

Article

Contributors' Withdrawal from Online Collaborative Communities: The Case of OpenStreetMap

Daniel Bégin ^{1,*}, Rodolphe Devillers ^{1,2} and Stéphane Roche ²

¹ Department of Geography, Memorial University of Newfoundland, St. John's, NL, A1B 3X9, Canada; rdeville@mun.ca

² Centre de Recherche en Géomatique, Université Laval, Québec, QC, G1V 0A6, Canada; stephane.roche@scg.ulaval.ca

* Correspondence: d.begin@mun.ca; Tel.: +1-709-864-8412

Received: 1 September 2017; Accepted: 2 November 2017; Published: 4 November 2017

Abstract: Online collaborative communities are now ubiquitous. Identifying the nature of the events that drive contributors to withdraw from a project is of prime importance to ensure the sustainability of those communities. Previous studies used ad hoc criteria to identify withdrawn contributors, preventing comparisons between results and introducing interpretation biases. This paper compares different methods to identify withdrawn contributors, proposing a probabilistic approach. Withdrawals from the OpenStreetMap (OSM) community are investigated using time series and survival analyses. Survival analysis revealed that participants' withdrawal pattern compares with the life cycles studied in reliability engineering. For OSM contributors, this life cycle would translate into three phases: "evaluation," "engagement" and "detachment." Time series analysis, when compared with the different events that may have affected the motivation of OSM participants over time, showed that an internal conflict about a license change was related to largest bursts of withdrawals in the history of the OSM project. This paper not only illustrates a formal approach to assess withdrawals from online communities, but also sheds new light on contributors' behavior, their life cycle, and events that may affect the length of their participation in such project.

Keywords: Chebyshev's inequality; circadian cycle; time series analysis; survival analysis; life cycle; OSM history; contributors' behavior

1. Introduction

With the advent of the Web 2.0, large communities have developed around online collaborative projects that allow people to contribute data. Examples include platforms that allow sharing of in situ observations (e.g., the Audubon Society for birdwatching), identification of features from images (e.g., Zooniverse), the sharing of general (e.g., Wikipedia) and technical knowledge (e.g., PostgreSQL), and the mapping of people's neighborhoods (e.g., OpenStreetMap). Every day, millions of people visit websites from online communities like Wikipedia.org or OpenStreetMap.org [1]. Researchers are increasingly referring to these communities as a valuable work force and important source of data [2,3].

These successful communities may have hundreds of thousands of active contributors, but all do not contribute in the same way. Among those who contribute, a majority of them will only participate once [4,5], leaving most transactions to a small group of dedicated contributors [6,7]. Even if the proportions may slightly change between communities [5], this typical participation model is referred to as the 90–9–1 rule [6], stating that 90% of the members of a given online community will not contribute anything, 9% will contribute sporadically, and the remaining 1% will be dedicated contributors. In this context, the withdrawal of participants who maintained their participation beyond an initial period of engagement is a significant loss for a community [8].

Studies have looked at the life cycle of online contributors [5,9–12], but the results can be hard to compare. The use of ad hoc criteria to identify withdrawn contributors prevents comparisons between studies, in addition to introducing biases and interpretation errors. Most collaborative online projects have no formal mechanism to determine who withdrew from the project. Since participants freely decide when they contribute, based on their spare time, it is then difficult to distinguish between a participant who left a project from one who is waiting for some free time to contribute again.

Assessing withdrawals from online projects and identifying the nature of the events that drive contributors to leave a community is thus of prime importance. Such knowledge is required to monitor the health of an online community and to minimize contributor withdrawal, particularly when changes are to be made to the participatory environment.

In order to analyze this phenomenon, about 10 years of withdrawals from the OpenStreetMap (OSM) community were investigated. Different statistical approaches were explored to model participants' behavior based on the history of their daily contributions. Using the history of daily contributions required first eliminating potential biases caused by the location of contributors. A probabilistic procedure was then developed to identify the contributors who left the project according to their historical behavior. The resulting daily count of withdrawals was analyzed using both survival and time series analyses.

Survival analysis was used to model the proportion of OSM participants who were still considered active in the project after a given period of time (i.e., survival curve). The resulting model was also used to generate the hazard curve of OSM participants, Hazard curves are often used to characterize life cycles of different domains, such as demography or reliability engineering, and may provide similar insight about OSM contributors.

Time series analysis was used to decompose daily withdrawals in their different components (i.e., trend, seasonal and random). Once decomposed, significant variations of resulting components were compared with the different events that dotted the OSM history to identify which ones may have affected the motivation of OSM participants over time [13].

This paper describes the distribution functions used to characterize the frequency of contributions from participants and discusses the results. The origin of the bias induced when using UTC timestamps to determine the dates of the contributions is explained, and the method used to correct the dates is described. The life expectancy and the survival rates of OSM contributors are presented with the results of a time series analysis. Finally, the paper reports on the events in the OSM project that correlated with large numbers of withdrawals from the community over years.

2. Materials and Methods

The OpenStreetMap project was chosen because the project's history is well documented and the data are freely available. The OSM project aims to create a comprehensive map of the world built on the interests and the local knowledge of its community [14–16]. The project uses a Wiki approach to enable its community to create and improve the map. With currently more than 3 million registered users [17], it has become one of the most successful peer-production projects of the Web and is the largest mapping project in the world. The chronicle of the project's history (e.g., technical improvements, normative changes, social activities) is maintained in the project's wiki documentation [18] and a record of all the contributions is made available on a regular basis through OSM history dump files [19]. These files contain all transactions made since the first contribution and include the virtual containers (i.e., changesets) in which the edits were provided. These changesets identify the contributors who submitted changes, the temporal extent of each editing session, and a minimum bounding rectangle covering all the features edited during the session.

2.1. Data Retrieval

As part of a larger project, a history dump file released on 1 September 2014, was downloaded from the OSM web site to access the records of contributions made to the project since 9 April 2005

(i.e., the first edits). FME workbenches (Safe software 2015.0) were developed to extract and load the data contained in the history dump file to a PostgreSQL (9.3) database. The resulting 2 TB database included 25 M changesets that were used in this study. Statistical analyses and visualizations presented in this paper were carried out using R software (v.3.2.1).

The frequency of contributions (i.e., the number of continuous time intervals an individual has invested in the project) cannot be determined from the number of changesets a contributor provided. The number of changesets and the time span of each of these changesets largely depend on the OSM application interface (API) and the mapping application used by the contributor. First, the OSM API applies constraints regarding the time over which a changeset has been opened by automatically closing them either after being inactive for one hour, or after being active for 24 h. Second, OSM mapping applications have different schema for creating changesets. The same editing session may then produce various numbers of changesets, according to the application used and its configuration. However, the changesets' creation timestamps were exploited to identify on which days a contributor was active.

In order to link potential bursts of withdrawals from the community with events from the project's history, a comprehensive event repository was built by retrieving the entire history of the project from OSM Wiki pages [18] and some OSM mailing lists [20] (i.e., talk, dev and legal mailing lists). The period covered by the repository matched the time span of the history dump file. The events were classified according to an adapted version of the Wiki page's nomenclature and OSM event classification [21] to include development milestones, media news and internal announcements (i.e., blogs and mailing lists).

2.2. Assessing the Frequency of Contributions

The frequency of contributions of each participant has been derived from the UTC timestamps of their changesets. UTC timestamps cannot be used directly to extract the dates of contributions as it could introduce a bias due to the contributor's geographic location and the local time at which the contributions were usually made. The number of distinct dates extracted from the changesets can double when the local time at which the contributions are made falls around midnight GMT. In order to circumvent the problem, we needed to aggregate individuals' contributions in 24-h units that would not be affected by this temporal reference. Two approaches were compared to define a daily contribution timeframe for each individual, the first one based on the proximity of contributions, the other based on contributors' circadian behavior.

