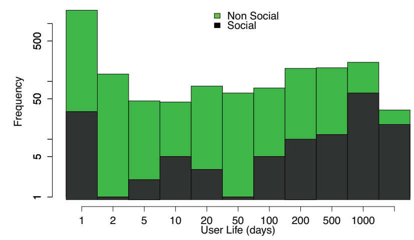


Figure 6: Distribution of user life for social and non-social users

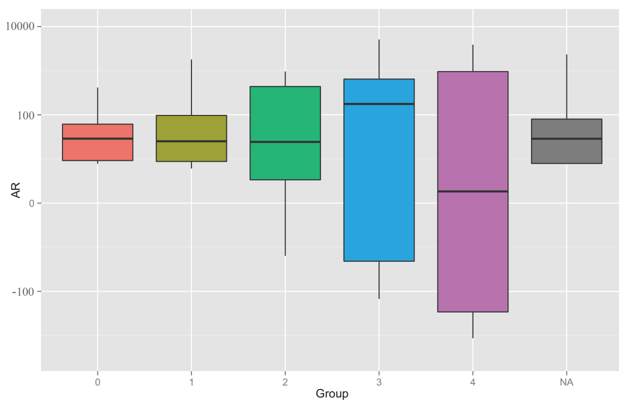


Figure 7: Box-and-whisker plot of Abnormal Returns during a party

history to be confidently placed in the above groups). The results for this group assess the impact of mapping parties on new-comers. Note that the same user may be placed in different groups when analyzing different parties, thus reflecting users' varying behavior over time. In order to consider genuine results and not be influenced by attendance to prior events, we excluded from the impact analysis of a party those users who took part in another mapping event during the $\delta = 6$ months preceding it.

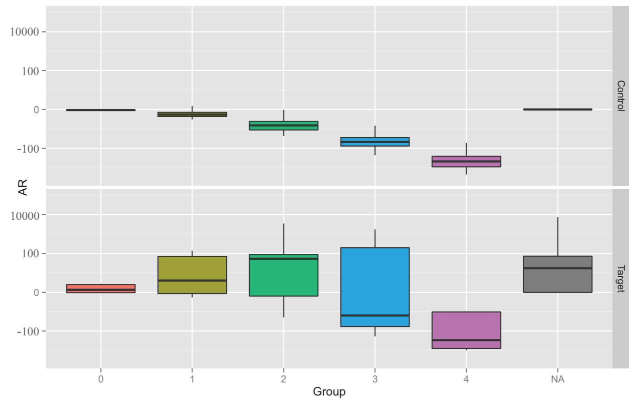
Figure 7 shows the average results across the 94 mapping parties that took place in London in the period under consideration, for each of these user groups. We use a box-and-whisker plot to display the spread of the AR results, with the thick black line within each box representing the median value and the 'whiskers' of the box representing the top and bottom quartile values. Median y axis values above zero indicate that most users within that group exhibit higher number of edits during the party time than before it, and vice versa (negative y values indicate reduced activity during the mapping as compared to the norm).

The results show that for Groups 0-2 (light to medium contributors) and Group NA (new-comers), mapping parties have a strong positive impact in terms of contributions, with their edits being significantly more than usual. Despite more variation within it (and some negative returns too), Group 3 experienced the overall highest AR, with more than 50% of its members (median value and above) contributing at least 100 edits *more than expected* in the observation period (i.e., party time). Perhaps surprisingly at first glance, only half of the heaviest editors (Group 4) contribute more than expected; the other half in fact perform much below par. We cross checked the names of some of these contributors against what is publicly available in OSM wikis, and found that many of these users take on organizational roles, visiting an area prior to the party, creating 'cake diagrams', and identifying 'problems' they wish the party to fix. We thus speculate that their reduced contribution during the event itself might be due to their engagement in organizational rather than editing activities (e.g., acting as demonstrators for less expert users).