The first approach aimed at aggregating contributions by using hierarchical clustering on the time interval (i.e., distance) between changesets. The approach was based on the fact that when the participants have some free time to contribute, the changesets generated during their editing sessions will form clusters in time as demonstrated by Halfaker [22] for different online communities. The closer the changesets, the higher the odds the edits were made on the same editing session and consequently on the same day (from contributors' point of view). For each contributor, clusters of changesets were formed by iteratively grouping the nearest changesets using the nearest-neighbor chain algorithm [23]. The algorithm was chosen because of its relative simplicity to implement as a recursive function in PostgreSQL. When a cluster was about to extend over more than 24 h, it was removed from the process and considered as a one-day contribution. After all the contribution clusters were removed (i.e., any new cluster would span over 24 h), the inter-cluster times were rounded to one-day units to obtain the number of days spent by a contributor between each contribution.

The second approach aimed at identifying the circadian cycle of each contributor in order to apply an offset to the UTC timestamps and consequently to adjust the date of contributions. The circadian cycle partition of a contributor was defined as the time (UTC) at which a contributor was usually inactive (i.e., potentially asleep) according to the history of its contributions. The UTC offset was computed by averaging hours over the longest contiguous interval of time for which the number of contributions was at its minimum. The number of contributions was counted over 24 one-hour bins

(0 h–23 h). Corresponding bins were duplicated over four hours on each side (−4 h, −3 h, ..., 26 h, 27 h) to smooth contributions' count with a nine-hour moving average window. Once a UTC offset was obtained for each contributor, it was applied to their changesets' UTC timestamps prior to extract the distinct dates of their contributions (i.e., active days). Changesets' creation timestamps were used since only participants can trigger them while closing timestamps could result from an API operation.

Both approaches were compared and assessed using a subset of about fifty contributors at both ends of the activity spectrum. The subsets covered both new (active days < 10) and accomplished (active days > 1000) contributors. The approach that provided the most reasonable estimate of contributors' active days for both subsets was used to identify the number of contributions (active days) and the number of days between these contributions. Since a reasonable estimate had to be compatible with human behavior, the time spent by participants contributing on each active day was measured for each method. The higher the number of days an outstanding time was spent contributing (i.e., 12–24 h), the less the method was considered compatible.

2.3. Identifying Withdrawn Contributors

Due to the irregular nature of contributions made by volunteers on online communities, it can be hard to discriminate participants who are waiting for time to contribute again from others who simply withdrew from a project. Results from the analysis described above were used to model the frequency of contributions and identify a time threshold after which an inactive contributor should be considered as being withdrawn from the project with say a 95% probability. Three models were used to identify such threshold. The first two used a global approach based on the contributions from all the participants while the last one considered the history of contributions of individual participants.

First, the potential theoretical distribution of delays was identified based on kurtosis and skewness methods. The 'descdist' procedure (from R's fitdistrplus package) was used to identify the distribution using a 'Cullen and Frey' graph for discrete values [24,25] with 100 bootstrap samples. The proposed distribution was examined to model the delays and identify withdrawn contributors.

Second, the 95th percentile of delays between each sequential contribution was computed and plotted on a log-log graph, providing threshold values that can be used to identify withdrawn contributors. The graph was assessed on both new and accomplished contributors.

Third, since the history of contributions of each individual is available, we used the Chebyshev inequality described in Equation (1) to assess the contributions of each participant and set individuals' threshold:

$$P(|X - u| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}. \quad (1)$$

On the left side of the inequality, P is the probability that the interval of time since the participant's last contribution (X) is larger or equal to a given value (ϵ) when compared to the average interval (u) between its contributions. The right side of the inequality shows that this probability is less or equal to the ratio of the variance of the intervals between contributions (σ^2) over the square of the value provided on the left side of the equation (ϵ^2).

Chebyshev's inequality was chosen because it can be applied to any arbitrary distribution, something expected in our context. However, Equation (1) determines the probability for both sides of the distribution while we are only interested in the upper bound (i.e., the maximum delay expected from a given contributor). Furthermore, the equation requires the population's mean and variance while we consider having only a sample of the delays a contributor will experience during its lifespan in the project, unless the contributor has already left the community. Consequently, we used a version of the one-sided Chebyshev inequality adapted to samples [26], as described by Equation (2):

$$P(X_n - \bar{X} \geq \epsilon s) \leq \frac{1}{1 + \frac{n}{n-1}\epsilon^2}. \quad (2)$$

In order to determine that a participant has withdrawn from a project with a given probability (P), the time since its last contribution (X_n) must differ by at least a given threshold (ϵ) from the average delays (\bar{X}) experienced by the participant. This probability is smaller than or equal to the right side of the inequality, which takes into account the size of the sample, where (n) is the number of delays, (s) is the standard deviation of the delays, and (ϵ) is a constant specific to each participant. The constant is obtained from Equation (3):

$$\epsilon \leq \sqrt{\frac{1 - P}{P \left(\frac{n}{n-1}\right)}}. \quad (3)$$

Equations (2) and (3) were used to determine individuals' thresholds for the time interval since their last contribution. The contributors were considered withdrawn with a 95% probability (P) when the interval between the creation of the history dump and their last contribution reached this threshold. In cases where the participants did not have enough contributions to compute delays' standard deviation (i.e., fewer than three contributions), we used the average threshold of people having made three contributions.

Finally, the subsets of participants from both ends of the activity spectrum were used again to assess the most appropriate method to identify withdrawn contributor from the distribution identified by the Cullen and Frey graph, the 95th percentile of delays, and the sample version of the one-sided Chebyshev inequality. The method was selected by comparing the proportions of contributions that happened outside the threshold established by each method using the history of contributions from our subset of participants. The nearer the proportion is to 5%, the more adequate the method.

2.4. Survival Analysis

Survival analysis provides a set of methods that allow for modeling the probability that an event occurred (e.g., death, withdrawal) over a given period of time. The methods deal with two types of observations, those for which the observed event occurred, and those for which the event did not occur during the period under consideration. In cases the event did not occur within this period, the observations must be censored. Censored data (i.e., a type of missing data) are observations for which the information was measured accurately within the studied period but for which we only know that the survival span was longer than the observed period. The survival analysis is preferred to standard regression models because it adequately handles censored observations, avoiding potential bias in such analysis.

A survival analysis [27,28] was run using the R 'survival' package to measure the probability that an OSM contributor would still be active after a given time in the project. We estimated and plotted survival curves using a non-parametric estimator of the survival function (i.e., the Kaplan–Meier method). The contributors not considered as withdrawn at the end of the period covered by our study (1 September 2014) were identified as censored observations.

Kaplan–Meier estimators were computed for the entire OSM population, and then for years at which participants first contributed (i.e., strata computation). Using the resulting survival curves, we computed and plotted the instantaneous rate of withdrawal over time, also known as the hazard function. This function provides the proportion of active contributors that are expected to withdraw from the project at a given point in time. It illustrates at which points in the life cycle of contributors the odds they withdraw from the project are higher, stable, or lower. Since the results vary on a daily basis, they were filtered using a moving average on a 30-day window.

2.5. Time Series Analysis

A time series analysis assumes the data result from a stochastic process, dividing the process into a deterministic trend, seasonal and centered random components [29,30]. The daily counts of withdrawn contributors were considered as resulting from such a stochastic process. Variations in the different components can show changes in the interest of the participants to contribute to the

project. However, one must consider the volume of new contributors in interpreting any variations because withdrawals depend on them, particularly since most participants contribute for only a very short period of time [4,5]. Consequently, a time series of both withdrawn and new contributors were computed.

The time series were divided into their components using the R package ‘decompose’ procedure [31]. The procedure first determines the trend component by using a moving average on observed data and removes it from the time series. The window used in this process is determined by the cyclical variations expected in the data (i.e., seasonal). The length of the seasonal variations was set to a year, resulting in 182 days without value on each side of the trends components. The seasonal variations were then computed by averaging resulting observations for each of the 365 time units and the results duplicated over the whole range of observations. Finally, the centered random component is what remains after having removed both the trend and the seasonal values from observed data. An additive decomposition was chosen over a multiplicative one to limit the influence of early years of the project in the analysis. Given the small number of participants at that time, any change represented a large proportion of the population using a multiplicative decomposition, which in turn would have had a large impact on the resulting seasonal and random components later in time [13].

Variations in withdrawals and the number of new contributors were compared for each component. Outstanding variations in withdrawal components that were not correlated with variations from new contributors were identified and linked to potential explanatory events found in our inventory. The number of participants who withdrew from the project was estimated by adding positive random component values over 21 days surrounding each event.

3. Results

We identified 464,858 distinct contributors from the 25.1 M changesets found in an OSM history dump retrieved on 1 September 2014. The dump spanned a period of 3433 days (almost 10 years), from first to last registered contributions. The 8381 changesets created by anonymous users were not used in the analyses. This option to remain anonymous was removed for new contributors in fall 2007 and for all participants with the advent of API 0.6 in spring 2009. Furthermore, 400–450 contributors who declined the CT/ODbL license implemented in 2012 [32,33] were not considered either since their data were removed from the database and their contributions did not appear in the dump.

Over 3570 events related to the history of the OSM project were retrieved from the OSM Wiki and from forums’ threads, covering the project’s history from 2005 to 2014. Events were classified into seven categories (Table 1).

Table 1. Classification of events related to the OSM project (2005–2014).

Category	Category Description	Number
Meeting	Administrative, development and social activities.	1350
Upgrade	Infrastructure and software upgrade implementation.	135
Forum	Mailing lists announcements and OSM Foundation blog.	52
License	Contributor terms and OdbL ¹ license change milestones.	8
Mapping	Mapping parties/efforts, including humanitarian activities.	725
Conference	Conferences mentioning/discussing the OSM project.	369
Media	Media coverage about OSM or related topics.	939

¹ OSM switched to an Open Database License (ODbL) after a lengthy process that lasted almost four years.

3.1. Assessing the Frequency of Contributions in Days

Results from the nearest-neighbor chain algorithm estimated to 4.52 M the number of days OSM participants contributed, with an average of 9.72 days per contributor, and up to 2373 days for the most active ones. Results from the circadian cycle algorithm estimated to 5.03 M the number of days

OSM contributors were active, with an average of 10.83 days per contributor, and a maximum of 2465 days for one of the contributors.

The comparison of both approaches shows that the nearest-neighbor chain algorithm generated five times more occurrences of contribution spans longer than 12 h for a day (50,579 days) than the circadian cycle (10,875 days). This was further analyzed by comparing activities over long contribution span clusters with the UTC offsets of their contributors. The result shown that the changesets grouped under long span clusters were usually split by a period of inactivity around contributors' UTC offsets (i.e., contributors' middle of the night). Using our subset of new and accomplished contributors, we found the average daily contribution span was 58% longer for the nearest-neighbor chain algorithm in the first group and 44% longer for the second group. Similarly, the longest daily contribution span was of 24 h for the nearest-neighbor chain algorithm and of 20 h for the circadian cycle algorithm. The circadian cycle algorithm then provided results that were more compatible with expected human behaviors for both new and accomplished participants. Consequently, the circadian cycle algorithm was used to identify contributors' active days and then compute the time they waited between two consecutive active days (i.e., contributors' delays).

3.2. Identifying Withdrawn Contributors

The first approach used the skewness and kurtosis of contributors' delays (i.e., the Cullen and Frey graph) to suggest potential models of distributions for the delays and identify withdrawal thresholds (Figure 1).

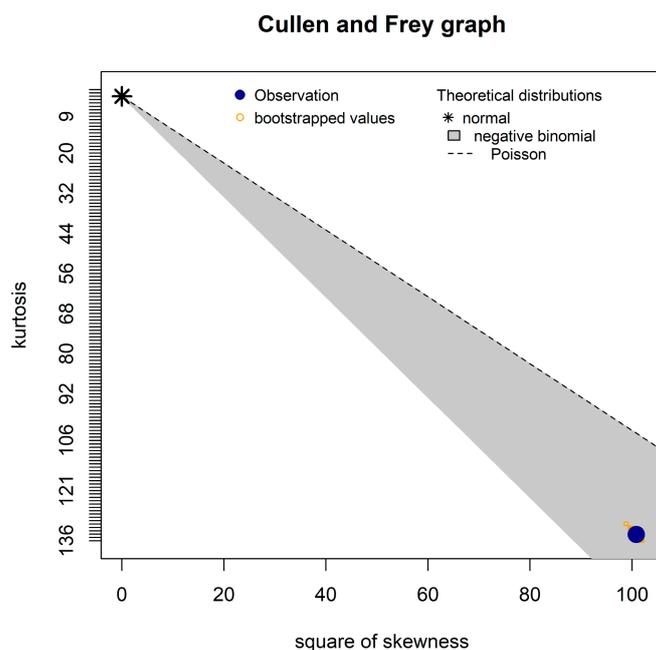


Figure 1. Cullen and Frey graph of delays between contributions of OSM participants with 100 bootstrap samples.

Results suggested a negative binomial distribution (Figure 1). A negative binomial distribution is the distribution of a random variable that gives the expected number of trials required prior a given number of successes (r) to happen (for instance, obtaining a given result twice when throwing dice). Since in our case the number of trials, failures, and successes are integers (days), and we are waiting for a next contribution to happen ($r = 1$), the data would have a geometric distribution (i.e., a special case of the negative binomial distribution), as long as the probability remains the same over all trials. In other words, contributing on a given day could be seen as the successful result of a dice game, in which all OSM participants would use the same dice.

In the case of a geometric distribution, the probability of being successful (i.e., to contribute on a given day) is inversely related to the average number of trials required, which in our case is the average delay between contributions (in days). Using the 4.57 M delays experienced by those who contributed at least twice to the OSM project, we found that on average, an OSM contributor waited 19.51 days between two consecutive contributions, with the longest delay being of 3118 days (i.e., over 8.5 years).

Using the dice game analogy, OSM participants did not use the same dice since they show a broad spectrum of frequency of contributions. Furthermore, assuming that each participant would keep playing the same game with the same number of dice all over their life span in a project is not realistic. Consequently, identifying withdrawn contributors from the above statistical model was not considered realistic either.

The second approach used the 95th percentile of the delays between each sequential contribution illustrated here in a log-log plot (Figure 2).

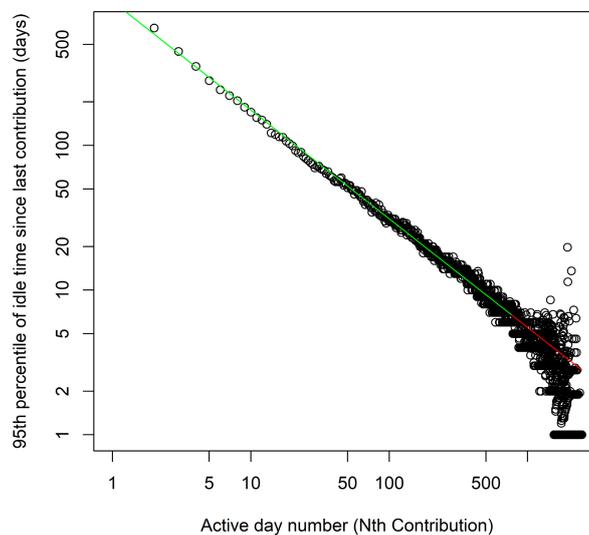


Figure 2. The 95th percentile of delays (days) between a Nth contribution and the previous one. An exponential model of the distribution covering 99.9% of contributors (i.e., a subset) is drawn on the log-log graph (green line). The model was extrapolated for the remaining 0.1% of contributors (red line) where delays were diverging.

The curve shows that new participants may take years before contributing again since at least 5% of them waited more than a year between one of their first four active days. It also shows that, as the number of active days gets higher, the delays between contributions become smaller. An exponential decay model was built by fitting a linear equation on the log transform of both the percentiles and active day numbers to characterize the behavior of 99.9% of contributors (green line). We chose to exclude from the model the percentiles derived from the remaining 0.1% of contributors since their values started to disperse unevenly after about 765 active days. These values were affecting the adjustment of the model with 69% of available measurements representing only 0.1% of contributors. The resulting equation is shown below:

$$P_{95} = e^{-0.75 \log(N) + 6.898}, \quad (4)$$

where P_{95} is the number of days after which 95% of participants will have contributed again after a previous active day, and N is the current contribution (active day). The resulting model coefficients ($p < 0.001$) produced an adjusted R-squared of 0.986 (green line). The model was extrapolated to cover the remaining contributions (red line). However, we found that the graph tends to underestimate

actual delays experienced by individual participants. For new participants, 26% experienced a delay longer than the 95th percentiles defined in above Equation (4), while we were expecting around 5%. For accomplished contributors, this proportion rises to 74%. Since the 95th percentiles were determined from the delays of all participants (which count a few bots), those who kept contributing for a larger number of days pulled the model to shorter delays as the frequencies of their contributions were higher (as defined by the model). Interestingly, the fact that the more the participants have contributed, the less time they wait until their next contribution may suggest behavior that is typical of an addictive process [34–36].