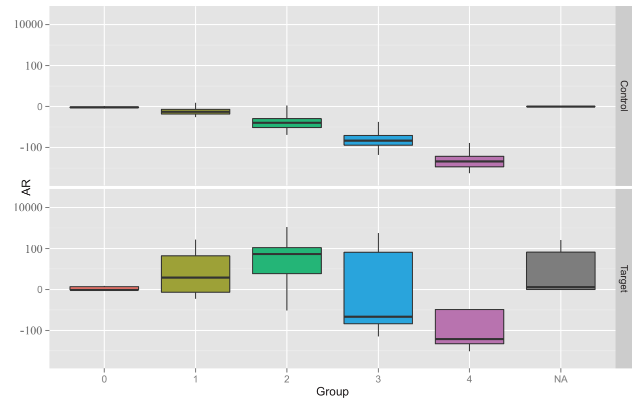
Hypothesis 2: Indirect Impact of Mapping Parties

The second hypothesis aims to quantify the impact that mapping parties have on users' contributions *after* they took part in an event. As before, we do so by computing the AR for the 6 user categories (from light to heavy editors – Groups 0-4, and new-comers – Group NA). To distinguish between the impact caused by attending a party from the impact potentially caused by external events (e.g., weather, OSM advertising), we constructed control groups for each of the 6 study groups. Each respective control group includes users who (*i*) have had a similar number of contributions as users in the corresponding study group in the $\delta = 6$ months prior to the party under examination and (*ii*) who did not take part in it or any other event in that time period. We then computed the AR for each control group too.

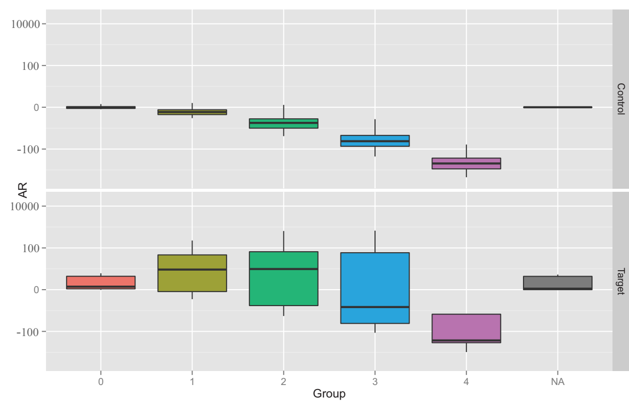
To quantify both short and long term effects of mapping party attendance, we computed AR on four non-overlapping observation windows τ: *(i)* up to one week following the event, *(ii)* between one week and one month following the event, *(iii)* between one and three months following the event, and *(iv)* between three and six months following the event. All observations exclude the contributions made *during* the event. For an easy comparison across all plots, we chose Δt equal to one week as the unit of time to compute AR across all cases. Results for each observation window are shown in Figures 8a to 8d. Once again, we use box-and-whisker plots, with boxes in the upper part of the plot illus-



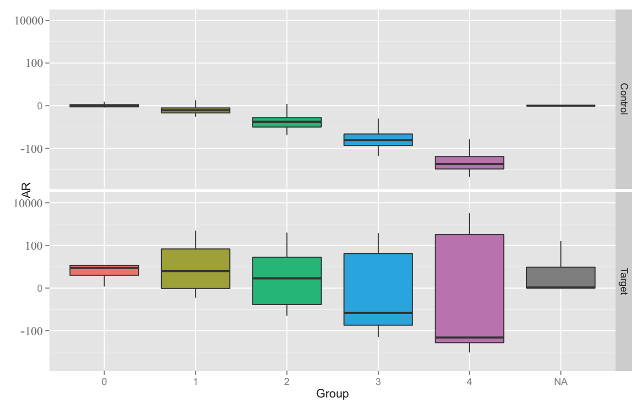
(a) 1 week following a mapping party



(b) 1 week to 1 month following a mapping party



(c) 1 to 3 months following a mapping party



(d) 3 to 6 months following a mapping party

Figure 8: Box-and-whisker plots of abnormal returns

trating the behavior of the control groups, and the bottom part displaying the behavior of the study groups (referred to as 'Target' group in plots).

First of all, we observe a decline in contributions (negative AR) by all *control* groups across all observation windows: users who do not take part in a mapping party tend to become more and more disengaged as time passes. This loss of engagement is more pronounced for users who were previously heavily contributing to OSM (Groups 3 and 4).

Let us now turn our attention to the study groups instead, beginning with light contributors (Groups 0 and 1). In the short and short-to-medium term (Figures 8a and 8b), these contributors have positive AR, with 25% of users in these groups showing AR values between 10 and 100 edits more per week. This increased engagement seems to be sustained over time (AR is still positive in both 1-to-3 and 3-to-6 months window – see Figures 8c and 8d).