The Chebyshev inequality determined the time threshold after which a contributor should be considered as being withdrawn with a 95% probability. Since Chebyshev’s inequality requires at least two observations to compute a threshold, participants having fewer than three contributions had their thresholds set to 598 days, the average threshold value of participants having three contributions. The resulting thresholds were compared to the time actually spent by the participants between each contribution. We found that 7% of new contributors experienced at least one delay longer than the estimated threshold, and 3.8% of accomplished contributors could have been identified as being withdrawn from the project more often than expected (i.e., 5% of the delays). These results are consistent with the proportion expected from the analysis and were considered appropriate to run the remaining analyses.

The Chebyshev inequality built on individuals’ history has provided a better estimate of the thresholds than those obtained from statistics using the whole OSM population. Individuals’ thresholds obtained from Chebyshev’s inequality were then compared to the time lapse between contributors’ last participation and 1 September 2014. Participants for which the time lapse was longer than their individual thresholds were considered withdrawn from the project.

3.3. Survival Analysis

The Kaplan-Meier estimator used to model survival rates of participants in the OSM project reveals variations in withdrawals of participants over years (Table 2).

Table 2. Withdrawals per year of first contribution. For each year, “Joined” is the number of people who made a first edit in that year, “Quit” is the number of concerned people who withdrew from the project over years, “Rate” is the resulting proportion of contributors who withdrew over years, and “Median” is the number of days over which at least 50% of participants contributed to the project.

Year	Joined	Quit	Rate	Median
2005	83	41	49%	3143
2006	432	218	50%	2733
2007	4820	3240	67%	1036
2008	26,545	20,409	77%	111
2009	61,566	52,044	85%	1
2010	58,547	49,698	85%	1
2011	65,516	55,917	85%	1
2012	87,582	73,833	84%	1
2013	86,319	9278	11%	NA *
2014	73,447	4220	6%	NA *
All	464,857	268,898	58%	28

* Participants who made a first contribution after January 2013 should not be considered since the majority of them were assigned a threshold of 598 days as they contributed fewer than three times. Consequently, their thresholds were not reached yet at the time the history dump was created.

Table 2 shows that half of participants who enrolled during the 2005–2007 period were still active in September 2014, while 85% of those who enrolled after 2009 withdrew from the project prior to that date. Similar turning points in participants’ behavior were found in OSM’s enrollment history [13] and were linked to early stages of the Diffusion of Innovation theory [37]. After 2009; half

of withdrawn participants contributed only once, as shown by the median values. Combining all the above participants, the analysis produced a survival curve that is shown in Figure 3.

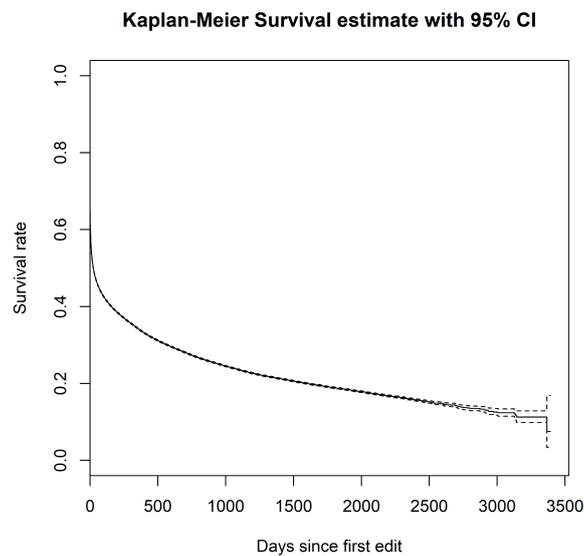


Figure 3. Survival curve of OSM contributors with 95% confidence intervals.

The model estimated that 64% of OSM participants “survived” their first active day, while 11% would have been active after almost 10 years (3335 days). After a steep drop of the survival rate, the slope rapidly decreases to eventually become constant. This characteristic is more easily understood from the hazard function that assesses the rate of withdrawal of participants who keep contributing to the project. The plot of the hazard function is presented in Figure 4.

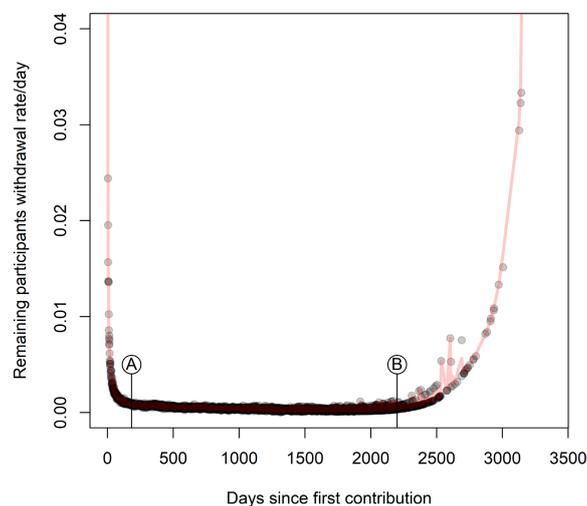


Figure 4. Hazard function of OSM participants, where dark dots are the proportion of remaining participants who withdrew at a given time and the red line is a moving average of the data. The first and last points of the distribution are not shown. Tags A and B delimit a segment of the curve where withdrawal rates are low and almost constant.

The curve shows a bathtub profile familiar to reliability engineering and system safety domains [38]. These curves are used to characterize the rate of failure of different systems or manufactured objects and are used to split life cycles into three stages. The first stage is called “early failures” and shows an initial steep drop in the failure rates, where weaker components rapidly

fail after an item is put into service. The next stage is referred to as the “useful life” of equipment, where failure rates are low and relatively constant and result from random events. The last one is called the “wear out” stage, in which cumulative damages eventually trigger cascade failures of the components.

When using similar definitions with OSM (Figure 4), one can observe that the early defect rates are high with 36% of withdrawals happening on the first day (not shown on the graph). The daily rates then drop rapidly to stabilize around 0.1% after six months. By this time, about 60% of contributors will have left the project. The second stage, delimited by tags A and B (Figure 4), shows stabilized daily rates. These rates slightly decrease over time to reach a minimum of 0.023% (i.e., 8% on an annual basis) after 1670 active days. The rates then increase to reach 0.04% after six years (2192 days). By this time, about 80% of contributors will have left the project. The last stage sees the rates of withdrawal increasing exponentially to reach 33% (not shown on the graph). This rate results from the withdrawal of one of the three oldest participants who quit the project after having contributed over 3367 days. This last stage concerns early OSM contributors since the span of the history dump used in this research was 3432 days and the longest individual span was 3381 days.

3.4. Time Series Analysis

The data used in the analysis were a continuous sequence of discrete time-ordered number of withdrawals from the OSM project, as identified previously. A first analysis was run on all OSM participants who withdrew from OSM. The variations in the number of both withdrawals and new contributors proved to be highly correlated (Spearman’s rank correlation $\rho = 0.721$ $p < 0.001$), which means that the events that triggered a large volume of new contributors did the same for withdrawals since 36% of these new contributors withdrew on the same day. In order to reduce this correlation, the same analysis was run with participants who contributed more than once to the project. The resulting analysis presented an outstanding peak of withdrawals in mid-2011, which was not visible on results from all participants. The height of the peak affected the computation of seasonal and the random components. To remove the effect from the seasonal component, observed values were replaced by trend values over the event interval. A second analysis was run and the peak was added back on observed and random components. Figure 5 presents the time series of new contributors and the adjusted time series of the withdrawn contributors.

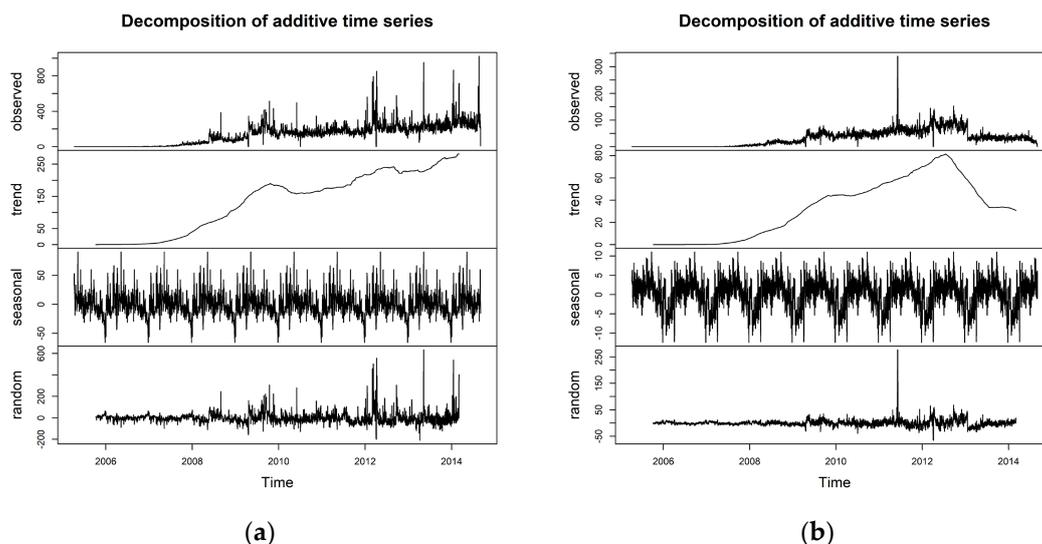


Figure 5. Compared time series analysis plots for participants who contributed more than once where (a) shows the time series for new contributors and (b) shows the time series for withdrawn contributors with seasonal and random components adjusted for the peak event. Both graphs show the observed values, trend, seasonal, and random components that indicate the estimated number of contributors.

As expected, seasonal and trend variations look similar on both graphs, although the trend of withdrawals (Figure 5b) should not be considered after it started declining in mid-2012. This decline resulted from participants who began contributing after this date and for whom the probability of withdrawal had not yet reached 95% when the history dump file was created. Random variations show numerous peaks on both distributions. These peaks identify days when unusual volumes of participants (i.e., small or large) first contributed or withdrew from the project. These unusual volumes of withdrawals were manually identified on the graph, and potential explanations were searched from the event inventory. Outstanding variations of withdrawals that were synchronized with variations of the number of new contributors were excluded from our selection. These included all negative peaks of withdrawals since they were all related to OSM database downtime and the events that potentially brought a burst of new participants as identified by the literature [13]. The remaining outstanding withdrawal events are identified in Figure 6.

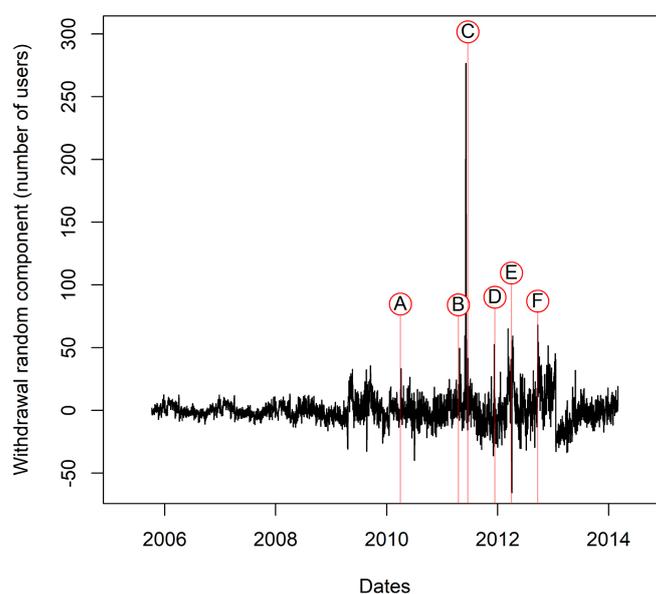


Figure 6. Random components of withdrawals from the OSM project and largest outstanding events (A–F). The sharp drop seen after the last event (F) is an artefact of the 598-day threshold assigned to new contributors, and the time at which the history dump file was created.

In addition to the main peak (C), five other peaks were identified in the graph. The potential explanatory events of these peaks are identified in Table 3. The cyclic variations visible at the left of the first event (A) are residual from the seasonal variations (Figure 5a seasonal) and the large withdrawals correlate with bursts of new contributors following large mapping parties after the implementation of API 0.6.

Table 3. Outstanding random variations of withdrawals from OSM with associated explanatory events. ‘Id’ refers to the labels of Figure 6. ‘Quit’ is the estimated number of withdrawn contributors.

Id	Date	Quit	Associated Explanatory Event Description
A	1 April 2010	136	Ordnance Survey began releasing data for free reuse.
B	17 April 2011	255	ODbL: Unsettled users must make their choice in order to contribute.
C	19 June 2011	1117	ODbL: Users who did not agree with the new license were blocked.
D	13 December 2011	111	ODbL: Treads about what data should be removed from the database.
E	1 April 2012	501	ODbL: Planned non-ODbL data removal and Blog. announcements
F	20 September 2012	419	Import guidelines now require dedicated accounts.

Interestingly, the first peak of withdrawals (A) seems related to the origin of the OSM project itself [39,40]. The last peak (F) could be related to participants who have imported or were to import

data to the OSM database. In such a case, the volume of withdrawn contributors should correspond to those who have changed the nature of their activities at this time or before since at the same time the number of new contributors increased without any other explanation according to the event inventory.

The remaining peaks of withdrawals correlate with specific milestones or discussions about the license change. The largest peak (C) happened in the days before the accounts of users who did not agree to the CT/ODbL license were to be deactivated. It is important to recall that the data from these contributors were later removed from the databases and consequently do not appear in our results. These peaks could represent contributors who accepted the new license in order not to see their work removed from the database [41], or subsequently lost their motivation to contribute when the process resulted in a data loss.

4. Discussion

The results obtained from the different analyses and procedures have not only allowed for identifying withdrawn contributors from an online community, but also suggest potential explanations about the origin of collective withdrawals from OSM. Those results have also shed some light on OSM contributors' behavior and life cycle.

4.1. Assessing Withdrawals from an Online Community

According to communities' conventions about withdrawals, if any, contributors may announce their decisions to quit using templates or messages in their personal profiles. However, in order for the decision to be made public, contributors must care about respecting community conventions and their decision must be taken consciously. We suspect this happens mostly on specific circumstances such as health problems, personal obligations or a conflict with the community (e.g., OSM license change), as illustrated in some OSM users' profiles [41]. The vast majority of contributors rather withdraw from a project by simply postponing their next contributions indefinitely because the priority they give to the activity slowly dropped, along with their motivation to contribute. This supports the need to use a statistical approach that depends only on actual contributions made by participants.

The challenge in identifying withdrawn contributors was twofold. First, using statistical models derived from the contributions of a whole population would not have permitted an analysis of individuals' behavior. The use of Chebyshev's inequality to assess the contributions of each participant has proven to provide accurate decisions about individuals' withdrawal. The main drawback of the method is that it took 798 days before confirming a one-time OSM contributor had left the project with 95% certainty, which is much shorter in most of the cases. According to Figure 3, about 75% of contributors have left the project at this time but the status of these one-time contributors cannot be confirmed with a 95% certainty until the threshold is reached. However, the length of this threshold for one-time contributors will vary according to the studied community and the required level of certainty. Second, in order to identify withdrawn participants based on the history of their contribution, one must identify the frequency at which they contributed to the project. We demonstrated that the UTC timestamps used to make such an assessment can lead to very different results depending on contributors' location and the time at which they usually contribute. The resulting frequency of contributions may even double in certain circumstances, something that has to our knowledge not been mentioned in the literature. Such bias could induce interpretation error when assessing contributions based on participants' locations (i.e., country, continent). Determining individuals' circadian cycle based on the UTC timestamps of their contribution proved to be a simple and efficient approach. Identifying the time at which the volume of contributions is at its minimum for each contributor better reflects individuals' natural cycles, even with fewer than 10 contributions, as we found when assessing changesets' clustering using nearest-neighbor algorithm.