Let us now turn our attention to medium contributors (Group 2). In the short and short-to-medium term (Figures 8a and 8b), these contributors have high positive AR, indicating a strong impact of mapping party attendance on their editing behavior. This increased engagement is not sustained over time though, and while AR is still positive in the 1-to-3 month window (Figure 8c), it becomes near zero in the longer term (3-to-6 months, Figure 8d). Note however, that users in the corresponding control group exhibit negative AR consistently. By comparison with the control group, the effect of mapping parties is more pronounced.

Let us now turn our attention to heavy contributors (Groups 3 and 4). Their behavior is somewhat surprising, as their median AR values are negative across all time periods. A possible explanation is that heavy contributors do not sustain this level of engagement continuously over time; rather, they do so for periods of time, then falling back to less active editing patterns. In this case, a more insightful understanding of the impact of mapping parties on these users can be gained by comparing their AR with regards to those in the corresponding control groups, rather than considering the positive or negative sign of AR values per se. In doing so, we observe a slightly increased engagement in the short term (Figure 8a) for Group 3, and comparable engagement for Group 4. However, as time progresses (Figure 8b), we observe 25% of Group 4 participants now exhibiting positive abnormal returns, whilst the AR of its control group remains consistently low. Finally, in the longer term (3-6 months, Figure 8d), Group 4 is indeed the only study group exhibiting significantly more engagement than what is observed in the corresponding control group, with 25% of its members showing AR values in between 0 and 100 edits per week, against AR close to negative -1000 edits per week for the control group.

Newcomers (Group NA) do not have a previous history of edits and are thus naturally observed to experience a positive impact, if any. A strong positive AR is indeed evident in the first week following participation in a mapping event (Figure 8a). However, as time progresses, the median AR value for this study group approaches zero, as observed in the corresponding control group. Indeed, after the first week following the event, 50% of new-comers stop contributing completely, with further complete disengagement as time passes.

To gain further confidence in these results, we measured the Pearson correlation between the AR metric and attendance to a mapping party, on a per group basis. Table 2 shows the results, with statistically significant correlations in bold. All correlations are indeed positive, even if only mildly so, confirming that mapping parties have both short and long term positive impact on user contributions, relative to control groups of similar characteristics.

Group	*1 week*	*1 week-1 month*	*1-3 months*	*3-6 months*
0	**0.10**	0.05	0.19	**0.12**
1	**0.16**	**0.14**	**0.15**	**0.14**
2	**0.22**	**0.20**	**0.11**	**0.08**
3	**0.15**	**0.11**	**0.11**	0.08
4	0.09	0.11	0.12	**0.22**

Table 2: Pearson correlation values between AR and mapping party attendance (in bold statistically significant results – p-value < 0.01).

Hypothesis 3: Retention of Participants

The last hypothesis we test aims to quantify the impact of mapping parties in retaining attendees, that is, in making them come back and take part in following events. The retention ratio is calculated using Formula 2 for each mapping party and considering observation windows τ of 0–1 month, 1–3 months, and 3–6 months after the event. In total, we computed the retention ratio for 82 out of 94 mapping parties (excluding the last 11, for which there does not exist sufficient post-event history, and the first one in 2007, which was not followed by a mapping party for a whole year). We present the total retention results across all 82 parties on a per user group basis (Groups 0-4 and Group NA). As not all groups are equally represented within each mapping party, we computed the weighted average, based on the number of participants from each group.