4.2. Withdrawals from the OSM Project

Examining the withdrawals from the OSM project over time proved to be more complex than expected considering the relationship between withdrawal and enrollment rates. However, although the origin of long-term variations of withdrawals could not be differentiated from those of enrollment, we were able to identify specific events that correlated with collective withdrawals of participants.

The first outstanding event originated from outside the project when the original *raison d'être* of the project disappeared for many contributors after the British national mapping agency (i.e., the Ordnance Survey) began releasing data for free use. This is a risk any crowdsourcing projects can face, when participants' needs can suddenly be better met through another source. In this case, a new authoritative source of free geographic data has potentially caused some local contributors to leave the project. However, considering the number of withdrawals directly related to this event, the individual needs the OSM project was meeting must have been larger for most participants, as suggested in the literature about the motivations of online participants [41–46].

The main source of withdrawals from the OSM project was related with events that were internal to the project. The license change process and related discussions in OSM forums may have resulted in the withdrawal of about 2000 contributors (Table 3) to which we must add the 400–450 contributors who declined the CT/ODbL license [31,32]. Overall, 1% of OSM contributors left the project during burst of withdrawals that seemed related to this process.

If shared interests, values, and beliefs bring contributors together in a collaborative project like OSM [46,47], it necessarily translates into a collective identity [48] that in turn should result in collective behavior regarding the events that pave the way to the project. The license change may have highlighted differences in the values and beliefs of participants, resulting in the collective withdrawal of people whose values were jostled in the process (Table 3 and Figure 6). The fact that these withdrawals happened over different events simply reflects differences in the collective identity of those people [48].

The last event identified in Table 3 may have shed light on the volume of participants who are concerned by data imports. When a change to the import guidelines required contributors to use dedicated accounts for import and for casual mapping, a large number of users seem to have withdrawn from the project (Table 3). Since this event simultaneously generated an increase in both the number of new and withdrawn contributors, the latest is probably not related to people that left the project, but rather people that considered not having the same type of contribution anymore (i.e., imports or casual mapping) and decided to leave their previous account to adjust to the new guidelines.

The withdrawals from the OSM project may reveal situations where a community is confronted to new challenges that cannot be overcome by all its participants [8]. The challenges online communities face in preventing contributors from withdrawing are twofold. First, changes related to the technical aspects of the participation (e.g., new rules, technical requirements) may trigger withdrawals even when changes can be considered as being positive for the community. This is not necessarily because the learning curve could be too steep, but also because the motivation to adapt from some contributors may not be there anymore (the wear-out stage). Second, interventions and changes that may hurt personal values or beliefs of the participants (e.g., changes in project's objectives, better alternatives, internal conflicts) seem to have triggered large numbers of withdrawals in an otherwise strong and healthy community. In this case alternatives are limited since our results have shown that multiple collective identities can coexist in the same project, where going towards one group means moving away from another one.

4.3. Contributors' Behavior

As shown by Vázquez and Barabási [42,43], people contribute through bursts of rapidly occurring events separated by long periods of inactivity. The main difference between new and accomplished contributors should then be the length of their activity bursts, this length being much longer for the

latter. Figure 2 reveals such long periods of inactivity for new contributors and the long periods of rapidly occurring contributions from accomplished ones.

When participants engage in the project, they seem to assess the project to determine whether they find it relevant, enjoyable, or both [13]. The contributors will consider a project as relevant if it meets their needs, desires, or aspirations, whether because of the project's objectives [44–47] or because of the nature of the tasks [48–50]. They will find a project enjoyable if their participation provide them distraction or even fun [45,46,51]. According to the Self-Determination Theory [52], an important motivation to keep contributing is self-efficacy [50,53]. This is the perception the individuals gain about their capacity to fulfill the required tasks as they contribute. When they are successful, individuals gain a feeling of control, competency, and autonomy that motivates them to keep contributing, while unsuccessful attempts may lead them to lose their motivation and stop contributing.

Figure 4 shows that this phase seems to last up to six months, where the daily rates of withdrawals fall from 35% to 0.1% when they stabilize. During this phase, about 60% of the participants will have withdrawn from the project. We would call this period the “assessment” phase, a period over which participants are estimating the costs and benefits of contributing to the project. During this phase, the knowledge and skills required to contribute geographical information [54–56] can certainly be an obstacle for OSM contributors, which makes the project's learning curve steeper than the average collaborative project. One would expect the rate of withdrawal to be higher with such a project than with other projects such as Wikipedia. However, the literature suggests the contrary, since about 60% of Wikipedia contributors withdraw within the first day [4,12], while a similar rate was found only after six months for OSM. An explanation might be that while learning to contribute, participants are less inclined to withdraw from a project. Such behavior may be seen in communities of practice where legitimate peripheral participation [57,58] is an important learning mechanism in which new participants slowly move from the periphery to the core of an activity. The longer it takes to grasp the nature of an activity, the longer it may take to assess the costs and benefits of engaging in such an activity. Interestingly, a similar assessment phase has been illustrated in another volunteered geographical information (VGI) project where the rates of withdrawals seemed to stabilize after about six months [11] (Figure 5).

If the project meets the needs of the participants, they seem to engage with the project for the long term since daily rates of withdrawal stay low for a period of about six years. Given that such long-term engagement is frequent in collaborative projects [4,12,59,60], we have called this period the “engagement” phase. Over the first half of the period, the daily rates dropped from 0.1% to almost nothing (0.004%) before rising again over the second half to reach 0.04%. Referring to concepts used in reliability engineering, we consider the time at which the rates reached their minimum (i.e., 3.5 years after the first contribution) as a pivotal point where contributors seem to switch from an adaptation-dominated process to a cumulative-damage-dominated process [38]. During the adaptation-dominated process, contributors adapt to the community's norms and rules, learn how to contribute and master available tools, and develop a feeling of self-efficacy. During the cumulative-damage-dominated process, the many events that over years brought irritation or annoyance to the participants start affecting their motivation to keep contributing. It is a period in which contributors may become less inclined to adapt to an evolving project and a never-ending flow of unexperienced contributors. This type of behavior (adaptation–conservatism) has already been mentioned in the literature regarding the vocabulary used by participants in online communities [59].

We called the last period experienced by participants, after having contributed to the project for over six years, the “detachment” phase. Results have shown that the daily rates of withdrawal increase exponentially over this period (Figure 4). However, the analyses also revealed that only half of early contributors (2005–2006) withdrew from the project (Table 2). This special commitment to the project contrasts with withdrawals from later participants, which reached 85% after 2009. According to Budhathoki [61], a large proportion of these early contributors were also project developers or people who had an impact on its development, which could explain the discrepancy.

Another interesting finding made about contributors' behavior is the time they spent between contributions, as the number of their contributions increases (Figure 2). The fact that this pattern of participation is similar to what would be expected from an addictive process should be linked to contributors' motivation. Providing geographic data to a project like OSM is a complex task [54,55], which may increase the pleasure gained by participants from fulfilling the task (learning, self-efficacy, self-actualization, self-expression), contemplating the outcome (fun, instrumentality), or using the result (meeting own need), as described by Budhathoki [51]. The more they contribute and master the process, the more pleasure they derive from it, and the higher priority they will give to the activity during their free time. The latest mechanism has even been used to explain the "bursty" nature of human behavior when engaging in online activities [43]. However, since the number of active days (Figure 2) and the time span of the project are related, some have suggested that new participants may have had fewer opportunities to contribute (lower frequency) than older participants (higher frequency) because of the OSM map saturation [62] in many Western countries [63]. An analysis of the number of participants who contributed frequently (more than once a week) against their years of enrollment revealed that there was no such relationship, the number of recurring contributors being even higher in recent years.

Finally, the rates of withdrawal have shown variations over the years, a phenomenon similar to that identified within OSM enrollment and linked to the early phases of the Diffusion of Innovation theory [13,56]. This might result from a stronger engagement of early participants who developed the project, while the latest participants got involved once the project's infrastructure was mostly set up [37,64].

5. Conclusions

Online collaborative communities have grown in importance, with millions of people visiting or consulting their websites every day. For this reason, assessing withdrawals from online projects and identifying events that drive the contributors to leave a community is of prime importance.

This study compared different methods to identify the contributors who have left a community. All these methods required assessing the frequency of contributions over time but the literature had not yet assessed the biases that could result from assessing this frequency according to participants' location and schedules. We developed a method based on contributors' circadian cycles that proved to be a simple and efficient approach to avoid such biases when using UTC timestamps. Our results show that assessing the withdrawal of individual participants required estimating individual behavior from the history of their own contributions. Accurately identifying withdrawn contributors should have provided reliable results when assessing withdrawals from the OSM community over time. Contrarily to previous studies that relied on ad hoc criteria to identify withdrawn contributors, the use of both the participants' circadian cycles and Chebyshev's inequality provides a transparent and reproducible approach when analyzing and comparing the behavior of contributors within and between online communities.

The different procedures and analyses achieved in this research have not only illustrated an effective approach to assess withdrawals from online communities, but also shed light on contributors' behavior, their life cycle, and the events that may affect the length of their participation in such a project. Our results suggest the origin of withdrawals from an online community is twofold.

First, collective withdrawal can result from changes in the environment that cause participants to question their primary motivation for enrolling in a given community. These changes may lessen the need for the participants to contribute to a project, either because the need does not exist anymore or the need is better fulfilled elsewhere. Internal conflicts seem to be a major threat to the well-being of a community. Such conflicts often result from differences in values and beliefs between the members of a community, and these disagreements may be difficult to resolve collectively. Other changes that are internal to a project may also trigger withdrawals on a smaller scale in the event of a change in the community's norms and rules, contribution tools, or communication interfaces.

Second, contributors' withdrawal has also proven to be determined by three different phases of their life cycle. There is first a short "assessment" phase, when contributors probe the project and determine if they will engage in the long term. A large majority of the participants will withdraw from a project during this phase. A longer "engagement" phase follows, during which withdrawal rates are low and relatively constant. Finally, a "detachment" phase will come when years of wear and tear have exhausted the determination of many remaining participants. However, we were not able to establish a maximum lifespan for OSM contributors since half of those who engaged in the early years of the project were still active.

This research has highlighted very simple mechanisms that can explain most withdrawals from an online collaborative project, from both individual and collective perspectives. Understanding the processes that determine withdrawals from an online community can help with intervening and minimizing their effects. It may then be possible to minimize withdrawals by directing efforts to appropriate phases of the life cycle of the contributors, or to transform the life of a project without generating conflicts, taking into account that all contributors do not have the same sensibilities, values, and beliefs.

Acknowledgments: This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant awarded to Rodolphe Devillers and by Memorial University of Newfoundland.

Author Contributions: Daniel Bégin conceived and performed the experiments, analyzed the data, and wrote the paper. Rodolphe Devillers and Stéphane Roche provided substantial support in structuring and editing the final document.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. SimilarWeb Ltd. Analyze any Web site or App—Home page. Available online: <https://www.similarweb.com/> (accessed on 6 January 2017).
2. Kimura, A.H.; Kinchy, A. Citizen Science: Probing the Virtues and Contexts of Participatory Research. *Engag. Sci. Technol. Soc.* **2016**, *2*, 331–361. [[CrossRef](#)]
3. Michelucci, P.; Dickinson, J.L. The power of crowds. *Science* **2016**, *351*, 32–33. [[CrossRef](#)] [[PubMed](#)]
4. Panciera, K.; Halfaker, A.; Terveen, L. Wikipedians are born, not made: A study of power editors on Wikipedia. In Proceedings of the ACM 2009 International Conference on Supporting Group Work, Sanibel Island, FL, USA, 10–13 May 2009; ACM: New York, NY, USA, 2009; pp. 51–60.
5. 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](#)]
6. Nielsen, J. The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities. Available online: http://www.useit.com/alertbox/participation_inequality.html (accessed on 26 October 2012).
7. Ochoa, X.; Duval, E. Quantitative analysis of user-generated content on the web. In Proceedings of the First International Workshop on Understanding Web Evolution (WebEvolve2008): A prerequisite for Web Science, Beijing, China, 22 April 2008; pp. 1–8.
8. Balestra, M.; Cheshire, C.; Arazy, O.; Nov, O. Investigating the Motivational Paths of Peer Production Newcomers. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; ACM: New York, NY, USA, 2017; pp. 1–5.
9. Ciampaglia, G.L.; Vancheri, A. Empirical Analysis of User Participation in Online Communities: The Case of Wikipedia. In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, Washington, DC, USA, 23–26 May 2010; The AAAI Press: Menlo Park, CA, USA, 2010; pp. 219–222.
10. Ortega, F.; Izquierdo-Cortazar, D. Survival analysis in open development projects. In Proceedings of the 2009 ICSE Workshop on Emerging Trends in Free/Libre/Open Source Software Research and Development, Vancouver, BC, Canada, 18 May 2009; IEEE Computer Society: Washington, DC, USA, 2009; pp. 7–12.
11. Panciera, K.; Priedhorsky, R.; Erickson, T.; Terveen, L. Lurking? cyclopaths?: A quantitative lifecycle analysis of user behavior in a geowiki. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010; ACM: New York, NY, USA, 2010; pp. 1917–1926.

12. Zhang, D.; Prior, K.; Levene, M. How long do Wikipedia editors keep active? In Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration, Linz, Austria, 27–29 August 2012; ACM: New York, NY, USA, 2012; pp. 1–4.
13. Bégin, D.; Devillers, R.; Roche, S. Contributors' Enrollment in Collaborative Online Communities: The Case of OpenStreetMap. *Geo-Spat. Inf. Sci.* **2017**, *19*, 282–295. [[CrossRef](#)]
14. Mooney, P.; Corcoran, P. Who are the contributors to OpenStreetMap and what do they do? In Proceedings of the GIS Research UK 20th Annual Conference, Lancaster, UK, 11–13 April 2012; pp. 355–360.
15. Napolitano, M.; Mooney, P. MVP OSM: A Tool to identify Areas of High Quality Contributor Activity in OpenStreetMap. *Bull. Soc. Cartogr.* **2012**, *45*, 10–18.
16. Bright, J.; De Sabbata, S.; Lee, S. Geodemographic biases in crowdsourced knowledge websites: Do neighbours fill in the blanks? *GeoJournal* **2017**, 1–14. [[CrossRef](#)]
17. OpenStreetMap contributors Stats. Available online: <http://wiki.openstreetmap.org/wiki/Stats> (accessed on 15 January 2013).
18. OpenStreetMap contributors Main Page. Available online: http://wiki.openstreetmap.org/wiki/Main_Page (accessed on 18 May 2017).
19. OpenStreetMap contributors Complete OSM Data History. Available online: <http://planet.openstreetmap.org/planet/full-history/> (accessed on 3 September 2014).
20. OpenStreetMap contributors OSM mailing lists. Available online: http://wiki.openstreetmap.org/wiki/Mailing_lists (accessed on 7 April 2017).
21. OpenStreetMap contributors Events category template. Available online: <http://wiki.openstreetmap.org/wiki/Template:Cal/doc> (accessed on 7 April 2017).
22. Halfaker, A.; Keyes, O.; Kluver, D.; Thebault-Spieker, J.; Nguyen, T.; Shores, K.; Uduwage, A.; Warncke-Wang, M. User session identification based on strong regularities in inter-activity time. In Proceedings of the 24th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, Florence, Italy, 18–22 May 2015; pp. 410–418.
23. Day, W.H.; Edelsbrunner, H. Efficient algorithms for agglomerative hierarchical clustering methods. *J. Classif.* **1984**, *1*, 7–24. [[CrossRef](#)]
24. Cullen, A.C.; Frey, H.C. *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, 1st ed.; Plenum Press: New York, NY, USA, 1999; ISBN 0-306-45957-4.
25. Delignette-Muller, M.L.; Dutang, C.R. Fitdistrplus Package—An R package for fitting distributions. *J. Stat. Softw.* **2015**, *64*, 1–34. [[CrossRef](#)]
26. User: Cardinal Does a sample version of the one-sided Chebyshev inequality exist? Available online: <https://stats.stackexchange.com/a/82694/82725> (accessed on 5 September 2016).
27. Kleinbaum, D.G.; Klein, M. Statistics for Biology and Health. In *Survival Analysis: A Self-Learning Text*, 2nd ed.; Springer Science & Business Media: New York, NY, USA, 2006; ISBN 0-387-23918-9.
28. Therneau, T.M.; Lumley, T.R. *Survival Package—Survival Analysis*; CRAN: Fermanagh, Northern Ireland, 2017; pp. 1–143.
29. McLeod, A.I.; Yu, H.; Mahdi, E. Time Series Analysis: Methods and Applications. In *Time Series Analysis with R*; Rao, C.R., Ed.; Elsevier: Oxford, UK, 2011; Volume 30, pp. 661–707. ISBN 978-0-444-53858-1.
30. Hyndman, R.J.; Athanasopoulos, G. Open access book from OTexts. In *Forecasting: Principles and Practice*, 1st ed.; OTexts: Melbourne, Australia, 2014; ISBN 978-0-9875071-0-5.
31. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2016; Volume 3.2.1, ISBN 3-900051-07-0.
32. Weait, R. OSM License Upgrade—Phase 4 coming soon. Available online: <https://blog.openstreetmap.org/2011/06/14/osm-license-upgrade-phase-4-coming-soon/> (accessed on 8 May 2016).
33. OpenStreetMap administrator ODbL disagreed users Ids. Available online: http://planet.openstreetmap.org/users_agreed/users_disagreed.txt (accessed on 6 July 2017).
34. Rozaire, C.; Landreat, M.G.; Grall-Bronnec, M.; Rocher, B.; Vénisse, J. Qu'est-ce que l'addiction? *Arch. Politique Crim.* **2009**, *31*, 9–23.
35. Vaghefi, I.; Lapointe, L. When too much usage is too much: Exploring the process of it addiction. In Proceedings of the 2014 47th Hawaii International Conference on System Sciences, Hawaii, HI, USA, 6–9 January 2014; IEEE Computer Society: Washington, DC, USA, 2014; pp. 4494–4503.