Figure 9 illustrates the weighted averages across groups

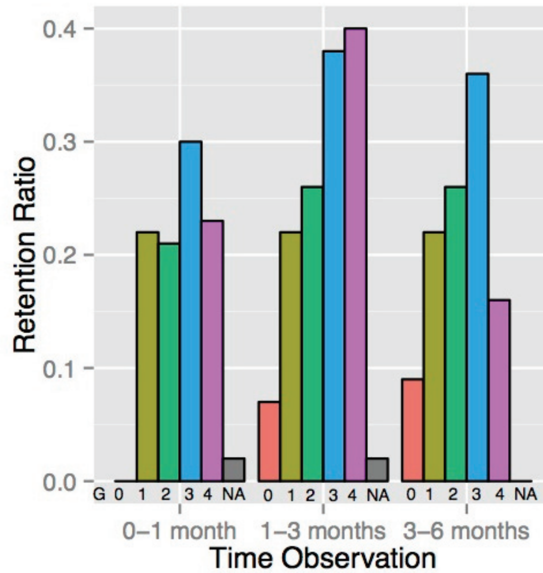


Figure 9: Weighted average retention ratio per user group over time

and time bins. As we can observe, mapping parties fail to retain new-comers (Group NA) almost completely, with no retention at all in the long term; similarly, less than 10% of very light contributors (Group 0) re-attend a mapping party within 6 months of having taken part in one. As we move our attention from lighter to more engaged contributors, we observe an increase in retention ratio: more than 20% of users in Groups 1-2 attend a mapping party again within 6 months, and this percentage doubles up for Groups 3 and 4, with near 40% retention in the 1-3 month period. In the 3-6 month period all groups remain consistent except the retention for the heaviest contributors - Group 4, which drops by half. We hypothesize that this unexpected result may be a consequence of the limitations to our inference method, discussed in the following section. Nonetheless, it appears that mapping parties, although unsuccessful in engaging new-comers, become increasingly more appealing to experienced users.

Discussion

Summary of Contributions. In this work we have quantified the impact that mapping parties have for the OSM community in Greater London. First, we have verified that mapping parties do cause participants to edit more than usual; we have quantified this effect across different user groups, and observed that the only group where this does not hold true is that of the heaviest editors (Group 4), who we have hypothesized are engaged in organizational activities during the party itself. This is in line with studies of Wikipedia (Bryant, Forte, and Bruckman 2005), where heavier contributors take on more administrative roles. Although this is not formally possible in OSM, there are self-identified members of the community who take on similar roles as geowikis are largely self-managed and mapping parties are lead by enthusiast mappers.

Second, we have measured the impact of attending mapping parties in terms of editing activities after attending an event, both in the short and long term. By comparing results against control groups, we have measured an overall positive impact, which in the short-term is stronger for light to medium contributors (Groups 0-2), and in the long term is more pronounced for high contributors (Group 4). These heavy contributors do not appear to be significantly impacted by the mapping parties - they are classified as heavy contributors before the mapping party and continue to edit heavily in the future long term. This suggests that OSMers, like Wikipedians, 'are born, not made' (Panciera, Halfaker, and Terveen 2009): their activity starts intensely, tails off a little but then remains strong consistently. These findings suggest that mapping parties *cause* light to medium contributors to edit more, both during an event and in the short and medium term afterwards; on the contrary, for the group of heaviest contributors (Group 4), mapping parties are an *effect* of their need of having a common social ground.

Finally, we have quantified the effectiveness of these social events in retaining attendees, and observed failure in doing so for new-comers but success in retaining the more experienced users instead. This may be linked to issues of socialization in collaborative projects, where the integration of new-comers is halted by 'tribe' membership behavior (Ducheneaut 2005). Similarly in OSM, we assume that it may be more difficult for less experienced mappers to integrate with the community socially, leading to lower re-attendance.

Limitations. The results presented in this paper cannot be verified by comparison to a ground truth dataset of who took part in which party, as such a dataset does not exist. The spatio-temporal inference technique we have adopted appears sufficiently robust, based on manual inspection of a random sample of events. However, we acknowledge its limitations, especially in dealing with Group 4 (the most engaged users in the time-frame prior to an event). These users may not edit at all during the event itself, in which case our inference would fail to capture their participation, thus disregarding them from the analysis of future retention, as well as direct and indirect impact. Unlike Wikipedia, OSM has a very informal organizational structure, with no explicit role differentiation; the only way to distinguish OSM editors is to analyze the effort they vest into the community. To gain further confidence in the conclusions we drew above, we would thus need to complement this quantitative study with a qualitative one. It is also part of our ongoing work to improve the spatial inference processing, so to use space syntax theory to determine the party area, instead of pre-defined ward units.

Implications. Mapping parties are being organized with the specific aim to *map*, *socialize* and *engage new-comers*. How successful are they in attaining these goals? In this paper, we have proposed a methodology to quantify the impact that mapping parties have on contributors of OSM in London. Our findings suggest that these goals are only partly achieved: in particular, mapping parties do cause an increased editing activity during the events themselves; they also sustain engagement over time, though mostly for already active contributors; however, they largely fail on their third goal of engaging new-comers. After just a week following the party, these users stop contributing to OSM and do not come back to other mapping parties again. We do not know the reasons behind this, though we may expect new-comers to have very different needs and motivations than experienced users: the latter may be called by an intrinsic desire to exhaustively map an area; the former may be attracted by mapping tasks they can relate to, either because they target an area of relevance to them (e.g., where they live or work), or because of the focus of the mapping itself (e.g., POIs of a specific category, like motorbike parking spots, public benches, etc.). More focused interest and local groups can make integration easier for inexperienced users with specific geographic interests. Beginner-friendly mapping tools and emphasizing the 'fun' aspect of mapping as suggested in (Nov 2007) would also benefit the socialization of new-comers.

We believe that communities like OSM, which completely rely on volunteered contributions, must be able to measure how successful their range of activities is, both in attracting new users and retaining old ones. The methodology we have proposed in this paper goes one step in this direction, offering a way to quantitatively monitor the impact that these events have on the long term sustainability of the community.

Acknowledgements

The research leading to these results has received funding from the European Community (FP7-SST-2008-RTD-1) under Grant Agreement n. 234239.

References

Beenen, G.; Ling, K.; Wang, X.; Chang, K.; Frankowski, D.; Resnick, P.; and Kraut, R. E. 2004. Using Social Psychology to Motivate Contributions to Online Communities. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, 212–221. New York, NY, USA: ACM.

Bryant, S. L.; Forte, A.; and Bruckman, A. 2005. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, GROUP '05, 1–10.

Budhathoki, N., and Haythornthwaite, C. 2012. Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *American Behavioral Scientist*.

Coast, S. 2011. How OpenStreetMap is changing the world. In *Proceedings of the 10th international conference on Web and wireless geographical information systems*, W2GIS'11, 4–4. Berlin, Heidelberg: Springer-Verlag.

Cosley, D.; Frankowski, D.; Terveen, L.; and Riedl, J. 2006. Using Intelligent Task Routing and Contribution Review to

Help Communities Build Artifacts of Lasting Value. In *In Proc. CHI. 2006*, 1037–1046.

Ducheneaut, N. 2005. Socialization in an Open Source Software Community: A Socio-Technical Analysis. *Computer Supported Cooperative Work (CSCW)* 14(4):323–368.

GISPro. 2007. The GISPro Interview with OpenStreetMap Founder Steve Coast. In *GIS Professional*, number 18, 20–23.

Gomez Rodriguez, M., and Rogati, M. 2012. Bridging Offline and Online Social Graph Dynamics. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, CIKM '12, 2447–2450. New York, NY, USA: ACM.

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

Harper, F. M.; Li, S. X.; Chen, Y.; and Konstan, J. A. 2007. Social Comparisons to Motivate Contributions to an Online Community. In *Proceedings of the 2nd international conference on Persuasive technology*, PERSUASIVE'07, 148–159. Berlin, Heidelberg: Springer-Verlag.

Kraut, R., and Resnick, P. 2012. *Building Successful Online Communities: Evidence-Based Social Design*. The MIT Press. 42–43.

Lampe, C.; Wash, R.; Velasquez, A.; and Ozkaya, E. 2010. Motivations to Participate in Online Communities. In *Proc. CHI*, volume 10, 1927–1936.

MacKinlay, A. C. 1997. Event Studies in Economics and Finance. *Journal of Economic Literature* 35(1):13–39.

Mockus, A.; Fielding, R.; and Herbsleb, J. 2002. Two Case Studies of Open Source Software Development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 11(3):309–346.

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

Panciera, K.; Priedhorsky, R.; Erickson, T.; and Terveen, L. 2010. Lurking? Cyclopaths? A Quantitative Lifestyle Analysis of User Behaviour in a Geowiki. In *Proceedings of CHI 2010: 28th ACM Conference on Human Factors in Computing Systems*, 1917–1926.

Panciera, K.; Halfaker, A.; and Terveen, L. 2009. Wikipedians Are Born, Not Made: A Study of Power Editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*, GROUP '09, 51–60. New York, NY, USA: ACM.

Panciera, K.; Masli, M.; and Terveen, L. 2011. "How should I go from ___ to ___ without getting killed?:" Motivation and Benefits in Open Collaboration. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, 183–192. New York, NY, USA: ACM.

Perkins, C., and Dodge, M. 2008. The Potential of User-Generated Cartography: A Case Study of the OpenStreetMap project and Mapchester. *North West Geography* 8:19–32.

Priedhorsky, R.; Masli, M.; and Terveen, L. 2010. Eliciting and focusing geographic volunteer work. In *Proceedings of the International Conference on Supporting Group Work*. ACM.