36. OpenStreetMap contributors OSM purity self-test. Available online: http://wiki.openstreetmap.org/wiki/OSM_purity_self-test (accessed on 7 April 2017).
37. Rogers, E.M. *Diffusion of Innovations*, 3rd ed.; The Free Press: New York, NY, USA, 1983; ISBN 0-02-926650-5.
38. Wang, K.; Hsu, F.; Liu, P. Modeling the bathtub shape hazard rate function in terms of reliability. *Reliab. Eng. Syst. Saf.* **2002**, *75*, 397–406. [[CrossRef](#)]
39. Al-Bakri, M.; Fairbairn, D. User generated content and formal data sources for integrating geospatial data. In Proceedings of the 25th International Cartographic Conference, Paris, France, 3–8 July 2011; International Cartographic Association: Paris, France, 2011; pp. 1–8.
40. Koukoletsos, T. A Framework for Quality Evaluation of VGI Linear Datasets. Ph.D. Thesis, University College London, London, UK, 2012.
41. OpenStreetMap contributors User: TimSC/Quit. Available online: <http://wiki.openstreetmap.org/wiki/User:TimSC/Quit> (accessed on 7 April 2017).
42. Vázquez, A.; Oliveira, J.G.; Dezsö, Z.; Goh, K.; Kondor, I.; Barabási, A. Modeling bursts and heavy tails in human dynamics. *Phys. Rev. E* **2006**, *73*, 1–19. [[CrossRef](#)] [[PubMed](#)]
43. Barabási, A. The origin of bursts and heavy tails in human dynamics. *Nature* **2005**, *435*, 207–211. [[CrossRef](#)] [[PubMed](#)]
44. Chacon, F.; Vecina, M.L.; Davila, M.C. The Three-Stage Model of Volunteers' Duration of Service. *Soc. Behav. Personal.* **2007**, *35*, 627–642. [[CrossRef](#)]
45. Nov, O.; Arazy, O.; Anderson, D. Technology-Mediated Citizen Science Participation: A Motivational Model. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011; The AAAI Press: Menlo Park, CA, USA, 2011; pp. 249–256.
46. Aknouche, L.; Shoan, G. Motivations for Open Source Project Entrance and Continued Participation. Master's Thesis, Lund University, Lund, Sweden, 2013.
47. von Hippel, E.; von Krogh, G. Open source software and the private-collective innovation model: Issues for organization science. *Organ. Sci.* **2003**, *14*, 209–223. [[CrossRef](#)]
48. Houle, B.B.J. A Functional Approach to Volunteerism: Do Volunteer Motives Predict Task Preference? *Basic Appl. Soc. Psychol.* **2005**, *27*, 337–344. [[CrossRef](#)]
49. Borst, W.A.M. *Understanding Crowdsourcing—Effects of Motivation and Rewards on Participation and Performance in Voluntary Online Activities*, 1st ed.; Erasmus University of Rotterdam: Rotterdam, The Netherlands, 2010; ISBN 978-90-5892-262-5.
50. Hemetsberger, A.; Pieters, R. When consumers produce on the internet: The relationship between cognitive-affective, socially-based, and behavioral involvement of prosumers. *J. Soc. Psychol.* **2003**, *2*, 274–291.
51. Budhathoki, N.R.; Nedovic-Budic, Z.; Bruce, B. An interdisciplinary frame for understanding volunteered geographic information. *Geomatica* **2010**, *64*, 11–26.
52. Ryan, R.M.; Deci, E.L. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemp. Educ. Psychol.* **2000**, *25*, 54–67. [[CrossRef](#)] [[PubMed](#)]
53. Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* **1989**, *13*, 319–340. [[CrossRef](#)]
54. DiBiase, D.; DeMers, M.N.; Johnson, A.; Kemp, K.; Luck, A.T.; Plewe, B.; Wentz, E. *Geographic Information Science & Technology—Body Of Knowledge*, 1st ed.; Association of American Geographers: Washington, DC, USA, 2006; ISBN 978-0-89291-267-4.
55. Downs, R.M.; DeSouza, A. *Learning to Think Spatially: GIS as A Support System in the K-12 Curriculum*, 1st ed.; The National Academies Press: Washington, DC, USA, 2006; ISBN 978-0-309-09208-1.
56. Jones, C.E.; Weber, P. Towards Usability Engineering for Online Editors of Volunteered Geographic Information: A Perspective on Learnability. *Trans. GIS* **2012**, *16*, 523–544. [[CrossRef](#)]
57. Lave, J.; Wenger, E. *Situated Learning: Legitimate Peripheral Participation*; Cambridge University Press: Cambridge, UK, 1991.
58. Wenger, E. *Communities of Practice: Learning, Meaning, and Identity*; Cambridge University Press: Cambridge, UK, 1998.
59. Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; Potts, C. No country for old members: User lifecycle and linguistic change in online communities. In Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; ACM: New York, NY, USA, 2013; pp. 307–318.

60. Arazy, O.; Lifshitz-Assaf, H.; Nov, O.; Daxenberger, J.; Balestra, M.; Cheshire, C. On the “how” and “why” of emergent role behaviors in Wikipedia. In Proceedings of the Conference on Computer-Supported Cooperative Work and Social Computing, Portland, OR, USA, 25 February–1 March 2017; pp. 2039–2051.
61. Budhathoki, N.R. Participants’ Motivations to Contribute Geographic Information in an Online Community. Ph.D. Thesis, Graduate College of the University of Illinois, Urbana, IL, USA, 2010.
62. Rehrl, K.; Gröchenig, S. A Framework for Data-Centric Analysis of Mapping Activity in the Context of Volunteered Geographic Information. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 37. [[CrossRef](#)]
63. Neis, P.; Zielstra, D.; Zipf, A. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Futur. Internet* **2013**, *5*, 282–300. [[CrossRef](#)]
64. Shepherd, D.A.; Kuratko, D.F. The death of an innovative project: How grief recovery enhances learning. *Bus. Horiz.* **2009**, *52*, 451–458. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).